Proceedings of the 12th Student Conference in Interaction Technology and Design 2016

Edited by Thomas Mejtoft

Nr 1/2016

Interaction and Media Technology Report Series
Preface

The Student Conference in Interaction Technology and Design is the grand finale of the course *Current Topic in Interaction Technology and Design* at the Department of Applied Physics and Electronics, Umeå University. The idea and objective of the course is to give the students a forum where they can actively participate in scientific research and development through developing their own ideas. The course introduces students to independently researching an interesting topic, using a foreign language orally and in writing, writing a scientific article on their work, peer-review and presenting their work at a conference. The conference format was chosen to provide a realistic environment for the presentation of the results. The work has been reviewed both by other participant on the course and members of the department. If the reviews are favorable, the paper is accepted as a full paper at the conference and included in the proceedings. Research that has an interesting topic and potential for future publication is presented as work-in-progress at the conference and the abstract is included in the conference proceedings.

This year 19 papers were accepted at the conference as full papers and are included in this proceedings. Furthermore, 2 papers was accepted as work-in-progress with included abstract.

Umeå, May 2016

Thomas Mejtoft and Ulrik Söderström
Program Committee
Table of Content

Full papers
Choice of fonts and price expectations
Anna Österlund

Evaluating Google material design with a focus on web application animations
Jakob Åberg

Presenting content to restless receivers: evaluating 360 video and interactive street view formats
Sandra Ingerhage

Using smartwatch as an aid for people with diabetes
Elin Hörfeldt

Examining the potential of smart glasses as an assistive technology for dementia patients
Karolina Jonzén

UX guidelines for accessibility in action shooter games
Anna Pääni

The effect the number of columns has on information perception on a web page
Petra Andersson

Paper vs. simple digital prototypes: differences in user feedback when testing
Hanna Söderström

Guidelines for designing context-aware fitness applications
Julia Westman

The role of feedback in educational applications for smart devices
Lena Lindberg

Push-notifying the end-user at the most appropriate time
Mattias Edin

Aesthetic value versus function in mobile applications
Robin Schulze
Feeling safe in autonomous vehicles  
*Gustav Nordlander*

Using augmented reality for learning a second language  
*Tove Barsk*

Biohacking: Evaluating the benefits of near field communication in the human body.  
*Carl Holmertz*

Ditching The Hamburger Menu: An analysis of current menu design  
*Anton Forsberg*

Interaction limitations on mobile devices related to screen size  
*David Bergvik*

Evaluating the potential and limitations of Google Glass  
*Joakim Zakrisson*

Design of brain-computer interface to ease everyday life for people with severe motor impairments: A metanalysis  
*Ellinor Rönn*

**Work-in-progress**  
Limitations and possibilities of using eye-tracking technology with smart glasses  
*Anton Bäckström*

Analysis of text based information in virtual reality  
*Isak Dunér Lundberg*
Choice of fonts and price expectations

Anna Österlund
Department of Applied Physics and Electronics
Umeå University, Sweden
id10aod@cs.umu.se

Abstract.
The market for on-line shopping has expanded as an increasing number of customers are connected to the Internet. Fonts are one of the factors that influence the visual impression of the description of a product. The aim of this study was to assess if the choice of fonts had any effect on price expectations of a product.
The participants in the study were randomly subjected to one of the fonts describing three different products (TV, jeans, and running shoes). The three most viewed Google fonts in the categories serif, sans serif and handwriting were studied (Slabo 27px, Open Sans, Indie Flower). 219 persons participated in the study distributed equally in each font group. There were no significant differences in price expectations between the font groups. Disregarding fonts, there were significant differences between men and women, and between the younger and the older participants in price estimation for all three products. The students estimated a significantly lower price for a TV than the employed group, and significantly lower price for the jeans than the retired group. In this study there were no differences reaching significance in price expectations for the three products described in three different fonts. However, in the age group 60-69 there was a strong tendency to estimate a lower price for a TV described in Indie Flower than in Slabo 27px. For a TV described in Slabo 27px, women tended to estimate a higher price than men did.

Keywords. font, typeface, price expectation, price estimation, typography

1 Introduction

The trading of goods has been a big part of human society since the start of civilization. But now, with the worldwide expansion of the Internet, the trading has found a new platform, online shops. In six years, from the year 2008 to 2014, the e-commerce nearly doubled in Sweden [1]. According to Statista [2], the number of Internet users worldwide that has purchased a product online is 40 percent, a number that is projected to keep growing.

With a growing online market, companies are starting to expand their businesses by offering the possibility of buying products online. Before, their customers were the ones that came in to the physical store. Now, with online sales,
a customer could be anyone with access to a computer connected to the Internet. However, with a bigger market comes increased competition. In 2014, almost 50 percent of the business owners in Sweden said that they are experiencing an increasing competition from foreign companies [1].

To be successful on the online market, it is important to design a site that attracts the customers. When asked which features played an important role in whether or not they would purchase a product from a web shop, 94 percent of the Swedish consumers said it was important that the information about the product was good and accurate [1]. 89 percent of the same consumers also said that it was important that it was evident what the total price would be.

There are several factors to consider while designing the product information on a website. Choosing an appropriate layout, a good picture of the product and writing a good product description are some factors that need to be considered. Even though the content of the text might be the most important aspect to make the product description good and accurate, the visual impression of the actual letters can add meaning to the text as well. Previous research has shown that font selection adds additional information to a text, like personality traits [3, 4] and trustworthiness [5].

The aim of this study was to determine how the choice of font affects the price expectation of a product. Research questions to be answered in the study are:

- Does the choice of font affect the perceived price of a product?
- What are the main differences between the serif, sans serif and handwriting fonts, with respect to price expectations?

2 Font typography

Since hundreds of years typography has developed a large amount of knowledge and rules. We use serified fonts to improve readability of body text, and for posters and advertising we use sans serif. These rules are not necessarily true but rather highly subjective, based on empirical knowledge. Modern research is questioning these statements [6].

The effect of fonts on readability has been examined in several studies [7–10]. However, even though good readability is an important part of conveying information through text, it is far from the only effect that comes with the choice of font. As readers, we are not simply passive recipients of information presented to us. In fact, our experiences, prior knowledge and expectations shape our interaction with the visual language [3].

A study by Lewis and Walker [11] reveal that when the appearance of the word is consistent with the meaning of the word, it is processed faster. When asked to press a button when the words "strong" and weak appeared on a screen, the study showed that people reacted faster if the word "strong" appeared in bold and if "weak" appeared in light weight. If the word was presented in the opposite font weight, "weak" in bold and "strong" in light weight, the reaction times were significantly slower.
While visual features of words influence the processing of a written message they may also actually add their own semantic representations [12]. Studies show that different fonts are perceived as having different personality traits [3, 4]. Serifed fonts can be perceived as being elegant, interesting, distinct, formal, mature, practical, stable, charming, emotional and beautiful [4, 13]. Sans serif fonts, on the other hand, can be perceived as being smart, powerful, manly, upper-class and loud [4]. Script fonts, such as handwriting, are perceived as happy, creative, feminine, casual, youthful and cuddly [4].

Despite adding information to the text, the choice of font can also have an impact on the readers’ impression of the writer. When applying for a job, choosing an appropriate font for the resume affects whether or not the applicant gets called in for an interview [5], and thereby still being in the running of getting hired. Choosing an inappropriate font may actually cost the applicant the job.

In business, choosing an appropriate font for the brand is important since it affects how the consumers perceive it. Semantic associations linked to the type font used are transferred to the brand. A formal type font is associated with a luxurious brand while the use of a casual type font results in a more casual brand perception [12]. Thus, the consumers take the type font characteristics and transfer them to brand personality perceptions [14].

If the appearance of the font contradicts the meaning of the writing, it is perceived as less appealing [15]. A study by Doyle and Bottomley [15] show that subjects were more likely to choose a chocolate from a box where the brand name was written in an appropriate font. Thus, choosing an inappropriate font for a brand can make the consumer reject the product and, instead, choose a product from a competing brand.

The ability to affect the meaning, emotions and convincing qualities attached to the content of a text, without the readers noticing, has implications for advertising, marketing and persuasive literature [16].

3 Method

The aim of this study was to evaluate the perception of font type and how a product is perceived. A survey was thought to be most suitable for collecting information about the participants’ personal opinions and behavior.

The survey was conducted in Swedish during March 2016. It was designed to assess if the choice of fonts had any effect on the price expectations of a product. The participant was asked to read three different product descriptions, one at a time, and estimate what they believed the product would cost to buy. They were also asked about their age, sex, occupation, how often they research products online, their online shopping experience and if they had ever bought one of the products described in the survey.

Three copies were made of the survey where the only difference between the copies was the font style of the text describing the products. The font styles used in the survey were serif (Slabo 27px), sans serif (Open sans) and handwriting font (Indie Flower).
Although three fonts were tested in this study, each participant was only subjected to one type of font and therefore only asked to answer one of the surveys. A shared first page was constructed to provide all participants with the same information and instructions before being subjected to one of the surveys. After reading the instructions the participants were randomly subjected to one of the three surveys.

Since the participants were not to focus on the font, but the product description as a whole, they were not told that the survey was conducted to evaluate fonts. The participants were only told that the survey concerned price estimation.

3.1 Choice of fonts

Table 1. Fonts used in the survey

<table>
<thead>
<tr>
<th>Most viewed Google fonts in January 2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serif: Slabo 27px</td>
</tr>
</tbody>
</table>

The fonts used in the study (Table 1) are Slabo 27px, Open Sans and Indie Flower. These fonts are the most viewed Google fonts [17], January 2016, in the categories, Serif, Sans Serif and handwriting.

3.2 Choice of products

Since the study focused on products purchased online and the survey was conducted in Swedish, the products described in the survey were chosen from the top three product categories for online shopping in the Nordic countries. According to PostNords compilation of the e-commerce, e-barometern [1], in the Nordic countries for the year 2014, the top three product categories were home electronics, clothes/shoes and sporting goods.

Three products, one from each category, were chosen to be used in the survey. The products selected were a TV [18], jeans [19] and running shoes [20].

3.3 Survey

The first page provided the participant with instructions and information about the survey. The participants were informed that the purpose of the survey was to
identify how a written product description effects price estimation of products. Information about what the first three questions would entail was presented. The participants were told that they were to estimate the price of three different products after only reading the description of the products. They were also informed that the study was conducted as a part of a course at Umeå University.

After receiving information about the survey, the participants were presented with a button labeled "Do the survey". When the button was clicked they were randomly forwarded to one of the three surveys.

![Product 1](image)

**Fig. 1.** The first question of the survey, in Slabo 27px font [18]

The first three questions of the survey consisted of a product description, one from each product category, and a text field with a prompt for the participant to specify the estimated price of the product. The products were described in the following order, TV, jeans, running shoes. An example of the first question in the survey can be seen in figure 1, the product description of a TV. The option for the participant to enter the estimated price in free text was chosen to avoid affecting the answer by restricting it to an interval.

After completing the three questions about product description and price estimation some supplementary questions were presented. These questions were asked to get additional information about the participants, their sex, age, occupation, how often they research products online, their online shopping experience and if they had ever bought one of the products described in the survey.

The first three questions regarding price estimation were mandatory, while filling out the additional information was voluntary.
### 3.4 Statistical analysis

A quantitative approach of collecting data through a survey was chosen. Data were analyzed using an independent two sample \( t \) test, \( \alpha = 0.05 \), to detect differences between two samples. F-test was used to evaluate the variance of the samples.

### 4 Result

The surveys were performed by 219 persons. The participants were equally distributed with approximately 70 participants in each group (Table 2).

<table>
<thead>
<tr>
<th>Study population</th>
<th>Slabo 27px</th>
<th>Open Sans</th>
<th>Indie Flower</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>n = 73</td>
<td>n = 74</td>
<td>n = 72</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>29</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>Male</td>
<td>43</td>
<td>41</td>
<td>37</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-19</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20-29</td>
<td>25</td>
<td>17</td>
<td>20</td>
</tr>
<tr>
<td>30-39</td>
<td>21</td>
<td>31</td>
<td>21</td>
</tr>
<tr>
<td>40-49</td>
<td>8</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>50-59</td>
<td>6</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>60-69</td>
<td>11</td>
<td>12</td>
<td>17</td>
</tr>
<tr>
<td>70-</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Occupation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>56</td>
<td>56</td>
<td>54</td>
</tr>
<tr>
<td>Student</td>
<td>13</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>Retired</td>
<td>4</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

All ages were represented except for the age groups, 19 and younger, and 70 and older. The majority were employed and there were slightly more men in each group.
Most participants made research on the Internet before buying a product (Fig. 2a) and the majority was using the Internet for shopping (Fig. 2b). The question “Have you ever bought one of these products?” had to be excluded. Feedback from some of the participants indicated that they thought that only online purchases were to be considered.

Table 3. Mean resp. min-max values for the estimated price in SEK of the products evaluated

<table>
<thead>
<tr>
<th></th>
<th>Slabo 27px</th>
<th>Open Sans</th>
<th>Indie Flower</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
<td>n = 73</td>
<td>n = 74</td>
<td>n = 72</td>
</tr>
<tr>
<td><strong>TV</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>7502 ± 4608</td>
<td>7652 ± 3126</td>
<td>6675 ± 4254</td>
</tr>
<tr>
<td>Min - Max</td>
<td>3000 - 35000</td>
<td>2000 - 18000</td>
<td>300 - 20000</td>
</tr>
<tr>
<td><strong>Jeans</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>691 ± 338</td>
<td>647 ± 298</td>
<td>665 ± 314</td>
</tr>
<tr>
<td>Min - Max</td>
<td>199 - 1700</td>
<td>40 - 1200</td>
<td>50 - 1500</td>
</tr>
<tr>
<td><strong>Running shoes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>1361 ± 571</td>
<td>1300 ± 399</td>
<td>1311 ± 635</td>
</tr>
<tr>
<td>Min - Max</td>
<td>349 - 4000</td>
<td>399 - 2500</td>
<td>200 - 5000</td>
</tr>
</tbody>
</table>

There were no significant differences between the price estimation of the three products, regarding the fonts. However, the age group 60-69 showed a tendency to estimate a higher price for a TV described in Slabo 27px than in Indie Flower (p=0.059). In the Slabo 27px group, the result indicates that women thought that the TV was more expensive than the men did (p=0.066).

Disregarding fonts, there were significant differences in the price estimation of products. The women estimated the price for a TV (p=0.046) and running shoes (p=0.027) higher than the men did. But for a pair of jeans it was the opposite (p=0.00003).
Price estimations for a TV and running shoes did not differ between the age groups. For jeans, however, there was a tendency to estimate a higher price with increasing age. For example, the age group 60-69 estimated the price for a pair of jeans higher than the age group 20-29 (p=0.0048).

The students estimated a significantly lower price for the TV than the employed group did (p=0.0158). The students also thought that a pair of jeans were less expensive than the retired group thought (p=0.0085).

5 Discussion

In this study, font choice did not significantly affect the price expectation of a product. However, it showed that age, sex and occupation are all factors that seemed to have an impact on how we estimate prices. Slightly more men responded to the survey but this did not appear to have any effect on the results.

Although the differences between the fonts did not reach significance there could be some possible reasons for this. The size of the study groups or the chosen products could have had an impact on the results.

Since fonts are known to add their own semantic representations [12], a formal or powerful font might be more appropriate than a casual font when describing a technical product, like a TV [4,13]. In this study, there was a tendency to estimate a lower price for a TV described in Indie Flower than in both Slabo 27px and Open Sans. No such tendency was apparent for a pair of jeans and running shoes.

The choice of products could have had an impact on the result since prior knowledge about the products may have affected the participants’ price estimations. If they already had a clear view regarding the price level of the product, it is possible that the appearance of the font did not have an impact on the results. Thus, choosing an unusual product, one which the participants had limited knowledge about, might have resulted in a different outcome.

The fact that there were differences between the sexes in the price expectations for all three products is not easy to explain. Even though it might be a controversial statement, one possibility could be that women are less interested in technical devices [21] and therefore might perceive them as more expensive. For the jeans, it might be the differences in shopping habits that explains the diverse results.

Age did not have an impact on the price estimations of a TV and running shoes. However, the age group 60-69 tended to estimate a lower price for a TV described in Indie Flower than in Slabo 27px. This could be due to a difference in how fonts are perceived by different generations, based on their empirical knowledge [6].

The younger participants thought that a pair of jeans was less expensive than the older ones thought. In fact, the price estimation increased with increasing age. A possible explanation for this could be the increasing market for jeans during the last century. As a result of this, the production cost has decreased.
and thereby also the market price. The generations might have different historical perspectives.

A higher standard of living seemed to play a role for the price expectations, both for a TV and a pair of jeans. The employed group estimated a significantly higher price for a TV than the students. The retired group thought that the jeans were more expensive that the students thought. This seems difficult to explain. It might be that students presumably buy cheaper TVs and jeans since their budget is tighter during the study period than after getting a job.

The study showed that most of the participants researched a product on the Internet before buying it, and the majority had bought a product online at least once a year. Today, most swedes have access to the Internet which has made online shopping possible. In fact, a survey conducted in the last quarter of 2014 by PostNord [1], showed that 76 percent of the swedes were buying products online.

5.1 Future work

There are some limitations in this study. Larger test groups might have led to differences between the chosen fonts, since some tendencies for difference in price estimation were observed between the groups. It is also possible that the products chosen were too familiar to the participants, and that previous knowledge about the products might have overpowered the effects of the font appearance. With a less known product the result might have been different. Further studies are needed to clarify this.

6 Conclusion

In conclusion, this study cannot show any significant differences in price estimations for a TV, a pair of jeans and running shoes, regarding the fonts chosen for product descriptions. However, the subgroup of participants over the age of 60 tended to estimate a lower price for a TV described in Indie Flower than the in Slabo 27px. For a TV described in Slabo 27px, women showed a tendency to estimate a higher price than the men did.

References

Evaluating Google material design with a focus on web application animations

Jakob Åberg
Department of Applied Physics and Electronics
Umeå University
jaab0010@student.umu.se

Abstract. With an expanding access of the Internet and a growing use of dynamic content on websites, web-animation has become a subject for discussion. There are different opinions on how animations should be implemented to aid users instead of distracting them. This paper analyses the use of web-animations and specifically the animations given in the guidelines suggested by Google material design. The usability of the animation guidelines in Google material design was tested, and the effects they had on information seeking performance. The findings of this paper were based on a SUS-questionnaire and two web applications: one that follows the animations suggested by Google (application A), and another that implements the animations that Google advice against (application B). 20 persons participated in the study, 10 on each application. The participants of application A seemed to execute all tasks faster with an exception of the first task. The scores of the SUS-questionnaire indicates that application A was perceived as over average, in regards of usability, while application B was perceived as passable. The results of this paper cannot show any significant differences in time efficiency between the two applications. However, the scores received from this study suggests that there might be a link between usability and the animations suggested by Google.

1 Introduction

Ever since the beginning of civilization people have been fascinated by moving images and of animation. The desire to give motion and life to pictures has played a big part in the developments of culture, science and technology. Now animation is widely used in applications and on websites, both to capture viewer attention but also to enhance usability [1]. The use of visualized information has been increased with approximately 10000 percent on the world wide web since 2007 [2], a percentage that just keeps on growing.

With an expanding access of the Internet and a growing use of dynamic content on websites, web-animation has become a subject for discussion and studies have been conducted on the topic [3,4,5]. Earlier
work by Zhang [3] claims that web-animations interfere with users’ primary tasks when using web-pages. This happens when animation as a secondary stimulus deteriorates the information seeking performance, and when the animation is similar to the task [3]. On the other hand, there is also a study that suggests that motion design and animation enriches the web-page and makes it comfortable and easy to use [1]. In an article from 2014 [6] the authors state that web-animation and dynamic web-sites make users happy, and can surprise in a positive way. Mori, Paterno and Furci [6] found that it is very important for the animations to facilitate interactions, without being perceived as annoying.

To help developers and designers implement animations, companies such as Google, Apple and Facebook release design guidelines to follow. Several studies have been conducted on topics such as animation in web applications [1,3,4,5] but none have focused on evaluating the front-runners of technology and interfaces today. This topic deserves our attention as there is a tendency to put trust in technology organizations [7].

The aim of this study was to get a better understanding of how people react towards animation in web applications. Furthermore, the objective was to evaluate Google material design with a focus on web application animation. Research questions to be answered in this study are:

- What is the correlation between usability and the Google material design animation guidelines?
- What is the difference in time efficiency regarding information seeking performance using the animations suggested by Google and the animations they advice against?

## 2 Google material design

In 2014, Google released their new design language called material design. The main goal was to create a visual language that combines the classical principles of successful design with the innovation and possibility of technology and science [8]. Google also wanted to make it easier for developers and designers to build beautiful and consistent applications that allows for a unified experience across all platforms [8].

Just one year after the big release, Google had implemented the visual language on most android platforms and on all the Google Play-branded applications, such as Google Mail, Drive, Sheets, Docs, Maps and Youtube [9]. Google’s material design is not only reserved for mobile applications
there are also a number of desktop web-interfaces that has incorporated it as well, such as Google Drive, Mail, Docs and much more. The language is consistently updated and maintained [10].

The material design framework is constructed by three main principles: (1) Material is a metaphor; the unifying theory of a rationalized space and a system of motion taken from the real world. (2) Bold, graphic, intentional; typography, space, colors and use of imagery is not only pleasing to the eye. Implemented with the right approach it can create hierarchy, meaning and focus in a beautiful way. (3) Motion provides meaning; the idea that motion helps the user navigate and to understand how elements belong together [8]. This paper will mainly discuss the latter of these three principles, focusing on animation.

2.1 Animations

In this study five of the material design animation guidelines will be tested. The animation guidelines chosen are the ones that often relates to navigation elements within a website and the elements that take up the most visible space in the browser. Thus, the ones that is believed to give the most tangible results in this paper. The figures in this section are all based on figures given in Google material design [8].

2.1.1 Point of origin. In the real world, feedback is received in such a way that output origins from the input; output elements expands from input elements. This is the best way to do it in the digital world as well (Fig. [1] [8].

2.1.2 Visual Continuity. Transitions between two visual elements should be smooth, clear and non-distractive [8]. Irrelevant content should be removed and newly generated items should be highlighted. Animations should direct user attention, and create connections between transitioning states [8]. If movement is used it is important that it is precise.

2.1.3 Hierarchical timing. Consider the timing and the movement of elements. Motion should follow the information hierarchy; the most important content should be highlighted [8]. This is easily done by creating an animation path through the content (Fig. [2]). This makes it easier for the user to know where to put their attention [8].
2.1.4 **Consistent choreography.** Transitioning items should behave in the same manner. The paths that the elements follow should be orderly and create meaning. If items move in different ways (Fig. 3b), it can be distracting for the user. 

---

**Fig. 1.** a) The best way to animate material response: output should originate from input [8]. b) How not to animate: output origins from the middle of the screen from no particular source [8].

**Fig. 2.** a) How to animate hierarchical timing [8]. A path through the content helps the user to know where to first put their attention [8]. The most important content should come first. b) How not to animate motion; independently and without a path [8].
2.1.5 Delightful details. With this guideline Google [8] wants to acknowledge the importance of creating applications that are fun by adding small transitional details, like turning a hamburger icon into a go-back-arrow when clicked. In this way, Google believes that an application can go from a good to a superb one [8]. There is a big risk with this since it is hard to draw a line where animations and transitions become too distracting.

3 Method

To analyze the animation guidelines given in Google material design, two simple web applications were created; application A - based on the animations suggested by Google, and application B - based on the animations that Google advice against. They were both built with the help of Angular Material [11] where CSS3 and JavaScript animations were used.

Application A and B were both plain in terms of color and general design so that the test participants focused primarily on animations. The applications contained four pages, each page with a task for the participants to solve in the fastest way possible. To measure the effects of the animations on information seeking performance, the effectiveness of
the test participants was evaluated as they were going to identify target strings from other strings [3]. Random strings were chosen in this study in order to eliminate any automatic processing of familiar target strings [3]. The process of identifying strings is one of the most common information seeking tasks on a web-site [12] and thus, very relevant to this study. Application A and B were tested individually on ten test subjects each, with two pilot tests conducted per application.

To analyze how effective the test subjects were, the time was taken for each task. The tests were conducted in Swedish, since all the subjects understood written and spoken Swedish fluently.

3.1 Application set up

The introduction page provided the participants with general instructions about the test. They were informed that the objective of the test was to measure the usability of a website with different animations. The participants were also told that they were going to perform a task for each page on the web-site and afterwards answer a short questionnaire. A button was then displayed, with the text "Begin test". When this button was clicked the users were forwarded to one of the applications.

Before the participants were directed to each page in the application, a brief explanation was given about that specific page and the task they were going to solve. The only difference in the pages of application A and B was the animations.

3.1.1 First page. The participants were told that this page had a button labeled "Show". They were informed that if this button was clicked a grid of squares was going to be displayed, each square with a random string of letters. The participants were then asked to find the string "pagd" as fast as possible and click on it. In application A the grid animated in such a way that it followed a path (Fig. 2a), in application B the squares in the grid animated independently and without a path (Fig. 2b). The page of both applications are based on the guideline Hierarchical timing.

3.1.2 Second page. The second page of the application is based on the guideline Consistent choreography. The participants were told that a grid of squares was going to be displayed and that each square contained a random string of letters. The participants were asked to find the string "mui" as fast as possible and then click on it. In application A all the
squares faded in from bottom and up, and in application B the squares faded in from different directions without a clear choreography.

3.1.3 Third page. In this page of the application the focus was on the animation guideline *Point of origin*. The participants were told that they were going to see a page with a button, when this button was clicked a container would be displayed. The participants were then asked to find a dot as quick as possible and then click on it. In application A the container expands from the button that is pressed, in application B the container expands from the bottom of the page without a clear point of origin.

3.1.4 Fourth page. The last page focused primarily on *Visual continuity* and *Delightful details*. The participants were told that the page was going to contain six cards, each card with a number displayed on it. They were asked to click on these cards, to find the random string of ”opasd” as quick as possible. In application A the card was flipped over and highlighted in the middle of the screen to show the test subjects the string of letters. In application B the card did not move, when clicked it just displayed the other side of the card.

3.2 Questionnaire

After finishing the tasks, a questionnaire was conducted to get a sense of how the participants perceived the usability of the application. The method chosen in this study was The System Usability Scale (SUS) \cite{13,14}. SUS is a likert scale consisting of ten statements, each statement can be answered on a scale ranging from ”Strongly Disagree” to ”Strongly Agree” \cite{14}. The statements that were given in the questionnaire are shown in Table 1.

Final scores from the SUS can range from 0-100. A score over 70 means that a product is passable, a score over 80 means that a product is over average and a really good product scores over 90 \cite{14}.

Information about the test participant’s sex, age and occupation were also gathered in this questionnaire. All questions and statements were mandatory.

3.3 Statistical analysis

To analyze the data both from the time efficiency test and the questionnaire, a two sample t test was used to detect differences in application A and B.
Table 1. SUS statements given to the test participants in the questionnaire. The statements are changed a bit from the SUS-standard to fit the study conducted.

<table>
<thead>
<tr>
<th>SUS statements given in the questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I think that I would like to use the animations in this test frequently.</td>
</tr>
<tr>
<td>2. I found the animations in this test unnecessarily complex.</td>
</tr>
<tr>
<td>3. I thought the animations in this test were easy to use.</td>
</tr>
<tr>
<td>4. I think that I would need the support of a technical person to be able to use an application with these animations.</td>
</tr>
<tr>
<td>5. I found the various animations in this test well integrated.</td>
</tr>
<tr>
<td>6. I thought there was too much inconsistency in the animations of this test.</td>
</tr>
<tr>
<td>7. I would imagine that most people would learn to understand the animations in this test very quickly.</td>
</tr>
<tr>
<td>8. I found the animations in this test very awkward to use.</td>
</tr>
<tr>
<td>9. I felt very confident navigating through this test.</td>
</tr>
<tr>
<td>10. I needed to know a lot in order to use the animations in this test.</td>
</tr>
</tbody>
</table>

4 Result

The study was conducted by 20 people, 10 women and 10 men with an average age of 29 years (ranging from 20 - 77). The participants were equally distributed with 10 test subjects on each application. The time was taken for each participant as they tried to solve the tasks. The mean time to solve each task are shown in Table 2.

There was not a significant difference in how fast the test participants of application A solved the tasks compared to how fast the participants of application B solved them ($p=0.66$). In task 2, 3 and 4 the participants of application A were faster and in task 1 the participants of application B were faster, but not significantly ($p_1=0.63$, $p_2=0.23$, $p_3=0.49$, $p_4=0.90$). The biggest time difference was in task 2 where participants of application A were 3.217 seconds faster than participants of application B ($p_2=0.23$). Participants of application A solved all the tasks 1.067 seconds faster than B.

The results from the SUS-questionnaire are given in Table 3. There were no significant difference in perceived usability of the two applications ($p=0.14$). The SUS Scores were slightly higher for application A than
Table 2. Mean time for the participants to solve each task in the test. The tasks were solved slightly faster in Application A than B.

<table>
<thead>
<tr>
<th>Task</th>
<th>Application A</th>
<th>Application B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8.543</td>
<td>7.053</td>
</tr>
<tr>
<td>2</td>
<td>4.356</td>
<td>7.573</td>
</tr>
<tr>
<td>3</td>
<td>5.535</td>
<td>7.319</td>
</tr>
<tr>
<td>4</td>
<td>12.372</td>
<td>13.132</td>
</tr>
<tr>
<td>Total</td>
<td>7.702</td>
<td>8.769</td>
</tr>
</tbody>
</table>

Table 3. Calculated SUS Scores for each participant in application A and B.

<table>
<thead>
<tr>
<th>SUS Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application A</td>
</tr>
<tr>
<td>87.5</td>
</tr>
<tr>
<td>70.0</td>
</tr>
<tr>
<td>70.0</td>
</tr>
<tr>
<td>95.0</td>
</tr>
<tr>
<td>80.0</td>
</tr>
<tr>
<td>82.5</td>
</tr>
<tr>
<td>87.5</td>
</tr>
<tr>
<td>60.0</td>
</tr>
<tr>
<td>80.0</td>
</tr>
<tr>
<td>65.0</td>
</tr>
</tbody>
</table>

Average: 77.8  Average: 70.3

B, where application A scored an average of 77.8 and application B an average of 70.3.

5 Discussion

In this study, a significant difference in time efficiency could not be proven between the two applications. However, we can see a variation in time in
how fast all the tasks were solved. In task 1, the participants of application B were 1.490 seconds faster than the participants of application A. This could mean that Hierarchical timing does not have any effect when it comes to information seeking performance in web-animations. One thing that was observed during test runs was that the participants of application A were much faster at placing the mouse in the first square of the grid than the participants of application B. A possible explanation for this is that Hierarchical timing actually makes it easier for the user to know where to focus.

The biggest time difference can be seen in task 2 where the participants of application B were more than 3 seconds slower than the participants of application A. This could mean that Google [8] is right in their statement that random motion in applications and on web-sites can be distracting for the users. The result indicates that when all elements follow a certain movement it might actually aid user understanding of the application, instead of getting the users confused on where to put attention.

The result from task 3 and 4 did not show any statistical evidence that there is a difference in time efficiency either. Google’s main objective with the guideline Point of origin might just be to help developers create applications that have high connectivity to the real world, focusing on the usability aspects of applications instead of their time efficiency. The same discussion might apply to task 4 and the the guidelines Delightful details and Visual continuity. The guidelines may not have any effect when it comes to information seeking performance, but rather a higher emotional value. It was clear from the study that the participants of application A enjoyed task 4 a lot more than the subjects of application B.

Although there was not a significant difference in how the participants of the two applications perceived usability, there could be some logical reasons for this. Two of the test participants stated that they had not noticed the animations in the applications. This could be due to the fact that some users can block the distraction or the events caused by animation [3]. There is a possibility that some of the test participants were so focused on the tasks that they did not notice the animations. If this is the case, there is a chance that the participants did not notice the usability of the animations and thus, the result from the SUS-questionnaire can not be seen as reliable.
6 Future work

There are some limitations in this study. Since some of the participants stated that they had not noticed the animations in the tasks, there is a possibility that the study was conducted in a misleading way. It may have been the wrong approach to measure perceived usability in a test that also measures time efficiency on information seeking performance. A better approach for this study could have been to build two applications; one task-application that measures the time efficiency and another application that the participants could navigate through. This may have separated the two factors of the test and the participants might have put equal focus on both the usability and to solve the tasks as quick as possible.

The study could be extended in a lot of ways. A larger amount of test participants might have led to bigger differences in time efficiency and perceived usability. A test with more pages and tasks could also result in more thorough conclusions about the guidelines. Another unexamined factor is web-pages with multiple animations, which are commonly used in web applications today.

7 Conclusions

This study cannot show any significant differences in time efficiency in information seeking performance as well as usability when it comes to the Google material design guidelines. However, the scores received from the SUS-questionnaire suggests that their might be a link between usability and the animations suggested by Google. This research provides a base for future investigations about design guidelines in web-environments.

References

9. V. Savov, “Google’s next big android redesign is coming in the fall.” [Online; accessed 25-February-2016].
Presenting Content to Restless Receivers: Evaluating 360 Video and Interactive Street View Formats

Sandra Ingerhage
Department of Applied Physics and Electronics
Umeå University, Sweden
sain0012@student.umu.se

Abstract. Digital lifestyles with increased access to information and needs for extracting the gist have been found to change how people focus their attention. This paper evaluates the 360 video and the interactive street view formats, with the aim to identify advantages and disadvantages of their ability to capture the attention of people with digital lifestyles. The formats were evaluated by letting users consume information through them, and the subjective experience of the users’ focus was primarily considered. Findings indicated that the rich media and large scope of stimuli provided by the formats is an advantage, but that it also is important that the rich media is cohesively arranged and contributes to getting the message through, or the rich media may instead have a negative effect on the focus. Furthermore both advantages and disadvantages were found in the tempo, scene switches and navigation of the formats. In conclusion both formats have potential to capture the attention of people with digital lifestyles, but they also both come with challenges.

1 Introduction

In the information age there is always something new to take part of and discover. Increasing digitalization and access to information in everyday environments have changed how people focus and allocate attention. A decade ago people reported that a shift from reading printed documents to reading electronic documents on digital devices changed their reading behavior [1]. When reading on a digital device, more time was spent on selected reading and less time was spent on in-depth reading. Hence, the behavior changed from focusing prolonged attention following a text to alternating focus between different parts of the text and extracting key sections.

Today, these changes seem to have continued to develop on the same path. Evolving technology does not only let people do more things digitally, it also encourages people to take part of more information since it is easily accessed through always present media streams. Digital lifestyles are characterized by a high volume of media consumption as well as concurrent consumption of several media streams. Studies have shown that people with digital lifestyles have harder
to focus on the same task for longer periods, especially in less interactive environments [2, 3]. They strive to find the highlights and are driven by the thrill of finding the next one, which is always certain to be found in the streams of awaiting information. Concurrent consumption of several media streams, referred to as media multitasking [3–5], also decrease the ability to filter out irrelevant information and maintain focus on a single primary task [3]. Although people with digital lifestyles are more easily distracted and have problems to maintain focus, they have at the same time an increased ability to faster process and assimilate information [2].

Digital lifestyles are common today, especially among the young generation of teenagers who reflect the future to come [6, 7]. If digital lifestyles will continue to gain ground as the typical lifestyle, it is important to acknowledge the change in attentional behavior that comes with it. Having problems to maintain prolonged focus but possess skills to faster assimilate information, people seek to sort out the gist, bring forth the most important conclusions and move on to the next thing [2]. This should be taken into consideration when it comes to reaching out with information to a receiver, it is an advantage to keep it short, clear and get to the point. In many situations this is a good strategy to let receivers handle heavy information loads from multiple channels, but some information may be hard to shorten down and compress without losing essential parts and may need some more time to be comprehended and understood. The change in attentional behavior brings both challenges and opportunities to reach out with information to receivers. While ever evolving technology is contributing to multitasking tendencies and decreased ability to maintain attention, it may also be utilized to enhance these behaviors in order to convey information. Two interesting ways of presenting information in this sense are the 360 video format and what is referred to in this paper as the interactive street view format. These formats are interesting since they provide a rich media experience and have a way of keeping receivers active.

This study aims to evaluate the ability of the 360 video format and the interactive street view format to enhance the attentional behavior of people with digital lifestyles. The objective is to identify advantages and disadvantages related to the characteristics of these formats when it comes to keeping the focus for a longer period of time.

2 Background

The 360 video format and the interactive street view format have similarities but also disparities which makes it interesting to analyze them in relation to each other. The formats and their characteristics will be described further in this section, after a consideration of earlier studies and concepts of human attention.
2.1 Human attention

Human attention is a wide area of research, and this section will attempt to bring forth some pieces that are especially interesting in relation to digital lifestyles, and that can be used as a reference ground for analysis of the formats.

2.1.1 Various types of attention

There are different approaches to describe cognitive processes related to human attention. One model proposed by Sohlberg and Mateer [8] distinguish five categories that describe different types of attention.

- **Focused attention** is the ability to respond to stimuli, which may be a visual, auditory or tactile or any other type of stimuli.
- **Sustained attention** is the ability to maintain focus and respond consistently while engaging in a continuous prolonged activity.
- **Selective attention** refers to the ability to stay focused on a task while exposed to concurrent or competing stimuli.
- **Alternating attention** is the ability to shift focus between tasks that demands different cognitive skills.
- **Divided attention** refers to the ability to simultaneously respond to two or more task demands.

These different types can describe the attentional demands on a person when taking part of information that requires prolonged attention. The person essentially needs to be able to focus attention to assimilate any information, and need to maintain sustained attention to get to the end of the story. It is also crucial to be able to allocate attention to the selected task and not get distracted by other stimulus. Depending on the content and how it is presented, the person may also need to divide or alternate attention between different parts of the information. Hence, all these aspects of attention is important to consider when trying to convey information that needs to keep the receiver’s attention for a longer period.

2.1.2 Factors that affect attentional focus

**Wide attentional filters.** One branch in research on attentional behavior and digital lifestyles focus on selective attention. Among such work are studies that have shown that people with digital lifestyles tend to have a wide attentional focus and take in much of the information available [3, 5]. This has been described as breadth-biased attention behavior [3, 9], meaning that larger amounts of information is processed rather than that a particular piece of information is selected and processed in detail. It also means that all stimuli is treated as equally important and are equally attended to [9]. These findings suggest that attention is best captured by larger amounts of stimulus that can catch the breadth of the attentional filter.

**Perceptual and cognitive load.** There is also research that consider perceptual processing and its relation to cognitive processing and workload. With
Load Theory, Lavie, Hirst, De Fockert and Viding [10] propose that the ability to maintain attention increases when a task involves high perceptual load, i.e. when the level of perceptual stimuli is high. It also proposes that it decreases when a task involves high cognitive workload, i.e. in situations when the processing requires much effort and cognitive resources. Hence, this means that increasing the perceptual stimuli in a task should increase the ability to focus on the task, as long as it does not also imply more cognitive load. It has also been shown that increased perceptual load in tasks can reduce the processing of task-irrelevant information [11], and this could be considered as a strategy to help people with digital lifestyles to keep away from distractions.

Multitasking. Research do also concern attention in relation to multitasking, and multitasking by dividing and switching attention within and between tasks is a core characteristic of people with digital lifestyles. A model that takes a broad approach describing different aspects of multitasking is the Multiple Resource Theory proposed by Wickens [12]. The model is based on four dimensions of cognitive resources, described below.

The Stages of processing dimension implies that perceptual and cognitive processing uses different resources than selection and execution of action.

The Codes of processing dimension implies that spatial processing uses different resources than linguistic processing.

The Modalities dimension implies that auditory perception uses different resources than visual perception.

The Visual channels dimension implies that focal and ambient vision is distinguished visual channels.

The point with these dimensions is that they show how cognitive resources best can be utilized, and describe what kind of tasks that can be done concurrently with good performance. The better tasks use different resources along each of the dimensions, the easier they will be to perform concurrently. The more tasks compete for different resources along the dimensions, the harder they will be to perform. It is for example easier to focus on listening on a verbal instruction and looking at a picture, than it is to listen to a verbal instruction and reading a text. This is because listening to the instruction and reading the text requires the same resources from the coding dimension, i.e. linguistic coding. This as opposed to listening to the instruction and looking at a picture, which requires different resources from the coding dimension, i.e. linguistic and spatial coding respectively. Hence, considering these dimensions when designing how tasks are to be performed can help exploit cognitive resources in an efficient way and avoid exhausting performers attending to a large amount of information concurrently.

2.2 Media formats

There are different kinds of information and content that requires maintained focus for a longer period of time to be assimilated, and different types of content providers also experiment with strategies and formats to be able to reach out to their target groups. Hence, the 360 video and interactive street view formats are just two among many other ways to present information. What makes
these formats interesting is that they provide the receiver with a rich and large amount of stimuli through various media types, which should suit people with digital lifestyles with media multitasking tendencies. Having this in common, the formats do also differ on other parts which makes them interesting to examine together. Following is a description of the two formats.

### 2.2.1 360 video

A format of presenting content that is increasing in digital contexts is video [13, 14]. The 360 video can be described as an extension of the video format, which adds a new feature. Using technology that captures the whole 360 degrees perspective of scenes, 360 videos can then also be viewed in 360 degree perspectives. As opposed to "common" video where only one perspective is captured and viewed, the one directed by the person filming, 360 video lets the viewer decide the perspective to be shown by controlling the direction of the view using simple interactions. Figure 1 shows what it may look like when shifting from one perspective to another. 360 video is also gaining usage and big media platforms, such as YouTube [15] and Facebook [16], initiated support for the format in 2015. Another example is news agencies that have started to experiment with 360 videos as a format of presenting news stories [17–19].

### 2.2.2 Interactive street view

The approach of the interactive street view format is similar to that of the 360 video. Material for this format is also generated by capturing 360 degrees perspectives of scenes, but in still images. These are then put together to a resulting 3D world that can be explored by "walking around" to different places in the world. Places can be explored by "looking around" using simple interactions in the same way as in the 360 video, which is illustrated in Figure 1. One extensive example of how this technique has been used is the Google street view application [20], which lets the viewer wander through many parts of the planet. It is not focused on a concentrated content in the same way and the world is huge, but it is an example that illustrates the "walking around" feature. Hence, a difference from the 360 video format is that there are several paths to take through the story, and it is the viewer who decides which paths to take between places. The designer of the content does lay out the paths through the content, but the viewer decides how to move along them. In addition to the "walking around" feature, places can also be explored by interacting with things present in places and a speaker voice and/or sound can also give additional information about a place when stopping at one.

### 3 Method

To identify advantages and disadvantages of the formats’ ability to keep people with digital lifestyles focused, the study was conducted by letting users consume
information through the formats. This was done by primarily consider the users’ subjective experience of their focus. As a complement objective data that gave an indication of how the content reached out to the users was also gathered, by considering what the users could recall of the content. The material selected for testing the formats were two news reportages about Burma, a 360 video [22] and a street view reportage [21] both made by professional journalists wanting to communicate the situation in the country. The 360 video is about five minutes long and the street view reportage was delimited to a section that also took about five minutes to go through. Focusing on the same thing for five minutes should be a quite long time for people with digital lifestyles [2].

3.1 Setup and structure

A test leader guided the user through the test session, and one user at the time viewed the reportages in a calm environment using a laptop. Hence, the full focus was on the reportages and the study did not try to take into consideration external distraction that may occur in everyday life, but concentrated on the characteristics of the formats.

The study was divided into four parts, where the first two parts consisted of the user testing each of the two formats, the third part consisted of the user answering questions about their media consumption and the fourth part consisted of a short interview.

The first two parts evaluating the formats were structured in the same way. The user was first given an introduction to the reportage, both shortly to the subject of the reportage and how to use and navigate the format. For the 360 video reportage the user was told to use the keyboard and mouse to "look around" as much as the user felt like, and for the street view reportage the user was told to freely explore the reportage and tell the test leader when feeling done with viewing the reportage. After taking part of the reportage, the user answered questions about what had been said in the reportages and about the experienced ability to focus on the reportages.
3.2 Measurements

To measure the user’s focus the user’s subjective experience of the focus was primarily considered, and to objectively see that the user also did attend to the content the user also answered questions about the content. Since the study focus on people with digital lifestyles the user also estimated their media usage.

3.2.1 The user’s subjective experience

The user’s subjective experience of the formats was identified by the interview at the end of the test session. The interview was semi structured and grounded in the questions (1) How did you experience your focus during the reportages? (2) Would you prefer one of the reportages in front of the other?

The subjective experience was also identified by letting the user estimate their perceived workload when viewing the reportages. This part was constructed with inspiration from the lighter version of the NASA Task Load Index [23], which lets the user rank different aspects of how heavy the workload is when doing a task, such as level of effort and frustration. Some modifications of the questions the index uses were done in order to make them more explanatory in relation to the task of viewing the reportages. A complementary question about the user’s interest in the content of the reportage was also added, to make it possible to see the potential effect of this on the perceived task load. The resulting questions are stated with the results in the following section. The user answered these questions directly after viewing the reportage.

3.2.2 Comprehension of the content

To measure how much the user assimilated of the information presented in the reportages, the user answered questions about things that had been said in the reportages. The questions were formulated as multiple choice questions and the user also ranked the confidence in the answers, this to get an indication of the degree of guessing. To focus on the user’s attention questions targeted facts stated, and not things that needed to be interpreted and deeper reflected upon. There were four questions for each of the formats, and the user answered these directly after viewing the reportage.

3.2.3 Media consumption

To get an idea of the user’s media consumption habits, a scaled down version of the Media Multitasking Index developed by Ophir, Nass and Wagner [3] was used. The user estimated the amount of usage of different types of media, and how often several types of media was used at the same time.
4 Results

Eleven users participated in the study, which was conducted in 2016 in Sweden. The study was conducted in Swedish, and the results have been translated to English.

4.1 The users’ subjective experience

The following lists describes experiences of the formats expressed during the interviews. When asked about which format the users preferred, four users did not want to choose one in front of the other, four users chose the 360 video format and three users chose the interactive street view format.

Experiences of the 360 video reportage
- It was nice to be guided through the video and at the same time be able to "look around"
- It was hard to focus when the tempo in the video sometimes felt stressful
- Scene switches was sometimes stressful, and could happen while panning the view
- Video gives a real time experience which is fun and engaging

Experiences of the interactive street view reportage
- It was good to be able to decide the tempo and decide when move between the places in the reportage
- Making decisions about where and when to move was laborious
- Keeping track of what sections that had been viewed was confusing
- It would be easier to focus if being more used to how to navigate the reportage
- The speaker voice was confusing, it was hard to know when it started and was finished and it sometimes interrupted one section of speech with another

Experiences of both formats
- Being active by "looking around" made it easier to focus
- Even if it was tempting to "look around", doing so sometimes resulted in lost focus on what the speaker voice said, especially when there were no connection between the visual view and what was said
- The reportages were fun and engaging
- Viewing the reportages requires the user to focus on the reportage, it is hard to do something else at the same time in a feasible way

Task load ratings

Figure 2 - Figure 7 presents the results of the task load ratings of viewing the reportages, and the results for the additional question on the interest in the subject of the reportages are presented in Figure 8.
Fig. 2. How was it to concentrate?

Fig. 3. How demanding was the physical activity (using the keyboard and mouse)?

Fig. 4. How stressed were you about the pace of the reportage?

Fig. 5. How was it to grasp what the reportage presented?

Fig. 6. How did you feel?

Fig. 7. How much effort did the reportage demand?
Fig. 8. Did you find the reportage interesting?

4.2 Comprehension of the content

For the questions about the content of the reportages, the average for the 360 video reportage was 2.4 right answers out of 4 possible, and the average for the interactive street view format was 2.7 right answers out of 4 possible. An average confidence level of the answers is presented in Figure 9.

4.3 Media consumption

The user reporting the highest consumption of media usage estimated 22 hours of total media usage per day. The user reporting the lowest consumption of media estimated 4.7 hours of total media usage per day. The median media usage was 11.9 hours per day. When asked about how often they used two or more media types concurrently on a scale from 1 to 5 where 1 is "seldom" and 5 is "often", all users gave an answer of 3 or higher.

5 Discussion

The result shows that the experienced focus when viewing the reportages varied much among the users. Advantages and disadvantages of the formats can be recognized in the results and are associated with characteristics of the formats. Although the users’ reported media usage also varied, no user reported very low media usage. It can be compared to an extensive study of the media usage of American teenagers [6] where 8.4 hours of media usage per day was considered to be high. All users also reported often using several media types concurrently, and thus the results should be relevant for the purpose of targeting users with digital lifestyles.
5.1 Rich media elements

A characteristic for both of the formats is that they use rich media elements to present information to the user, with images, sound and speaker voices. For both of the reportages, an experience expressed was that they were engaging and that they kept the user’s focus in a way that made it hard to do other things at the same time in a feasible way. Thus, such an experience may indicate that the way the formats present information well suits a breadth-biased attentional filter, since the user is given a large scope of information to pay attention to, a scope that is large enough to not make room for more information from other sources, and does not require the user to focus on a single particular piece of information in detail. Though, it is important to emphasize that the study was conducted without consideration for possible other competing stimulus that may be present in everyday life. The results from how well the users comprehended the content show that users also seem to have assimilated information from the reportages and that they succeeded to convey at least some of the content to the users.

Users did not express difficulties taking part of the different media elements concurrently, and according to Wickens [12] the elements should be a good multi-tasking combination since they use different resources along the different dimensions in the Multiple Resource Theory. One exception was that "looking around" sometimes could result in lost focus of the speaker voice when these two did not have a clear perceived connection. Since these activities as well are well spread out along the different dimensions, by together utilizing all the different resources of cognitive processing (listening to the speaker voice and perceiving the visual content), selection and execution of action (select where to "look" and execute action), spatial processing (perceiving the visual content), linguistic processing (listening to the speaker voice), auditory perception (listening to the speaker voice), visual perception (viewing the visual content) and focal and ambient vision (viewing the visual content), the difficulty may occur because of limited resources rather than poorly utilized resources.

Hence, since the rich media elements indicated to enhance a breadth-biased attentional filter and together compound a good multitasking combination, they seem to be an advantage for the formats.

5.2 Characteristics affecting perceived task load

The task load ratings are widely spread along the scales of all questions for both formats, and this shows that the perceived task load when viewing the reportages differed much among the users. The interviews identified some characteristics that were experienced to increase the task load, and as proposed by Lavie et al. [10] these were also perceived to decrease the ability to maintain focus.

"Looking around". A characteristic shared by the formats is the ability to "look around" while a speaker voice is talking, and thereby gather more perceptual stimuli. This was experienced as an advantage for keeping focus, but it was also expressed that it sometimes could result in lost focus of what the
speaker voice talked about when what was shown visually was perceived as more arbitrary content, and not strongly connected to what the speaker voice said. In these situations the user could start thinking about what to look at and trying to make sense of why the specific visual content was viewed, and then the "looking around" resulted in cognitive processing rather than just getting more perceptual stimuli. Thus, this resulted in increased cognitive load and also decreased ability to focus. This as opposed to views where the visual content and what the speaker voice said was perceived as clearly connected, where "looking around" gave the user more perceptual stimuli without also increasing the cognitive load, since the user then did not question the purpose of the visual content. Hence, it is an advantage that the visual content of views strongly relates to what the speaker voice says to help the user to focus.

**Tempo.** The tempo in which the reportages proceeds is another factor that was found to affect the focus of the user, and this is a thing that varies between the two formats. The characteristic of the 360 video format is that the video goes on and presents new information as time lapses, while the characteristic of the interactive street view format is that the user decides when to proceed with new information. Some users found it stressful when the 360 video was too fast, and they felt that is was easier to focus when they were able to decide the tempo themselves, as they did with the interactive street view format. Some other users found it stressful to decide when to move on when viewing the interactive street view reportage, and found it easier to focus when viewing the 360 video where they did not need to bother about that. Hence, an advantage of the 360 video format is that the user can be guided through the story but with this comes the difficulty to find a tempo that different users are comfortable with and that is not perceived to increase workload. An advantage of the interactive street view format is instead that the user can adapt the tempo, but it then becomes a challenge to make the decision easier so that the decision process do not result in too much extra workload.

**Scene switches.** Another thing that was expressed to interrupt the focus while watching the 360 video reportage, was that scene switches sometimes happened while the user was panning the view. Even if it did not happened many times during viewing the reportage, it still annoyed the user and forced the user to reset for the next scene before being done with the current. Hence, it is a disadvantage for the 360 video format if the user’s panning do not time well with the scene switches, and those situations may put extra cognitive load on the user in the form of irritation and stress which also results in lost focus.

**Navigation.** For the interactive street view reportage, the navigation of the reportage was expressed to be strenuous and result in lost focus. It was hard to know what places that had already been visited, and since this information was not provided in the interface of the reportage, the user needed to keep track of where they had been and not and which information that had already been viewed. A thing that worsened this was the speaker voice, which confused the user about what information the user had already heard by being ambiguous about when it started and was finished, and it also sometimes interrupted one
section of speech with another when the user interacted with the reportage. These things were expressed to increase the perceived workload and result in lost focus of the content of the reportage, and hence they are a disadvantage of the interactive street view format. It was also expressed that it would be easier to focus if being more used to navigating the reportage, which further indicates that the navigation of the interactive street view format was experienced as non-intuitive and may require its user to be experienced with how the navigation work if it should not result in increased perceived cognitive load.

5.3 Outer factors affecting the focus

The users' interest in the content of the reportages did also vary quite much, though few people had very low interest in the reportages. The interest in the reportages may have had a positive effect on the users' ability to focus, which should not be credited to the formats and the way the content were presented. It is also important to emphasize the settings in which the users viewed the reportages, a test situation may affect the results and some users are more affected by a test situation than others.

6 Conclusion

Both of the formats have potential to capture users' focus for a longer period of time, but that also depends much on how the content is using the rich media elements and how contended the individual user is with the tempo and the activity that is demanded when consuming information through the formats. Hence, although it often is important that the content is deliberately arranged and to acknowledge users' different preferences, this should be especially emphasized when using these formats to present information. The disadvantages of the scene switching in the 360 video format and the navigation of the interactive street view format are however things built into the formats and not affected by the content or the user in the same way, but are also things to be aware of when using the formats to present information.

This study only recognize some aspects of presenting content with the formats, and there are many further things that would be interesting to evaluate. One thing would be to investigate potential improvements on the scene switches and navigation problems of the formats. It would also be interesting to evaluate further interface possibilities and how they affect the user's focus, such as interfaces on tablets or phones that use the sensors of the device to let the user "look around" by moving the device. Except from the focus of the user, it would also be interesting to widen the perspective and consider content providers' thoughts on the formats.

References

Using smartwatch as an aid for people with diabetes

Elin Hörfeldt
Department of Applied Physics and Electronics
Umeå university, Sweden
id12eht@cs.umu.se

Abstract. 450 million people in the world are living with diabetes. Living with diabetes will include blood glucose tests and injections with insulin to keep a healthy blood glucose level. There are medical aids such as insulin pumps and continuous glucose monitoring systems (CGM) that simplify the life for a person with diabetes. These contain a lot of information and might sometimes also be placed inaccessible for the user that can face troubles with controlling their blood glucose. With a questionnaire, 16 people with diabetes answered what information they see as most important from the insulin pump and the CGM. Remaining insulin, active insulin, and the basal rate was the three most important functions the test subject wanted to know from the insulin pump. Blood glucose curve, current blood glucose level, and increasing/decreasing blood glucose were the most important information from the CGM according to the test subjects. This information can be shown on a smartwatch and can simplify the blood glucose control for a person with diabetes. A prototype of a smartwatch with the resulting information was made by design guidelines from literature research.

1 Introduction

Diabetes mellitus is a chronic disease, and there exist different kinds of diabetes. The two most common ones are called diabetes type-1 and diabetes type-2 and the major difference between these two is that people with type-1 diabetes has no insulin production at all, and people with type-2 diabetes have some small production but not enough to meet the needs of the whole body [1].

According to the Swedish Diabetes Association, 4% of the Swedish population is diagnosed with diabetes mellitus [2]. Around 450 million people in the world are living with diabetes, and the International Diabetes Federation predicts that there will be 642 million inhabitants with diabetes in 2040 [3]. These numbers affect a lot of people that will have to rearrange their lives. Living with diabetes will include injections of insulin, blood glucose tests and adjustment to avoid getting too high (hyperglycemia) or too low (hypoglycemia) in blood glucose, and this is not always easy tasks to perform. In Olohan’s and Zappitelli’s work [4] they tell the story of 15-year old Kirsten that has had diabetes type-1 for five years. Kirsten often gets a high blood glucose level before dinner which probably
occurs because she usually eats pizza with her friends after school. Even if she try to correct her blood glucose with injections with insulin, it is hard to find the right amount to inject which results in high blood glucose. Kirsten also exercises after the dinner and then she needs to eat something extra before to avoid getting too low in blood glucose. But even this is a hard adjustment for Kirsten, and she often ends up with hypoglycemia (too low blood glucose) a few hours after the exercising or during the night. Hypoglycemia after bedtime makes Kirsten wake up sweaty and with a headache. Luckily there are technical aids that can help Kirsten and other with diabetes to keep stabilized blood glucose levels. With an insulin pump, people with diabetes can calculate how much insulin they need to the number of carbohydrates they eat and a rearrange in the basal insulin (the background insulin) in the pump can minimize low or high blood glucose levels. With a Continuous Glucose Monitoring (CGM) a person with diabetes will also be able to see and follow their blood glucose curve [4].

People using insulin pumps and CGM’s show to have lower glycated hemoglobin levels (an average of the blood glucose level over the past 2 to 3 months) compared with people with injections [5]. And a lower glycated hemoglobin level is shown to give a better quality of life for people with diabetes [6]. Insulin pumps’ and CGMs’ simplifies the life for people with diabetes but can it be even more simplified? The insulin pump is usually hidden in the pocket or might even be inaccessible hidden underneath the clothes which can complicate the usage of the insulin pump in the public [7]. How would it be for the user if the information was shown on a smart watch instead?

1.1 Objective

The objective of this study is to investigate how information from the CGM and the insulin pump can be visualized on a smartwatch in a good usability way for people with diabetes. The insulin pump contains a lot of information like current active insulin amount, history, the condition of the pump and much more. The CGM also have different information about the blood glucose levels. Designing for a smartwatch will include a large downscaling of information [8], so what is the most important information for a person with diabetes to know at the moment? To find out what information the user want to see on the smart watch, we will study what the test subjects see as most important to know at the moment. We also want to find out what the guidelines for designing smart watches are and from the result, create a prototype. This study is focused on the people with insulin-depended treatment, i.e. people with diabetes type-1.

2 Theory

For designing this approach, we will have to consider two major areas. What is diabetes and what are the medical aids? And how do we implement good usability on smart watches with small screens?
2.1 Diabetes

Diabetes can occur in different kind of types, but the common factor is that the blood glucose is higher than normal (hyperglycemia). Latent Autoimmune Diabetes of the Adult (LADA), Gestational Diabetes, Diabetes Type-1 and Diabetes Type-2 are the known diabetes diagnoses. Diabetes type-2 is the most common type, about 90-95% of the people diagnosed with diabetes, has type-2. Type-2 occur when the cells in the body get insulin resistance and will therefore suffer from some relative insulin deficiency. Type-2 usually occur by overweight, sedentary lifestyle, and inheritance. People with diabetes type-2 do not usually need to inject insulin for survival and will best stabilize their blood glucose through exercising and weight loss. About 5-10% of the people with diabetes has type-1. It is an autoimmune disease where the body self-destruct the beta-cells of the pancreas (the ones who produce the insulin) and the cells will therefore get absolute insulin deficiency and insulin by injection is required for survival [9].

2.2 Insulin pump

An insulin pump is connected via a tubing to a small needle that is placed under the skin [10]. The pump delvers insulin at a basal rate, also called background insulin. Every person needs background insulin. A healthy person gets it from its pancreas, but for a person with diabetes, the insulin pump is the closest alternative to a real pancreas. The basal insulin is given through the night and day to get a healthy body function. By using a temporary basal insulin, the amount of the background insulin can be lowered or higher when the person, for example, is exercising or is sick. When the person eats, he or she takes a bolus which is an amount of insulin to regulate the higher blood glucose level that occurs after eating, or whenever the blood glucose level is too high. With personal settings on the insulin pump the pump can calculate how much insulin the person should take to the meal depending on blood glucose and carbohydrates, how much insulin from the bolus dose that is still active in the body and much more [11]. The advantages with an insulin pump compared with normal injections are that the small needle under the skin that is connected to the insulin pump can be changed every 2-3 day when a person with only insulin injections have to take around 3-4 doses a day. The basal rate from the insulin pump optimizes the glycemic control overnight and decreases the risk of low blood glucose levels at activities and exercising. The person also gets more freedom of eating when the insulin pump can give different kinds of boluses depending on the food. These are the normal bolus (when the dose is given directly), square-wave (when the dose is distributed over a particular time, usually between 30 min and 2 hours) and dual-wave bolus (which is a combination of a normal bolus and a square-wave bolus) [12].

Pickup et al. [13] show that the blood glucose control is better at a person with an insulin pump than insulin injections. With an insulin pump, the person also needs less insulin to achieve the blood glucose levels.
2.3 Continuous Glucose Monitoring (CGM)

A frequent self-monitoring of blood glucose tests show a better glycated hemoglobin level (HbA1c) [14]. HbA1c is an average of the blood glucose level over the past 2 to 3 months. A Continuous Glucose Monitoring system (CGM) receives up to 288 glucose values every day and will fill in the gap between the self-monitored fingersticks glucose tests [15]. The CGM consists of three parts (see Figure 1), a transmitter (1), a glucose sensor (2), and the insulin pump (3) or another monitoring device [16].

Fig. 1. The continuous glucose monitoring system from Medtronic MiniMed 640G System. The three components are the transmitter (1), a glucose sensor (2), and the insulin pump (3) [15].

The glucose sensor is placed under the skin and measures the glucose levels in the body. The transmitter connects to the sensor and sending the glucose values to the insulin pump that shows it for the user in a graph. The CGM needs to be calibrated every twelve hours with a fingerstick blood glucose test and can be worn for six days before the sensor need to be changed. With personal settings, the CGM can give alarms when the blood glucose levels are increasing or decreasing too fast or when a certain glucose level is reached [15].

2.4 Smartwatch

A smartwatch is similar to a smartphone except that it is smaller and are worn on the wrist. Usually, the smartwatch is connected to a smartphone and can have apps, receive notification and even calls [17]. They are sold in different shapes (i.e. round, square, rectangular) and sizes. The size of an Apple Watch is 38 mm or 42 mm and Samsung Gear S that is 42.3 x 49.8 mm. A smartwatch can
have both touchscreen, buttons and also other functions as heart rate monitor, accelerometer, speakers, and microphone [18][19].

![Apple Watch and Samsung Gear S2](image)

**Fig. 2.** The Apple Watch and Samsung Gear S2 [18][19].

### 2.5 Design principles for smartwatches

One positive aspect of the smartwatch is that it naturally comes in the area of the eyesight [20], which makes it suitable for notifications. A study made by Giang et al. [21] show that driver glance more at their smartwatch than their smartphone during driving. A more challenging aspect of the smartwatch is that it has a small screen so i.e. zooming will not work. The primary function for a smartwatch is to serve a smartphone and only show the most important features, and then redirect to the smartphone [20].

To design a good usability interface, it is important to reduce the amount of information showing on the screen and just visualize the most important information. By decrease the amount of information the need of scrolling will reduce and make a device with a small screen more usable [8].

Apple and Samsung have developed own UI guidelines for their smartwatches. Apples’ design principles are [22],

- Lightweight interactions: the information is easy to access and dismiss.
- Holistic design: the boundaries between the device and software should be blurred.
- Personal communication: be aware of that the smartwatch is carried close to the user.
- Gestures: Taps is action-based, vertical swipe is scroll and horizontal swipe is for showing next or previous page.

Samsungs' design principles are [23],
– Glanceable: should give the user the most important information in a consistent and informative way.
– Context-rich: make it easy for the user to follow the pages.
– Fast and responsive.
– Attractive: make the app draw attention.

3 Method

By using a questionnaire, 16 people with diabetes type-1 answered questions about what information on the insulin pump or CGM that is most important for them. The test subjects were older than 18, and they have either both insulin pump and CGM, just one of them or none. The questionnaires were made in Google forms, and the majority of them was sent out to known test subjects through Facebook. A few questionnaires were given at the diabetes clinic at Norrlands Universitetssjukhus (the university hospital) in Umeå, Sweden.

The questionnaire had multiple answer questions and first the test subjects stated their gender, age and if they have or have had an insulin pump or CGM. The questions they answered was what information they see as the most important information for them to know at the moment. The options the test subject could choose between was based on the functions and information that already exists in the insulin pump and CGM. From the insulin pump, the options were: the status of the battery, the amount of insulin that is left, time and date, basal insulin, active insulin, temporary basal change, other temporary changes, carbohydrate counting, history, the insulin pump version/age/model and own option. The test subject could choose maximum four options but was asked to choose only three if it was possible. The second question was what information from the CGM that was most important for the test subjects. The options were: next calibration, the age of the sensor, transmitter battery, transmitter version/id, blood glucose curve over the latest hours, the current blood glucose level, if the blood glucose is rising or falling and own options. This question also allowed four options to choose but if the test subject could choose only three option they were asked to do that. If the test subject did not have either an insulin pump or CGM, they got the same question but were asked what they thought was the most valuable information for them. The test subjects also answered a question about where they stored their insulin pump or CGM. They were asked if the device was easy to get from, for example, a pocket, or if the device were more hidden, for example, underneath clothes or an own option. A supplementary question was if the test subject saw the location of the insulin pump or CGM as an obstacle for them to look at their blood glucose curve or some other function or information. The options were yes, no or sometimes. At the questionnaire given at the clinic a supplementary question about what kind of diabetes the person had was added, to be sure that the test subjects had diabetes type-1 when they were unknown. By a literature research, guidelines for designing smartwatches was collected. After receiving the answer from the questionnaire, a prototype was made in Adobe Illustrator by the design guidelines.
To notice in this division is that all test subject using blood glucose meters. These should not be confused with CGM’s who measure the blood glucose continuously. Glucose meters is necessary for 100% certain blood glucose values and is a necessary tool for using CGM’s.

4 Result

Total 16 test subjects answered the questionnaire. 11 through Google form and 5 through the clinic. It was 4 men and 12 women where 13 was between 18-30 years old, and 3 was over 30 years old. 3 had an insulin pump, 4 had CGM, 8 had both an insulin pump and CGM and 1 had injections.

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;18 years</td>
<td>18-30 years</td>
<td>&gt;30 years</td>
</tr>
<tr>
<td>Insulin pump</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGM</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>CGM and insulin pump</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Injections</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Women</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;18 years</td>
<td>18-30 years</td>
<td>&gt;30 years</td>
</tr>
<tr>
<td>Insulin pump</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGM</td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>CGM and insulin pump</td>
<td>6</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Injections</td>
<td></td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. The distribution of the test subjects.

The test subjects could choose 3-4 alternative from the insulin pump functions and the CGM information. The three options that got most votes from the insulin pump was remaining insulin amount with 14 votes, the basal rate with 8 votes and active insulin with 8 votes (see Figure 3). The three options that got most votes from the CGM was blood glucose curve over the latest hours with 14 votes, current blood glucose level and increasing/decreasing blood glucose with 15 votes each (see Figure 4).

With the downscaling of information from the insulin pump and CGM, the three chosen options from each device can be shown on an own page on a smart-watch. With design guidelines as horizontal swipe is for showing next page, the information should be glanceable, follow the pages should be easy and the application should be attractive, one example of how the smartwatch could visualize the information is made. The result is shown in Figure 5 and Figure 6. Figure 5 show the three most important information from the CGM. One example shows a stabilized blood glucose level and one when the blood glucose is rising. In
Fig. 3. The most important functions in the insulin pump for people with diabetes type-1. The alternative was battery status (1), remaining insulin amount (2), time and date (3), basal rate (4), active insulin (5), temporary basal changes (6), another temporary change (7), carbohydrates calculations (8), history (9), the pump’s version/age/model (10) and other (11).

Fig. 4. The most important information from the CGM for people with diabetes type-1. The alternative was next calibration (1), age of the sensor (2), the battery of the transmitter (3), transmitter version/id (4), blood glucose curve over the latest hours (5), current blood glucose level (6), increasing/decreasing blood glucose (7) and other (8).
Figure 6 the most important functions from the insulin pump are shown. In one example the remaining insulin amount is still high when the other example indicate that the insulin in the pump is almost finished. Both Figure 5 and Figure 6 represent an own page on the smartwatch where the user will swipe horizontally to switch page.

Fig. 5. An example of how blood glucose curve, current blood glucose level and increasing/decreasing blood glucose can be visualized on a smartwatch. The watch to the left show a stabilized blood glucose progress and the smartwatch to the right show a fast increasing blood glucose progress.

To see if a smartwatch could simplify the blood glucose control for a person with diabetes, a question of how often the test subjects check their blood glucose, where the test subject placed their insulin pump or CGM and if that was an obstacle for the test subject were asked. The result showed that most of the test subjects were looking at their blood glucose curve sometimes, especially when they felt that something was not right. One test subject choose other option which was that person could not see that without a CGM (see Figure 7). 7 test subjects place their insulin pump or CGM in an easily accessible place, and 6 placed it somewhere that it was harder to access. 2 test subjects choose other option that was in an omnipad (a belt special for CGM monitors) and in their bag (see Figure 8). 7 test subject did not see the placement of the insulin pump or monitor as an obstacle to looking at their blood glucose curve. 4 saw it as a barrier sometimes, and 3 choose other option. From the other options, two answered that their CGM was connected to a monitor and they therefore had the monitor and the insulin pump separately, and the third one did not give any answer (see Figure 9).
**Fig. 6.** An example of how remaining insulin, active insulin and basal rate can be visualized on a smartwatch. The watch to the left show that it is a lot of remaining insulin left and the smartwatch to the right show that the insulin in the insulin pump is almost finished.

**Fig. 7.** The result of how often the test subjects check their blood glucose. The alternative was only at warning (1), somtimes and espacially when they feel that the blood glucose level isn’t right (2), the whole time (3) and other (4).
Fig. 8. The result of where the test subjects is placing their insulin pump or CGM monitor. The alternative was easily accessible in a pocket or a belt (1), somewhere it’s a little bit harder to access (e.g. underneath the clothes) (2) and other (3).

Fig. 9. The result of if the test subjects see the placement of the insulin pump or CGM monitor as an obstacle to check their blood glucose curve.
5 Discussion

The result from the questionnaire shows that the test subjects were very united about what information that was most important from the CGM. The votes were a little bit more spread out over the functions in the insulin pump. When the CGMs' primary mission is to measure the blood glucose levels it is not a very surprisingly result. The insulin pump has much more functionality to choose between for the test subjects. Some of the answers can also be a little bit misleading when the test subjects that only had a CGM has a monitor that is "free" e.g. not connected to the body like an insulin pump are. It is therefore much easier to place the monitor in a bag where it will be more accessible than an insulin pump placed closer to the body would be.

After the compilation of the questionnaire, a prototype was made. The prototype of the smartwatch achieved in this study is only one example of how a smartwatch can be designed for this purpose. This prototype was downscaled in information, and tried to be done as simple as possible and following the guidelines. When the screen of a smartwatch is so small, the smartwatch should not have so much "purpose". When the screen is not that easy to interact with it is good that the functions are not too many. That is why only the most important information is visualized on the watch for the user. In real life, when a user, for example, gets an SMS, the user will see the notification on the watch but probably use the smartphone to answer the SMS, since the smartphone has a bigger screen as well as a keyboard. The same will apply for people with diabetes. The smartwatch will show the most important information, but active choices will be made on the insulin pump. The smartwatch can be perfect for the users to get notifications about their blood glucose levels and also be more aware of it. Vervloet et al. [24] conducted a study were people with chronic medication got reminders from electronic devices to take their medication. Electronic devices (SMS, pager messages, electronic reminder devices) showed to be good reminders. This type of reminder system could be associated with a smartwatch when it is worn on the wrist and in eyesight. A smartwatch could therefore help people with diabetes to be "reminded" about their blood glucose control. Another positive aspect of a smartwatch is that it is accessible to the user. Even if the majority of the test subjects did not see the placement of the CGM monitor or insulin pump as an obstacle to check their blood glucose, there was still some test subjects that did. And for these, the placement of a smartwatch could simplify the blood glucose control. To consider in this study is that the prototype is only done by design guidelines and by asking people with diabetes what they see as important to know from their insulin pump and CGM. The prototype is not tested, but just a suggestion of how it could look like. A future work in this area could be to see how the blood glucose control affects in real life by using a smartwatch.

5.1 Limitations

Some of the questions may have been a bit unclear for the test subjects when some answers did not respond to the question. I.e. for the question about how
often the test subject looked at their blood glucose curve, one answer was that the test subject had not used the combination of CGM and an insulin pump. This answer came from the questionnaire that was left at the clinic. The difference with the questionnaire online and the ones at the clinic was that online, the questionnaire form could give different questions depending on the earlier answers. This could not be done on the physical questionnaire, and the question should have been formulated as "where do you place your insulin pump/CGM monitor?" instead of just "where do you place your insulin pump?".

6 Conclusion

A smartwatch can be used to visualize blood glucose curve, current blood glucose level, and increasing/decreasing blood glucose from a CGM. On another page, the smartwatch can visualize remaining insulin, active insulin, and basal rate for the user. This is a complement to an insulin pump and a CGM and can simplify the blood glucose control for a person with diabetes type-1. A smartwatch is accessible and in eyesight, and the user will easily see when the blood glucose is not right.

References


Examining the potential of smart glasses as an assistive technology for dementia patients

Karolina Jonzén
Department of Applied Physics and Electronics
Umeå University, Sweden
kajo0136@student.umu.se

Abstract. With the rapid growth of the world’s elderly population, dementia has become a global health care priority, due to its disabling and dependency-increasing effects on the sufferer. At the same time, the recent years of fast advancement in computing science and interface design has opened up new opportunities for developing more powerful technologies to assist these individuals in maintaining their independence. This paper examines the potential benefits, challenges and future possibilities of using smart glasses as an assistive technology for dementia patients, based on a literature review and an interview with domain experts. While several potential benefits were found, such as decreased interaction complexity and promising uses for navigational support, memory assistance and cognitive training, addressing ethical concerns and involving dementia patients in the studies remain as important future challenges to overcome.

1 Introduction

Dementia is a brain condition that causes a gradual decline in cognitive abilities, such as memory, language and orientation, and is one of the most common causes of disability and dependency among elderly [1]. In fact, there are estimations of 47 million people all over the world suffering from it today, and with the rapid aging of population, this number is expected to be tripled by 2050 [2]. Since a cure for dementia related diseases is yet to be found, this has motivated greater investments and research to find more effective ways of assisting both patients, caregivers and family members [3].

One way of increasing independence in everyday activities among disabled people is through the use of Assistive Technologies (AT), defined by Marshall [4] as "any item, piece of equipment, product or system, whether acquired commercially, off-the-shelf, modified or customized, that is used to increase, maintain or improve functional capabilities of individuals with cognitive, physical or communication disabilities". ATs can support the cognitive disabilities of individuals suffering from dementia in many different ways. For example, automated prompts and reminders, such as voice messages to remind of appointments and medication taking, locator devices, like electronic tags on keys and wallets to
find misplaced items, and communication aids that simplify contact with relatives and friends [5].

The recent years of fast development in computing science and interface design has opened up for new opportunities and technical innovations with possible applications in this field - one being smart glasses. In this paper, smart glasses are referred to as a miniaturized wearable computer that adds digital information to what the wearer sees [6]. By being hands-free, always on and able to recognize the user's surroundings and activities, smart glasses could become a powerful aid for cognitively impaired people and their caregivers. However, their potential in the field of dementia care is still relatively unexplored.

1.1 Objective

This paper aims to make a contribution to the development of modern technology in dementia care, by examining the potential benefits, challenges and future possibilities of using smart glasses as an AT for dementia patients. More specifically, it seeks to do this by not only looking into what the technology has to offer, but also by taking to account the human factors that need to be considered when designing for this target group. Furthermore, the paper provides an overview of related work conducted in this field, as well as experts' opinions on the subject given in an in-depth interview.

2 Background

This section gives a brief introduction to dementia, assistive technologies and smart glasses.

2.1 Characteristics of dementia

To be able to provide a good care for the patients, it is important to understand the nature of the disease. Sandilyan and Dening [7] divide the symptoms of dementia into two broad groups: cognitive and non-cognitive. This study will however focus on the cognitive symptoms. While there are individual differences and the symptoms vary depending on what type and stage the dementia is at, the authors describe several common problem areas affecting different everyday activities:

- **Memory**: difficulty learning new information, misplacing items around the house, repeating questions or forgetting appointments.
- **Orientation**: forgetting what time it is, where they are or who they are with.
- **Language**: difficulties naming objects or people, following a conversation or trouble reading and writing.
- **Actions**: inability to perform complex actions, e.g. buttoning a shirt or brushing teeth.
- **Recognizing sensory information**: problems recognizing objects or people, despite having no visual defects.
The most common form of dementia is Alzheimer’s Disease (AD), accounting for 60 to 80 percent of all cases, according to the Alzheimer’s Association [8].

### 2.2 Assistive technologies in dementia care

In recent years, there have been a lot of research on more innovative and modern approaches to the in-home care of dementia patients by the development of Intelligent Assistive Technologies (IATs). Bharucha et al. [9] define IATs as technologies that support physical or cognitive impairments by sensing and responding to the user’s needs, and adapting to changing situations. In 2009, the authors conducted an extensive research and found a total of 58 examples of such emerging technologies with potential application in dementia care. The examples included everything from automated fall detectors and fabrics with embedded sensors for physiologic monitoring, to adaptive memory aids and wearable devices for task identification and guidance. Although a wide range of IATs was found, the authors concluded both technological and ethical challenges needed to be overcome and requested more clinical studies involving persons with dementia.

The ethical issues regarding the use of Assistive Technologies (ATs) in the field of dementia have been pointed out in several other studies as one of its main future challenges. Zwijsen et al. [10] argue that research on ATs has focused more on improving technical aspects of the devices rather than prioritizing their ethical implications. The authors raise ethical concerns such as privacy issues in recording, monitoring and tracking devices, the possible loss of human contact and the problems of informed consent, in which the patient gives its voluntary acceptance of a medical action, after being told and aware of the plan, its risks, benefits and alternative approaches [11].

Cash [12] discusses how the cognitive problems in dementia can complicate the process of informed consent, but points out that the communication skills of people with dementia should not be foreseen. She additionally emphasizes the importance of involving them in the decision making to make sure the technology meets their needs.

**Guidelines** Designing ATs for people with dementia can be especially challenging due to their cognitive impairments and the degrading nature of the disease. This section presents design guidelines that consider these issues, found in previous researches where dementia patients were involved - an important ethical consideration as stated earlier.

Hyry et al. [13] present the following requirements for the design of ATs for dementia patients, based on a two-year long case study by Nordic countries [14], that included 29 people suffering from dementia.

- *Introduce the technology in the early stages of the disease.*
- *Combine device functionalities.*
- *Less the amount of devices.*
- *Create adaptive and tailored user interfaces.*
Implement tele-presence assistance.

Mayer and Zach [15] developed a mobile touch screen to assist people with dementia using the concept of participatory design to actively involve the users in the design process through interviews, observations and workshops. A summary of the guidelines resulting from the study is seen below.

- Use familiar and non-abstract design components to evoke understandable associations.
- Minimize complexity and choice and emphasize clarity and simplicity to ease decision making.
- Evoke curiosity and use a non-stigmatizing design (i.e. don’t put the disease in focus) to ease acceptance.
- Use positive and supportive feedback that addresses all senses to create comfort.
- Promote learnability and memorability since people with dementia might have to re-learn the interface each time they use it.
- Involve caretakers and relatives to help the patients use the system as the disease progresses.

2.3 Smart glasses characteristics

The most influential and commercially available model of smart glasses today is Google Glass. The Google Glass system consists of an optical head-mounted display that shows information in the user’s field of vision and includes functions like taking pictures, making video calls and running apps [16]. The user interacts with the system through voice commands, head-gestures and a touchpad placed on the side of the frame [16]. Figure 1 below illustrates its main components.

![Google Glass components](image)

Fig. 1: Illustration of Google Glass components based on original image by Tim Reckmann [17] [CC BY-SA 3.0] and infographic by Missfeldt [18].
Techniques Following are explanations of various technical concepts related to smart glasses.

- **Wearable technology**: electronics and computers integrated into different accessories, such as watches, glasses, clothing and jewelry, designed to be worn comfortably on the body and used in people’s daily life [19].
- **Computer-mediated reality**: technologies where information is either added, subtracted or manipulated from one’s perception through the use of a wearable computer or hand-held device [20]. Following are two examples of such technologies, as described by Gaukrodger and Lintott [21]:
  - **Virtual Reality** (VR): allows the user to move and interact in a simulated three-dimensional environment.
  - **Augmented Reality** (AR): incorporates the real environment with virtual objects, opposed from VR where a complete virtual environment is constructed. The Google Glass system uses this technique to present digital information to the user’s field of view.
- **Context awareness**: the use of non-explicit user and environmental input, often acquired through sensors, to determine the user’s current activity, which can help to reduce interaction complexity in task-support applications [22].

Applications Smart glasses and other head-mounted display systems have, to date, mainly been developed for the general public and for more specific purposes in fields such as the military and healthcare [23, 24].

3 Method

The study was carried out by analyzing findings from a literature review and an interview with domain experts, described in more detail below.

3.1 Literature review

A literature review was conducted to find existing research regarding the use of smart glasses as an assistive technology for dementia patients. The review was carried out through an online search in databases where peer-reviewed material was prioritized, in order to ensure high quality and creditability.

Due to the limited number of identified articles in this area, the search was broadened to include studies on similar technical solutions, using either computer-mediated reality or any kind of head-mounted display system, preferably targeting dementia patients, but elderly and cognitive impaired people in general were also considered.

3.2 Interview

To gain further insights about the potential of using smart glasses in dementia care, an interview with domain experts were conducted. Except from getting
expert opinions regarding the use of smart glasses for dementia patients, the interview was also useful in gaining a deeper knowledge of how assistive technologies in general are being used and developed in dementia care.

A semi-structured format was chosen to guarantee collection of relevant data without limiting the possibilities of new leads, as suggested by Bernard [25]. The participants were interviewed in group, mainly due to time limitations on both parts. Group interviews can also create a group dynamic that helps to facilitate the discussion [26]. With permission from each participant, the interview was audio recorded in order to capture all details and be able to remain focused on the dialogue.

Participants Two occupational therapists with experience of working with assistive technologies for elderly and people suffering from dementia or cognitive impairments participated in the interview and were recruited through email communication. The invitation to participate was given to a larger number of therapists, which resulted in these two being able and willing.

Questions A set of prepared questions, presented below, were constructed and used as an interview guide to ensure the main goal of the interview was addressed.

- What do you think of the idea of using smart glasses as an assistive technology for dementia patients?
  - What do you think could be the benefits of using such a technology?
  - What do you think could be the challenges of using such a technology?
- How are assistive technologies being developed and used for dementia patients?
  - What factors are important to consider when designing for this target group?

Procedure The interview was carried out in Swedish, at the participants’ workplace, and lasted for approximately one hour, which also included a demonstration of their current repertoire of cognitive assistive technologies. As recommended by Doody and Noonan [27], the interview started with the participants being informed about the purpose of the study and assured of their confidentiality and anonymity.

Before the question session began, a promo video of the Google Glass application Trace, by Ward et al. [28], was shown to give a more comprehensible introduction to smart glasses and their possible potential and application in dementia care. The creators describe the app as a tool aiming at supporting people in the early stages of Alzheimer’s disease, by combining Google Glass and a tablet device. In the 3-minute long video, it is demonstrated how the system can be used to ease everyday life, as the glasses provide the user with reminders of meetings, face recognition of relatives and friends, navigational directions and a camera to capture moments during the day that later can be viewed in a timeline on the tablet (see Figure 2 on next page).
4 Results

This section presents the main findings from the literature review and the conducted interview with domain experts.

4.1 Literature review

While no published research on the specific use of smart glasses in the field of dementia care was found, several studies were identified exploring the use of related techniques with computer-mediated reality and Head-Mounted Display (HMD) systems for cognitive assistance and training.

In 2015, Garcia et al. [29] reviewed the recent and current use of Virtual Reality (VR) for Alzheimer’s Disease (AD) applications. The authors present different studies where VR systems are developed for diagnosis and cognitive training of mild cognitive impairment and dementia patients, where focus lies on everyday activities like navigation and orientation, face recognition and cognitive functionality. The authors suggest future applications in this field should exploit the available modern techniques and prioritize the specific needs of AD patients and how their symptoms develop over time. Furthermore, the authors point out Augmented Reality (AR) as an upcoming useful tool for AD.

GenVirtual, by Corea et al. [30], is an AR musical game proposed to help people with learning disabilities by stimulating the memorization of colors and
sounds. Virtual cubes in different colors are projected on markers, each associated with a specific musical note. During the game, the cubes light on according to a musical sequence that gets expanded as the user stepwise is to remember the musical notes played, by using their hands or feet to interact with the cubes. The authors suggest that using VR and AR systems in health care is especially beneficial, opposed to traditional treatments, since they allow the user to interact in a natural way. Additionally, the authors point out that with the interfaces being virtual, the risks of patients hurting themselves are reduced and the stimulus generation can be controlled by therapists.

Another example of an AR game in this field is CogArc, under development by Boletsis and McCallum [31, 32]. The authors describe the game as a tool for cognitive training and screening, primarily targeting elderly players above 60 years old. CogArc consists of a collection of cognitive mini-games, such as matching shapes, memorizing patterns and forming words through manipulation of physical cubes placed on a desktop. The AR game content is viewed through a tablet PC mounted on an adjustable arm to enable use of both hands, though the authors point out AR glasses would be the ideal solution. The game was tested on five elderly individuals, however not diagnosed with any kind of dementia disease. The results showed overall positive responses, but the authors also found specific problems related to AR, usability, interaction and game design that lead to some negative reactions.

Quintana and Favela [33] developed a system called Ambient aNnotation System, aimed at supporting people with AD and their caregivers by using AR and a tagging system to provide text-based and auditory memory aids. The user wears a mobile phone, hanging from the neck, a wearable cam recorder on one ear and an earphone on the other. The system involves both the person with AD and the primary caregiver, who creates annotations that automatically get presented to the user when approaching a certain object in the environment. While the authors describe how the system requirements were set up based on AD characteristics, it was not tested on the specific target group, but on six graduate students, though under conditions purposed to simulate those of the final user.

Some studies exploring the use of HMDs were also found. Firouzian et al. [34] developed a prototype of a HMD system designed to assist seniors suffering from memory loss in daily navigational tasks. Though not implemented with any computer-mediated reality technique, the system consists of a pair of eyeglasses with a mounted camera, a Global Positioning System (GPS) tracker, a Bluetooth headset for voice communication and sensors like accelerometer and step counter. The main idea of the system is to provide the user with visual cues from a set of blinking LED indicators located on the frames of the lenses, controlled by a remote caretaker. Preliminary results from user tests of the prototype showed that the performance was highly related to the user’s severity of dementia. Personalizing the visual cues based on the user’s visual and orientation abilities was suggested to improve the system.
AI Googles, developed by Nakayama et al. [35], is another HMD system with possible application in memory assistance. The goggles, implemented with a camera and image recognition techniques, identify and name objects in the environment to be displayed for the user in real-time. The authors describe how it for example could be used to find misplaced items, since the system can provide images of where they were last seen. While no user tests on the target group were conducted, real world testing of the system showed great effectiveness in the object identification.

4.2 Interview with domain experts

A summary of the interview with the two occupational therapists is presented below, categorized in the main themes identified.

**Smart glasses: Possible applications and challenges** The possibilities of using smart glasses as an assistive technology for dementia patients was discussed after having watched the video of the Google Glass application Trace. Both therapists found the idea very interesting and could think of many different applications. Capturing daily moments on picture, as shown in the Trace video, was one of the suggested uses: "Think about those being on daily activities and then when they come home and their spouse might ask ‘What have you been doing today?’, then they can easily look together and discuss and get a topic of conversation that works, opposed to not being able to retell anything.”

Furthermore the possible use of navigational support was discussed. While being positive to this, as it could help persons with dementia to stay active without the risk of getting lost, the therapists discussed how the guidance should be presented to the user. One of the them argued that a map, as used in the Trace app (previously shown in Figure 2c), would be to abstract for a person with dementia to understand: "That would need to be complemented with a voice that verbally gives instructions." The other therapist suggested those instructions could advantageously be combined with displaying arrows in the real environment.

Additionally, the therapists mentioned how the GPS function and navigational guidance would be a safety not only for the person with dementia, but also for their relatives, if they could have the location of their demented family member displayed on their phones. However, this kind of tracking products are not currently supported by the municipal and county: "There are no decisions on who is responsible. You can buy [the product] yourself, but not as a refund, so that you could get it as an assistive technology.” The therapists explained how this is mainly due to slow bureaucracy and ethical concerns regarding the right to track someone who may not be fully aware of it.

Using the glasses to recognize people was another function the therapists found interesting: "Because there are many who pretty early suffer from not being able to recognize people and acquaint persons. The family usually works, but when meeting someone out in the city that says like ‘Oh hi, how are you?’"
and then [the person with dementia] can think like 'Oh my God, who is that?' and then they could get some support like 'This is Jill.' Ethical dilemmas were though once again brought up as a challenge, since the therapists felt it would be a bit creepy to have people walking around with recording devices that secretly can track someone’s identity.

When asked about how suiting the smart glasses’ visual interface would be for a person with dementia, i.e. getting digital information displayed in their field of vision, the therapists answered it could work really well for some people, while not at all for others: "That is something that needs to be tested individually. I also think that it probably would not be that suitable for those with visual hallucinations."

Introducing assistive technologies: Challenges and considerations Assistive technologies (ATs) were also discussed in a broader context during the interview - especially the challenges in integrating them into the patients’ daily lives. The therapists emphasized on the importance of introducing ATs as early into the disease as possible: "That is something we have discussed quite a lot when it comes to applying cognitive aids, that you need to start really early, almost before they even feel the need of it, for it to work and be incorporated into their everyday life."

The therapists also mentioned how motivation plays an important role in this, stressing the importance of considering the ethical aspects: "Each patient’s needs and what is meaningful and what they want to do is what should be in focus." Furthermore, they described that even though caregivers and family can see how an AT would be really helpful, some patients are not aware of their cognitive impairments, and hence do not see the need of getting any kind of support: "It is a challenge to introduce an aid and to make them see the purpose of it and still not be insulted by feeling like 'They think I am stupid!' or something like that."

The patient’s previous experience with technology before the onset of the disease was another important factor mentioned by one of the therapists: "When it comes to these technical aspects and it becomes several technical steps, it can get difficult pretty early on. For example using an iPad can be challenging quite early on, but if you have previous knowledge it is easier." The other therapist suggested relatives or other people close could help to handle more complicated steps in the device.

Designing for dementia: Human factors and considerations Human factors to consider when designing for people with dementia was another reappearing theme in the interview. One of the more prominent subjects brought up on this was the importance of looking at each patient individually, and seeing their whole disease picture: "Often, as you get older, you usually not only have dementia, you might have had a stroke before and be half-side paralyzed, or maybe you have Parkinson or a cardiac failure that have made you weak and tired since before [the dementia disease]."
Aside from suggesting the ATs to be personalized so they can adapt to each individual, one of the therapists also proposed to make them adjustable to different situations: "So that you can deactivate a stimuli that you might need in a different situation." Having too many inputs at the same time can be too stressful and confusing for a person with dementia: "The idea when you have these cognitive issues and similar is that you should have an aid that only helps you with what you really need. So it should be as clean and simple as possible around."

When being asked how information is best presented to a user suffering from dementia, the therapists explained that using visual information is the most effective way: "Reading is possible, but the visual memory is better than voice and read information, so pictures are optimal." They additionally suggested using a combination of picture and text or voice, if possible.

The therapists also mentioned the importance of supporting people with dementia to live in their own homes as long as possible: "[At home] you are competent and can find the bathroom with your eyes closed, instead of moving into an apartment or something, where you can be totally handicapped. It was though brought up how it is important for many sufferers to not feel stigmatized by using assistive devices.

5 Discussion

This study shows that there is an increasing interest in using modern technology as a way of supporting the world’s growing elderly population and people in need of cognitive assistance. Even though no studies specifically involving smart glasses were found, their potential in the field of dementia care can still be discussed based on the presented expert opinions and related research on Assistive Technologies (ATs) that use computer-mediated reality and head-worn technology. The potential benefits, challenges and future possibilities of using smart glasses as an AT for dementia patients are discussed separately below.

5.1 Benefits

Wearable devices in general can be easily integrated into people’s everyday life, since they in a way becomes a part of the user. Combining such devices with intelligent and modern techniques, such as computer-mediated reality and context awareness, reduces the interaction complexity, since there is no need for the user to explicitly give input to the application. Instead, these applications can assist the users by sensing their current resources and providing feedback automatically. Having that kind of cognitive support system, implemented in such a common everyday accessory as a pair of glasses, could be especially beneficial in dementia patients, as the therapists described how it can be challenging to integrate new devices into their everyday life.

Another major benefit with the glasses being intelligent is the fact that they can be more easily adapted to different situations and users, which the therapists
described as key design considerations, since the symptoms in dementia vary individually and during the development of the disease. Having an individual and tailored approach to the AT agrees with one of the guidelines proposed by Hyry et al. [13] and findings in the study by Firozian et al. [34].

Smart glasses also have the ability to address various senses, by using both visual and auditory information, which should be exploited according to the guidelines by Mayer and Zach [15] and the therapists’ opinions. The system furthermore enables different ways of user interaction; except from touch and gestures, it provides a hands-free interaction through voice-commands and a visual interface being presented in the user’s field of vision. This is another powerful feature considering the individual differences in this target group and the therapists’ experience that dementia rarely is the only present impairment or diagnosis, possibly further limiting either cognitive or physical ability. Using computer-mediated reality systems in health and dementia care was also suggested as a beneficial solution in several studies presented in the literature review [29–32].

On a functional level, this study presented a wide range of possible uses for smart glasses and similar solutions in dementia care. The literature review showed examples of applications for cognitive training [30–32], memory assistance [33, 35] and navigational support [34], while the therapists also discussed possible uses of storing memories and recognizing people.

All of these support systems could help to increase or maintain independence in everyday activities, and hence improve quality of life and reduce the burden of caregivers and family. This would also enable the sufferers to remain living in their own homes for an extended period, which the therapists stated as a critical aspect in keeping their independence.

5.2 Challenges

Developing and introducing any AT for this target group can be challenging, since learning new things is one of the cognitive abilities being highly affected by the memory degradation. Applying ATs in this target group can also be obstructed by the lack of motivation, due to patients not being aware of their symptoms, as explained by the therapists. Promoting learnability and memorability, as suggested by Mayer and Zach [15], as well as introducing the technology in the early stages of the disease, proposed by both Hyry et al. [13] and the therapists, are therefore essential for the AT to be usable and able to reach its full potential.

Other important challenges found in this study are related to ethical issues. Since smart glasses and intelligent devices in general typically are equipped with a camera and incorporate sensors that can be used for monitoring and tracking purposes, it is important that the patients give their informed consent, which can be complicated by the cognitive symptoms in dementia diseases [12]. Additionally, the therapists expressed how their county and municipal do not accept assistive technologies with tracking functionality, which limits the navigational
support only to those who can afford to buy the product themselves. Similar problems and ethical dilemmas go for the facial recognition technology.

Despite their cognitive impairments, dementia patients should be involved in the design and decision making process for the technology to meet their needs, as stated in previous studies [9,10,12]. The therapists also describe how it is the patient’s needs and wishes that always need to be in focus. Although challenging, participatory design in this target group have been carried out with good results, as shown by Mayer and Zach [15]. However, most of the applications presented in this study are lacking in this aspect, which unfortunately appears to be a common issue, as previously shown by Bharucha et al. [9].

5.3 Future possibilities

Since smart glasses still are such a recent innovation, the current lack of research and studies exploring their potential in this specific target group is not surprising. However, this is likely to change as they become more accessible and commonly known. In the longer perspective, when they potentially also are being more accepted by society and commonly used, people getting a dementia diagnosis will already have previous experience with the technology - an important factor for an AT to be integrated into everyday life, according to the therapists. With an increased use and a discrete design in the future, the glasses could also be worn without being a symbol for the disease, and hence reduce the stigmatization and ease acceptance, as suggested by Mayer and Zach [15] and the therapists.

6 Conclusion

This study has examined the potential benefits, challenges and future possibilities of using smart glasses as an assistive technology for dementia patients, taking into account both what the technology has to offer and what human factors to consider. Results from a literature review and an interview with two occupational therapists, showed that smart glasses have many promising applications and features that could be beneficial in this target group. However, their real potential has yet to be proven in clinical studies, which hopefully will happen as smart glasses become more accessible. Including dementia patients in the studies and prioritizing the ethical concerns are important challenges to address in the continued development of smart glasses and other modern assistive technologies in dementia care.

References


UX Guidelines for Accessibility in Action Shooter Games

Anna Päärni

Department of Applied Physics and Electronics
Umeå University, Sweden
annpaa91@gmail.com

Abstract. Gaming has moved away from being a pop-culture phenomenon to an integral part of culture, but not everyone can partake. When game developers forgets that all future players of their game do not possess the same abilities, some are left out. This article aims to identify where cognitively disabled players are left out in the game development process, focusing on their user experience in action shooter games. This was done by evaluating the UX in three action shooter games, using a template built on HTA and the game interaction model to perform and evaluate a set of tasks. The resulting data, paired with commonly reported problems by cognitively disabled players, was used to identify where accessibility mistakes had been committed by the developers. The pinpointed problematic areas were then used to develop an additional set of UX guidelines, building on existing accessibility guidelines for games. The results showed that the detected issues were centred around the main menus, disruption, and perceiving and interpreting danger. The research concluded with four additional guidelines; rely on design conventions, mix visual hints with other input, avoid undesired multitasking, and allow for customization in combat.

1 Introduction

The UN reports that 15% percent of the worlds population are disabled to some degree, making it the world’s largest minority[1]. Thus it is important to move away from the idea of a "standard" user, since all user have different levels of experience and abilities (for example, about 8% of all males are colourblind[2]). The United Nations Committee on the Rights of Persons with Disabilities states that a part of the rights as a disabled person is the right to be able to participate in cultural activities on an equal basis[3]. The evolution of IT, design and technology does not only mean tools for improving things like public transportation or hearing aids; it can also be an important bridge to culture. These bridges can consist of screen readers to read blog posts, SDH-subtitles (subtitles for the deaf and hard-of-hearing) during a movie, or books written in simple accessible English.

Video games are an integral part of our culture[4] and no longer just a pop culture phenomenon[5]. 2015 51% of all American households had a dedicated
game console, and 42% played video games at least 3 hours per week[4]. Economically, the game industry might benefit from being invested in making sure that gaming should be accessible to all of those who want to participate, since more players would result in a bigger revenue[6].

Multiple groups are working for a more inclusive world of games, for example Games Accessibility Special Interest Group\(^1\)[7], and the AbleGamers Foundation, an interest group for disabled players and relatives to disabled players[8]. At the Gamer Developers Conference the 16th of march 2016 a lecture titled Inclusification: How to Make Your Game(s) More Inclusive to Millions was held by the director of AbleGamers, Mark Barleton. Barleton described the obstacles disabled players face, and advised game developers to increase the accessibility in games to reach a broader audience and become more inclusive[9].

A coalition of game studios, specialists and academics have published an extensive online resource with accessibility guidelines for game studios, and both AbleGamers and Games Accessibility Special Interest Group have also proposed accessibility guidelines for games [10]. The guidance provided in these is not divided into different game genres, and is with a few exceptions not specific regarding UX implementations. Examples of instances where cognitively disabled gamers are left out or hindered because of poor UX are lack of a tutorial mode, poor documentation and no indications of dangerous situations [7].

The objective of this paper is to identify where cognitively disabled users are left out in the design process of games, and propose new UX guidelines for accessibility when developing action shooter games played on PC. Three important aspects of the games were included in the analysis; the in game menu, the main menu, and the combat experience (usage of weapons and abilities). The guidelines were developed with data collected in an accessibility evaluation of three action shooter games, as well as previously reported problems for cognitively disabled players in games. They are intended to be seen as an addition to or altering of existing guidelines, proposed from a UX-designers point of view. The research questions in this article are as follows:

- Which common pitfalls in the UX, that hinder cognitively disabled players, exist in action shooter PC games?
- How can existing guidelines for UX in game development be altered to better include cognitively disabled players in action shooter games?

2 Theory

Games Accessibility Special Interest Group states that video game players are far from a homogeneous group with the same abilities (and lack of disabilities), income, age and gender. Disabled player’s potential is often underestimated, as well as their willingness to participate in games[7]. Accessibility in video games means including all players who wish to participate, and not limiting the experience to only be enjoyed by "able" players.

\(^1\) GASIG is a part of the International Game Developers Association, San Fransisco.
The concept of *universal design*[3] is another important and interesting aspect to take into account. Instead of viewing accessibility for disabled users as something that needs to be added on and tailored to fit a certain group, design can from the beginning strive to be as inclusive as possible. In game development this can be done by considering accessibility in early stages of the development, but also

2.1 Cognitive disability

Cognitive disability is a broad term that can be defined in various depths, where one description given in a medical journal is the following,

> Causal factors related with cognitive disability are multiples and can be classified as follows: Genetic, acquired (congenital and developmental), environmental and sociocultural. Likewise, in relation to the classification, cognitive disability has as a common denominator a subnormal intellectual functioning level [...] [11]

Examples of diagnoses that fall into the category cognitive disability are ADD, dyslexia, dyscalculia, autism, Downs syndrome and TBI (traumatic brain injury), where the diagnoses can be divided further into different subcategories. Dyslexia, ADD and dyscalculia are examples of diagnoses that are considered to be less severe conditions than Downs syndrome and TBI[12].

2.2 Common issues reported in game UX

A study with 500 participants from 2010[13] collected and compiled common issues that players reported facing when playing video games. The participants consisted of group A (490 individuals), members of a large computer festival, and B (10 individuals), a mailing list for disabled (not game-related). 46% of cognitively the disabled players (B) reported problems with speed, language, and understanding the task they were given in the game, compared to 60% of the larger "gamer" group (A). However, it is important to note that group B had a significantly lower number of responses. The types of problems cognitively disabled players reported having are summarized in table 1.

2.3 Existing guidelines for accessibility in games

A coalition of game studios, interest groups that work for disabled gamers, academics and specialists on the topic have developed guidelines for accessibility in games that are available on their website[14] as an online resource. They are divided into basic, intermediate and advanced, as well as the categories hearing,

---

1 This is marked as a problem encountered by players who are hearing impaired. However, the author argues that this is a problem that is common for all players, since important information given during cut scenes can easily be missed.
<table>
<thead>
<tr>
<th>Problem</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inability to follow a storyline.</td>
<td>Story is very complex and difficult to follow</td>
</tr>
<tr>
<td>Unable to determine how game is played.</td>
<td>Lack of a tutorial mode.</td>
</tr>
<tr>
<td></td>
<td>Poor documentation.</td>
</tr>
<tr>
<td></td>
<td>Documentation written at too high a level for intended audience.</td>
</tr>
<tr>
<td></td>
<td>Lack of hints in game.</td>
</tr>
<tr>
<td></td>
<td>Lack of Informative UI.</td>
</tr>
<tr>
<td>Player’s character gets killed/injured repeatedly in game.</td>
<td>No indication of dangerous situation.</td>
</tr>
<tr>
<td></td>
<td>Unable to alter game speed.</td>
</tr>
<tr>
<td></td>
<td>Unable to adjust difficulty levels.</td>
</tr>
<tr>
<td>Unable to complete a puzzle or task.</td>
<td>Vital clues given in cut scenes with no text or visual queues available.</td>
</tr>
<tr>
<td></td>
<td>The puzzle is too hard or complex, and cannot be adjusted with difficulty modes or speed adjustments.</td>
</tr>
</tbody>
</table>

Table 1: Based on the table presented by Games Accessibility Special Interest Group[6], issues related to cognitive disability have been selected and presented.[7]

vision, cognitive, motor, speech and general. The guidelines for cognitively disabled players often provide examples of UX implementations and further reading, as well as examples of games with "best practice".

AbleGamers game accessibility guidelines presented in their accessibility guide [10] are divided into four separate type of players; visually disabled, mobility disabled, hearing disabled and cognitively disabled. Each disability category is then further divided into three levels of how much of an impact these implementations would do. They are general regarding the implementation of the UX, where the most detailed guidance is that menus should be intuitive and that enemies should be clearly separated from friends[10].

Games Accessibility Special Interest Group present their guidelines[15] in a similar manner as AbleGamers. The guidelines are divided in to the categories visual disabilities, cognitive disabilities, mobility disabilities and auditory disabilities. The ones regarding cognitively disabled players are mostly short and general, with little focus on how to implement accessibility in the UX. Each guideline also explains the benefits of implementing this accessibility feature. Many have a short question the developer can asks themselves to estimate the
accessibility, for example "Test: Does the player need to read the manual or look up how to properly play your game? [15]."

The three set of guidelines often overlapped, but differ some in the amount of detail, topics brought up and number of paragraphs. None of the sets of guidelines provided advice specific to game genre, and often left room for elaboration and more concrete guidelines regarding the UX.

3 Method

The results were based on two steps; during the first step the author evaluated three existing action shooter games, identifying and reporting commonly reported problems for cognitively disabled players. In the second step the data from the analysis was used to propose guidelines specific to the UX for cognitively disabled players in action shooter games, extending existing accessibility guidelines.

3.1 Evaluating the games

Since the author did not have the means to conduct a observational analysis with cognitively disabled video game players, HTA[16] and game interaction model[17] was used to perform an analysis of the UX without involving users.

The first step in the evaluation was to devise a simple template to estimate the accessibility for players with cognitive disabilities. The template would then be used as a tool for an objective evaluation that would determine if the chosen action shooter games, Left 4 Dead, Bioshock I, and Dead Space 1 had fallen into any UX pitfalls regarding accessibility. Three aspects of these games were taken into consideration; the player’s in-game-menu, the main-menu, and the combat experience (usage of weapons and abilities).

3.1.1 Developing the evaluation template

To develop a simple but exhaustive template to evaluate the selected games’ UX, two separate but similar approaches to evaluating the user experience were used. One specific to games and accessibility, the game interaction model[17], and one general for user interface evaluations, HTA [16]. The template was devised to evaluate the performance of a number of tasks, breaking them down into constituents to detect problematic solutions. To connect the HTA to cognitively disabled players frequently reported issues with games[13][6], previous problems were added to the template, so that they could be correctly noted if they were identified.

The finished template consisted of an HTA-analysis for each task (the tasks are defined below in Tasks to be performed), where the evaluator wrote down the required sub-tasks to achieve the main task. The received stimuli, determined response and provided input from/to the game were also given a column for each task, using the principle of the game interaction model-analysis. Additional
notes were taken on each factor when problems and errors were detected, in accordance with the HTA-principle\cite{16}.

### 3.2 Tasks to be performed

**Main menu**

1. Change the difficulty setting.
2. Switch on the subtitles.

**In-game-menu**

1. Pick up an item. Find it in the inventory, and then select the item. Determine what the item can be used for.
2. Select and drop an item.
3. Determine which task currently needs to be performed to move forward in the main quest.
4. Regenerate health.

**Combat**

1. Shoot the first weapon that is found and equipped.
2. Improve aiming, or determine if this is not possible.
3. Switch weapons.
4. Update a weapon or an ability.

### 3.3 Choice of games

Games with large budgets have resources to make changes\cite{6} and educate their employees, as well as the possibility to reach a large audience. Sequels have something the first does not; the advantage of player’s input, such as critique towards clumsy menus, or a demand for more of a favourable feature. Games with large budgets, that were either stand-alone games or the first one in a series, were therefore chosen, resulting in the selection *Left 4 Dead 1* by Valve, *Bioshock 1* by 2Kgames and *Dead Space 1* by EA Redwood Shores.

### 3.4 The aspects of the games to be evaluated

To clarify what this paper defines as the vital building blocks of the three chosen aspects (the in-game-menu, the main-menu, and the combat experience), they are below defined individually. The reasoning behind choosing these aspects were that the author believes them to be integral parts of the game UX, strongly linked to the action shooter-experience. The player relies on these three aspects to make adjustments when possible, as well as to perform necessary actions to proceed in the game. Video games are not an exact science, and this was a decision made by the author based on previous experiences of action shooter games.
3.4.1 Main menu

The main menu is the first menu that the player sees when the game is started. It usually contains settings (volume, brightness, difficulty, subtitles and similar), options to continue or start a new game, and an option to exit.

3.4.2 In-game-menu

The in-game-menu is the menu that can be reached and used inside the game, as opposed to the main menu that is reached outside the game when it is paused, or not yet started. Generally the purpose of the in-game-menu is to contain the items obtained by the player, ranging from health packs and weapons to crucial items for solving puzzles or making progress, information about current quest(s), maps, the players stats (health, level, status of abilities and similar) and other information relevant to the specific game being played.

3.4.3 Combat

Combat is when the player interacts with enemies to either incapacitate or kill them. This can be done with special abilities (electric shocks or stunning, for example), firing a weapon or using melee. Different action shooter games offers different weapons, abilities and aiming, and different enemies require different tactics. Often weapons and abilities can be altered within the game, for example by using workbenches or assigning earned points to skill trees.

3.5 Proposing guidelines

The information gathered by the game evaluation was considered separately for each game. Reoccurring problems that could be connected to a specifically defined accessibility issue for cognitively disabled players were noted to be further discussed. This was the presented result of the article as identified pitfalls.

When the detected problems had been noted and written down for each task, they were considered task for task and compared to existing guidelines[15][10][18] to determine if this area had already been addressed. If it had been suitably addressed, the issue was not further considered. If it had not been addressed at all, or was too general to provide tangible advice, a new guideline or an altering of an existing was proposed to solve this.

4 Results

The issues that were detected by the evaluation were chosen to be shown below, which means that the result of each task is not presented. The results from the evaluation were considered for each game separately, and are presented in that way below. None of the games had a tryout-level, or the option to alter the game speed, but did not have complex puzzles either (see table 1).
4.1 Dead Space 1

The font in Dead Space’s main menu had a coherent UI with the same font through all levels, high contrast between text and background to the text, and clear, loud feedback sounds. Scrolling, selecting and backing had different, distinguished sounds. However, the background behind the menu was flashing and changing, showing streams of texts, numbers and pictures, overlaid with static. The background music was loud, changing and also mixed with static.

The in game menu was projected in front of the player, and could be angled differently depending on where the player looked. It did not pause the game, and took up a considerable amount of the screen. The different available options were presented as icons, where the icons representing missions and database were not standard icons, as well as very similar. Text logs often contained long detailed information that was not relevant to the main objective, and was sometimes written in technical language. There was no indication given to distinguish important text logs from less important ones. Audio logs could be interrupted by transmissions from other characters, cutting them off.

The enemies in Dead Space 1 alerted the player about their presence with growling. The game sometimes played dramatic music when multiple enemies attacked, which could drown out the sounds made by them. Unless the player faced the enemies and saw them it was generally difficult to pinpoint their position, allowing them to sneak up on the player. The players health gauge was placed on the back of the main character, and was out of view if enemies blocked it, if the camera angle was wrong or if the character twisted his upper body. The only constant feedback alerting the player that they were hurt were grunts from the main character, making the assessment of how dangerous the situation was difficult. Audio and video transmissions from other characters did not pause the game, and were shown in front of the player. This meant that a part of the screen could be hidden behind the transmission, making enemies more difficult to both hear and see. Enemies were killed by cutting off limbs, but could "survive" without a head, for example, and it was often difficult to know when the "final blow" was dealt.

4.2 BioShock 1

The main menu in BioShock was inconsistent. Two different fonts were used, shading was sometimes placed above words, sometimes below, and the background image changed drastically between different menu levels. The highest level in the menu had a logo that took up almost a third of the screen, leaving little room for actual menu options.

Optional lore in BioShock was offered in audio-logs that could be found throughout the game. They did not pause the game, and required the player to stand still and be vigilant to avoid being jumped. The in-game-menu offered the messages to be replayed, but this did not mean that the player could not be interrupted again. The audio logs were often interrupted by an announcement from the game, or from a transmission from another character.
BioShock had no visual indication of the position of the attacker unless the player faced them. Enemies muttered to themselves if they had not spotted the player, and shouted at the player if they discovered them, but were often difficult to pinpoint if they were not spotted. When the player was hurt the vision was blurred. If the player’s health became dangerously low, a loud heartbeat was played and a text-flash urging the player to regenerate health was shown.

4.3 Left 4 Dead 1

The main menu in Left 4 Dead relied heavily on simple lists. Only the word Start was in a different colour (turquoise, instead of white or grey). The start-option was placed above Back, in a list of option that was left-aligned. The HTA also showed that the action that required most steps was to start a single-player game in Left 4 Dead. The subtitles were in thin italics, making them quite hard to make out. Sounds from zombies were also subtitled, which meant that when a large group of zombies were nearby the subtitle feed was flooded with subtitled sounds the enemies, drowning out helpful lines from other characters and instructions from the game.

Left 4 Dead does not rely on lore and different objectives, since the objective is always the same, and the only messages that were given in audio form were short sentences said by characters. These were often interrupted by another character yelling a different line.

Unless the player faced zombies they were often difficult to spot if it was dark, or if the player was attacked by multiple zombies from different directions. The zombies in Left 4 Dead attack in big groups to overwhelm the player and it was often difficult to pinpoint where the zombies were. Red triangles near the middle of the screen were shown when the player was hurt, pointing in the general direction of the attacker.

5 Discussion

With only a few exceptions, the HTA and the game interaction model showed that the steps to perform actions and find information were few and straightforward. The issues were rarely too many steps, but instead poor choices of interaction, colours, fonts and layout. The HT-analysis and game interaction model proved to be helpful to pinpointing these errors, since it required the evaluator to perform actions systematically. The problems that were selected and elaborated on below were those that occurred in all three games, and were linked to previously reported accessibility obstacles. These problems are not presented for the games separately, but are instead discussed in areas of interest.

5.0.1 Main menu

One of the most problematic parts of both BioShock 1 and Left 4 Dead 1 were their main menus, an issue related to the commonly reported problem lack of
Informative UI (see table 1). They often did not follow UX design conventions, and especially the main menu in *BioShock* was inconsistent.

The menu in *Dead Space* had a considerably big flaw. The background kept changing and flashing, showing streams of texts, numbers and pictures, overlain with static. This brought down the readability, was distracting and could even affect players with epilepsy[19]. The background sounds in *Dead Space* were loud, changing and mixed with static, making a bit stressful to take the time to scroll through the menu options.

Menu layouts (as well as general advice for fonts) are addressed in the on-line accessibility guidelines [18], as well as *Games Accessibility Special Interest Group*’s guidelines [15] and *AbleGamers*’[10]. An addition to this might be to advise that when designing UX it is important to keep accepted, commonly used design conventions as a guidance. Examples of these Nielsen’s heuristics[20] and Shneiderman’s ”Golden rules”[21]. It is also highly relevant when the player is in dangerous situations or is given important information. Using conventions will help to avoid confusions, poor usage of space and other UX-mistakes.

### 5.0.2 In-game-menu

A problem regarding both story is very complex and difficult to follow and lack of hints in game (see table 1) were the constant interruptions in monologues and quest updates in *Dead Space* and *BioShock*. The interruptions did not make the story more complex, but it made it significantly more difficult to follow. While the game or a character was conveying important information to the player (either ”in person”, a text flash on the screen, or an audio/video/text-message), they were often either abruptly cut off or interrupted, while the message kept playing, by attacking enemies.

Some players might enjoy the challenge to try to listen to detailed description of the current mission while fighting off five monsters, but to some it might mean a less enjoyable playing-session. This somewhat addressed in the online guidelines [18] that recommends instructions and narrative to be possible to replay, but none of the guidelines advises to avoid allowing information be interrupted. It could be an additional instruction to allow the player to pause the game when information is given, or to only play instructions and narritations when the player is in a safe space.

### 5.0.3 Combat

A detected problem that is strongly linked to the problem no indication of dangerous situations, and partially lack of hints in game (see table 1), was the difficulty to read dangerous situations. None of the three games had a clear, visual indication of danger, instead only alerting the player with visual feedback once they were hurt. The enemies in all games either yelled or growled, but this could be drowned out by a message being played or dramatic music indicating an incoming attack. *Left 4 Dead* had an interesting visual solution to telling the
player that they are hurt, but it was not especially helpful if a large group of enemies hurt the player at the same time, which was the most common case.

It can be annoying to be given hints and instructions when they are undesired, but too few hints is a very commonly reported error by cognitively disabled players. A possible solution is to allow for players to decide for themselves the extent of hints. This is not discussed specifically in any of the guidelines, but the online guidelines [18] bring up the options to clearly show that interactive elements are interactive, and to remove non-interactive elements completely. In action shooters this can be expanded, and also include showing positions of enemies before they attack. Two possible solutions to this are to give the player a radar or to elaborate on the red triangles in *Left 4 Dead*, making them pinpoint enemies *before* they attack. An additional visual indication of danger, instead of only relying on sound, could help avoid surprise attacks and dying repeatedly.

6 Conclusion - Proposed guidelines

Action shooters need to find a balance where the players are given equal opportunities to enjoy the combat oriented game play. There are almost endless possibilities to alter the UX to fit as many players of action shooters as possible, if the developers are both aware of accessibility-guidelines and given the correct information. Judging from the material this article is based on, the industry seems to be including more and more of the accessibility-guidelines provided by interests groups for disabled players.

To add substantial content to the existing guidelines, the proposed ones are putting cognitively disabled players in the middle, strictly regarding UX and developed with the intention to be used in action shooter games. The proposed guidelines are meant to be seen as an addition to existing guidelines, and not the sole advice.

6.1 Rely on design conventions

Using conventions will help to avoid confusions, poor usage of space and other UX-mistakes.

6.2 Avoid undesired multitasking

Allow the player to have the choice to let the game pause when important information is given, or give information only when the player is in a safe space without enemies. Avoid to allow multiple monologues or dialogues to be played at the same time, interrupting or cutting each other of.

6.3 Allow for customization in combat

Let the player add additional hints for combat that show the position of enemies prior to attacking.
6.4 Mix visual hints with other input

A mixture of different types of inputs will make the hints easier to perceive (as well as benefiting deaf players). A visual hint instead of only an auditory when enemies are near can help the player to avoid constantly being surprised and hurt or killed.

6.5 Limitations and future work

Action shooters consist of many aspects, and far from all of them have been examined. The author had to determine which aspects to evaluate to limit the study and deemed the main menu, the in-game menu and combat to be the most important parts of an action shooter game. This inevitably adds some subjectivity to the study. Additional evaluation of the other parts of the game might even this out, but the author argues that games and UX are rarely a completely objective subject.

The analysis would have benefited from an observational analysis where test subjects played the games, with the same directives as the author. The best case scenario would have been to let cognitively disabled players perform these walk throughs. It would have drawn attention to problems in the games that were missed during the analysis performed by the author. The developed guidelines would also have benefited from being discussed with UX-developers at game studios, as well as representatives from interest groups for disabled players.
References


The Effect the Number of Columns has on Information Perception on a Web Page

Petra Andersson
Department of Applied Physics and Electronics
Umeå University, Sweden
id12pan@cs.umu.se

Abstract. It becomes more and more popular to use a column-based layout in web pages, but how does the number of columns affect the user and how the information on the web page is perceived? This paper analyses the use of a web page with two different layouts, one with three columns and one with five columns, both containing the same information. The study is based on a test containing interviews and a questionnaire related to the two web pages. The result shows that there is a correlation between the number of columns and the amount of information a user were able to perceive during a limited time. Participants testing the web page with three columns obtained better results regarding perceiving information on the web page, and they also rated the web pages as more visual appealing considering seven design characteristics.

1 Introduction

The internet has become the number one platform for many businesses, and the trend is increasing. E-commerce and marketing are two examples out of many sectors that utilise the internet as a tool to reach out and profile themselves towards both current and future customers, and markets. Communication between companies and users or clients commonly takes place on the internet, via the companies’ web pages. Web pages are frequently used as promotional tool and a large proportion of a company’s competitiveness rely on their web page and if the visitor finds the company trustworthy [1]. If a company is not convincing enough on their web page a user can, with one single click, open the next search result to find a different web page that belongs to a competitor. The competitor web page provides a similar service only communicated in a manner where the visitor more easily can see the value. This may determine which company the visitor chooses to proceed with, and in the long term become a revenue-generating customer.

How companies choose to design their web page varies, but from benchmarking e-commerce web pages it is found that one common design is to use columns as a way to present information. The benchmark shows that the number of columns frequently used lies between two to six. Figure 1 shows one example of how four columns are used on a web page for presenting information. Columns
Fig. 1. The figure is a wire frame showing how information can be displayed on a web page with three equal columns containing images, a headline and a body text.

is one type of layout to use when designing a web pages. By dividing and categorising the information in smaller sections, the user might get an overview more easily. There is a broad range of possible layouts with regard to columns, for example width, height and number of columns. All these different aspects may have an impact on how the user will perceive the web page and the information presented. Are columns used in the most effective way today, or could companies generate even higher sales or more clients with the right number of columns?

1.1 Objective

The aim of this study was to identify the number of columns that let the user perceive and scan a desktop web page in the most effective way. The effectiveness was measured in the amount of information gathered by the test subject at a pre-specified time. Information gathered consciously and unconsciously was considered in order to see if any of the subjects were affected by the number of columns. By analysing the amount of information the test subjects collected, it was possible to identify potential differences in column usage and establish guidelines for the most effective way of using columns on a web page. The research questions are as follows,

- What is the most satisfying number of columns which helps the user gather the most information in a defined time?
- How does the number of columns affect users’ information gathering?

2 Theory

There are numerous of components to consider when designing a layout for a web page and some of them are more important for this study since they affect how to compose columns. Described below are these components and also aspects such as how a user reads on the web and how much time a designer has to make a first impression.
2.1 Columns

Columns originated from printed or written material such as newspapers, magazines, folders and similar printed media. In the mid-fifteenth century, books were the primary printed material and mainly used one column. In a document found from the early sixteenth century (an English Bible) it shows that the printers had adapted symmetrical typing; the text was justified \(^1\) and placed in two symmetrical columns \(^2\). Further, the industrial revolution marked the beginning of graphic design. As advertising grew in a creative direction, the ads evolved and came to look as what today is referred to as graphic design. Columns, as a part of the graphic design, are today frequently used all over the internet. When applying graphic design to a web page, a frequently used technique is referred to as the ‘grid system’. This system is a broader term for graphical layout and consists of columns \(^2\). The ‘grid system’ is used for dividing a paper or page vertically and horizontally into margins, columns, inter-column space, lines of type, and space between blocks of type and images. With these subdivisions, the grid forms the basis of a modular, systematic and effective system for designing pages.

2.2 Read patterns

In a study concerning reading patterns Nielsen \(^3\) discovered that the users had fairly consistent behaviour when scanning the different websites. The dominant reading pattern is referred to as the F-shaped pattern, since it adopted the shape of the letter ‘F’, see Figure 2. The user firstly reads in a horizontal movement, usually across the upper part of the web page. The user then moves further down the page to continue reading horizontally, but this typically covers a shorter horizontal area than the previous movement. Finally, the user scans the left side of the web page in a vertical movement. From this study it is clear that users focus more on the left than the right side of a web page. Results similar to those

![Fig. 2. An illustration of the F-shaped reading pattern.](image)

\(^1\) Text aligned along the left margin, with letter- and word-spacing adjusted so that the text falls even with both margins.
of Nielsen’s [3] are also shown in an eye tracking study conducted by Buscher et al. [4] that examined users’ behavior to predict salient and muted regions of web pages. Salient and muted regions are the fields of a web page where the user do not seem to pay as much attention as to other fields. The entire right side of the web page above the fold 2 is neglected for both information searching- and page recognition- based tasks. Participants took approximately ten times longer when searching for information to fixate and pay attention to the right sided regions of a web page - 4 874 milliseconds compared to 339 milliseconds for regions on the top left side, see Figure 3. When users did eventually fixate on the right side, they did not stay there for a long time.

![Fig. 3. The median time to first fixation on the 10 different areas of a screen in milliseconds from a study conducted by Buscher et al. [4].](image)

### 2.3 Typography on the web

**Optimal line length** There is a lot to consider when designing information in text format. Earlier studies have been conducted regarding optimal line length - considered to be 50-60 characters per line including spaces [5]. If the line is too wide (around 70 characters and above) the user will have a hard time to focus on the text. This occurs because a long line length makes it difficult for the user to know where the line starts and ends. Too narrow lines break the user’s rhythm since the eyes have to travel back and forth too often. This, according to Lupton, tends to stress some users and can lead them to unwillingly skipping valuable information. Another variable to take into consideration is the fact that a user loses the focus along the line and therefore too long lines may cause a

2 All elements that are not visible without scrolling after opening a page are below the fold.
loss of focus. At the start of a line, and in the break between two lines, the focus is highest; at this optimal line length the user can maintain focus on the whole line and paragraph of text.

**Font and size** Lupton [6] has developed guidelines for choosing the right font size for screens. Users often have a larger distance to screens than to a magazine or book when reading. For comfortable reading on screens, a larger font size than used in print is required. A good starting point is 17pt\(^3\)[6] and from there go smaller or larger depending on web page requirements and importance of information.

The width of the column and font size are dependent up on each other. To achieve a good design layout and high usability, both the font size and the width have to match. A guideline is to have approximately ten words per line [7], and the narrower the column, the smaller the font size is required to be to maintain readability. Applying and working with the grid system implies submitting to the laws of universal validity according to Müller [7], such as the will to systematise and clarify, rationalise the creative and technical production processes, as well as the will to cultivate objectivity instead of subjectivity.

### 2.4 Time aspect

A study with 17 test subjects observing their daily use of the internet for four months resulted in visits on 17242 different URL addresses [8]. 52\% of all visits were shorter than ten seconds (median: 9.4s), 25\% lasted for less than four seconds, and nearly 10\% of all visits lasted for longer than 2 minutes. If only first-time visits were considered then the distribution over time did not differ much - the peak value for all new visits was still under 4 seconds. The results from this study are similar to, and partly based on, an earlier study conducted with 107 participants [9]. The time-aspect is important to take into consideration when designing a web page. The fact that a visitor generally will not give a web page as much of attention or time (often less than 4s) as needed for them to comprehend the whole web page sets high demands on the design. This indicates that the first impression is a crucial and a very important matter. The first seconds are the most critical since the average visit is less than that (9.4 seconds)[9]. Visitors are extremely sceptical to new web pages since they have previously experienced poorly designed pages. They are well aware of this probability when entering a page, and therefore try to minimise the time spent on poorly design pages, making the smallest design error increase the bounce rate \(^4\) dramatically[11]. If the web page survives the first ten seconds the possibility for a longer session increases dramatically as the user might stay and have a look around.

\(^3\) In typography the point (pt) is the smallest unit of measure. It is used for measuring font size

\(^4\) The bounce rate is defined as the percentage of visitors who turn around at the entry page and immediately leave the site. Such visitors "bounce" out and never see additional pages [10].
2.5 Benchmark study

A benchmark was made at an early stage of this study to analyse the actual use of columns on web pages which focused on sales, marketing or e-commerce. Both Swedish and international companies were considered. Five well-known companies' websites containing columns were evaluated to see how they used columns in their design: ICA, IKEA, Momondo, Kissmetric and Telia. The five companies were all using columns in one or more pages on their website and had between two and six columns in one row altogether. The most frequent number of columns were two, three and four (in no particular order), followed by five, and six was the least used.

3 Method

A study conducted in 2006 by Lindgaard et al.[12] forms the foundation of this study method. The study was performed using three different strategies to test the same thing - the first impression of websites. All three tests were executed one-by-one during the study. In each step, data were collect and through analysis a result was produced: how many milliseconds a designer has to make a first impression on a visitor. In the two first steps, the subjects were allowed to see the web pages for a predetermined amount of time (500ms) and rated the visual appeal of each page. Two different groups undertook this test. In the last step, the subjects repeated the same procedure as above and then were allowed to spend as much time as they needed to offer their opinion about seven paired design characteristics. The seven design characteristics were: simple - complex; interesting - boring; clear - confusing; well designed - poorly designed; good use of colour - bad use of colour; good layout - bad layout; imaginative - unimaginative. The characteristics were used to analyse what could affect the first impression, and a scale-based form was used for evaluation. In this study one coherent test was constructed, with two parts similar to the method that Lindgaard [12] used in 2006. First, the test leader read a case aloud to prepare the test subject for what type of web page they were going to enter and also to create a somewhat realistic situation with a spacious task. Further, the subject were shown one of the two websites for 11.4 seconds (adjusted after the pilot study, see 3.2) and after that verbally answer questions about the content of the page, both about

![Characteristic-scale](Image)

Fig. 4. Characteristic-scale, Simple - Complex, the participants were told to put a mark on the line.

was constructed, with two parts similar to the method that Lindgaard [12] used in 2006. First, the test leader read a case aloud to prepare the test subject for what type of web page they were going to enter and also to create a somewhat realistic situation with a spacious task. Further, the subject were shown one of the two websites for 11.4 seconds (adjusted after the pilot study, see 3.2) and after that verbally answer questions about the content of the page, both about
the information (text), graphics and the layout. The interview after the first part of the test was based on both open-ended and closed-ended questions [13]. After the interview, the tested page was visible to the subject for as long as they needed to offer their opinion about the seven design characters. A survey with a 10cm scale belonging to each principle was filled out by the subject. With data from the first part of the test, a conclusion was drawn over which number of columns gave the test subjects the best possibility to perceive the web page. The analysis was made by giving the results different score and compare those scores against each other. With data from the second part of the survey conclusions about if the page with the highest visual appeal, according to the design principals, were then related to the page with the best result in the first test.

The tests, the questionnaire, the survey and the web page, were performed in Swedish. This method focused on quantitative data. 16 persons conducted the test, all students within five different departments at Umeå university Sweden, equal number of men and women. All used the internet on a daily basis. To simulate a realistic environment, the test was conducted in the field [14] in open areas where the subjects attention was divided with activity outside of the test. This environment was chosen since it represents the daily life of the thought test group.

### 3.1 The web pages

In this study web pages with three and five columns were tested. These specific numbers were chosen dependent on the benchmark which showed that two, three, four and five were the most common, and the fact that some were using six columns. Due to the limitations of this study and prognostication by the author determined that differences between two and three columns and four and five columns are too small to analyse with this chosen method, therefore the two and four column alternatives were eliminated. Due to the infrequently use of six columns, this too was eliminated. Two sketches of the web pages were constructed, see Figure 5 and Figure 6. The sketches were created concerning the subjects in the theory chapter and additional design elements to make the web page as realistic as possible. Similar to when performing a usability test, the test environment and the task should reflect the users normal use to get a reliable result[14]. With a realistic web page, the test subjects should be able to feel confident and recognise themselves in the task given; this will make them act as natural as possible and provide more accurate results. The web pages contained general objects found in web pages from the benchmark such as header, menu with links, call to action and background.

A set of two web pages with a different number of columns, width, and height were tested. The same amount of information, text, images and figures were shown in a different number of columns. The height and width of the columns differed depending on the number of columns. The alternative with three columns contained about 65 characters, approximately ten words per line, and had font size 17pt, see Section 3, Earlier related work, for an explanation. It contained three smaller icons and three headlines, one per column. In the second
alternative with five columns, one column contained about 35 characters in font size 15pt, approximately 6-7 words per line, five smaller icons and five headlines. The columns together in both alternatives had the same width in total.

The text-based information in the web pages was exact the same except from the headlines. There were two more headlines in the alternative with five columns, one of them were entirely new and the other one was constructed from splitting a headline from the alternative with three columns. In total, the difference in text between the two alternatives was one word. Further differences were in the number of small icons, there were two more icons in the alternative with five columns, this to maintain the web pages realistically. The web page and the text simulated a travel agent and the text contained general information to support the company’s offer. The language in the web page was Swedish.

![Fig. 5. The web page used during tests with three columns.](image)

![Fig. 6. The column layout for the web page used during tests with five columns.](image)
3.2 Pilot study

A pilot study was conducted to ensure that the intended method was viable before the real study begun [13]. The pilot study contained two test subjects each testing one alternative to control the method addressing both alternatives, three and five columns. The aim of this pilot study was also to see if the time where the web page was visible to the subject was suitable for them to perceive enough information to be capable of answering questions about the content, and also to verify the case. From the pilot study, it appeared that the time was too short, and the test subjects had a hard time answering the questions. With this result, the difference between the two alternatives could be so small to measure with this method. By increasing the time to 11.4 sec, from 9.4, the test subjects were going to get a higher chance to perceive information and this was hopefully going to make the difference easier to observe. No changes were made regarding the case.

4 Result

The test contained two parts, one with questions based on the content and the second part where the subject considered and scored the web page by seven design characteristics. From the first part, the participants were able to perceive and reproduce the web page slightly better when using three columns, in total 3.5 correct answers with a mean of 0.4, compared to a total of 2.5 correct answers with a mean of 0.3, see Table 1. The result from the first part of the test with questions based on information in the text was relatively low. Few participants answered the questions correctly, and many of them had no clue. 10 of the 16 participants did not answer at all or wrong. The type value for both alternatives, three and five, was zero correct answer.

In the second part, the option with three columns generated a better result in all characteristics except one, imaginative - unimaginative, see Table 2. The characteristics were evaluated on a scale from 0.0 to 10.0, where 0.0 represents the best score. The web page with three columns got a better mean score at 24.0 compared to 29.7 regarding all characteristics. The characteristic with the largest difference between the two alternatives was clear - confusing, the web page with three columns was rated as more simple than the other one, even if both were considered as quite clear. There was also a noticeable difference considering the characteristic good layout - bad layout. The alternative with three columns received better scores, a mean of 2.6, compared to five columns with a mean of 3.5.

5 Discussion

The test conducted in this study did not only take into consideration the amount of information gathered from the text but also information found in other elements on the web page. Through that, the study still gives a good indication of
Table 1. The total for eight participants and the mean value for each participants of the correct answers for three and five columns.

<table>
<thead>
<tr>
<th>Questions</th>
<th>3 columns</th>
<th>5 columns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8 participants</td>
<td>Mean value</td>
</tr>
<tr>
<td>About content</td>
<td>3.5</td>
<td>0.4</td>
</tr>
<tr>
<td>About design &amp; layout</td>
<td>23</td>
<td>2.9</td>
</tr>
<tr>
<td>Total</td>
<td>26.5</td>
<td>3.3</td>
</tr>
</tbody>
</table>

Table 2. The mean score of the eight tests for each characteristic and column alternative. A lower number means better result.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>3 columns Mean value</th>
<th>5 columns Mean value</th>
</tr>
</thead>
<tbody>
<tr>
<td>simple - complex</td>
<td>1.9</td>
<td>2.6</td>
</tr>
<tr>
<td>interesting - boring</td>
<td>4.3</td>
<td>5.1</td>
</tr>
<tr>
<td>clear - confusing</td>
<td>1.9</td>
<td>3.7</td>
</tr>
<tr>
<td>well - poorly designed</td>
<td>2.2</td>
<td>2.6</td>
</tr>
<tr>
<td>good - bad use of colour</td>
<td>4.6</td>
<td>5.9</td>
</tr>
<tr>
<td>good - bad layout</td>
<td>2.6</td>
<td>3.5</td>
</tr>
<tr>
<td>imaginative - unimaginative</td>
<td>6.6</td>
<td>6.5</td>
</tr>
<tr>
<td>Total</td>
<td>29.7</td>
<td>24.0</td>
</tr>
</tbody>
</table>
how a visitor reacts considering the number of columns on the web page. There could be many reasons to why the participants could gather more information from the web page with fewer columns. The line length is one of them, in the alternative with three columns the line length is optimal [5] and could have had an impact on the result. It is hard to draw conclusions about the reading pattern referred to as the F-shaped pattern [3] and the findings of silent and muted regions on a web page [4] in this study since it did not contain higher technology as eye tracking. Despite, the fact that 10 of the 16 participants answered wrong or not at all could have a connection to the F-shaped reading pattern and the time it takes to fixate all regions of a web page. Since the columns were placed at the bottom of the page (to preserve a realistic layout) it may have taken longer time for the participants to fixated them and they may have read the text at the upper part first, according to Nielsen’s F-shape. From the result, it is clear that many of the participants did not read the text on the web page even if the case was constructed to encourage reading. Ten out of 16 did not answer correctly or at all on any of the questions based on the text. A common answer was "I did not read that, I saw a section about it, but I did not read it". Further, some responded with fictional information not found on the web page just to provide an answer, and some admitted they had no clue. During the pilot study the same issue was revealed, the participants did not read the text. With the intention to give them more time to read, the time was extended by 2 seconds. But the increased time did not have a major impact on the participants behavior related to their reading habits. The text was still left unread by almost 63% of the participants. But considering this as a model for how it works in reality when a visitor is entering a web page, the result also reflects the reality. A typical visitor does not read a lot of text, they rather look at headings, images, and icons and base their decisions on that. This conclusion is important to consider when designing a web page, does the web page reflect the company based only on the headings, images, and icons since this is what a user will have time to perceive before they decide to leave or stay within the first crucial 10 seconds [8, 11].

6 Conclusion

The result in this study indicates that the web page with three columns makes it easier for the visitor to grasp and perceive information, both information placed in the text and design elements. The participants who used the web page with three columns could answer the questions better. They had gathered more information, and also, they ranked the web page better considering all the characteristics except one, see Table 2. From the result of this study a guideline regarding the number of columns can be formed; With fewer columns the visitor will both have a better chance of perceiving the information, and they will also obtain a more positive attitude towards the website which could lead to a positive attitude towards the owning company.
References


Abstract. Prototyping software is becoming more accessible and can replace the well established paper prototyping method. Differences in user perception when testing paper prototypes vs. simple digital prototypes was investigated in this paper. Research was carried out in the form of twelve usability tests with two mobile application prototypes; one paper and one digital tested in a somewhat controlled environment. Findings from the research and post questionnaires showed that the paper prototype was given a higher score from users (paper 75.025 and digital 63.3) when evaluating the usability. The paper prototype had more interactions and comments during the usability test than the digital prototype. The results imply that the paper prototype was perceived more positively by the users with an increased willingness to collaborate with the prototype.

1 Introduction

When a new software product evolves, or when an existing product is in need of an update and redesign, the designer and the development team usually keep to a design method. This method is used to create an idea, evaluate this against users, and at the end of the process develop a final product. Paper prototyping has been a part of this method for many years and has helped to evaluate ideas during the early design phase [1, 2, 3]. The design method usually includes creating a rapid prototype, regularly drawn on or made out of paper but sometimes made digitally, to quickly generate ideas and evaluate these against the target group and users.

During the last few years, a considerable amount of different web applications and software for prototyping has evolved to facilitate the process, and there are now applications for all different purposes and fidelity levels. Some examples of these programs include Balsamiq, Axure RP, POP, InVison, Marvel and Pixate. These applications let the designer create digital prototypes with different fidelities. The prototypes can take the form of a pen and paper prototype all the way to a fully functional prototype with the right functionality, look and feel. There are studies and literature focusing on the effect of prototyping and the benefits of testing a product at an early stage and with different fidelities [1, 2, 3]. The average user has daily involvement with computers and the Internet according
to Statistics Sweden [4, 5]. In 2015, 89% of the Swedish population between 16 - 85 years old had access to the Internet. More than 75% use the Internet daily and 70% use a computer daily. A paper prototype can be far less intimidating than a digital prototype and can also eliminate the risk of feeling foolish during a test session according to Snyder [1, p. 57]. This paper will investigate if the use of simple digital prototypes can be received the same as well-established paper prototype, without losing any valuable feedback, satisfaction or collaborative willingness from subjects.

1.1 Objective

The objective of this study is to identify and analyse differences in user feedback and user satisfaction when testing simple paper and digital prototypes. The two types of prototypes used are paper prototypes and screen based prototypes. This study is an attempt to determine any differences between the two test options and to distinguish if any option is more favourable when it comes to:

- The amount of qualitative user feedback.
- Subject satisfaction regarding the application.
- Willingness to collaborate during tests with prototypes.
- Attitude towards the prototype.

2 Previous work

Previous work within this field has been carried out with slight differences and aims compared to this study’s goal. A study from George Mason University [6] compares two identical prototypes made from a Photoshop image, one print out and one using POP\(^1\) for realistic smartphone navigation. The study looked at finding usability issues when testing this on a group of people and claims that there are no significant differences between the two methods. Another study, comparing low- vs. high-fidelity prototypes, found that both prototypes are equally effective at finding usability issues [7]. A study from University of Magdeburg and University of Berlin [8] proves that a sketch like rendered image brings up more creativity and discussion from test persons than a wireframe, or shaded image does when it comes to architectural buildings. Usability issues are not this study’s aim; the focus is on how the subjects feel and their willingness to contribute to the usability test.

3 Theory

To clarify the terminology used in this paper, short explanations and definitions will follow below.

\(^1\) www.popapp.in
3.1 Paper prototype

A prototype is a designer’s or developers first attempt to examine, communicate and demonstrate ideas to various stakeholders. Paper prototypes are being used in various fields, but this article will refer to a paper prototype as a prototype of a digital system. A paper prototype can have different fidelity levels depending on content. The look, for example, can offer low fidelity while the depth of the prototype can be classed as high [1, p. 259-260].

A paper prototype is a simple prototype built from cheap materials including cardboard, paper, post-its and similar materials. The aim is to investigate different options and evaluate the conceptual design rather than graphical design and to be used as a support when communicating ideas between team members and users [2]. These prototypes are often hand drawn but can also be created by using a prototyping software. The paper prototypes can be displayed on paper or as a simple sketch on a screen with added interactive fields to create realistic navigation between existing screens. Paper prototypes are used to carry out usability tests at an early stage to gain valuable insights from the user. Below some advantages of paper prototypes are listed according to Snyder [1, p. 12].

- Provides sustainable user feedback early in the development process.
- Provides rapid iterative development. It is possible to experiment with many ideas.
- Facilitated communication within the development team and between the development team and customers.
- Do not require any technical skills, so a multidisciplinary team can work together.
- Encourages creativity in the product development process.

3.2 Digital prototype

A prototype made digitally can be used in a similar way as a paper prototype and shares the same objectives as paper prototypes [9]. The digital prototype, however, can range between narrative low-fidelity to a high-fidelity with final graphical elements and interactions. The different types of digital prototypes are used during different stages of the design process. Choosing which digital prototype to use depends on the main area to evaluate and where the product is in the design process. A screen based prototype made in a prototyping software is sometimes referred to as medium fidelity when it comes to interactions and looks, but in this study, the used digital prototype will be designed with the same fidelity levels as the paper prototype.

3.3 Fidelity

Levels of fidelity can sometimes be misused and applied to classify a single prototype as a whole rather than the current prototype’s content [1, p. 259]. A short explanation will follow to clarify the definition used in this paper. In
a prototype, there are different prototype contents to take into consideration; depth of the prototype, length, interaction, navigation structure, look, etc. All these aspects can, for example in a paper prototype, range between different fidelity levels depending on the content used [9, p. 85-95]. An example would be the aesthetics of the prototype that most probably will hold low-fidelity in comparison to the depth of the prototype that can range from a medium to a high-fidelity level. The two prototypes used in this article will both hold a low-fidelity look with medium fidelity interaction and depth. The ambition is to present these prototypes with the same fidelities throughout the sessions to be able to use the user feedback for comparison and evaluation.

3.4 Usability test

A usability test can be conducted at any time during the design and development phase with the help of a prototype, preferably started in an early stage. The purpose of a usability test is to collect the user’s feedback and identify any pitfalls, errors or mistakes to improve the overall usability. During a usability test real users will evaluate the prototype with sessions usually carried out in a natural setting or sometimes in a controlled environment. Collected feedback will be evaluated, and the prototype updated before a new session of usability tests are carried out. The usability test can be seen as successful if the test helps to improve the product for the better [10].

4 Method

The findings presented in this report are based on research collected from 2016, Umeå, Sweden. The test was conducted on twelve participants, divided into two groups of similar size. These groups had more than five subjects each, which is, according to Nielsen [11, 12], enough participants for a usability test. The participants age were between 17 and 56, and an equal number of men and women participated. The language used during the usability test was the local language Swedish. The test was a 10 to 20-minute long one-to-one session where the subject received information about the test, what the task would be and that the sessions would finish with two different questionnaires. The questionnaire followed the standard of The System Usability Scale (SUS) [13, 14].

All subjects were first presented with a verbal explanation about the session followed by a verbal description of the tasks to conduct. Throughout the session, the subject was encouraged to think aloud to communicate thought, feelings and assumptions to the test leader. This method also gave the test leader information to collect and later, evaluate. The test leader looked carefully at attitude when speaking, eagerness, tone in voice and more to make a conclusion on how the subject felt during the session. All communication during the session was recorded and categorised to estimate how many different comments the two groups made on average. To distinguish any differences of the type of feedback given during
the two tests, the recordings were analysed to categorise the comments into positive and negative statements. The total amount of remarks, amount of times when the subject said "This is good/fun/easy" or "This is hard/difficult" and how many times the test leader had to remind the subject to press the buttons on the prototype was also categorised. Due to the nature of the test persons’ comments, the categorisation of the collected data was an attempt to see any signs of differences between the two groups and used to support the collected feedback in the questionnaire.

4.1 Used prototypes for test session

Two similar prototypes were created, one made out of hand-drawn sketches on paper, Fig. 1, and one screen based prototype designed with software Axure RP\(^2\) to get a sketched feeling and appearance, Fig. 2. The two prototypes had the same text, graphics, and interactive depth and the only difference was the medium used to present the prototypes. In this test, an Android smartphone was used during the test session to display the digital prototype. The model was a OnePlus One from 2014 running on Android 6.0.1.

![Fig. 1. Examples of the hand drawn application interface.](image)

The application used in the test session is an application delivered by a national coffee shop, available on Google Play\(^3\). This application was selected due to its interactive depth, and somewhat unclear navigations and includes both

\(^2\) www.axure.com  
\(^3\) https://play.google.com/store/apps/details?id=com.mobivending.espressohouse
a loyalty card, inspiration and information about coffee. The thought was that most participants would not have used this application before the test session to minimise any recalls from previous app usage.

The paper prototype consisted of 16 screenshots to cater for the use cases the user was told to perform. The look of the paper prototype is somewhat rough with hand drawn lines and handwritten text and can, therefore, be classified as low-fidelity. The size was the actual mobile size to represent the interface as accurately as possible. The digital prototype created in Axure RP had the same amount of pages and graphic look to mimic the paper version, and the aim was to make these as similar as possible. The lines were consciously made uneven and a handwriting font style used to again, create a hand drawn feel.

4.2 Use case

During the test sessions, two test scenarios, and a prototype was presented to the current subject. The subject did not receive any information about the fact that the test had two different groups, testing two different types of prototypes to collect adequate data without interfering with knowledge about the second type of prototype. The data and questionnaire answers collected from the usability tests could, therefore, be compared.

All users were presented with two use case scenarios with related tasks to complete. During the execution of the task, the test leader took notes about the numbers of comments, if it was a positive or negative comment and how well the subject accomplished the given task. The comfort of the subject, tone in voice and other, small signals to identify the subject’s mood was observed. All test sessions were conducted in a somewhat controlled area to minimise disturbing noise and other disturbances.
4.3 Questionnaire

To complement the usability test with quantitative data, questions were given both before and after the session. The method chosen in this study is a standardised usability questionnaire called System Usability Scale, SUS [13, 14]. SUS is based on a Likert scale with ten predefined question, all with five answer options each varying from "Strongly Agree" to "Strongly Disagree". The questions were translated into Swedish for the test session due to the subject’s native language. For clarification, the use of "system" in the questionnaire is replaced in the standardised SUS with "application" [15]. Along with SUS, a second questionnaire was presented to the subject after the SUS to cover how the user felt during the test session. SUS was given first to eliminate the risk of affecting subject’s thoughts and feelings.

The SUS questionnaire covers the ISO standard of usability which includes effectiveness, efficiency, and satisfaction [14, p. 190]. The focus of the second questionnaire is how the subjects felt during the test and their attitude towards the prototype rather than the prototype and application itself. The questions where (1) Did you give ideas of new solutions for the application during the test session? (2) Did you give suggestions for improvement for the application during the test session? (3) How much do you think your ideas and suggestions mean for the application in the future? (4) What attitude did you have towards the prototype during the test session? All questions had five-grade Likert scale to use when answering the question.

5 Results

A mean SUS score, according to the established method [14, p. 194], will be presented here. The results of comment type, subject attitude and other findings will also be presented in this section.

5.1 System Usability Scale (SUS)

Data collected from the SUS showed a difference between the two types of prototypes. The paper prototype scored a mean of 75.025 in the SUS. The digital prototype scored a much lower score, a mean of 63.3, even though the prototypes were very similar except for the medium used to present them. The paper prototype was perceived to have better usability than the digital one.

5.2 Post SUS questionnaire

In the post-SUS questionnaire, the question "What attitude did you towards the prototype during the test session? / Vilken inställning hade du gentemot prototypen under testets gång?" differed the most and stood out from the rest. The two test groups showed that the subjects had a more positive attitude towards the paper prototype, Fig. 3, than towards the digital prototype, Fig. 4. The other three remaining questions in the post-SUS questionnaire did not show any remarkable differences.
Fig. 3. Scoring the paper prototype after the test, the answers ranged between (1) Very Negative and (5) Very Positive. (n = 6).

Fig. 4. Scoring the digital prototype after the test, the answers ranged between (1) Very Negative and (5) Very Positive. (n = 6).
5.3 Types and amount of comments

Nine out of twelve usability tests were recorded and from these nine recorded sessions, a count of positive and negative statements, how many times the subjects expressly said "This is good/fun/easy" or "This is hard/difficult" was made. In addition to this, the number of times the test leader had to remind the subject to press on the prototype to move on to the next page was counted.

Table 1. Mean number of comment per person on the two different types of prototypes.

<table>
<thead>
<tr>
<th></th>
<th>Paper prototype</th>
<th>Digital prototype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive comments</td>
<td>18.8</td>
<td>14.25</td>
</tr>
<tr>
<td>Negative comments</td>
<td>21.2</td>
<td>17</td>
</tr>
<tr>
<td>&quot;This is good/fun/easy&quot;</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>&quot;This is hard/difficult&quot;</td>
<td>1.2</td>
<td>0.25</td>
</tr>
<tr>
<td>Reminder to press on buttons on the prototype</td>
<td>1.6</td>
<td>0.75</td>
</tr>
</tbody>
</table>

The table above, Table 1, shows that the subjects in the paper prototype test gave, on average, 4.55 more positive comments and 4.2 more negative comments than subjects from the digital prototype test session. The number of comments regarded "This is good/fun/easy" and "This is hard/difficult" showed that subjects from the paper prototype session on average had twice the amount of positive statements and 4.8 times more negative statements than people from the digital prototype test session. The test leader reminded the subjects to press on the prototype buttons in order to move to the next screen on average 1.6 times per subject in the paper prototype session and 0.75 times on average during the digital prototype session. During the session, it was noticed that subjects testing the paper prototype read the application text more carefully than in the test with a digital prototype. No number or quantitative data was collected and it is only based on the test leaders perception.

6 Discussion

The result from the study can be compared to the article from Schumann et al. [8] who claimed that more creativity from test persons was found when presenting a rendered sketch image to clients rather than a wireframe or shaded images. It might, therefore, be beneficial for designers to consider using a paper prototype in the design phase to increase creativity and willingness to collaborate in the process.

The SUS score presented from the two prototypes gave a clear indication that the paper prototype was perceived as more usable than the digital one. This result may be because when a user holds a phone displaying a digital prototype, it is easy to compare it to a fully functional application without even realising it which affect the users thoughts and feelings about the prototype. The
paper prototype’s simplicity could reduce the likelihood of being subconsciously influenced by the knowledge of how a smartphone application is used, making it easier to separate the paper prototype from a more developed prototype or fully functional application.

The subjects testing the paper prototype became more creative than in the digital prototype tests and some even started to move the paper pieces by themselves instead of letting the test leader run the prototype. Results of the study could imply that the paper prototype improved the subject’s creativity and collaboration willingness. A higher number of comments, both positive and negative were received from the paper prototype sessions compared to the digital prototype test sessions. Some subjects had minor problems to navigate the digital prototype due to the prototype’s slow response.

Results indicate that designers should be wary in trusting new technology and should consider to stick with well-established methods for the best outcome. Findings from this research should be seen as indications on prototype choice and a mix of different mediums and fidelities might be in favour to cover all aspects of evaluating a product.

6.1 Limitations

No information was given about the prototype or application before the usability test started. In hindsight, it could have been mentioned that the test leader did not design the application, to remove the element of apathy from the subject.

Due to technical issues, three usability tests were not recorded and therefore could affect the accuracy in terms of the count regarding the different types of comments. Five paper prototype tests and four digital tests were documented and analysed. The classifications and quantity of the different comments were difficult to determine and should, therefore, be seen as a complement to the rest of the results.

The lack of observers during the test session made it difficult for the test leader to both run the prototype and analyse the subject completing the task. Due to this, small but meaningful signals might have been missed.

7 Conclusions

This study aimed to find differences between a paper prototype and a simple digital prototype. When adding up the results from the SUS, the type of comments, how the subjects acted and their attitude towards the prototype, it is clear that the paper prototype gave a better result than the digital prototype. The paper prototype gave more qualitative feedback than the digital one and the users rated the paper prototype as more positive than the digital. The fact that the paper prototype had a higher SUS score and was perceived as more positive could mean that the subject would be more eager and willing to collaborate during a usability test session even though during the test sessions it was difficult to distinguish any major differences between the two groups. The
subject satisfaction can be read from the SUS score and because the paper prototype scored a higher SUS than the digital prototype, it might be classed as more satisfying, at least when it comes to usability.

References


Guidelines for designing context-aware fitness applications

Julia Westman
Department of Applied Physics and Electronics
Umeå University, Sweden
id12jwn@cs.umu.se

Abstract. This paper has investigate what kind of information that is relevant to display in a context-aware fitness application. This was researched through a web-based survey with 29 individuals who responded to it. The survey was divided into two different parts. The first part was about fitness and job conditions and the second part was about information and features. The Participants were divided in two groups. One group for those who worked and one for those who studied or were unemployed. The findings from the research included that those who work in an office environment or study easily become sedentary during their day. It also includes that a context-aware fitness application should follow some guidelines to make the user more motivated to exercise. The guidelines that was found was for statistic, context, encouragement, notifications and community.

1 Introduction

According to a study by Alter, we become increasingly sedentary, both in our work environment but, also in everyday contexts of our life. Sedentary means that that you perform activities that do not require physical exertion, for example sitting down or lying down.

This may contribute to health problems such as diabetes, overweight and in the worst case cancer [1]. To reduce the time that we are sedentary can we motivate the user to perform some kind of physical activity. This could be done by using a fitness application that is using the method of ubiquitous computing and the concept of context awareness [2]. This kind of application can help users become more motivated, by knowing how far a user has moved, where the user has moved and also how many calories that the user has burnt while it has performed the physical activity.

The relation between a user and a computer has changed during the evolution of technology, at first it was many people per computer, and then it was one person per computer and now it is many computers per person.

The main idea of ubiquitous computing, is to focus on relation: many computers per person [3]. The word ubiquitous means "existing everywhere at anytime" [4] thus ubiquitous computing means that a wireless device is completely connected
Context-awareness is one of the main areas of ubiquitous computing [5] and it is defined as any information that a device can use to characterise the situation of an entity. An entity can be a person, place or an object. Thus context awareness means that a device can use information from the environment to provide task-relevant information or services to a user [4].

Wireless devices can transmit their location and status over the network when they are in a context-aware environment. The transmission can be done through various sensors that will capture the task-relevant information and then displays it to the user through the screen. For example, if a user has a daily goal to walk five kilometer and reached the goal, then a context-aware application can notify the user with a user message on the screen.

1.1 Objective
The purpose of the study was to answer the research question: What kind of health and fitness information is relevant for the user to receive from the environment to make the user more motivated. The question was asked to identify guidelines for how to design a fitness application that use the concept of context-awareness.

2 Theory
This section gives a brief introduction to sedentary, context, context-awareness and ubiquitous computing.

2.1 Sedentary
According to Warburton we are becoming more sedentary in today’s society, both during the time when we are working, but also in our own time and our leisure [1]. That means that we spend more time sitting still, other than to perform some sort of physical activity. Which means that the risk of health problems such as weight gain, metabolic syndrome, diabetes, and heart disease increases [6].
Sedentary behavior means that you perform activities that do not require physical exertion and some examples of these activities is sleeping, sitting, lying down, and watching some kind of screen-based entertainment [6].

2.2 What is Context?
The word context can simply be explained with Deys definition: "Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and application themselves" [7].
Thus, if some kind of information can be used to represent and characterize a situation in an interaction, then that information can be used as context information [8].

An example of how to use the context is if you want to know how many degrees it is outside, then you can by using your location find out the weather at the location.

There is plenty of information around us all the time, which is considered to be used as context information [9]. Some examples of such information:

- Environment information, e.g. temperature, light-level or noise-level.
- Social situation, e.g. who you are with and who is nearby you.
- Spatial information, e.g. location, orientation, and speed.
- Temporal information, e.g. time of the day, date, and season of the year.
- Identity, e.g. who you are as a user.

There are basically four certain types of context, that are more important. These are identity, time, location and activity. These types of context responds not only to the questions of what, who, when, but can also can be used together as other sources of contextual information [10].

2.3 What is context-awareness?

In 1994 Schilit and Their discussed for the first time Context-aware computing, they meant that it was a software that "Adapt of according to its location of use, the collection of nearby people and objects, as well as changes to Those objects overtime" [11].

Dey and Abowds definition of Context-aware computing, means that context-aware computing is âany information that can be used to characterise the situation of an entity, where an entity can be a person, place or physical or computational objectâ. They also mean that context-aware computing is âthe use of context to provide task-relevant information and/or services to the user, wherever they may beâ [10]. Thus Context-awareness means that a device knows its owns location and the environment around it. That means that a system that is using the concept of context-awareness can interact with the environment and also interpret and utilize context information [10].

Where you are, who you are with and what the resources are nearby, are three important aspects that a context-aware system is using. Thus the context includes more than just the userâs position [10]. There are also three important behaviours in context-awareness that an application can be identified with [13]:

- The automatic execution of a service.
- The presentation of information and services to a user.
- Labeling of context information for later retrieval.
There are two ways for an application to use the context of the environment. One way is to use the context active and the other way is to use it passively [14]. These two ways can simply be explained by Chen and Kotz’s two definitions [14]:

"Active context awareness: an application automatically adapts to discovered context, by changing the application’s behavior."

"Passive context awareness: an application presents the new or updated context to an interested user or makes the context persistent for the user to retrieve later."

2.4 Ubiquitous computing

During the last 50 years of computer science the interaction between a computer and a human has been relevant for the evolution of technology. During these 50 years, there have been three waves of technology. The first wave focused on the relationship between many people and one computer. The second wave focused on the relationship between one person and one computer. The last wave also known as ubiquitous computing, focused on the relationship between one person and many computers [3].

According to Rainie and Wellman ubiquitous computing can sometimes be called “Internet of things” and can be described as human-computer interaction that goes beyond personal computing to an environment of object processing information and networking with each other and humans [12].

Ubiquitous computing is a part of different research areas. Some examples of these areas are: including distributed computing, mobile computing, context-aware computing, location computing, sensor networks, mobile networking, and human-computer interaction [5].

3 Method

A web-based survey was designed to find guidelines for how to design a context aware fitness application. The target group for this research was a person that was between 18 - 65 years old and used different kinds of applications on their smartphone. The survey was disseminated through Facebook. This decision was made to get in touch with the right people who fit into the target group.

The expected result of this study will be answering the question:

– What type of information would the participants want to know from a context-aware fitness application, when they are working out?
3.1 Survey

The survey was divided into two parts, one part about job conditions and fitness. The second part was about what kind of information and features that should be shown in a fitness application. The first part was also divided into two different response groups, those who work and those who are studying or were unemployed. This decision was made so that the questions would be more suited for their circumstances.

The survey included 21 questions about occupation, general health, and what information they think should be displayed in a fitness application. The survey was comprised of different types of questions, such as: alternative questions, yes or no questions followed by open questions and scaling questions.

The choice to yes or no questions was made to get a quick view of the answer. Depending on what the person answered on the question, the person would get a supplementary question that is an open question. The choice to include open questions in the survey was made to get a more freely answers from the participating.

The choice to have scale questions was made to get an overall measurement of sentiment around a particular topic, such as what time of day they are most likely to perform a light physical activity.

The likert scale of the scale questions was made in a scale from one to six. In this way, the person had to have an opinion, and could not be neutral because they and could not choose the midpoint of the scale.

Test group The survey was sent to 40 persons through Facebook. Those who received the request to participate in the survey, first got a question if they had a smart phone that they use active. So they was using different applications. If the answer was yes then they belonged to the target group for the survey.

4 Results

The result from the user study was divided into two parts, one part about job conditions and fitness. The second part was about what kind of information and features that should be in a context-aware fitness application.

It was 29 persons that responded to the survey, 10 of them was students, 18 of them had a job and one of them was unemployed. The age of the participants ranged from 18 to 65 years old, but the majority of the respondents were between 26 - 35 years old.

4.1 Part one: Job conditions and fitness

The first part of the survey was divided into two different response groups, the first one is: those who work and the second one is: those who are studying or doing something else.
Group one: participants that was working Those who participated in the survey worked for example with nursing care, industrial, IT and finance. Of those who worked 29% felt that they easily become sedentary during the day while 35% did not think they were sedentary during their work day. The difference between the two groups is showed in table 1.

<table>
<thead>
<tr>
<th>Type of job environment</th>
<th>Answered between 1-3</th>
<th>Answered between 4 - 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Office environment</td>
<td>33 %</td>
<td>67 %</td>
</tr>
<tr>
<td>Remaining environments</td>
<td>78 %</td>
<td>22 %</td>
</tr>
</tbody>
</table>

Table 1. The result of group one that answered on the question If they feel that they easily become sedentary during there workday. If they answered between 1-3, then it means that they did not feel that. But if they answered between 4-6 then they did feel that.

When the participating in the study would estimate how many hours they sit down during a normal working day. The result was that 29% estimated that they sit down for more than five hours during the day and 47% estimated that they sit down between 1-5 hours. While 29% estimated that they sit down to less than 1 hour during the day. This is shown in Figure 2.

The participants was asked to respond on a scale from 1-6 how likely it was that they would perform a light physical activity, during certain times of the day. Where one was unlikely and 6 were likely. The times they would consider were: before they get to work, during their lunch break and after work.
9 persons responded between 1-3 on the scale, and 9 persons responded between 4-6 on the scale of how likely it is that they perform a light physical activity before they get to work.

The majority of participants felt that it is not likely that they will perform a lighter physical activity during their lunch break.

They were five people who felt that they are not likely they will perform a physical activity after their job. while they were 13 that would likely perform a physical activity after their job.

Group two: Participants who was students or was unemployed All who participated who study or was unemployed felt that they easily become sedentary during the day. (See Figure 3) Of those it was 82% who estimated that they sit down for more than five hours a day.

When participants were asked to respond on a scale from 1-6 how likely it was that they would perform a light physical activity, during certain times of the day. Where 1 was unlikely and 6 were likely. The times they would consider were: during the day and during the night.
One person replied that it was not likely that it would perform an activity during the day, while ten persons felt that it was unlikely that they would perform a physical activity during the day. When they were asked how likely is it that you
perform a physical activity in the evening, answered ten people that it is not likely that they will do it. While one person replied that it is likely they do.

4.2 Part two: information and features

Of the 29 persons that was participating in the study, had the majority of them used a fitness application earlier. Some of the applications that they had used were: Runkeeper, Apple’s Health app, miCoach Adidas, Vivo fit, Google Fit, Endomondo, Fitbit and FunBeat. Five persons of those who participating in the study had never used a fitness application. Some of the reasons was that they do not exercise or that they do not think it is a necessary application.

When the participants were asked if they would feel more motivated if they knew more about their training and their training results? Majority answered that they would become more motivated by knowing.

Some things that they would like to know about their training was: How far they have left to reach their goal, see how their training went, how much calories they have burned during the exercise, improvements, average speed, top speed, track of weight and track of steps.

The participants were asked if they would feel more motivated by receiving notifications when it is time to work out again. 17 persons felt that they would not be more motivated by receive notifications while 12 people felt that they would feel more motivated.

Six individuals would like to have notifications each day, four persons would like to have notifications a few times a week and two persons were like to have it within a range that they can decide. A summery of this two questions and the result showed in table 2.

The majority would not be more motivated by see their friends activities. While nine individuals felt that they would be more motivated by seeing that.
Would you feel more motivated if you knew more about your training and training results?

- Answered Yes: 90 %
- Answered No: 10 %

Some things that they would like to know about their training was: How far they have left to reach their goal, see how their training went, how much calories they have burned during the exercise, improvements, average speed, top speed track of weight and track of steps.

Would you feel more motivated if you received notifications from your smartphone, when it is time to work out again?

- Answered Yes: 41 %
- Answered No: 59 %

Some wanted notifications every day, some wanted notifications a few times a week and some of them wanted to decide for themselves.

Table 2. The result of questions about information and required functionality from the user study.

Some reasons to why they would feel more motivated was: that it would feel like a competition between you and your friends and in that way they would feel forced to exercise more. Another reason was that they would feel "if others can, so can I".

The majority also felt that they would be more motivated to see statistics about their performed activities. Other things that the participants felt would motivate them was if you get some kind of reward from the application, when you become better or when you achieved your goals. A summery of this two questions and the result will be showed in table 3.

Table 3. The result of questions about information and required functionality from the user study.
5 Discussion

It is not news that there is now a strong focus on how long and how often we exercise. Today’s technology and work situations mean that we easily become sedentary and not moving as much as we should. However, it depends entirely of what type of work you have.

The study found that if you, for example work with health care, then you are very active during your workday, which means that you do not sit down that much. But if you work in an office environment or study at the university, you will sit down large parts of your day. Which means that there are those kind of people who needs more daily exercise.

The survey also showed that the majority had used some of any kind fitness application. Which was very positive, when it is that type of user that knows what kind of information that is interested to know more about and what is not. The study showed that majority of the participants felt that a context-aware fitness application should be able to:

− View information about a the users training and their results.
− View statistics of past activities, weight and number of steps.
− View context information such as distance, maximum speed and average speed.
− Provide some sort of reward to the user when they reach their goals and become better.
− Give praise when you’ve done something better than before.

If an application would have all these features, the user would feel more motivated to exercise and will perform the daily exercise that they need.

To make the application complete and satisfy a wider audience, some other some features should be added. For example to get notifications about when it’s time to workout again, or get view friends past activities. These features would be there to make the user feel more compelled to actually exercise more.

However, these functions should be optional to choose to use. When some users thinks that they become irritated by them.

In further research on the subject it would have been interesting to align the target group towards the younger generation. because I feel that they easily become sedentary since today’s technology is a big part of their lives. Another thing that would be interesting for further research is to perform user testing on any existing context-aware application.

6 Guidelines

The conclusion of this study is that a context-aware fitness application, should follow certain guidelines to fulfill their purpose to motivate their users to take more exercise.
Statistics A context-aware fitness application should show information about the users past activities. The information shall include information about the user improvements, max speed during the activity, the average speed during the activity, and calories they burned during the activity.

Context A context-aware fitness application should know about their environment and context. In order to present the relevant information to the user. So as the total distance, how far it is to the destination, the user’s location, and the speed of the user.

Encouragement A context-aware fitness application should give the user some kind of reward when you have become better, such as points or different awards depending on what you have become better at. Because when the user receives some kind of reward then the user would feel proud of them self.

Notifications A context-aware fitness application should give the user an opportunity to choose to use notices when you think it’s time to work again. Through all users are different, some feel more motivated when they are reminded that they should exercise while others get annoyed by the pop up notices all the time on their smart phone, that’s why it should be optional.

Community A context-aware fitness application should give the user an opportunity to choose to connect to their friends, so that users can see their friends’ past activities. In order to be inspired by their friends and feel able to work out. But there can also be a negative effect of this, when it becomes an irritation for the user. That’s why it should be optional.

References


The role of feedback in educational applications for smart devices

Lena Lindberg
Department of Applied Physics and Electronics
Umeå University, Sweden
oil21en@cs.umu.se

Abstract. New technologies are affecting the way of teaching in primary school. Today iPads and educational applications are growing in popularity and is commonly used in schools. These new technologies have a potential beneficial effect on the learning, but studies show that some applications have flaws that hinder these positive effects, for example regarding the feedback provided to the user. Since feedback is an important part of the learning process, it is of great importance that the educational applications used in schools manage to provide feedback that support the learning process. This study has determined a number of guidelines for good feedback in a learning situation and further showed how well these guidelines can be reached in an educational application today. The study showed that there are both advantages and disadvantages of feedback in an application compared to giving feedback in a more traditional learning situation. Many of the guidelines for good feedback were found to be possible to reach in an application, and the greatest advantage found in applications is that they can provide feedback timely on each action taken by the pupil. The most critical disadvantage found in this study was that an application can not provide any conversational feedback.

1 Introduction

As the technology evolves, the way of teaching is changing and new technology becomes part of the every day classroom situation. Since tablets became available they have steadily increased in popularity in primary schools as a teaching material. Today there is a large number of applications designed for an educational purpose and to be used in primary school, and the number of available applications is constantly growing. According to Apple [1] there were 1.5 million iPads used in education and over 20.000 educational applications available in 2012.

According to a study on how tablets are being used in education [2], most teachers explained that “Using iPad increases the students motivation” and “The students are able to comprehend more information” as the two major positive aspects of using tablets in their education. Their findings showed that the use of tablets also lead to time saving, efficiency and increased mobility and flexibility. A recent study [3] with focus on students and teachers experiences of applications
showed that both groups saw that these technologies could for example make education more fun, collaborative and engaging. One of the greatest prospects in the use of tablets and application is that they can help improve the students motivation. Both teachers, students and scientists see a great potential in the use of tablets, but studies also show that this potential is not fully actualized [4–6].

1.1 Problems

According to a study by Vaala [6] where the highest rated and most popular literacy apps for children were studied, it was found that the most prominent applications are missing key teaching opportunities. Another evaluating study [5], show that the design of educational applications sometimes seems to focus more on superficial aspects than on supporting deeper learning.

One identified reason for why the applications do not fulfill their educational purpose is that many suffer from shortcomings in the feedback provided to the user. Since feedback is an important part of the learning process, the feedback has a deep impact in how well an application can support learning.

In a study made by Blair [7] preschool math-apps were studied with focus on the provided feedback. This study showed that in 82 % of the cases a so called “Answer Until Correct (AUC) feedback” was used. This kind of feedback means that the program indicates that the answer is incorrect by for example a red color, buzzing sound or similar, and the user must keep trying until the right answer is given. Further this report implied that the most frequently used AUC feedback is not the most effective feedback to support learning, and suggest that the learning could be improved by using different sorts of feedback that gives more information than only showing if the answer was right or wrong.

The fact that many educational applications have shortcomings considering the feedback provided to the user is also supported by Agneta Gultz, a Swedish scientist and leader of the The Educational Technology Group at Lund University. According to her studies only about 2-3 percent of the educational applications used today gives enough feedback to support deeper learning [8]. She claims that the majority of applications used today are rather a way of testing students current knowledge than supporting learning, and that some applications could even have a negative effect on the students learning because of their lack of feedback.

1.2 Objective

This study will focus on feedback given to the user in educational applications. The goal of the study is to determine what "good" feedback is and how this kind of feedback can be reached in educational applications, in order to ensure that the feedback truly contributes to the learning process. The following research question has been formulated:

- What defines good feedback in a learning situation and how can these theories be applied to feedback in educational applications?
This study will focus on children in first to third grade, since educational applications are being used more frequently in early primary school, than among older pupils.

2 Theory

In this article the concept of feedback is defined as information given about how someone is doing in their efforts to reach a certain goal, in this case as information given from an agent (a teacher or an application) to the user (the pupil) regarding aspects of the users performance or understanding of a task.

2.1 Feedback and learning

Feedback is an important part of the learning process [9,10]. Feedback can help pupils through providing motivation, enhance learning, enhance reflection and as a performance indicator [11]. Latham [12] explains that feedback is an important and necessary step in the learning process since it can give information and explanation beyond the original task or question that the pupil is dealing with. This means that feedback is a way for the pupil to gain a deeper understanding and a wider knowledge than what is possible without feedback. Latham mentions so called instructive feedback, referring to feedback that gives extra information in response to the students work, and states that this sort of feedback produces "quantifiable strides in learning".

Hattie [10] has made an analysis in order to find rate numbers for the effectiveness of different influences on students achievement, such as attributes in classrooms, teachers and curricula. By comparing the rates of these different influences it was shown that the average effect size was 0.4, while average effect size of feedback was 0.79. This meant that feedback was one of the top 5 to 10 influences on learning along with aspects like students prior cognitive ability (rate 0.71) and direct instruction (rate 0.93). Hattie also states that feedback is most effective when it provides information about correct rather than incorrect responses. An earlier study [13] talks about the relation between feedback and goals when it comes to learning. Hattie here explains that "Feedback without goal setting is less effective, and goal setting without feedback is ineffective" and that a combination of the two is necessary for an effective learning process. This has to do with the fact that a greater challenge gives a higher probability of the student seeking, receiving and assimilating feedback information.

Even though feedback has a proved positive and important impact on learning, there are also situations were feedback can have a negative impact. This is explained by Askew [9] who states that to much feedback can be overwhelming and inhibitory, particularly if there is no dialouge or discussion and no help considering how to make changes are given.
2.2 Types of feedback

How well the feedback actually provides the potential positive impact of learning has a lot to do with the type of feedback. There are many different thoughts on how to categorize feedback. Tunstall and Gipps [14] has made a study with the goal to identify a typology for teacher feedback. This study resulted in four categories of feedback with a number of subtypes. Their four types of feedback are termed A, B, C and D and these types have been placed across a sort of scale from evaluative to descriptive approach, where each type has subdivisions. As feedback moves from type A to D it becomes more descriptive and less evaluative. The evaluative types can be defined as either positive or negative feedback, while the descriptive types can not be described in terms of positive or negative, and are rather focused on improvement or achievement (see figure 1).

<table>
<thead>
<tr>
<th>Type A</th>
<th>Type B</th>
<th>Type C</th>
<th>Type D</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positive feedback</strong></td>
<td><strong>Positive feedback</strong></td>
<td><strong>Positive feedback</strong></td>
<td><strong>Positive feedback</strong></td>
</tr>
<tr>
<td>Rewarding</td>
<td>Approving</td>
<td>Specifying attainment</td>
<td>Constructing achievement</td>
</tr>
<tr>
<td>Rewards</td>
<td>Positive personal expression</td>
<td>Specific acknowledgment of attainment</td>
<td>Mutual articulation of achievement</td>
</tr>
<tr>
<td>Warm expression of feeling</td>
<td>Use of criteria in relation to work/behaviour/teacher models</td>
<td>Additional use of emerging criteria; child role in presentation</td>
<td></td>
</tr>
<tr>
<td>General praise</td>
<td>More specific praise</td>
<td>Praise integral to description</td>
<td></td>
</tr>
<tr>
<td>Positive non-verbal feedback</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Negative feedback</strong></td>
<td><strong>Negative feedback</strong></td>
<td><strong>Negative feedback</strong></td>
<td><strong>Negative feedback</strong></td>
</tr>
<tr>
<td>Punishing</td>
<td>Disapproving</td>
<td>Specifying improvement</td>
<td>Constructing the way forward</td>
</tr>
<tr>
<td>Punishments</td>
<td>Negative personal expression</td>
<td>Correction of errors</td>
<td>Mental critical appraisal</td>
</tr>
<tr>
<td>Reprimands; negative generalizations</td>
<td>More practice given; training in self-checking</td>
<td>Provision of strategies</td>
<td></td>
</tr>
<tr>
<td>Negative non-verbal feedback</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 1: Types of feedback according to Tunstall [14]**

**Positive and negative** One way to look at feedback is in terms of positive and negative. Negative feedback is mostly defined as feedback focused on incorrect actions, while positive feedback focuses on correct actions taken by the learner. In a study made by Ashby et al. [15] the effects of positive and negative feed-
back was investigated. This was done by providing different types of feedback to three different groups. The first group was given positive feedback only, the second group was given only negative feedback and third group was given a combination of the two feedback types. This study showed that the third group performed better than the other two, and the conclusion that a combination of both negative and positive feedback was to prefer in order to enhance learning could be drawn. Another study [16] on positive and negative feedback stated that there are positive aspects of using both types of feedback, and that the best type to use depends on the nature of the task.

A study made Van Duijvenvoorde et al. [17] used functional magnetic resonance imaging to examine the neural developmental changes related to feedback-based learning. Behavioral results from three age groups here showed that, compared to adults, 8- to 9-year-old children performed more inaccurately after receiving negative feedback relative to positive feedback. This study further implied a difference in how children compared to adults use feedback and the neural activation pattern found indicated a transition around the age 11 to 13 toward an increased positive influence on performance adjustment of negative feedback. These results are also supported in a similar study [18] on the effect of positive and negative feedback on preschoolers and suggest that young children in preschool and early school years are more responsive to positive feedback than to negative.

**Modalities** Another way to categorize feedback is by different modes, which is defined by Brookhart [19] as either written feedback, oral feedback or feedback given through demonstration. In a feedback study made by Werts et al. [20] also visual feedback is discussed as one possible feedback mode.

According to Orsmond [11] a majority of students prefer oral/verbal feedback since this type of feedback gives the opportunity to discuss the received feedback with the teacher and thereby clarify misunderstandings. This study also shows that some students finds written feedback unhelpful if it is not descriptive enough, a simple word telling that something needs to be improved is not enough. To enhance learning, feedback that provides a deeper explanation of the problems is needed.

The study by Werts et al. [20] further promotes the so called instructive feedback as a way to enhance learning. Instructive feedback is here defined as feedback that adds additional information to the pupil that might be unrelated to the original task. One example of this could be that if the task is to spell a word correctly, the feedback is not only telling if the spelling was right or wrong, is also adds more information such as telling an antonym to the spelled word. The positive aspects and effectiveness of so called instructive feedback is also supported by Griffen et al. [21] who shows that instructive feedback is also beneficial for pupils with mild intellectual disabilities.

**Levels of feedback** Hattie and Timperly [10] talks about four different levels of feedback as a way to categorize based on the content of the feedback. The
first of this four levels is called Task Level and is focused on the task performed. This level aims to tell whether an action is correct or incorrect for example by simply writing the word correct or incorrect for the student to read. This type of feedback is also known as corrective feedback and is the most frequently used type of feedback. The task level feedback could also include directions and more information about what needs to be done in order to improve. The biggest problem with feedback on the task level is that it often does not apply to other tasks.

The other level determined by Hattie and Timperly is the Process Level [10]. This type of feedback is directed to the process used to accomplish a task. This means that the feedback gives a wider information than only if the task is completed or not. The feedback also considers the understanding needed to complete a certain task. Studies have shown that feedback given on the process level seems to be more effective than feedback at the task level when it comes to enhancing deeper learning.

Third Hattie mentions a so called Self-regulation Level [10]. Self-regulation has to do with commitment, control and confidence and is a kind of internal feedback referring to the way the student monitor and regulate actions in order to reach the learning goal. How well this type of feedback works is depending on the students personal learning effectiveness, since less effective learners have problems developing self-regulation strategies and are in greater need of external feedback to reach the goal.

The fourth and last level of feedback explained by Hattie is called Self Level [10]. Feedback on this level focus on the student as a person and is often unrelated to the performance on the task. Hattie explains that this level should be avoided for two major reasons. First this type of feedback does not give any information on how to improve and reach further learning. The other reason is that this kind of feedback can contribute to the false picture of knowledge and intelligence as something static, which could have an impact on pupils understanding for the connection between effort and success.

2.3 Key elements of good feedback

Based on the different types of feedback it is clear that some types are to prefer over others when it comes to determining the meaning of good feedback. Wiggins [22] has defined a number of essentials of effective feedback that could also work as a framework for good feedback. According to Wiggins study the base of good feedback is that a goal exists and actions are taken to achieve this goal. The feedback should consist of information related to the goal about the actions performed. Apart from a clear goal it is also important that there are tangible results related to the goal. Effective feedback also needs to be concrete, specific and useful and well adapted to the recipient in a way that makes it understandable to them. The sooner the feedback can be given, the better the recipient can use it. It is especially important that the task is still ongoing so that the student have time and reason to act upon the given feedback. Last Wiggins [22] states that feedback must be consistent even if feedback is given from different sources.
A similar list of feedback aspects has been determined by Brookhart [19] who lists four important feedback strategies in order to reach good feedback. First Brookhart mentions the aspect of **timing**. Effective feedback should be given immediately to make sure that the feedback is given while the pupil is still working towards a learning goal, and not after they have finished. If the feedback is given immediately, they still have a reason to act upon the feedback and keep working towards the goal with the feedback in mind.

The **amount** of feedback given also have a great impact on how well the pupil receive it according to Brookhart [19]. To only list all of the issues or areas where the pupil needs to improve, is not effective even though it provides a great amount of feedback. A more effective way is to select a few main points and be very clear about what to do next in order to improve on these points. To be able to select which feedback to give, it is important to look at individual aspects. It is important to know the pupil to be able to sort out what information is relevant for each single person. The pupils personal goals and interests is one thing to consider when it comes to sorting out the most relevant feedback to given. It is also of great importance that the feedback have a balance between positive and negative aspects.

Feedback can also be given in different **modalities**, such as written, oral or demonstrations. When giving feedback to younger pupils Brookhart [19] explains that it is especially important to consider the pupils abilities, such as reading abilities, to make sure that feedback is given in a mode that they can understand. A conversational feedback is to prefer in most cases, where the pupil get a chance to talk to the teacher, ask questions and explain how they think about the task and their own performance.

Brookhart [19] further explains that good feedback demands appropriate sense of the **audience**. Depending on the content of the feedback it could either work best addressed to the individual person or in other cases to a group. Generally speaking, individual feedback is to prefer since it does not only give information about the how the student is doing, but also indicates that the teacher really cares about the pupils personal progress. By giving individual feedback, the feedback becomes more specific and well adapted to the individual.

A study made by Kelly and McLaughlin [23] have also shown the importance of a feedback that is based on the learners cognitive resources, abilities and prior experience. Their study suggests that feedback should be given as a function of the learners characteristics. The study states that it is of great importance to know the learner, know the task and match the feedback to the learner and the task.

### 3 Method

This study was conducted in two parts as follows:
3.1 Literature review

To achieve the objective of this study a literature review was conducted where feedback was studied from different perspectives based on the research question. The literature review served to collect information and theories found in earlier studies about feedback, with focus on feedback in a school environment and how feedback should be given to young students in an effective way. Through this review the role of feedback in the learning process could be identified as well as different types of feedback and how these different types affect the receiver. Finally theories and frameworks considering key elements of good feedback in a learning situation could be found and a list of guidelines for good feedback could be created based on these findings. The main keywords used to find relevant articles were: feedback, feedback and learning, feedback in education, types of feedback and effects of feedback. Among the articles found, the ones that had been peer reviewed were prioritized to make sure that articles were reliable.

3.2 Application study

The second part of the study aimed to determine how the theories about feedback found in the literature review could be applied in educational applications. In order to reach this goal two applications was studied with focus on the feedback provided to the user. The applications were chosen based on number of downloads and rating, where applications with high rating and high number of downloads were chosen to make sure that the evaluated applications were relevant and frequently used. One math application and one language application were chosen since these are the most common used types of educational applications. These applications were downloaded and studied on an Apple Ipad tablet. The study of these two applications was conducted in following steps:

1. Identification of the feedback provided in the application.
2. Categorization of the identified feedback according to different feedback types found in step 1.
3. Identification of successful and less successful examples of feedback in the applications based on the definitions of good feedback found in step 1.

By following these steps conclusions could be drawn on how well the chosen applications fulfills the framework for good feedback and thereby more general principles for how good feedback can be reached in educational applications could be determined. With ground in the study conclusions was drawn considering advantages and shortcomings of applications compared to traditional teaching, when it comes to providing good feedback.

Zcooly Affären 1  The first application in the study is called Zcooly Affären 1. This is a math application aimed for children between 6 to 8 years. The application contains a few different games and exercises that takes place in a store. By doing these exercises the user is supposed to learn addition, subtraction,
sequence number, pattern recognition, even and odd numbers and number sense. The exercise that contained most feedback was chosen for the study. The goal of this exercise was to practice counting items and calculating addition. The task for the user was to take the role of a cashier in a store. Customers entered the store and told what he wanted to buy, and the task was to give him the right kind and amount of items, count the price and put the right amount of money in the cash register. The task was divided in subtasks for each customer entering the store (see figure 2).

![Fig. 2: Zcooly Affären 1](image)

**Fun English** The second application in the study is named *Fun English*. This is an application aimed for children between 3 to 10 years. It consists of 14 different learning exercises with focus on learning English words. In this applications two different exercises were chosen for the study in order to cover as much feedback as possible, since the exercises in this application were more basic. In the first exercise (*Animal Band*) the goal was to learn to recognize the names of different animals. The users task was to listen to a voice that said an animal and then click on the animal that was named on an image with several different animals (see figure 3). In the second exercise (*Colour Buggies*) the goal was to learn how to spell different colours. The word that the user was supposed to spell was shown but faded and a voice read it out when the game begun. The letters of the word were circulating above the faded word and the task was to swipe a car around, and hit the letters with the car in the right order (see figure 4).
4 Result

The first part of the study identified guidelines for good feedback in a traditional learning situation. The second part found that in applications used today, both examples that follow and do not follow these guidelines could be found.

4.1 Literature review

According to the earlier studies that has been reviewed in this report a number of elements that are crucial for effective feedback have been found. The following
list summarize the aspects that were found to be the most common mentioned and thereby considered the most important guidelines in order to reach good feedback in a learning situation.

- **In right time** - This aspect was mentioned in almost all of the earlier studies as one of the most important aspects. Feedback must be given while the task is still ongoing, so that it is still relevant for the pupil to act upon the given feedback and improve their result. The sooner the feedback can be given, the better it will be received [10,19,22].

- **Informative** - Effective feedback needs to provide concrete and deep information about what the pupil need to to in order to improve as well as clear information about what the pupil has done well and why. The feedback should indicate both how the pupil is doing at the moment and what to do next. The feedback should not only refer to the specific task, a more effective feedback focuses on a higher level that makes it possible for the pupil to apply the feedback to other similar contexts [10,12,19].

- **Goal-based** - The feedback must be related to a clear goal if the pupil is going to be able to act upon the feedback in an effective way. If the goal is unclear, the pupil and the teacher might have different views of the goal, which can lead to misunderstanding of the given feedback [13,22].

- **Adapted** - Pupils different abilities and characteristics must be considered. The feedback must be given in an individualized way based on the learners personal characteristics [19,22,23].

- **Predominantly positive** - There are positive aspects of both positive and negative feedback and many studies indicated that a combination is to prefer [15,16,19]. It was though shown that negative feedback affect children in a less beneficial way than it affects adults [17,18].

- **Conversational** - Pupils often prefer a verbal feedback, especially if there is room for a dialog. If the student gets the opportunity to ask questions or explain his or her thoughts about their own performance, the student will receive a better understanding and can more easily adapt the given feedback [9,11,19].

- **Instructive** - When giving feedback to a pupil, additional information that answers question outside of the original task can be added to the feedback to support a deeper learning [12,20,21].

4.2 Application study

The use of feedback in the studied applications was identified as follows:

**Timing**

- **Zcooly** - Feedback is given as soon as the user has performed an action and marked that he/she is finished.

- **Fun English** - Feedback is given immediately as soon as the user has performed an action. There are a few exceptions were no feedback is given if the action performed was not correct.
Information level

Zcooly - The feedback given in this application is mostly on a task level but also gives further information about what went wrong and guides the user on how to move on in order to succeed with the task. Detailed information is given about exactly what part of the task that was performed incorrectly. If the task is performed correctly no further information is given.

Fun English - In this application the feedback provides uninformative feedback. All feedback is on a task level and simply indicates whether the task was performed correctly or not. No further information is provided in any of the exercises.

Clarness of goal

Zcooly - The goal of the task is not explained in detail as standard, but a more detailed instruction can be found by clicking a question mark. When choosing an exercise a voice and an image indicates what the task is about (image of a cash register and voice asking you to help out at the counter). As customers enter it is explained what this customer wants to purchase by a voice and a written text. Money and items are also visible to indicate the goal of the task.

Fun English - Each exercise is chosen by clicking a square were the name of the exercise and a very short description can be read. When this square has been clicked, no additional instructions about the goal is given. The exercise begins with a voice pronouncing a word of a color that is supposed to be spelled, or pronouncing an animal that is supposed to be clicked. There is no information button or other way to gain more information about the task.

Adaption

Zcooly - The feedback is given in the same way throughout this application. There is no way to adapt feedback based on the users abilities or preferences. Feedback is mostly given by a voice explaining what went wrong in combination with a sound and a visualisation through a red colored cross or green colored square. The task is given in both written and spoken form.

Fun English - Also in this application the feedback is given in the same standardized way at all times. This application is using a combination of visual feedback (stars) and a sound that indicates right or wrong.

Positive or negative

Zcooly - If a task is performed correctly, positive feedback is given by voice, sound and color. If the task is performed incorrectly, negative feedback is provided in the same way. Even if a task is performed partly correct but not completely, only negative feedback is given.

Fun English - Positive feedback is always given when a task is performed correctly. If a mistake occurs, negative feedback is given though a buzzing sound, or no feedback is given.
Conversation
Zcooly - No conversational feedback is provided.
Fun English - No conversational feedback is provided.

Instructiveness
Zcooly - No constructive feedback is provided.
Fun English - One type of instructive feedback is provided. Pronunciation is added to the spelling task. When the task is to spell a word, each letter is pronounced by a voice and the whole letter is pronounced when the task is completed.

5 Discussion
By looking at these two different applications it becomes clear that feedback can be designed in many different ways in an application. The two studied applications provide feedback differently and have different areas were they succeed and were they do not. They both have a lot to learn from each other. Both applications give feedback mostly on a task level, but the Zcooly application manage to provide a deeper and more informative negative feedback that can guide the user about what to do in order to improve and succeed with the task.

The fact that the Zcooly application is very clear and informative in the feedback given could, however, also have a negative effect. As shown by Van Duijvenvoorde et al. [17] feedback to children of this age should be mainly positive since children handles negative feedback different than adults and seems to decrease their performance if the feedback is negative. As this application is designed, a pupil who has troubles with the tasks will receive a lot of negative feedback. Even though some part of the task may have been performed correctly, this is not noted as long as the whole task is not performed correctly. The Fun English application gives less negative feedback, since it sometimes gives no feedback at all as long as the task is not performed correctly. Even though it is a good thing to reduce the amount of negative feedback, the absence of feedback is not the best solution according to other studies implying the importance of informative feedback [10].

When it comes to providing instructive feedback none of the studied application seems to focus on this. Zcooly gives no instructive feedback at all while the Fun English gives a kind of instructive feedback as letters and words are pronounced when the task is simply to spell the word. The pronunciation is not necessary to complete the task and can thereby be seen as an instructive feedback that extends the learning goal of the task. In the other exercise no instructive feedback is given, but could have been added in the same way. Here a word is only pronounced, but could also have been shown written in order to add a visual recognition to the hearing recognition. Also in the Zcooly application it should be possible to add instructive information by adding additional
information when feedback is given. By adding this kind of additional feedback the learning outcome of each exercise could be enhanced [20, 21].

One area where applications seems to have an advantage is the aspect of timing. When it comes to more traditional ways of teaching, the teacher can not follow each pupil's every step and give feedback timely on every action performed by the pupil. An application, on the other hand, can provide instant feedback at all times, which both of the studied applications manage to do. The feedback is given timely and regularly throughout the exercises in both Zcooly and Fun English. The difference between the two is that Zcooly does not give feedback until the user has implied that he/she is finished with the task. This is a way to minimize feedback given on accidental actions, and can thereby be a way to reduce negative feedback. As shown in the studied applications, the application can also make sure that a task is performed correctly by not allowing the user to move on to the next task before the first one is finished. Both applications together shows that it is possible to reach many of the guidelines for effective feedback in an application. Zcooly manages to give informative feedback that explains in detail what went wrong and gives a guidance of what to do in order to improve. The application does not only provide feedback when something is done incorrect, it also provides positive feedback when the user has performed a task correctly. This application also provides a clear description about the goal of the task. By providing a combination of different feedback types this application can also work for students with different abilities. By a combination of both visual, written and talked feedback, the feedback can be understood despite things like reading or hearing disabilities. Even though the application does not actively adapt to different users, this is a way to handle the aspect of adaption. Fun English manages to give some instructive feedback and thereby enhance the learning. This application also gives little negative feedback, which is important for the aimed age group. Both applications provide feedback on a task level according to Hattie's definition [10], even though the Zcooly application provides a more informative version of this feedback. As Hattie's study further implied it should be more effective to give feedback on a higher level that provides a deeper understanding and makes the feedback transferable to other similar tasks.

One of the main issues with feedback from an application compared to feedback from a teacher is that they do not provide the opportunity to discuss the given feedback or ask questions about it, it does not provide a conversational feedback. As shown in the earlier studies, a conversational feedback makes it easier for the student to understand the given feedback and gives the opportunity to clear out misunderstandings of both teachers and students [11].

A counter-argument to the lack of conversation may be that the tasks to be solved in an application is not always fully comparable to a student task that a teacher corrects. As seen in the applications, feedback is mostly given for very specific and small subtasks, where the answer is an unequivocal right or wrong. In situations where the teacher gives feedback to the student, it may cover larger tasks where the answer may not always be as obviously is right or wrong, and thereby the need of asking questions or discussing the feedback.
may differ. Still a conversational feedback has great advantages as shown in the literature review [9, 11, 19]. Even though the tasks are often quite simple in the applications, misunderstandings could occur considering the goal of the task, or there might be deeper misunderstandings that hinders the user to perform the task correctly. These things might be hard to discover and correct when there is no conversation available. Perhaps a solution to this problem could be for pupils to use these applications in close collaboration with the teacher. If the teacher is well familiar with the application, he or she could answer questions that occur and detect misunderstandings by the pupil.

Another great disadvantage with applications is the lack of individualization and personal adaption [23]. The application does not have knowledge about the user the way that a teacher does about a student. A teacher can adapt his/her way of giving feedback based on individual aspects, the current situation and the current task in a way that the applications today can not since they follow a programed, standardized feedback system. Perhaps it is also possible that both a better adaption and detection of misunderstandings could be reached within the application by using more advanced techniques and other ways to design these applications in the future.

The study further showed the importance of being clear about the goal, and when it comes to an application a technical aspect is also added to the understanding of the task. Even if the goal is clear and easy to understand, the user must also understand how to technically proceed to solve the task. One example of this is the task in the Zcooky application. The task contains many steps, like dragging items to a certain place on the touch screen, clicking at right places and so on, and it is not completely obvious how to solve this tasks even if you know the mathematical answer (which is the goal). The Zcooky application has a quite clear description of the main goal, explaining step by step what the user is supposed to do, but this description is shown only if the user actively choose to click the question mark. The standard information given if this button is not clocked is not very detailed or descriptive and could probably easily be misunderstood. Also the Fun English application is lacking in how the goal of the task is presented and there is no way to find more information. If the main goal is unclear as in the studied cases, the user could get confused and the errors might be caused simply by a misunderstanding of the goal.

6 Conclusions

The literature review found that there are many theories and studies defining good feedback in a learning situation. Some theories differ from each other, but a number of guidelines that most studies seemed to support could be identified. The second part of the study showed that many of these guidelines for good feedback can be reach in an educational application. Feedback in an application compared to a traditional teacher to student feedback was shown to have both advantages and disadvantages. Advantages of feedback in applications is first of all the aspect of timing and the possibility to provide a combination of different
feedback modes, such as written, visual and auditive. The most critical shortcomings in applications today is the lack of individualization and conversational feedback.

Further studies are needed to determine whether these shortcomings could be overcome with the help of new technologies or new ways to create and design educational applications.

References

11. Orsmond, P., Merry, S., Reiling, K.: The student use of tutor formative feedback in their learning, Learning Communities and Assessment Cultures Conference, Education-line database (2002)
Push-Notifying the End-User at the Most Appropriate Time

Mattias Edin
Umeå University, Sweden
maed0101@student.umu.se

Abstract. Smartphones have notifications for various kinds, for instance emails, social network updates, messages. Although little is known how stress affects how receptive smartphone users are to notifications. In this article, an in-situ study consisting of 16 smartphone users where response time was collected for 211 notifications in total. An Android application were installed on the users to conduct the survey. We found that higher stressed participants have a longer respond time to notifications. The notifications that are attended to within 1 minute consists of mostly low stressed participants. These findings show that low stressed participants are more susceptible to notifications.

Keywords: health, monitor, measure, stress, technology, interruptions, context aware, notifications, disruption, digital stress, technostress.

1 Introduction

While the smartphone market has grown in the past, the prediction is made that it will continue to do so in the near future [1]. One big purpose of these devices is to make one and other more connected to each other. Unfortunately, this connective also comes with a price. The price is not only the vendors price tag on the product but the mental health that can get affected by the use of the products. The mobile use is causing stress and sleep disorder on its users [2]. While this mainly is affecting users with an unhealthy use of their devices it does, in fact, show that there is an underlying problem with our use of them. The smartphone devices are enabling us to use social media services wherever we are and at any time. This, in turn, have shown to have a negative effect on the well being of the user [3].

There are commonly used applications such as Facebook, Twitter or most of the messaging or mail applications that utilise push notification services to send notifications to the user. A push notification is a notification that shows when the client user did not explicitly request the
information, the notification is, therefore, a message that shows up without warning and can surprise the user. This could be the reason for the interruptive behaviour of the push notifications. If the service that pushes the notification could know when the user is more susceptible for notifications, the interruptive behaviour may be minimised. A study done by Spira and Feintuch [4] in 2006 shows that approximately 28% of a workday is taken up by different interruptions. To make the matter worse when the average interruption is 2.8 seconds the interrupted person tends to do twice as many errors in the current task [5]. One explanation for this could be why digital stress has a significant negative impact on productivity as resulting in a study carried out in 2007 [6]. Although this data is already showing the problems of interruptions but fails to analyse how stress affects our receptivity of push notifications.

1.1 Aims and objective

The goal of this study is to evaluate the current push notification system by looking at how likely the user is to attend to notifications during different times during the day and how stressed the user feels at these specific times. In turn, this, could be used to find possible times of the day when the user is more susceptible to notifications and also take into account of how stressed the user feels at the same time. With this study, a conclusion about specific times the user will be more likely to attend to the notification was drawn. This information can be used for the implementation of future mobile applications that utilise the push notification system to send notifications when the user wants them rather when the service want you to have them. This, in turn, could make the user experience better by eliminating annoyances of receiving notifications when not wanted.

- At which time of the day the users of smartphones are more susceptible attend to push-notifications?
- Is there a correlation between the time it takes for the user to look at the notification and how stressed the user feels?

2 Theoretical framework

In this study, the focus of what precise type of stress the user is experiencing is not of most importance but the negative feeling of it the user has. Therefore, the user is free to define its stress as a negative feeling that affects the mental health at the current time. As described under the
section Aims and objective the study does not focus on the stress level itself but how it affects its user of how push notifications are perceived. It is, for this reason, the absent of explanation in stress. According to the American Institute of Stress, stress in common sense a term that cannot be defined as it is a subjective phenomenon and is described instead as follows "a condition or feeling experienced when a person perceives that demands exceed the personal and social resources the individual is able to mobilize" [7]. According to Selye [8] "Stress is the non-specific response of the body to any demand for change." Stress is, therefore, vague in the sense of definition. Stress has according to American Institute of Stress both good and bad impact on the human body [7]. In this article we have chosen to focus on the bad impact stress has on the user and in this case the user of a smartphone.

2.1 Technostress

The concept of technostress is stress caused by an inability to cope and process with all the information that is surrounding the user. This is also described as when there is a decrease in productivity and satisfaction of the end product of different tasks. The unhealthy usage that this article is referring to is when the smartphone is distracting the end user from its primary task. Therefore, affecting productivity, as described in the related work of Brooks and Stoney where stress has a correlation with the use of social media, where the use of social media had a negative impact on the performance of the current task, as well as and lower happiness [3].

As described by Arnetz and Wiholm [9], technostress can be defined as follows; "the state of mental and physiological arousal observed in certain employees who are heavily dependent on computers in their work". This definition is from the year 1997 and could therefore perhaps extend to not only employees because it can be easy to think that people have become more dependent on computers in the form of smartphones in recent years that is not only work related.

2.2 Push notification

In the simplest form, a notification is defined in this article as a message coupled with either sound or/and light flashes from the clients device. The message is brief and implies that the user should interact with it by either opening it for the full message or be dismissed. Either way, the message has had an interactive effect on the user by being in attention. Simple notifications do not describe what the cause of the notification is
so the notification could be a result of the user’s action. For example, if the user starts a download of a File, the user will also create a notification that shows the downloading progress. An extension of a simple notification is Push Notifications (PNs) which also describes the initiation of the notification. The PNs is initiated without the end users explicit action and is often done remotely from within a subscription service that the user has explicitly signed up for. This could in turn possibly interrupt the end user and increasing mental, cognitive workload.

2.3 Related work

There are a lot of related work done in the topic of notifications where one study published in 2014 [10] showed that people have to deal with 63.5 notifications on average per day who also found that notifications were most often attended to within one minute. Another related study published in 2010 showed that desktop users receiving email notifications showed that the users thought the notifications to be disruptive but still wanted them for more awareness. Another article published in 2014 showed that about 49% of the notifications on smartphone users were attended to within one minute. The survey also showed that if the notification was coupled with sound, vibration or a light cue or any combination of them is 12 times more likely to attend to the notification immediately while the time of day had no evident effect. A related study conducted by Microsoft Research [11] in 2010 showed that instant messaging had a disruptive effect on the current tasks the user is doing. The study also showed that if the message is highly relevant to the current task the descriptiveness effect on performance was not as strong. Another study published in 2007 [12] showed that instant messaging such as short message service (SMS) had the effect of prolonged stress and sleep disorder on some of the subjects.

This article distances itself from these studies by looking at how stress has the direct impact of how fast we attend to PNs. The above studies all has their corresponding topic but with this survey, we found a connection between stress and how prone one is to attend to a PN.

3 Method

To get an understanding of how likely one is to attend to push notifications (PNs) while being affected by different amount stress and workloads, an Android application was constructed and installed on the subjects smartphones to monitor PNs. The application randomly received PNs
sent by an external service to the test subject at different times of day, a service in the application monitored the time it takes the user to take a look at the notification. It is important to note that the notifications from the notification did not try to mimic other applications as explained later in the section: Discussion. When the subject looks at the notification, they will give a value of how busy they are feeling at the current time and the time is recorded. When the test subject first installs the application, they will register for the survey. During the time of registration the test subject gives a value of how stressed they are feeling at the time of the registration see section 3.1 for how stressed is measured. The data was then relayed to a web server. The test subjects had the application installed for seven days and during this time the notifications were received to their smartphones at random times but within 8.00am to 9.00pm. There are four different timestamps that were collected by the survey, first when the notification was sent, when the notification is received by the smartphone, when the subject opens the notification and finally when the notification is later received and stored on the web server. The important time stamp is when the subject receives the PN in their smartphones which are going to be used as reference time throughout this study. It is important to note that the user can dismiss the PN sent from the server which in turn will not be taken into consideration when calculating the time subject takes to attend to the PN. The participants were not told that the survey also looks at the time it takes for the participant to attend to the PN.

For each PN there are two questions the participants had to answer, one is defined in section 3.2 and the other question is a randomly chosen question from a set of five different questions. The second question is not used for the survey and is there only for making the survey more varied and interesting for the participant. If each PN was exactly the same the participant could lose interest in the survey which could potentially have an impact on the result. The other reason why we chose to have this question is also to make the survey more realistic because PNs are from our experience seldom contains the same information all the time in a realistic situation. All questions that were asked used the same answering system, that is pre-defined responses choices with a drop down menu. The reason for this is for making the survey as less time consuming as possible for the participant. This is important because if each PN were associated with a time-consuming task there would be a bigger risk for the survey to make an impact on itself, where the response time could increase just because the participant knows that if the notification is clicked they have to postpone their current task.
The number of participants N=16 consists of 11 men and 5 women with occupations division of 13 students, 2 employed and 1 other with age range from 20 to 56.

In this study, we have chosen not to include the study of specific applications like Facebook or e-mail notifications, because of the comprehension of the survey. Each application needs its independent study to find differences between them.

### 3.1 Stress measurement

The stress measurement was derived from the QPS Nordic questionnaire [13] which has good validity as studied by related work [14]: "By stress is meant a state when a person feels tense, restless or worried or cannot sleep at night because he/she is thinking about problems all the time. Are you experiencing this kind of stress currently?" The response scale was 1–5: 1=not at all, 2=just a little, 3=to some extent, 4=rather much, 5=very much. The question was originally in Swedish and translated by *Scandinavian Journal of Work, Environment & Health*. The responses were also divided furthermore into two groups; High (response sets 4–5) and a Low (response sets 1–3) to minimise the complexity of the data set. To measure stress with only one question with good validity is necessary for the simplicity of the application. More work had to be put down in developing the application if another stress measurement would be chosen that weight in more factors. For instance, one could choose to measure the stress with an activity band with sensory data. Although for this study it was sufficient to have confidence in the test subjects own answer range from 1 to 5. Another reason that one question is a good choice is that of making the survey easier on the participants. Fewer questions would not have a smaller impact on the participants choice of participating in the study.

### 3.2 Busy measurement

To measure how busy the test subjects are feeling one question was constructed as follows; "When a person feels busy, one feels that some work has to be prioritised and left out because not all can be done in a limited time span. Are you feeling this kind busyness?". This question was derived from the tested stress question to maintain the same type of question where the reader defines one’s current situation according to a definition which is defined in the question itself. The response scale is the same as for the stress question.
3.3 Procedure

Recruitment of participants to the survey was mainly done through self voluntarily assigning from an announcement on a social media page on Facebook. The participant that volunteered were then contacted with both e-mail and personal messages on Facebook. The test subject’s were then added to the testing group and given a link to download the application for this survey. Once installed the test subject made a one-time registration from inside the application. The test subjects were then asked to have the application installed for 7 days. During the test period each subject received in average about 13 notifications which are scattered at different times of the day.

4 Results

During the test 275 notifications were sent and received by the test subjects and from those 42 notifications was dismissed. The dismissed notifications were removed from the dataset. Because of extreme values, the charts in the figures (1, 2, 3) had to be zoomed to make the it easier to see the differences between the datasets. According to the median in figure 2b the stress have a clear impact on the response time. Test subjects that answered low stress have a median of 4.9 minutes and high stress with a median of 17.3 minutes in response time. The upper quartile (25% of the dataset) in figure 2b for high and low stressed participants consists of data points within an interval consists of 46 and 133 seconds in respond time while the high stressed participants have the quartile positioned about 10 minutes above that of the low stressed participants. This means that high stressed participants have a tendency of having a longer response time when considering the upper half of the results in those datasets. Although the lower quartile of the data set for high stressed participant also cover the same region as low stressed participant which shows that there is a slight lower certainty for this result. The same finding could be found for how busy the participants are feeling. According to figure 3b the participants with high busyness have if considering the median a difference in 12.4 minutes where high busyness tends to have a higher response time.

27% (63 data points) of all notifications were attended to within 1 minute (see figure 4) where the rest are scattered to as high as 26.7 hours as respond time at the most. To elaborate the data set which was attended to within one minute one could find a big difference from how stressed the participants was. A difference of 364% more of the data points attended
to within 1 minute has low stress. This means that participants with low stress have a higher probability of attending to a notification than high stressed participant during the first minute (see figure 5). From figure 6 there is a hint of better receptivity with PN during noon although more tests have to be carried out to confirm this.

One abnormal finding in the result is how the test subjects that answered the question of how stressed they feel (see section 3.1 for questionnaire), see figure 3, with the answer 'Not at all' and 'Just a little' all reported to have a very fast respond time, median of only 12 and 43.5 seconds where the other datasets have a median of 13.9, 29.5, 12.6 minutes (see figure 2a). This could be interpreted that the very low stressed participant are very fast at attending to PNs.

One more interesting finding is how the response time median relates to different hours of the day. As can be seen in figure 7 the hours 7 and 17 have an obvious higher response time if considering only the median.

**Response time for each weekday**

![Response time for each weekday](image)

Fig. 1: All data points grouped by weekday. Datasets in (b) is displayed without extreme values with values in range of 0-40 minutes to make the differences more readable.
Fig. 2: Box plot over response time grouped by stress. Datapoints in (b) is displayed with grouped values where datasets 1-3 is low and 4-5 is high.

Fig. 3: Box plot over response time grouped by busyness. Datapoints in (b) is displayed with grouped values where datasets 1-3 is low and 4-5 is high.
Fig. 4: Histogram over datasets with respond time within 0 and 30 minutes.

Fig. 5: Extracted datapoints which responded within 1 minute grouped by stress.
Fig. 6: Percentage responded within 1 minute for each hour.

Fig. 7: Median response time for each hour where annotation for each bar is the number of values the median is calculated from.
5 Discussion

According to related work\cite{10} there is significant difference between the
time it takes for users to attend to notifications during the weekdays and
weekends. This is not the case of the findings of this survey where the
result has no noticeable difference between the two (see figure 1).

Because the test subject was not instructed to know that the response
time is important for the survey the result should therefore not affect the
data. One thing that could be affecting the data is the fact that the
participant knows that all the PNs is coming from the survey. This fact
could alter the importance of attending to the PN and therefore, be more
important for the user than other PN, this, in turn, could make the user
more prone to attend to it. The participant could therefore in other cases
not be as prone to attend to other notifications because the importance
is not as high so the response time could be affected.

There is a subjective matter of what notifications that are more urgent
than others which will, in turn, have an effect on the result. The urgency
of the notification can, in turn, explain the reaction time that the user
has when to look at the notifications. For example, some user may argue
the fact that a regular text message is more important than a Facebook
message equivalent and the reason for this is mainly subjective on behalf
of the client user. The point of this is that different people value different
things and that can explain the deviation in the result. But the user
can be theorised not to change their personal importance level of the test
applications notification during the survey so the fluctuations in the result
are greater than the average response time. What it is that makes these
differences can have to do with the notification type, how it notifies the
user. In many cases, the text message has user-defined sounds to it while
Facebook has pre-defined sounds. If the user has configured the sound,
he or she may be more prone to get interrupted by the notification in
comparison to the generic pre-defined, simply because it is more personal.
There may also be a difference in how the user values the content of
the application the notification originated from, for instance, an e-mail
notification may have a high user value because it has a high chance of
being work related, and could, therefore, be important. In these cases, the
originated application is of importance. Also, there may be a difference
in the reception of the notification depending on the type of notification
(sound or vibration). As Gallud and Tesorier\cite{15} found there is a slight
difference in how the user has defined their notification setting. Also, there
is a noticeable difference in how the users handle notifications depending on the quantity they receive.

This presented study has a limitation where the number of participants is 16 Android users, where only 5 of those are female where the collected data have a span of 7 days. Further work may have to get a greater number of participants to make more accurate statistical evaluations. A second limitation is the fact that the survey is limited to Android users, where other platforms may notify the user in a different way that could therefore have an impact on the result. The resulting conclusions are therefore only limited to how the current Android operation system handles notifications.

6 Conclusion

To answer the objectives laid out in section 1.1, the time the user is more susceptible has two factors to take into consideration. How stressed or busy the user is and time it takes for the user to attend to the notification. The result of the study showed that there is a big proportion of the notification that is attended to within 1 minute, namely 27% and the time is considered to be the noon of the day (see section 4 for further explanation). For the second objective, to find out if there are any differences for stressed participants or busy participants, the survey showed that there is a difference in the two. Participants with high stress tend to have a longer respond time for notifications as well as busy participants.

References

Aesthetic value versus function in mobile applications

Robin Schulze
Department of Applied Physics and Electronics
Umeå University, Sweden
id12rse@cs.umu.se

Abstract.

Mobile applications are going through a lot of changes. Developers are always trying to improve their aesthetic value and functionality. That is why the objective of the study is to identify whenever aesthetic value or function matters most for users of mobile applications. To be able to identify this, researching what defines aesthetic value and a function is necessary. It is also need to understand how mobile application works today. The study was made by letting seven participants preform three different tasks. They had to preform each task twice once with an application that was built to be more functional and the other to be more aesthetic pleasing. The tasks that were chosen were: "Searching in a list", "Clearing all input" and "Retyping same text". After the participants had preformed all task they were asked which one they preferred better. They were also asked if the application they did no prefer could be improved to make them prefer it instead. The results from the study shows that in two out of three tasks the participants prefers the functional application over the aesthetic pleasing and most did not think the aesthetic pleasing application could be improved to be chosen instead. Although it is a highly preference what the users prefer, the results might be unreliable. But it should at least give a better understanding to what the users prefers most of aesthetic value and function.

1 Introduction

Throughout the 20th century, mobile applications have been through big changes. Developers are continuously trying to improve what applications can offer the users to satisfy their needs. According to M. Enis, the popularity and functionality of mobile applications are continuously growing [1]. In the study "The evolution of mobile apps: An exploratory study", it was concluded that the complexity of mobile applications is continuously increasing [2].

Aesthetic value is also increasingly important in mobile applications. According to J. Silvennoine et.al "Visual attractiveness is increasingly seen as an essential factor in perceived usability, interaction, and overall appraisal of user interfaces." [3]. Which makes the increasing aesthetic value of the application a cornerstone in making it more visual attractive [3]. In the The Encyclopedia of Human-Computer Interaction(HCI), 2nd Ed.: Visual Aesthetics, N. Tractinsky
argues the importance of visual aesthetic in HCI applications from three different perspectives, the design perspective, the psychological perspective, and the practical perspective [4]. He states "I believe that, taken together they cover the lion share of arguments for the inclusion of visual aesthetics as a major aspect of HCI practice, research and education [4].

With the increase of popularity, complexity and functionality one could not help to wonder if it is the functionality or aesthetic value that matters most for users in applications. This study was conducted to get a deeper understanding of what users value the most of aesthetic value and function and if there is any difference between mobile applications.

1.1 Objective

The goal of this study was to identify and analyze what users value the most, aesthetic value or function, in mobile applications. The purpose was to conclude what matters the most for users today of the two. Furthermore, the study examined what defines aesthetic value. The functions that were chosen were "Searching in a list", "Clearing all input with a click" and "Remembering inputs and give suggestion for next input". A user study was conducted to gather information about the users preferences and thoughts on aesthetic value and function. The information and thoughts were analyzed to conclude what users favors the most of aesthetic value and function in both mobile applications.

The objectives of the study were as follows:

(a) Identify whenever aesthetic value or function matters most for users of mobile applications.

2 Theory

This study is making use of other articles to better understand and grasp the definition of some concepts. Furthermore other articles were used to find important information that was later used in the user study of this study.

2.1 Aesthetic value

What is aesthetic value? According to L. Plato and A. Meskin one definition of aesthetic value could be: "Aesthetic value is the value that an object, event or state of affairs (most paradigmatically an art work or the natural environment) possesses in virtue of its capacity to elicit pleasure (positive value) or displeasure (negative value) when appreciated or experienced aesthetically." [5]. This could be interpreted as aesthetic value is something that gives the user pleasure or "positive value" when experienced.

This definition of aesthetic value was used when conducting the user study. Another common definition of aesthetic is “an artistically beautiful or pleasing appearance” or “a pleasing appearance or effect: Beauty” which N. Tractinsky uses when he talks about aesthetic value [3].
2.2 Functions

Defining what a function is has proven difficult. But according to V. Beal one definition of a function could be that it is a section of an application that preforms a specific task \[6\]. That would mean in example that a function could be a button changing colour of a specified text. The functions that were chosen will be described more thoroughly below.

2.2.1 Searching in a list

The function searching in a list is implemented by having a text field where the user can input a text string and the application will there after search for that text string in the list. The application will then show any items that match the input text. If there are no items that matches the input text, no items will be shown \[7\]. Google’s search engine can be seen as searching trough a list of sites, that is why this function was chosen.

2.2.2 Clearing all input with a click

Clearing all input with a click is implemented in the way that when a user clicks a button all the input into a field is cleared. The field will then be empty and new text can be input. Google’s mobile site uses this function to make it faster for the user to clear the text area.

2.2.3 Remembering inputs and give suggestion for next input

Remembering inputs and give suggestions for next input is implemented by having the application remember what the user input and suggest it when the user is typing in another field that should contain the same information. Google’s search engine gives suggestions on what you have searched for before and that is why this function was chosen.

2.3 Applications

In "Computer Basics: Understanding Applications" the Goodwill Community Foundation is defining applications as a software that allows the user to complete specific tasks \[8\]. That would mean that an application consists of one or more functions as defined earlier in 2.2 that completes different task for the user. Mobile applications run inside the mobile operating system \[8\]. A mobile application can for example be used by a user to manage health and wellness \[9\]. A mobile application can also contain ads to help the developer of the application generate revenue \[10\].

3 Method

Here the method of how the user study’s testing applications were built and how the study itself was conducted will be presented. The study was done in the span of two weeks.
3.1 User study

The user study was conducted by first creating two different testing applications for mobile devices. One that were aesthetic pleasing for the user but without a function that eased the working process, one that had a function too ease the working process but did not include the aesthetic pleasing elements. The applications were minimalistic done and had only the elements to simulate one of the two application scenarios.

3.1.1 Participants

The participants in this user study were males and females from 20 to 30 years old. The thought behind the choice of target group were that the target group will be using applications in the future and they have been around applications longer then the ones younger than them. The participants was not picked by any other criteria. Participants were randomly selected from students at Umeå University. They were first approached and asked about their age to see if they met the criteria and then asked if they wanted to participate in a study. The study was conducted on seven persons.

3.1.2 User tests

All information gathered from the study was recorded in a text document. When the participants agreed to take part in the study they were first presented with ethics information which they read. After that they were asked what age and gender they were. They then continued to test the two applications. Which application they started with was set so half of the people started with one and the other half with the other one. They were asked to perform the stated tasks on the two applications. Before they started with the "Aesthetic pleasing application" they were asked if they found the application to be aesthetic pleasing or beautiful. At the end, the users were asked if the element (functional or aesthetic pleasing) of any of the three applications they did not choose could be heightened to be chosen instead. This was done to see if there could be some scenario where they would have chosen the other application. The results from each user were recorded to a survey which the observer filled in.

Tasks

(a) **Searching in a list.**

The users were asked to search for five items in a 100 item long list. The items which they were to search for were randomized.

(b) **Clearing all input.**

In this task the users were asked to type in five sentences with 5 to 20 words that were randomly chosen from a list of 20 sentences. After they had written one sentence they had to clear the field before writing the next one.

(c) **Retyping same text.**

In the last task the users were ask to fill out five forms. The forms were identical and the same information were needed in the forms. When one form was filled the users moved on to the next form.
3.1.3 Aesthetic pleasing application

The Aesthetic pleasing application was simulated by using the existing application "MobisleNotes". It had minimalistic elements to simulate aesthetic value for the user. The application was used in all three parts to test, "Searching in a list", "Clearing all input" and "Retyping same text".

3.1.4 Application with function to ease the working process

The application with function to ease the working process was written in html. It had minimalistic elements to simulate functionality for the user. The application was split into three parts to test, "Searching in a list", "Clearing all input" and "Retyping same text".

3.1.5 Environment

The tests was carried out on an iPhone 6 with iOS 8.1.2 installed. The tests were conducted in a small room where there was no immediate distractions for the participants.

3.1.6 Evaluating the test applications

When the applications were created they were tested by two chosen test persons to see if they yielded the right results. When the applications were tweaked so they yielded the right results the study continued to making user tests.

4 Results

Here the results of the study is presented. First the chosen application and built application will be shown and how the built application were altered after the test of the application were conducted. Thereafter the user study’s results will be presented.

4.1 Applications

Here the results regarding the application will be shown. The results of evaluating the application and the final application will be presented.

4.1.1 Evaluating the test applications

The application was tested on two persons to see if it yielded the right results and that the applications was built with the right elements. The test showed that the functional application for evaluating "Searching in a list" was clunky and were hard to use. To solve that, the result when searching for a item was instead shown at the top and the user did not need to scroll through the whole
list just to find the item highlighted. Another thing that was brought to light when evaluating the functional application that tested "Retyping same text" and "Clearing all input" was that the user used the mobile devices built in function to remove all text. To solve this the user is now instructed to not use this method.

4.1.2 Final applications

The final application that were used for the user tests will be shown below. **Searching in a list**

![Image](image.png)

**Fig. 1.** Final application for evaluating functionality for the task "Searching in a list".
Fig. 2. Final application for evaluating aesthetic value for the task "Searching in a list".

Clearing all input

Fig. 3. Final application for evaluating functionality for the task "Clearing all input".
Fig. 4. Final application for evaluating aesthetic value for the task "Clearing all input".

Retyping same text

Fig. 5. Final application for evaluating functionality for the task "Retyping same text".
4.2 User study

The results from the user study is presented here. The study was conducted on 9 users. Although one user did not think the aesthetic pleasing application was aesthetic pleasing so the results from that user were not included in the results.

4.2.1 Searching in a list

Below the results from the task "Searching in a list" is presented. Six of seven users preferred the functional application on this test. None of the seven thought that the aesthetic value could be increased to make them switch. Although the user who chose the aesthetic pleasing though that the functional aspect of the test application could be improved to make the user chose the functional application instead. The users who chose the functional application did not think that the aesthetic pleasing application could be improved to make them choose it instead.
4.2.2 Clearing all input

Here the results from the task "Clearing all input" will be presented. In this task four out of seven chose the functional application. And when on the question if the user thought the application they did not chose could be improved to be chosen instead only one answered that the aesthetic pleasing application could be
improved to be chosen instead and the other seven said that it was not possible that other application they did not chose could be improved to be chosen instead.

![Pie chart of what application the users liked the most in the task "Clearing all input".](image)

**Fig. 9.** Pie chart of what application the users liked the most in the task "Clearing all input".

![Pie chart over if the users thought that the application they did not chose in task "Clearing all input" could be improved to make them chose that one instead.](image)

**Fig. 10.** Pie chart over if the users thought that the application they did not chose in task "Clearing all input" could be improved to make them chose that one instead.
4.2.3 Retyping same text

Below the results from the task "Retyping same text" is shown in pie charts. All seven users chose the functional application in this task. Although two users though that the aesthetic pleasing application could be improved to make them chose it instead.

**Fig. 11.** Pie chart of what application the users liked the most in the task "Retyping same text".

**Fig. 12.** Pie chart over if the users thought that the application they did not chose in task "Retyping same text" could be improved to make them chose that one instead.
5 Discussion

In this section of the report I will voice my own thoughts about the study, the results and the conclusions I drew from those results.

5.1 Study

Here I voice my thoughts and opinions from the study. I will go through the subjects: Participants, Tests and Results. I will be talking about function and aesthetic levels to somewhat compare them.

5.1.1 Participants

I would first like to say that I would really have seen that I had the time to test more users. While I feel that seven users still may yield a taste of what all the users prefer, I would want to have tested at least 20 more users to see that my seven users were no special cases and that they are in line with the rest. What I also would have like was to have users from other universities to get a broader spectrum of users and that their opinions is not local opinions. Otherwise I am pleased with the chosen age. I feel like it was the right generation to make the study on since they are one of the generations that will be using applications in the future, they have good knowledge of how things work and gave constructive and complete answers in the study.

5.1.2 Tests

I am pleased with how the tests went. The users were really interested and made good points for their choices. I made a mistake to not take notes of all the good points they made. It would have been great to have in the results to compare to each user. But on the other hand, then I would have wanted all users to say something similar or else I feel that I can not compare users who was not as elaborate as some other users. One thing that was a challenge in creating the test was to balance them. I noticed that this was impossible due to the fact that there is no way to tell if the aesthetic value in one application equals to the functionality in another application. This is why I asked them the follow up question if they thought that it was possible to improve the application they did not chose to make them chose that one instead. Although I feel that they at least are close to being on the same level.

5.1.3 Results

Here I will discuss the results and the thoughts about the results from the three tasks.

Searching in a list
Starting with the first task "Searching in a list" we can clearly see that the majority of users chose the functional application. Also none of the ones who chose the functional thought that the aesthetic could be improved to out weight the function of a search box that returned the number of the typed word. Also the one user who valued the aesthetic pleasing application thought that the functional application could be improved to be chosen instead. When just looking at the first question one could think that the functionality just was on a different level then the aesthetic value. But when considering the second question too, we can clearly determine that in this task the users would much rather prefer the functional application. Which means that the function is valued over aesthetic value in this task.

Clearing all input
Moving on to the second task "Clearing all input" we can see that it is quite even between the two applications, but the functional application is still in the lead. Four of seven chose the functional application. And looking at the follow up question no user but one thought the other application could be improved. The last one thought that aesthetic could be improved to make the user chose that one instead. In this task the both application is even and no application is highly favored. I feel neither aesthetic value or function is valued higher in this task.

Retyping same text
In the last task "Retyping same text" all the users chose the functional application. This may be because the function introduced in the functional application saves a lot of time writing so it could be that it is just on a higher level then the aesthetic application. Continuing to the second question, we see that only two users thought that the aesthetic pleasing application could be improved to be chosen instead. Which makes a big majority chose the functional application. Which puts function over aesthetic value in this task.

5.2 Conclusion
In two of the three task the functional application was highly favored compared to the aesthetic pleasing. And most user did not think that the aesthetic pleasing application could be improved to a point where they would chose it instead. In the last task we can not see that any of the two application is favored. This would mean that function matter the most for users in a mobile application. But this does not mean that this is true for all more functionally applications and aesthetic pleasing applications. Even though I did ask them if they thought the applications could be improved to make them switch application, they might not have though of a scenario or an improvement that would make them change. Also, I feel what users value the most is also highly individual. Some have a higher love for functional applications in general and some have higher for aesthetic pleasing. Therefor more work needs to be done in the subject to give a deeper understanding. But I do think these two elements go hand in hand when
designing a mobile application. To have a successful application you will need to have both, with some exceptions. But now we have a general idea to what is more important.

5.3 Future work

For future work in the subject we should make tests with a higher amount of participants, more applications with different functions and aesthetic pleasing elements and from different geographic locations to get a broader spectrum. What we can learn from this study is that remember to record all the thoughts from the user meanwhile they make the test.
References

Feeling Safe in Autonomous Vehicles

Gustav Nordlander
Department of Applied Physics and Electronics
Umeå University, Sweden
contact@gustav-nordlander.se

Abstract. This study was conducted to examine what was needed to feel safe in an autonomous car and how existing models can be modified to help the user feel safe. This was done by researching why people might not feel safe in an autonomous vehicle (AV) by discussing certain elements with potential users through interviews. Research was also conducted to understand what a person needs to feel safe and what could be the reason why the feeling is lacking. The results showed that most people want to have a steering wheel and a foot brake available at all times, even when the car is fully autonomous. Almost everyone also wanted to be able to quickly take control over the vehicle at any time and with that automatically shut down the autonomous part. The conclusion was drawn that people will feel safe in AV when they feel that they are in control of the situation, this includes that they need to feel that they will know what the AV will do in most situations as that allows you to anticipate and plan accordingly.

1 Introduction

Much of the technology around us is evolving to require little or no interaction but still be useful for us. This is possible due to location services and sensors like thermometers and pulse measurement. Our devices should give us information and help us depending on what data it gets from the different sensors. Cars are following the same evolution with them sending data between vehicles to i.e. change the route of your GPS to decrease traffic [1].

Today, a hot topic within automobile technology is autonomous vehicles (AV), which means that the vehicle, in some extent, drive without the help of the actual driver. There are several different takes on this and the levels of automation varies but the main goal is, as with all technology to make it easier, more sustainable and removing the human error that can be caused for whatever reason to reduce traffic accidents. The Insurance Institute for Highway Safety (IIHS) [2] predicts that with autonomous functions such as blind spot assist, adaptive headlights and forward collision warning systems one third of all crashes and fatalities could be prevented.

As technology becomes increasingly common for almost everyone in the industrialized world, technophobia becomes a relevant subject as close to one third of the industrialized population harbor some technophobia, the fear of technology [3]. In addition to this, humans have a tendency to feel safe when in control
and having knowledge of about what is about to happen [4]. So how can we use our knowledge to design an autonomous car in which we feel safe?

1.1 Objective

The aim of this study is to identify what makes a human willing to trust a computer in such a way that he or she can relax while the computer is doing something that can have severe consequences if done recklessly or incorrectly, such as driving. The questions that will be answered in this study is as follows;

- What kind of assurance is needed for drivers to be comfortable with autonomous vehicles?
- How can we design such a product, which elements are essential?

2 Theory

Autonomous Vehicles is today one of the most challenging research topics in intelligent transportation and has been a hot topic of research for the past few years. AV’s that control their own steering and speed are expected to account for 75% of vehicles on the road by 2040 [5].

2.1 Levels of Autonomy

National Highway Traffic Safety Administration [2] have defined different levels of autonomy in vehicles by a hierarchy of five levels to more easily define what kind of product it is.

**Level 0** - No automation at all, the driver is in full control of everything in the vehicle.

**Level 1** - Specific functions are automated, and even if more than one function is automated they operate independent from each other.

**Level 2** - At least two primary control functions operate in unison to relieve the driver of control. At this level the system and the driver share authority over the car.

**Level 3** - At this level the vehicle is autonomous in all normal conditions but may require the driver to control the vehicle if something unexpected occurs, such as abnormal weather conditions.

**Level 4** - The vehicle is designed to operate fully without help from the driver, even in special conditions such as a traffic accident or roadblock which creates heavy traffic.

Note that the highest level of automation is not always desired and the benefits that comes with it depends on the level of automation. For example, the safety can be improved by just adding an automatic breaking system which would fall under category Level 1. Although according to the National Highway Traffic Safety Administration level 3 or 4 automation may need to be applied to achieve noticeable environmental effects.
2.2 Adaptive Cruise Control

One type of AV is the Adaptive Cruise Control (ACC) which is the evolution of normal cruise control where you set a speed for the vehicle to keep. ACC makes often use of laser or radar to pinpoint the location of a vehicle in front. With this information it is possible to adjust speed accordingly so that your vehicle always is at a safe distance from the vehicle in front and therefore i.e. avoid collision from sudden breaks [6].

![Example state machine for an ACC system.](image)

Figure 1. Example state machine for an ACC system.

2.3 Limitations in AV

The main limitation of any computer generated response is that you can not pre-program every event, that being said a person would not know what to do in any given situation either. But with the human survival instincts most know how to reduce damage from, say a natural disaster such as an earthquake. On the other hand intelligent vehicles that can communicate also opens up possibilities for security personnel, such as ambulance to get to the point of the accident quickly by determining a good route with less traffic.
2.4 Anthropomorphism

"Anthropomorphism is a process of inductive inference whereby people attribute to nonhumans distinctively human characteristics, particularly the capacity for rational thought (agency) and conscious feeling" [7]. An example of this is if you were to put a name of the guide for a website. This can look like a happy icon or something saying "Welcome, I'm Jeff, how may I help you" rather than a little question mark that would indicate help. Someone named Jeff is something most can relate to as it is a normal name, and that feeling makes the object itself feel less alien.

Studies in anthropomorphism show that a machine with human features such as a name, a human voice or gender makes it more trustworthy. By adding a name, gender and voice to an autonomous car the drivers were less likely to blame the car for faults in an accident [5].

2.5 Safety

While the feeling of safety varies for each individual, Jeanne Priesler at fosteringperspectives [8] defines it as "Safe can be defined as free from harm or hurt. So feeling safe means you do not anticipate either harm or hurt, emotionally or physically.". Taking this definition into consideration, feeling safe in a car equals that you do not anticipate either you, the driver or the car itself to make any errors that would result in harm or hurt. The ability to predict what is about to happen next is key when not controlling the situation, e.g. a fully autonomous vehicle. This ties to trust, which McKnight & Chervany [9] defines such as "Trust is a multifaceted concept that can refer to belief that another will behave with benevolence, integrity, predictability or competence".

Trust and safety is also tied to repetition according to David G. Kitron [10] who states that it is only via repetition that risk taking may begin. This happens in steps, at first small risks are taken to avoid bigger risks, which in turn repeats itself, leading to bigger risks. Connecting this to Jeanne Prieslers definition of safety, the same situation is not expected to be different the next time it occurs, which makes someone not anticipate harm or hurt unless that was the case the previous time as well.

2.6 Technophobia

Technophobia is the fear of technology and while there is many definitions, one commonly used was coined Jay T [11] in 1981 and is as follows;

1. A resistance to talking about computers or even thinking about computers
2. Fear or anxiety towards computers
3. Hostile or aggressive thoughts about computers

According to Daniel Dinello [12] one of the reasons for technophobia being increasingly common is the theme in many movies, the post apocalyptic earth, ruled by artificial intelligence. The belief that once a computer is able to "fully
think for itself" it will manifest a human mind and destroy humanity in order to be the sole species.

Another reason behind the fear of technology is that people become comfortable using it and therefore not thinking themselves, an example of this is when you start using the calculator for simple additions [13] that you easily can do quickly without any help. This creates the fear that people, in time, not will be able to think for themselves as they are used to something else to do the, for example, calculations for them.

3 Method

The research was primarily conducted by analyzing and comparing existing models and by modifying certain elements of said model to pinpoint the function of each element. This was be done by interview along with low-fidelity prototypes in order to create a overview for the user. The interviews were conducted by showing the test subject pictures of interiors of current prototypes. During the interview, the interviewee were asked about specific elements, namely the steering wheel, dashboard and brake, how they feel with the current solution and how it can be modified for them to feel safer.

3.1 Target group

The chosen test group for this study was young males and females aged 23-30. They all needed to have a drivers license and at some point in the last two years been driving regularly. The main reason this specific group was chosen is that they are the ones that will most likely encounter and use Level 3 or 4 automation in the future if the development of said cars will continue.

The test group was be found by asking students at Umeå University in Sweden that fulfills the requirements. The results from the test group was not differentiated based on gender or academic backgrounds.

3.2 Lo-fi Prototypes

Seeing this study was conducted without the use of proper prototypes from companies that is currently developing autonomous cars, pictures that illustrates the view of the driver were used. These lo-fi prototypes were created by using AV’s currently in development and by altering those. The components that were studied in this article were steering wheel, brake and dashboard.

3.2.1 Steering wheel

One major component of contributing to the feeling of safety is the ability to take over the control of the car yourself, with other words, the ability to use manual control. And since the steering wheel is a cornerstone for controlling any vehicle, how would the removal of the steering wheel affect the drivers?
Figure 2 shows the interior of Nissans self driving car which they chose to construct in such a way that when you enable the AD technology, here called PD for Piloted Drive, the interior changes such that the steering wheel and dashboard changes into something more enjoyable like the weather or a social media feed and the foot brakes retract so that they are not used without intention.

When Piloted Drive is disabled the steering wheel appears, another dashboard is shown and the break lights up and ascend to their original position as shown in figure 3. Giving the driver the ability to use the vehicle to drive manually.
3.2.2 Dashboard

This study also included whether the dashboard adds a feeling of safety or not. The dashboard allows the driver to get information about the car, such as engine temperature and current speed. During the interview, the test subjects were asked what kind of information they would like to have on a dashboard to feel safe.

![Fig. 4. Interior of a Ford F-150 using ActivePilot. [16]](image)

Figure 4 shows Ford's solution of a dashboard from their car F-150 which uses ActivePilot, their autonomous technology. Unlike Nissan, they have chosen to retain tachometer and speed meter at all times as well as information about the car such as engine temperature, fuel and oil level.

3.2.3 Brake

The ability to manually brake the car, either with a pedal or via hand brake were also examined during the interviews. The test subjects were asked how they felt in different scenarios, such as removal of the brake. As previously mentioned, Nissan's solution to this when in Pilot Drive they remove the light from the pedals and retract them so that they are not a burden when stretching your legs as can be seen in Figure 2. With the removal of Pilot Drive though the pedals light up and descend to their original position so that you easily can access them as seen in Figure 3.

4 Results

The interview was conducted on seven people and the results have been organized depending on which element of the car it concerns.
4.1 Steering Wheel

A bit more than half (5/7) of the people interviewed wanted the steering wheel to remain in its place while autonomous drive (AD) was active and both of the other two needed the steering wheel unless there was some sort of button or way to abort AD quickly and take control over the vehicle again. A vast majority, 6 of 7, wanted the AD to completely be turned off when using the steering wheel rather than just adjusting the course of the vehicle and then, when letting go of the steering wheel, the car would automatically go back to AD. The steering wheel was also prominently (6/7) voted as the most important element for feeling safe when they had to choose one out of the three (steering wheel, brake and dashboard).

4.2 Brake

As the steering wheel, the same five people wanted to have a foot brake available at all times. Even when asked if a hand brake with features such as ABS (Anti-lock braking system), which may reduce braking distance by preventing wheel lock, they felt that a foot brake is safer because it feels reliable. The two people who did not want the steering wheel and brake wanted, along with three that did, an emergency stop button that would force the AD to shut down and also stop the car in an efficient manner. This also had to be mechanical rather than electronic so that it is usable without power. The placement of this button varied between the dashboard or in the middle of the car so that everyone could reach it. Four out of seven thought that it should be in the middle of the car which would enable the driver and/or the passenger to not pay attention to the road because another person in the same car could.

4.3 Dashboard

The only element at the dashboard where everyone agreed was the necessity of the speedometer. Although only four persons needed a fuel gauge to feel safe, everyone wanted the standard set of warning lights for errors within the car, such as engine temperature warning, or a warning that a light is broken. Six out of seven agreed that the tachometer were abundant, as well as the meter for engine temperature, where, as previously mentioned a warning is enough. All seven of the interviewees ranked dashboard as the least important of the three previously mentioned elements (steering wheel, brake and dashboard).

4.4 Improvements

Among the interviewees, a few improvements came up that they felt would make them feel more safe using AD; the idea that the steering wheel would retract a bit but still be close enough to grab. When taking the steering wheel and pulling it out to its original position the AD would stop and you would again have full
control over the vehicle. This would result in a combination of an emergency stop with always having the steering wheel at hand.

Another addition that six out of seven felt would help were anthropomorphism. By adding e.g. a voice to the car that welcomes you when you enter the car would help it feel more intelligent and advanced which in turn would make it more trustworthy.

5 Discussion

Based on the results, the feeling of safety and which elements a person will need to feel safe depends on whether they trust the technology or not. Based on this assumption, the results of the study may differ if the test was redone in a few years, when supposedly AV’s is a common thing. From my own experience, you take more risks and become less prone to anxiety or fear when the same action is repeated several times. For example, even though it is still dangerous to look away from the road when speaking to another person in the car, experienced drivers do that because they are used to nothing happening. It is of course, more times than not, less dangerous since an experienced driver often can act in a manner that minimizes damage. Using this theory, the only way to make people feel safe in autonomous vehicles is to slowly convert all cars to autonomous so that people always feels like it still is the same usual task, only with a new feature. This would also apply to people who harbor technophobia, although the curve would be slower, it would most likely take more time for them to feel safe in an AV. Perhaps they themselves need to test AV’s rather than seeing that it works when other people are using it.

Another common answer for everyone interviewed was that they felt they needed to be in control of the situation to feel safe, or at least feel that they can influence the outcome of the situation at any given time. This is probably the main reason why most people chose the steering wheel as the most important part for feeling safe during AD. This because the steering wheel also allows the driver to avert certain dangers that the brake can not, of course this goes the other way around as well, but it is more common that you need to evade a wild animal running across the road rather than that you need to brake for a wall or something you cannot drive around.

The results from studying the dashboard showed that the dashboard today is something that has been carefully developed over a long time span which means that the information currently shown is what people want to see. Although, the tachometer is most likely something that lives on from old times and is not really needed today, especially since quite a few cars uses automatic gear rather than manual.

The fact that almost everyone felt that anthropomorphism would help them trust the car is quite fascinating and it makes you consider what other areas it can be applied to to help them bloom. The question also rises regarding Daniel Dinellos [12] theory about technophobia being born from movies where machines get so advanced that they will eventually rule the world.
Solely considering the definition of safety previously mentioned by Jeanne Priesler; "Safe can be defined as free from harm or hurt. So feeling safe means you do not anticipate either harm or hurt, emotionally or physically.". With this definition we probably feel safer by having a steering wheel, brake or a dashboard as it allows us to anticipate that we are less likely to get hurt serious if we can control the situation to some extent. But the question to why anthropomorphism makes us feel safer may be harder to answer. It might be because the feeling that it is more advanced makes us believe that we don’t need to do as much to produce the best possible outcome in a dire situation.

Even though the results were similar enough from person to person to come to a conclusion, it would have been positive for the study to broaden the range of the test group by including e.g. older people who does not use technology every day the same way younger people do. Although I reckon the result would be much alike the current as almost everyone wanted both the steering wheel and the brake to remain, this would strengthen the conclusion. It would also be interesting to see if there was a certain group of people who acted like two of the interviewees right now, trusting the technology completely and therefore not wanting brake nor steering wheel when using AD. Is there any common factor between these people like education or is it perhaps based on environment?

6 Conclusions

A majority of people will feel safe in AV when they are common and used everyday. This is because they will know how the AV will react in every situation once it’s been used enough, which in turn creates a feeling of control since dangers can be anticipated and planned for accordingly. This in addition to the fact that most people wanted to keep the steering wheel and brake shows that the feeling of control over a situation is very important.

According to the interviews, anthropomorphism helped the user feel safe as it makes the object feel more advanced, which would lead to the user thinking it is better. It also helps the user relax and make it feel comfortable as a human voice is something that they are used to hearing.

The definition previously stated by Jeanne Priesler [8]; "Safe can be defined as free from harm or hurt. So feeling safe means you do not anticipate either harm or hurt, emotionally or physically." is a good definition that covers almost everything, but I feel like it could be modified to: "Safe can be defined as free from harm or hurt. Feeling safe means that you can control any given situation in such a degree that you can prevent or minimize both harm and hurt, emotionally and physically.".
References

Using augmented reality for learning a second language

Tove Barsk
Department of Computing Science
Umeå University, Sweden
tovebarsk@gmail.com

Abstract. When studying a new language there are several parameters that affects how fast the learning process are. Technology is used a lot for educational purposes, and augmented reality(taking an exciting picture and blending new information to it) is becoming more and more common. This article investigate how to best use AR to learn a new language, or more specific how to use it to improve someone's sentence structure and vocabulary. A questionnaire was formed to investigate how students prefer to learn and which methods they use to remember new information better. The questionnaire contained both likert statements and open questions and was sent to 22 people over email and Facebook. The results of the study showed areas where Augmented reality can be of big help. One area is to make the learning situation more lifelike, it can also help improve immediacy and motivation. Mobile AR applications and AR books are two kinds of aid that can make the language learning process more effective. AR can make the user more active and motivated and therefore help them remember words and sentence structure better.

1 Introduction

Learning a new language can be an important part of social integration. When studying a new language there are several parameters that affects how fast the learning process is. For example it is easier to learn a language that are similar to a language you already speak and it is easier when you are young [1]. One big help is if you have someone who knows the language you are trying to learn, who can help you, and who you can talk to. This is because you learn quicker by "doing the real thing, in this case speaking the new language [2]. But it is not always possible to do the real thing, and you might not always have someone to talk the new language with.

How fast you learn also depends on what kind of aid you have. Technology is used a lot in educational purposes, both in classroom situations and at home. Augmented reality, hereafter called AR, can bring a new dimensions to this. AR is a visualization technology that combines real-time three-dimensional computer-generated imagery and real-life footage to create an enhanced representation of reality [3]. In other words you can add a computer-generated layer to the reality by using different kinds of AR technology. Applications already exists that
helps you translate both words and sentences by instant visual translation. By pointing your camera on for example a street sign it can translate it to a chosen language. By adding a virtual layer to the sign, it looks like it is written in the new language. There also exists AR books, where the user can look at the book through some kind of AR device and see three dimensional models.

One important part when learning something new is to actively learn instead of just listening or watching. The use of AR in educational purposes is increasing, there are a lot of areas where AR can be used to improve the learning situation. For example it can be used to study the human body, in a three dimensional and lifelike environment. Augmented reality can help create situated learning even if it is not possible to do the real thing. As Mehmet and Yasins says "Displaying information by using virtual things that the user cannot directly detect with his own senses can enable a person to interact with the real world in ways never before possible" [4]. The focus in this report is to determine the effects it has on the learning process when learning a new language. Learning a new language is not just remembering words, but it also involves studying sentence structure.

1.1 Objective

The objective of this paper is to investigate the usage of augmented reality for learning a second language. Both the effect it has when learning new words, and the effect it has when trying to improve sentence structure will be studied. The following research questions have been formulated to investigate the problem:

• How can an augmented reality be used to improve vocabulary?
• How can augmented reality be used to improve sentence structure?

By making a survey about learning techniques and combining the results with previously done studies about learning a second language, these questions will be answered.

2 Theory

There has been a lot of advances in the use of technology in educational purposes. As Jayson [5] says technological innovation is changing the way that students learn and universities teach. Augmented reality is one technology that can bring a new dimension to this.

2.1 Augmented reality

Augmented Reality (AR) is a visualization technology that combines real-time three-dimensional computer-generated imagery and real-life footage to create an enhanced representation of reality [3]. Unlike Virtual Reality, where users are immersed in an environment that is entirely computer generated (i.e. the real
world is replaced by a virtual world), AR allows users to experience a modified version of the real world by superimposing virtual images onto real views of the world (the real world is “augmented,” or supplemented, by virtual elements) [6]. In other words, augmented reality takes an existing picture and blends new information into it. As Granmar [7] writes, AR was made as an protest against VR, they thought VR locked up the user, so they created AR to open the user to an improved reality. She also says that it is the simplicity to use AR for everyone that will make it big, and that AR had more practical use than VR.

2.2 Augmented reality and education

Augmented Reality is changing education. What started out as something that was simply “cool” has become a way to engage learners like never before [8]. It is not always possible to study something in its natural environment, it may be dangerous or expensive which is why education is one important domain for augmented reality. AR can provide situated study environments for students. AR has matured enough to perform more and more important functions in education field with its distinct superiority [9].

Some areas where AR can make a big difference in education is augmented biology and physics education. For example it can be used to study the human body, what the organs exists of and how they look [10]. It already exists educational augmented reality applications and educational augmented reality books. An augmented reality book (also called “the magic book”) is a book where you can look through some kind of AR-device and three dimensional models will appear [9]. The possibilities are many, the book can contain interactive elements in a text, hyperlinks, images or even interactive assignments. In other words there are many ways of improving the students learning experience by using augmented reality. As Wu et al [11] says AR technologies help learners engage in authentic exploration in the real world. They also says that ”AR could enable learning content in 3D perspectives, ubiquitous, collaborative and situated learning, learners’ senses of presence, immediacy, and immersion, visualizing the invisible, and bridging formal and informal learning. Kesim and Ozarslan says [4] it can even be applied to help with collaborative tasks since it is possible to develop innovative computer interfaces that merge virtual and real worlds to enhance face-to-face and remote collaboration. They also say that Augmented reality makes the impossible possible and its potential in education is just beginning. Augmented reality interfaces offer seamless interaction between the real and virtual worlds.

2.3 The learning process

There are many ways of learning something new. Different individuals feel they learn best in different ways. Something that helps is repetition. Even if your are really focused when learning something, the knowledge can be quickly forgotten if not repeated. After 24 hours you have forgotten about 50-80 percent of what
you learned [12]. Our brains are constantly recording information on a temporary basis: scraps of conversation heard on the sidewalk, what the person in front of you is wearing. Because the information isn’t necessary, and it doesn’t come up again, our brains dump it all off, along with what was learned in the lecture that you actually do want to hold on to [12].

Spaced repetition is a method of learning which is proven to embed new knowledge in the long-term memory [13]. The idea involves repeating the learning events with space in between instead of trying to learn it in one session. For example you could start by learning a couple of words in a new language one day, then repeat a second time in two days and so on. This is an effective method to make new knowledge stick [13].

Another effective method is active learning. Active learning is generally defined as any instructional method that engages students in the learning process [14]. So instead of just listening or seeing, the student is speaking or doing something active to learn faster. By using spaced repetition and active learning the student have a better chance of getting the information to the long-term memory.

**Learning strategies for second language** When learning a second language, there are seven broad aspects to consider, you have to work on reading, writing, vocabulary, grammar, sound acquisition and listening comprehension [15]. Different learning strategies work best on different aspects. Since this article focuses on sentence structure and vocabulary, only strategies to acquire these will be mentioned below. Neiman et al (as cited in [15]) mentions some strategies:

- **Grammar:**
  - Following rules given by text
  - Inferring grammar rules from text
  - Comparing a language you know to the second language
  - Memorizing structures and using them often
- **Sound acquisition:**
  - Repeating aloud after teacher, native speaker or tape
  - Listening carefully
  - Talking aloud, including role playing
- **Vocabulary**
  - Making up charts and memorizing them
  - Learning words in context
  - Learning words that are associated
  - Using new words in phrases
  - Using a dictionary
  - Carrying a notebook to note new items
**The cone of experience** Edgar Dale came up with a cone of learning (also called the cone of experience, see figure 1). The cone is a graphic figure that shows the hierarchy of learning through the programming of real experiences. It shows that to the best way of learning is at the bottom of the cone, where you "do the real thing". It also says that we remember 10 percent of what we read, 20 percent of what we hear, 30 percent of what we see, 50 percent when we combine seeing and hearing, 70 percent of what we say and write and at last 90 percent of what we do [2]. In other words, you want to structure your learning so that it falls under some of the lower sections in the cone. When learning a new language it is therefore better to both say and write it, than to just hear it. The cone also says that if the teacher show someone how to do something, they will probably remember, but if the teacher involve them in a meaningful way, they are more likely to understand it better [2]. Augmented reality can give more opportunities to train or study in lifelike environments.

![Dale's cone of experience](image)

Fig. 1. Dale’s cone of learning [2]

### 3 Method

A questionnaire was designed to assess the best way to use augmented reality applications when learning a new language. The questionnaire was a web-based form and was distributed over social media and email. The statements and questions asked about learning and remembering in general. The results of the questionnaire was then combined with previously done studies about learning a second language in particular, to determine the usage of augmented reality when learning new words and sentence structure.
The questionnaire consisted of 9 likert scale statements and 2 open questions. The choice to use open questions was made so that the person answering would have some questions where they could answer more freely. The choice to use likert scale questions too was made to minimize the risk that the people answering the survey did not answer at all which can be a risk with open-question-surveys.

On the likert scale the choice was made not to have a mid-point since studies [16] show that it can make a person answer more positively to please the ”interviewer”. As Ron Garland says ”Research provides some evidence that social desirability bias, arising from respondents’ desires to please the interviewer or appear helpful or not be seen to give what they perceive to be a socially unacceptable answer, can be minimized by eliminating the mid-point” [16]. By not having a mid-point the person answering the survey has to take a stand at every statement and cannot just choose to be neutral. The number of points on the scale then was decided to be 6 since research confirms that data from Likert scales becomes significantly less accurate when the number of points drops below five or above seven [17].

The statements was written in an “agree or disagree” way. The 6 answers were as follows:
- Agree Very Strongly
- Agree Strongly
- Agree
- Disagree
- Disagree Strongly
- Disagree Very Strongly

3.1 Test group

The survey was sent to 22 students in Umeå and Luleå. Gender was of no importance so it was sent to both female and male participants. The target group consisted of students. This decision was made because the questions asked a lot about learning and how people prefer to learn, and people who study learn new information on a daily basis.

3.2 Questionnaire

The questionnaire had 9 Likert scale statements which are shown below in figure 2. The statements was meant to determine how important different methods of learning were to people who are trying to learn and remember new information, for example words and sentence structure. For example if people preferred to speak and listen rather than write and listen.

The Questionnaire contained a few Yes or No questions that led to different
open questions, this was to check what the person knew before redirecting. For example, those who knew what augmented reality is (77.3 percent of the participants), a following question appeared about how they thought augmented reality could be used in educational purposes. Examples of the open questions are shown in figure 3.

4 Results

The first three likert statements was constructed to research how people preferred to learn new words and sentence structure, by listening and speaking, by listening and writing or by reading and writing. The statement that got the most positive results was listening and speaking. 86.3 percent answered that they agreed that felt that they learned best by listening and speaking(fig 4) On the statement ”listening and writing” the result was very scattered, and the statement that got the least agreement was reading and writing (see figure 5). So ordering them by preference:
1. Listening and speaking (49 percent)
2. Listening and writing (31 percent)
3. Reading and writing (20 percent)
I remember new information better if I discuss them with someone.

Manipulating objects helps me remember what someone says.

Helpful to have something visual supporting the information being presented (see figure 8).

4.1 Open questions

One of the open questions were if the participants use any method to remember new information better and if so, what kind of method. The most common answer was repeating information in their head, both in the moment and over a longer period of time. Another common answer was taking notes while listening to new information. Two of the participants answered that they use mind
maps and wrote questions to quiz themselves. One of the participants also wrote
that being able to connect the information to something helped them remember
better. An example written by one participant was “when learning a new noun,
seeing a picture of it makes it easier to remember.”

On the question asking how the participants thought augmented reality could be
used in educational purposes, most answers talked about visualizing things, for
element “visualize things that can be hard to grasp with text and voice only”. The participants talked about using AR as a complement to help students understand and also getting a more “hands on” learning situation. Some also wrote
that AR can give the student a perspective that is not possible without it. One
answer also mentioned that it could help motivate students and make them more
focused.

5 Discussion

Steve et al [18] points out that educators often utilize games to assist students
to easily grasp concepts. They also say that with AR technology, educational
games in the real world can be designed. These kind of games can add a 3d layer
to real world things. In other words, AR can be used in educational games to
help motivate the student. Combining this with the strategies mentioned above
to learn sound acquisition and grammar, educational games where you memorize
structures or participates in role play can help improve the students language
skills. This is supported by the study since all the participants remembered new
information better by discussing it with someone and if they had some kind of
visual aid, which a game can provide.

As mentioned earlier in this article, the cone of experience shows that the more
active we are, the bigger the chance is that we remember the new information
including words and grammar. Since AR can create 3D layers on the real world,
and can let us manipulate this layer, it can be of big help to help the user be-
come more active. When learning new words, AR can translate things in it’s
natural environment and that way help us make connections which will help
us remember. This was supported by the results in the questionnaire since the
participants felt that they remembered things better if they could make some
connections.

One application for using augmented reality in educational purposes in gen-
eral is to make the learning situation as lifelike as possible. When learning a
new language, being able to talk to someone, or discuss with someone, can be
the purpose for wanting to learn. The study showed that a majority of the par-
ticipants felt that they remembered new information better if they discussed
it with someone else. And Neiman et al (as cited in [15]) supports this when
talking about strategies to learn grammar and sound acquisition, for example
repeating aloud after teacher, native speaker or tape, listening carefully, talk-
ing aloud, including role playing. When studying a new language the user often has some kind of literature. By using AR books when studying, the user can see videos, translations and conversations by using some kind of AR technology. Supported by the study, most participants preferred to learn by speaking and listening instead of just reading, and this is one place AR can be of big use. Instead of just using books to write and read information, AR can be applied to let the user view videos and listen to people speaking the language. AR books will also increase the activity of the user.

Another thing AR can help with when learning new words is immediacy. Instead of having to look up a word when wondering what something is called, AR technology can add a layer with words in the real world. This also makes the learning more situated, when spotting an orange for example and wondering what it is called, the user can immediately see the orange and the name through some kind of AR technology. In other words, the user learns words in context and learns words that are associated, which was two of the strategies Neiman et al (as cited in [15] mentions for improving vocabulary. This can also be applied to translate words, for example signs and notes, AR can add the translation immediately on the sign or note. Applications for translating words directly on the AR-layer of the real world already exists, for example google translate application allows the user to point their camera to a street sign and get an instant translation on the screen. [19] One of the strategies for learning a second language is to compare the with a known language. By using AR to translate words and sentences directly on the screen allows the user to directly compare and make connections to the first language.

Another strategy that Neiman et al (as cited in [15] talks about is carrying a notebook to note new words or sentences. By having a instant translation AR application and being able to save the words or the pictures from the camera, this strategy will be easier to carry out.

6 Conclusions

Both the questionnaire and previously done studies on learning a second language shows that Augmented reality can indeed make a difference when learning a second language. One area where it can be of big help is to increase the activation level, so that instead of just reading and writing new words and sentences, AR can make the user learn while integrating with the real world. It can also make the learning situation more lifelike and more situated by using AR technology to help with translations, words and even videos. The user will have easier to make connections and remember words and sentence structure better. By using the AR layer to show translations, sentences and maybe even videos it can also increase immediacy. Supported by the study, visualization is an important part of remembering new information better, and by using AR technology you get more possibilities to do this.
Another area where AR can be applied to help learn a second language is motivation. Augmented reality can help make the learning situation more stimulating and fun. Just as gamification can be used on regular applications to increase motivation, it can be combined with augmented reality to make it even more interactive and stimulating [20]. Augmented books is another example where the user, instead of just reading, can use AR technology to interact with the book. These books can contain elements which (when using AR technology) for example can show videos of people talking.

Here follows some suggestions on future work. By using some kind of prototype, tests can be performed to determine how big the affects are when using augmented reality instead of the regular learning methods. The test could be extended to people who study a second language instead of students since they probably have more specific methods to remember new words and sentence structure and not just new information in general.
References


Biohacking: Evaluating the benefits of Near Field Communication in the human body.

Carl Holmertz
Department of Applied Physics and Electronics
Umeå University, Sweden
carlholmertz@gmail.com

Abstract. Objective: Near Field Communication (NFC) is a technology that is growing together with the development of technologies such as smartphones. A new way to use NFC today in 2016 is to implant a NFC chip in the human body, often the hand, in order to provide services, functions and possibilities to the user and to ease tasks in the daily life. The aim of this study was to determine what different functions, with NFC implants, that would make people want to implant NFC chips into their bodies.

Method: The selected participants of this study were given a survey to fill out which contained examples of both existing functions and functions that could possibly exist in the future. The participants were then asked to rate these different functions and add additional functions that would make them willing to consider implanting a NFC chip in their own body.

Results: 35 persons participated in the study and the results show that the majority of the participants preferred the functions that made their daily life easier such as opening and accessing doors or paying in stores which created a life were they did not need to carry around keys or a wallet etc. A lot of the answers and comments showed that some participants were worried about the security with the NFC ship.

Conclusion: The companies should start with improving the security with NFC to prevent the information on the NFC chips from being hacked and stolen. When security no longer is a problem they should start focusing on functions that eases the daily life of the user like payments, instead of high-tech functions like a smart fridge etc.

1 Introduction

Imagine if we could open all doors at work, at home or even our car with just our hand. Imagine if we could pay at the store, or if the doctors could see our medical records in case of need, by just reading a chip in the hand.

NFC stands for Near Field Communication and is a wireless technology that uses Radio Frequency Identification (RFID) technology and short range radio waves in order to provide communication between devices. This communication requires a distance shorter than 4-10 cm between the devices in order for it to work [1]. The possibility to use NFC technology with payments in the grocery store already exists today in 2016. We can pay by just touching the cashier
terminal with a NFC equipped credit card. NFC is also popular for identification systems at universities and at companies where the students or employees have cards that work as keys to open doors in order to gain access to different rooms.

Biohacking is a term that can be described as a practice of biological experimentation for a self-defined purpose using a variety of Do It Yourself (DIY) devices and techniques in a nontraditional way. The term was first used online in 2008 and describes a broad set of activities such as DNA manipulation and implanting of computer chips in the human body [2].

In 2016 there are at least 200 people in Sweden that use NFC in biohacking with implants of NFC chips in their hands (U. Swanström, Netlight, personal communication, April 28, 2016). These people use this new technology to do different things such as opening doors at work or at the gym, or storing a digital portfolio or a business card. One can wonder if this is enough, or does NFC allow us to create new ideas and usage areas.

1.1 Objective

The objective of this study was to identify the different usage areas that makes an individual willing to implant an NFC chip into the body.

People have already started with NFC implants in their bodies for different reasons, but what applications should the companies focus on in order to get this technology to be used by the general public? In this study the objective is to answer the following question:

- What different functions will make people implant an NFC chip in their body?

The results of this study were used to create a guideline for markets and companies in order to know where their focus should lie in the future development of NFC. After the objective was finished, the results were presented in a way that helped specify what usage areas the market and companies should focus on, in order to get users.

2 Theory

In order to understand the possibilities with NFC, it is important to understand the technology behind it. This section includes information about the technology and how it is used today and possibly tomorrow.

2.1 The technology

Near field communication (NFC) is one of the most promising and emerging technologies [3]. It is a wireless technology built on Radio Frequency Identification (RFID) [3]. NFC works through short range radio waves in order to communicate with another NFC device. Most of the new smart phones have NFC implemented. Since NFC tags can store data, the device (e.g. phone) scans
the data and then acts upon it. Hence NFC is built on RFID systems and creates a system that allows a two-way communication between devices instead of a one-way communication as earlier systems had [3].

The purpose of NFC is to make it easier and more convenient to exchange digital information, make transactions, and connect digital devices with a touch. Thanks to the NFC’s ability to both read and write information to devices, the future usage of NFC is believed to be bigger than e.g. standard smart cards [4].

NFC protocol exists of two statuses; active and passive mode. Tags and cards are always passive and therefore they do not need any power in order to work. In a passive mode only one side of the communication needs power in order to read and write information to the tag or card. Active mode is when both the initiating device and the target device generates their own radio fields in order to communicate e.g. NFC connection between two smart phones with their own power supplies [4].

2.2 The use of NFC today

The most common use of NFC today occurs in the passive mode, often presented in key card systems or in credit card payments. If we look at the past, shopping has been associated with cash or credit card payments, but later different possibilities to pay with a smart phone have been developed. Today we have the possibility to use our phone to pay for goods or to pay our bills [3]. The development of mobile payments has grown fast and over the last ten years we have been introduced to various different payment methods, such as Short Messaging Service (SMS). Even though these different methods provide good service, they are still not good enough or ideal if we view it from the traditional payment context [6]. A fresh way of paying is with mobile credit cards which is a contactless credit card payment using only a mobile phone with Near Field Communication technology (NFC). Through this technology we do not need a physical connection between the payment terminal reader and the consumer since the transaction can take place with just simple wave (a couple of centimeters from the terminal) or a simple touch [5].

Another usage area of NFC today is for Mobile Coupons. Instead of cutting out a piece of a newspaper, the user can now redeem their coupon through their NFC enabled smartphone [7]. The user does not need to carry a rewards card for every store, he or she can now load the coupons on their phone and the reward points go straight to the account without ever digging in the wallet for a card [3]. Functions like these show that NFC has the ability to change the market in a way that the coupons and commercials can be targeted towards a specific user depending on the users personal preferences, instead of a more general targeting such as in the newspapers [7].

Unlocking doors or computers is another way to use NFC technology. This function makes companies, universities, hotels and homes etc. more secure and more available for the user. This can be done through both a smartphone or NFC keycards. The technology allows users to open doors or even their computers by just waving their phone or card towards a reader [7].
Transportation is a necessity whether you live in Sweden or any other country. With NFC we have the possibility to pay for our rides just by waving our phone or card on the bus, at the train gates or at the platforms. Today, this is more often used with keycards, with NFC tags in them, instead of a smartphone [7].

Mobile gaming is an industry with tens of millions of users worldwide playing games daily [8]. In 2011 a popular mobile game called Angry Birds implemented an NFC function where two devices could connect to each other by tapping the devices together while having the game running. When this was done, new levels were unlocked on the phones. Even though this game is not fully functioning on NFC, it gives the game developers an idea of what is possible to do with NFC in gaming. A Smart Home is created by different solutions for carrying out basic functions like opening doors, turning off alarms, or recognizing objects through sensors [9]. When all of these functions are combined the result becomes useful and exciting, this has created the idea and practice of Smart Homes. One way to achieve this is to strategically place NFC tags on different places around the house. When scanning these tags with your phone they will automatically activate a process, such as changing the temperature in the house, turning off and on lights etc. The user can choose what different functions the tags should activate to create the wanted Smart Home [10].

Smart Homes can include a varied amount and types of functions. For example, a smart refrigerator with NFC technology could detect if any compartments are open, tell the temperature and energy consumption etc. Different sensors collect this information and can then be sent to the user via a smartphone with NFC just by waving the phone in front of the fridge [11].

3 The use of NFC tomorrow

The future of NFC has the potential to be a pillar in our daily lives. Thanks to the possibility for NFC to connect with such a wide variety of devices, it creates a possibility for NFC usage to grow in an unlimited way [7]. Some functions are not really "future functions" anymore, because different companies already have them in development but they are not yet out in the open market.

Imagine if the fridge could, through NFC technology, keep track of what groceries the user has in the fridge. The fridge could then tell the user the amount of food in the fridge, and if it has expired or not. If NFC works together with Wi-Fi in this task, the fridge could then show shopping lists depending on the user’s preferences for food and drinks [7].

NFC technology can be used in healthcare for identification of patients in hospitals, which will be very useful since patients have different illnesses. If a doctor would confuse the illness of a patient with the wrong patient a fatal error could occur. With NFC we can easily protect patients from ever getting mixed up. Even if a person will faint or be unconscious when arriving to the hospital, the doctor could access the patients’ medical record to avoid further problems such as giving penicillin to a patient that is allergic to it, which could be fatal.
4 Safety

The era of the Internet and technology such as smartphones has taught us about security, and about the risks of being hacked and getting our information stolen. This is an aspect to process and discuss. There are different aspects to the safety issue and a big question is if the pros outweighs the cons in technology development.

4.1 Eavesdropping

Because of NFC being a wireless communication technology it creates a security problem with eavesdropping. NFC uses radio frequency waves in order to communicate and this allows an attacker to use an antenna to receive the same signals [12]. There are different factors that affects the possibility to eavesdrop and at what distance it is possible [12]. The only way to prevent these kind of attacks is within the application itself and not in the technology NFC. A possibility is to create a secure channel for NFC, which means that two NFC devices agree to a standard key protocol in order to establish a shared secret between the two devices. This secret is then used in order to provide confidentiality, integrity and authenticity of the transmitted data [12].

5 Method

First the identification of the future usage areas with NFC implants were evaluated and identified through research about NFC today and the future thoughts about the usage of NFC in general. The results of this constructed the foundation for developing the questions for the objective.

With the information gathered from the identification of the future usage areas the objective was researched. The different usage areas, that will make an individual willing to implant an NFC chip into the body instead of just using a chip in the wallet or such, were identified through a survey. Before the the real study took place, two pilot studies were made in order to evaluate and develop the form itself. In the first section of the finished survey, the users were asked to answer some basic questions about their age, gender and occupation.

In the second section the participants were asked to rate the usefulness of examples of NFC functions that are available today. The participants were asked to rate the usefulness of the functions on a one-to-five scale where number one represented "Not Useful" and number five represented "Very Useful". At the end of the second section, they were asked if there were any other NFC functions (that were not in the list) that they thought were useful.

The third section was very much like the first. The participants were asked to rate the usefulness of some future functions. At the end of this section they were asked to think freely and use their imagination to think of future usage areas with NFC that they thought could be useful.
In the fourth section in the survey for this study the participants were given a list of functions with NFC, some existing today in 2016 and some that can exist in the future. The participants were asked to mark which of these alternatives that together could create enough possibilities in order for the user to consider implanting an NFC chip in their hand. They were then asked if they would consider doing an NFC implant if their marked functions existed. At the end, they were given the possibility to add some functions that they would like to have in the future.

The fifth and last section of the survey contained some basic questions about if they knew what NFC was before this study, if they ever used NFC in some way, and how they used it.

5.1 Choice of participants and language

The participants for the study were choosen based upon their field of work or study. The goal was to have participants that already had some knowledge about what NFC is, how it works, and the risks and possibilities with the technology. Therefore the study first and foremost included students of the Interaction and Design program at Umeå University in Sweden. These participants study engineering with focus on interaction design with programming, user experience (UX) and user interface (UI), and the majority of these students are well up-to-date about technology development. The study was created in Swedish because the participants were all from Sweden. These choices were made in order to ease the study and to get more creative responses, compared to doing a survey with people not studying or working in this field or doing a study in another language than Swedish.

6 Results

There were 35 participants in this study and the average age was 25 years of age. There were 64.3 % men and 35.7 % women. As planned in the method there where 78.6 % students and 21.4 % working people. The participants were asked to rate existing functions of how important they thought they were and the result is illustrated below.
Fig. 1: Payments: As shown above, the participants rated payments as very useful. Only one test participant rated payments as "not useful", but 60 percent (21 participants) rated payments as "very useful".

Fig. 2: Coupons: The view on the possibility to get rewards and points to get discounts is quite widely spread as shown above.

Fig. 3: Access: The test participants rated Access as a useful function.
The third section of the survey included functions that are not available today for the users but could be in the future. The participants were asked to rate these different functions with respect to usefulness.

Fig. 4: Mobile Gaming: The usefulness in connecting phones in order to play together and to earn extra points.

Fig. 5: Smart Home - Example: book the laundry room with a NFC tag.

Fig. 6: Smart Fridge that keeps information about what groceries that are in it.
Fig. 7: Healthcare: could be used as identification of patients at hospitals and the doctor could get quick access to the patients medical records if needed.

Fig. 8: Verification and identification: for safety (Ex: creditcard + code + NFC chip)

The participants were in the fourth part of the survey given a list of functions. Some already existing today, and some that could exist in the future. They were asked to mark which of the alternatives that together creates enough possibilities for the user in order to consider implanting a NFC chip into their hand. The result of this can be seen in figure 9.

After the participants marked their wanted functions they were asked if they would consider doing an NFC implant in their body given that their wanted functions existed.
Fig. 9: The alternatives that together create enough possibilities for the user in order to consider implanting a NFC chip into their hand.

Fig. 10: The results of the test participants’ answers to the question if they would consider implanting a NFC chip into their bodies given that their wanted functions existed.

At the end of this section, in the survey, the participants were asked if there were any other usage areas or functions that they would like to have, in order to
consider implanting a NFC chip into their hand. They were asked to think freely and to use their imagination. Examples of answers, translated from Swedish:

- "The functions that can make me leave my wallet and keys at home."
- "100 % guaranteed security, so that there is no possibility to hack information with an NFC reader."
- "Opening doors, paying etc"
- "Identification, for example when getting medical help, flying or such."
- "If I know that the chip can be implanted in my body in a safe way, and that I can be sure that these chips are not gonna be outdated in a near future."
- "Possibly for identification, in order to avoid carrying around driving licenses or other ID cards."
- "Unlock my phone and maybe measure my body temperature."
- "There are a lot of existing and future functions that are cool which would make a lot of daily tasks easier, but I care too much about my personal integrity to think this implant would be worth it. Maybe sometime in the future when we know more about how it works."
- "... to be honest I would never consider implanting a NFC chip in my hand today. As long as there are organization like NSA and CSEC. As long as there are no extreme laws to prevent these organizations to do surveillance on people I would never feel safe with this kind of implant...."

The last and finishing section of this survey included some questions about the participants' knowledge and usage of NFC today. There were 89.3 % that knew what NFC was before they took part in the study, 7.1 % did not know about NFC in advance and 3.6 % answered "maybe". The participants in the survey were asked if they had ever used NFC before and if they had they were asked to give an example. 35.7 % had never used NFC before, 7.1 % did not know if they had and 57.1 % had used NFC. Some examples of how the participants had used NFC before:

- "Connect phone to speaker system, booking the laundry room, open doors at work, SL-card (Train-card)"
- "Payed in the grocery store."
- "Sent pictures and contacts between phones."
- "To pare two phones for file transfer and to pare a phone and a speaker."

7 Discussion

From the results, it is quite clear that the majority of the test participants are willing (52.4%) or "maybe" willing (38.1%) to consider an NFC implant in their bodies. Only 9.5% answered that they would not consider implanting an NFC chip.

About the different examples of functions with NFC that were given in the survey it was three that stood out as less useful than the rest. Those three were coupons, mobile gaming and the smart fridge example. Payments, access,
health care and verification and identification were rated as very useful. This is confirmed by the comments and answers the participants wrote in text. The examples show that the majority of the answers contained ordinary functions such as opening doors, access to the phone, identification etc; functions that could make the daily life easier for the participants.

The author of this paper did an interview with Unn Swanström that has an NFC chip implanted in her hand. The interview gave information about how NFC was used today but also ideas of how it could be used in the future. One major opinion from Unn that corresponds to a lot of the answers, was the opinion about security. She was scared to store sensitive information on her NFC chip today with respect to the security and safety problems. In the comments from the test participants we can see answers such as ".. I care to much about my integrity to think this implant would be worth it".

To further examine this topic, research could be done on other ways to use NFC, like in a bracelet or such, in order to reach the same usage areas and usability. It could also be interesting to look at the possibility to use fingerprints instead, but that would make this study too broad and it would create a lot of other questions such as safety issues etc. NFC could be used in various ways, but this study was focused on NFC as an implant in the hand because of its media exposure lately and the increasing number of people that are doing the implant.

(U. Swanström, Netlight, personal communication, April 28, 2016)

8 Conclusion

Before the companies start to develop and market new functions, they should start to improve the security with NFC. This study shows that a lot of people are interested in NFC and its possible functions, but they are concerned about the safety of its use.

If there were no problem with the security and safety, the companies should focus on functions that could ease the daily life of the user. Functions like these could be the simple ones like accessing doors at work, home or even to the car. It could also be functions that could access the users’ medical records if needed or identification like a passport or a driving license. A way that NFC is used today and that people still need, are easy ways to pay in stores.
References

Ditching The Hamburger Menu

- An analysis of current menu design

Anton Forsberg

Department of Applied Physics and Electronics
Umeå University, Sweden
id12afg@cs.umu.se

Abstract. Menu design has become somewhat standardized and the use of what is called a hamburger menu is seen in apps and websites frequently, but some are calling for change. The menu layout has come under scrutiny in recent years for being inefficient and for losing user engagement. This study investigated the validity of these claims and tested the menu layout against another newly proposed layout, bottom navigation, to see how the two fared with experienced users. It also tried to determine which of the options the users preferred and to identify possible issues with the two layouts. The results show that while there is no immediate need to redesign the hamburger menu it is indeed very replaceable as long enough attention is paid to the choice of icons.

1 Introduction

The hamburger menu has gone from famous innovation to infamous user experience (UX) nightmare over the course of its lifetime. Most smartphone users see this many times over the course of a day, but few know it by name. In short, the hamburger menu (Figure 1) is an icon symbolized by three horizontal stripes found in the top left corner of most mobile web pages and apps. It is there to show us that there is off screen content available on click. The icon’s resemblance to a bun-meat-bun constructed hamburger has led to its playful yet deceiving name.

Fig. 1. The Hamburger Menu
The recognizable icon was first introduced in the Xerox Star operating system in 1981 and was intended to be a simple way to design menus with limited use of screen real estate. The design stuck and has become the industry standard when designing app and website layouts for mobile platforms. In recent years the design has, however, received much criticism for its low discoverability and efficiency. It is supposedly responsible for lost user engagement and making simple tasks difficult for new users. This study intends to find out what impact this menu design has on everyday users and if it can be done better.

1.1 Objective

The primary focus of this study lies in determining whether or not the hamburger menu layout has any negative affects on usability when compared to other layout designs. Furthermore, if this turns out to be the case, this study aims to provide suggestions for replacing this layout. More specifically this study will attempt to answer the following questions:

– Do experienced smartphone users experience issues with the hamburger menu compared to other common alternatives?
– When provided with another option, will users still prefer the hamburger menu?

2 Theory

The following subsections explain some of the terminology used as well as motivating the need for the study.

2.1 Mobile menu design

There are several different approaches to menu design when it comes to portable devices and smaller screens. In the early days of smartphones the most common layout in apps would be a springboard or launchpad (Figure 2) design. This type of layout acts as a navigation page where users can navigate to different sub-pages. Later the list menu was introduced, which is similar to the springboard layout but elements are displayed in a list layout with some hierarchical order.
While still in use widely on an operating system level, the layout has since been abandoned by most apps after Aza Raskin designed the Side Drawer. To access this menu he placed an icon with three horizontal stripes, the design we now know as the hamburger menu [4,5]. This design is now the most common in mobile apps and is also very present in mobile web pages. The design choice has lately been criticised however and several major companies have recently started to distance themselves from the icon and are figuring out different ways to navigate within an application [3,6].

### 2.2 The issues with the hamburger menu

The main issue with the hamburger menu lies in its low discoverability. The menu in itself offers no clue to what lies within it. For all the user knows it could be contain a list of dog species, friends birthdays or simply be empty. It is silent and secretive. An interface should be focused on
recognition rather than recall in the sense that users should immediately be aware of what action a button will perform rather than have to have this memorized from previous attempts. In this regard the hamburger menu offers only a simple clue, the three horizontal stripes that somewhat resemble a list. The hamburger menu has its contents hidden off screen in what is called and "off canvas" approach to menu design. According to Krug websites should be designed with a focus on scanning rather than reading. The idea is based on that users rarely read through entire pages but rather scan the page looking for a specific item. In this regard the hamburger menu, hiding its content, does not fare well. Furthermore, the minimalist implementation of the icon does generally not imply being clickable. It does in no way resemble a "real world button". This, however, is outside the scope of this article and is more due to the advancement of minimalist design than to the use of this icon.

2.3 Usability

In 1995 Jakob Nielsen wrote down ten design heuristics which are nowadays considered paramount for good usability design. This study builds a major part of its motivation on the premise of these heuristics. However, some of the terminology used in this study may need further explanation.

*Discoverability* is a term that aims to describe an item's capacity to be found or discovered. A person can for instance research their own discoverability by performing an internet search for their own name and seeing what shows up. More specifically it refers to how likely a user is to find something in an app or in our case, a web page.

3 Method

To determine if the hamburger menu is replaceable by more efficient menu layouts and how it affects users compared to those layouts a test was performed. The test subjects were let to use a web page using either the hamburger menu or another type of navigation, namely *bottom navigation*. The users were then given a set of tasks to perform within the environment and were encouraged to share their thoughts during the process. After the test was complete the users were shown the web page with the other menu layout and were asked questions about which one they would prefer. This study also aimed to determine whether or not the hamburger menu was the preferred layout within a system if the user could choose another form of navigation.
3.1 The Environment

To get viable data a mock web page was designed and developed with two different menu layouts baked into it. One layout was the hamburger menu which was fixated in the top left corner and the other one was a bottom navigation menu with only icons on it (Figure 3). As the bottom navigation is now a proposed way of designing navigation within apps by Google [6] with limited menu items, this was chosen as the alternative layout. The hamburger menu had the items ordered in it with some hierarchy but the bottom navigation was, as per the design principles, without hierarchy or descriptive text alongside the icons.

![Figure 3](image)

**Fig. 3.** The application used to test the Users. Here it is showing both types of navigation, but the users were only subjected to either one of them. The hamburger menu (top) and bottom navigation (bottom).

The web page was created using the Angular framework [10] and the scaffolding tool Yeoman [11] was used to provide quick and easy set up and testing. The menus used contained five different items: profile, news, friends, events and settings.
3.2 The Testing

The ten test subjects were all chosen from Umeå University and they all had more than two years of experience using smartphones and apps. Furthermore they also considered themselves very comfortable using smartphone apps. The tests were carried out in the students’ native tongue Swedish.

Each user was presented with a web page using either of the two menu layouts, to mimic an A/B testing scenario [12]. To test what issues arose when the subjects would navigate within the environment they were given different assignments of varying difficulty. For each assignment the accompanying researcher would note each action the subjects took and they were also encouraged to explain what they were thinking as they were performing the tasks. This provided qualitative data for the study and was gathered to show possible flaws with either of the two layouts.

After the test was finished the test subjects were asked a few follow-up questions to determine if they had any issues with the navigation within the app. The users were also given the opportunity to familiarize themselves with the other menu layout and decide which one they were most comfortable with using.

The questions at the end of the test were as follows, (translated from Swedish):

- What was your experience navigating through the app menu?
- Did you encounter any difficulties while navigating within the application?
- Which of the two menu options do you feel is the easiest to use?
- Which of the two menu options is the most appealing to you?

4 Results

The users who were presented with the hamburger menu had no problems at all completing their assignments. All subjects completed their assignments without performing any mistakes. This was however not the case with the subjects who had the bottom navigation layout. 100% of these subjects navigated to the wrong item while performing a certain task.

The users had varying feedback on the different layouts but most commonly the users with the bottom navigation layout mentioned the icons being misleading or hard to interpret.

However, after being allowed to familiarize themselves with both layouts, all users considered the bottom navigation being the simplest layout
and a majority of users, (70%), also considered the bottom navigation the most appealing.

5 Discussion

The fact that the subject with the hamburger menu had no issues with navigating the application might seem overwhelming at first, but it needs to be put in perspective. Since all users were experienced smartphone users they were not expected to have any trouble. This layout is after all one of the most occurring menu layouts seen today and users should be somewhat familiar with it. The fact that the other subjects had trouble with the bottom navigation shows us something important however. Since this layout did not have the descriptive text in the menu items, the icons were supposed to give the users enough information alone. This is where the layout failed and it shows us the importance of having good icons with a strong symbolism to what action they perform when clicked. The users also mentioned this themselves, which only further emphasises the importance of a designers choice of icons. Another solution for the problem could be to have the text showing underneath the icons. This may provide some difficulties however, due to limited amount of space.

The fact that most users do consider the bottom navigation to be the easiest to use of the two layouts illustrates how the issues they had with it are of little importance after repeated use. The fact that a majority of users considered the same layout to be the most appealing could either be due to the ease of use it provides or it is simply the novelty of the design.

6 Conclusion

The study can determine that experienced users do not have any issues using the hamburger menu, meaning that there is no immediate need to replace it. The results however indicate that the hamburger menu can be successfully replaced in many applications by implementing a bottom navigation. Not only was this way of navigating considered easier to use by all subjects, it was also considered as more appealing by a majority of the respondents. However, usage of this layout requires a limited amount of navigation options as it is not recommended to use more than 5\cite{6}.

It can also be concluded that first time users are heavily reliant on the usage of icons and how well they represent their performed actions. This shows that if a designer opts to use icons without explanatory text
within a menu design then the user experience of the navigation is heavily reliant on the chosen icons. With this information the bottom navigation can be used as an on-screen alternative to the hamburger menu in certain applications, as long as icons and explanatory text are considered.

References

Interaction limitations on mobile devices related to screen size

David Bergvik
Department of Applied Physics and Electronics
Umeå University, Sweden
dabe0052@cs.umu.se

Abstract. With the current trend of mobile devices increasing in size, this paper tries to identify possible issues related to screen size when performing one-handed interactions on mobile devices. Furthermore it evaluates if there are constraints that arise when users’ reach become insufficient for the desired interaction. The interactions tested were based on placements recommended by Android’s design principles, with focus on one-handed interactions. A testing environment was created and installed on Android devices with different screen sizes, and the effective time between interactions could be measured during sessions of qualitative testing followed by a post test interview. The results showed that the effective time when interacting with a larger mobile device is significantly longer. Furthermore it indicates that as screens become larger, the physical reach of the user becomes more relevant. This paper can serve as a baseline for further investigations on the effects of screen size within the spectrum of one-handed interactions.

Keywords: Screen-size, one-handed interaction, mobile devices, interaction limitations

1 Introduction

Personal digital devices are as of this point a big part of everyday life, with more than 4 billion mobile subscriptions in 2014 [1]. Mobile phones has emerged as a major communication and interaction resource between each other and our environment. With a steep development curve, mobile phones are getting bigger and faster by the minute. But with quick changes and new trends within mobile development, principles that used to be a key part tends to be forgotten.

In 2007, Steve Jobs and Apple introduced the first generation IPhone [2]. With a screen-size of 3.5 inches, the company claimed that it is the optimal screen-size for a handheld smartphone. Jobs described how a larger screen will have physical size that is too big for a human hand. Regardless, the rest of the phone industry started to produce bigger phones each year. Apple themselves also had to keep up with this new trend, and left their 3.5 inch screen size. The question that
remains is: Was Steve Jobs right back in 2007? Does the demand for bigger and faster, make the industry forget about the interaction abilities necessary for a handheld device. Is it all up to the manufacturer, or is it the software that need to adapt to these changes?

2 Objective

The objective of this study is to identify the interactive issues caused by the physical size of the device. The main area for this will be one-handed interaction sequences with smartphones, where the testing interfaces are guided by industry standard principles [3]. It will also try to evaluate how the size and screen size of the device will affect the ability for the user to interact correctly. This study will also try estimate possible differences in effective time when interacting with bigger devices, and establish when the devices prevents certain interactions. The questions that will be used to evaluate this are:

- Are there interactions that gets affected by the size of the mobile interface?
- How is the effective time of common intercations affected by the size of the interfaces?
- How does hand-size (reach) of the user effect the one-handed interaction?

To answer these questions a testing software will be developed. Using the software, qualitative testing will be done with subjects with experience of mobile devices. After analyzing the results of the tests, possible solutions to these problems will be presented.

3 Theory

In 2015, more than 84 percent of all mobile devices run some version of the Android operating system [4]. Even though that also includes tablets and watches, it is enough of a majority for basing this study on. It also comes with an open sourcebase for creating free applications. With this operating system (OS) comes design principles [3] necessary for creating an Android application as well as hardware buttons (see Figure 1). These buttons will be harder to include in testing software seeing as they also interact outside of the created application.

Fig. 1: Navigation bar of a nexus 4, made by LG and Google.
3.1 Core design concepts

When mobile developers create a new application, they have to obey some of the rules and concepts that have been constructed by the OS owners [5] [6]. The owners can reject an application that does not fully follow these concepts, in order to achieve a higher standard and convergence in terms of design within their application-pool. This study will be conducted on Android devices and will therefore focus on the design principles of that OS.

**Android Screen sizes** In today’s market, a vast variety of different mobile devices exists with different screen-sizes [7]. In order to help mobile developers with creating applications that work with all these devices, Android recommends designing for four different versions in terms of user interface (UI) [8]:

- Small: Screens below 3” (Wearables and Phones)
- Normal: Screens between 3” to 4.5” (Phones)
- Large: Screens between 4.5” to 7” (Phones and tablets)
- Extra large: Screens larger than 7” (tablet)

**Android Key Elements** Android applications have some navigation elements and components that classify as OS standards [9], and as such their position on screen will be key to find interaction combinations that are executed in a majority of applications. Back navigation is used to navigate back to a previous view, and is to be located in the top left corner of the screen (see Figure 2a). In the main view, a menu button is generally used to display a side menu, also displayed in the top left corner of the display (see Figure 2b). The menu-button will be replaced by the back-button when the user navigates to deeper parts of the application.

In the top right corner, a lot of different buttons are displayed. Usually a search-button (see Figure 2a), a button to display a right-side menu or drawer or a navigation to go to a view deeper in the application. As an overlaying element in the lower right corner, Google presents a floating "add"- or new-button (see Figure 2b) [10]. This element is considered in this study for its position always on top of the rest of content in that view, usually located in the lower right corner.

3.2 Default hand placements

How a mobile phone is held does at first glance seem trivial, but depending on how the hand is positioned, the user’s reach will be limited by the length of the finger and the hand placement used when interacting [11]. This forces the user to change the phone’s placement in order to perform certain interactions. The position of hardware buttons on the phone is primarily the source of the phone’s hand-placement, seeing as some of the hardware-buttons usually have to be pressed in order to unlock and start interacting with the phone. In this study the user is allowed to use any starting hand placement, as the tests concerns interactions only within the digital interface.
3.3 One hand usage

In a field study done by Hoober [12], observing 1333 people in public locations, 49% of the people that was using a mobile device was only using it with one hand.
This correlates with what Karlson et al. [13] shows in their study, that using a single hand to interact with a mobile device is common practice for users that are traveling while interacting with the device. Furthermore, they explain how users want to be able to interact with one hand only, but that some mobile devices, especially touchscreen interfaces, are not designed for this type of interaction. Karlson's study did not focus solely on touchscreen devices, but determined that the hardest interaction to execute for a right-handed user is from NorthWest <-> SouthEast (see Figure 2).

Fig. 2: Visualization of the hardest interaction on a mobile held with users right hand

3.4 User benefits of different screen-sizes

In 2011, a study conducted in South Korea researched the effects that screen size has on perceived mobility and enjoyment [14]. The authors Ki Joon Kim, S. Shyam Sundar, Eunil Park explain how they with a video-based test evaluate these parameters based on three different screen sizes (3.5", 5.7", 9.7"). Their results conclude that a smaller screen gives the user a greater perceived mobility while a bigger screen offers the user more enjoyment.

4 Method

By examining the typical elements in an Android application interface, a quantitative user study was conducted. The study consisted of two parts; a user test followed up by an interview. With this approach, the results consisted of measurable test data and in-depth responses to that data. The tests were performed on ten subjects that will be randomly picked students from Umeå University, that used a smartphone on a regular basis.
4.1 Interactive testing software

A testing software is created using Android Studio to be used on different Android-devices. The goal of this part is to establish statistical support for possible differences in execution-time of identical tasks with varying screen sizes. Since a major factor in this type of testing is the subject’s hand and finger size, this must also be addressed in the tests in order to collect viable results. The testing process consists of two parts;

Setup

Credentials The testsubject is presented with a series of screens, where personal credentials are to be answered. This data will be correlated with the result of the rest of the test. The testsubject will state their name (see Figure 3a) and which hand the device will be held in during the test (see Figure 3b). Selecting which hand will be used, is a critical piece of data when analyzing the current design principles that mobile applications are to follow.

![Input field for name](image1)

![Hand selection for tests](image2)

(a) Input field for name  
(b) Hand selection for tests

Fig. 3: User interface of credentials input

Measurement of reach In order for the results to be consistent over different subjects over different hand and finger sizes, the subjects reach capabilities over the screen is measured. To accurately measure reach, the hand should be flat with the thumb pointing upwards. The device is then placed on top of the hand (on the index-finger), with the usb-connector against the first joint of the thumb. The reach is then measured by interacting with the slider on the interface of the device (see Figure 4a), sliding it as far as possible towards the top of the screen.
Intro to test  The setup is after those steps complete. A screen is displayed that briefly explains the test, and encourages the test subject to start the test when ready (see Figure 4b)

Test: Sequence of interactions  The subject will undergo a series of tests where the objective is to perform a sequence of different interactions, which will be measured by miss-clicks and time to completion. When the user is ready and starts the test from the initial layout (see Figure 5a) buttons will appear one at a time, and will disappear on press. The test will focus on the outline of the screen, so buttons will be placed along the outline in corners (see Figure 5c), and will always be followed by a button in the center (see Figure 5b). With this approach each sequence can be singled out without bias to placement of previous buttons. When the test is complete (see Figure 5d) the score can be sent to a server where all test results are saved. All results can be viewed thru that server via a web-client.

4.2 Interview

To follow up on the user’s thoughts and experiences when undergoing the tests, an interview will be conducted. The interview consists of a few standard questions for all subjects, and a more subjective part where the focus is on the question why and how? Why did certain situations occur during the tests? Why did you perform certain actions? How would you have wanted to interact instead? The results from the testings software and the interview will be compared and analyzed in order to find patterns and possibly a higher bound for screen-size
(a) Start screen of the test

(b) A button at the center of the screen

(c) A button on the outline of the screen

(d) Final screen displayed after test is completed

Fig. 5: Testing interface with buttons

(dependning on hand/thumb size). This higher bound will mark where the screen-size makes interactions awkward or hard to perform, and will be calculated by finding long execution-times and comparing that with the post-test interviewie.

**Limitations** These test will measure execution-time over different devices and screensizes, however it will not cover how hand-placement will impact the results. The phones that will be used are unlocked when handed to the test subject, and they will be told that no hardware-buttons will be used for the tests. Since the software do not recognize if one hand is used, the tests will be supervised to verify this.

5 Result

5.1 Test: Sequence of interactions

The results from these tests are conducted on three (3) different devices within the large-spectrum (that was mentioned earlier) [8]:

- Samsung Galaxy Core: 4.5 inch (Core)
- LG Nexus 5X: 5.2 inch (5X)
- LG G4: 5.5 inch (G4)

All of the test subjects chose to use their right hand for the tests, and the results are divided and presented according to the specifications in the Method-section:

**Total time over different screen sizes** The effective time for the test subjects to complete all sequences over all three devices are summerized (see Figure 6)
The results shows that performing interactions with the Core is faster than the other two devices. Furthermore the majority of test subjects had a faster execution-time with the 5X inch screen compared to the device with the largest screen (G4). The abnormality in this result is test-subject 2 performing the test with the 5X almost double the time compared to the median time. This occurrence is from the test-subject almost dropping the device due to the strain from reaching towards the next button.

Combined time over each sequence Following are the results of effective time in each section (between each button press), combined over all devices and test subjects (see Figure 7a) and combined over all test using the Core device (see Figure 7b). Due to the abnormality in one of the tests, that result is not included in this chart, but will be included as an effect of the test itself. The reason is for the results of the cart to reflect general succesful tests measuring effective time.

These result highlights three major spikes in the time-data, sections 2, 4 and 6. We can establish that the button placed in the top opposite corner of the hand used is the hardest one to reach, followed by bottom opposite corner and top corresponding corner (see Figure 8).

Reach over totaltime All subjects measured their reach on the device before conducting the sequence tests, in order to determine if reach has an impact on the effective time when interacting with mobile devices (see Figure 9). The scale on which the reach is measured is abstract, and does not transfer between the three test devices. Because of that, the results of this part is only measured on a device basis.

Fig. 6: Total time for subject to perform all sequences of the test
Looking the results obtained from the Core (see Figure 9a) it is not possible to determine any obvious pattern, as the data is scattered over the entire reach spectrum. After examining the results from 5X and G4 (Fig 9b and 9c respectively) there is a resemblance to a pattern that the effective time is decreased when the reach increases.
5.2 Interviews

The interviews that were performed after each test resulted in 80

6 Discussion

This study requires the test subjects to be somewhat equal, even if factors like reach are taken into consideration. It relies on that all subjects have the same skillset when handling a mobile device. This is not easy to determine, and from watching the tests it became easier and easier to spot the subjects who perform these kind of interactions on big screens on a regular basis. They have certain techniques that allows them to reach further than with a steady grip. This did however not change the fact that the effective time still was slowed down because of the large screen size. And even in this controlled test environment there was one subject who almost dropped the device while trying to interact, and many subjects that came to a stall when they had to change the positioning of the device within the hand itself. When out "in the wild" (in the real world) it becomes a much bigger problem to not have a firm grip of your device. The reason for this is all due to the physical size, that forces a shift in hand placement. By observing the test subjects performing these tests, it became clear that when they were not able to reach a button from the default hand placement, they came to a stall while readjusting the device within their hand.

When going deeper into the test results (see Figure 7a), it is very clear that the time lost on the larger devices originates from interactions within the South and NorthWest region of the screen (see Figure 2 for details). This correlates well with the study by Karlson et al. citeOneHandedUse that was mentioned earlier. All test subjects managed to complete the tests, which shows that the screens used in this study did not prevent any interactions. It did however force many subjects to stop interacting with the screen and change their hand placement before the interaction was possible. From the results from the reach test, the data is not as clearly stated. There is a resemblance of pattern for the 5X and G4, but for the Core there is no pattern at all (see Figure 9). What this shows is that when the screen size of the mobile device is increasing, the reach and thumbsize becomes a more significant factor. This test group does however only consist of ten subjects, which by its small size makes it harder to recognize a deviation from an actual finding.

The tests performed in this study are limited to simple interactions such as pressing buttons, as this is the way that users interact with the elements and components mentioned earlier in Section 3.1. It does however not include more complex interactions such as swiping or dragging objects. In order to go further into the issues that this article is focusing on, test with those kind of elements should be considered in order to get a deeper understanding of the users behaviour and potential issues that arise. The number of devices that were tested could also increase. Because of technical issues, the test environment was not installed on a forth 4.0 inch device. This size is not within the large spectrum, but would be a good benchmark to compare with. After observing how the Core
was used, it showed that not a single test subject had any issues interacting with a device with a screen size of 4.5 inch. There where no adjustment of hand placement or awkward postioning. As we can see from the Core results (see Figure 7b) the biggest total time difference between the sections are 2.5 seconds for 9 tests. If we compare that with the overall results that have almost 12 seconds over 27 tests, this is a significant difference that shows how the issues with interactions that are affected by screen size actually get enlarged. These findings should affect the mobile software developers in a way that they will have to take this into consideration when developing for a wide range of devices. It should also be a part of the design process when manufacturers creates new devices, in order to meet as many needs of their customers as possible.

7 Conclusion

Using a larger device with one hand is not an easy task nowadays. It's becoming harder and harder for users to interact with their mobile device in a way that they have before. From the data obtained in this study, it is certain that the screen size of mobile devices do affect interactions with it. Not only is it more time consuming, it also creates a physical strain on the hand. From the results (see Figure 6) it is clear that the screensize do affect the time it takes to interact with it. It also point to that even a minor enlargement (from 5.2 to 5.5 inch) changes the effective time when interacting.

Interactions located at positions 2, 4 and 6 (see Figure 8) are directly affected by the screensize of the mobile device, and this should be accounted when designing mobile applications that supports different screen sizes. The users reach when interacting with mobile devices becomes a more significant factor when the screen becomes larger, as the user will be forced to change hand postioning in order to perform said interaction. Mobile devices have over the years become larger, both as a part of our daily lives and also as the physical size keeps growing, and the effects of this are clearly shown in this study. Steve Jobs stated that a 3.5 inch screen is optimal for handheld interactions, and with the results from this study this statement still stands strong.

References

Evaluating the potential and limitations of Google Glass

Joakim Zakrisson
Department of Applied Physics and Electronics
Umeå University, Sweden
zakrisson88@gmail.com

Abstract. Google Glass was originally predicted to become a very popular device amongst the general public. This turned out to not be the case as Google has since January 2015 halted it’s production of the Google Glass device. It was found that Google Glass had issues with its current devices, such as hardware limitations, no clear usage area, high price tag and users experience they are clunky. Although Google Glass has also been used together with experiments within the fields of telemedicine and teletoxicology where they might become useful. But first the device would need to resolve some of its issues. To evaluate the potential and limitations of Google Glass different methods were used, such as theoretical examination and a survey. The theoretical examination was looking into the background of the Google Glass, which area of interest the device might have and what kind of limitations the device had. And the survey was sent to persons who are generally interested in technology which improves the chances that they have used a wearable device and is more likely to leave relevant feedback to the survey. The survey had open questions about wearable devices in general where the person could leave feedback for devices they had used. Also the survey examined the pro’s and con’s with each wearable device and was investigating what made a person not be interested in the Google Glass. The data gathered from the theoretical examination and the survey was used together to examine why Google Glass did not become as popular as other wearable devices. It was concluded that Google Glass did have limitations with the current devices which had a negative effect on the user experience, also people do prefer to wear a wrist-worn device instead of glasses. Furthermore it was concluded that Google Glass might have usage areas within different work-related fields such as telemedicine.
1 Introduction

Google Glass is originally developed by Google X, which is a semi-secret research and development facility created by Google. Google Glass is a popular device within the category of smart glasses. The original intended purpose of the smart glasses was to be a hands-free display of information available to most smartphone users. According to Cellan-Jones [1] Google announced as of January 2015 that they would stop producing the Google Glass prototype but is still committed to keep developing the product where project Glass has been “graduated” out of Google X.

Looking at other wearable devices it is something that has increased in popularity during the last few years. The total amount of unit sales of wearable devices reached 36 million units during the year of 2014. Smart glasses did account for 1 million units and smart watches 4 million units. During 2015 the total amount of wearable devices being sold was predicted to 72 million units with smart glasses at 2 million and smart watches at 26 million units [2]. Wearable sales is really taking off and it is something that is becoming very popular and attractive to the general public. With sales for wearable devices increasing this
much especially smart watches, why is smart glasses not increasing in popularity the same way?

1.1 Objective

The purpose of this research is to identify what is making wearable devices attractive and why Google Glass has not increased in popularity the same way other wearable devices (e.g. smart watches) have. During this research the following questions was taken into consideration:

- Are there any hardware or software limitations that users have experienced with Google Glass which might restrict the device’s popularity?
- What functionality is a user generally looking for at a wearable device, and can such functionality be added or any other improvements be made for Google Glass to make them more attractive to the general public?

Theoretical examination was made to examine these questions and also a survey was made with purpose to figure out what is making a person become interested in a wearable device and to analyse what could possibly make the person more interested in Google Glass.

2 Theory

Smart watches has increased in popularity the last few years with the release of devices like Apple Watch. During the second quarter of 2014 Samsung stood for 73.6 percent of the global market share of smart watches with 0.7 million devices sold [3]. However, during the same period in 2015 Apple had 75.5 percent of the global market share for smart watches with 4 million devices sold [3]. This data shows how popularity has increased for smart watches between 2014-2015 and how much Apple is dominating the market since Apple Watch was released.

Smart glasses did not become as popular as predicted; In late 2013 Google Glass was predicted to by 2016 become a mainstream product and sell million of devices [4]. This however did not turn out to be true since Google halted it’s production of Google Glass as of January 15, 2015 [1]. Also while Google Glass together with other smart glasses where still being produced and sold in 2014 they only accounted for 1 million devices sold while smart watches sold by 4 million devices [2].

In a scientific paper about the importance of well designed applications it is stated that an important aspect with wearable devices is not just the aesthetic value of the device itself but also the design of the applications [5]. They also conclude that well designed applications with good functionality will make a person more likely to enjoy the wearable even though it does not have as good aesthetic value as other devices.
2.1 Background

In an article by D’Onfro [6] which is about the first release of Google Glass devices it was stated that Google Glass was announced as the first smart glass device offered for personal use back in 2013. D’Onfro continues, that in 2013 a handful of people could get their hands on Google Glass where they could pick up a pair of the device at designated Glass "Basecamps" located in San Francisco, New York, Los Angeles and London. Furthermore D’Onfro states that the device had a price tag of 1500 USD by then and the first employees responsible for working at those "Basecamps" were hired without knowing what they would be doing because Google Glass had been a secret project. She continues that during training those employees had to practice by pitching random Google products. Also in a scientific paper about who might be interested in Google Glass it is stated that Google started distributing a limited number of Google Glass to selected people such as developers and researchers in 2014 [7]. They also state that Google then announced they would start selling Google Glass to the general public during the fall of 2014.

Google themselves stated in January 15, 2015 that the Google Glass project is moved out of Google x because Glass at work has been growing and they are seeing big developments with Glass in the workplace [8]. Google also state that Google Glass will instead have their own team at Google and is now focusing to build for the future where they eventually will release new versions of Google Glass. According to Google the last pair of Google Glass explorer edition could be obtained until the 19th of January 2015, thereafter Google’s explorers program was closed.

2.2 Area of interest

According to a research article about clinical and surgical applications of smart glasses [9], there is currently great interest in using Google Glass together with health information technology within health care to improve patient care, decrease cost, and increase efficiency of health care professionals. The article also states that hands-free web connected devices have great utility to health care professionals by providing access to patient records for viewing or input. Furthermore they continue by stating that studies show that smart glasses have great potential to reduce the time doctors spend managing Electric Health Record systems and data entry where the front facing camera also has potential to help with surgeries, improve documentation of clinical data and analyse diagnostic tests. Furthermore the article states that additionally the camera can be used with telemedicine. The article also shows that smart glasses still have flaws and before they can be recommended for general medical use improvements to the hardware and specialized software need to be developed.

Google Glass have also been examined for use in telemedicine as stated in an article where a study was made where two paramedic teams triaged the same 20 patients [10]. Telemedicine is considered the use of telecommunication and information technologies to provide clinical health care at a distance. In
this study one team was using telemedicine together with Google Glass which they used to communicate with an off-site physician disaster expert. The other team was using no telemedicine and was acting as they would normally do. The conclusion of this study showed that there were no particular increase in triage accuracy when paramedics were evaluating disaster victims using telemedicine with Google Glass than paramedics who did not use the device. Furthermore the study that the team of paramedics who used telemedicine required more time to evaluate and triage the patients than the team who did not. The study also states that this might change in the future since Google Glass at this time have some hardware limitations which might had influence on the result and also the paramedics were unfamiliar with this kind of technology.

Google Glass have also been examined within the area of teletoxicology. Teletoxicology is like telemedicine but with focus on chemicals and to counsel persons who might have been exposed to chemicals at a distance. A study was conducted where 19 consults were attempted by a medical toxicology consult staff [11]. Google Glass were used at this study and provided real-time video communication with the patient and medical staff. The same study was also made without the use of Google Glass. The study concluded that in the case where teletoxicology were used together with Google Glass 89 percent of the cases were considered successful according to a supervisory consultant whereas without the use of teletoxicology 59 percent of the cases were seen as successful. This study shows that Google Glass do have potential within this area of use.

Health and wellness areas will see the biggest early jump in market numbers since sensors in wearable devices can be cross-purposed for everything from fitness to technology for aging which is stated in an article by Ranck [12]. Ranck also concludes that many customers are unsatisfied with medical devices that look like medical devices and prefer design aesthetics of a company like Apple. Also according to Ranck wearable technologies developed for military use are also making their way into the sports and outdoor markets.

2.3 Google Glass limitations

Some limitations and issues have been encountered with the Google Glass device. Described below are some limitations that was encountered with a study where Google Glass was used as an experiment together with telemedicine [10].

- The need for internet connectivity with dependence on either Wi-Fi network or Bluetooth connection with a smart phone.
- Short battery life which further decreased during video streaming.
- Microphone position which also picked up lots of background noise and was not always responsive to verbal commands
- The screen was very bad outdoors, the contrast made the screen very hard to see in outdoor light.

Furthermore the study concluded that there were also software problems which occurred. Furthermore the study concluded that there were a significant lag in
live-video broadcasting with a delay of more than 20 seconds between the disaster drill site and the off-site physician, although audio communications was almost instantaneous. They continue by the software also had no lock-functionality to prevent the possibility of inadvertently halting the streaming. The study also states that the device was unable to opt-out the frequent OS-updates which complicates appropriate technology selection and research planning.

Another study also states similar issues with the Google Glass device. During this study Google Glass were researched within use of teletoxiology [11], here a total of 19 consults were made where Google Glass provided real-time video communication with the patient and medical staff. The persons using the device experienced hardware limitations such as video lag and interruptions which occurred during video-feed. One of the consultants also experienced audio-lag and in another case the consultant were unusable to use the device because of both video and audio lag occurred.

In a research paper where Google Glass was used together with video-streaming it was concluded that the device also has temperature problems [13]. The results from this research paper shows that Google Glass has a threshold at 55 degree celsius from where it will ask the user to stop the current activity for the Glasses to cool down. They continue by stating that after exceeding this threshold the glasses will operate at a lower frequency for the CPU, also lower the voltage. This will result in less power drainage and slower temperature change but at the cost of user experience. Furthermore to test these temperature problems the authors used the Google Glass to live stream a 5 minute video via bluetooth and Wi-fi. Their results from this test shows that the temperature for the glasses rises significantly the first 3 minutes of live-streaming and then gradually stabilizes. The article further states: "Our results suggest that Google Glasses do not seem to comply to user expectations in terms of good user experience for the streaming applications. Streaming applications are not suitable for longer duration of use on Glass with its current hardware configuration as it leads to high power consumption and temperature rise. High power consumption leads to low battery life and high temperature which can be dangerous to user skin" The same research paper also does web browsing experiments together with the Google Glass where 16 different web pages are measured to check the power and heat consumption of Google Glass [13]. The research concluded that for example wsj.com needs 3 redirects before going to the main mobile webpage of the site. This would increase the page load time by 2-3 seconds on the Google Glass. They also state that apart from time, redirects also leads to wastage of power on the glasses. The article also concludes: "Extra time to load and extra power dissipation due to redirections is not good for the quality of user experience on wearables. We think that the number of redirections should be minimized and better be avoided by better webpage design for such sites." From the websites tested youtube.com was the most optimal for the Google Glass. It uses efficient network protocols and image formats to deliver the content faster. Ted.com could similary to youtube convert all .jpg images to .webp to reduce the power consumption on the Google Glass by 12 percent with no change to the user experience.
3 Method

To fulfill this research a survey was made to find out why smart phone users also want to use a wearable device and how they use it to complement their smartphone. The survey was also used to find out what is making a wearable device interesting for the user and what kind of functionality they are looking for in such device. Furthermore the survey was used to figure out which properties are most important and what kind of functionality or characteristics is the smartphone users looking for in a wearable device. Also the survey was used to identify why a user would not be interested in Google Glass.

Similar studies or research articles were used as a source to help answer this. Furthermore, other research papers will be used with focus on smart glasses to find out why they did not become as attractive to the general public as other wearable devices e.g. smart watches and why Google Glass did not become as popular as people first had predicted.

3.1 Survey

The survey was used to identify what functions a smart-phone user is generally looking for in a wearable device. The survey was also examining what the user did find lacking with Google Glass and possibly what functionality or other characteristics could change or be added to make them more attractive. Another objective with this survey was to identify why the person is currently using or not using a wearable device and what was the person's reasoning for this.

The survey was split into two different parts. The first part consists of questions about Google Glass where the survey will try to gather pros and cons of people who did use them. Furthermore some general questions about Google Glass was asked, for example if the person would be interested in owning their own pair of Google Glass and what the person would be willing to pay for the device.

The other part is about other wearable devices. Also here the general opinion about the wearable devices a person has used was asked including pro’s and con’s. Furthermore the survey was identifying why a person would prefer another wearable device instead of Google Glass. It did also examine which wearable device are most appealing to the person and which property in such device is most important.

The aim was for the survey to reach 10-20 persons who were interested in technology and preferably have used some kind of wearable device to be able to share their general opinions about them.

3.2 Survey target group

The target group for the survey was persons who currently own a smart phone and are within the age of 20-30 years old. The persons within the target group was university students who were currently studying within the field of technology or computer science. This specific target group was selected mainly because
a relatively small range of age indicates that the persons have similar experience with smart phone and wearable technology. And persons studying within the field of technology would more likely be interested in wearable devices and would therefore be more likely to have used some wearable device themselves. Also persons who currently own a smart phone was mandatory because a wearable device is often used as a complement to the user’s smart phone.

3.3 Theoretical examination

To complement the survey, theoretical examination was done to review scientific articles with focus on wearable devices or smart glasses. The purpose was to evaluate what current wearable devices are being used for and why smart glasses, such as Google Glass, did not take off the way it was predicted to. To identify this, theoretical examination was made for other wearable devices such as smart watches to analyse what characteristics were making such device popular. Furthermore, Google Glass current usage areas and limitations with the current version of Google Glass was examined.

4 Results

4.1 Theoretical examination

In this section conclusions gathered from theoretical examination will be shown. Below is a list of some hardware related limitations experienced by the users in the experiments with Telemedicine [10], Teletoxicology [11] and video-streaming [13].

- Dependence on either Wi-Fi network or Bluetooth connection.
- Short battery life.
- Microphone not always responsive and picking up background noise.
- The screen was very bad outdoors, the contrast made the screen very hard to see in outdoor light.
- Overheating issues.
- Lag while video- or voice streaming.

The hardware related limitations are not the only problems which users did experience with Google Glass. In the video-streaming experiment they also identified that due to websites being poorly optimized for devices such as Google Glass some websites drains more power than needed looking from a user-experience perspective [13]. The experiment show that for example Ted.com could lower the power consumption by loading the website by 12 percent only if they did change all images on the site from .jpg to .webp. It’s also important that websites use efficient network protocols to become more efficient. This shows that limitations experienced with the Google Glass could be altered by providing optimized content, like applications and websites.
4.2 Survey

In this section data gathered from the survey is shown.

![Bar chart](image1.png)

**Fig. 2.** Data from survey: Whether a person has used Google Glass

![Bar chart](image2.png)

**Fig. 3.** Data from survey: Whether a person would be interested in Google Glass
**Fig. 4.** Data from survey: How much the person would be willing to pay for Google Glass

**Fig. 5.** Data from survey: Which wearable devices has the person used
**Fig. 6.** Data from survey: Weighted average for what score each wearable device had. The scores range from 1-5 where 5 was given to the most interesting device.

**Fig. 7.** Data from survey: Weighted average for which property is most important. The scores range from 1-4 where 4 was given to the most important property.
Furthermore the survey had some open questions where the most relevant answers will be shown below:

For the 36 percent of the persons taking the survey who are not interested in owning a pair of Google Glass there was an follow-up question as of why not. The results are listed below:

- Limitations, such as battery time and overheating issues.
- No clear usage area.
- Technology is not fascinating.
- The device being too expensive.
- They feel clunky.

Also the survey had another question for people who have used a wearable device such as smart glasses, smart watches or wrist-worn activity trackers. In this question they could state their general opinion about such device. The results are listed below:

Smart glasses:
- Easy to obtain information.
- No clear usage area.
- Can get tiresome to use after a while.

Smart watches
- Easy to access information from smart phone directly on the wrist.
- No clear usage area.
- Too small which cause irritation.
- Too much focus on fitness applications.

Wrist-worn activity trackers
- Help motivate a more active lifestyle due to result logging and being able to compete with old results.
- Relatively low price tag.
- Easy to access information from smart phone directly on the wrist.
- Clear usage area
- Can get tiresome to use after a while.

There was also a general question about what would make the person taking the survey prefer another wearable device instead of Google Glass. The results are listed below:

- A lower price-tag
- Prefer wearing a wristband over glasses
- A clear usage area
- Better hardware specifications
- Appearance
- Wrist worn devices is more subtle and is easier to wear
5 Discussion

From the survey and theoretical examination it can be concluded that Google Glass had some limitations, especially hardware related ones. These limitations were experienced by users when the device was being used in different experiments such as within telemedicine and teletoxicology. Such limitations include problems with poor battery time and the device overheating. Also other limitations such as application lag while video- or voice-streaming, microphone picking up background noise and the screen having a bad contrast which made it hard to see, especially in outdoor light.

Other than hardware related issues it is also challenging to find a clear usage area for the Google Glass device. This was shown as a result in the survey where the persons who took the survey did find wearable devices helpful since they can easily provide the user information without having to pick up their smart phone. But other than that many stated that the Google Glass had no clear area of use which made them not as interesting as smart watches or activity trackers. Persons who took the survey did find that activity trackers had a very clear usage area which was to motivate the user to a more active lifestyle where smart watches was like Google Glass, a device to help provide the user information without having to look at their smart phone. Although people did prefer wearing a smart watch compared to smart glasses because a wrist-worn device is more subtle and can easily be brought anywhere by the user. This would mean that even though the smart glasses and smart watches had similar functionality most people would prefer wearing the watch.

Another problem with the Google Glass is the high price tag they initially had. As stated in an article by Business Insider [6] the Google Glass had an initial price tag of 1500 USD which is equal to 12100 SEK [14]. However, from the survey we can gather the following results for what a person would be interested to pay for Google Glass: 42.86 percent would be willing to pay between 0-2500 SEK, 50.00 percent would be willing to pay between 2500-5000 SEK and 7.14 percent would be willing to pay between 5000-7500 SEK. No person who took the survey would be willing to pay more than that. This clearly shows that persons are not willing to pay as much as 12100 SEK for the Google Glass device but rather somewhere between 0-5000 SEK.

From the text above it can be concluded that Google Glass has issues and is in need of some refining before the device will become popular to the general public. It seems Google would need to add functionality other devices do not have, lower the price tag and fix the devices hardware limitations before the glasses would become interesting to the general public.

6 Conclusion

The questions stated by the objective in section 1.1 is taken into consideration and will try to be answered. To begin with the limitations with the Google Glass device is being looked at. As seen in the results the Google Glass device
had some limitations which were both hardware and software related. Many users experienced issues like the device becoming overheated after a some time of use. The battery time was very short. The device was always dependent on either Wi-Fi network or Bluetooth connection. The screen had poor contrast which made it hard to see clearly, especially in outdoor light. Other than that it had lag related issues which was experienced together with video- or voice streaming and the functionality offered by the device makes it hard to find a clear usage area and users would prefer wearing a wrist-worn device.

To continue the functionality a user is looking for in a wearable device is being looked at. And was there any improvements that can be made for Google Glass to make them as attractive to the general public as other wearable devices. It can be concluded that users mainly wants to use a wearable device to easily obtain information without having to reach out for their smart phone. Also while smart watches has no clear usage area besides providing the user information from their smart phone, much like the Google Glass, it was concluded that people would still prefer wearing a wrist worn device instead of having to wear a pair of glasses. Data from the survey shows that most people could see a clear usage area for activity trackers which is mainly used to help motivate the user to a more active lifestyle. People also do prefer to wear a more subtle device which both activity trackers and smart watches are compared to smart glasses, this also makes them more neat to wear in their everyday life.

It is challenging to conclude what kind of functionality could be added to Google Glass to make them more attractive but to start with it would need to resolve the limitations the users have experienced with the device. To lower the price tag would most likely benefit the popularity of Google Glass, which was seen from the survey that people were not willing to pay as much for the device as it was initially priced at. To make smart glasses as attractive to the general public as smart watches or activity trackers could still prove to be challenging as people do prefer to wear wrist-worn devices. Perhaps smart glasses should not have its usage area within the general public but rather be used within work-related fields which it has already been experimented with. It may become very popular within fields such as telemedicine where the biggest issue was the hardware limitations which can always be resolved.
References

Design of brain-computer interface to ease everyday life for people with severe motor impairments: A metanalysis

Ellinor Rönn
Department of Applied Physics and Electronics
Umeå University, Sweden
elra0017@student.umu.se

Abstract. For people who are fully motor impaired, brain-computer interface (BCI) may provide the only means of communication. However, there is a gap between the BCI performance and the users’ desires for a meaningful communication. The goal of this study was to develop guidelines of how to design user interfaces of BCI applications that meet the users’ desires. This metanalysis evaluated scientific articles with focus on users’ desires and present design solutions in order to identify design solutions that meet the users’ desires. The results show that the users want a spelling speed of at least 15-19 characters per minute (cpm) and an accuracy of 90%. The accepted training time to be able to use the BCI independently is approximately five sessions. Many of the studied designs achieved the desired training time and accuracy (max = 100%). However, only a few reached the desired speed (max = 20.6 cpm). Because of the trade-off between accuracy and speed, none of the designs reached the desired performance in both accuracy and speed. Both P300 and SSVEP was found to be suitable with a performance close to the users’ desires. The spellers, at least the P300 based spellers, can be improved with predictive text entry to achieve better efficiency and user satisfaction. The trade-off between speed and accuracy should be carefully considered when designing BCI spellers. Furthermore, to design good BCIs for users with severe motor impairments it is also important to consider differences in e.g. vision, cognitive functions in working memory, etc.

1 Introduction

Individuals with severe motor impairments often lack the means of communications, such as speech, gestures and voluntary facial expressions. Happ studied in 2000 how voicelessness affects both the patient and their family members and loved ones [1]. In this study Happ showed that the inability to speak made the patients terrified and frustrated, while the patients’ family members frequently expressed regret and anxiety over the voicelessness. Happ further showed that even for family members it was nearly impossible to interpret the voiceless patient’s feelings, experiences and desires. This indicates the difficulties for the
patients to express their opinions and decisions regarding clinical treatment or their living and biological wills.

Brain-computer interface (BCI) has been described as a possible tool to enable these patients to communicate and control their environment [2,3]. BCI is a technology that enables humans to control devices through their mind, independent of muscle movements [4]. A BCI translates the users brain activity into commands, usually by measuring the user’s electroencephalogram (EEG) acquired from the scalp [5]. There has been a lot of work in research and development of BCI, both in health care and entertainment. For example, there exists BCI technology that enables people with motor impairments to spell messages [6–9], operate a cursor on a screen [7] and control environmental features [10].

1.1 Objective

This study examines previous papers and existing BCIs from a usability aspect in order to find “good” designs from the user’s point of view. In this study, a “good” design is measured by Rubin and Chisnell’s [11] description of usability. That is usefulness, efficiency, effectiveness, learnability and satisfaction.

The goal of this metanalysis is to use these attributes to develop guidelines for how to design user interfaces of BCI applications to ease everyday life for people with severe motor disabilities. Since it has been shown that communicating is one of the most important things locked-in patients want from a BCI system [3], this study focuses on this specific task. Questions to answer are:

- What do BCI users with motor impairments want in a BCI system?
- How should BCIs be designed to meet the users’ desires?

To answer these questions, previous studies will be analysed in order to identify which properties motor impaired people consider most important in a BCI system. Furthermore, to analyse and compare present design solutions, their advantages and disadvantages.

2 Theory

Brain-computer interface (BCI) is a technology that enables humans to control devices through their mind. In 1999, in an international meeting between different research groups, BCI was defined as “a communication system that does not depend on the brain’s normal output pathways of peripheral nerves and muscles” [4]. This is usually accomplished by measuring the user’s electroencephalogram (EEG) acquired from the scalp, and translate the response into commands [5]. There are several commercially available BCI devices with a variety of sensor interpretations, for example the Emotiv EPOC headset and Neurosky MindWave headset, which both have shown to be suitable for use in BCI research [12].
2.1 BCI stimuli

BCIs use different forms of stimuli for communication. Three regularly used forms are P300, steady-state visually evoked potentials (SSVEP), and the imagined movement paradigm (also known as motor imagery, event-related resynchronisation or event-related synchronisation) [13].

**P300** is an event-related potential, manifested as a positive response in the EEG around 300 ms post stimulus [8, 9, 14]. The user focuses their attention on a rare stimulus among many irrelevant stimuli [9]. An example is a matrix of characters where the rows and columns flash in a random manner, and the user focuses on the character they wish to select [8, 9, 13, 15].

**SSVEP** is a periodic response evoked by a visual stimulus that flashes at a constant rate, evoking a response in the user’s EEG that matches the frequency of the flash rate [13, 14]. By using multiple stimuli flashing in different rates, a number of choices can be made [13, 14]. An example of SSVEP is to use four light emitting diodes as stimuli for navigating up, down, left and right. SSVEP based BCI spellers usually use a decision-tree method or moving the cursor position to select the target [14].

**Imagined movement** is controlled by the user imagining moving a certain body part, such as a hand or foot. This will induce sensory motor rhythms over the sensory-motor cortex that can be detected and mapped onto different commands [14, 16].

2.2 BCI for motor disabled people

Lately there has been a lot of research in the field of BCI and how it can be used to help people with motor disabilities in activities of everyday living. For example through spelling programs, prosthesis and robotic arms that can be controlled by the use of BCI. People can suffer from lack of motor function and severe levels of paralysis for many reasons, e.g. amyotrophic lateral sclerosis (ALS), brain injury and stroke. In the late stage of the neuromuscular disease ALS, the patients have no possible way to communicate due to loss of speech and motor functions. BCI has been described as a possible tool for ALS patients, to enable them to control and communicate with their surrounding [2, 3]. In 2011 Huggins, Wren and Gruis [3] showed that being able to communicate with other people is one of the most important things ALS-patients want in a BCI-system.

It has been shown that for people who are partially motor impaired, it is in many cases more effective with devices that detect voluntary motions such as head movements, eye gazing, or eye blink, than using BCIs [6]. In another study, however, it was reported that painting with BCI was less exhausting than using eye tracking [17]. For fully motor impaired people however, BCI may offer the only means of communication [6, 8, 18, 19].
2.3 Character by character versus predictive text entry

In a predictive text entry (PTE) system the user is provided with examples after writing the first letter in a word. For example, after spelling “y” the alternative “your” can appear with the number “1” in a list of possible words, and the user can spell the word “your” by focusing on “1” instead of spelling the whole word character by character [9]. Since a predictive speller may enable the user to write more text with fewer selections it could enhance communication for people using a P300 BCI. However, in a study by Ryan et al. [8] it is suggested that predictive spellers in non-BCI context increases cognitive demand, which can lead to a lower accuracy.

3 Methods

Since there were no BCI devices available for this study, it was conducted by analysing previous articles. When selecting articles, search terms used were; “brain-computer interface(s)” or “BCI” in combination with “motor impairments”, “motor disabilities”, “design(s)”, “speller(s)”, “user-centred (design)” or “usability”. The reference list of the retrieved articles were further looked into to identify additional relevant articles. From the search result, all relevant articles that had been peer-reviewed was selected. However, due to the time limitation of the study, it was not possible to analyse all of the articles. The articles were analysed in random order until the time limit was reached. One selected article was excluded from the study due to numerical errors in their results.

The information extracted from the articles was training time, output characters per minute (OCM), accuracy, information about the test subjects, user evaluation and any other relevant information. Some of the articles did not provide the result of OCM, which was needed in order to compare the users’ desires to the performance of the studied BCIs. When no OCM was provided, it was estimated by the time it takes to enter one character. If this was not given, it was estimated by the given formula used to calculate the information transfer rate, which depends on the typing speed. With a given time for each selection and an accuracy above 50%, the correctly entered characters per minute was estimated with the formula

\[ OCM = \frac{60(2P - 1)}{T} \]

where P is the accuracy and T is the time needed to spell one character. For \( P \leq 0.5 \), \( OCM = 0 \). This is based on the formulas Park et al. [21], and Yin et al. [25] used in their studies.

3.1 Metrics

The different designs was measured by Rubin and Chisnell’s [11] description of usability:
**Usefulness** - The degree to which a product enables a user to achieve their goals and the user’s willingness to use it at all. This was measured by self-reported opinions in the design evaluation.

**Efficiency** - The quickness with which the user can accomplish their goals accurately and completely. This was measured by the selection rate, i.e. the number of commands/characters per minute (cpm) when writing a message. Several BCIs used predictive text entry, which means that one selection results in more than one character. Therefore the output characters per minute (OCM), was also used. OCM refers to total number of characters spelled, i.e. “your” is interpreted as four characters in both character-by-character and PTE, in contrast to the selection rate that interpreted “your” as two characters when selecting “y” followed by the predicted word “your”.

**Effectiveness** - The extent to which the product behaves the way the user expects it to, and the ease with which the user can use it to do what they intend. This was measured by the the accuracy, that is the number of successfully completed commands divided by the total number of commands.

**Learnability** - The user’s ability to operate the system to a defined level of competence after a predetermined amount of training time. This was measured by the training time needed to be able to operate the BCI independently.

**Satisfaction** - The user’s perceived comfort and acceptability while using the product. This was measured by evaluating the users’ self-reported satisfaction.

To evaluate which of these properties is most important from a user’s perspective, ratings from earlier studies were collected and compiled. Usefulness however, was not further investigated when evaluating the users’ desires since it is assumed that the users want a useful system. Other important aspects mentioned in the articles were also investigated.

## 4 Results

The study identified 2 articles regarding what BCI users with motor impairments would want in a BCI. Furthermore, it identified 11 articles that presents one or more design solutions with data from user studies.

### 4.1 What users want in a BCI

**Huggins et al. 2011** Huggins et al. [3] conducted a survey to evaluate what ALS patients would want in a BCI. 63 participants with ALS responded to a telephone survey, with 61 participants answering BCI questions. Huggins et al. showed that the highest rated BCI design feature is ‘Accuracy of BCI operation’, followed by ‘simplicity of BCI setup’, ‘Functions the BCI provides’ and ‘reliability of BCI standby mode’. Appearance was the least important feature.

The result in their study shows that the users want a typing speed at 15-19 letters per minute (see Table 1). Furthermore they want an accuracy at minimum
90%, that is the minimum percentage of the time the BCI has to recognize your commands correctly. Acceptable training time was 2-5 sessions. Another important finding is that the common occurrence of visual and auditory impairments should be considered during BCI design.

**Huggins et al. 2015** Huggins et al. [22] conducted a survey to establish what BCI users with spinal cord injury would want in a BCI. 40 subjects participated in the study. The results showed that the patients want a spelling speed at 20-24 letters per minute, and an accuracy at minimum 90%. Furthermore, the patients would accept a training time at 6-10 sessions to be able to use the BCI independently.

<table>
<thead>
<tr>
<th>Table 1. Results of what the users would want</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Article</strong></td>
</tr>
<tr>
<td>Huggins et al. 2011</td>
</tr>
<tr>
<td>Huggins et al. 2015</td>
</tr>
</tbody>
</table>

### 4.2 Present design solutions

**Kaufmann et al. 2012** Kaufmann et al. [9] compared character-by-character paradigm versus predictive text entry (PTE). 19 healthy students, naive to BCI, participated in the study. The participants were instructed to spell German sentences with a P300 based BCI with a $6 \times 6$ matrix that flashed row/column wise. Selecting a predictive word was performed, like selecting a character, by counting each highlighting of the desired word to spell. The participants also reported a forced choice questionnaire.

Their results (see Table 2) showed that the output characters per minute (OCM) was significantly higher in the PTE condition ($M = 20.6$ characters per minute, cpm) than the character-by-character condition ($M = 12.0$ cpm). In the character-by-character condition the performance ranged from 100% accuracy to 76%, with an average of 91.2%. This was not significantly higher than the PTE condition that ranged from 100% to 74%. All participants managed to handle the system on their own, without expert support, from the moment the electrodes were mounted. The questionnaire revealed that, on a scale on 1-4, the participants estimated in average 3.84 that they would be able to manage the BCI application on their own in the future and 3.74 that they could explain it to others. The questionnaires also revealed that all participants would prefer to use PTE instead of spelling character by character, when exposed to BCI-based spelling in the future. One participant reported higher effort in selecting complete words compared to single character.
Table 2. Results of Kaufmann et al. 2012 [9]

<table>
<thead>
<tr>
<th>Condition</th>
<th>OCM</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTE</td>
<td>20.6</td>
<td>76-100</td>
</tr>
<tr>
<td>CBC</td>
<td>12.0</td>
<td>74-100 (M = 91.2)</td>
</tr>
</tbody>
</table>

**Ryan et al. 2010** Ryan et al. [8] conducted a study where 24 able-bodied adults were instructed to accurately copy a sentence with a P300 based BCI. All the participants were naive to BCI use and they had no uncorrected visual impairments or any known cognitive impairments. The participants completed two sessions (one session in PTE condition and one character-by-character condition) on separate days within the period of one week. Both sessions started with a calibration and there was no stated training for the participants before the sessions. In both conditions the keyboard consisted of an 8 × 9 matrix, with the characters A-Z, 0-9, and other keyboard commands.

The results (see Table 3) did not show any significant difference in selections per minute between the character-by-character condition (M = 3.76) and the PTE condition (M = 3.71). However, the PTE condition reached an average OCM of 5.28, which is significantly higher than the 3.76 selections/minute in the character-by-character condition. The accuracy was higher in the character-by-character condition (M = 89.80%) than the PTE condition (M = 84.88%).

Table 3. Results of Ryan et al. 2010 [8]

<table>
<thead>
<tr>
<th>Condition</th>
<th>OCM</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTE</td>
<td>5.28</td>
<td>84.88</td>
</tr>
<tr>
<td>CBC</td>
<td>3.76</td>
<td>89.80</td>
</tr>
</tbody>
</table>

**Pan et al. 2013** Pan et al. [23] conducted an experiment to compare two different P300 based BCI spellers. The first interface used a 6 × 6 matrix of characters whereas the second had a region based interface, with six circles, each containing six characters/actions, arranged in a circle. 12 healthy subjects participated in this experiment. Ten of the participants were naive to BCI, the other two had limited prior experience from the system development. There was no stated training for the participants before the sessions. The results (see Table 4) shows that both conditions reached the same average speed, however, the region based interface reached a higher accuracy. Region based condition showed a higher accuracy (M = 93.49%) and OCM (M = 2.42 cpm) than single character (M = 89.32%, 2.42 cpm). The region based speller showed a greater increase of accuracy for each trial.
**Table 4.** Results of Pan et al. 2013 [23]

<table>
<thead>
<tr>
<th>Condition</th>
<th>OCM</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matrix based</td>
<td>2.19</td>
<td>89.32</td>
</tr>
<tr>
<td>Region based</td>
<td>2.42</td>
<td>93.49</td>
</tr>
</tbody>
</table>

**Ahi et al. 2011** Ahi et al. [24] evaluated two different P300 based PTE spellers, both with a 5×6 matrix of characters. The difference between them was the order of the characters. One had the letters arranged in alphabetic order whereas the other used a modified order based on the words in the dictionary. 14 able-bodied subjects, with little or no previous experience of BCI, participated in their study. The subjects spelled 15 four-letter words, once in each condition. The five first in each test were training words. The results (see Table 5) shows that the accuracy is increasing with the number of trials for each condition. With the conventional order and typing character-by-character, 5 trials per character was also best regarding OCM (M = 2.85). The highest OCM was achieved with 2 trials per character for both conventional order with PTE (M = 12.96) and Modified order with PTE (M = 19.70).

**Table 5.** Results of Ahi et al. 2011 [24]. The results from 3 trials was was excluded since they did not count the time between two consecutive selections

<table>
<thead>
<tr>
<th>Number of trials</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Conventional without PTE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy (%)</td>
<td>8.57</td>
<td>29.29</td>
<td>62.86</td>
<td>72.86</td>
</tr>
<tr>
<td>OCM</td>
<td>0</td>
<td>0</td>
<td>2.00</td>
<td>2.85</td>
</tr>
<tr>
<td><strong>Conventional with PTE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy (%)</td>
<td>39.29</td>
<td>67.14</td>
<td>95.00</td>
<td>95.71</td>
</tr>
<tr>
<td>OCM</td>
<td>0</td>
<td>12.96</td>
<td>11.40</td>
<td>9.26</td>
</tr>
<tr>
<td><strong>Modified with PTE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy (%)</td>
<td>53.57</td>
<td>87.14</td>
<td>95.00</td>
<td>96.43</td>
</tr>
<tr>
<td>OCM</td>
<td>4.92</td>
<td>19.70</td>
<td>11.40</td>
<td>9.32</td>
</tr>
</tbody>
</table>

**Park et al. 2012** Park et al. [21] compared a P300 speller with the frequently used random flashing pattern, to P300 spellers with flashing pattern decided from a partially observable Markov decision process (POMDP). Ten able-bodied students participated in the study. It was not clear whether they had any previous experience with BCIs. The participants did not get any more training than the calibration phase, where 20 trials randomly assigned a target letter. In their study, the subjects used a 6×6 matrix to spell “MACHINE LEARNING”. Their result (see Table 6) shows that both average accuracy and spelling speed was highest for the improved POMDP (96.88%, 4.74 cpm), followed by POMDP (92.50%, 4.31 cpm), and last random flashing pattern (85.63%, 3.18 cpm).
Table 6. Results of Park et al. 2012 [21]

<table>
<thead>
<tr>
<th>Flashing pattern</th>
<th>OCM</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>3.18</td>
<td>85.63</td>
</tr>
<tr>
<td>POMDP</td>
<td>4.31</td>
<td>92.50</td>
</tr>
<tr>
<td>Improved POMDP</td>
<td>4.74</td>
<td>96.88</td>
</tr>
</tbody>
</table>

Yin et al. 2015 Yin et al. [25] designed a SSVEP based speller with a 6 × 6 matrix of characters. Their goal was to enhance the performance of the speller, compared to previous SSVEP based BCIs, in terms of the number of selectable items on the interface, accuracy and speed. They used two different interfaces, one with fixed approach (the speller provides result once the optimal stimulus time is met) and one using a dynamic approach (the stimuli stop and give spelling result once the SSVEP response reaches a threshold). 11 healthy subjects participated in their study. The participant did not get more training than the offline session (approximately 50 min), where the proposed approach were validated. Four of the participants had previous experience of SSVEP based BCIs. The spelling speed using dynamic optimization (M = 16.98 cpm) was significantly faster than the fixed optimization (M = 12.27 cpm).

Chen et al. 2014 Chen et al. [14] constructed a SSVEP based speller with focus on high information transfer rate. They used a 5 × 9 matrix of characters and stimulus time of 2 and 3 seconds. 10 healthy subjects participated in their study. Two of the participant had previous experience of SSVEP BCI experiments, the others were naive to BCI experiments. The only training the participants had was during the offline sessions (approximately 2 × 5 minutes). The results (see Table 7) show that the 2 s time window resulted in higher average information transfer rate while the 3 s time window resulted in higher average accuracy. The authors therefore suggests that the trade-off between the time-window length/efficiency and effectiveness should be carefully considered.

Table 7. Results of Chen et al. 2014 [14]

<table>
<thead>
<tr>
<th>Stimulus (s)</th>
<th>OCM</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>17.94</td>
<td>84.1</td>
</tr>
<tr>
<td>3</td>
<td>14.67</td>
<td>90.2</td>
</tr>
</tbody>
</table>

D’Albis et al. 2012 D’Albis et al. [16] suggested a user interface for BCI spelling based on imagined movement with PTE. Three able-bodied subjects, naive to BCI, participated in the study. They did not get any more training than the offline session, that was used to train the algorithm. Two of the subjects used three classes (left hand, right hand, and both feet motor imagery), whereas one
subject had an additional class, namely both hands. All of the subjects used both three and four classes in the offline session. In the online test, the participants were instructed to perform seven repetitions of the phrase “what a wonderful day”. The result (see Table 8) shows that the subject using 4 classes performed better in both accuracy (M = 80%) and speed (M = 3 cpm).

<table>
<thead>
<tr>
<th># classes</th>
<th>OCM</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>3</td>
<td>80.20</td>
</tr>
<tr>
<td>3</td>
<td>2.35</td>
<td>70.15</td>
</tr>
</tbody>
</table>

Perdikis et al. 2014 Perdikis et al. [27] conducted an evaluation of a motor imagery hybrid speller, using both EEG and electromyographic signals. The authors compared three different interfaces, using either context awareness (CA), hybrid based ‘undo’ error handling, or both. In the study 16 subjects participated, of which six were severely disabled and ten were able-bodied. With all functions enabled, there was a completion rate of 100% for both user groups. Across all conditions and subjects the average accuracy was 94.2% (98.2% for the severely disabled test subjects). The results of the comparison (See Table 9) show that the CA condition reached the highest average speed of 1.76 cpm, but the lowest accuracy of 81.8%. CA + hybrid reached the highest average accuracy of 100%. 24 severely disabled subjects participated in the training phase, where approximately 50% reached 70% accuracy within nine sessions. None of the six motor disabled subjects had undergone more than five training sessions.

<table>
<thead>
<tr>
<th>Condition</th>
<th>OCM</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA</td>
<td>1.76</td>
<td>81.8</td>
</tr>
<tr>
<td>Hybrid</td>
<td>1.34</td>
<td>96.90</td>
</tr>
<tr>
<td>Hybrid + CA</td>
<td>1.68</td>
<td>100</td>
</tr>
</tbody>
</table>

Xu et al. 2014 Xu et al. [28] developed a parallel-BCI speller of four independent P300+SSVEP-B (P300 plus SSVEP blocking) with different flashing frequencies. Each sub-speller had a 3 × 3 matrix of characters. The interface had no support for error handling. 11 healthy subjects participated in the study. One subject had previous experience of a P300+SSVEP-B-based BCI system, the others were naïve to BCI. The only training the subjects got before the test was in the offline session, where they were required to pay attention to a specified character and count silently how many times it was extinguished. The subjects
spelled a short, predefined message, once with five trials for each character selection and once with only three trials for each selection. The results (see Table 10) show, like Ahi et al. [24], that more trials provides a higher accuracy whereas the OCM was higher with three trials (M = 11.45) compared to five trials (M = 8.80).

<table>
<thead>
<tr>
<th># trials</th>
<th>OCM</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>11.45</td>
<td>87.8</td>
</tr>
<tr>
<td>5</td>
<td>8.80</td>
<td>93.9</td>
</tr>
</tbody>
</table>

**Table 10. Results of Xu et al. 2014 [28]**

Yin et al. [29] also evaluated hybrid BCI spellers based on P300 and SSVEP. 12 healthy subjects participated in the study, where they were required to input 14 symbols in random order. The results (see Table 11) show that the hybrid speller performed much better in both Accuracy (M = 93.85%) and OCM (M = 10.95 cpm), than the P300 (68.65% and 4.73 cpm respectively.

<table>
<thead>
<tr>
<th>Condition</th>
<th>OCM</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P300</td>
<td>4.73</td>
<td>68.65</td>
</tr>
<tr>
<td>Hybrid</td>
<td>10.95</td>
<td>93.85</td>
</tr>
</tbody>
</table>

**Table 11. Results of Yin et al. 2013 [29]**

5 Discussion

The results show that the users with motor impairments see accuracy as the most important feature in a BCI, it should translate at least 90% of the commands correctly. Furthermore they want the BCI to provide a spelling speed of at least 15-19 characters per minute (cpm), however, a large group of people want a spelling speed of 20-24 cpm. Many of the solutions meet the users’ desires in accuracy, whereas only a few of the studied BCIs meet the users’ desires when it comes to efficiency.

The design solutions that best meet the users’ desires were P300 using Predictive text entry (PTE) by Kaufmann et al. [9], and Ahi et al. [24] using modified order and 2 trials per character. Yin et al. [25] developed a speller using dynamic optimization that reached the users’ desires in speed, however the average accuracy was not stated in this study, which makes it difficult to say whether this is a good approach. The SSVEP speller by Chen et al. [14] was also relatively
close to the users’ desires, however, when using 2 s stimuli the OCM reached the users’ desires whereas the accuracy was below the accepted limit. With the 3 s stimuli the accuracy reached the accepted limit, however, this time extension forced the OCM below the user’s desires.

This trade-off between accuracy and speed is common when designing BCIs. However, in the study of Park et al. it was shown that the choice of flashing pattern can improve both accuracy and speed, which suggests that the flashing pattern is an important aspect to consider in order to achieve both high speed and high accuracy. This is probably because the higher accuracy is achieved without changing the required time to make an action. This leads to fewer errors and thereby a higher speed since the user does not have to delete the mistake and type the correct character.

The time windows for stimuli causes a trade-off between accuracy and speed. The longer stimuli time the higher accuracy, at the cost of speed. The number of trials per selection causes a similar trade-off between accuracy and speed. More trials per selection provides a higher accuracy at the cost of speed. Since spelling speed showed the largest gap between the users’ desires and the performance of the BCIs, it might be a good idea to choose the shortest time window and lowest number of trials that provides an acceptable accuracy.

5.1 P300, SSVEP or imagined movement

Two out of five proposed P300 spellers could provide performance close to the users’ desires. This was only achieved with PTE, however, one of the P300 based spellers with PTE did not show satisfactory results in accuracy or OCM. The subjects were able to use the P300 spellers independently immediately or after one short training session. This suggests that the learnability matches the users’ desires to be able to use system independently within 5 training sessions.

Two out of two studied articles using SSVEP based spellers could provide performance close to the users’ wishes. One study used a dynamic optimization, which could increase or decrease the spelling speed dynamically. This approach resulted in a higher OCM than the fixed optimization, however no specific numbers for the accuracy was given. This makes it difficult to analyse the results in a good way. I feel that this could be a good approach to achieve a higher OCM if the algorithm to decide how to adjust the spelling speed is well implemented. The trade-off between speed and accuracy should be well considered to provide both satisfactory accuracy and speed. The studied SSVEP based speller with the highest OCM (M = 17.94) used a 5 × 9 matrix and 2 s stimulus time. However, this resulted in an accuracy below the desired limit.

Both spellers based on imagined movement were far from reaching the users’ desires regarding spelling speed, with the highest speed of 3 cpm. The designs with the highest OCM, did not reach the desired accuracy either. With a hybrid, using both EEG and electromyographic signals, the accuracy (M = 96.90) was higher than any of the other studies could achieve. Combining this with context awareness, improved the accuracy to 100%. However, this only resulted in a spelling speed of 1.34 cpm and 1.68 cpm respectively. In one study, only around
50% of the 24 severely motor disabled subjects that participated in the test session managed to reach a 70% accuracy after within nine training sessions. This means that 50% of the subjects did not reach a satisfying accuracy after the accepted training time. Based on these findings, BCI spellers using imagined movement can not meet the users’ desires.

5.2 character by character versus predictive text entry

Three articles compared P300 based BCI spellers with predictive text entry (PTE), to character-by-character spellers. The results show that PTE provides a higher spelling speed in all cases. One study suggests that the PTE does not reach as high accuracy as the character-by-character speller, whereas another study suggests the opposite. In the study of Ryan et al. [8] in 2010, the accuracy for character-by-character was much lower than the best results from Ahi et al. 2011 [24]. However, the results from the study by Ryan et al. were not satisfactory for the character-by-character condition either. This could indicate that the low results was caused by some other factor, for example the large amount of possible actions. Kaufmann et al. [9] found no significant difference in accuracy between the two conditions.

Furthermore, the subjects in one study stated that they would prefer a speller with PTE if they would be exposed to BCI-based spelling in the future. Based on these findings I would argue for using PTE in at least P300 based BCI spellers. For SSVEP and imagined movements, however, it is not possible to draw the conclusion whether PTE or character-by-character is most suitable, from the results in this study. This is because none of the studied articles used SSVEP together with PTE, and the only study that used imagined movement with PTE did not compare the results to character-by-character. This means that the bad performance in the speller using imagined movement with PTE could as well be caused by some other parameter.

5.3 Limitations

The study did not differ between different causes of the impairment and focused on people with intact cognitive abilities. However, some deceases cause impairments even in the cognitive abilities. For example, ALS patients often show cognitive impairment in working memory, attention response inhibition, naming and other functions [13]. These differences need to be accounted for when designing BCIs for motor impaired people.

This study focuses on letter based alphabets and takes little consideration to how the BCI application would work to spell messages with a different kind of alphabet, such as the Chinese alphabet with a lot more characters. This means that the solutions found here is not necessary applicable for all languages.

Furthermore, the time limitations of this study caused a limitation on the number of studied articles. The results could be more accurate if a larger number of articles were studied for P300, SSVEP and imagined movement combined with PTE, dynamic optimization, region based versus matrix based, etc. Moreover,
the results could have been affected by the mixture of subjects, with different
different motor impairments or none at all. In one study, the subjects with motor impair-
ment reached a higher average accuracy (98.2%) than the able-bodied subjects.
However, only 50% of the motor disabled subjects could reach an accuracy of
70% within the given training time.

6 Conclusions

Both P300 and SSVEP was found to be suitable with a performance close to
the users’ desires. Imagined movement however, did not perform good in effec-
tiveness, efficiency or training time. Combined with electromyographic signals
however, the accuracy was very satisfying. P300 based spellers could be improved
with predictive text entry to achieve better efficiency and user satisfaction. A
well chosen flashing pattern can also improve the P300 speller by increasing
both accuracy and speed. When selecting number of trials and stimulus time, it
is important to consider the trade-off between speed and accuracy in order to
provide satisfactory results in both. Since spelling speed showed the largest gap
between the users’ desires and the performance of the BCIs, the time window and
number of trials could be set to the shortest time window and lowest number of
trials that provides an acceptable accuracy. Dynamic optimization could provide
good results in both accuracy and speed, by adjusting the time required to type
one character, if this trade-off is carefully considered. Furthermore, when design-
ing a BCI for motor impaired people, it is important to account for differences
in vision, cognitive functions in working memory, etc.

References

1. M. B. Happ, “Interpretation of nonvocal behavior and the meaning of voicelessness
2. M. Marchetti and K. Priftis, “Braincomputer interfaces in amyotrophic lateral
   sclerosis: A metanalysis,” Clinical Neurophysiology, vol. 126, no. 6, pp. 155–1263,
   2014.
   users want? opinions and priorities of potential users with amyotrophic lateral
4. J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, H. P. Peckham,
   G. Schalk, E. Douchin, L. A. Quatrano, C. J. Robinson, and T. M. Vaughan,
   “Brain-computer interface technology: a review of the first international meeting,”
5. A. Kübler, A. Furdea, S. Halder, E. M. Hammer, F. Nijboer, and B. Kotchoubey,
   “A braincomputer interface controlled auditory event-related potential (p300)
   spelling system for locked-in patients,” Annals of the New York Academy of Sci-
Limitations and Possibilities of Using Eye-Tracking Technology with Smart Glasses

Anton Bäckström
Department of Applied Physics and Electronics
Umeå University, Sweden
anba0126@student.umu.se

Abstract. As wearables are expected to become the next generation of mobile devices, head-worn displays are becoming more affordable and available for commercial use. Although the emergence of Google Glass, the culmination of smart glasses as a successful device is yet to happen. Smart glasses, such as Google Glass, offer a lot of potential advantages in front other mobile devices, with ambition of providing a hands-free experience and always being accessible. In practice however, it is not as convenient. By being limited to the environment and the set of voice commands, furthermore the size and placement of the capacitive touch sensor, limits the range of possible inputs. Additionally, the placement can be physically demanding while doing regular tasks under an extended period of time. This paper examines the potential advantages, challenges and future possibilities of integrating eye-tracking technology with smart glasses, in order to achieve a more convenient, natural and hands-free experience, but also to consider the social aspects, based on a literature review with experts in this field of study, and a supervised user study to evaluate the social aspects of using such device. While potential advantages were found in eye-tracking, such as providing an increased bandwidth from the brain and the computer, and the human eye to be distinctly faster and naturally controlled, challenges were found in the difficulty of interpreting the user’s intentions, due to the eyes are always “on” and there are no natural way of making selections with them, which can easily lead to a gruelling experience of not being able to look anywhere without triggering another command. More sensors could be added in order to surpass the problem, by combining information from the eye and other lightweight inputs to help understand the user’s intentions and give the right response. Findings from the user study indicated that the conceptual solution of integrating eye-tracking with smart glasses, could be helpful and more socially accepted in way of being more discrete while interacting, but it does also introduce the temptation of using it discreetly, resulting in the feeling of being interfered while communicating with people who are wearing such device, knowing; they can always start and stop interacting with the device, without being fully able to notice it.
Analysis of text based information in virtual reality

Isak Dunér Lundberg
Department of Applied Physics and Electronics
Umeå University, Sweden
dv12i1r@cs.umu.se

Abstract. Virtual reality (VR) is a technology with a long history dating back to the 1960’s. It is a technology that let humans get immersed in virtual worlds and experience things that would not other wise be possible.

Today the field of VR have gained traction with the release of VR consumer products such as the Oculus Rift and HTC Vive. This time around many of the previous limitations of older hardware have been addressed and improved for a better user experience. Regardless of the use case for the VR devices there will almost always be a need for relaying information to the end user. A solution for this problem is to integrate text elements into the virtual world, but one have to consider the aspects of the users feeling of immersion and comfort when developing such interfaces.

The goal of this paper was to perform an analysis of placement of text on a two dimensional plain in a VR user interface (UI). The paper considered problems regarding readability of the text and how the text was integrated within the virtual world as well as the UI. The research question posed was as follows: To identify differences in text display methods in the interface of virtual reality applications.

What is the important differences that can arise when looking at the UI from an perspective of user comfort and the ability for the user to understand the relayed information? Understanding the information is regarded as how easy the text is to read without too much effort from the user.

To limit the scope of the paper, the focus was solely on the positioning of the text. This paper did not consider researching different fonts, sizes or colors applied to the text. Nor evaluating differences with backgrounds and environments where the user could be placed. Which possibly could have some effect on the users ability to read the text presented to them.

To evaluate the research question an heuristic evaluation method was used and applied on three different types of text integration that is common in current applications. The types asses was fixed text, the text is fixed to the users view, which means no matter where the user looks the text will follow. Partially fixed text, text locked to the environment in 3D space. The text will therefor only obstruct a partial part of the world and can be moved out of view if desired. Finally integrated text in the world as in text that is a part of the world objects not just text projected on top of the screen.
The result showed only minor differences between the display methods where the most obstructive and potentially immersion breaking method was using fixed text. None of the methods had any major differences in readability and performed about the same. The conclusion was that what affected the readability the most was how close the text was to the user and when the text was too far away they blurred into small pixels. This suggests that the screen resolution even with the new hardware today, is of vital concern for the ability to read text. In other words the different display methods considered in this paper had no major effect on the readability.
Table of Content

Full papers
Choice of fonts and price expectations
Anna Österlund

Evaluating Google material design with a focus on web application animations
Jakob Åberg

Presenting content to restless receivers: evaluating 360 video and interactive street view formats
Sandra Ingerhage

Using smartwatch as an aid for people with diabetes
Elin Hörfeldt

Examining the potential of smart glasses as an assistive technology for dementia patients
Karolina Jonzén

UX guidelines for accessibility in action shooter games
Anna Pääni

The effect the number of columns has on information perception on a web page
Petra Andersson

Paper vs. simple digital prototypes: differences in user feedback when testing
Hanna Söderström

Guidelines for designing context-aware fitness applications
Julia Westman

The role of feedback in educational applications for smart devices
Lena Lindberg

Push-notifying the end-user at the most appropriate time
Mattias Edin

Aesthetic value versus function in mobile applications
Robin Schulze
Feeling safe in autonomous vehicles
_Gustav Nordlander_

Using augmented reality for learning a second language
_Tove Barsk_

Biohacking: Evaluating the benefits of near field communication in the human body.
_Carl Holmertz_

Ditching The Hamburger Menu: An analysis of current menu design
_Anton Forsberg_

Interaction limitations on mobile devices related to screen size
_David Bergvik_

Evaluating the potential and limitations of Google Glass
_Joakim Zakrisson_

Design of brain-computer interface to ease everyday life for people with severe motor impairments: A metanalysis
_Ellinor Rönn_

**Work-in-progress**

Limitations and possibilities of using eye-tracking technology with smart glasses
_Anton Bäckström_

Analysis of text based information in virtual reality
_Isak Dunér Lundberg_
Choice of fonts and price expectations

Anna Österlund

Department of Applied Physics and Electronics
Umeå University, Sweden
id10aod@cs.umu.se

Abstract.
The market for on-line shopping has expanded as an increasing number of customers are connected to the Internet. Fonts are one of the factors that influence the visual impression of the description of a product. The aim of this study was to assess if the choice of fonts had any effect on price expectations of a product.
The participants in the study were randomly subjected to one of the fonts describing three different products (TV, jeans, and running shoes). The three most viewed Google fonts in the categories serif, sans serif and handwriting were studied (Slabo 27px, Open Sans, Indie Flower). 219 persons participated in the study distributed equally in each font group. There were no significant differences in price expectations between the font groups. Disregarding fonts, there were significant differences between men and women, and between the younger and the older participants in price estimation for all three products. The students estimated a significantly lower price for a TV than the employed group, and significantly lower price for the jeans than the retired group. In this study there were no differences reaching significance in price expectations for the three products described in three different fonts. However, in the age group 60-69 there was a strong tendency to estimate a lower price for a TV described in Indie Flower than in Slabo 27px. For a TV described in Slabo 27px, women tended to estimate a higher price than men did.

Keywords. font, typeface, price expectation, price estimation, typography

1 Introduction

The trading of goods has been a big part of human society since the start of civilization. But now, with the worldwide expansion of the Internet, the trading has found a new platform, online shops. In six years, from the year 2008 to 2014, the e-commerce nearly doubled in Sweden [1]. According to Statista [2], the number of Internet users worldwide that has purchased a product online is 40 percent, a number that is projected to keep growing.

With a growing online market, companies are starting to expand their businesses by offering the possibility of buying products online. Before, their customers were the ones that came in to the physical store. Now, with online sales,
a customer could be anyone with access to a computer connected to the Internet. However, with a bigger market comes increased competition. In 2014, almost 50 percent of the business owners in Sweden said that they are experiencing an increasing competition from foreign companies [1].

To be successful on the online market, it is important to design a site that attracts the customers. When asked which features played an important role in whether or not they would purchase a product from a web shop, 94 percent of the Swedish consumers said it was important that the information about the product was good and accurate [1]. 89 percent of the same consumers also said that it was important that it was evident what the total price would be.

There are several factors to consider while designing the product information on a website. Choosing an appropriate layout, a good picture of the product and writing a good product description are some factors that need to be considered. Even though the content of the text might be the most important aspect to make the product description good and accurate, the visual impression of the actual letters can add meaning to the text as well. Previous research has shown that font selection adds additional information to a text, like personality traits [3, 4] and trustworthiness [5].

The aim of this study was to determine how the choice of font affects the price expectation of a product. Research questions to be answered in the study are:

– Does the choice of font affect the perceived price of a product?
– What are the main differences between the serif, sans serif and handwriting fonts, with respect to price expectations?

2 Font typography

Since hundreds of years typography has developed a large amount of knowledge and rules. We use serified fonts to improve readability of body text, and for posters and advertising we use sans serif. These rules are not necessarily true but rather highly subjective, based on empirical knowledge. Modern research is questioning these statements [6].

The effect of fonts on readability has been examined in several studies [7–10]. However, even though good readability is an important part of conveying information through text, it is far from the only effect that comes with the choice of font. As readers, we are not simply passive recipients of information presented to us. In fact, our experiences, prior knowledge and expectations shape our interaction with the visual language [3].

A study by Lewis and Walker [11] reveal that when the appearance of the word is consistent with the meaning of the word, it is processed faster. When asked to press a button when the words "strong" and weak appeared on a screen, the study showed that people reacted faster if the word "strong" appeared in bold and if "weak" appeared in light weight. If the word was presented in the opposite font weight, "weak" in bold and "strong" in light weight, the reaction times were significantly slower.
While visual features of words influence the processing of a written message, they may also add their own semantic representations [12]. Studies show that different fonts are perceived as having different personality traits [3, 4]. Serif fonts can be perceived as elegant, interesting, distinct, formal, mature, practical, stable, charming, emotional, and beautiful [4, 13]. Sans serif fonts, on the other hand, can be perceived as smart, powerful, manly, upper-class, and loud [4]. Script fonts, such as handwriting, are perceived as happy, creative, feminine, casual, youthful, and cuddly [4].

Despite adding information to the text, the choice of font can also have an impact on the readers’ impression of the writer. When applying for a job, choosing an appropriate font for the resume affects whether or not the applicant gets called in for an interview [5], and thereby still being in the running of getting hired. Choosing an inappropriate font may cost the applicant the job.

In business, choosing an appropriate font for the brand is important since it affects how the consumers perceive it. Semantic associations linked to the type font used are transferred to the brand. A formal type font is associated with a luxurious brand, while using a casual type font results in a more casual brand perception [12]. Thus, the consumers take the type font characteristics and transfer them to brand personality perceptions [14].

If the appearance of the font contradicts the meaning of the writing, it is perceived as less appealing [15]. A study by Doyle and Bottomley [15] shows that subjects were more likely to choose a chocolate from a box where the brand name was written in an appropriate font. Thus, choosing an inappropriate font for a brand can make the consumer reject the product and, instead, choose a product from a competing brand.

The ability to affect the meaning, emotions, and convincing qualities attached to the content of a text, without the readers noticing, has implications for advertising, marketing, and persuasive literature [16].

3 Method

The aim of this study was to evaluate the perception of font type and how a product is perceived. A survey was thought to be most suitable for collecting information about the participants’ personal opinions and behavior.

The survey was conducted in Swedish during March 2016. It was designed to assess if the choice of fonts had any effect on the price expectations of a product. The participant was asked to read three different product descriptions, one at a time, and estimate what they believed the product would cost to buy. They were also asked about their age, sex, occupation, how often they research products online, and their online shopping experience and if they had ever bought one of the products described in the survey.

Three copies were made of the survey where the only difference between the copies was the font style of the text describing the products. The font styles used in the survey were serif (Slabo 27px), sans serif (Open sans) and handwriting font (Indie Flower).
Although three fonts were tested in this study, each participant was only subjected to one type of font and therefore only asked to answer one of the surveys. A shared first page was constructed to provide all participants with the same information and instructions before being subjected to one of the surveys. After reading the instructions the participants were randomly subjected to one of the three surveys.

Since the participants were not to focus on the font, but the product description as a whole, they were not told that the survey was conducted to evaluate fonts. The participants were only told that the survey concerned price estimation.

3.1 Choice of fonts

Table 1. Fonts used in the survey

| Most viewed Google fonts in January 2016 |
|------------------|------------------|------------------|
| Serif:           | Sans Serif:      | Handwriting:     |
| Slabo 27px       | Open Sans        | Indie Flower     |

The fonts used in the study (Table 1) are Slabo 27px, Open Sans and Indie Flower. These fonts are the most viewed Google fonts [17], January 2016, in the categories, Serif, Sans Serif and handwriting.

3.2 Choice of products

Since the study focused on products purchased online and the survey was conducted in Swedish, the products described in the survey were chosen from the top three product categories for online shopping in the Nordic countries. According to PostNords compilation of the e-commerce, e-barometern [1], in the Nordic countries for the year 2014, the top three product categories were home electronics, clothes/shoes and sporting goods.

Three products, one from each category, were chosen to be used in the survey. The products selected were a TV [18], jeans [19] and running shoes [20].

3.3 Survey

The first page provided the participant with instructions and information about the survey. The participants were informed that the purpose of the survey was to
identify how a written product description affects price estimation of products. Information about what the first three questions would entail was presented. The participants were told that they were to estimate the price of three different products after only reading the description of the products. They were also informed that the study was conducted as a part of a course at Umeå University.

After receiving information about the survey, the participants were presented with a button labeled "Do the survey". When the button was clicked they were randomly forwarded to one of the three surveys.

![Product 1](image.png)

**Fig. 1.** The first question of the survey, in Slabo 27px font [18]

The first three questions of the survey consisted of a product description, one from each product category, and a text field with a prompt for the participant to specify the estimated price of the product. The products were described in the following order, TV, jeans, running shoes. An example of the first question in the survey can be seen in figure 1, the product description of a TV. The option for the participant to enter the estimated price in free text was chosen to avoid affecting the answer by restricting it to an interval.

After completing the three questions about product description and price estimation some supplementary questions were presented. These questions were asked to get additional information about the participants, their sex, age, occupation, how often they research products online, their online shopping experience and if they had ever bought one of the products described in the survey.

The first three questions regarding price estimation were mandatory, while filling out the additional information was voluntary.
3.4 Statistical analysis

A quantitative approach of collecting data through a survey was chosen. Data were analyzed using an independent two sample t test, $\alpha = 0.05$, to detect differences between two samples. F-test was used to evaluate the variance of the samples.

4 Result

The surveys were performed by 219 persons. The participants were equally distributed with approximately 70 participants in each group (Table 2).

<table>
<thead>
<tr>
<th>Study population</th>
<th>Slabo 27px</th>
<th>Open Sans</th>
<th>Indie Flower</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>n = 73</td>
<td>n = 74</td>
<td>n = 72</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>29</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>Male</td>
<td>43</td>
<td>41</td>
<td>37</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-19</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20-29</td>
<td>25</td>
<td>17</td>
<td>20</td>
</tr>
<tr>
<td>30-39</td>
<td>21</td>
<td>31</td>
<td>21</td>
</tr>
<tr>
<td>40-49</td>
<td>8</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>50-59</td>
<td>6</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>60-69</td>
<td>11</td>
<td>12</td>
<td>17</td>
</tr>
<tr>
<td>70+</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Occupation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>56</td>
<td>56</td>
<td>54</td>
</tr>
<tr>
<td>Student</td>
<td>13</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>Retired</td>
<td>4</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

All ages were represented except for the age groups, 19 and younger, and 70 and older. The majority were employed and there were slightly more men in each group.
Most participants made research on the Internet before buying a product (Fig. 2a) and the majority was using the Internet for shopping (Fig. 2b). The question “Have you ever bought one of these products?” had to be excluded. Feedback from some of the participants indicated that they thought that only online purchases were to be considered.

Table 3. Mean resp. min-max values for the estimated price in SEK of the products evaluated.

<table>
<thead>
<tr>
<th></th>
<th>Slabo 27px</th>
<th>Open Sans</th>
<th>Indie Flower</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Price estimations (SEK)</strong></td>
<td>n = 73</td>
<td>n = 74</td>
<td>n = 72</td>
</tr>
<tr>
<td><strong>TV</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>7502 ± 4608</td>
<td>7652 ± 3126</td>
<td>6675 ± 4254</td>
</tr>
<tr>
<td>Min - Max</td>
<td>3000 - 35000</td>
<td>2000 - 18000</td>
<td>300 - 20000</td>
</tr>
<tr>
<td><strong>Jeans</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>691 ± 338</td>
<td>647 ± 298</td>
<td>665 ± 314</td>
</tr>
<tr>
<td>Min - Max</td>
<td>199 - 1700</td>
<td>40 - 1200</td>
<td>50 - 1500</td>
</tr>
<tr>
<td><strong>Running shoes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>1361 ± 571</td>
<td>1300 ± 399</td>
<td>1311 ± 635</td>
</tr>
<tr>
<td>Min - Max</td>
<td>349 - 4000</td>
<td>399 - 2500</td>
<td>200 - 5000</td>
</tr>
</tbody>
</table>

There were no significant differences between the price estimation of the three products, regarding the fonts. However, the age group 60-69 showed a tendency to estimate a higher price for a TV described in Slabo 27px than in Indie Flower (p=0.059). In the Slabo 27px group, the result indicates that women thought that the TV was more expensive than the men did (p=0.066).

Disregarding fonts, there were significant differences in the price estimation of products. The women estimated the price for a TV (p=0.046) and running shoes (p=0.027) higher than the men did. But for a pair of jeans it was the opposite (p=0.00003).
Price estimations for a TV and running shoes did not differ between the age groups. For jeans, however, there was a tendency to estimate a higher price with increasing age. For example, the age group 60-69 estimated the price for a pair of jeans higher than the age group 20-29 (p=0.0048).

The students estimated a significantly lower price for the TV than the employed group did (p=0.0158). The students also thought that a pair of jeans were less expensive than the retired group thought (p=0.0085).

5 Discussion

In this study, font choice did not significantly affect the price expectation of a product. However, it showed that age, sex and occupation are all factors that seemed to have an impact on how we estimate prices. Slightly more men responded to the survey but this did not appear to have any effect on the results.

Although the differences between the fonts did not reach significance there could be some possible reasons for this. The size of the study groups or the chosen products could have had an impact on the results.

Since fonts are known to add their own semantic representations [12], a formal or powerful font might be more appropriate than a casual font when describing a technical product, like a TV [4,13]. In this study, there was a tendency to estimate a lower price for a TV described in Indie Flower than in both Slabo 27px and Open Sans. No such tendency was apparent for a pair of jeans and running shoes.

The choice of products could have had an impact on the result since prior knowledge about the products may have affected the participants’ price estimations. If they already had a clear view regarding the price level of the product, it is possible that the appearance of the font did not have an impact on the results. Thus, choosing an unusual product, one which the participants had limited knowledge about, might have resulted in a different outcome.

The fact that there were differences between the sexes in the price expectations for all three products is not easy to explain. Even though it might be a controversial statement, one possibility could be that women are less interested in technical devices [21] and therefore might perceive them as more expensive. For the jeans, it might be the differences in shopping habits that explains the diverse results.

Age did not have an impact on the price estimations of a TV and running shoes. However, the age group 60-69 tended to estimate a lower price for a TV described in Indie Flower than in Slabo 27px. This could be due to a difference in how fonts are perceived by different generations, based on their empirical knowledge [6].

The younger participants thought that a pair of jeans was less expensive than the older ones thought. In fact, the price estimation increased with increasing age. A possible explanation for this could be the increasing market for jeans during the last century. As a result of this, the production cost has decreased.
and thereby also the market price. The generations might have different historical perspectives.

A higher standard of living seemed to play a role for the price expectations, both for a TV and a pair of jeans. The employed group estimated a significantly higher price for a TV than the students. The retired group thought that the jeans were more expensive than the students thought. This seems difficult to explain. It might be that students presumably buy cheaper TVs and jeans since their budget is tighter during the study period than after getting a job.

The study showed that most of the participants researched a product on the Internet before buying it, and the majority had bought a product online at least once a year. Today, most swedes have access to the Internet which has made online shopping possible. In fact, a survey conducted in the last quarter of 2014 by PostNord [1], showed that 76 percent of the swedes were buying products online.

5.1 Future work

There are some limitations in this study. Larger test groups might have led to differences between the chosen fonts, since some tendencies for difference in price estimation were observed between the groups. It is also possible that the products chosen were too familiar to the participants, and that previous knowledge about the products might have overpowered the effects of the font appearance. With a less known product the result might have been different. Further studies are needed to clarify this.

6 Conclusion

In conclusion, this study cannot show any significant differences in price estimations for a TV, a pair of jeans and running shoes, regarding the fonts chosen for product descriptions. However, the subgroup of participants over the age of 60 tended to estimate a lower price for a TV described in Indie Flower than the Slabo 27px. For a TV described in Slabo 27px, women showed a tendency to estimate a higher price than the men did.

References

Evaluating Google material design with a focus on web application animations

Jakob Åberg
Department of Applied Physics and Electronics
Umeå University
jaab0010@student.umu.se

Abstract. With an expanding access of the Internet and a growing use of dynamic content on web-sites, web-animation has become a subject for discussion. There are different opinions on how animations should be implemented to aid users instead of distracting them. This paper analyses the use of web-animations and specifically the animations given in the guidelines suggested by Google material design. The usability of the animation guidelines in Google material design was tested, and the effects they had on information seeking performance. The findings of this paper were based on a SUS-questionnaire and two web applications; one that follows the animations suggested by Google (application A), and another that implements the animations that Google advice against (application B). 20 persons participated in the study, 10 on each application. The participants of application A seemed to execute all tasks faster with an exception of the first task. The scores of the SUS-questionnaire indicates that application A was perceived as over average, in regards of usability, while application B was perceived as passable. The results of this paper cannot show any significant differences in time efficiency between the two applications. However, the scores received from this study suggests that there might be a link between usability and the animations suggested by Google.

1 Introduction

Ever since the beginning of civilization people have been fascinated by moving images and of animation. The desire to give motion and life to pictures has played a big part in the developments of culture, science and technology. Now animation is widely used in applications and on websites, both to capture viewer attention but also to enhance usability [1]. The use of visualized information has been increased with approximately 10000 percent on the world wide web since 2007 [2], a percentage that just keeps on growing.

With an expanding access of the Internet and a growing use of dynamic content on web-sites, web-animation has become a subject for discussion and studies have been conducted on the topic [1][3][4][5]. Earlier
work by Zhang [3] claims that web-animations interfere with users primary tasks when using web-pages. This happens when animation as a secondary stimulus deteriorates the information seeking performance, and when the animation is similar to the task [3]. On the other hand, there is also a study that suggests that motion design and animation enriches the web-page and makes it comfortable and easy to use [1]. In an article from 2014 [6] the authors state that web-animation and dynamic web-sites make users happy, and can surprise in a positive way. Mori, Paterno and Furci [6] found that it is very important for the animations to facilitate interactions, without being perceived as annoying.

To help developers and designers implement animations, companies such as Google, Apple and Facebook release design guidelines to follow. Several studies have been conducted on topics such as animation in web applications [1,3,4,5] but none have focused on evaluating the front-runners of technology and interfaces today. This topic deserves our attention as there is a tendency to put trust in technology organizations [7].

The aim of this study was to get a better understanding of how people react towards animation in web applications. Furthermore, the objective was to evaluate Google material design with a focus on web application animation. Research questions to be answered in this study are:

- What is the correlation between usability and the Google material design animation guidelines?
- What is the difference in time efficiency regarding information seeking performance using the animations suggested by Google and the animations they advice against?

2 Google material design

In 2014, Google released their new design language called material design. The main goal was to create a visual language that combines the classical principles of successful design with the innovation and possibility of technology and science [8]. Google also wanted to make it easier for developers and designers to build beautiful and consistent applications that allows for a unified experience across all platforms [8].

Just one year after the big release, Google had implemented the visual language on most android platforms and on all the Google Play-branded applications, such as Google Mail, Drive, Sheets, Docs, Maps and Youtube [9]. Google’s material design is not only reserved for mobile applications.
there are also a number of desktop web-interfaces that has incorporated it as well, such as Google Drive, Mail, Docs and much more. The language is consistently updated and maintained [10].

The material design framework is constructed by three main principles: (1) Material is a metaphor; the unifying theory of a rationalized space and a system of motion taken from the real world. (2) Bold, graphic, intentional; typography, space, colors and use of imagery is not only pleasing to the eye. Implemented with the right approach it can create hierarchy, meaning and focus in a beautiful way. (3) Motion provides meaning; the idea that motion helps the user navigate and to understand how elements belong together [8]. This paper will mainly discuss the latter of these three principles, focusing on animation.

2.1 Animations

In this study five of the material design animation guidelines will be tested. The animation guidelines chosen are the ones that often relates to navigation elements within a website and the elements that take up the most visible space in the browser. Thus, the ones that is believed to give the most tangible results in this paper. The figures in this section are all based on figures given in Google material design [8].

2.1.1 Point of origin. In the real world, feedback is received in such a way that output origins from the input; output elements expands from input elements. This is the best way to do it in the digital world as well (Fig. 1) [8].

2.1.2 Visual Continuity. Transitions between two visual elements should be smooth, clear and non-distractive [8]. Irrelevant content should be removed and newly generated items should be highlighted. Animations should direct user attention, and create connections between transitioning states [8]. If movement is used it is important that it is precise.

2.1.3 Hierarchical timing. Consider the timing and the movement of elements. Motion should follow the information hierarchy; the most important content should be highlighted [8]. This is easily done by creating an animation path through the content (Fig. 2). This makes it easier for the user to know where to put their attention [8].
Fig. 1. a) The best way to animate material response: output should origin from input [8]. b) How not to animate: output origins from the middle of the screen from no particular source [8].

Fig. 2. a) How to animate hierarchical timing [8]. A path through the content helps the user to know where to first put their attention [8]. The most important content should come first. b) How not to animate motion; independently and without a path [8].

2.1.4 Consistent choreography. Transitioning items should behave in the same manner. The paths that the elements follow should be orderly and create meaning. If items move in different ways (Fig. 3b), it can be distracting for the user [8].
2.1.5 Delightful details. With this guideline Google \footnote{8} wants to acknowledge the importance of creating applications that are fun by adding small transitional details, like turning a hamburger icon into a go-back-arrow when clicked. In this way, Google believes that an application can go from a good to a superb one \footnote{8}. There is a big risk with this since it is hard to draw a line where animations and transitions become too distracting.

3 Method

To analyze the animation guidelines given in Google material design, two simple web applications were created; application A - based on the animations suggested by Google, and application B - based on the animations that Google advice against. They were both built with the help of Angular Material \footnote{11} where CSS3 and JavaScript animations were used.

Application A and B were both plain in terms of color and general design so that the test participants focused primarily on animations. The applications contained four pages, each page with a task for the participants to solve in the fastest way possible. To measure the effects of the animations on information seeking performance, the effectiveness of
the test participants was evaluated as they were going to identify target strings from other strings [3]. Random strings were chosen in this study in order to eliminate any automatic processing of familiar target strings [3]. The process of identifying strings is one of the most common information seeking tasks on a website [12] and thus, very relevant to this study. Application A and B were tested individually on ten test subjects each, with two pilot tests conducted per application.

To analyze how effective the test subjects were, the time was taken for each task. The tests were conducted in Swedish, since all the subjects understood written and spoken Swedish fluently.

3.1 Application set up

The introduction page provided the participants with general instructions about the test. They were informed that the objective of the test was to measure the usability of a website with different animations. The participants were also told that they were going to perform a task for each page on the website and afterwards answer a short questionnaire. A button was then displayed, with the text ”Begin test”. When this button was clicked the users were forwarded to one of the applications.

Before the participants were directed to each page in the application, a brief explanation was given about that specific page and the task they were going to solve. The only difference in the pages of application A and B was the animations.

3.1.1 First page. The participants were told that this page had a button labeled ”Show”. They were informed that if this button was clicked a grid of squares was going to be displayed, each square with a random string of letters. The participants were then asked to find the string ”pagd” as fast as possible and click on it. In application A the grid animated in such a way that it followed a path (Fig. 2a), in application B the squares in the grid animated independently and without a path (Fig. 2b). The page of both applications are based on the guideline Hierarchical timing.

3.1.2 Second page. The second page of the application is based on the guideline Consistent choreography. The participants were told that a grid of squares was going to be displayed and that each square contained a random string of letters. The participants were asked to find the string ”mui” as fast as possible and then click on it. In application A all the
squares faded in from bottom and up, and in application B the squares faded in from different directions without a clear choreography.

3.1.3 Third page. In this page of the application the focus was on the animation guideline Point of origin. The participants were told that they were going to see a page with a button, when this button was clicked a container would be displayed. The participants were then asked to find a dot as quick as possible and then click on it. In application A the container expands from the button that is pressed, in application B the container expands from the bottom of the page without a clear point of origin.

3.1.4 Fourth page. The last page focused primarily on Visual continuity and Delightful details. The participants were told that the page was going to contain six cards, each card with a number displayed on it. They were asked to click on these cards, to find the random string of "opasd" as quick as possible. In application A the card was flipped over and highlighted in the middle of the screen to show the test subjects the string of letters. In application B the card did not move, when clicked it just displayed the other side of the card.

3.2 Questionnaire

After finishing the tasks, a questionnaire was conducted to get a sense of how the participants perceived the usability of the application. The method chosen in this study was The System Usability Scale (SUS) [13,14]. SUS is a likert scale consisting of ten statements, each statement can be answered on a scale ranging from "Strongly Disagree" to "Strongly Agree" [14]. The statements that were given in the questionnaire are shown in Table 1.

Final scores from the SUS can range from 0-100. A score over 70 means that a product is passable, a score over 80 means that a product is over average and a really good product scores over 90 [14].

Information about the test participant’s sex, age and occupation were also gathered in this questionnaire. All questions and statements were mandatory.

3.3 Statistical analysis

To analyze the data both from the time efficiency test and the questionnaire, a two sample t test was used to detect differences in application A and B.
Table 1. SUS statements given to the test participants in the questionnaire. The statements are changed a bit from the SUS-standard to fit the study conducted.

<table>
<thead>
<tr>
<th>SUS statements given in the questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I think that I would like to use the animations in this test frequently.</td>
</tr>
<tr>
<td>2. I found the animations in this test unnecessarily complex.</td>
</tr>
<tr>
<td>3. I thought the animations in this test were easy to use.</td>
</tr>
<tr>
<td>4. I think that I would need the support of a technical person to be able to use an application with these animations.</td>
</tr>
<tr>
<td>5. I found the various animations in this test well integrated.</td>
</tr>
<tr>
<td>6. I thought there was too much inconsistency in the animations of this test.</td>
</tr>
<tr>
<td>7. I would imagine that most people would learn to understand the animations in this test very quickly.</td>
</tr>
<tr>
<td>8. I found the animations in this test very awkward to use.</td>
</tr>
<tr>
<td>9. I felt very confident navigating through this test.</td>
</tr>
<tr>
<td>10. I needed to know a lot in order to use the animations in this test.</td>
</tr>
</tbody>
</table>

4 Result

The study was conducted by 20 people, 10 women and 10 men with an average age of 29 years (ranging from 20 - 77). The participants were equally distributed with 10 test subjects on each application. The time was taken for each participant as they tried to solve the tasks. The mean time to solve each task are shown in Table 2.

There was not a significant difference in how fast the test participants of application A solved the tasks compared to how fast the participants of application B solved them ($p=0.66$). In task 2, 3 and 4 the participants of application A were faster and in task 1 the participants of application B were faster, but not significantly ($p_1=0.63$, $p_2=0.23$, $p_3=0.49$, $p_4=0.90$). The biggest time difference was in task 2 where participants of application A were 3.217 seconds faster than participants of application B ($p_2=0.23$). Participants of application A solved all the tasks 1.067 seconds faster than B.

The results from the SUS-questionnaire are given in Table 3. There were no significant difference in perceived usability of the two applications ($p=0.14$). The SUS Scores were slightly higher for application A than
Table 2. Mean time for the participants to solve each task in the test. The tasks were solved slightly faster in Application A than B.

<table>
<thead>
<tr>
<th>Task</th>
<th>Application A</th>
<th>Application B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8.543</td>
<td>7.053</td>
</tr>
<tr>
<td>2</td>
<td>4.356</td>
<td>7.573</td>
</tr>
<tr>
<td>3</td>
<td>5.535</td>
<td>7.319</td>
</tr>
<tr>
<td>4</td>
<td>12.372</td>
<td>13.132</td>
</tr>
<tr>
<td>Total</td>
<td>7.702</td>
<td>8.769</td>
</tr>
</tbody>
</table>

Table 3. Calculated SUS Scores for each participant in application A and B.

<table>
<thead>
<tr>
<th>SUS Scores</th>
<th>Application A</th>
<th>Application B</th>
</tr>
</thead>
<tbody>
<tr>
<td>87.5</td>
<td>75.0</td>
<td></td>
</tr>
<tr>
<td>70.0</td>
<td>65.0</td>
<td></td>
</tr>
<tr>
<td>70.0</td>
<td>85.0</td>
<td></td>
</tr>
<tr>
<td>95.0</td>
<td>67.5</td>
<td></td>
</tr>
<tr>
<td>80.0</td>
<td>77.5</td>
<td></td>
</tr>
<tr>
<td>82.5</td>
<td>62.5</td>
<td></td>
</tr>
<tr>
<td>87.5</td>
<td>62.5</td>
<td></td>
</tr>
<tr>
<td>60.0</td>
<td>52.5</td>
<td></td>
</tr>
<tr>
<td>80.0</td>
<td>70.0</td>
<td></td>
</tr>
<tr>
<td>65.0</td>
<td>85.5</td>
<td></td>
</tr>
<tr>
<td><strong>Average:</strong> 77.8</td>
<td><strong>Average:</strong> 70.3</td>
<td></td>
</tr>
</tbody>
</table>

B, where application A scored an average of 77.8 and application B an average of 70.3.

5 Discussion

In this study, a significant difference in time efficiency could not be proven between the two applications. However, we can see a variation in time in
how fast all the tasks were solved. In task 1, the participants of application B were 1.490 seconds faster than the participants of application A. This could mean that **Hierarchical timing** does not have any effect when it comes to information seeking performance in web-animations. One thing that was observed during test runs was that the participants of application A were much faster at placing the mouse in the first square of the grid than the participants of application B. A possible explanation for this is that **Hierarchical timing** actually makes it easier for the user to know where to focus.

The biggest time difference can be seen in task 2 where the participants of application B were more than 3 seconds slower than the participants of application A. This could mean that Google \[8\] is right in their statement that random motion in applications and on web-sites can be distracting for the users. The result indicates that when all elements follow a certain movement it might actually aid user understanding of the application, instead of getting the users confused on where to put attention.

The result from task 3 and 4 did not show any statistical evidence that there is a difference in time efficiency either. Google’s main objective with the guideline **Point of origin** might just be to help developers create applications that have high connectivity to the real world, focusing on the usability aspects of applications instead of their time efficiency. The same discussion might apply to task 4 and the guidelines **Delightful details** and **Visual continuity**. The guidelines may not have any effect when it comes to information seeking performance, but rather a higher emotional value. It was clear from the study that the participants of application A enjoyed task 4 a lot more than the subjects of application B.

Although there was not a significant difference in how the participants of the two applications perceived usability, there could be some logical reasons for this. Two of the test participants stated that they had not noticed the animations in the applications. This could be due to the fact that some users can block the distraction or the events caused by animation \[3\]. There is a possibility that some of the test participants were so focused on the tasks that they did not notice the animations. If this is the case, there is a chance that the participants did not notice the usability of the animations and thus, the result from the SUS-questionnaire can not be seen as reliable.
6 Future work

There are some limitations in this study. Since some of the participants stated that they had not noticed the animations in the tasks, there is a possibility that the study was conducted in a misleading way. It may have been the wrong approach to measure perceived usability in a test that also measures time efficiency on information seeking performance. A better approach for this study could have been to build two applications; one task-application that measures the time efficiency and another application that the participants could navigate through. This may have separated the two factors of the test and the participants might have put equal focus on both the usability and to solve the tasks as quick as possible.

The study could be extended in a lot of ways. A larger amount of test participants might have led to bigger differences in time efficiency and perceived usability. A test with more pages and tasks could also result in more thorough conclusions about the guidelines. Another unexamined factor is web-pages with multiple animations, which are commonly used in web applications today.

7 Conclusions

This study cannot show any significant differences in time efficiency in information seeking performance as well as usability when it comes to the Google material design guidelines. However, the scores received from the SUS-questionnaire suggests that their might be a link between usability and the animations suggested by Google. This research provides a base for future investigations about design guidelines in web-environments.

References

8. Google, “Google material design.” [https://www.google.com/design/spec/material-design](https://www.google.com/design/spec/material-design) [Online; accessed 15-February-2016].
10. M. Duarte, “Google i/o 2015.” [https://www.youtube.com/watch?v=8UicJ0SxBwA](https://www.youtube.com/watch?v=8UicJ0SxBwA) [Online; accessed 29-February-2016].
Presenting Content to Restless Receivers:
Evaluating 360 Video and Interactive Street View Formats

Sandra Ingerhage
Department of Applied Physics and Electronics
Umeå University, Sweden
sain0012@student.umu.se

Abstract. Digital lifestyles with increased access to information and needs for extracting the gist have been found to change how people focus their attention. This paper evaluates the 360 video and the interactive street view formats, with the aim to identify advantages and disadvantages of their ability to capture the attention of people with digital lifestyles. The formats were evaluated by letting users consume information through them, and the subjective experience of the users' focus was primarily considered. Findings indicated that the rich media and large scope of stimuli provided by the formats is an advantage, but that it also is important that the rich media is cohesively arranged and contributes to getting the message through, or the rich media may instead have a negative effect on the focus. Furthermore both advantages and disadvantages were found in the tempo, scene switches and navigation of the formats. In conclusion both formats have potential to capture the attention of people with digital lifestyles, but they also both come with challenges.

1 Introduction

In the information age there is always something new to take part of and discover. Increasing digitalization and access to information in everyday environments have changed how people focus and allocate attention. A decade ago people reported that a shift from reading printed documents to reading electronic documents on digital devices changed their reading behavior [1]. When reading on a digital device, more time was spent on selected reading and less time was spent on in-depth reading. Hence, the behavior changed from focusing prolonged attention following a text to alternating focus between different parts of the text and extracting key sections.

Today, these changes seem to have continued to develop on the same path. Evolving technology does not only let people do more things digitally, it also encourages people to take part of more information since it is easily accessed through always present media streams. Digital lifestyles are characterized by a high volume of media consumption as well as concurrent consumption of several media streams. Studies have shown that people with digital lifestyles have harder
to focus on the same task for longer periods, especially in less interactive environments [2, 3]. They strive to find the highlights and are driven by the thrill of finding the next one, which is always certain to be found in the streams of awaiting information. Concurrent consumption of several media streams, referred to as media multitasking [3–5], also decrease the ability to filter out irrelevant information and maintain focus on a single primary task [3]. Although people with digital lifestyles are more easily distracted and have problems to maintain focus, they have at the same time an increased ability to faster process and assimilate information [2].

Digital lifestyles are common today, especially among the young generation of teenagers who reflect the future to come [6, 7]. If digital lifestyles will continue to gain ground as the typical lifestyle, it is important to acknowledge the change in attentional behavior that comes with it. Having problems to maintain prolonged focus but possess skills to faster assimilate information, people seek to sort out the gist, bring forth the most important conclusions and move on to the next thing [2]. This should be taken into consideration when it comes to reaching out with information to a receiver, it is an advantage to keep it short, clear and get to the point. In many situations this is a good strategy to let receivers handle heavy information loads from multiple channels, but some information may be hard to shorten down and compress without losing essential parts and may need some more time to be comprehended and understood. The change in attentional behavior brings both challenges and opportunities to reach out with information to receivers. While ever evolving technology is contributing to multitasking tendencies and decreased ability to maintain attention, it may also be utilized to enhance these behaviors in order to convey information. Two interesting ways of presenting information in this sense are the 360 video format and what is referred to in this paper as the interactive street view format. These formats are interesting since they provide a rich media experience and have a way of keeping receivers active.

This study aims to evaluate the ability of the 360 video format and the interactive street view format to enhance the attentional behavior of people with digital lifestyles. The objective is to identify advantages and disadvantages related to the characteristics of these formats when it comes to keeping the focus for a longer period of time.

2 Background

The 360 video format and the interactive street view format have similarities but also disparities which makes it interesting to analyze them in relation to each other. The formats and their characteristics will be described further in this section, after a consideration of earlier studies and concepts of human attention.
2.1 Human attention

Human attention is a wide area of research, and this section will attempt to bring forth some pieces that are especially interesting in relation to digital lifestyles, and that can be used as a reference ground for analysis of the formats.

2.1.1 Various types of attention

There are different approaches to describe cognitive processes related to human attention. One model proposed by Sohlberg and Mateer [8] distinguish five categories that describe different types of attention.

Focused attention is the ability to respond to stimuli, which may be a visual, auditory or tactile or any other type of stimuli.

Sustained attention is the ability to maintain focus and respond consistently while engaging in a continuous prolonged activity.

Selective attention refers to the ability to stay focused on a task while exposed to concurrent or competing stimuli.

Alternating attention is the ability to shift focus between tasks that demands different cognitive skills.

Divided attention refers to the ability to simultaneously respond to two or more task demands.

These different types can describe the attentional demands on a person when taking part of information that requires prolonged attention. The person essentially needs to be able to focus attention to assimilate any information, and need to maintain sustained attention to get to the end of the story. It is also crucial to be able to allocate attention to the selected task and not get distracted by other stimulus. Depending on the content and how it is presented, the person may also need to divide or alternate attention between different parts of the information. Hence, all these aspects of attention is important to consider when trying to convey information that needs to keep the receiver’s attention for a longer period.

2.1.2 Factors that affect attentional focus

Wide attentional filters. One branch in research on attentional behavior and digital lifestyles focus on selective attention. Among such work are studies that have shown that people with digital lifestyles tend to have a wide attentional focus and take in much of the information available [3, 5]. This has been described as breadth-biased attention behavior [3, 9], meaning that larger amounts of information is processed rather than that a particular piece of information is selected and processed in detail. It also means that all stimuli is treated as equally important and are equally attended to [9]. These findings suggest that attention is best captured by larger amounts of stimulus that can catch the breadth of the attentional filter.

Perceptual and cognitive load. There is also research that consider perceptual processing and its relation to cognitive processing and workload. With
Load Theory, Lavie, Hirst, De Fockert and Viding [10] propose that the ability to maintain attention increases when a task involves high perceptual load, i.e. when the level of perceptual stimuli is high. It also proposes that it decreases when a task involves high cognitive workload, i.e. in situations when the processing requires much effort and cognitive resources. Hence, this means that increasing the perceptual stimuli in a task should increase the ability to focus on the task, as long as it does not also imply more cognitive load. It has also been shown that increased perceptual load in tasks can reduce the processing of task-irrelevant information [11], and this could be considered as a strategy to help people with digital lifestyles to keep away from distractions.

**Multitasking.** Research do also concern attention in relation to multitasking, and multitasking by dividing and switching attention within and between tasks is a core characteristic of people with digital lifestyles. A model that takes a broad approach describing different aspects of multitasking is the Multiple Resource Theory proposed by Wickens [12]. The model is based on four dimensions of cognitive resources, described below.

The Stages of processing dimension implies that perceptual and cognitive processing uses different resources than selection and execution of action.

The Codes of processing dimension implies that spatial processing uses different resources than linguistic processing.

The Modalities dimension implies that auditory perception uses different resources than visual perception.

The Visual channels dimension implies that focal and ambient vision is distinguished visual channels.

The point with these dimensions is that they show how cognitive resources best can be utilized, and describe what kind of tasks that can be done concurrently with good performance. The better tasks use different resources along each of the dimensions, the easier they will be to perform concurrently. The more tasks compete for different resources along the dimensions, the harder they will be to perform. It is for example easier to focus on listening on a verbal instruction and looking at a picture, than it is to listen to a verbal instruction and reading a text. This is because listening to the instruction and reading the text requires the same resources from the coding dimension, i.e. linguistic coding. This as opposed to listening to the instruction and looking at a picture, which requires different resources from the coding dimension, i.e. linguistic and spatial coding respectively. Hence, considering these dimensions when designing how tasks are to be performed can help exploit cognitive resources in an efficient way and avoid exhausting performers attending to a large amount of information concurrently.

### 2.2 Media formats

There are different kinds of information and content that requires maintained focus for a longer period of time to be assimilated, and different types of content providers also experiment with strategies and formats to be able to reach out to their target groups. Hence, the 360 video and interactive street view formats are just two among many other ways to present information. What makes
these formats interesting is that they provide the receiver with a rich and large amount of stimuli through various media types, which should suit people with digital lifestyles with media multitasking tendencies. Having this in common, the formats do also differ on other parts which makes them interesting to examine together. Following is a description of the two formats.

2.2.1 360 video

A format of presenting content that is increasing in digital contexts is video [13, 14]. The 360 video can be described as an extension of the video format, which adds a new feature. Using technology that captures the whole 360 degrees perspective of scenes, 360 videos can then also be viewed in 360 degree perspectives. As opposed to "common" video where only one perspective is captured and viewed, the one directed by the person filming, 360 video lets the viewer decide the perspective to be shown by controlling the direction of the view using simple interactions. Figure 1 shows what it may look like when shifting from one perspective to another. 360 video is also gaining usage and big media platforms, such as YouTube [15] and Facebook [16], initiated support for the format in 2015. Another example is news agencies that have started to experiment with 360 videos as a format of presenting news stories [17–19].

2.2.2 Interactive street view

The approach of the interactive street view format is similar to that of the 360 video. Material for this format is also generated by capturing 360 degrees perspectives of scenes, but in still images. These are then put together to a resulting 3D world that can be explored by "walking around" to different places in the world. Places can be explored by "looking around" using simple interactions in the same way as in the 360 video, which is illustrated in Figure 1. One extensive example of how this technique has been used is the Google street view application [20], which lets the viewer wander through many parts of the planet. It is not focused on a concentrated content in the same way and the world is huge, but it is an example that illustrates the "walking around" feature. Hence, a difference from the 360 video format is that there are several paths to take through the story, and it is the viewer who decides which paths to take between places. The designer of the content does lay out the paths through the content, but the viewer decides how to move along them. In addition to the "walking around" feature, places can also be explored by interacting with things present in places and a speaker voice and/or sound can also give additional information about a place when stopping at one.

3 Method

To identify advantages and disadvantages of the formats' ability to keep people with digital lifestyles focused, the study was conducted by letting users consume
Fig. 1. Two images illustrating the "look around" feature of the 360 video and interactive street view formats, the images show the same place but from slightly different perspectives. The images are from an interactive street view reportage [21], and the visible round, black symbols (which are not present in the 360 video format) are used to move from one place to another or explore things in the view.

information through the formats. This was done by primarily consider the users’ subjective experience of their focus. As a complement objective data that gave an indication of how the content reached out to the users was also gathered, by considering what the users could recall of the content. The material selected for testing the formats were two news reportages about Burma, a 360 video [22] and a street view reportage [21] both made by professional journalists wanting to communicate the situation in the country. The 360 video is about five minutes long and the street view reportage was delimited to a section that also took about five minutes to go through. Focusing on the same thing for five minutes should be a quite long time for people with digital lifestyles [2].

3.1 Setup and structure

A test leader guided the user through the test session, and one user at the time viewed the reportages in a calm environment using a laptop. Hence, the full focus was on the reportages and the study did not try to take into consideration external distraction that may occur in everyday life, but concentrated on the characteristics of the formats.

The study was divided into four parts, where the first two parts consisted of the user testing each of the two formats, the third part consisted of the user answering questions about their media consumption and the fourth part consisted of a short interview.

The first two parts evaluating the formats were structured in the same way. The user was first given an introduction to the reportage, both shortly to the subject of the reportage and how to use and navigate the format. For the 360 video reportage the user was told to use the keyboard and mouse to "look around" as much as the user felt like, and for the street view reportage the user was told to freely explore the reportage and tell the test leader when feeling done with viewing the reportage. After taking part of the reportage, the user answered questions about what had been said in the reportages and about the experienced ability to focus on the reportages.
3.2 Measurements

To measure the user’s focus the user’s subjective experience of the focus was primarily considered, and to objectively see that the user also did attend to the content the user also answered questions about the content. Since the study focus on people with digital lifestyles the user also estimated their media usage.

3.2.1 The user’s subjective experience

The user’s subjective experience of the formats was identified by the interview at the end of the test session. The interview was semi structured and grounded in the questions (1) How did you experience your focus during the reportages? (2) Would you prefer one of the reportages in front of the other?

The subjective experience was also identified by letting the user estimate their perceived workload when viewing the reportages. This part was constructed with inspiration from the lighter version of the NASA Task Load Index [23], which lets the user rank different aspects of how heavy the workload is when doing a task, such as level of effort and frustration. Some modifications of the questions the index uses were done in order to make them more explanatory in relation to the task of viewing the reportages. A complementary question about the user’s interest in the content of the reportage was also added, to make it possible to see the potential effect of this on the perceived task load. The resulting questions are stated with the results in the following section. The user answered these questions directly after viewing the reportage.

3.2.2 Comprehension of the content

To measure how much the user assimilated of the information presented in the reportages, the user answered questions about things that had been said in the reportages. The questions were formulated as multiple choice questions and the user also ranked the confidence in the answers, this to get an indication of the degree of guessing. To focus on the user’s attention questions targeted facts stated, and not things that needed to be interpreted and deeper reflected upon. There were four questions for each of the formats, and the user answered these directly after viewing the reportage.

3.2.3 Media consumption

To get an idea of the user’s media consumption habits, a scaled down version of the Media Multitasking Index developed by Ophir, Nass and Wagner [3] was used. The user estimated the amount of usage of different types of media, and how often several types of media was used at the same time.
4 Results

Eleven users participated in the study, which was conducted in 2016 in Sweden. The study was conducted in Swedish, and the results have been translated to English.

4.1 The users’ subjective experience

The following list describes experiences of the formats expressed during the interviews. When asked about which format the users preferred, four users did not want to choose one in front of the other, four users chose the 360 video format and three users chose the interactive street view format.

Experiences of the 360 video reportage

- It was nice to be guided through the video and at the same time be able to "look around"
- It was hard to focus when the tempo in the video sometimes felt stressful
- Scene switches was sometimes stressful, and could happen while panning the view
- Video gives a real time experience which is fun and engaging

Experiences of the interactive street view reportage

- It was good to be able to decide the tempo and decide when move between the places in the reportage
- Making decisions about where and when to move was laborious
- Keeping track of what sections that had been viewed was confusing
- It would be easier to focus if being more used to how to navigate the reportage
- The speaker voice was confusing, it was hard to know when it started and was finished and it sometimes interrupted one section of speech with another

Experiences of both formats

- Being active by "looking around" made it easier to focus
- Even if it was tempting to "look around", doing so sometimes resulted in lost focus on what the speaker voice said, especially when there were no connection between the visual view and what was said
- The reportages were fun and engaging
- Viewing the reportages requires the user to focus on the reportage, it is hard to do something else at the same time in a feasible way

Task load ratings

Figure 2 - Figure 7 presents the results of the task load ratings of viewing the reportages, and the results for the additional question on the interest in the subject of the reportages are presented in Figure 8.
Fig. 2. How was it to concentrate?

Fig. 3. How demanding was the physical activity (using the keyboard and mouse)?

Fig. 4. How stressed were you about the pace of the reportage?

Fig. 5. How was it to grasp what the reportage presented?

Fig. 6. How did you feel?

Fig. 7. How much effort did the reportage demand?
4.2 Comprehension of the content

For the questions about the content of the reportages, the average for the 360 video reportage was 2.4 right answers out of 4 possible, and the average for the interactive street view format was 2.7 right answers out of 4 possible. An average confidence level of the answers is presented in Figure 9.

4.3 Media consumption

The user reporting the highest consumption of media usage estimated 22 hours of total media usage per day. The user reporting the lowest consumption of media estimated 4.7 hours of total media usage per day. The median media usage was 11.9 hours per day. When asked about how often they used two or more media types concurrently on a scale from 1 to 5 where 1 is "seldom" and 5 is "often", all users gave an answer of 3 or higher.

5 Discussion

The result shows that the experienced focus when viewing the reportages varied much among the users. Advantages and disadvantages of the formats can be recognized in the results and are associated with characteristics of the formats. Although the users’ reported media usage also varied, no user reported very low media usage. It can be compared to an extensive study of the media usage of American teenagers [6] where 8.4 hours of media usage per day was considered to be high. All users also reported often using several media types concurrently, and thus the results should be relevant for the purpose of targeting users with digital lifestyles.
5.1 Rich media elements

A characteristic for both of the formats is that they use rich media elements to present information to the user, with images, sound and speaker voices. For both of the reportages, an experience expressed was that they were engaging and that they kept the user’s focus in a way that made it hard to do other things at the same time in a feasible way. Thus, such an experience may indicate that the way the formats present information well suits a breadth-biased attentional filter, since the user is given a large scope of information to pay attention to, a scope that is large enough to not make room for more information from other sources, and does not require the user to focus on a single particular piece of information in detail. Though, it is important to emphasize that the study was conducted without consideration for possible other competing stimulus that may be present in everyday life. The results from how well the users comprehended the content show that users also seem to have assimilated information from the reportages and that they succeeded to convey at least some of the content to the users.

Users did not express difficulties taking part of the different media elements concurrently, and according to Wickens [12] the elements should be a good multi-tasking combination since they use different resources along the different dimensions in the Multiple Resource Theory. One exception was that "looking around" sometimes could result in lost focus of the speaker voice when these two did not have a clear perceived connection. Since these activities as well are well spread out along the different dimensions, by together utilizing all the different resources of cognitive processing (listening to the speaker voice and perceiving the visual content), selection and execution of action (select where to "look" and execute action), spatial processing (perceiving the visual content), linguistic processing (listening to the speaker voice), auditory perception (listening to the speaker voice), visual perception (viewing the visual content) and focal and ambient vision (viewing the visual content), the difficulty may occur because of limited resources rather than poorly utilized resources.

Hence, since the rich media elements indicated to enhance a breadth-biased attentional filter and together compound a good multitasking combination, they seem to be an advantage for the formats.

5.2 Characteristics affecting perceived task load

The task load ratings are widely spread along the scales of all questions for both formats, and this shows that the perceived task load when viewing the reportages differed much among the users. The interviews identified some characteristics that were experienced to increase the task load, and as proposed by Lavie et al. [10] these were also perceived to decrease the ability to maintain focus.

"Looking around". A characteristic shared by the formats is the ability to "look around" while a speaker voice is talking, and thereby gather more perceptual stimuli. This was experienced as an advantage for keeping focus, but it was also expressed that it sometimes could result in lost focus of what the
speaker voice talked about when what was shown visually was perceived as more arbitrary content, and not strongly connected to what the speaker voice said. In these situations the user could start thinking about what to look at and trying to make sense of why the specific visual content was viewed, and then the "looking around" resulted in cognitive processing rather than just getting more perceptual stimuli. Thus, this resulted in increased cognitive load and also decreased ability to focus. This as opposed to views where the visual content and what the speaker voice said was perceived as clearly connected, where "looking around" gave the user more perceptual stimuli without also increasing the cognitive load, since the user then did not question the purpose of the visual content. Hence, it is an advantage that the visual content of views strongly relates to what the speaker voice says to help the user to focus.

**Tempo.** The tempo in which the reportages proceeds is another factor that was found to affect the focus of the user, and this is a thing that varies between the two formats. The characteristic of the 360 video format is that the video goes on and presents new information as time lapses, while the characteristic of the interactive street view format is that the user decides when to proceed with new information. Some users found it stressful when the 360 video was too fast, and they felt that this was easier to focus when they were able to decide the tempo themselves, as they did with the interactive street view format. Some other users found it stressful to decide when to move on when viewing the interactive street view reportage, and found it easier to focus when viewing the 360 video where they did not need to bother about that. Hence, an advantage of the 360 video format is that the user can be guided through the story but with this comes the difficulty to find a tempo that different users are comfortable with and that is not perceived to increase workload. An advantage of the interactive street view format is instead that the user can adapt the tempo, but it then becomes a challenge to make the decision easier so that the decision process do not result in too much extra workload.

**Scene switches.** Another thing that was expressed to interrupt the focus while watching the 360 video reportage, was that scene switches sometimes happened while the user was panning the view. Even if it did not happened many times during viewing the reportage, it still annoyed the user and forced the user to reset for the next scene before being done with the current. Hence, it is a disadvantage for the 360 video format if the user’s panning do not time well with the scene switches, and those situations may put extra cognitive load on the user in the form of irritation and stress which also results in lost focus.

**Navigation.** For the interactive street view reportage, the navigation of the reportage was expressed to be strenuous and result in lost focus. It was hard to know what places that had already been visited, and since this information was not provided in the interface of the reportage, the user needed to keep track of where they had been and not and which information that had already been viewed. A thing that worsened this was the speaker voice, which confused the user about what information the user had already heard by being ambiguous about when it started and was finished, and it also sometimes interrupted one
section of speech with another when the user interacted with the reportage. These things were expressed to increase the perceived workload and result in lost focus of the content of the reportage, and hence they are a disadvantage of the interactive street view format. It was also expressed that it would be easier to focus if being more used to navigating the reportage, which further indicates that the navigation of the interactive street view format was experienced as non-intuitive and may require its user to be experienced with how the navigation work if it should not result in increased perceived cognitive load.

5.3 Outer factors affecting the focus

The users’ interest in the content of the reportages did also vary quite much, though few people had very low interest in the reportages. The interest in the reportages may have had a positive effect on the users’ ability to focus, which should not be credited to the formats and the way the content were presented. It is also important to emphasize the settings in which the users viewed the reportages, a test situation may affect the results and some users are more affected by a test situation than others.

6 Conclusion

Both of the formats have potential to capture users’ focus for a longer period of time, but that also depends much on how the content is using the rich media elements and how contended the individual user is with the tempo and the activity that is demanded when consuming information through the formats. Hence, although it often is important that the content is deliberately arranged and to acknowledge users’ different preferences, this should be especially emphasized when using these formats to present information. The disadvantages of the scene switching in the 360 video format and the navigation of the interactive street view format are however things built into the formats and not affected by the content or the user in the same way, but are also things to be aware of when using the formats to present information.

This study only recognize some aspects of presenting content with the formats, and there are many further things that would be interesting to evaluate. One thing would be to investigate potential improvements on the scene switches and navigation problems of the formats. It would also be interesting to evaluate further interface possibilities and how they affect the user’s focus, such as interfaces on tablets or phones that use the sensors of the device to let the user "look around" by moving the device. Except from the focus of the user, it would also be interesting to widen the perspective and consider content providers’ thoughts on the formats.

References

Using smartwatch as an aid for people with diabetes

Elin Hörnfeldt
Department of Applied Physics and Electronics
Umeå university, Sweden
id12eht@cs.umu.se

Abstract. 450 million people in the world are living with diabetes. Living with diabetes will include blood glucose tests and injections with insulin to keep a healthy blood glucose level. There are medical aids such as insulin pumps and continuous glucose monitoring systems (CGM) that simplify the life for a person with diabetes. These contain a lot of information and might sometimes also be placed inaccessible for the user that can face troubles with controlling their blood glucose. With a questionnaire, 16 people with diabetes answered what information they see as most important from the insulin pump and the CGM. Remaining insulin, active insulin, and the basal rate was the three most important functions the test subject wanted to know from the insulin pump. Blood glucose curve, current blood glucose level, and increasing/decreasing blood glucose were the most important information from the CGM according to the test subjects. This information can be shown on a smartwatch and can simplify the blood glucose control for a person with diabetes. A prototype of a smartwatch with the resulting information was made by design guidelines from literature research.

1 Introduction

Diabetes mellitus is a chronic disease, and there exist different kinds of diabetes. The two most common ones are called diabetes type-1 and diabetes type-2 and the major difference between these two is that people with type-1 diabetes has no insulin production at all, and people with type-2 diabetes have some small production but not enough to meet the needs of the whole body [1].

According to the Swedish Diabetes Association, 4% of the Swedish population is diagnosed with diabetes mellitus [2]. Around 450 million people in the world are living with diabetes, and the International Diabetes Federation predicts that there will be 642 million inhabitants with diabetes in 2040 [3]. These numbers affect a lot of people that will have to rearrange their lives. Living with diabetes will include injections of insulin, blood glucose tests and adjustment to avoid getting too high (hyperglycemia) or too low (hypoglycemia) in blood glucose, and this is not always easy tasks to perform. In Olohan’s and Zappitelli’s work [4] they tell the story of 15-year old Kirsten that has had diabetes type-1 for five years. Kirsten often gets a high blood glucose level before dinner which probably
occurs because she usually eats pizza with her friends after school. Even if she try to correct her blood glucose with injections with insulin, it is hard to find the right amount to inject which results in high blood glucose. Kirsten also exercises after the dinner and then she needs to eat something extra before to avoid getting too low in blood glucose. But even this is a hard adjustment for Kirsten, and she often ends up with hypoglycemia (too low blood glucose) a few hours after the exercising or during the night. Hypoglycemia after bedtime makes Kirsten wake up sweaty and with a headache. Luckily there are technical aids that can help Kirsten and other with diabetes to keep stabilized blood glucose levels. With an insulin pump, people with diabetes can calculate how much insulin they need to to the number of carbohydrates they eat and a rearrange in the basal insulin (the background insulin) in the pump can minimize low or high blood glucose levels. With a Continuous Glucose Monitoring (CGM) a person with diabetes will also be able to see and follow their blood glucose curve [4].

People using insulin pumps and CGM’s show to have lower glycated hemoglobin levels (an average of the blood glucose level over the past 2 to 3 months) compared with people with injections [5]. And a lower glycated hemoglobin level is shown to give a better quality of life for people with diabetes [6]. Insulin pumps’ and CGMs’ simplifies the life for people with diabetes but can it be even more simplified? The insulin pump is usually hidden in the pocket or might even be inaccessible hidden underneath the clothes which can complicate the usage of the insulin pump in the public [7]. How would it be for the user if the information was shown on a smart watch instead?

1.1 Objective

The objective of this study is to investigate how information from the CGM and the insulin pump can be visualized on a smartwatch in a good usability way for people with diabetes. The insulin pump contains a lot of information like current active insulin amount, history, the condition of the pump and much more. The CGM also have different information about the blood glucose levels. Designing for a smartwatch will include a large downscaling of information [8], so what is the most important information for a person with diabetes to know at the moment? To find out what information the user want to see on the smart watch, we will study what the test subjects see as most important to know at the moment. We also want to find out what the guidelines for designing smart watches are and from the result, create a prototype. This study is focused on the people with insulin-depended treatment, i.e. people with diabetes type-1.

2 Theory

For designing this approach, we will have to consider two major areas. What is diabetes and what are the medical aids? And how do we implement good usability on smart watches with small screens?
2.1 Diabetes

Diabetes can occur in different kinds of types, but the common factor is that the blood glucose is higher than normal (hyperglycemia). Latent Autoimmune Diabetes of the Adult (LADA), Gestational Diabetes, Diabetes Type-1 and Diabetes Type-2 are the known diabetes diagnoses. Diabetes type-2 is the most common type, about 90-95% of the people diagnosed with diabetes, has type-2. Type-2 occur when the cells in the body get insulin resistance and will therefore suffer from some relative insulin deficiency. Type-2 usually occur by overweight, sedentary lifestyle, and inheritance. People with diabetes type-2 do not usually need to inject insulin for survival and will best stabilize their blood glucose through exercising and weight loss. About 5-10% of the people with diabetes has type-1. It is an autoimmune disease where the body self-destruct the beta-cells of the pancreas (the ones who produce the insulin) and the cells will therefore get absolute insulin deficiency and insulin by injection is required for survival [9].

2.2 Insulin pump

An insulin pump is connected via a tubing to a small needle that is placed under the skin [10]. The pump delivers insulin at a basal rate, also called background insulin. Every person needs background insulin. A healthy person gets it from its pancreas, but for a person with diabetes, the insulin pump is the closest alternative to a real pancreas. The basal insulin is given through the night and day to get a healthy body function. By using a temporary basal insulin, the amount of the background insulin can be lowered or higher when the person, for example, is exercising or is sick. When the person eats, he or she takes a bolus which is an amount of insulin to regulate the higher blood glucose level that occurs after eating, or whenever the blood glucose level is too high. With personal settings on the insulin pump the pump can calculate how much insulin the person should take to the meal depending on blood glucose and carbohydrates, how much insulin from the bolus dose that is still active in the body and much more [11]. The advantages with an insulin pump compared with normal injections are that the small needle under the skin that is connected to the insulin pump can be changed every 2-3 day when a person with only insulin injections have to take around 3-4 doses a day. The basal rate from the insulin pump optimizes the glycemic control overnight and decreases the risk of low blood glucose levels at activities and exercising. The person also gets more freedom of eating when the insulin pump can give different kinds of boluses depending on the food. These are the normal bolus (when the dose is given directly), square-wave (when the dose is distributed over a particular time, usually between 30 min and 2 hours) and dual-wave bolus (which is a combination of a normal bolus and a square-wave bolus) [12].

Pickup et al. [13] show that the blood glucose control is better at a person with an insulin pump than insulin injections. With an insulin pump, the person also needs less insulin to achieve the blood glucose levels.
2.3 Continuous Glucose Monitoring (CGM)

A frequent self-monitoring of blood glucose tests show a better glycated hemoglobin level (HbA1c) [14]. HbA1c is an average of the blood glucose level over the past 2 to 3 months. A Continuous Glucose Monitoring system (CGM) receives up to 288 glucose values every day and will fill in the gap between the self-monitored fingersticks glucose tests [15]. The CGM consists of three parts (see Figure 1), a transmitter (1), a glucose sensor (2), and the insulin pump (3) or another monitoring device [16].

![Fig. 1. The continuous glucose monitoring system from Medtronic MiniMed 640G System. The three components are the transmitter (1), a glucose sensor (2), and the insulin pump (3) [15].](image)

The glucose sensor is placed under the skin and measures the glucose levels in the body. The transmitter connects to the sensor and sending the glucose values to the insulin pump that shows it for the user in a graph. The CGM needs to be calibrated every twelve hours with a fingerstick blood glucose test and can be worn for six days before the sensor need to be changed. With personal settings, the CGM can give alarms when the blood glucose levels are increasing or decreasing too fast or when a certain glucose level is reached [15].

2.4 Smartwatch

A smartwatch is similar to a smartphone except that it is smaller and are worn on the wrist. Usually, the smartwatch is connected to a smartphone and can have apps, receive notification and even calls [17]. They are sold in different shapes (i.e. round, square, rectangular) and sizes. The size of an Apple Watch is 38 mm or 42 mm and Samsung Gear S that is 42.3 x 49.8 mm. A smartwatch can
have both touchscreen, buttons and also other functions as heart rate monitor, accelerometer, speakers, and microphone [18][19].

![Apple Watch and Samsung Gear S2](image)

**Fig. 2.** The Apple Watch and Samsung Gear S2 [18][19].

### 2.5 Design principles for smartwatches

One positive aspect of the smartwatch is that it naturally comes in the area of the eyesight [20], which makes it suitable for notifications. A study made by Giang et al. [21] show that driver glance more at their smartwatch than their smartphone during driving. A more challenging aspect of the smart watch is that it has a small screen so i.e. zooming will not work. The primary function for a smartwatch is to serve a smartphone and only show the most important features, and then redirect to the smartphone [20].

To design a good usability interface, it is important to reduce the amount of information showing on the screen and just visualize the most important information. By decrease the amount of information the need of scrolling will reduce and make a device with a small screen more usable [8].

Apple and Samsung have developed own UI guidelines for their smartwatches. Apples’ design principles are [22],

- Lightweight interactions: the information is easy to access and dismiss.
- Holistic design: the boundaries between the device and software should be blurred.
- Personal communication: be aware of that the smartwatch is carried close to the user.
- Gestures: Taps is action-based, vertical swipe is scroll and horizontal swipe is for showing next or previous page.

Samsungs' design principles are [23],

---

307
– Glanceable: should give the user the most important information in a consistent and informative way.
– Context-rich: make it easy for the user to follow the pages.
– Fast and responsive.
– Attractive: make the app draw attention.

3 Method

By using a questionnaire, 16 people with diabetes type-1 answered questions about what information on the insulin pump or CGM that is most important for them. The test subjects were older than 18, and they have either both insulin pump and CGM, just one of them or none. The questionnaires were made in Google forms, and the majority of them was sent out to known test subjects through Facebook. A few questionnaires were given at the diabetes clinic at Norrlands Universitetssjukhus (the university hospital) in Umeå, Sweden.

The questionnaire had multiple answer questions and first the test subjects stated their gender, age and if they have or have had an insulin pump or CGM. The questions they answered was what information they see as the most important information for them to know at the moment. The options the test subject could choose between was based on the functions and information that already exists in the insulin pump and CGM. From the insulin pump, the options were: the status of the battery, the amount of insulin that is left, time and date, basal insulin, active insulin, temporary basal change, other temporary changes, carbohydrate counting, history, the insulin pump version/age/model and own option. The test subject could choose maximum four options but was asked to choose only three if it was possible. The second question was what information from the CGM that was most important for the test subjects. The options were: next calibration, the age of the sensor, transmitter battery, transmitter version/id, blood glucose curve over the latest hours, the current blood glucose level, if the blood glucose is rising or falling and own options. This question also allowed four options to choose but if the test subject could choose only three option they were asked to do that. If the test subject did not have either an insulin pump or CGM, they got the same question but were asked what they thought was the most valuable information for them. The test subjects also answered a question about where they stored their insulin pump or CGM. They were asked if the device was easy to get from, for example, a pocket, or if the device were more hidden, for example, underneath clothes or an own option. A supplementary question was if the test subject saw the location of the insulin pump or CGM as an obstacle for them to look at their blood glucose curve or some other function or information. The options were yes, no or sometimes. At the questionnaire given at the clinic a supplementary question about what kind of diabetes the person had was added, to be sure that the test subjects had diabetes type-1 when they were unknown. By a literature research, guidelines for designing smartwatches was collected. After receiving the answer from the questionnaire, a prototype was made in Adobe Illustrator by the design guidelines.
To notice in this division is that all test subject using blood glucose meters. These should not be confused with CGM’s who measure the blood glucose continuously. Glucose meters is necessary for 100% certain blood glucose values and is a necessary tool for using CGM’s.

4 Result

Total 16 test subjects answered the questionnaire. 11 through Google form and 5 through the clinic. It was 4 men and 12 women where 13 was between 18-30 years old, and 3 was over 30 years old. 3 had an insulin pump, 4 had CGM, 8 had both an insulin pump and CGM and 1 had injections.

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;18 years 18-30 years &gt;30 years</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Insulin pump</td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>CGM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGM and insulin pump</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Injections</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Women</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;18 years 18-30 years &gt;30 years</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Insulin pump</td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>CGM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGM and insulin pump</td>
<td></td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Injections</td>
<td></td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. The distribution of the test subjects.

The test subjects could choose 3-4 alternative from the insulin pump functions and the CGM information. The three options that got most votes from the insulin pump was remaining insulin amount with 14 votes, the basal rate with 8 votes and active insulin with 8 votes (see Figure 3). The three options that got most votes from the CGM was blood glucose curve over the latest hours with 14 votes, current blood glucose level and increasing/decreasing blood glucose with 15 votes each (see Figure 4).

With the downscaling of information from the insulin pump and CGM, the three chosen options from each device can be shown on an own page on a smart-watch. With design guidelines as horizontal swipe is for showing next page, the information should be glanceable, follow the pages should be easy and the application should be attractive, one example of how the smartwatch could visualize the information is made. The result is shown in Figure 5 and Figure 6. Figure 5 show the three most important information from the CGM. One example shows a stabilized blood glucose level and one when the blood glucose is rising.
Fig. 3. The most important functions in the insulin pump for people with diabetes type-1. The alternative was battery status (1), remaining insulin amount (2), time and date (3), basal rate (4), active insulin (5), temporary basal changes (6), another temporary change (7), carbohydrates calculations (8), history (9), the pumps version/age/model (10) and other (11).

Fig. 4. The most important information from the CGM for people with diabetes type-1. The alternative was next calibration (1), age of the sensor (2), the battery of the transmitter (3), transmitter version/id (4), blood glucose curve over the latest hours (5), current blood glucose level (6), increasing/decreasing blood glucose (7) and other (8).
Figure 6 the most important functions from the insulin pump are shown. In one example the remaining insulin amount is still high when the other example indicate that the insulin in the pump is almost finished. Both Figure 5 and Figure 6 represent an own page on the smartwatch where the user will swipe horizontally to switch page.

![Figure 6](image)

**Fig. 5.** An example of how blood glucose curve, current blood glucose level and increasing/decreasing blood glucose can be visualized on a smartwatch. The watch to the left show a stabilized blood glucose progress and the smartwatch to the right show a fast increasing blood glucose progress.

To see if a smartwatch could simplify the blood glucose control for a person with diabetes, a question of how often the test subjects check their blood glucose, where the test subject placed their insulin pump or CGM and if that was an obstacle for the test subject were asked. The result showed that most of the test subjects were looking at their blood glucose curve sometimes, especially when they felt that something was not right. One test subject choose other option which was that person could not see that without a CGM (see Figure 7). 7 test subjects place their insulin pump or CGM in an easily accessible place, and 6 placed it somewhere that it was harder to access. 2 test subjects choose other option that was in an omnipad (a belt special for CGM monitors) and in their bag (see Figure 8). 7 test subject did not see the placement of the insulin pump or monitor as an obstacle to looking at their blood glucose curve, 4 saw it as a barrier sometimes, and 3 choose other option. From the other options, two answered that their CGM was connected to a monitor and they therefore had the monitor and the insulin pump separately, and the third one did not give any answer (see Figure 9).
Fig. 6. An example of how remaining insulin, active insulin and basal rate can be visualized on a smartwatch. The watch to the left show that it is a lot of remaining insulin left and the smartwatch to the right show that the insulin in the insulin pump is almost finished.

Fig. 7. The result of how often the test subjects check their blood glucose. The alternative was only at warning (1), sometimes and especially when they feel that the blood glucose level isn’t right (2), the whole time (3) and other (4).
**Fig. 8.** The result of where the test subjects is placing their insulin pump or CGM monitor. The alternative was easily accessible in a pocket or a belt (1), somewhere it’s a little bit harder to access (e.g. underneath the clothes) (2) and other (3).

**Fig. 9.** The result of if the test subjects see the placement of the insulin pump or CGM monitor as an obstacle to check their blood glucose curve.
5 Discussion

The result from the questionnaire shows that the test subjects were very united about what information that was most important from the CGM. The votes were a little bit more spread out over the functions in the insulin pump. When the CGMs’ primary mission is to measure the blood glucose levels it is not a very surprisingly result. The insulin pump has much more functionality to choose between for the test subjects. Some of the answers can also be a little bit misleading when the test subjects that only had a CGM has a monitor that is ”free” e.g. not connected to the body like an insulin pump are. It is therefore much easier to place the monitor in a bag where it will be more accessible than an insulin pump placed closer to the body would be.

After the compilation of the questionnaire, a prototype was made. The prototype of the smartwatch achieved in this study is only one example of how a smartwatch can be designed for this purpose. This prototype was downscaled in information, and tried to be done as simple as possible and following the guidelines. When the screen of a smartwatch is so small, the smartwatch should not have so much ”purpose”. When the screen is not that easy to interact with it is good that the functions are not too many. That is why only the most important information is visualized on the watch for the user. In real life, when a user, for example, gets an SMS, the user will see the notification on the watch but probably use the smartphone to answer the SMS, since the smartphone has a bigger screen as well as a keyboard. The same will apply for people with diabetes. The smartwatch will show the most important information, but active choices will be made on the insulin pump. The smartwatch can be perfect for the users to get notifications about their blood glucose levels and also be more aware of it. Vervloet et al. [24] conducted a study were people with chronic medication got reminders from electronic devices to take their medication. Electronic devices (SMS, pager messages, electronic reminder devices) showed to be good reminders. This type of reminder system could be associated with a smartwatch when it is worn on the wrist and in eyesight. A smartwatch could therefore help people with diabetes to be ”reminded” about their blood glucose control. Another positive aspect of a smartwatch is that it is accessible to the user. Even if the majority of the test subjects did not see the placement of the CGM monitor or insulin pump as an obstacle to check their blood glucose, there was still some test subjects that did. And for these, the placement of a smartwatch could simplify the blood glucose control. To consider in this study is that the prototype is only done by design guidelines and by asking people with diabetes what they see as important to know from their insulin pump and CGM. The prototype is not tested, but just a suggestion of how it could look like. A future work in this area could be to see how the blood glucose control affects in real life by using a smartwatch.

5.1 Limitations

Some of the questions may have been a bit unclear for the test subjects when some answers did not respond to the question. I.e. for the question about how
often the test subject looked at their blood glucose curve, one answer was that the test subject had not used the combination of CGM and an insulin pump. This answer came from the questionnaire that was left at the clinic. The difference with the questionnaire online and the ones at the clinic was that online, the questionnaire form could give different questions depending on the earlier answers. This could not be done on the physical questionnaire, and the question should have been formulated as “where do you place your insulin pump/CGM monitor?” instead of just ”where do you place your insulin pump?”.

6 Conclusion

A smartwatch can be used to visualize blood glucose curve, current blood glucose level, and increasing/decreasing blood glucose from a CGM. On another page, the smartwatch can visualize remaining insulin, active insulin, and basal rate for the user. This is a complement to an insulin pump and a CGM and can simplify the blood glucose control for a person with diabetes type-1. A smartwatch is accessible and in eyesight, and the user will easily see when the blood glucose is not right.

References

Experiencing the potential of smart glasses as an assistive technology for dementia patients

Karolina Jonzén
Department of Applied Physics and Electronics
Umeå University, Sweden
kajo0136@student.umu.se

Abstract. With the rapid growth of the world’s elderly population, dementia has become a global health care priority, due to its disabling and dependency-increasing effects on the sufferer. At the same time, the recent years of fast advancement in computing science and interface design has opened up new opportunities for developing more powerful technologies to assist these individuals in maintaining their independence. This paper examines the potential benefits, challenges and future possibilities of using smart glasses as an assistive technology for dementia patients, based on a literature review and an interview with domain experts. While several potential benefits were found, such as decreased interaction complexity and promising uses for navigational support, memory assistance and cognitive training, addressing ethical concerns and involving dementia patients in the studies remain as important future challenges to overcome.

1 Introduction

Dementia is a brain condition that causes a gradual decline in cognitive abilities, such as memory, language and orientation, and is one of the most common causes of disability and dependency among elderly [1]. In fact, there are estimations of 47 million people all over the world suffering from it today, and with the rapid aging of population, this number is expected to be tripled by 2050 [2]. Since a cure for dementia related diseases is yet to be found, this has motivated greater investments and research to find more effective ways of assisting both patients, caregivers and family members [3].

One way of increasing independence in everyday activities among disabled people is through the use of Assistive Technologies (AT), defined by Marshall [4] as ”any item, piece of equipment, product or system, whether acquired commercially, off-the-shelf, modified or customized, that is used to increase, maintain or improve functional capabilities of individuals with cognitive, physical or communication disabilities”. ATs can support the cognitive disabilities of individuals suffering from dementia in many different ways. For example, automated prompts and reminders, such as voice messages to remind of appointments and medication taking, locator devices, like electronic tags on keys and wallets to
find misplaced items, and communication aids that simplify contact with relatives and friends [5]. The recent years of fast development in computing science and interface design has opened up for new opportunities and technical innovations with possible applications in this field - one being smart glasses. In this paper, smart glasses are referred to as a miniaturized wearable computer that adds digital information to what the wearer sees [6]. By being hands-free, always on and able to recognize the user’s surroundings and activities, smart glasses could become a powerful aid for cognitively impaired people and their caregivers. However, their potential in the field of dementia care is still relatively unexplored.

1.1 Objective

This paper aims to make a contribution to the development of modern technology in dementia care, by examining the potential benefits, challenges and future possibilities of using smart glasses as an AT for dementia patients. More specifically, it seeks to do this by not only looking into what the technology has to offer, but also by taking to account the human factors that need to be considered when designing for this target group. Furthermore, the paper provides an overview of related work conducted in this field, as well as experts’ opinions on the subject given in an in-depth interview.

2 Background

This section gives a brief introduction to dementia, assistive technologies and smart glasses.

2.1 Characteristics of dementia

To be able to provide a good care for the patients, it is important to understand the nature of the disease. Sandilyan and Dening [7] divide the symptoms of dementia into two broad groups: cognitive and non-cognitive. This study will however focus on the cognitive symptoms. While there are individual differences and the symptoms vary depending on what type and stage the dementia is at, the authors describe several common problem areas affecting different everyday activities:

- **Memory**: difficulty learning new information, misplacing items around the house, repeating questions or forgetting appointments.
- **Orientation**: forgetting what time it is, where they are or who they are with.
- **Language**: difficulties naming objects or people, following a conversation or trouble reading and writing.
- **Actions**: inability to perform complex actions, e.g. buttoning a shirt or brushing teeth.
- **Recognizing sensory information**: problems recognizing objects or people, despite having no visual defects.
The most common form of dementia is Alzheimer's Disease (AD), accounting for 60 to 80 percent of all cases, according to the Alzheimers Association [8].

2.2 Assistive technologies in dementia care

In recent years, there have been a lot of research on more innovative and modern approaches to the in-home care of dementia patients by the development of Intelligent Assistive Technologies (IATs). Bharucha et al. [9] define IATs as technologies that support physical or cognitive impairments by sensing and responding to the user's needs, and adapting to changing situations. In 2009, the authors conducted an extensive research and found a total of 58 examples of such emerging technologies with potential application in dementia care. The examples included everything from automated fall detectors and fabrics with embedded sensors for physiologic monitoring, to adaptive memory aids and wearable devices for task identification and guidance. Although a wide range of IATs was found, the authors concluded both technological and ethical challenges needed to be overcome and requested more clinical studies involving persons with dementia.

The ethical issues regarding the use of Assistive Technologies (ATs) in the field of dementia have been pointed out in several other studies as one of its main future challenges. Zwijsen et al. [10] argue that research on ATs has focused more on improving technical aspects of the devices rather than prioritizing their ethical implications. The authors raise ethical concerns such as privacy issues in recording, monitoring and tracking devices, the possible loss of human contact and the problems of informed consent, in which the patient gives its voluntary acceptance of a medical action, after being told and aware of the plan, its risks, benefits and alternative approaches [11].

Cash [12] discusses how the cognitive problems in dementia can complicate the process of informed consent, but points out that the communication skills of people with dementia should not be foreseen. She additionally emphasizes the importance of involving them in the decision making to make sure the technology meets their needs.

Guidelines Designing ATs for people with dementia can be especially challenging due to their cognitive impairments and the degrading nature of the disease. This section presents design guidelines that consider these issues, found in previous researches where dementia patients were involved - an important ethical consideration as stated earlier.

Hyry et al. [13] present the following requirements for the design of ATs for dementia patients, based on a two-year long case study by Nordic countries [14], that included 29 people suffering from dementia.

– Introduce the technology in the early stages of the disease.
– Combine device functionalities.
– Less the amount of devices.
– Create adaptive and tailored user interfaces.
– Implement tele-presence assistance.

Mayer and Zach [15] developed a mobile touch screen to assist people with dementia using the concept of participatory design to actively involve the users in the design process through interviews, observations and workshops. A summary of the guidelines resulting from the study is seen below.

– Use familiar and non-abstract design components to evoke understandable associations.
– Minimize complexity and choice and emphasize clarity and simplicity to ease decision making.
– Evoke curiosity and use a non-stigmatizing design (i.e. don’t put the disease in focus) to ease acceptance.
– Use positive and supportive feedback that addresses all senses to create comfort.
– Promote learnability and memorability since people with dementia might have to re-learn the interface each time they use it.
– Involve caretakers and relatives to help the patients use the system as the disease progresses.

2.3 Smart glasses characteristics

The most influential and commercially available model of smart glasses today is Google Glass. The Google Glass system consists of an optical head-mounted display that shows information in the user’s field of vision and includes functions like taking pictures, making video calls and running apps [16]. The user interacts with the system through voice commands, head-gestures and a touchpad placed on the side of the frame [16]. Figure 1 below illustrates its main components.

![Google Glass components](image)

Fig. 1: Illustration of Google Glass components based on original image by Tim Reckmann [17] [CC BY-SA 3.0] and infographic by Missfeldt [18].
**Techniques** Following are explanations of various technical concepts related to smart glasses.

- *Wearable technology*: electronics and computers integrated into different accessories, such as watches, glasses, clothing and jewelry, designed to be worn comfortably on the body and used in people’s daily life [19].

- *Computer-mediated reality*: technologies where information is either added, subtracted or manipulated from one’s perception through the use of a wearable computer or hand-held device [20]. Following are two examples of such technologies, as described by Gaukrodger and Lintott [21]:
  - *Virtual Reality* (VR): allows the user to move and interact in a simulated three-dimensional environment.
  - *Augmented Reality* (AR): incorporates the real environment with virtual objects, opposed from VR where a complete virtual environment is constructed. The Google Glass system uses this technique to present digital information to the user’s field of view.

- *Context awareness*: the use of non-explicit user and environmental input, often acquired through sensors, to determine the user’s current activity, which can help to reduce interaction complexity in task-support applications [22].

**Applications** Smart glasses and other head-mounted display systems have, to date, mainly been developed for the general public and for more specific purposes in fields such as the military and healthcare [23, 24].

3 Method

The study was carried out by analyzing findings from a literature review and an interview with domain experts, described in more detail below.

3.1 Literature review

A literature review was conducted to find existing research regarding the use of smart glasses as an assistive technology for dementia patients. The review was carried out through an online search in databases where peer-reviewed material was prioritized, in order to ensure high quality and creditability.

Due to the limited number of identified articles in this area, the search was broadened to include studies on similar technical solutions, using either computer-mediated reality or any kind of head-mounted display system, preferably targeting dementia patients, but elderly and cognitive impaired people in general were also considered.

3.2 Interview

To gain further insights about the potential of using smart glasses in dementia care, an interview with domain experts were conducted. Except from getting
expert opinions regarding the use of smart glasses for dementia patients, the interview was also useful in gaining a deeper knowledge of how assistive technologies in general are being used and developed in dementia care.

A semi-structured format was chosen to guarantee collection of relevant data without limiting the possibilities of new leads, as suggested by Bernard [25]. The participants were interviewed in group, mainly due to time limitations on both parts. Group interviews can also create a group dynamic that helps to facilitate the discussion [26]. With permission from each participant, the interview was audio recorded in order to capture all details and be able to remain focused on the dialogue.

**Participants** Two occupational therapists with experience of working with assistive technologies for elderly and people suffering from dementia or cognitive impairments participated in the interview and were recruited through email communication. The invitation to participate was given to a larger number of therapists, which resulted in these two being able and willing.

**Questions** A set of prepared questions, presented below, were constructed and used as an interview guide to ensure the main goal of the interview was addressed.

- What do you think of the idea of using smart glasses as an assistive technology for dementia patients?
  - What do you think could be the benefits of using such a technology?
  - What do you think could be the challenges of using such a technology?
- How are assistive technologies being developed and used for dementia patients?
  - What factors are important to consider when designing for this target group?

**Procedure** The interview was carried out in Swedish, at the participants’ workplace, and lasted for approximately one hour, which also included a demonstration of their current repertoire of cognitive assistive technologies. As recommended by Doody and Noonan [27], the interview started with the participants being informed about the purpose of the study and assured of their confidentiality and anonymity.

Before the question session began, a promo video of the Google Glass application Trace, by Ward et al. [28], was shown to give a more comprehensible introduction to smart glasses and their possible potential and application in dementia care. The creators describe the app as a tool aiming at supporting people in the early stages of Alzheimer’s disease, by combining Google Glass and a tablet device. In the 3-minute long video, it is demonstrated how the system can be used to ease everyday life, as the glasses provide the user with reminders of meetings, face recognition of relatives and friends, navigational directions and a camera to capture moments during the day that later can be viewed in a timeline on the tablet (see Figure 2 on next page).
4 Results

This section presents the main findings from the literature review and the conducted interview with domain experts.

4.1 Literature review

While no published research on the specific use of smart glasses in the field of dementia care was found, several studies were identified exploring the use of related techniques with computer-mediated reality and Head-Mounted Display (HMD) systems for cognitive assistance and training.

In 2015, Garcia et al. [29] reviewed the recent and current use of Virtual Reality (VR) for Alzheimer’s Disease (AD) applications. The authors present different studies where VR systems are developed for diagnosis and cognitive training of mild cognitive impairment and dementia patients, where focus lies on everyday activities like navigation and orientation, face recognition and cognitive functionality. The authors suggest future applications in this field should exploit the available modern techniques and prioritize the specific needs of AD patients and how their symptoms develop over time. Furthermore, the authors point out Augmented Reality (AR) as an upcoming useful tool for AD.

GenVirtual, by Corea et al. [30], is an AR musical game proposed to help people with learning disabilities by stimulating the memorization of colors and
sounds. Virtual cubes in different colors are projected on markers, each associated with a specific musical note. During the game, the cubes light on according to a musical sequence that gets expanded as the user stepwise is to remember the musical notes played, by using their hands or feet to interact with the cubes. The authors suggest that using VR and AR systems in health care is especially beneficial, opposed to traditional treatments, since they allow the user to interact in a natural way. Additionally, the authors point out that with the interfaces being virtual, the risks of patients hurting themselves are reduced and the stimulus generation can be controlled by therapists.

Another example of an AR game in this field is CogArc, under development by Boletsis and McCallum [31, 32]. The authors describe the game as a tool for cognitive training and screening, primarily targeting elderly players above 60 years old. CogArc consists of a collection of cognitive mini-games, such as matching shapes, memorizing patterns and forming words through manipulation of physical cubes placed on a desktop. The AR game content is viewed through a tablet PC mounted on an adjustable arm to enable use of both hands, though the authors point out AR glasses would be the ideal solution. The game was tested on five elderly individuals, however not diagnosed with any kind of dementia disease. The results showed overall positive responses, but he authors also found specific problems related to AR, usability, interaction and game design that lead to some negative reactions.

Quintana and Favela [33] developed a system called Ambient aNnotation System, aimed at supporting people with AD and their caregivers by using AR and a tagging system to provide text-based and auditory memory aids. The user wears a mobile phone, hanging from the neck, a wearable cam recorder on one ear and an earphone on the other. The system involves both the person with AD and the primary caregiver, who creates annotations that automatically get presented to the user when approaching a certain object in the environment. While the authors describe how the system requirements were set up based on AD characteristics, it was not tested on the specific target group, but on six graduate students, though under conditions purposed to simulate those of the final user.

Some studies exploring the use of HMDs were also found. Firouzian et al. [34] developed a prototype of a HMD system designed to assist seniors suffering from memory loss in daily navigational tasks. Though not implemented with any computer-mediated reality technique, the system consists of a pair of eyeglasses with a mounted camera, a Global Positioning System (GPS) tracker, a Bluetooth headset for voice communication and sensors like accelerometer and step counter. The main idea of the system is to provide the user with visual cues from a set of blinking LED indicators located on the frames of the lenses, controlled by a remote caretaker. Preliminary results from user tests of the prototype showed that the performance was highly related to the user’s severity of dementia. Personalizing the visual cues based on the user’s visual and orientation abilities was suggested to improve the system.
AI Googles, developed by Nakayama et al. [35], is another HMD system with possible application in memory assistance. The goggles, implemented with a camera and image recognition techniques, identify and name objects in the environment to be displayed for the user in real-time. The authors describe how it for example could be used to find misplaced items, since the system can provide images of where they were last seen. While no user tests on the target group were conducted, real world testing of the system showed great effectiveness in the object identification.

4.2 Interview with domain experts

A summary of the interview with the two occupational therapists is presented below, categorized in the main themes identified.

**Smart glasses: Possible applications and challenges**

The possibilities of using smart glasses as an assistive technology for dementia patients was discussed after having watched the video of the Google Glass application Trace. Both therapists found the idea very interesting and could think of many different applications. Capturing daily moments on picture, as shown in the Trace video, was one of the suggested uses: "Think about those being on daily activities and then when they come home and their spouse might ask 'What have you been doing today?', then they can easily look together and discuss and get a topic of conversation that works, opposed to not being able to retell anything."

Furthermore the possible use of navigational support was discussed. While being positive to this, as it could help persons with dementia to stay active without the risk of getting lost, the therapists discussed how the guidance should be presented to the user. One of the them argued that a map, as used in the Trace app (previously shown in Figure 2c), would be to abstract for a person with dementia to understand: "That would need to be complemented with a voice that verbally gives instructions." The other therapist suggested those instructions could advantageously be combined with displaying arrows in the real environment.

Additionally, the therapists mentioned how the GPS function and navigational guidance would be a safety not only for the person with dementia, but also for their relatives, if they could have the location of their demented family member displayed on their phones. However, this kind of tracking products are not currently supported by the municipal and county: "There are no decisions on who is responsible. You can buy [the product] yourself, but not as a refund, so that you could get it as an assistive technology." The therapists explained how this is mainly due to slow bureaucracy and ethical concerns regarding the right to track someone who may not be fully aware of it.

Using the glasses to recognize people was another function the therapists found interesting: "Because there are many who pretty early suffer from not being able to recognize people and acquaint persons. The family usually works, but when meeting someone out in the city that says like 'Oh hi, how are you?'
and then [the person with dementia] can think like 'Oh my God, who is that?' and then they could get some support like 'This is Jill.' Ethical dilemmas were though once again brought up as a challenge, since the therapists felt it would be a bit creepy to have people walking around with recording devices that secretly can track someone’s identity.

When asked about how suiting the smart glasses’ visual interface would be for a person with dementia, i.e. getting digital information displayed in their field of vision, the therapists answered it could work really well for some people, while not at all for others: "That is something that needs to be tested individually. I also think that it probably would not be that suitable for those with visual hallucinations."

**Introducing assistive technologies: Challenges and considerations**

Assistive technologies (ATs) were also discussed in a broader context during the interview - especially the challenges in integrating them into the patients’ daily lives. The therapists emphasized on the importance of introducing ATs as early into the disease as possible: "That is something we have discussed quite a lot when it comes to applying cognitive aids, that you need to start really early, almost before they even feel the need of it, for it to work and be incorporated into their everyday life.".

The therapists also mentioned how motivation plays an important role in this, stressing the importance of considering the ethical aspects: "Each patient’s needs and what is meaningful and what they want to do is what should be in focus." Furthermore, they described that even though caregivers and family can see how an AT would be really helpful, some patients are not aware of their cognitive impairments, and hence do not see the need of getting any kind of support: "It is a challenge to introduce an aid and to make them see the purpose of it and still not be insulted by feeling like 'They think I am stupid!' or something like that.

The patient’s previous experience with technology before the onset of the disease was another important factor mentioned by one of the therapists: "When it comes to these technical aspects and it becomes several technical steps, it can get difficult pretty early on. For example using an iPad can be challenging quite early on, but if you have previous knowledge it is easier." The other therapist suggested relatives or other people close could help to handle more complicated steps in the device.

**Designing for dementia: Human factors and considerations**

Human factors to consider when designing for people with dementia was another reappearing theme in the interview. One of the more prominent subjects brought up on this was the importance of looking at each patient individually, and seeing their whole disease picture: "Often, as you get older, you usually not only have dementia, you might have had a stroke before and be half-side paralyzed, or maybe you have Parkinson or a cardiac failure that have made you weak and tired since before [the dementia disease]."
Aside from suggesting the ATs to be personalized so they can adapt to each individual, one of the therapists also proposed to make them adjustable to different situations: "So that you can deactivate a stimuli that you might need in a different situation." Having too many inputs at the same time can be too stressful and confusing for a person with dementia: "The idea when you have these cognitive issues and similar is that you should have an aid that only helps you with what you really need. So it should be as clean and simple as possible around."

When being asked how information is best presented to a user suffering from dementia, the therapists explained that using visual information is the most effective way: "Reading is possible, but the visual memory is better than voice and read information, so pictures are optimal." They additionally suggested using a combination of picture and text or voice, if possible.

The therapists also mentioned the importance of supporting people with dementia to live in their own homes as long as possible: "[At home] you are competent and can find the bathroom with your eyes closed, instead of moving into an apartment or something, where you can be totally handicapped. It was though brought up how it is important for many sufferers to not feel stigmatized by using assistive devices.

5 Discussion

This study shows that there is an increasing interest in using modern technology as a way of supporting the world’s growing elderly population and people in need of cognitive assistance. Even though no studies specifically involving smart glasses were found, their potential in the field of dementia care can still be discussed based on the presented expert opinions and related research on Assistive Technologies (ATs) that use computer-mediated reality and head-worn technology. The potential benefits, challenges and future possibilities of using smart glasses as an AT for dementia patients are discussed separately below.

5.1 Benefits

Wearable devices in general can be easily integrated into people’s everyday life, since they in a way becomes a part of the user. Combining such devices with intelligent and modern techniques, such as computer-mediated reality and context awareness, reduces the interaction complexity, since there is no need for the user to explicitly give input to the application. Instead, these applications can assist the users by sensing their current resources and providing feedback automatically. Having that kind of cognitive support system, implemented in such a common everyday accessory as a pair of glasses, could be especially beneficial in dementia patients, as the therapists described how it can be challenging to integrate new devices into their everyday life.

Another major benefit with the glasses being intelligent is the fact that they can be more easily adapted to different situations and users, which the therapists
described as key design considerations, since the symptoms in dementia vary individually and during the development of the disease. Having an individual and tailored approach to the AT agrees with one of the guidelines proposed by Hyry et al. [13] and findings in the study by Firouzian et al. [34].

Smart glasses also have the ability to address various senses, by using both visual and auditory information, which should be exploited according to the guidelines by Mayer and Zach [15] and the therapists’ opinions. The system furthermore enables different ways of user interaction; except from touch and gestures, it provides a hands-free interaction through voice-commands and a visual interface being presented in the user’s field of vision. This is another powerful feature considering the individual differences in this target group and the therapists’ experience that dementia rarely is the only present impairment or diagnosis, possibly further limiting either cognitive or physical ability. Using computer-mediated reality systems in health and dementia care was also suggested as a beneficial solution in several studies presented in the literature review [29–32].

On a functional level, this study presented a wide range of possible uses for smart glasses and similar solutions in dementia care. The literature review showed examples of applications for cognitive training [30–32], memory assistance [33, 35] and navigational support [34], while the therapists also discussed possible uses of storing memories and recognizing people.

All of these support systems could help to increase or maintain independence in everyday activities, and hence improve quality of life and reduce the burden of caregivers and family. This would also enable the sufferers to remain living in their own homes for an extended period, which the therapists stated as a critical aspect in keeping their independence.

5.2 Challenges

Developing and introducing any AT for this target group can be challenging, since learning new things is one of the cognitive abilities being highly affected by the memory degradation. Applying ATs in this target group can also be obstructed by the lack of motivation, due to patients not being aware of their symptoms, as explained by the therapists. Promoting learnability and memorability, as suggested by Mayer and Zach [15], as well as introducing the technology in the early stages of the disease, proposed by both Hyry et al. [13] and the therapists, are therefore essential for the AT to be usable and able to reach its full potential.

Other important challenges found in this study are related to ethical issues. Since smart glasses and intelligent devices in general typically are equipped with a camera and incorporate sensors that can be used for monitoring and tracking purposes, it is important that the patients give their informed consent, which can be complicated by the cognitive symptoms in dementia diseases [12]. Additionally, the therapists expressed how their county and municipal do not accept assistive technologies with tracking functionality, which limits the navigational
support only to those who can afford to buy the product themselves. Similar problems and ethical dilemmas goes for the facial recognition technology.

Despite their cognitive impairments, dementia patients should be involved in the design and decision making process for the technology to meet their needs, as stated in previous studies [9, 10, 12]. The therapists also describe how it is the patient’s needs and wishes that always need to be in focus. Although challenging, participatory design in this target group have been carried out with good results, as shown by Mayer and Zach [15]. However, most of the applications presented in this study are lacking in this aspect, which unfortunately appears to be a common issue, as previously shown by Bharucha et al. [9].

5.3 Future possibilities

Since smart glasses still are such a recent innovation, the current lack of research and studies exploring their potential in this specific target group is not surprising. However, this is likely to change as they become more accessible and commonly known. In the longer perspective, when they potentially also are being more accepted by society and commonly used, people getting a dementia diagnosis will already have previous experience with the technology - an important factor for an AT to be integrated into everyday life, according to the therapists. With an increased use and a discrete design in the future, the glasses could also be worn without being a symbol for the disease, and hence reduce the stigmatization and ease acceptance, as suggested by Mayer and Zach [15] and the therapists.

6 Conclusion

This study has examined the potential benefits, challenges and future possibilities of using smart glasses as an assistive technology for dementia patients, taking into account both what the technology has to offer and what human factors to consider. Results from a literature review and an interview with two occupational therapists, showed that smart glasses have many promising applications and features that could be beneficial in this target group. However, their real potential has yet to be proven in clinical studies, which hopefully will happen as smart glasses become more accessible. Including dementia patients in the studies and prioritizing the ethical concerns are important challenges to address in the continued development of smart glasses and other modern assistive technologies in dementia care.

References

UX Guidelines for Accessibility in Action Shooter Games

Anna Päärni
Department of Applied Physics and Electronics
Umeå University, Sweden
annpaa91@gmail.com

Abstract. Gaming has moved away from being a pop-culture phenomenon to an integral part of culture, but not everyone can partake. When game developers forget that all future players of their game do not possess the same abilities, some are left out. This article aims to identify where cognitively disabled players are left out in the game development process, focusing on their user experience in action shooter games. This was done by evaluating the UX in three action shooter games, using a template built on HTA and the game interaction model to perform and evaluate a set of tasks. The resulting data, paired with commonly reported problems by cognitively disabled players, was used to identify where accessibility mistakes had been committed by the developers. The pinpointed problematic areas were then used to develop an additional set of UX guidelines, building on existing accessibility guidelines for games. The results showed that the detected issues were centred around the main menus, disruption, and perceiving and interpreting danger. The research concluded with four additional guidelines; rely on design conventions, mix visual hints with other input, avoid undesired multitasking, and allow for customization in combat.

1 Introduction

The UN reports that 15% percent of the world's population are disabled to some degree, making it the world's largest minority[1]. Thus it is important to move away from the idea of a "standard" user, since all user have different levels of experience and abilities (for example, about 8% of all males are colourblind[2]). The United Nations Committee on the Rights of Persons with Disabilities states that a part of the rights as a disabled person is the right to be able to participate in cultural activities on an equal basis[3]. The evolution of IT, design and technology does not only mean tools for improving things like public transportation or hearing aids; it can also be an important bridge to culture. These bridges can consist of screen readers to read blog posts, SDH-subtitles (subtitles for the deaf and hard-of-hearing) during a movie, or books written in simple accessible English.

Video games are an integral part of our culture[4] and no longer just a pop culture phenomenon[5]. 2015 51% of all American households had a dedicated
game console, and 42% played video games at least 3 hours per week[4]. Economically, the game industry might benefit from being invested in making sure that gaming should be accessible to all of those who want to participate, since more players would result in a bigger revenue[6].

Multiple groups are working for a more inclusive world of games, for example Games Accessibility Special Interest Group\(^1\)[7], and the AbleGamers Foundation, an interest group for disabled players and relatives to disabled players[8]. At the Gamer Developers Conference the 16th of March 2016 a lecture titled Inclusion: How to Make Your Game(s) More Inclusive to Millions was held by the director of AbleGamers, Mark Barleton. Barleton described the obstacles disabled players face, and advised game developers to increase the accessibility in games to reach a broader audience and become more inclusive[9].

A coalition of game studios, specialists and academics have published an extensive online resource with accessibility guidelines for game studios, and both AbleGamers and Games Accessibility Special Interest Group have also proposed accessibility guidelines for games [10]. The guidance provided in these is not divided into different game genres, and is with a few exceptions not specific regarding UX implementations. Examples of instances where cognitively disabled gamers are left out or hindered because of poor UX are lack of a tutorial mode, poor documentation and no indications of dangerous situations [7].

The objective of this paper is to identify where cognitively disabled users are left out in the design process of games, and propose new UX guidelines for accessibility when developing action shooter games played on PC. Three important aspects of the games were included in the analysis; the in game menu, the main menu, and the combat experience (usage of weapons and abilities). The guidelines were developed with data collected in an accessibility evaluation of three action shooter games, as well as previously reported problems for cognitively disabled players in games. They are intended to be seen as an addition to or altering of existing guidelines, proposed from a UX-designers point of view. The research questions in this article are as follows:

- Which common pitfalls in the UX, that hinder cognitively disabled players, exist in action shooter PC games?
- How can existing guidelines for UX in game development be altered to better include cognitively disabled players in action shooter games?

2 Theory

Games Accessibility Special Interest Group states that video game players are far from a homogeneous group with the same abilities (and lack of disabilities), income, age and gender. Disabled player’s potential is often underestimated, as well as their willingness to participate in games[7]. Accessibility in video games means including all players who wish to participate, and not limiting the experience to only be enjoyed by "able" players.

\(^1\) GASIG is a part of the International Game Developers Association, San Francisco.
The concept of *universal design*[3] is another important and interesting aspect to take into account. Instead of viewing accessibility for disabled users as something that needs to be added on and tailored to fit a certain group, design can from the beginning strive to be as inclusive as possible. In game development this can be done by considering accessibility in early stages of the development, but also

2.1 Cognitive disability

Cognitive disability is a broad term that can be defined in various depths, where one description given in a medical journal is the following,

Causal factors related with cognitive disability are multiples and can be classified as follows: Genetic, acquired (congenital and developmental), environmental and sociocultural. Likewise, in relation to the classification, cognitive disability has as a common denominator a subnormal intellectual functioning level [...]. [11]

Examples of diagnoses that fall into the category cognitive disability are ADD, dyslexia, dyscalculia, autism, Downs syndrome and TBI (traumatic brain injury), where the diagnoses can be divided further into different subcategories. Dyslexia, ADD and dyscalculia are examples of diagnoses that are considered to be less severe conditions than Downs syndrome and TBI[12].

2.2 Common issues reported in game UX

A study with 500 participants from 2010[13] collected and compiled common issues that players reported facing when playing video games. The participants consisted of group A (490 individuals), members of a large computer festival, and B (10 individuals), a mailing list for disabled (not game-related). 46% of cognitively the disabled players (B) reported problems with speed, language, and understanding the task they were given in the game, compared to 60% of the larger "gamer" group (A). However, it is important to note that group B had a significantly lower number of responses. The types of problems cognitively disabled players reported having are summarized in table 1.

2.3 Existing guidelines for accessibility in games

A coalition of game studios, interest groups that work for disabled gamers, academics and specialists on the topic have developed guidelines for accessibility in games that are available on their website[14] as an online resource. They are divided into basic, intermediate and advanced, as well as the categories hearing, 1

---

1 This is marked as a problem encountered by players who are hearing impaired. However, the author argues that this is a problem that is common for all players, since important information given during cut scenes can easily be missed.
Inability to follow a storyline.

Story is very complex and difficult to follow.

Unable to determine how game is played.

Lack of a tutorial mode.

Poor documentation.

Documentation written at too high a level for intended audience.

Lack of hints in game.

Lack of Informative UI.

Player’s character gets killed/injured repeatedly in game.

No indication of dangerous situation.

Unable to alter game speed.

Unable to adjust difficulty levels.

Unable to complete a puzzle or task.

Vital clues given in cut scenes with no text or visual queues available.

The puzzle is too hard or complex, and cannot be adjusted with difficulty modes or speed adjustments.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inability to follow a storyline.</td>
<td>Story is very complex and difficult to follow.</td>
</tr>
<tr>
<td>Unable to determine how game is played.</td>
<td>Lack of a tutorial mode.</td>
</tr>
<tr>
<td>Poor documentation.</td>
<td></td>
</tr>
<tr>
<td>Documentation written at too high a level for intended audience.</td>
<td></td>
</tr>
<tr>
<td>Lack of hints in game.</td>
<td></td>
</tr>
<tr>
<td>Lack of Informative UI.</td>
<td></td>
</tr>
<tr>
<td>Player’s character gets killed/injured repeatedly in game.</td>
<td>No indication of dangerous situation.</td>
</tr>
<tr>
<td>Unable to alter game speed.</td>
<td></td>
</tr>
<tr>
<td>Unable to adjust difficulty levels.</td>
<td></td>
</tr>
<tr>
<td>Unable to complete a puzzle or task.</td>
<td>Vital clues given in cut scenes with no text or visual queues available.</td>
</tr>
<tr>
<td>The puzzle is too hard or complex, and cannot be adjusted with difficulty modes or speed adjustments.</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Based on the table presented by Games Accessibility Special Interest Group[6]. Issues related to cognitive disability have been selected and presented.[7]

vision, cognitive, motor, speech and general. The guidelines for cognitively disabled players often provide examples of UX implementations and further reading, as well as examples of games with "best practice".

AbleGamers game accessibility guidelines presented in their accessibility guide [10] are divided into four separate type of players; visually disabled, mobility disabled, hearing disabled and cognitively disabled. Each disability category is then further divided into three levels of how much of an impact these implementations would do. They are general regarding the implementation of the UX, where the most detailed guidance is that menus should be intuitive and that enemies should be clearly separated from friends[10].

Games Accessibility Special Interest Group present their guidelines[15] in a similar manner as AbleGamers. The guidelines are divided in to the categories visual disabilities, cognitive disabilities, mobility disabilities and auditory disabilities. The ones regarding cognitively disabled players are mostly short and general, with little focus on how to implement accessibility in the UX. Each guideline also explains the benefits of implementing this accessibility feature. Many have a short question the developer can asks themselves to estimate the
accessibility, for example "Test: Does the player need to read the manual or look up how to properly play your game? [15]".

The three set of guidelines often overlapped, but differ some in the amount of detail, topics brought up and number of paragraphs. None of the sets of guidelines provided advice specific to game genre, and often left room for elaboration and more concrete guidelines regarding the UX.

3 Method

The results were based on two steps; during the first step the author evaluated three existing action shooter games, identifying and reporting commonly reported problems for cognitively disabled players. In the second step the data from the analysis was used to propose guidelines specific to the UX for cognitively disabled players in action shooter games, extending existing accessibility guidelines.

3.1 Evaluating the games

Since the author did not have the means to conduct a observational analysis with cognitively disabled video game players, HTA[16] and game interaction model[17] was used to perform an analysis of the UX without involving users.

The first step in the evaluation was to devise a simple template to estimate the accessibility for players with cognitive disabilities. The template would then be used as a tool for an objective evaluation that would determine if the chosen action shooter games, Left 4 Dead, Bioshock I and Dead Space 1 had fallen into any UX pitfalls regarding accessibility. Three aspects of these games were taken into consideration; the player’s in-game-menu, the main-menu, and the combat experience (usage of weapons and abilities).

3.1.1 Developing the evaluation template

To develop a simple but exhaustive template to evaluate the selected games’ UX, two separate but similar approaches to evaluating the user experience were used. One specific to games and accessibility, the game interaction model[17], and one general for user interface evaluations, HTA [16]. The template was devised to evaluate the performance of a number of tasks, breaking them down into constituents to detect problematic solutions. To connect the HTA to cognitively disabled players frequently reported issues with games[13][6], previous problems were added to the template, so that they could be correctly noted if they were identified.

The finished template consisted of an HTA-analysis for each task (the tasks are defined below in Tasks to be performed), where the evaluator wrote down the required sub-tasks to achieve the main task. The received stimuli, determined response and provided input from/to the game were also given a column for each task, using the principle of the game interaction model-analysis. Additional
notes were taken on each factor when problems and errors were detected, in accordance with the HTA-principle\cite{16}.

3.2 Tasks to be performed

Main menu

1. Change the difficulty setting.
2. Switch on the subtitles.

In-game-menu

1. Pick up an item. Find it in the inventory, and then select the item. Determine what the item can be used for.
2. Select and drop an item.
3. Determine which task currently needs to be performed to move forward in the main quest.
4. Regenerate health.

Combat

1. Shoot the first weapon that is found and equipped.
2. Improve aiming, or determine if this is not possible.
3. Switch weapons.
4. Update a weapon or an ability.

3.3 Choice of games

Games with large budgets have resources to make changes\cite{6} and educate their employees, as well as the possibility to reach a large audience. Sequels have something the first does not; the advantage of player’s input, such as critique towards clumsy menus, or a demand for more of a favourable feature. Games with large budgets, that were either stand-alone games or the first one in a series, were therefore chosen, resulting in the selection Left 4 Dead 1 by Valve, Bioshock I by 2K games and Dead Space 1 by EA Redwood Shores.

3.4 The aspects of the games to be evaluated

To clarify what this paper defines as the vital building blocks of the three chosen aspects (the in-game-menu, the main-menu, and the combat experience), they are below defined individually. The reasoning behind choosing these aspects were that the author believes them to be integral parts of the game UX, strongly linked to the action shooter-experience. The player relies on these three aspects to make adjustments when possible, as well as to perform necessary actions to proceed in the game. Video games are not an exact science, and this was a decision made by the author based on previous experiences of action shooter games.
3.4.1 Main menu

The main menu is the first menu that the player sees when the game is started. It usually contains settings (volume, brightness, difficulty, subtitles and similar), options to continue or start a new game, and an option to exit.

3.4.2 In-game-menu

The in-game-menu is the menu that can be reached and used inside the game, as opposed to the main menu that is reached outside the game when it is paused, or not yet started. Generally the purpose of the in-game-menu is to contain the items obtained by the player, ranging from health packs and weapons to crucial items for solving puzzles or making progress, information about current quest(s), maps, the players stats (health, level, status of abilities and similar) and other information relevant to the specific game being played.

3.4.3 Combat

Combat is when the player interacts with enemies to either incapacitate or kill them. This can be done with special abilities (electric shocks or stunning, for example), firing a weapon or using melee. Different action shooter games offers different weapons, abilities and aiming, and different enemies require different tactics. Often weapons and abilities can be altered within the game, for example by using workbenches or assigning earned points to skill trees.

3.5 Proposing guidelines

The information gathered by the game evaluation was considered separately for each game. Reoccurring problems that could be connected to a specifically defined accessibility issue for cognitively disabled players were noted to be further discussed. This was the presented result of the article as identified pitfalls.

When the detected problems had been noted and written down for each task, they were considered task for task and compared to existing guidelines[15][10][18] to determine if this area had already been addressed. If it had been suitably addressed, the issue was not further considered. If it had not been addressed at all, or was too general to provide tangible advice, a new guideline or an altering of an existing was proposed to solve this.

4 Results

The issues that were detected by the evaluation were chosen to be shown below, which means that the result of each task is not presented. The results from the evaluation were considered for each game separately, and are presented in that way below. None of the games had a tryout-level, or the option to alter the game speed, but did not have complex puzzles either (see table 1).
4.1 Dead Space 1

The font in Dead Space main menu had a coherent UI with the same font through all levels, high contrast between text and background to the text, and clear, loud feedback sounds. Scrolling, selecting and backing had different, distinguished sounds. However, the background behind the menu was flashing and changing, showing streams of texts, numbers and pictures, overlain with static. The background music was loud, changing and also mixed with static.

The in-game menu was projected in front of the player, and could be angled differently depending on where the player looked. It did not pause the game, and took up a considerable amount of the screen. The different available options were presented as icons, where the icons representing missions and database were not standard icons, as well as very similar. Text logs often contained long detailed information that was not relevant to the main objective, and was sometimes written in technical language. There was no indication given to distinguish important text logs from less important ones. Audio logs could be interrupted by transmissions from other characters, cutting them off.

The enemies in Dead Space 1 alerted the player about their presence with growling. The game sometimes played dramatic music when multiple enemies attacked, which could drown out the sounds made by them. Unless the player faced the enemies and saw them it was generally difficult to pinpoint their position, allowing them to sneak up on the player. The players health gauge was placed on the back of the main character, and was out of view if enemies blocked it, if the camera angle was wrong or if the character twisted his upper body. The only constant feedback alerting the player that they were hurt were grunts from the main character, making the assessment of how dangerous the situation was difficult. Audio and video transmissions from other characters did not pause the game, and were shown in front of the player. This meant that a part of the screen could be hidden behind the transmission, making enemies more difficult to both hear and see. Enemies were killed by cutting off limbs, but could "survive" without a head, for example, and it was often difficult to know when the "final blow" was dealt.

4.2 BioShock 1

The main menu in BioShock was inconsistent. Two different fonts were used, shading was sometimes placed above words, sometimes below, and the background image changed drastically between different menu levels. The highest level in the menu had a logo that took up almost a third of the screen, leaving little room for actual menu options.

Optional lore in BioShock was offered in audio-logs that could be found throughout the game. They did not pause the game, and required the player to stand still and be vigilant to avoid being jumped. The in-game-menu offered the messages to be replayed, but this did not mean that the player could not be interrupted again. The audio logs were often interrupted by an announcement from the game, or from a transmission from another character.
**BioShock** had no visual indication of the position of the attacker unless the player faced them. Enemies muttered to themselves if they had not spotted the player, and shouted at the player if they discovered them, but were often difficult to pinpoint if they were not spotted. When the player was hurt the vision was blurred. If the player’s health became dangerously low, a loud heartbeat was played and a text-flash urging the player to regenerate health was shown.

### 4.3 Left 4 Dead 1

The main menu in *Left 4 Dead* relied heavily on simple lists. Only the word *Start* was in a different colour (turquoise, instead of white or grey). The start-option was placed above *Back*, in a list of option that was left-aligned. The HTA also showed that the action that required most steps was to start a single-player game in *Left 4 Dead*. The subtitles were in thin italics, making them quite hard to make out. Sounds from zombies were also subtitled, which meant that when a large group of zombies were nearby the subtitle feed was flooded with subtitled sounds the enemies, drowning out helpful lines from other characters and instructions from the game.

*Left 4 Dead* does not rely on lore and different objectives, since the objective is always the same, and the only messages that were given in audio form were short sentences said by characters. These were often interrupted by another character yelling a different line.

Unless the player faced zombies they were often difficult to spot if it was dark, or if the player was attacked by multiple zombies from different directions. The zombies in *Left 4 Dead* attack in big groups to overwhelm the player and it was often difficult to pinpoint where the zombies were. Red triangles near the middle of the screen were shown when the player was hurt, pointing in the general direction of the attacker.

### 5 Discussion

With only a few exceptions, the HTA and the game interaction model showed that the steps to perform actions and find information were few and straightforward. The issues were rarely too many steps, but instead poor choices of interaction, colours, fonts and layout. The HT-analysis and game interaction model proved to be helpful to pinpointing these errors, since it required the evaluator to perform actions systematically. The problems that were selected and elaborated on below were those that occurred in all three games, and were linked to previously reported accessibility obstacles. These problems are not presented for the games separately, but are instead discussed in areas of interest.

#### 5.0.1 Main menu

One of the most problematic parts of both *BioShock 1* and *Left 4 Dead 1* were their main menus, an issue related to the commonly reported problem *lack of*
Informative UI (see table 1). They often did not follow UX design conventions, and especially the main menu in *BioShock* was inconsistent.

The menu in *Dead Space* had a considerably big flaw. The background kept changing and flashing, showing streams of texts, numbers and pictures, overlain with static. This brought down the readability, was distracting and could even affect players with epilepsy[19]. The background sounds in *Dead Space* were loud, changing and mixed with static, making a bit stressful to take the time to scroll through the menu options.

Menu layouts (as well as general advice for fonts) are addressed in the online accessibility guidelines [18], as well as *Games Accessibility Special Interest Group*’s guidelines [15] and AbleGamers’[10]. An addition to this might be to advise that when designing UX it is important to keep accepted, commonly used design conventions as a guidance. Examples of these Nielsen’s heuristics[20] and Shneiderman’s ”Golden rules”[21]. It is also highly relevant when the player is in dangerous situations or is given important information. Using conventions will help to avoid confusions, poor usage of space and other UX-mistakes.

5.0.2 In-game-menu

A problem regarding both story is very complex and difficult to follow and lack of hints in game (see table 1) were the constant interruptions in monologues and quest updates in *Dead Space* and *BioShock*. The interruptions did not make the story more complex, but it made it significantly more difficult to follow. While the game or a character was conveying important information to the player (either "in person", a text flash on the screen, or an audio/video/text-message), they were often either abruptly cut off or interrupted, while the message kept playing, by attacking enemies.

Some players might enjoy the challenge to try to listen to detailed description of the current mission while fighting off five monsters, but to some it might mean a less enjoyable playing-session. This somewhat addressed in the online guidelines [18] that recommends instructions and narrative to be possible to replay, but none of the guidelines advises to avoid allowing information be interrupted. It could be an additional instruction to allow the player to pause the game when information is given, or to only play instructions and narratives when the player is in a safe space.

5.0.3 Combat

A detected problem that is strongly linked to the problem no indication of dangerous situations, and partially lack of hints in game (see table 1), was the difficulty to read dangerous situations. None of the three games had a clear, visual indication of danger, instead only alerting the player with visual feedback once they were hurt. The enemies in all games either yelled or growled, but this could be drowned out by a message being played or dramatic music indicating an incoming attack. *Left 4 Dead* had an interesting visual solution to telling the
player that they are hurt, but it was not especially helpful if a large group of enemies hurt the player at the same time, which was the most common case. It can be annoying to be given hints and instructions when they are undesired, but too few hints is a very commonly reported error by cognitively disabled players. A possible solution is to allow for players to decide for themselves the extent of hints. This is not discussed specifically in any of the guidelines, but the online guidelines [18] bring up the options to clearly show that interactive elements are interactive, and to remove non-interactive elements completely. In action shooters this can be expanded, and also include showing positions of enemies before they attack. Two possible solutions to this are to give the player a radar or to elaborate on the red triangles in Left 4 Dead, making them pinpoint enemies before they attack. An additional visual indication of danger, instead of only relying on sound, could help avoid surprise attacks and dying repeatedly.

6 Conclusion - Proposed guidelines

Action shooters need to find a balance where the players are given equal opportunities to enjoy the combat oriented game play. There are almost endless possibilities to alter the UX to fit as many players of action shooters as possible, if the developers are both aware of accessibility-guidelines and given the correct information. Judging from the material this article is based on, the industry seems to be including more and more of the accessibility-guidelines provided by interests groups for disabled players.

To add substantial content to the existing guidelines, the proposed ones are putting cognitively disabled players in the middle, strictly regarding UX and developed with the intention to be used in action shooter games. The proposed guidelines are meant to be seen as an addition to existing guidelines, and not the sole advice.

6.1 Rely on design conventions

Using conventions will help to avoid confusions, poor usage of space and other UX-mistakes.

6.2 Avoid undesired multitasking

Allow the player to have the choice to let the game pause when important information is given, or give information only when the player is in a safe space without enemies. Avoid to allow multiple monologues or dialogues to be played at the same time, interrupting or cutting each other of.

6.3 Allow for customization in combat

Let the player add additional hints for combat that show the position of enemies prior to attacking.
6.4 Mix visual hints with other input

A mixture of different types of inputs will make the hints easier to perceive (as well as benefiting deaf players). A visual hint instead of only an auditory when enemies are near can help the player to avoid constantly being surprised and hurt or killed.

6.5 Limitations and future work

Action shooters consist of many aspects, and far from all of them have been examined. The author had to determine which aspects to evaluate to limit the study and deemed the main menu, the in-game menu and combat to be the most important parts of a action shooter game. This inevitably adds some subjectivity to the study. Additional evaluation of the other parts of the game might even this out, but the author argues that games and UX are rarely a completely objective subject.

The analysis would have benefited from an observational analysis where test subjects played the games, with the same directives as the author. The best case scenario would have been to let cognitively disabled players perform these walk throughs. It would have drawn attention to problems in the games that were missed during the analysis performed by the author. The developed guidelines would also have benefited from being discussed with UX-developers at game studios, as well as representatives from interest groups for disabled players.
References


The Effect the Number of Columns has on Information Perception on a Web Page

Petra Andersson

Department of Applied Physics and Electronics
Umeå University, Sweden
id12pan@cs.umu.se

Abstract. It becomes more and more popular to use a column-based layout in web pages, but how does the number of columns affect the user and how the information on the web page is perceived? This paper analyses the use of a web page with two different layouts, one with three columns and one with five columns, both containing the same information. The study is based on a test containing interviews and a questionnaire related to the two web pages. The result shows that there is a correlation between the number of columns and the amount of information a user were able to perceive during a limited time. Participants testing the web page with three columns obtained better results regarding perceiving information on the web page, and they also rated the web pages as more visual appealing considering seven design characteristics.

1 Introduction

The internet has become the number one platform for many businesses, and the trend is increasing. E-commerce and marketing are two examples out of many sectors that utilise the internet as a tool to reach out and profile themselves towards both current and future customers, and markets. Communication between companies and users or clients commonly takes place on the internet, via the companies’ web pages. Web pages are frequently used as promotional tool and a large proportion of a company’s competitiveness rely on their web page and if the visitor finds the company trustworthy [1]. If a company is not convincing enough on their web page a user can, with one single click, open the next search result to find a different web page that belongs to a competitor. The competitor web page provides a similar service only communicated in a manner where the visitor more easily can see the value. This may determine which company the visitor chooses to proceed with, and in the long term become a revenue-generating customer.

How companies choose to design their web page varies, but from benchmarking e-commerce web pages it is found that one common design is to use columns as a way to present information. The benchmark shows that the number of columns frequently used lies between two to six. Figure 1 shows one example of how four columns are used on a web page for presenting information. Columns
Fig. 1. The figure is a wire frame showing how information can be displayed on a web page with three equal columns containing images, a headline and a body text.

is one type of layout to use when designing a web pages. By dividing and categorising the information in smaller sections, the user might get an overview more easily. There is a broad range of possible layouts with regard to columns, for example width, height and number of columns. All these different aspects may have an impact on how the user will perceive the web page and the information presented. Are columns used in the most effective way today, or could companies generate even higher sales or more clients with the right number of columns?

1.1 Objective

The aim of this study was to identify the number of columns that let the user perceive and scan a desktop web page in the most effective way. The effectiveness was measured in the amount of information gathered by the test subject at a pre-specified time. Information gathered consciously and unconsciously was considered in order to see if any of the subjects were affected by the number of columns. By analysing the amount of information the test subjects collected, it was possible to identify potential differences in column usage and establish guidelines for the most effective way of using columns on a web page. The research questions are as follows,

– What is the most satisfying number of columns which helps the user gather the most information in a defined time?
– How does the number of columns affect users’ information gathering?

2 Theory

There are numerous of components to consider when designing a layout for a web page and some of them are more important for this study since they affect how to compose columns. Described below are these components and also aspects such as how a user reads on the web and how much time a designer has to make a first impression.
2.1 Columns

Columns originated from printed or written material such as newspapers, magazines, folders and similar printed media. In the mid-fifteenth century, books were the primary printed material and mainly used one column. In a document found from the early sixteenth century (an English Bible) it shows that the printers had adapted symmetrical typing; the text was justified \(^1\) and placed in two symmetrical columns \(\text{[2].}\) Further, the industrial revolution marked the beginning of graphic design. As advertising grew in a creative direction, the ads evolved and came to look as what today is referred as to graphic design. Columns, as a part of the graphic design, are today frequently used all over the internet. When applying graphic design to a web page, a frequently used technique is referred to as the ‘grid system’. This system is a broader term for graphical layout and consists of columns \(\text{[2].}\) The ‘grid system’ is used for dividing a paper or page vertically and horizontally into margins, columns, inter-column space, lines of type, and space between blocks of type and images. With these subdivisions, the grid forms the basis of a modular, systematic and effective system for designing pages.

2.2 Read patterns

In a study concerning reading patterns Nielsen \(\text{[3].}\) discovered that the users had fairly consistent behaviour when scanning the different websites. The dominant reading pattern is referred to as the F-shaped pattern, since it adopted the shape of the letter ‘F’, see Figure 2. The user firstly reads in a horizontal movement, usually across the upper part of the web page. The user then moves further down the page to continue reading horizontally, but this typically covers a shorter horizontal area than the previous movement. Finally, the user scans the left side of the web page in a vertical movement. From this study it is clear that users focus more on the left than the right side of a web page. Results similar to those

\[\text{Fig. 2. An illustration of the F-shaped reading pattern.}\]

\(^1\) Text aligned along the left margin, with letter- and word-spacing adjusted so that the text falls even with both margins.
of Nielsen’s [3] are also shown in an eye tracking study conducted by Buscher et al. [4] that examined users’ behavior to predict salient and muted regions of web pages. Salient and muted regions are the fields of a web page where the user do not seem to pay as much attention as to other fields. The entire right side of the web page above the fold \(^2\) is neglected for both information searching- and page recognition- based tasks. Participants took approximately ten times longer when searching for information to fixate and pay attention to the right sided regions of a web page - 4,874 milliseconds compared to 339 milliseconds for regions on the top left side, see Figure 3. When users did eventually fixate on the right side, they did not stay there for a long time.

![Table showing median time to first fixation](table.png)

**Fig. 3.** The median time to first fixation on the 10 different areas of a screen in milliseconds from a study conducted by Buscher et al. [4].

### 2.3 Typography on the web

**Optimal line length** There is a lot to consider when designing information in text format. Earlier studies have been conducted regarding optimal line length - considered to be 50-60 characters per line including spaces [5]. If the line is too wide (around 70 characters and above) the user will have a hard time to focus on the text. This occurs because a long line length makes it difficult for the user to know where the line starts and ends. Too narrow lines break the user’s rhythm since the eyes have to travel back and forth too often. This, according to Lupton, tends to stress some users and can lead them to unwillingly skipping valuable information. Another variable to take into consideration is the fact that a user loses the focus along the line and therefore too long lines may cause a

\(^2\) All elements that are not visible without scrolling after opening a page are below the fold.
loss of focus. At the start of a line, and in the break between two lines, the focus is highest; at this optimal line length the user can maintain focus on the whole line and paragraph of text.

**Font and size** Lupton [6] has developed guidelines for choosing the right font size for screens. Users often have a larger distance to screens than to a magazine or book when reading. For comfortable reading on screens, a larger font size than used in print is required. A good starting point is 17pt[3][6] and from there go smaller or larger depending on web page requirements and importance of information.

The width of the column and font size are dependent up on each other. To achieve a good design layout and high usability, both the font size and the width have to match. A guideline is to have approximately ten words per line [7], and the narrower the column, the smaller the font size is required to be to maintain readability. Applying and working with the grid system implies submitting to the laws of universal validity according to Müller [7], such as the will to systematise and clarify, rationalise the creative and technical production processes, as well as the will to cultivate objectivity instead of subjectivity.

### 2.4 Time aspect

A study with 17 test subjects observing their daily use of the internet for four months resulted in visits on 17 242 different URL addresses [8]. 52% of all visits were shorter than ten seconds (median: 9.4s), 25% lasted for less than four seconds, and nearly 10% of all visits lasted for longer than 2 minutes. If only first-time visits were considered then the distribution over time did not differ much - the peak value for all new visits was still under 4 seconds. The results from this study are similar to, and partly based on, an earlier study conducted with 107 participants [9]. The time-aspect is important to take into consideration when designing a web page. The fact that a visitor generally will not give a web page as much of attention or time (often less than 4s) as needed for them to comprehend the whole web page sets high demands on the design. This indicates that the first impression is a crucial and a very important matter. The first seconds are the most critical since the average visit is less than that (9.4 seconds) [9]. Visitors are extremely sceptical to new web pages since they have previously experienced poorly designed pages. They are well aware of this probability when entering a page, and therefore try to minimise the time spent on poorly design pages, making the smallest design error increase the bounce rate 4 dramatically [11]. If the web page survives the first ten seconds the possibility for a longer session increases dramatically as the user might stay and have a look around.

---

3 In typography the point (pt) is the smallest unit of measure. It is used for measuring font size

4 The bounce rate is defined as the percentage of visitors who turn around at the entry page and immediately leave the site. Such visitors "bounce" out and never see additional pages [10].
2.5 Benchmark study

A benchmark was made at an early stage of this study to analyse the actual use of columns on web pages which focused on sales, marketing or e-commerce. Both Swedish and international companies were considered. Five well-known companies’ websites containing columns were evaluated to see how they used columns in their design: ICA, IKEA, Momondo, Kissmetric and Telia. The five companies were all using columns in one or more pages on their website and had between two and six columns in one row altogether. The most frequent number of columns were two, three and four (in no particular order), followed by five, and six was the least used.

3 Method

A study conducted in 2006 by Lindgaard et al.[12] forms the foundation of this study method. The study was performed using three different strategies to test the same thing - the first impression of websites. All three tests were executed one-by-one during the study. In each step, data were collected and through analysis a result was produced: how many milliseconds a designer has to make a first impression on a visitor. In the two first steps, the subjects were allowed to see the web pages for a predetermined amount of time (500ms) and rated the visual appeal of each page. Two different groups undertook this test. In the last step, the subjects repeated the same procedure as above and then were allowed to spend as much time as they needed to offer their opinion about seven paired design characteristics. The seven design characteristics were: simple - complex; interesting - boring; clear - confusing; well designed - poorly designed; good use of colour - bad use of colour; good layout - bad layout; imaginative - unimaginative. The characteristics were used to analyse what could affect the first impression, and a scale-based form was used for evaluation. In this study one coherent test was constructed, with two parts similar to the method that Lindgaard [12] used in 2006. First, the test leader read a case aloud to prepare the test subject for what type of web page they were going to enter and also to create a somewhat realistic situation with a spacious task. Further, the subject were shown one of the two websites for 11.4 seconds (adjusted after the pilot study, see 3.2) and after that verbally answer questions about the content of the page, both about

![Simple - Complex Characteristic-Scale](image-url)
the information (text), graphics and the layout. The interview after the first part of the test was based on both open-ended and closed-ended questions [13]. After the interview, the tested page was visible to the subject for as long as they needed to offer their opinion about the seven design characters. A survey with a 10cm scale belonging to each principle was filled out by the subject. With data from the first part of the test, a conclusion was drawn over which number of columns gave the test subjects the best possibility to perceive the web page. The analysis was made by giving the results different score and compare those scores against each other. With data from the second part of the survey conclusions about if the page with the highest visual appeal, according to the design principals, were then related to the page with the best result in the first test.

The tests, the questionnaire, the survey and the web page, were performed in Swedish. This method focused on quantitative data. 16 persons conducted the test, all students within five different departments at Umeå university Sweden, equal number of men and women. All used the internet on a daily basis. To simulate a realistic environment, the test was conducted in the field [14] in open areas where the subjects attention was divided with activity outside of the test. This environment was chosen since it represents the daily life of the thought test group.

3.1 The web pages

In this study web pages with three and five columns were tested. These specific numbers were chosen dependent on the benchmark which showed that two, three, four and five were the most common, and the fact that some were using six columns. Due to the limitations of this study and prognostication by the author determined that differences between two and three columns and four and five columns are too small to analyse with this chosen method, therefore the two and four column alternatives were eliminated. Due to the infrequently use of six columns, this too was eliminated. Two sketches of the web pages were constructed, see Figure 5 and Figure 6. The sketches were created concerning the subjects in the theory chapter and additional design elements to make the web page as realistic as possible. Similar to when performing a usability test, the test environment and the task should reflect the users normal use to get a reliable result[14]. With a realistic web page, the test subjects should be able to feel confident and recognise themselves in the task given; this will make them act as natural as possible and provide more accurate results. The web pages contained general objects found in web pages from the benchmark such as header, menu with links, call to action and background.

A set of two web pages with a different number of columns, width, and height were tested. The same amount of information, text, images and figures were shown in a different number of columns. The height and width of the columns differed depending on the number of columns. The alternative with three columns contained about 65 characters, approximately ten words per line, and had font size 17pt, see Section 3, Earlier related work, for an explanation. It contained three smaller icons and three headlines, one per column. In the second
alternative with five columns, one column contained about 35 characters in font size 15pt, approximately 6-7 words per line, five smaller icons and five headlines. The columns together in both alternatives had the same width in total.

The text-based information in the web pages was exact the same except from the headlines. There were two more headlines in the alternative with five columns, one of them were entirely new and the other one was constructed from splitting a headline from the alternative with three columns. In total, the difference in text between the two alternatives was one word. Further differences were in the number of small icons, there were two more icons in the alternative with five columns, this to maintain the web pages realistically. The web page and the text simulated a travel agent and the text contained general information to support the company’s offer. The language in the web page was Swedish.

Fig. 5. The web page used during tests with three columns.

Fig. 6. The column layout for the web page used during tests with five columns.
3.2 Pilot study

A pilot study was conducted to ensure that the intended method was viable before the real study began [13]. The pilot study contained two test subjects each testing one alternative to control the method addressing both alternatives, three and five columns. The aim of this pilot study was also to see if the time where the web page was visible to the subject was suitable for them to perceive enough information to be capable of answering questions about the content, and also to verify the case. From the pilot study, it appeared that the time was too short, and the test subjects had a hard time answering the questions. With this result, the difference between the two alternatives could be too small to measure with this method. By increasing the time to 11.4 sec, from 9.4, the test subjects were going to get a higher chance to perceive information and this was hopefully going to make the difference easier to observe. No changes were made regarding the case.

4 Result

The test contained two parts, one with questions based on the content and the second part where the subject considered and scored the web page by seven design characteristics. From the first part, the participants were able to perceive and reproduce the web page slightly better when using three columns, in total 3.5 correct answers with a mean of 0.4, compared to a total of 2.5 correct answers with a mean of 0.3, see Table 1. The result from the first part of the test with questions based on information in the text was relatively low. Few participants answered the questions correctly, and many of them had no clue. 10 of the 16 participants did not answer at all or wrong. The type value for both alternatives, three and five, was zero correct answer.

In the second part, the option with three columns generated a better result in all characteristics except one, imaginative - unimaginative, see Table 2. The characteristics were evaluated on a scale from 0.0 to 10.0, where 0.0 represents the best score. The web page with three columns got a better mean score at 24.0 compared to 29.7 regarding all characteristics. The characteristic with the largest difference between the two alternatives was clear - confusing, the web page with three columns was rated as more simple than the other one, even if both were considered as quite clear. There was also a noticeable difference considering the characteristic good layout - bad layout. The alternative with three columns received better scores, a mean of 2.6, compared to five columns with a mean of 3.5.

5 Discussion

The test conducted in this study did not only take into consideration the amount of information gathered from the text but also information found in other elements on the web page. Through that, the study still gives a good indication of
Table 1. The total for eight participants and the mean value for each participants of the correct answers for three and five columns.

<table>
<thead>
<tr>
<th>Questions</th>
<th>3 columns</th>
<th></th>
<th>5 columns</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8 participants</td>
<td>Mean value</td>
<td>8 participants</td>
<td>Mean value</td>
</tr>
<tr>
<td>About content</td>
<td>3.5</td>
<td>0.4</td>
<td>2.5</td>
<td>0.3</td>
</tr>
<tr>
<td>About design &amp; layout</td>
<td>23</td>
<td>2.9</td>
<td>20.5</td>
<td>2.6</td>
</tr>
<tr>
<td>Total</td>
<td>26.5</td>
<td>3.3</td>
<td>23</td>
<td>2.9</td>
</tr>
</tbody>
</table>

Table 2. The mean score of the eight tests for each characteristic and column alternative. A lower number means better result.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>3 columns</th>
<th></th>
<th>5 columns</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean value</td>
<td>Mean value</td>
<td>Mean value</td>
<td>Mean value</td>
</tr>
<tr>
<td>simple - complex</td>
<td>1.9</td>
<td>2.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>interesting - boring</td>
<td>4.3</td>
<td>5.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>clear - confusing</td>
<td>1.9</td>
<td>3.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>well - poorly designed</td>
<td>2.2</td>
<td>2.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>good - bad use of colour</td>
<td>4.6</td>
<td>5.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>good - bad layout</td>
<td>2.6</td>
<td>3.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>imaginative - unimaginative</td>
<td>6.6</td>
<td>6.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>29.7</td>
<td></td>
<td>24.0</td>
<td></td>
</tr>
</tbody>
</table>
how a visitor reacts considering the number of columns on the web page. There could be many reasons to why the participants could gather more information from the web page with fewer columns. The line length is one of them, in the alternative with three columns the line length is optimal [5] and could have had an impact on the result. It is hard to draw conclusions about the reading pattern referred to as the F-shaped pattern [3] and the findings of silent and muted regions on a web page [4] in this study since it did not contain higher technology as eye tracking. Despite, the fact that 10 of the 16 participants answered wrong or not at all could have a connection to the F-shaped reading pattern and the time it takes to fixate all regions of a web page. Since the columns were placed at the bottom of the page (to preserve a realistic layout) it may have taken longer time for the participants to fixated them and they may have read the text at the upper part first, according to Nielsen’s F-shape. From the result, it is clear that many of the participants did not read the text on the web page even if the case was constructed to encourage reading. Ten out of 16 did not answer correctly or at all on any of the questions based on the text. A common answer was "Oh, I did not read that. I saw a section about it, but I did not read it". Further, some responded with fictional information not found on the web page just to provide an answer, and some admitted they had no clue. During the pilot study the same issue was revealed, the participants did not read the text. With the intention to give them more time to read, the time was extended by 2 seconds. But the increased time did not have a major impact on the participants behavior related to their reading habits. The text was still left unread by almost 63% of the participants. But considering this as a model for how it works in reality when a visitor is entering a web page, the result also reflects the reality. A typical visitor does not read a lot of text, they rather look at headings, images, and icons and base their decisions on that. This conclusion is important to consider when designing a web page, does the web page reflect the company based only on the heading, images, and icons since this is what a user will have time to perceive before they decide to leave or stay within the first crucial 10 seconds [8, 11].

6 Conclusion

The result in this study indicates that the web page with three columns makes it easier for the visitor to grasp and perceive information, both information placed in the text and design elements. The participants who used the web page with three columns could answer the questions better. They had gathered more information, and also, they ranked the web page better considering all the characteristics except one, see Table 2. From the result of this study a guideline regarding the number of columns can be formed; With fewer columns the visitor will both have a better chance of perceiving the information, and they will also obtain a more positive attitude towards the website which could lead to a positive attitude towards the owning company.
References

Paper vs. Simple Digital Prototypes: Differences in User Feedback when Testing

Hanna Söderström
Department of Applied Physics and Electronics
Umeå University, Sweden
hasa0063@student.umu.se

Abstract. Prototyping software is becoming more accessible and can replace the well established paper prototyping method. Differences in user perception when testing paper prototypes vs. simple digital prototypes was investigated in this paper. Research was carried out in the form of twelve usability tests with two mobile application prototypes; one paper and one digital tested in a somewhat controlled environment. Findings from the research and post questionnaires showed that the paper prototype was given a higher score from users (paper 75.025 and digital 63.3) when evaluating the usability. The paper prototype had more interactions and comments during the usability test than the digital prototype. The results imply that the paper prototype was perceived more positively by the users with an increased willingness to collaborate with the prototype.

1 Introduction

When a new software product evolves, or when an existing product is in need of an update and redesign, the designer and the development team usually keep to a design method. This method is used to create an idea, evaluate this against users, and at the end of the process develop a final product. Paper prototyping has been a part of this method for many years and has helped to evaluate ideas during the early design phase [1, 2, 3]. The design method usually includes creating a rapid prototype, regularly drawn on or made out of paper but sometimes made digitally, to quickly generate ideas and evaluate these against the target group and users.

During the last few years, a considerable amount of different web applications and software for prototyping has evolved to facilitate the process, and there are now applications for all different purposes and fidelity levels. Some examples of these programs include Balsamiq, Axure RP, POP, InVison, Marvel and Pixate. These applications let the designer create digital prototypes with different fidelities. The prototypes can take the form of a pen and paper prototype all the way to a fully functional prototype with the right functionality, look and feel. There are studies and literature focusing on the effect of prototyping and the benefits of testing a product at an early stage and with different fidelities [1, 2, 3]. The average user has daily involvement with computers and the Internet according
to Statistics Sweden [4, 5]. In 2015, 89% of the Swedish population between 16 - 85 years old had access to the Internet. More than 75% use the Internet daily and 70% use a computer daily. A paper prototype can be far less intimidating than a digital prototype and can also eliminate the risk of feeling foolish during a test session according to Snyder [1, p. 57]. This paper will investigate if the use of simple digital prototypes can be received the same as well-established paper prototype, without losing any valuable feedback, satisfaction or collaborative willingness from subjects.

1.1 Objective

The objective of this study is to identify and analyse differences in user feedback and user satisfaction when testing simple paper and digital prototypes. The two types of prototypes used are paper prototypes and screen based prototypes. This study is an attempt to determine any differences between the two test options and to distinguish if any option is more favourable when it comes to:

- The amount of qualitative user feedback.
- Subject satisfaction regarding the application.
- Willingness to collaborate during tests with prototypes.
- Attitude towards the prototype.

2 Previous work

Previous work within this field has been carried out with slight differences and aims compared to this study’s goal. A study from George Mason University [6] compares two identical prototypes made from a Photoshop image, one print out and one using POP\(^1\) for realistic smartphone navigation. The study looked at finding usability issues when testing this on a group of people and claims that there are no significant differences between the two methods. Another study, comparing low- vs. high-fidelity prototypes, found that both prototypes are equally effective at finding usability issues [7]. A study from University of Magdeburg and University of Berlin [8] proves that a sketch like rendered image brings up more creativity and discussion from test persons than a wireframe, or shaded image does when it comes to architectural buildings. Usability issues are not this study’s aim; the focus is on how the subjects feel and their willingness to contribute to the usability test.

3 Theory

To clarify the terminology used in this paper, short explanations and definitions will follow below.

\(^1\) www.popapp.in
3.1 Paper prototype

A prototype is a designer's or developers' first attempt to examine, communicate and demonstrate ideas to various stakeholders. Paper prototypes are being used in various fields, but this article will refer to a paper prototype as a prototype of a digital system. A paper prototype can have different fidelity levels depending on content. The look, for example, can offer low fidelity while the depth of the prototype can be classed as high [1, p. 259 -260].

A paper prototype is a simple prototype built from cheap materials including cardboard, paper, post-its and similar materials. The aim is to investigate different options and evaluate the conceptual design rather than graphical design and to be used as a support when communicating ideas between team members and users [2]. These prototypes are often hand drawn but can also be created by using a prototyping software. The paper prototypes can be displayed on paper or as a simple sketch on a screen with added interactive fields to create realistic navigation between existing screens. Paper prototypes are used to carry out usability tests at an early stage to gain valuable insights from the user. Below some advantages of paper prototypes are listed according to Snyder [1, p. 12].

- Provides sustainable user feedback early in the development process.
- Provides rapid iterative development. It is possible to experiment with many ideas.
- Facilitated communication within the development team and between the development team and customers.
- Do not require any technical skills, so a multidisciplinary team can work together.
- Encourages creativity in the product development process.

3.2 Digital prototype

A prototype made digitally can be used in a similar way as a paper prototype and shares the same objectives as paper prototypes [9]. The digital prototype, however, can range between narrative low-fidelity to a high-fidelity with final graphical elements and interactions. The different types of digital prototypes are used during different stages of the design process. Choosing which digital prototype to use depends on the main area to evaluate and where the product is in the design process. A screen based prototype made in a prototyping software is sometimes referred to as medium fidelity when it comes to interactions and looks, but in this study, the used digital prototype will be designed with the same fidelity levels as the paper prototype.

3.3 Fidelity

Levels of fidelity can sometimes be misused and applied to classify a single prototype as a whole rather than the current prototype's content [1, p. 259]. A short explanation will follow to clarify the definition used in this paper. In
a prototype, there are different prototype contents to take into consideration; depth of the prototype, length, interaction, navigation structure, look, etc. All these aspects can, for example in a paper prototype, range between different fidelity levels depending on the content used [9, p. 85-95]. An example would be the aesthetics of the prototype that most probably will hold low-fidelity in comparison to the depth of the prototype that can range from a medium to a high-fidelity level. The two prototypes used in this article will both hold a low-fidelity look with medium-fidelity interaction and depth. The ambition is to present these prototypes with the same fidelities throughout the sessions to be able to use the user feedback for comparison and evaluation.

3.4 Usability test

A usability test can be conducted at any time during the design and development phase with the help of a prototype, preferably started in an early stage. The purpose of a usability test is to collect the user’s feedback and identify any pitfalls, errors or mistakes to improve the overall usability. During a usability test real users will evaluate the prototype with sessions usually carried out in a natural setting or sometimes in a controlled environment. Collected feedback will be evaluated, and the prototype updated before a new session of usability tests are carried out. The usability test can be seen as successful if the test helps to improve the product for the better [10].

4 Method

The findings presented in this report are based on research collected from 2016, Umeå, Sweden. The test was conducted on twelve participants, divided into two groups of similar size. These groups had more than five subjects each, which is, according to Nielsen [11, 12], enough participants for a usability test. The participants age were between 17 and 56, and an equal number of men and women participated. The language used during the usability test was the local language Swedish. The test was a 10 to 20-minute long one-to-one session where the subject received information about the test, what the task would be and that the sessions would finish with two different questionnaires. The questionnaire followed the standard of The System Usability Scale (SUS) [13, 14].

All subjects were first presented with a verbal explanation about the session followed by a verbal description of the tasks to conduct. Throughout the session, the subject was encouraged to think aloud to communicate thought, feelings and assumptions to the test leader. This method also gave the test leader information to collect and later, evaluate. The test leader looked carefully at attitude when speaking, eagerness, tone in voice and more to make a conclusion on how the subject felt during the session. All communication during the session was recorded and categorised to estimate how many different comments the two groups made on average. To distinguish any differences of the type of feedback given during
the two tests, the recordings were analysed to categorise the comments into positive and negative statements. The total amount of remarks, amount of times when the subject said "This is good/fun/easy" or "This is hard/difficult" and how many times the test leader had to remind the subject to press the buttons on the prototype was also categorised. Due to the nature of the test persons' comments, the categorisation of the collected data was an attempt to see any signs of differences between the two groups and used to support the collected feedback in the questionnaire.

4.1 Used prototypes for test session

Two similar prototypes were created, one made out of hand-drawn sketches on paper, Fig. 1, and one screen based prototype designed with software Axure RP\(^2\) to get a sketched feeling and appearance, Fig. 2. The two prototypes had the same text, graphics, and interactive depth and the only difference was the medium used to present the prototypes. In this test, an Android smartphone was used during the test session to display the digital prototype. The model was a OnePlus One from 2014 running on Android 6.0.1.

![Fig. 1. Examples of the hand drawn application interface.](image)

The application used in the test session is an application delivered by a national coffee shop, available on Google Play\(^3\). This application was selected due to its interactive depth, and somewhat unclear navigations and includes both

\(^2\) www.axure.com

\(^3\) https://play.google.com/store/apps/details?id=com.mobivending.espressohouse
a loyalty card, inspiration and information about coffee. The thought was that most participants would not have used this application before the test session to minimise any recalls from previous app usage.

The paper prototype consisted of 16 screenshots to cater for the use cases the user was told to perform. The look of the paper prototype is somewhat rough with hand drawn lines and handwritten text and can, therefore, be classified as low-fidelity. The size was the actual mobile size to represent the interface as accurately as possible. The digital prototype created in Axure RP had the same amount of pages and graphic look to mimic the paper version, and the aim was to make these as similar as possible. The lines were consciously made uneven and a handwriting font style used to again, create a hand drawn feel.

4.2 Use case

During the test sessions, two test scenarios, and a prototype was presented to the current subject. The subject did not receive any information about the fact that the test had two different groups, testing two different types of prototypes to collect adequate data without interfering with knowledge about the second type of prototype. The data and questionnaire answers collected from the usability tests could, therefore, be compared.

All users were presented with two use case scenarios with related tasks to complete. During the execution of the task, the test leader took notes about the numbers of comments, if it was a positive or negative comment and how well the subject accomplished the given task. The comfort of the subject, tone in voice and other, small signals to identify the subject’s mood was observed. All test sessions were conducted in a somewhat controlled area to minimise disturbing noise and other disturbances.
4.3 Questionnaire

To complement the usability test with quantitative data, questions were given both before and after the session. The method chosen in this study is a standardised usability questionnaire called System Usability Scale, SUS [13, 14]. SUS is based on a Likert scale with ten predefined question, all with five answer options each varying from "Strongly Agree" to "Strongly Disagree". The questions were translated into Swedish for the test session due to the subject’s native language. For clarification, the use of "system" in the questionnaire is replaced in the standardised SUS with "application" [15]. Along with SUS, a second questionnaire was presented to the subject after the SUS to cover how the user felt during the test session. SUS was given first to eliminate the risk of affecting subject’s thoughts and feelings.

The SUS questionnaire covers the ISO standard of usability which includes effectiveness, efficiency, and satisfaction [14, p. 190]. The focus of the second questionnaire is how the subjects felt during the test and their attitude towards the prototype rather than the prototype and application itself. The questions where (1) Did you give ideas of new solutions for the application during the test session? (2) Did you give suggestions for improvement for the application during the test session? (3) How much do you think your ideas and suggestions mean for the application in the future? (4) What attitude did you have towards the prototype during the test session? All questions had five-grade Likert scale to use when answering the question.

5 Results

A mean SUS score, according to the established method [14, p. 194], will be presented here. The results of comment type, subject attitude and other findings will also be presented in this section.

5.1 System Usability Scale (SUS)

Data collected from the SUS showed a difference between the two types of prototypes. The paper prototype scored a mean of 75.025 in the SUS. The digital prototype scored a much lower score, a mean of 63.3, even though the prototypes were very similar except for the medium used to present them. The paper prototype was perceived to have better usability than the digital one.

5.2 Post SUS questionnaire

In the post-SUS questionnaire, the question "What attitude did you towards the prototype during the test session? / Vilken inställning hade du gentemot prototypen under testets gång?" differed the most and stood out from the rest. The two test groups showed that the subjects had a more positive attitude towards the paper prototype, Fig. 3, than towards the digital prototype, Fig. 4. The other three remaining questions in the post-SUS questionnaire did not show any remarkable differences.
Fig. 3. Scoring the paper prototype after the test, the answers ranged between (1) Very Negative and (5) Very Positive. (n = 6).

Fig. 4. Scoring the digital prototype after the test, the answers ranged between (1) Very Negative and (5) Very Positive. (n = 6).
5.3 Types and amount of comments

Nine out of twelve usability tests were recorded and from these nine recorded sessions, a count of positive and negative statements, how many times the subjects expressly said "This is good/fun/easy" or "This is hard/difficult" was made. In addition to this, the number of times the test leader had to remind the subject to press on the prototype to move on to the next page was counted.

<table>
<thead>
<tr>
<th></th>
<th>Paper prototype</th>
<th>Digital prototype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive comments</td>
<td>18.8</td>
<td>14.25</td>
</tr>
<tr>
<td>Negative comments</td>
<td>21.2</td>
<td>17</td>
</tr>
<tr>
<td>&quot;This is good/fun/easy&quot;</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>&quot;This is hard/difficult&quot;</td>
<td>1.2</td>
<td>0.25</td>
</tr>
<tr>
<td>Reminder to press on buttons</td>
<td>1.6</td>
<td>0.75</td>
</tr>
</tbody>
</table>

The table above, Table 1, shows that the subjects in the paper prototype test gave, on average, 4.55 more positive comments and 4.2 more negative comments than subjects from the digital prototype test session. The number of comments regarded "This is good/fun/easy" and "This is hard/difficult" showed that subjects from the paper prototype session on average had twice the amount of positive statements and 4.8 times more negative statements than people from the digital prototype test session. The test leader reminded the subjects to press on the prototype buttons in order to move to the next screen on average 1.6 times per subject in the paper prototype session and 0.75 times on average during the digital prototype session. During the session, it was noticed that subjects testing the paper prototype read the application text more carefully than in the test with a digital prototype. No number or quantitative data was collected and it is only based on the test leaders perception.

6 Discussion

The result from the study can be compared to the article from Schumann et al. [8] who claimed that more creativity from test persons was found when presenting a rendered sketch image to clients rather than a wireframe or shaded images. It might, therefore, be beneficial for designers to consider using a paper prototype in the design phase to increase creativity and willingness to collaborate in the process.

The SUS score presented from the two prototypes gave a clear indication that the paper prototype was perceived as more usable than the digital one. This result may be because when a user holds a phone displaying a digital prototype, it is easy to compare it to a fully functional application without even realising it which affect the users thoughts and feelings about the prototype. The
paper prototype’s simplicity could reduce the likelihood of being subconsciously influenced by the knowledge of how a smartphone application is used, making it easier to separate the paper prototype from a more developed prototype or fully functional application.

The subjects testing the paper prototype became more creative than in the digital prototype tests and some even started to move the paper pieces by themselves instead of letting the test leader run the prototype. Results of the study could imply that the paper prototype improved the subject’s creativity and collaboration willingness. A higher number of comments, both positive and negative were received from the paper prototype sessions compared to the digital prototype test sessions. Some subjects had minor problems to navigate the digital prototype due to the prototype’s slow response.

Results indicate that designers should be wary in trusting new technology and should consider to stick with well-established methods for the best outcome. Findings from this research should be seen as indications on prototype choice and a mix of different mediums and fidelities might be in favour to cover all aspects of evaluating a product.

6.1 Limitations

No information was given about the prototype or application before the usability test started. In hindsight, it could have been mentioned that the test leader did not design the application, to remove the element of apathy from the subject.

Due to technical issues, three usability tests were not recorded and therefore could affect the accuracy in terms of the count regarding the different types of comments. Five paper prototype tests and four digital tests were documented and analysed. The classifications and quantity of the different comments were difficult to determine and should, therefore, be seen as a complement to the rest of the results.

The lack of observers during the test session made it difficult for the test leader to both run the prototype and analyse the subject completing the task. Due to this, small but meaningful signals might have been missed.

7 Conclusions

This study aimed to find differences between a paper prototype and a simple digital prototype. When adding up the results from the SUS, the type of comments, how the subjects acted and their attitude towards the prototype, it is clear that the paper prototype gave a better result than the digital prototype. The paper prototype gave more qualitative feedback than the digital one and the users rated the paper prototype as more positive than the digital. The fact that the paper prototype had a higher SUS score and was perceived as more positive could mean that the subject would be more eager and willing to collaborate during a usability test session even though during the test sessions it was difficult to distinguish any major differences between the two groups. The
subject satisfaction can be read from the SUS score and because the paper prototype scored a higher SUS than the digital prototype, it might be classed as more satisfying, at least when it comes to usability.

References


Guidelines for designing context-aware fitness applications

Julia Westman
Department of Applied Physics and Electronics
Umeå University, Sweden
id12jwn@cs.umu.se

Abstract. This paper has investigate what kind of information that is relevant to display in a context-aware fitness application. This was researched through a web-based survey with 29 individuals who responded to it. The survey was divided into two different parts. The first part was about fitness and job conditions and the second part was about information and features. The Participants were divided in two groups. One group for those who worked and one for those who studied or were unemployed. The findings from the research included that those who work in an office environment or study easily become sedentary during their day. It also includes that a context-aware fitness application should follow some guidelines to make the user more motivated to exercise. The guidelines that was found was for statistic, context, encouragement, notifications and community.

1 Introduction

According to a study by Alter, we become increasingly sedentary, both in our work environment but, also in everyday contexts of our life. Sedentary means that that you perform activities that do not require physical exertion, for example sitting down or lying down. This may contribute to health problems such as diabetes, overweight and in the worst case cancer [1]. To reduce the time that we are sedentary can we motivate the user to perform some kind of physical activity. This could be done by using a fitness application that is using the method of ubiquitous computing and the concept of context awareness [2]. This kind of application can help users become more motivated, by knowing how far a user has moved, where the user has moved and also how many calories that the user has burnt while it has performed the physical activity.

The relation between a user and a computer has changed during the evolution of technology, at first it was many people per computer, and then it was one person per computer and now it is many computers per person. The main idea of ubiquitous computing, is to focus on relation: many computers per person [3]. The word ubiquitous means "existing everywhere at anytime" [4] thus ubiquitous computing means that a wireless device is completely connected.
and constantly available [3]. Context-awareness is one of the main areas of ubiquitous computing [5] and it is defined as any information that a device can use to characterise the situation of an entity. An entity can be a person, place or an object. Thus context awareness means that a device can use information from the environment to provide task-relevant information or services to a user [4].

Wireless devices can transmit their location and status over the network when they are in a context-aware environment. The transmission can be done through various sensors that will capture the task-relevant information and then displays it to the user through the screen. For example, if a user has a daily goal to walk five kilometer and reached the goal, then a context-aware application can notify the user with a user message on the screen.

1.1 Objective

The purpose of the study was to answer the research question: What kind of health and fitness information is relevant for the user to receive from the environment to make the user more motivated. The question was asked to identify guidelines for how to design a fitness application that use the concept of context-awareness.

2 Theory

This section gives a brief introduction to sedentary, context, context-awareness and ubiquitous computing.

2.1 Sedentary

According to Warburton we are becoming more sedentary in today’s society, both during the time when we are working, but also in our own time and our leisure [1]. That means that we spend more time sitting still, other than to perform some sort of physical activity. Which means that the risk of health problems such as weight gain, metabolic syndrome, diabetes, and heart disease increases [6]. Sedentary behavior means that you perform activities that do not require physical exertion and some examples of these activities is sleeping, sitting, lying down, and watching some kind of screen-based entertainment [6].

2.2 What is Context?

The word context can simply be explained with Deys definition: "Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and application themselves" [7].
Thus, if some kind of information can be used to represent and characterize a situation in an interaction, then that information can be used as context information [8]. An example of how to use the context is if you want to know how many degrees it is outside, then you can by using your location find out the weather at the location.

There is plenty of information around us all the time, which is considered to be used as context information [9]. Some examples of such information:

- Environment information, e.g. temperature, light-level or noise-level.
- Social situation, e.g. who you are with and who is nearby you.
- Spatial information, e.g. location, orientation, and speed.
- Temporal information, e.g. time of the day, date, and season of the year.
- Identity, e.g. who you are as a user.

There are basically four certain types of context, that are more important. These are identity, time, location and activity. These types of context responds not only to the questions of what, who, when, but can also can be used together as other sources of contextual information [10].

2.3 What is context-awareness?

In 1994 Schilit and Their discussed for the first time Context-aware computing, they meant that it was a software that "Adapt of according to its location of use, the collection of nearby people and objects, as well as changes to Those objects overtime" [11].

Dey and Abowds definition of Context-aware computing, means that context-aware computing is "any information that can be used to characterise the situation of an entity, where an entity can be a person, place or physical or computational object. They also mean that context-aware computing is the use of context to provide task-relevant information and/or services to the user, wherever they may be" [10]. Thus Context-awareness means that a device knows its own location and the environment around it. That means that a system that is using the concept of context-awareness can interact with the environment and also interpret and utilize context information [10].

Where you are, who you are with and what the resources are nearby, are three important aspects that a context-aware system is using. Thus the context includes more than just the user’s position [10]. There are also three important behaviours in context-awareness that an application can be identified with [13]:

- The automatic execution of a service.
- The presentation of information and services to a user.
- Labeling of context information for later retrieval.
There are two ways for an application to use the context of the environment. One way is to use the context active and the other way is to use it passively [14]. These two ways can simply be explained by Chen and Kotzs two definitions [14]:

"Active context awareness: an application automatically adapts to discovered context, by changing the application’s behavior."

"Passive context awareness: an application presents the new or updated context to an interested user or makes the context persistent for the user to retrieve later."

2.4 Ubiquitous computing

During the last 50 years of computer science the interaction between a computer and a human has been relevant for the evolution of technology. During these 50 years, there have been three waves of technology. The first wave focused on the relationship between many people and one computer. The second wave focused on the relationship between one person and one computer. The last wave also known as ubiquitous computing, focused on the relationship between one person and many computers [3].

According to Rainie and Wellman ubiquitous computing can sometimes be called “Internet of things” and can be described as human-computer interaction that goes beyond personal computing to an environment of object processing information and networking with each other and humans [12].

Ubiquitous computing is a part of different research areas. Some examples of these areas are: including distributed computing, mobile computing, context-aware computing, location computing, sensor networks, mobile networking, and human-computer interaction [5].

3 Method

A web-based survey was designed to find guidelines for how to design a context aware fitness application. The target group for this research was a person that was between 18 - 65 years old and used different kinds of applications on their smartphone. The survey was disseminated through Facebook. This decision was made to get in touch with the right people who fit into the target group.

The expected result of this study will be answering the question:

– What type of information would the participants want to know from a context-aware fitness application, when they are working out?
3.1 Survey

The survey was divided into two parts, one part about job conditions and fitness. The second part was about what kind of information and features that should be shown in a fitness application. The first part was also divided into two different response groups, those who work and those who are studying or was unemployed. This decision was made so that the questions would be more suited for their circumstances.

The survey included 21 questions about occupation, general health, and what information they think should be displayed in a fitness application. The survey was comprised of different types of questions, such as: alternative questions, yes or no questions followed by open questions and scaling questions.

The choice to yes or no questions was made to get a quick view of the answer. Depending on what the person answered on the question, the person would get a supplementary question that is an open question. The choice to include open questions in the survey was made to get a more freely answers from the participating.

The choice to have scale questions was made to get an overall measurement of sentiment around a particular topic, such as what time of day they are most likely to perform a light physical activity.

The likert scale of the scale questions was made in a scale from one to six. In this way, the person had to have an opinion, and and could not be neutral because they and could not choose the midpoint of the scale.

Test group The survey was sent to 40 persons through Facebook. Those who received the request to participate in the survey, first got a question if they had a smart phone that they use active. So they was using different applications. If the answer was yes then they belonged to the target group for the survey.

4 Results

The result from the user study was divided into two parts, one part about job conditions and fitness. The second part was about what kind of information and features that should be in a context-aware fitness application.

It was 29 persons that responded to the survey. 10 of them was students, 18 of them had a job and one of them was unemployed. The age of the participants ranged from 18 to 65 years old, but the majority of the respondents were between 26 - 35 years old.

4.1 Part one: Job conditions and fitness

The first part of the survey was divided into two different response groups, the first one is: those who work and the second one is: those who are studying or doing something else.
Group one: participants that was working  Those who participated in the survey worked for example with nursing care, industrial, IT and finance. Of those who worked 29% felt that they easily become sedentary during the day while 35% did not think they were sedentary during their work day. The difference between the two groups is showed in table 1.

<table>
<thead>
<tr>
<th>Type of job environment</th>
<th>Answered between 1-3</th>
<th>Answered between 4 - 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Office environment</td>
<td>33 %</td>
<td>67 %</td>
</tr>
<tr>
<td>Remaining environments</td>
<td>78 %</td>
<td>22 %</td>
</tr>
</tbody>
</table>

Table 1. The result of group one that answered on the question If they feel that they easily become sedentary during there workday. If they answered between 1-3, then it means that they did not feel that. But if they answered between 4-6 then they did feel that.

When the participating in the study would estimate how many hours they sit down during a normal working day. The result was that 29% estimated that they sit down for more than five hours during the day and 47% estimated that they sit down between 1-5 hours. While 29% estimated that they sit down to less than 1 hour during the day. This is shown in Figure 2

The participants was asked to respond on a scale from 1-6 how likely it was that they would perform a light physical activity, during certain times of the day. Where one was unlikely and 6 were likely. The times they would consider were: before they get to work, during their lunch break and after work.

9 persons responded between 1-3 on the scale, and 9 persons responded between 4-6 on the scale of how likely it is that they perform a light physical activity before they get to work.

The majority of participants felt that it is not likely that they will perform a lighter physical activity during their lunch break.

They were five people who felt that they are not likely they will perform a physical activity after their job, while they were 13 that would likely perform a physical activity after their job.

Group two: Participants who was students or was unemployed  All who participated who study or was unemployed felt that they easily become sedentary during the day. (See Figure 3) Of those it was 82% who estimated that they sit down for more than five hours a day.

When participants were asked to respond on a scale from 1-6 how likely it was that they would perform a light physical activity, during certain times of the day. Where 1 was unlikely and 6 were likely. The times they would consider were: during the day and during the night.

One person replied that it was not likely that it would perform an activity during the day, while ten persons felt that it was unlikely that they would perform a physical activity during the day. When they were asked how likely is it that you
perform a physical activity in the evening, answered ten people that it is not likely that they will do it. While one person replied that it is likely they do.

4.2 Part two: information and features

Of the 29 persons that was participating in the study, had the majority of them used a fitness application earlier. Some of the applications that they had used were: Runkeeper, Apple’s Health app, miCoach Adidas, Vivo fit, Google Fit, Endomondo, Fitbit and FunBeat. Five persons of those who participating in the study had never used a fitness application. Some of the reasons was that they do not exercise or that they do not think it is a necessary application.

When the participants were asked if they would feel more motivated if they knew more about their training and their training results? Majority answered that they would become more motivated by knowing. Some things that they would like to know about their training was: How far they have left to reach their goal, see how their training went, how much calories they have burned during the exercise, improvements, average speed, top speed, track of weight and track of steps.

The participants were asked if they would feel more motivated by receiving notifications when it is time to work out again. 17 persons felt that they would not be more motivated by receive notifications while 12 people felt that they would feel more motivated. Six individuals would like to have notifications each day, four persons would like to have notifications a few times a week and two persons were like to have it within a range that they can decide. A summery of this two questions and the result showed in table 2.

The majority would not be more motivated by see their friends activities. While nine individuals felt that they would be more motivated by seeing that.
Would you feel more motivated if you knew more about your training and training results?

<table>
<thead>
<tr>
<th>Question</th>
<th>Answered Yes</th>
<th>Answered No</th>
<th>Required functionality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Would you feel more motivated if you received notifications from your smart phone, when it is time to work out again?</td>
<td>41 %</td>
<td>59 %</td>
<td>Some wanted notifications every day, some wanted notifications a few times a week and some of them wanted to decide for themselves.</td>
</tr>
</tbody>
</table>

Table 2. The result of questions about information and required functionality from the user study.

Some reasons to why they would feel more motivated was: that it would feel like a competition between you and your friends and in that way they would feel forced to exercise more. Another reason was that they would feel "if others can, so can I".

The majority also felt that they would be more motivated to see statistics about their performed activities. Other things that the participants felt would motivate them was if you get some kind of reward from the application, when you become better or when you achieved your goals. A summery of this two questions and the result will be showed in table 3.

<table>
<thead>
<tr>
<th>Question</th>
<th>Answered Yes</th>
<th>Answered No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Would you feel more motivated by seeing your friends past activities?</td>
<td>38 %</td>
<td>62 %</td>
</tr>
<tr>
<td>Would you feel more motivated if you saw the statistics of your activities?</td>
<td>86 %</td>
<td>14 %</td>
</tr>
<tr>
<td>Would you feel more motivated by knowing that an application knows things about the context?</td>
<td>65 %</td>
<td>35 %</td>
</tr>
</tbody>
</table>

Table 3. The result of questions about information and required functionality from the user study.
5 Discussion

It is not news that there is now a strong focus on how long and how often we exercise. Today’s technology and work situations mean that we easily become sedentary and not moving as much as we should. However, it depends entirely of what type of work you have.

The study found that if you, for example work with health care, then you are very active during your workday, which means that you do not sit down that much. But if you work in an office environment or study at the university, you will sit down large parts of your day. Which means that there are those kind of people who needs more daily exercise.

The survey also showed that the majority had used some of any kind fitness application. Which was very positive, when it is that type of user that knows what kind of information that is interested to know more about and what is not. The study showed that majority of the participants felt that a context-aware fitness application should be able to:

- View information about a the users training and their results.
- View statistics of past activities, weight and number of steps.
- View context information such as distance, maximum speed and average speed.
- Provide some sort of reward to the user when they reach their goals and become better.
- Give praise when you’ve done something better than before.

If an application would have all these features, the user would feel more motivated to exercise and will perform the daily exercise that they need.

To make the application complete and satisfy a wider audience, some other some features should be added. For example to get notifications about when it’s time to workout again, or get view friends past activities. These features would be there to make the user feel more compelled to actually exercise more.

However, these functions should be optional to choose to use. When some users thinks that they become irritated by them.

In further research on the subject it would have been interesting to align the target group towards the younger generation, because I feel that they easily become sedentary since today’s technology is a big part of their lives. Another thing that would be interesting for further research is to perform user testing on any existing context-aware application.

6 Guidelines

The conclusion of this study is that a context-aware fitness application, should follow certain guidelines to fulfill their purpose to motivate their users to take more exercise.
**Statistics** A context-aware fitness application should show information about the users past activities. The information shall include information about the user improvements, max speed during the activity, the average speed during the activity, and calories they burned during the activity.

**Context** A context-aware fitness application should know about their environment and context. In order to present the relevant information to the user. So as the total distance, how far it is to the destination, the user’s location, and the speed of the user.

**Encouragement** A context-aware fitness application should give the user some kind of reward when you have become better, such as points or different awards depending on what you have become better at. Because when the user receives some kind of reward then the user would feel proud of them self.

**Notifications** A context-aware fitness application should give the user an opportunity to choose to use notices when you think it’s time to work again. Through all users are different, some feel more motivated when they are reminded that they should exercise while others get annoyed by the pop up notices all the time on their smart phone, that’s why it should be optional.

**Community** A context-aware fitness application should give the user an opportunity to choose to connect to their friends, so that users can see their friends’ past activities. In order to be inspired by their friends and feel able to work out. But there can also be a negative effect of this, when it becomes an irritation for the user. That’s why it should be optional.

**References**


The role of feedback in educational applications for smart devices

Lena Lindberg
Department of Applied Physics and Electronics
Umeå University, Sweden
oli2len@cs.umu.se

Abstract. New technologies are affecting the way of teaching in primary school. Today iPads and educational applications are growing in popularity and is commonly used in schools. These new technologies have a potential beneficial effect on the learning, but studies show that some applications have flaws that hinder these positive effects, for example regarding the feedback provided to the user. Since feedback is an important part of the learning process, it is of great importance that the educational applications used in schools manage to provide feedback that support the learning process. This study has determined a number of guidelines for good feedback in a learning situation and further showed how well these guidelines can be reached in an educational application today. The study showed that there are both advantages and disadvantages of feedback in an application compared to giving feedback in a more traditional learning situation. Many of the guidelines for good feedback were found to be possible to reach in an application, and the greatest advantage found in applications is that they can provide feedback timely on each action taken by the pupil. The most critical disadvantage found in this study was that an application can not provide any conversational feedback.

1 Introduction

As the technology evolves, the way of teaching is changing and new technology becomes part of the every day classroom situation. Since tablets became available they have steadily increased in popularity in primary schools as a teaching material. Today there is a large number of applications designed for an educational purpose and to be used in primary school, and the number of available applications is constantly growing. According to Apple [1] there were 1.5 million iPads used in education and over 20,000 educational applications available in 2012.

According to a study on how tablets are being used in education [2], most teachers explained that “Using iPad increases the students motivation” and “The students are able to comprehend more information” as the two major positive aspects of using tablets in their education. Their findings showed that the use of tablets also lead to time saving, efficiency and increased mobility and flexibility. A recent study [3] with focus on students and teachers experiences of applications
showed that both groups saw that these technologies could for example make education more fun, collaborative and engaging. One of the greatest prospects in the use of tablets and application is that they can help improve the students motivation. Both teachers, students and scientists see a great potential in the use of tablets, but studies also show that this potential is not fully actualized [4–6].

1.1 Problems

According to a studie by Vaala [6] where the highest rated and most popular literacy apps for children were studied, it was found that the most prominent applications are missing key teaching opportunities. Another evaluating study [5], show that the design of educational applications sometimes seems to focus more on superficial aspects than on supporting deeper learning.

One identified reason for why the applications do not fulfill their educational purpose is that many suffer from shortcomings in the feedback provided to the user. Since feedback is an important part of the learning process, the feedback has a deep impact in how well an application can support learning.

In a study made by Blair [7] preschool math-apps were studied with focus on the provided feedback. This study showed that in 82 % of the cases a so called “Answer Until Correct (AUC) feedback” was used. This kind of feedback means that the program indicates that the answer is incorrect by for example a red color, buzzing sound or similar, and the user must keep trying until the right answer is given. Further this report implied that the most frequently used AUC feedback is not the most effective feedback to support learning, and suggest that the learning could be improved by using different sorts of feedback that gives more information than only showing if the answer was right or wrong.

The fact that many educational applications have shortcomings considering the feedback provided to the user is also supported by Agneta Gultz, a Swedish scientist and leader of the The Educational Technology Group at Lunds University. According to her studies only about 2-3 percent of the educational applications used today gives enough feedback to support deeper learning [8]. She claims that the majority of applications used today are rather a way of testing students current knowledge than supporting learning, and that some applications could even have a negative effect on the students learning because of their lack of feedback.

1.2 Objective

This study will focus on feedback given to the user in educational applications. The goal of the study is to determine what ”good” feedback is and how this kind of feedback can be reached in educational applications, in order to ensure that the feedback truly contributes to the learning process. The following research question has been formulated:

- What defines good feedback in a learning situation and how can these theories be applied to feedback in educational applications?
This study will focus on children in first to third grade, since educational applications are being used more frequently in early primary school, than among older pupils.

2 Theory

In this article the concept of feedback is defined as information given about how someone is doing in their efforts to reach a certain goal, in this case as information given from an agent (a teacher or an application) to the user (the pupil) regarding aspects of the user's performance or understanding of a task.

2.1 Feedback and learning

Feedback is an important part of the learning process [9,10]. Feedback can help pupils through providing motivation, enhance learning, enhance reflection and as a performance indicator [11]. Latham [12] explains that feedback is an important and necessary step in the learning process since it can give information and explanation beyond the original task or question that the pupil is dealing with. This means that feedback is a way for the pupil to gain a deeper understanding and a wider knowledge than what is possible without feedback. Latham mentions so-called instructive feedback, referring to feedback that gives extra information in response to the student's work, and states that this sort of feedback produces "quantifiable strides in learning".

Hattie [10] has made an analysis in order to find rate numbers for the effectiveness of different influences on student achievement, such as attributes in classrooms, teachers and curricula. By comparing the rates of these different influences it was shown that the average effect size was 0.4, while average effect size of feedback was 0.79. This meant that feedback was one of the top 5 to 10 influences on learning along with aspects like students prior cognitive ability (rate 0.71) and direct instruction (rate 0.93). Hattie also states that feedback is most effective when it provides information about correct rather than incorrect responses. An earlier study [13] talks about the relation between feedback and goals when it comes to learning. Hattie here explains that "Feedback without goal setting is less effective, and goal setting without feedback is ineffective" and that a combination of the two is necessary for an effective learning process. This has to do with the fact that a greater challenge gives a higher probability of the student seeking, receiving and assimilating feedback information.

Even though feedback has a proved positive and important impact on learning, there are also situations where feedback can have a negative impact. This is explained by Askew [9] who states that too much feedback can be overwhelming and inhibitory, particularly if there is no dialogue or discussion and no help considering how to make changes are given.
2.2 Types of feedback

How well the feedback actually provides the potential positive impact of learning has a lot to do with the type of feedback. There are many different thoughts on how to categorize feedback. Tunstall and Gipps [14] has made a study with the goal to identify a typology for teacher feedback. This study resulted in four categories of feedback with a number of subtypes. Their four types of feedback are termed A, B, C and D and these types have been placed across a sort of scale from evaluative to descriptive approach, where each type has subdivisions. As feedback moves from type A to D it becomes more descriptive and less evaluative. The evaluative types can be defined as either positive or negative feedback, while the descriptive types cannot be described in terms of positive or negative, and are rather focused on improvement or achievement (see figure 1).

<table>
<thead>
<tr>
<th>Type A</th>
<th>Type B</th>
<th>Type C</th>
<th>Type D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rewarding</td>
<td>Approving</td>
<td>Specifying attainment</td>
<td>Constructing achievement</td>
</tr>
<tr>
<td>Positive</td>
<td>Positive</td>
<td>Specific</td>
<td>mental articulation of achievement</td>
</tr>
<tr>
<td>non-verbal</td>
<td>personal</td>
<td>acknowledgement of achievement</td>
<td></td>
</tr>
<tr>
<td>feedback</td>
<td>expression</td>
<td>achievement</td>
<td></td>
</tr>
<tr>
<td>Punishing</td>
<td>Disapproving</td>
<td>Specifying</td>
<td>Constructing the way forward</td>
</tr>
<tr>
<td>Positive</td>
<td>Negative</td>
<td>improvement</td>
<td></td>
</tr>
<tr>
<td>non-verbal</td>
<td>personal</td>
<td>correction of errors</td>
<td></td>
</tr>
<tr>
<td>feedback</td>
<td>expression</td>
<td>correction of errors</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>Provision of strategies</td>
<td></td>
</tr>
<tr>
<td></td>
<td>non-verbal</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Positive and negative One way to look at feedback is in terms of positive and negative. Negative feedback is mostly defined as feedback focused on incorrect actions, while positive feedback focuses on correct actions taken by the learner. In a study made by Ashby et al. [15] the effects of positive and negative feed-
back was investigated. This was done by providing different types of feedback to three different groups. The first group was given positive feedback only, the second group was given only negative feedback and third group was given a combination of the two feedback types. This study showed that the third group performed better than the other two, and the conclusion that a combination of both negative and positive feedback was to prefer in order to enhance learning could be drawn. Another study [16] on positive and negative feedback stated that there are positive aspects of using both types of feedback, and that the best type to use depends on the nature of the task.

A study made Van Duijvenvoorde et al. [17] used functional magnetic resonance imaging to examine the neural developmental changes related to feedback-based learning. Behavioral results from three age groups here showed that, compared to adults, 8- to 9-year-old children performed more inaccurately after receiving negative feedback relative to positive feedback. This study further implied a difference in how children compared to adults use feedback and the neural activation pattern found indicated a transition around the age 11 to 13 toward an increased positive influence on performance adjustment of negative feedback. These results are also supported in a similar study [18] on the effect of positive and negative feedback on preschoolers and suggest that young children in preschool and early school years are more responsive to positive feedback than to negative.

**Modalities** Another way to categorize feedback is by different modes, which is defined by Brookhart [19] as either written feedback, oral feedback or feedback given through demonstration. In a feedback study made by Werts et al. [20] also visual feedback is discussed as one possible feedback mode.

According to Orsmond [11] a majority of students prefer oral/verbal feedback since this type of feedback gives the opportunity to discuss the received feedback with the teacher and thereby clarify misunderstandings. This study also shows that some students finds written feedback unhelpful if it is not descriptive enough, a simple word telling that something needs to be improved is not enough. To enhance learning, feedback that provides a deeper explanation of the problems is needed.

The study by Werts et al. [20] further promotes the so called instructive feedback as a way to enhance learning. Instructive feedback is here defined as feedback that adds additional information to the pupil that might be unrelated to the original task. One example of this could be that if the task is to spell a word correctly, the feedback is not only telling if the spelling was right or wrong, is also adds more information such as telling an antonym to the spelled word. The positive aspects and effectiveness of so called instructive feedback is also supported by Griffen et al. [21] who shows that instructive feedback is also beneficial for pupils with mild intellectual disabilities.

**Levels of feedback** Hattie and Timperly [10] talks about four different levels of feedback as a way to categorize based on the content of the feedback. The
first of this four levels is called \textit{Task Level} and is focused on the task performed. This level aims to tell whether an action is correct or incorrect for example by simply writing the word correct or incorrect for the student to read. This type of feedback is also known as \textit{corrective feedback} and is the most frequently used type of feedback. The task level feedback could also include directions and more information about what needs to be done in order to improve. The biggest problem with feedback on the task level is that it often does not apply to other tasks.

The other level determined by Hattie and Timperly is the \textit{Process Level} \cite{10}. This type of feedback is directed to the process used to accomplish a task. This means that the feedback gives a wider information than only if the task is completed or not. The feedback also considers the understanding needed to complete a certain task. Studies have shown that feedback given on the process level seems to be more effective than feedback at the task level when it comes to enhancing deeper learning.

Third Hattie mentions a so called \textit{Self-regulation Level} \cite{10}. Self-regulation has to do with commitment, control and confidence and is a kind of internal feedback referring to the way the student monitor and regulate actions in order to reach the learning goal. How well this type of feedback works is depending on the students personal learning effectiveness, since less effective learners have problems developing self-regulation strategies and are in greater need of external feedback to reach the goal.

The fourth and last level of feedback explained by Hattie is called \textit{Self Level} \cite{10}. Feedback on this level focus on the student as a person and is often unrelated to the performance on the task. Hattie explains that this level should be avoided for two major reasons. First this type of feedback does not give any information on how to improve and reach further learning. The other reason is that this kind of feedback can contribute to the false picture of knowledge and intelligence as something static, which could have an impact on pupils understanding for the connection between effort and success.

2.3 Key elements of good feedback

Based on the different types of feedback it is clear that some types are to prefer over others when it comes to determining the meaning of good feedback. Wiggins \cite{22} has defined a number of essentials of effective feedback that could also work as a framework for good feedback. According to Wiggins study the base of good feedback is that a goal exists and actions are taken to achieve this goal. The feedback should consist of information related to the goal about the actions performed. Apart from a clear goal it is also important that there are tangible results related to the goal. Effective feedback also needs to be concrete, specific and useful and well adapted to the recipient in a way that makes it understandable to them. The sooner the feedback can be given, the better the recipient can use it. It is especially important that the task is still ongoing so that the student have time and reason to act upon the given feedback. Last Wiggins \cite{22} states that feedback must be consistent even if feedback is given from different sources.
A similar list of feedback aspects has been determined by Brookhart [19] who lists four important feedback strategies in order to reach good feedback. First Brookhart mentions the aspect of **timing**. Effective feedback should be given immediately to make sure that the feedback is given while the pupil is still working towards a learning goal, and not after they have finished. If the feedback is given immediately, they still have a reason to act upon the feedback and keep working towards the goal with the feedback in mind.

The **amount** of feedback given also have a great impact on how well the pupil receive it according to Brookhart [19]. To only list all of the issues or areas where the pupil needs to improve, is not effective even though it provides a great amount of feedback. A more effective way is to select a few main points and be very clear about what to do next in order to improve on these points. To be able to select which feedback to give, it is important to look at individual aspects. It is important to know the pupil to be able to sort out what information is relevant for each single person. The pupils personal goals and interests is one thing to consider when it comes to sorting out the most relevant feedback to given. It is also of great importance that the feedback have a balance between positive and negative aspects.

Feedback can also be given in different **modalities**, such as written, oral or demonstrations. When giving feedback to younger pupils Brookhart [19] explains that it is especially important to consider the pupils abilities, such as reading abilities, to make sure that feedback is given in a mode that they can understand. A conversational feedback is to prefer in most cases, where the pupil get a chance to talk to the teacher, ask questions and explain how they think about the task and their own performance.

Brookhart [19] further explains that good feedback demands appropriate sense of the **audience**. Depending on the content of the feedback it could either work best addressed to the individual person or in other cases to a group. Generally speaking, individual feedback is to prefer since it does not only give information about the how the student is doing, but also indicates that the teacher really cares about the pupils personal progress. By giving individual feedback, the feedback becomes more specific and well adapted to the individual.

A study made by Kelly and McLaughlin [23] have also shown the importance of a feedback that is based on the learners cognitive resources, abilities and prior experience. Their study suggests that feedback should be given as a function of the learners characteristics. The study states that it is of great importance to know the learner, know the task and match the feedback to the learner and the task.

### 3 Method

This study was conducted in two parts as follows:
3.1 Literature review

To achieve the objective of this study a literature review was conducted were feedback was studied from different perspectives based on the research question. The literature review served to collect information and theories found in earlier studies about feedback, with focus on feedback in a school environment and how feedback should be given to young students in an effective way. Through this review the role of feedback in the learning process could be identified as well as different types of feedback and how these different types affect the receiver. Finally theories and frameworks considering key elements of good feedback in a learning situation could be found and a list of guidelines for good feedback could be created based on these findings. The main keywords used to find relevant articles were: feedback, feedback and learning, feedback in education, types of feedback and effects of feedback. Among the articles found, the ones that had been peer reviewed were prioritized to make sure that articles were reliable.

3.2 Application study

The second part of the study aimed to determine how the theories about feedback found in the literature review could be applied in educational applications. In order to reach this goal two applications was studied with focus on the feedback provided to the user. The applications were chosen based on number of downloads and rating, where applications with high rating and high number of downloads were chosen to make sure that the evaluated applications were relevant and frequently used. One math application and one language application were chosen since these are the most common used types of educational applications. These applications were downloaded and studied on an Apple Ipad tablet. The study of these two applications was conducted in following steps:

1. Identification of the feedback provided in the application.
2. Categorization of the identified feedback according to different feedback types found in step 1.
3. Identification of successful and less successful examples of feedback in the applications based on the definitions of good feedback found in step 1.

By following these steps conclusions could be drawn on how well the chosen applications fulfills the framework for good feedback and thereby more general principles for how good feedback can be reached in educational applications could be determined. With ground in the study conclusions was drawn considering advantages and shortcomings of applications compared to traditional teaching, when it comes to providing good feedback.

Zcooly Affären 1 The first application in the study is called Zcooly Affären 1. This is a math application aimed for children between 6 to 8 years. The application contains a few different games and exercises that takes place in a store. By doing these exercises the user is supposed to learn addition, subtraction,
sequence number, pattern recognition, even and odd numbers and number sense. The exercise that contained most feedback was chosen for the study. The goal of this exercise was to practice counting items and calculating addition. The task for the user was to take the role of a cashier in a store. Customers entered the store and told what he wanted to buy, and the task was to give him the right kind and amount of items, count the price and put the right amount of money in the cash register. The task was divided in subtasks for each customer entering the store (see figure 2).

Fun English

The second application in the study is named Fun English. This is an application aimed for children between 3 to 10 years. It consists of 14 different learning exercises with focus on learning English words. In this applications two different exercises were chosen for the study in order to cover as much feedback as possible, since the exercises in this application were more basic. In the first exercise (Animal Band) the goal was to learn to recognize the names of different animals. The users task was to listen to a voice that said an animal and then click on the animal that was named on an image with several different animals (see figure 3). In the second exercise (Colour Buggies) the goal was to learn how to spell different colours. The word that the user was supposed to spell was shown but faded and a voice read it out when the game begun. The letters of the word were circulating above the faded word and the task was to swipe a car around, and hit the letters with the car in the right order (see figure 4).
4 Result

The first part of the study identified guidelines for good feedback in a traditional learning situation. The second part found that in applications used today, both examples that follow and do not follow these guidelines could be found.

4.1 Literature review

According to the earlier studies that has been reviewed in this report a number of elements that are crucial for effective feedback have been found. The following
list summarize the aspects that were found to be the most common mentioned and thereby considered the most important guidelines in order to reach good feedback in a learning situation.

- **In right time** - This aspect was mentioned in almost all of the earlier studies as one of the most important aspects. Feedback must be given while the task is still ongoing, so that it is still relevant for the pupil to act upon the given feedback and improve their result. The sooner the feedback can be given, the better it will be received [10, 19, 22].

- **Informative** - Effective feedback needs to provide concrete and deep information about what the pupil needs to do in order to improve as well as clear information about what the pupil has done well and why. The feedback should indicate both how the pupil is doing at the moment and what to do next. The feedback should not only refer to the specific task, a more effective feedback focuses on a higher level that makes it possible for the pupil to apply the feedback to other similar contexts [10, 12, 19].

- **Goal-based** - The feedback must be related to a clear goal if the pupil is going to be able to act upon the feedback in an effective way. If the goal is unclear, the pupil and the teacher might have different views of the goal, which can lead to misunderstanding of the given feedback [13, 22].

- **Adapted** - Pupils different abilities and characteristics must be considered. The feedback must be given in an individualized way based on the learners personal characteristics [19, 22, 23].

- **Predominantly positive** - There are positive aspects of both positive and negative feedback and many studies indicated that a combination is to prefer [15, 16, 19]. It was though shown that negative feedback affect children in a less beneficial way than it affects adults [17, 18].

- **Conversational** - Pupils often prefer a verbal feedback, especially if there is room for a dialog. If the student gets the opportunity to ask questions or explain his or her thoughts about their own performance, the student will receive a better understanding and can more easily adapt the given feedback [9, 11, 19].

- **Instructive** - When giving feedback to a pupil, additional information that answers question outside of the original task can be added to the feedback to support a deeper learning [12, 20, 21].

### 4.2 Application study

The use of feedback in the studied applications was identified as follows:

**Timing**

- **Zcooly** - Feedback is given as soon as the user has performed an action and marked that he/she is finished.

- **Fun English** - Feedback is given immediately as soon as the user has performed an action. There are a few exceptions were no feedback is given if the action performed was not correct.
Information level

Zcooly - The feedback given in this application is mostly on a task level but also gives further information about what went wrong and guides the user on how to move on in order to succeed with the task. Detailed information is given about exactly what part of the task that was performed incorrectly. If the task is performed correctly no further information is given.

Fun English - In this application the feedback provides uninformative feedback. All feedback is on a task level and simply indicates whether the task was performed correctly or not. No further information is provided in any of the exercises.

Clearness of goal

Zcooly - The goal of the task is not explained in detail as standard, but a more detailed instruction can be found by clicking a question mark. When choosing an exercise a voice and an image indicates what the task is about (image of a cash register and voice asking you to help out at the counter). As customers enter it is explained what this customer wants to purchase by a voice and a written text. Money and items are also visible to indicate the goal of the task.

Fun English - Each exercise is chosen by clicking a square were the name of the exercise and a very short description can be read. When this square has been clicked, no additional instructions about the goal is given. The exercise begins with a voice pronouncing a word of a color that is supposed to be spelled, or pronouncing an animal that is supposed to be clicked. There is no information button or other way to gain more information about the task.

Adaption

Zcooly - The feedback is given in the same way throughout this application. There is no way to adapt feedback based on the users abilities or preferences. Feedback is mostly given by a voice explaining what went wrong in combination with a sound and a visualisation through a red colored cross or green colored square. The task is given in both written and spoken form.

Fun English - Also in this application the feedback is given in the same standardized way at all times. This application is using a combination of visual feedback (stars) and a sound that indicates right or wrong.

Positive or negative

Zcooly - If a task is performed correctly, positive feedback is given by voice, sound and color. If the task is performed incorrectly, negative feedback is provided in the same way. Even if a task is performed partly correct but not completely, only negative feedback is given.

Fun English - Positive feedback is always given when a task is performed correctly. If a mistake occurs, negative feedback is given though a buzzing sound, or no feedback is given.
Conversation

Zcooly - No conversational feedback is provided.
Fun English - No conversational feedback is provided.

Instructiveness

Zcooly - No constructive feedback is provided.
Fun English - One type of instructive feedback is provided. Pronunciation is added to the spelling task. When the task is to spell a word, each letter is pronounced by a voice and the whole letter is pronounced when the task is completed.

5 Discussion

By looking at these two different applications it becomes clear that feedback can be designed in many different ways in an application. The two studied applications provide feedback differently and have different areas were they succeed and were they do not. They both have a lot to learn from each other. Both applications give feedback mostly on a task level, but the Zcooly application manage to provide a deeper and more informative negative feedback that can guide the user about what to do in order to improve and succeed with the task.

The fact that the Zcooly application is very clear and informative in the feedback given could, however, also have a negative effect. As shown by Van Duijvenvoorde et al. [17] feedback to children of this age should be mainly positive since children handles negative feedback different than adults and seems to decrease their performance if the feedback is negative. As this application is designed, a pupil who has troubles with the tasks will receive a lot of negative feedback. Even though some part of the task may have been performed correctly, this is not noted as long as the whole task is not performed correctly. The Fun English application gives less negative feedback, since it sometimes gives no feedback at all as long as the task is not performed correctly. Even though it is a good thing to reduce the amount of negative feedback, the absence of feedback is not the best solution according to other studies implying the importance of informative feedback [10].

When it comes to providing instructive feedback none of the studied application seems to focus on this. Zcooly gives no instructive feedback at all while the Fun English gives a kind of instructive feedback as letters and words are pronounced when the task is simply to spell the word. The pronunciation is not necessary to complete the task and can thereby be seen as an instructive feedback that extends the learning goal of the task. In the other exercise no instructive feedback is given, but could have been added in the same way. Here a word is only pronounced, but could also have been shown written in order to add a visual recognition to the hearing recognition. Also in the Zcooly application it should be possible to add instructive information by adding additional
information when feedback is given. By adding this kind of additional feedback the learning outcome of each exercise could be enhanced [20, 21].

One area where applications seems to have an advantage is the aspect of timing. When it comes to more traditional ways of teaching, the teacher can not follow each pupils every step and give feedback timely on every action performed by the pupil. An application, on the other hand, can provide instant feedback at all times, which both of the studied applications manage to do. The feedback is given timely and regularly throughout the exercises in both Zcooly and Fun English. The difference between the two is that Zcooly does not give feedback until the user has implied that he/she is finished with the task. This is a way to minimize feedback given on accidental actions, and can thereby be a way to reduce negative feedback. As shown in the studied applications, the application can also make sure that a task is performed correctly by not allowing the user to move on to the next task before the first one is finished. Both applications together shows that it is possible to reach many of the guidelines for effective feedback in an application. Zcooly manage to give informative feedback that explains in detail what went wrong and gives a guidance of what to do in order improve. The application does not only provide feedback when something is done incorrect, it also provides positive feedback when the user has performed a task correctly. This application also provides a clear description about the goal of the task. By providing a combination of different feedback types this application can also work for students with different abilities. By a combination of both visual, written and talked feedback, the feedback can be understood despite things like reading or hearing disabilities. Even though the application does not actively adapt to different users, this is a way to handle the aspect of adaption. Fun English manages to give some instructive feedback and thereby enhance the learning. This application also gives little negative feedback, which is important for the aimed age group. Both applications provide feedback on a task level according to Hattie’s definition [10], even though the Zcooly application provides a more informative version of this feedback. As Hattie’s study further implied it should be more effective to give feedback on a higher level that provides a deeper understanding and makes the feedback transferable to other similar tasks.

One of the main issues with feedback from an application compared to feedback from a teacher is that they do not provide the opportunity to discuss the given feedback or ask questions about it, it does not provide a conversational feedback. As shown in the earlier studies, a conversational feedback makes it easier for the student to understand the given feedback and gives the opportunity to clear out misunderstandings of both teachers and students [11].

A counter-argument to the lack of conversation may be that the tasks to be solved in an application is not always fully comparable to a student task that a teacher corrects. As seen in the applications, feedback is mostly given for very specific and small subtasks, where the answer is an unequivocal right or wrong. In situations where the teacher gives feedback to the student, it may cover larger tasks where the answer may not always be as obviously is right or wrong, and thereby the need of asking questions or discussing the feedback.
may differ. Still a conversational feedback has great advantages as shown in the literature review [9, 11, 19]. Even though the tasks are often quite simple in the applications, misunderstandings could occur considering the goal of the task, or there might be deeper misunderstandings that hinders the user to perform the task correctly. These things might be hard to discover and correct when there is no conversation available. Perhaps a solution to this problem could be for pupils to use these applications in close collaboration with the teacher. If the teacher is well familiar with the application, he or she could answer questions that occur and detect misunderstandings by the pupil.

Another great disadvantage with applications is the lack of individualization and personal adaption [23]. The application does not have knowledge about the user the way that a teacher does about a student. A teacher can adapt his/her way of giving feedback based on individual aspects, the current situation and the current task in a way that the applications today can not since they follow a programed, standardized feedback system. Perhaps it is also possible that both a better adaption and detection of misunderstandings could be reached within the application by using more advanced techniques and other ways to design these applications in the future.

The study further showed the importance of being clear about the goal, and when it comes to an application a technical aspect is also added to the understanding of the task. Even if the goal is clear and easy to understand, the user must also understand how to technically proceed to solve the task. One example of this is the task in the Zcooly application. The task contains many steps, like dragging items to a certain place on the touch screen, clicking at right places and so on, and it is not completely obvious how to solve this tasks even if you know the mathematical answer (which is the goal). The Zcooly application has a quite clear description of the main goal, explaining step by step what the user is supposed to do, but this description is shown only if the user actively choose to click the question mark. The standard information given if this button is not clocked is not very detailed or descriptive and could probably easily be misunderstood. Also the Fun English application is lacking in how the goal of the task is presented and there is no way to find more information. If the main goal is unclear as in the studied cases, the user could get confused and the errors might be caused simply by a misunderstanding of the goal.

6 Conclusions

The literature review found that there are many theories and studies defining good feedback in a learning situation. Some theories differ from each other, but a number of guidelines that most studies seemed to support could be identified. The second part of the study showed that many of these guidelines for good feedback can be reach in an educational application. Feedback in an application compared to a traditional teacher to student feedback was shown to have both advantages and disadvantages. Advantages of feedback in applications is first of all the aspect of timing and the possibility to provide a combination of different
feedback modes, such as written, visual and auditive. The most critical shortcomings in applications today is the lack of individualization and conversational feedback.

Further studies are needed to determine whether these shortcomings could be overcome with the help of new technologies or new ways to create and design educational applications.

References

Push-Notifying the End-User at the Most Appropriate Time

Mattias Edin
Umeå University, Sweden
maed0101@student.umu.se

Abstract. Smartphones have notifications for various kinds, for instance emails, social network updates, messages. Although little is known how stress affects how receptive smartphone users are to notifications. In this article, an in-situ study consisting of 16 smartphone users where response time was collected for 211 notifications in total. An Android applications were installed on the users to conduct the survey. We found that higher stressed participants have a longer respond time to notifications. The notifications that are attended to within 1 minute consists of mostly low stressed participants. These findings show that low stressed participants are more susceptible to notifications.

Keywords: health, monitor, measure, stress, technology, interruptions, context aware, notifications, disruption, digital stress, technostress.

1 Introduction

While the smartphone market has grown in the past, the prediction is made that it will continue to do so in the near future [1]. One big purpose of these devices is to make one and other more connected to each other. Unfortunately, this connective also comes with a price. The price is not only the vendors price tag on the product but the mental health that can get affected by the use of the products. The mobile use is causing stress and sleep disorder on its users [2]. While this mainly is affecting users with an unhealthy use of their devices it does, in fact, show that there is an underlying problem with our use of them. The smartphone devices are enabling us to use social media services wherever we are and at any time. This, in turn, have shown to have a negative effect on the well being of the user [3].

There are commonly used applications such as Facebook, Twitter or most of the messaging or mail applications that utilise push notification services to send notifications to the user. A push notification is a notification that shows when the client user did not explicitly request the
information, the notification is, therefore, a message that shows up without warning and can surprise the user. This could be the reason for the interruptive behaviour of the push notifications. If the service that pushes the notification could know when the user is more susceptible for notifications, the interruptive behaviour may be minimised. A study done by Spira and Feintuch [4] in 2006 shows that approximately 28% of a workday is taken up by different interruptions. To make the matter worse when the average interruption is 2.8 seconds the interrupted person tends to do twice as many errors in the current task [5]. One explanation for this could be why digital stress has a significant negative impact on productivity as resulting in a study carried out in 2007 [6]. Although this data is already showing the problems of interruptions but fails to analyse how stress affects our receptivity of push notifications.

1.1 Aims and objective

The goal of this study is to evaluate the current push notification system by looking at how likely the user is to attend to notifications during different times during the day and how stressed the user feels at these specific times. In turn, this could be used to find possible times of the day when the user is more susceptible to notifications and also take into account of how stressed the user feels at the same time. With this study, a conclusion about specific times the user will be more likely to attend to the notification was drawn. This information can be used for the implementation of future mobile applications that utilise the push notification system to send notifications when the user wants them rather when the service want you to have them. This, in turn, could make the user experience better by eliminating annoyances of receiving notifications when not wanted.

- At which time of the day the users of smartphones are more susceptible to attend to push-notifications?
- Is there a correlation between the time it takes for the user to look at the notification and how stressed the user feels?

2 Theoretical framework

In this study, the focus of what precise type of stress the user is experiencing is not of most importance but the negative feeling of it the user has. Therefore, the user is free to define its stress as a negative feeling that affects the mental health at the current time. As described under the
section Aims and objective the study does not focus on the stress level itself but how it affects its user of how push notifications are perceived. It is, for this reason, the absent of explanation in stress. According to the American Institute of Stress, stress in common sense a term that cannot be defined as it is a subjective phenomenon and is described instead as follows "a condition or feeling experienced when a person perceives that demands exceed the personal and social resources the individual is able to mobilize" [7]. According to Selye [8] "Stress is the non-specific response of the body to any demand for change." Stress is, therefore, vague in the sense of definition. Stress has according to American Institute of Stress both good and bad impact on the human body [7]. In this article we have chosen to focus on the bad impact stress has on the user and in this case the user of a smartphone.

2.1 Technostress

The concept of technostress is stress caused by an inability to cope and process with all the information that is surrounding the user. This is also described as when there is a decrease in productivity and satisfaction of the end product of different tasks. The unhealthy usage that this article is referring to is when the smartphone is distracting the end user from its primary task. Therefore, affecting productivity, as described in the related work of Brooks and Stoney where stress has a correlation with the use of social media, where the use of social media had a negative impact on the performance of the current task, as well as and lower happiness [3].

As described by Arnetz and Wiholm [9], technostress can be defined as follows; "the state of mental and physiological arousal observed in certain employees who are heavily dependent on computers in their work". This definition is from the year 1997 and could therefore perhaps extend to not only employees because it can be easy to think that people have become more dependent on computers in the form of smartphones in recent years that is not only work related.

2.2 Push notification

In the simplest form, a notification is defined in this article as a message coupled with either sound or/and light flashes from the clients device. The message is brief and implies that the user should interact with it by either opening it for the full message or be dismissed. Either way, the message has had an interactive effect on the user by being in attention. Simple notifications do not describe what the cause of the notification is.
so the notification could be a result of the user’s action. For example, if the user starts a download of a File, the user will also create a notification that shows the downloading progress. An extension of a simple notification is Push Notifications (PNs) which also describes the initiation of the notification. The PNs is initiated without the end users explicit action and is often done remotely from within a subscription service that the user has explicitly signed up for. This could in turn possibly interrupt the end user and increasing mental, cognitive workload.

2.3 Related work

There are a lot of related work done in the topic of notifications where one study published in 2014 [10] showed that people have to deal with 63.5 notifications on average per day who also found that notifications were most often attended to within one minute. Another related study published in 2010 showed that desktop users receiving email notifications showed that the users thought the notifications to be disruptive but still wanted them for more awareness. Another article published in 2014 showed that about 49% of the notifications on smartphone users were attended to within one minute. The survey also showed that if the notification was coupled with sound, vibration or a light cue or any combination of them is 12 times more likely to attend to the notification immediately while the time of day had no evident effect. A related study conducted by Microsoft Research [11] in 2010 showed that instant messaging had a disruptive effect on the current tasks the user is doing. The study also showed that if the message is highly relevant to the current task the descriptiveness effect on performance was not as strong. Another study published in 2007 [12] showed that instant messaging such as short message service (SMS) had the effect of prolonged stress and sleep disorder on some of the subjects.

This article distances itself from these studies by looking at how stress has the direct impact of how fast we attend to PNs. The above studies all has their corresponding topic but with this survey, we found a connection between stress and how prone one is to attend to a PN.

3 Method

To get an understanding of how likely one is to attend to push notifications (PNs) while being affected by different amount stress and workloads, an Android application was constructed and installed on the subjects smartphones to monitor PNs. The application randomly received PNs
sent by an external service to the test subject at different times of day, a
service in the application monitored the time it takes the user to take a
look at the notification. It is important to note that the notifications from
the notification did not try to mimic other applications as explained later
in the section: Discussion. When the subject looks at the notification, they
will give a value of how busy they are feeling at the current time and the
time is recorded. When the test subject first installs the application, they
will register for the survey. During the time of registration the test subject
gives a value of how stressed they are feeling at the time of the registration
see section 3.1 for how stressed is measured. The data was then relayed to
a web server. The test subjects had the application installed for seven days
and during this time the notifications were received to their smartphones
at random times but within 8.00am to 9.00pm. There are four different
timestamps that were collected by the survey, first when the notification
was sent, when the notification is received by the smartphone, when the
subject opens the notification and finally when the notification is later
received and stored on the web server. The important time stamp is when
the subject receives the PN in their smartphones which are going to be
used as reference time throughout this study. It is important to note that
the user can dismiss the PN sent from the server which in turn will not
be taken into consideration when calculating the time subject takes to
attend to the PN. The participants were not told that the survey also
looks at the time it takes for the participant to attend to the PN.

For each PN there are two questions the participants had to answer,
one is defined in section 3.2 and the other question is a randomly chosen
question from a set of five different questions. The second question is not
used for the survey and is there only for making the survey more varied
and interesting for the participant. If each PN was exactly the same the
participant could lose interest in the survey which could potentially have
an impact on the result. The other reason why we chose to have this
question is also to make the survey more realistic because PNs are from
our experience seldom contains the same information all the time in a
realistic situation. All questions that were asked used the same answering
system, that is pre-defined responses choices with a drop down menu. The
reason for this is for making the survey as less time consuming as possible
for the participant. This is important because if each PN were associated
with a time-consuming task there would be a bigger risk for the survey
to make an impact on itself, where the response time could increase just
because the participant knows that if the notification is clicked they have
to postpone their current task.
The number of participants N=16 consists of 11 men and 5 women with occupations division of 13 students, 2 employed and 1 other with age range from 20 to 56.

In this study, we have chosen not to include the study of specific applications like Facebook or e-mail notifications, because of the comprehension of the survey. Each application needs its independent study to find differences between them.

3.1 Stress measurement

The stress measurement was derived from the QPS Nordic questionnaire [13] which has good validity as studied by related work [14]: "By stress is meant a state when a person feels tense, restless or worried or cannot sleep at night because he/she is thinking about problems all the time. Are you experiencing this kind of stress currently?" The response scale was 1–5: 1=not at all, 2=just a little, 3=to some extent, 4=rather much, 5=very much. The question was originally in Swedish and translated by Scandinavian Journal of Work, Environment & Health. The responses were also divided furthermore into two groups; High (response sets 4–5) and a Low (response sets 1–3) to minimise the complexity of the data set. To measure stress with only one question with good validity is necessary for the simplicity of the application. More work had to be put down in developing the application if another stress measurement would be chosen that weight in more factors. For instance, one could choose to measure the stress with an activity band with sensory data. Although for this study it was sufficient to have confidence in the test subjects own answer range from 1 to 5. Another reason that one question is a good choice is that of making the survey easier on the participants. Fewer questions would not have a smaller impact on the participants choice of participating in the study.

3.2 Busy measurement

To measure how busy the test subjects are feeling one question was constructed as follows; "When a person feels busy, one feels that some work has to be prioritised and left out because not all can be done in a limited time span. Are you feeling this kind busyness?". This question was derived from the tested stress question to maintain the same type of question where the reader defines one’s current situation according to a definition which is defined in the question itself. The response scale is the same as for the stress question.
3.3 Procedure

Recruitment of participants to the survey was mainly done through self voluntarily assigning from an announcement on a social media page on Facebook. The participant that volunteered were then contacted with both e-mail and personal messages on Facebook. The test subject’s were then added to the testing group and given a link to download the application for this survey. Once installed the test subject made a one-time registration from inside the application. The test subjects were then asked to have the application installed for 7 days. During the test period each subject received in average about 13 notifications which are scattered at different times of the day.

4 Results

During the test 275 notifications were sent and received by the test subjects and from those 42 notifications was dismissed. The dismissed notifications were removed from the dataset. Because of extreme values, the charts in the figures (1, 2, 3) had to be zoomed to make the it easier to see the differences between the datasets. According to the median in figure 2b the stress have a clear impact on the response time. Test subjects that answered low stress have a median of 4.9 minutes and high stress with a median of 17.3 minutes in response time. The upper quartile (25% of the dataset) in figure 2b for high and low stressed participants consists of data points within an interval consists of 46 and 133 seconds in respond time while the high stressed participants have the quartile positioned about 10 minutes above that of the low stressed participants. This means that high stressed participants have a tendency of having a longer response time when considering the upper half of the results in those datasets. Although the lower quartile of the data set for high stressed participant also cover the same region as low stressed participant which shows that there is a slight lower certainty for this result. The same finding could be found for how busy the participants are feeling. According to figure 3b the participants with high busyness have if considering the median a difference in 12.4 minutes where high busyness tends to have a higher response time.

27% (63 data points) of all notifications were attended to within 1 minute (see figure 4) where the rest are scattered to as high as 26.7 hours as respond time at the most. To elaborate the data set which was attended to within one minute one could find a big difference from how stressed the participants was. A difference of 364% more of the data points attended
to within 1 minute has low stress. This means that participants with low stress have a higher probability of attending to a notification than high stressed participant during the first minute (see figure 5). From figure 6 there is a hint of better receptivity with PN during noon although more tests have to be carried out to confirm this.

One abnormal finding in the result is how the test subjects that answered the question of how stressed they feel (see section 3.1 for questionnaire), see figure 3, with the answer 'Not at all' and 'Just a little' all reported to have a very fast respond time, median of only 12 and 43.5 seconds where the other datasets have a median of 13.9, 29.5, 12.6 minutes (see figure 2a). This could be interpreted that the very low stressed participant are very fast at attending to PNs.

One more interesting finding is how the response time median relates to different hours of the day. As can be seen in figure 7 the hours 7 and 17 have an obvious higher response time if considering only the median.

**Response time for each weekday**

![Fig. 1: All data points grouped by weekday. Datasets in (b) is displayed without extreme values with values in range of 0-40 minutes to make the differences more readable.](image)
Fig. 2: Box plot over response time grouped by stress. Datapoints in (b) is displayed with grouped values where datasets 1-3 is low and 4-5 is high.

Fig. 3: Box plot over response time grouped by busyness. Datapoints in (b) is displayed with grouped values where datasets 1-3 is low and 4-5 is high.
Fig. 4: Histogram over datasets with respond time within 0 and 30 minutes.

Fig. 5: Extracted datapoints which responded within 1 minute grouped by stress.
Fig. 6: Percentage responded within 1 minute for each hour.

Fig. 7: Median response time for each hour where annotation for each bar is the number of values the median is calculated from.
5 Discussion

According to related work[10] there is significant difference between the time it takes for users to attend to notifications during the weekdays and weekends. This is not the case of the findings of this survey where the result has no noticeable difference between the two (see figure 1).

Because the test subject was not instructed to know that the response time is important for the survey the result should therefore not affect the data. One thing that could be affecting the data is the fact that the participant knows that all the PNs is coming from the survey. This fact could alter the importance of attending to the PN and therefore, be more important for the user than other PN, this, in turn, could make the user more prone to attend to it. The participant could therefore in other cases not be as prone to attend to other notifications because the importance is not as high so the response time could be affected.

There is a subjective matter of what notifications that are more urgent than others which will, in turn, have an effect on the result. The urgency of the notification can, in turn, explain the reaction time that the user has when to look at the notifications. For example, some user may argue the fact that a regular text message is more important than a Facebook message equivalent and the reason for this is mainly subjective on behalf of the client user. The point of this is that different people value different things and that can explain the deviation in the result. But the user can be theorised not to change their personal importance level of the test applications notification during the survey so the fluctuations in the result are greater than the average response time. What it is that makes these differences can have to do with the notification type, how it notifies the user. In many cases, the text message has user-defined sounds to it while Facebook has pre-defined sounds. If the user has configured the sound, he or she may be more prone to get interrupted by the notification in comparison to the generic pre-defined, simply because it is more personal. There may also be a difference in how the user values the content of the application the notification originated from, for instance, an e-mail notification may have a high user value because it has a high chance of being work related, and could, therefore, be important. In these cases, the originated application is of importance. Also, there may be a difference in the reception of the notification depending on the type of notification (sound or vibration). As Gallud and Tesorier[15] found there is a slight difference in how the user has defined their notification setting. Also, there
is a noticeable difference in how the users handle notifications depending on the quantity they receive.

This presented study has a limitation where the number of participants is 16 Android users, where only 5 of those are female where the collected data have a span of 7 days. Further work may have to get a greater number of participants to make more accurate statistical evaluations. A second limitation is the fact that the survey is limited to Android users, where other platforms may notify the user in a different way that could therefore have an impact on the result. The resulting conclusions are therefore only limited to how the current Android operation system handles notifications.

6 Conclusion

To answer the objectives laid out in section 1.1, the time the user is more susceptible has two factors to take into consideration. How stressed or busy the user is and time it takes for the user to attend to the notification. The result of the study showed that there is a big proportion of the notification that is attended to within 1 minute, namely 27% and the time is considered to be the noon of the day (see section 4 for further explanation). For the second objective, to find out if there are any differences for stressed participants or busy participants, the survey showed that there is a difference in the two. Participants with high stress tend to have a longer respond time for notifications as well as busy participants.

References


Aesthetic value versus function in mobile applications

Robin Schulze
Department of Applied Physics and Electronics
Umeå University, Sweden
id12rse@cs.umu.se

Abstract.

Mobile applications are going through a lot of changes. Developers are always trying to improve their aesthetic value and functionality. That is why the objective of the study is to identify whenever aesthetic value or function matters most for users of mobile applications. To be able to identify this, researching what defines aesthetic value and a function is necessary. It is also need to understand how mobile application works today. The study was made by letting seven participants perform three different tasks. They had to perform each task twice once with an application that was built to be more functional and the other to be more aesthetic pleasing. The tasks that were chosen were: "Searching in a list", "Clearing all input." and "Retyping same text". After the participants had performed all task they were asked which one they preferred better. They were also asked if the application they did not prefer could be improved to make them prefer it instead. The results from the study shows that in two out of three tasks the participants prefers the functional application over the aesthetic pleasing and most did not think the aesthetic pleasing application could be improved to be chosen instead. Although it is a highly preference what the users prefer, the results might be unreliable. But it should at least give a better understanding to what the users prefers most of aesthetic value and function.

1 Introduction

Throughout the 20th century, mobile applications have been through big changes. Developers are continuously trying to improve what applications can offer the users to satisfy their needs. According to M. Enis, the popularity and functionality of mobile applications are continuously growing [1]. In the study "The evolution of mobile apps: An exploratory study", it was concluded that the complexity of mobile applications is continuously increasing [2].

Aesthetic value is also increasingly important in mobile applications. According to J. Silvennoine et.al "Visual attractiveness is increasingly seen as an essential factor in perceived usability, interaction, and overall appraisal of user interfaces." [3]. Which makes the increasing aesthetic value of the application a cornerstone in making it more visual attractive [3]. In the The Encyclopedia of Human-Computer Interaction(HCI), 2nd Ed.: Visual Aesthetics, N. Tractinsky
argues the importance of visual aesthetic in HCI applications from three different perspectives, the design perspective, the psychological perspective, and the practical perspective [3]. He states "I believe that, taken together they cover the lion share of arguments for the inclusion of visual aesthetics as a major aspect of HCI practice, research and education [3].

With the increase of popularity, complexity and functionality one could not help to wonder if it is the functionality or aesthetic value that matters most for users in applications. This study was conducted to get a deeper understanding of what users value the most of aesthetic value and function and if there is any difference between mobile applications.

1.1 Objective
The goal of this study was to identify and analyze what users value the most, aesthetic value or function, in mobile applications. The purpose was to conclude what matters the most for users today of the two. Furthermore, the study examined what defines aesthetic value. The functions that were chosen were "Searching in a list", "Clearing all input with a click" and "Remembering inputs and give suggestion for next input". A user study was conducted to gather information about the users preferences and thoughts on aesthetic value and function. The information and thoughts were analyzed to conclude what users favors the most of aesthetic value and function in both mobile applications.

The objectives of the study were as follows:
(a) Identify whenever aesthetic value or function matters most for users of mobile applications.

2 Theory
This study is making use of other articles to better understand and grasp the definition of some concepts. Furthermore other articles were used to find important information that was later used in the user study of this study.

2.1 Aesthetic value
What is aesthetic value? According to L. Plato and A. Meskin one definition of aesthetic value could be: "Aesthetic value is the value that an object, event or state of affairs (most paradigmatically an art work or the natural environment) possesses in virtue of its capacity to elicit pleasure (positive value) or displeasure (negative value) when appreciated or experienced aesthetically." [5]. This could be interpreted as aesthetic value is something that gives the user pleasure or "positive value" when experienced. This definition of aesthetic value was used when conducting the user study. Another common definition of aesthetic is "an artistically beautiful or pleasing appearance" or "a pleasing appearance or effect: Beauty" which N. Tractinsky uses when he talks about aesthetic value [3].
2.2 Functions

Defining what a function is has proven difficult. But according to V. Beal one definition of a function could be that it is a section of an application that performs a specific task \[^6\]. That would mean in example that a function could be a button changing color of a specified text. The functions that were chosen will be described more thoroughly below.

2.2.1 Searching in a list

The function searching in a list is implemented by having a text field where the user can input a text string and the application will thereafter search for that text string in the list. The application will then show any items that match the input text. If there are no items that matches the input text, no items will be shown \[^7\]. Google’s search engine can be seen as searching through a list of sites, that is why this function was chosen.

2.2.2 Clearing all input with a click

Clearing all input with a click is implemented in the way that when a user clicks a button all the input into a field is cleared. The field will then be empty and new text can be input. Google’s mobile site uses this function to make it faster for the user to clear the text area.

2.2.3 Remembering inputs and give suggestion for next input

Remembering inputs and give suggestions for next input is implemented by having the application remember what the user input and suggest it when the user is typing in another field that should contain the same information. Google’s search engine gives suggestions on what you have searched for before and that is why this function was chosen.

2.3 Applications

In "Computer Basics: Understanding Applications" the Goodwill Community Foundation is defining applications as a software that allows the user to complete specific tasks \[^8\]. That would mean that an application consists of one or more functions as defined earlier in 2.2 that completes different tasks for the user. Mobile applications run inside the mobile operating system \[^8\]. A mobile application can for example be used by a user to manage health and wellness \[^9\]. A mobile application can also contain ads to help the developer of the application generate revenue \[^10\].

3 Method

Here the method of how the user study’s testing applications were built and how the study itself was conducted will be presented. The study was done in the span of two weeks.
3.1 User study

The user study was conducted by first creating two different testing applications for mobile devices. One that were aesthetic pleasing for the user but without a function that eased the working process, one that had a function too ease the working process but did not include the aesthetic pleasing elements. The applications were minimalistic done and had only the elements to simulate one of the two application scenarios.

3.1.1 Participants

The participants in this user study were males and females from 20 to 30 years old. The thought behind the choice of target group were that the target group will be using applications in the future and they have been around applications longer then the ones younger than them. The participants was not picked by any other criteria. Participants were randomly selected from students at Umeå University. They were first approached and asked about their age to see if they met the criteria and then asked if they wanted to participate in a study. The study was conducted on seven persons.

3.1.2 User tests

All information gathered from the study was recorded in a text document. When the participants agreed to take part in the study they were first presented with ethics information which they read. After that they were asked what age and gender they were. They then continued to test the two applications. Which application they started with was set so half of the people started with one and the other half with the other one. They were asked to preform the stated tasks on the two applications. Before they started with the "Aesthetic pleasing application" they were asked if they found the application to be aesthetic pleasing or beautiful. At the end, the users were asked if the element (functional or aesthetic pleasing) of any of the three applications they did not choose could be heightened to be chosen instead. This was done to see if there could be some scenario where they would have chosen the other application. The results from each user were recorded to a survey which the observer filled in.

Tasks

(a) Searching in a list.

The users were asked to search for five items in a 100 item long list. The items which they were to search for were randomized.

(b) Clearing all input.

In this task the users were asked to type in five sentences with 5 to 20 words that were randomly chosen from a list of 20 sentences. After they had written one sentence they had to clear the field before writing the next one.

(c) Retyping same text.

In the last task the users were ask to fill out five forms. The forms were identical and the same information were needed in the forms. When one form was filled the users moved on to the next form.
3.1.3 Aesthetic pleasing application

The Aesthetic pleasing application was simulated by using the existing applica-
tion "MobisleNotes". It had minimalistic elements to simulate aesthetic value
for the user. The application was used in all three parts to test, "Searching in a
list", "Clearing all input" and "Retyping same text".

3.1.4 Application with function to ease the working process

The application with function to ease the working process was written in html. It
had minimalistic elements to simulate functionality for the user. The application
was split into three parts to test, "Searching in a list", "Clearing all input" and
"Retyping same text".

3.1.5 Environment

The tests was carried out on an iPhone 6 with iOS 8.1.2 installed. The tests
were conducted in a small room where there was no immediate distractions for
the participants.

3.1.6 Evaluating the test applications

When the applications were created they were tested by two chosen test persons
to see if they yielded the right results. When the applications were tweaked so
they yielded the right results the study continued to making user tests.

4 Results

Here the results of the study is presented. First the chosen application and built
application will be shown and how the built application were altered after the
test of the application were conducted. Thereafter the user study’s results will
be presented.

4.1 Applications

Here the results regarding the application will be shown. The results of evaluating
the application and the final application will be presented.

4.1.1 Evaluating the test applications

The application was tested on two persons to see if it yielded the right results
and that the applications was built with the right elements. The test showed
that the functional application for evaluating "Searching in a list" was clunky
and were hard to use. To solve that, the result when searching for a item was
instead shown at the top and the user did not need to scroll through the whole
list just to find the item highlighted.
Another thing that was brought to light when evaluating the functional application that tested "Retyping same text" and "Clearing all input" was that the user used the mobile devices built in function to remove all text. To solve this the user is now instructed to not use this method.

4.1.2 Final applications

The final application that were used for the user tests will be shown below.

**Searching in a list**

![Fig. 1. Final application for evaluating functionality for the task "Searching in a list".](image-url)
Fig. 2. Final application for evaluating aesthetic value for the task "Searching in a list".

Clearing all input

Fig. 3. Final application for evaluating functionality for the task "Clearing all input".
Fig. 4. Final application for evaluating aesthetic value for the task "Clearing all input".

Retyping same text

Fig. 5. Final application for evaluating functionality for the task "Retyping same text".
4.2 User study

The results from the user study is presented here. The study was conducted on 9 users. Although one user did not think the aesthetic pleasing application was aesthetic pleasing so the results from that user were not included in the results.

4.2.1 Searching in a list

Below the results from the task "Searching in a list" is presented. Six of seven users preferred the functional application on this test. None of the seven thought that the aesthetic value could be increased to make them switch. Although the user who chose the aesthetic pleasing though that the functional aspect of the test application could be improved to make the user chose the functional application instead. The users who chose the functional application did not think that the aesthetic pleasing application could be improved to make them choose it instead.
4.2.2 Clearing all input

Here the results from the task "Clearing all input" will be presented. In this task four out of seven chose the functional application. And when on the question if the user thought the application they did not chose could be improved to be chosen instead only one answered that the aesthetic pleasing application could be
improved to be chosen instead and the other seven said that it was not possible that other application they did not chose could be improved to be chosen instead.

Fig. 9. Pie chart of what application the users liked the most in the task "Clearing all input".

Fig. 10. Pie chart over if the users thought that the application they did not chose in task "Clearing all input" could be improved to make them chose that one instead.
4.2.3 Retyping same text

Below the results from the task "Retyping same text" is shown in pie charts. All seven users chose the functional application in this task. Although two users though that the aesthetic pleasing application could be improved to make them chose it instead.

![Pie chart of what application the users liked the most in the task "Retyping same text".](image)

**Fig. 11.** Pie chart of what application the users liked the most in the task "Retyping same text".

![Pie chart over if the users thought that the application they did not chose in task "Retyping same text" could be improved to make them chose that one instead.](image)

**Fig. 12.** Pie chart over if the users thought that the application they did not chose in task "Retyping same text" could be improved to make them chose that one instead.
5 Discussion

In this section of the report I will voice my own thoughts about the study, the results and the conclusions I drew from those results.

5.1 Study

Here I voice my thoughts and opinions from the study. I will go through the subjects: Participants, Tests and Results. I will be talking about function and aesthetic levels to somewhat compare them.

5.1.1 Participants

I would first like to say that I would really have seen that I had the time to test more users. While I feel that seven users still may yield a taste of what all the users prefer, I would want to have tested at least 20 more users to see that my seven users were no special cases and that they are in line with the rest. What I also would have like was to have users from other universities to get a broader spectrum of users and that their opinions is not local opinions. Otherwise I am pleased with the chosen age. I feel like it was the right generation to make the study on since they are one of the generations that will be using applications in the future, they have good knowledge of how things work and gave constructive and complete answers in the study.

5.1.2 Tests

I am pleased with how the tests went. The users were really interested and made good points for their choices. I made a mistake to not take notes of all the good points they made. It would have been great to have in the results to compare to each user. But on the other hand, then I would have wanted all users to say something similar or else I feel that I can not compare users who was not as elaborate as some other users. One thing that was a challenge in creating the test was to balance them. I noticed that this was impossible due to the fact that there is no way to tell if the aesthetic value in one application equals to the functionality in another application. This is why I asked them the follow up question if they thought that it was possible to improve the application they did not chose to make them chose that one instead. Although I feel that they at least are close to being on the same level.

5.1.3 Results

Here I will discuss the results and the thoughts about the results from the three tasks.

Searching in a list
Starting with the first task "Searching in a list" we can clearly see that the majority of users chose the functional application. Also none of the ones who chose the functional thought that the aesthetic could be improved to out weight the function of a search box that returned the number of the typed word. Also the one user who valued the aesthetic pleasing application thought that the functional application could be improved to be chosen instead. When just looking at the first question one could think that the functionality just was on a different level then the aesthetic value. But when considering the second question too, we can clearly determine that in this task the users would much rather prefer the functional application. Which means that the function is valued over aesthetic value in this task.

Clearing all input
Moving on to the second task "Clearing all input" we can see that it is quite even between the two applications, but the functional application is still in the lead. Four of seven chose the functional application. And looking at the follow up question no user but one thought the other application could be improved. The last one thought that aesthetic could be improved to make the user chose that one instead. In this task the both application is even and no application is highly favored. I feel neither aesthetic value or function is valued higher in this task.

Retyping same text
In the last task "Retyping same text" all the users chose the functional application. This may be because the function introduced in the functional application saves a lot of time writing so it could be that it is just on a higher level then the aesthetic application. Continuing to the second question, we see that only two users thought that the aesthetic pleasing application could be improved to be chosen instead. Which makes a big majority chose the functional application. Which puts function over aesthetic value in this task.

5.2 Conclusion
In two of the three task the functional application was highly favored compered to the aesthetic pleasing. And most user did not think that the aesthetic pleasing application could be improved to a point where they would chose it instead. In the last task we can not see that any of the two application is favored. This would mean that function matter the most for users in a mobile application. But this does not mean that this is true for all more functionally applications and aesthetic pleasing applications. Even though I did ask them if they thought the applications could be improved to make them switch application, they might not have though of a scenario or an improvement that would make them change. Also, I feel what users value the most is also highly individual. Some have a higher love for functional applications in general and some have higher for aesthetic pleasing. Therefor more work needs to be done in the subject to give a deeper understanding. But I do think these two elements go hand in hand when
designing a mobile application. To have a successful application you will need to have both, with some exceptions. But now we have a general idea to what is more important.

5.3 Future work

For future work in the subject we should make tests with a higher amount of participants, more applications with different functions and aesthetic pleasing elements and from different geographic locations to get a broader spectrum. What we can learn from this study is that remember to record all the thoughts from the user meanwhile they make the test.
References


Feeling Safe in Autonomous Vehicles

Gustav Nordlander
Department of Applied Physics and Electronics
Umeå University, Sweden
contact@gustav-nordlander.se

Abstract. This study was conducted to examine what was needed to feel safe in an autonomous car and how existing models can be modified to help the user feel safe. This was done by researching why people might not feel safe in an autonomous vehicle (AV) by discussing certain elements with potential users through interviews. Research was also conducted to understand what a person needs to feel safe and what could be the reason why the feeling is lacking. The results showed that most people want to have a steering wheel and a foot brake available at all times, even when the car is fully autonomous. Almost everyone also wanted to be able to quickly take control over the vehicle at any time and with that automatically shut down the autonomous part. The conclusion was drawn that people will feel safe in AV when they feel that they are in control of the situation, this includes that they need to feel that they will know what the AV will do in most situations as that allows you to anticipate and plan accordingly.

1 Introduction

Much of the technology around us is evolving to require little or no interaction but still be useful for us. This is possible due to location services and sensors like thermometers and pulse measurement. Our devices should give us information and help us depending on what data it gets from the different sensors. Cars are following the same evolution with them sending data between vehicles to i.e. change the route of your GPS to decrease traffic [1].

Today, a hot topic within automobile technology is autonomous vehicles (AV), which means that the vehicle, in some extent, drive without the help of the actual driver. There are several different takes on this and the levels of automation varies but the main goal is, as with all technology to make it easier, more sustainable and removing the human error that can be caused for whatever reason to reduce traffic accidents. The Insurance Institute for Highway Safety (IIHS) [2] predicts that with autonomous functions such as blind spot assist, adaptive headlights and forward collision warning systems one third of all crashes and fatalities could be prevented.

As technology becomes increasingly common for almost everyone in the industrialized world, technophobia becomes a relevant subject as close to one third of the industrialized population harbor some technophobia, the fear of technology [3]. In addition to this, humans have a tendency to feel safe when in control
and having knowledge of about what is about to happen [4]. So how can we use our knowledge to design an autonomous car in which we feel safe?

1.1 Objective

The aim of this study is to identify what makes a human willing to trust a computer in such a way that he or she can relax while the computer is doing something that can have severe consequences if done recklessly or incorrectly, such as driving. The questions that will be answered in this study is as follows:

- What kind of assurance is needed for drivers to be comfortable with autonomous vehicles?
- How can we design such a product, which elements are essential?

2 Theory

Autonomous Vehicles is today one of the most challenging research topics in intelligent transportation and has been a hot topic of research for the past few years. AV’s that control their own steering and speed are expected to account for 75% of vehicles on the road by 2040 [5].

2.1 Levels of Autonomy

National Highway Traffic Safety Administration [2] have defined different levels of autonomy in vehicles by a hierarchy of five levels to more easily define what kind of product it is.

**Level 0** - No automation at all, the driver is in full control of everything in the vehicle.

**Level 1** - Specific functions are automated, and even if more than one function is automated they operate independent from each other.

**Level 2** - At least two primary control functions operate in unison to relieve the driver of control. At this level the system and the driver share authority over the car.

**Level 3** - At this level the vehicle is autonomous in all normal conditions but may require the driver to control the vehicle if something unexpected occurs, such as abnormal weather conditions.

**Level 4** - The vehicle is designed to operate fully without help from the driver, even in special conditions such as a traffic accident or roadblock which creates heavy traffic.

Note that the highest level of automation is not always desired and the benefits that comes with it depends on the level of automation. For example, the safety can be improved by just adding an automatic breaking system which would fall under category Level 1. Although according to the National Highway Traffic Safety Administration level 3 or 4 automation may need to be applied to achieve noticeable environmental effects.
2.2 Adaptive Cruise Control

One type of AV is the Adaptive Cruise Control (ACC) which is the evolution of normal cruise control where you set a speed for the vehicle to keep. ACC makes often use of laser or radar to pinpoint the location of a vehicle in front. With this information it is possible to adjust speed accordingly so that your vehicle always is at a safe distance from the vehicle in front and therefore i.e. avoid collision from sudden breaks [6].

![Figure 1. Example state machine for an ACC system.](image)

Figure 1 shows the state machine for an ACC system, as it states you always go into normal driving if you use the manual break, where you alter have to enable ACC again for autonomous drive (AD).

2.3 Limitations in AV

The main limitation of any computer generated response is that you cannot pre-program every event, that being said a person would not know what to do in any given situation either. But with the human survival instincts most know how to reduce damage from, say a natural disaster such as an earthquake. On the other hand intelligent vehicles that can communicate also opens up possibilities for security personnel, such as ambulance to get to the point of the accident quickly by determining a good route with less traffic.
2.4 Anthropomorphism

"Anthropomorphism is a process of inductive inference whereby people attribute to nonhumans distinctively human characteristics, particularly the capacity for rational thought (agency) and conscious feeling" [7]. An example of this is if you were to put a name of the guide for a website. This can look like a happy icon or something saying "Welcome, I'm Jeff, how may I help you" rather than a little question mark that would indicate help. Someone named Jeff is something most can relate to as it is a normal name, and that feeling makes the object itself feel less alien.

Studies in anthropomorphism show that a machine with human features such as a name, a human voice or gender makes it more trustworthy. By adding a name, gender and voice to an autonomous car the drivers were less likely to blame the car for faults in an accident [5].

2.5 Safety

While the feeling of safety varies for each individual, Jeanne Priesler at fostering perspectives [8] defines it as "Safe can be defined as free from harm or hurt. So feeling safe means you do not anticipate either harm or hurt, emotionally or physically.". Taking this definition into consideration, feeling safe in a car equals that you do not anticipate either you, the driver or the car itself to make any errors that would result in harm or hurt. The ability to predict what is about to happen next is key when not controlling the situation, e.g. a fully autonomous vehicle. This ties to trust, which McKnight & Chervany [9] defines such as "Trust is a multifaceted concept that can refer to belief that another will behave with benevolence, integrity, predictability or competence".

Trust and safety is also tied to repetition according to David G. Kitron [10] who states that it is only via repetition that risk taking may begin. This happens in steps, at first small risks are taken to avoid bigger risks, which in turn repeats itself, leading to bigger risks. Connecting this to Jeanne Prieslers definition of safety, the same situation is not expected to be different the next time it occurs, which makes someone not anticipate harm or hurt unless that was the case the previous time as well.

2.6 Technophobia

Technophobia is the fear of technology and while there is many definitions, one commonly used was coined Jay T [11] in 1981 and is as follows;

1. A resistance to talking about computers or even thinking about computers
2. Fear or anxiety towards computers
3. Hostile or aggressive thoughts about computers

According to Daniel Dinello [12] one of the reasons for technophobia being increasingly common is the theme in many movies, the post apocalyptic earth, ruled by artificial intelligence. The belief that once a computer is able to "fully
Another reason behind the fear of technology is that people become comfortable using it and therefore not thinking themselves, an example of this is when you start using the calculator for simple additions [13] that you easily can do quickly without any help. This creates the fear that people, in time, not will be able to think for themselves as they are used to something else to do the, for example, calculations for them.

3 Method

The research was primarily conducted by analyzing and comparing existing models and by modifying certain elements of said model to pinpoint the function of each element. This was done by interview along with low-fidelity prototypes in order to create an overview for the user. The interviews were conducted by showing the test subject pictures of interiors of current prototypes. During the interview, the interviewee were asked about specific elements, namely the steering wheel, dashboard and brake, how they feel with the current solution and how it can be modified for them to feel safer.

3.1 Target group

The chosen test group for this study was young males and females aged 23-30. They all needed to have a drivers license and at some point in the last two years been driving regularly. The main reason this specific group was chosen is that they are the ones that will most likely encounter and use Level 3 or 4 automation in the future if the development of said cars will continue.

The test group was found by asking students at Umeå University in Sweden that fulfills the requirements. The results from the test group was not differentiated based on gender or academic backgrounds.

3.2 Lo-fi Prototypes

Seeing this study was conducted without the use of proper prototypes from companies that is currently developing autonomous cars, pictures that illustrates the view of the driver were used. These lo-fi prototypes were created by using AV’s currently in development and by altering those. The components that were studied in this article were steering wheel, brake and dashboard.

3.2.1 Steering wheel

One major component of contributing to the feeling of safety is the ability to take over the control of the car yourself, with other words, the ability to use manual control. And since the steering wheel is a cornerstone for controlling any vehicle, how would the removal of the steering wheel affect the drivers?
Figure 2 shows the interior of Nissans self driving car which they chose to construct in such a way that when you enable the AD technology, here called PD for Piloted Drive, the interior changes such that the steering wheel and dashboard changes into something more enjoyable like the weather or a social media feed and the foot brakes retract so that they are not used without intention.

When Piloted Drive is disabled the steering wheel appears, another dashboard is shown and the break lights up and ascend to their original position as shown in figure 3. Giving the driver the ability to use the vehicle to drive manually.
3.2.2 Dashboard

This study also included whether the dashboard adds a feeling of safety or not. The dashboard allows the driver to get information about the car, such as engine temperature and current speed. During the interview, the test subjects were asked what kind of information they would like to have on a dashboard to feel safe.

![Ford F-150 Interior with ActivePilot](image)

*Fig. 4. Interior of a Ford F-150 using ActivePilot. [16]*

Figure 4 shows Fords solution of a dashboard from their car F-150 which uses ActivePilot, their autonomous technology. Unlike Nissan, they have chosen to retain tachometer and speed meter at all times as well as information about the car such as engine temperature, fuel and oil level.

3.2.3 Brake

The ability to manually brake the car, either with a pedal or via hand brake were also examined during the interviews. The test subjects were asked how they felt in different scenarios, such as removal of the brake. As previously mentioned, Nissans solution to this when in Pilot Drive they remove the light from the pedals and retract them so that they are not a burden when stretching your legs as can be seen in Figure 2. With the removal of Pilot Drive though the pedals light up and descend to their original position so that you easily can access them as seen in Figure 3.

4 Results

The interview was conducted on seven people and the results have been organized depending on which element of the car it concerns.
4.1 Steering Wheel

A bit more than half (5/7) of the people interviewed wanted the steering wheel to remain in its place while autonomous drive (AD) was active and both of the other two needed the steering wheel unless there was some sort of button or way to abort AD quickly and take control over the vehicle again. A vast majority, 6 of 7, wanted the AD to completely be turned off when using the steering wheel rather than just adjusting the course of the vehicle and then, when letting go of the steering wheel, the car would automatically go back to AD. The steering wheel was also prominently (6/7) voted as the most important element for feeling safe when they had to choose one out of the three (steering wheel, brake and dashboard).

4.2 Brake

As the steering wheel, the same five people wanted to have a foot brake available at all time. Even when asked if a hand brake with features such as ABS (Anti-lock braking system), which may reduce braking distance by preventing wheel lock, they felt that a foot brake is safer because it feels reliable. The two people who did not want the steering wheel and brake wanted, along with three that did, an emergency stop button that would force the AD to shut down and also stop the car in an efficient manner. This also had to be mechanical rather than electronic so that it is usable without power. The placement of this button varied between the dashboard or in the middle of the car so that everyone could reach it. Four out of seven thought that it should be in the middle of the car which would enable the driver and/or the passenger to not pay attention to the road because another person in the same car could.

4.3 Dashboard

The only element at the dashboard where everyone agreed was the necessity of the speedometer. Although only four persons needed a fuel gauge to feel safe, everyone wanted the standard set of warning lights for errors within the car, such as engine temperature warning, or a warning that a light is broken. Six out of seven agreed that the tachometer were abundant, as well as the meter for engine temperature, where, as previously mentioned a warning is enough. All seven of the interviewees ranked dashboard as the least important of the three previously mentioned elements (steering wheel, brake and dashboard).

4.4 Improvements

Among the interviewees, a few improvements came up that they felt would make them feel more safe using AD; the idea that the steering wheel would retract a bit but still be close enough to grab. When taking the steering wheel and pulling it out to its original position the AD would stop and you would again have full
control over the vehicle. This would result in a combination of an emergency stop with always having the steering wheel at hand.

Another addition that six out of seven felt would help were anthropomorphism. By adding e.g. a voice to the car that welcomes you when you enter the car would help it feel more intelligent and advanced which in turn would make it more trustworthy.

5 Discussion

Based on the results, the feeling of safety and which elements a person will need to feel safe depends on whether they trust the technology or not. Based on this assumption, the results of the study may differ if the test was redone in a few years, when supposedly AV’s is a common thing. From my own experience, you take more risks and become less prone to anxiety or fear when the same action is repeated several times. For example, even though it is still dangerous to look away from the road when speaking to another person in the car, experienced drivers do that because they are used to nothing happening. It is of course, more times than not, less dangerous since an experienced driver often can act in a manner that minimizes damage. Using this theory, the only way to make people feel safe in autonomous vehicles is to slowly convert all cars to autonomous so that people always feels like it still is the same usual task, only with a new feature. This would also apply to people who harbor technophobia, although the curve would be slower, it would most likely take more time for them to feel safe in an AV. Perhaps they themselves need to test AV’s rather than seeing that it works when other people are using it.

Another common answer for everyone interviewed was that they felt they needed to be in control of the situation to feel safe, or at least feel that they can influence the outcome of the situation at any given time. This is probably the main reason why most people chose the steering wheel as the most important part for feeling safe during AD. This because the steering wheel also allows the driver to avert certain dangers that the brake can not, of course this goes the other way around as well, but it is more common that you need to evade a wild animal running across the road rather than that you need to brake for a wall or something you cannot drive around.

The results from studying the dashboard showed that the dashboard today is something that has been carefully developed over a long time span which means that the information currently shown is what people want to see. Although, the tachometer is most likely something that lives on from old times and is not really needed today, especially since quite a few cars uses automatic gear rather than manual.

The fact that almost everyone felt that anthropomorphism would help them trust the car is quite fascinating and it makes you consider what other areas it can be applied to to help them bloom. The question also rises regarding Daniel Dinellos [12] theory about technophobia being born from movies where machines get so advanced that they will eventually rule the world.
Solely considering the definition of safety previously mentioned by Jeanne Priesler; "Safe can be defined as free from harm or hurt. So feeling safe means you do not anticipate either harm or hurt, emotionally or physically.". With this definition we probably feel safer by having a steering wheel, brake or a dashboard as it allows us to anticipate that we are less likely to get hurt serious if we can control the situation to some extent. But the question to why anthropomorphism makes us feel safer may be harder to answer. It might be because the feeling that it is more advanced makes us believe that we don’t need to do as much to produce the best possible outcome in a dire situation.

Even though the results were similar enough from person to person to come to a conclusion, it would have been positive for the study to broaden the range of the test group by including e.g. older people who does not use technology every day the same way younger people do. Although I reckon the result would be much alike the current as almost everyone wanted both the steering wheel and the brake to remain, this would strengthen the conclusion. It would also be interesting to see if there was a certain group of people who acted like two of the interviewees right now, trusting the technology completely and therefore not wanting brake nor steering wheel when using AD. Is there any common factor between these people like education or is it perhaps based on environment?

6 Conclusions

A majority of people will feel safe in AV when they are common and used everyday. This is because they will know how the AV will react in every situation once it’s been used enough, which in turn creates a feeling of control since dangers can be anticipated and planned for accordingly. This in addition to the fact that most people wanted to keep the steering wheel and brake shows that the feeling of control over a situation is very important.

According to the interviews, anthropomorphism helped the user feel safe as it makes the object feel more advanced, which would lead to the user thinking it is better. It also helps the user relax and make it feel comfortable as a human voice is something that they are used to hearing.

The definition previously stated by Jeanne Priesler [8]; "Safe can be defined as free from harm or hurt. So feeling safe means you do not anticipate either harm or hurt, emotionally or physically." is a good definition that covers almost everything, but I feel like it could be modified to; "Safe can be defined as free from harm or hurt. Feeling safe means that you can control any given situation in such a degree that you can prevent or minimize both harm and hurt, emotionally and physically.".
References

Using augmented reality for learning a second language

Tove Barsk
Department of Computing Science
Umeå University, Sweden
tovebarsk@gmail.com

Abstract. When studying a new language there are several parameters that affects how fast the learning process are. Technology is used a lot for educational purposes, and augmented reality (taking an exciting picture and blending new information to it) is becoming more and more common. This article investigate how to best use AR to learn a new language, or more specific how to use it to improve someones sentence structure and vocabulary. A questionnaire was formed to investigate how students prefer to learn and which methods they use to remember new information better. The questionnaire contained both likert statements and open questions and was sent to 22 people over email and Facebook. The results of the study showed areas where Augmented reality can be of big help. One area is to make the learning situation more lifelike, it can also help improve immediacy and motivation. Mobile AR applications and AR books are two kinds of aid that can make the language learning process more effective. AR can make the user more active and motivated and therefore help them remember words and sentence structure better.

1 Introduction

Learning a new language can be an important part of social integration. When studying a new language there are several parameters that affects how fast the learning process is. For example it is easier to learn a language that are similar to a language you already speak and it is easier when you are young [1]. One big help is if you have someone who knows the language you are trying to learn, who can help you, and who you can talk to. This is because you learn quicker by "doing the real thing, in this case speaking the new language [2]. But it is not always possible to do the real thing, and you might not always have someone to talk the new language with.

How fast you learn also depends on what kind of aid you have. Technology is used a lot in educational purposes, both in classroom situations and at home. Augmented reality, hereafter called AR, can bring a new dimensions to this. AR is a visualization technology that combines real-time three-dimensional computer-generated imagery and real-life footage to create an enhanced representation of reality [3]. In other words you can add a computer-generated layer to the reality by using different kinds of AR technology. Applications already exists that
helps you translate both words and sentences by instant visual translation. By pointing your camera on for example a street sign it can translate it to a chosen language. By adding a virtual layer to the sign, it looks like it is written in the new language. There also exists AR books, where the user can look at the book through some kind of AR device and see three dimensional models.

One important part when learning something new is to actively learn instead of just listening or watching. The use of AR in educational purposes is increasing, there are a lot of areas where AR can be used to improve the learning situation. For example it can be used to study the human body, in a three dimensional and lifelike environment. Augmented reality can help create situated learning even if it is not possible to do the real thing. As Mehmet and Yasins says "Displaying information by using virtual things that the user cannot directly detect with his own senses can enable a person to interact with the real world in ways never before possible" [4]. The focus in this report is to determine the effects it has on the learning process when learning a new language. Learning a new language is not just remembering words, but it also involves studying sentence structure.

1.1 Objective

The objective of this paper is to investigate the usage of augmented reality for learning a second language. Both the effect it has when learning new words, and the effect it has when trying to improve sentence structure will be studied. The following research questions have been formulated to investigate the problem:

- How can an augmented reality be used to improve vocabulary?
- How can augmented reality be used to improve sentence structure?

By making a survey about learning techniques and combining the results with previously done studies about learning a second language, these questions will be answered.

2 Theory

There has been a lot of advances in the use of technology in educational purposes. As Jayson [5] says technological innovation is changing the way that students learn and universities teach. Augmented reality is one technology that can bring a new dimension to this.

2.1 Augmented reality

Augmented Reality (AR) is a visualization technology that combines real-time three-dimensional computer-generated imagery and real-life footage to create an enhanced representation of reality [3]. Unlike Virtual Reality, where users are immersed in an environment that is entirely computer generated (i.e. the real
world is replaced by a virtual world), AR allows users to experience a modified version of the real world by superimposing virtual images onto real views of the world (the real world is “augmented,” or supplemented, by virtual elements) [6]. In other words, augmented reality takes an existing picture and blends new information into it. As Granmar [7] writes, AR was made as an protest against VR, they thought VR locked up the user, so they created AR to open the user to an improved reality. She also says that it is the simplicity to use AR for everyone that will make it big, and that AR had more practical use than VR.

2.2 Augmented reality and education

Augmented Reality is changing education. What started out as something that was simply “cool” has become a way to engage learners like never before [8]. It is not always possible to study something in its natural environment, it may be dangerous or expensive which is why education is one important domain for augmented reality. AR can provide situated study environments for students. AR has matured enough to perform more and more important functions in education field with its distinct superiority [9].

Some areas where AR can make a big difference in educations is augmented biology and physics education. For example it can be used to study the human body, what the organs exists of and how they look [10]. It already exists educational augmented reality applications and educational augmented reality books. An augmented reality book (also called “the magic book”) is a book where you can look through some kind of AR-device and three dimensional models will appear [9]. The possibilities are many, the book can contain interactive elements in a text, hyperlinks, images or even interactive assignments. In other words there are many ways of improving the students learning experience by using augmented reality. As Wu et al [11] says AR technologies help learners engage in authentic exploration in the real world. They also says that ”AR could enable learning content in 3D perspectives, ubiquitous, collaborative and situated learning, learners’ senses of presence, immediacy, and immersion, visualizing the invisible, and bridging formal and informal learning. Kesim and Ozarslan says [4] it can even be applied to help with collaborative tasks since it is possible to develop innovative computer interfaces that merge virtual and real worlds to enhance face-to-face and remote collaboration. They also say that Augmented reality makes the impossible possible and its potential in education is just beginning. Augmented reality interfaces offer seamless interaction between the real and virtual worlds.

2.3 The learning process

There are many ways of learning something new. Different individuals feel they learn best in different ways. Something that helps is repetition. Even if your are really focused when learning something, the knowledge can be quickly forgotten if not repeated. After 24 hours you have forgotten about 50-80 percent of what
you learned [12]. Our brains are constantly recording information on a temporary basis: scraps of conversation heard on the sidewalk, what the person in front of you is wearing. Because the information isn’t necessary, and it doesn’t come up again, our brains dump it all off, along with what was learned in the lecture that you actually do want to hold on to [12].

Spaced repetition is a method of learning which is proven to embed new knowledge in the long-term memory [13]. The idea involves repeating the learning events with space in between instead of trying to learn it in one session. For example you could start by learning a couple of words in a new language one day, then repeat a second time in two days and so on. This is an effective method to make new knowledge stick [13].

Another effective method is active learning. Active learning is generally defined as any instructional method that engages students in the learning process [14]. So instead of just listening or seeing, the student is speaking or doing something active to learn faster. By using spaced repetition and active learning the student have a better chance of getting the information to the long-term memory.

**Learning strategies for second language** When learning a second language, there are seven broad aspects to consider, you have to work on reading, writing, vocabulary, grammar, sound acquisition and listening comprehension [15]. Different learning strategies work best on different aspects. Since this article focuses on sentence structure and vocabulary, only strategies to acquire these will be mentioned below. Neiman et al (as cited in [15]) mentions some strategies:

- **Grammar**:
  - Following rules given by text
  - Inferring grammar rules from text
  - Comparing a language you know to the second language
  - Memorizing structures and using them often

- **Sound acquisition**:
  - Repeating aloud after teacher, native speaker or tape
  - Listening carefully
  - Talking aloud, including role playing

- **Vocabulary**
  - Making up charts and memorizing them
  - Learning words in context
  - Learning words that are associated
  - Using new words in phrases
  - Using a dictionary
  - Carrying a notebook to note new items
The cone of experience  Edgar Dale came up with a cone of learning (also called the cone of experience, see figure 1). The cone is a graphic figure that shows the hierarchy of learning through the programming of real experiences. It shows that to the best way of learning is at the bottom of the cone, where you ”do the real thing”. It also says that we remember 10 percent of what we read, 20 percent of what we hear, 30 percent of what we see, 50 percent when we combine seeing and hearing, 70 percent of what we say and write and at last 90 percent of what we do [2]. In other words, you want to structure your learning so that it falls under some of the lower sections in the cone. When learning a new language it is therefore better to both say and write it, than to just hear it. The cone also says that if the teacher show someone how to do something, they will probably remember, but if the teacher involve them in a meaningful way, they are more likely to understand it better [2]. Augmented reality can give more opportunities to train or study in lifelike environments.

![Dale’s Cone of Experience](image)

**Fig. 1.** Dale’s cone of learning [2]

3 Method

A questionnaire was designed to assess the best way to use augmented reality applications when learning a new language. The questionnaire was a web based form and was distributed over social media and email. The statements and questions asked about learning and remembering in general. The results of the questionnaire were then combined with previously done studies about learning a second language in particular, to determine the usage of augmented reality when learning new words and sentence structure.
The questionnaire consisted of 9 likert scale statements and 2 open questions. The choice to use open questions was made so that the person answering would have some questions where they could answer more freely. The choice to use likert scale questions too was made to minimize the risk that the people answering the survey did not answer at all which can be a risk with open-question-surveys.

On the likert scale the choice was made not to have a mid-point since studies [16] show that it can make a person answer more positively to please the "interviewer". As Ron Garland says "Research provides some evidence that social desirability bias, arising from respondents' desires to please the interviewer or appear helpful or not be seen to give what they perceive to be a socially unacceptable answer, can be minimized by eliminating the mid-point" [16]. By not having a midpoint the person answering the survey has to take a stand at every statement and cannot just choose to be neutral. The number of points on the scale then was decided to be 6 since research confirms that data from Likert scales becomes significantly less accurate when the number of points drops below five or above seven [17].

The statements was written in an "agree or disagree" way. The 6 answers were as follows:
• Agree Very Strongly
• Agree Strongly
• Agree
• Disagree
• Disagree Strongly
• Disagree Very Strongly

3.1 Test group

The survey was sent to 22 students i Umeå and Luleå. Gender was of no importance so it was sent to both female and male participants. The target group consisted of students. This decision was made because the questions asked a lot about learning and how people prefer to learn, and people who study learn new information on a daily basis.

3.2 Questionnaire

The questionnaire had 9 Likert scale statements which are shown below in figure 2. The statements was meant to determine how important different methods of learning were to people who are trying to learn and remember new information, for example words and sentence structure. For example if people preferred to speak and listen rather than write and listen.

The Questionnaire contained a few Yes or No questions that led to different
open questions, this was to check what the person knew before redirecting. For example, those who knew what augmented reality is (77.3 percent of the participants), a following question appeared about how they thought augmented reality could be used in educational purposes. Examples of the open questions are shown in figure 3.

4 Results

The first three likert statements was constructed to research how people preferred to learn new words and sentence structure, by listening and speaking, by listening and writing or by reading and writing. The statement that got the most positive results was listening and speaking. 86.3 percent answered that they agreed that felt that they learned best by listening and speaking(fig 4) On the statement "listening and writing" the result was very scattered, and the statement that got the least agreement was reading and writing (see figure 5). So ordering them by preference:
Another statement was that "I prefer to learn by listening to a lecture rather than reading" and most of the answering answered on the agreeing side. But this statement did not have as overwhelming agreement as the statement "I remember new information better if I discuss them with someone". 72.7 percent answered Agree very strongly, and all the participants answered on the agreeing side of the scale (see figure 6). Another statement on the questionnaire was "Manipulating objects helps me to remember what someone says". The results of this question was very spread. There were more answers on the agreeing side of the scale, but the answers were concentrated to the middle, more neutral area (see figure 7). The last statement was about preferring visual aid when someone is presenting information, and almost all the participants agreed that it was
Fig. 6. I remember new information better if I discuss them with someone

Fig. 7. Manipulating objects helps me remember what someone says

helpful to have something visual supporting the information being presented (see figure 8).

Fig. 8. Manipulating objects helps me remember what someone says

4.1 Open questions

One of the open questions were if the participants use any method to remember new information better and if so, what kind of method. The most common answer was repeating information in their head, both in the moment and over a longer period of time. Another common answer was taking notes while listening to new information. Two of the participants answered that they use mind
maps and wrote questions to quiz themselves. One of the participants also wrote
that being able to connect the information to something helped them remember
better. An example written by one participant was “when learning a new noun,
seeing a picture of it makes it easier to remember.”

On the question asking how the participants thought augmented reality could be
used in educational purposes, most answers talked about visualizing things, for
example “visualize things that can be hard to grasp with text and voice only”. The participants talked about using AR as a complement to help students understand and also getting a more "hands on" learning situation. Some also wrote that AR can give the student a perspective that is not possible without it. One answer also mentioned that it could help motivate students and make them more focused.

5 Discussion

Steve et al [18] points out that educators often utilize games to assist students
to easily grasp concepts. They also say that with AR technology, educational
games in the real world can be designed. These kind of games can add a 3d layer
to real world things. In other words, AR can be used in educational games to
help motivate the student. Combining this with the strategies mentioned above
to learn sound acquisition and grammar, educational games where you memorize
structures or participates in role play can help improve the students language
skills. This is supported by the study since all the participants remembered new
information better by discussing it with someone and if they had some kind of
visual aid, which a game can provide.

As mentioned earlier in this article, the cone of experience shows that the more
active we are, the bigger the chance is that we remember the new information
including words and grammar. Since AR can create 3D layers on the real world,
and can let us manipulate this layer, it can be of big help to help the user be-
come more active. When learning new words, AR can translate things in it’s
natural environment and that way help us make connections which will help
us remember. This was supported by the results in the questionnaire since the
participants felt that they remembered things better if they could make some
connections.

One application for using augmented reality in educational purposes in gen-
eral is to make the learning situation as lifelike as possible. When learning a
new language, being able to talk to someone, or discuss with someone, can be
the purpose for wanting to learn. The study showed that a majority of the par-
ticipants felt that they remembered new information better if they discussed
it with someone else. And Neiman et al (as cited in [15]) supports this when
talking about strategies to learn grammar and sound acquisition, for example
repeating aloud after teacher, native speaker or tape, listening carefully, talk-
ing aloud, including role playing. When studying a new language the user often has some kind of literature. By using AR books when studying, the user can see videos, translations and conversations by using some kind of AR technology. Supported by the study, most participants preferred to learn by speaking and listening instead of just reading, and this is one place AR can be of big use. Instead of just using books to write and read information, AR can be applied to let the user view videos and listen to people speaking the language. AR books will also increase the activity of the user.

Another thing AR can help with when learning new words is immediacy. Instead of having to look up a word when wondering what something is called, AR technology can add a layer with words in the real world. This also makes the learning more situated, when spotting an orange for example and wondering what it is called, the user can immediately see the orange and the name through some kind of AR technology. In other words, the user learns words in context and learns words that are associated, which was two of the strategies Neiman et al (as cited in [15] mentions for improving vocabulary. This can also be applied to translate words, for example signs and notes, AR can add the translation immediately on the sign or note. Applications for translating words directly on the AR-layer of the real world already exists, for example google translate application allows the user to point their camera to a street sign and get an instant translation on the screen. [19] One of the strategies for learning a second language is to compare the with a known language. By using AR to translate words and sentences directly on the screen allows the user to directly compare and make connections to the first language.

Another strategy that Neiman et al (as cited in [15] talks about is carrying a notebook to note new words or sentences. By having a instant translation AR application and being able to save the words or the pictures from the camera, this strategy will be easier to carry out.

6 Conclusions

Both the questionnaire and previously done studies on learning a second language shows that Augmented reality can indeed make a difference when learning a second language. One area where it can be of big help is to increase the activation level, so that instead of just reading and writing new words and sentences, AR can make the user learn while integrating with the real world. It can also make the learning situation more lifelike and more situated by using AR technology to help with translations, words and even videos. The user will have easier to make connections and remember words and sentence structure better. By using the AR layer to show translations, sentences and maybe even videos it can also increase immediacy. Supported by the study, visualization is an important part of remembering new information better, and by using AR technology you get more possibilities to do this.
Another area where AR can be applied to help learn a second language is motivation. Augmented reality can help make the learning situation more stimulating and fun. Just as gamification can be used on regular applications to increase motivation, it can be combined with augmented reality to make it even more interactive and stimulating [20]. Augmented books is another example where the user, instead of just reading, can use AR technology to interact with the book. These books can contain elements which (when using AR technology) for example can show videos of people talking.

Here follows some suggestions on future work. By using some kind of prototype, tests can be performed to determine how big the affects are when using augmented reality instead of the regular learning methods. The test could be extended to people who study a second language instead of students since they probably have more specific methods to remember new words and sentence structure and not just new information in general.
References

Biohacking: Evaluating the benefits of Near Field Communication in the human body.

Carl Holmertz
Department of Applied Physics and Electronics
Umeå University, Sweden
carlhelmertz@gmail.com

Abstract. Objective: Near Field Communication (NFC) is a technology that is growing together with the development of technologies such as smartphones. A new way to use NFC today in 2016 is to implant a NFC chip in the human body, often the hand, in order to provide services, functions and possibilities to the user and to ease tasks in the daily life. The aim of this study was to determine what different functions, with NFC implants, that would make people want to implant NFC chips into their bodies.

Method: The selected participants of this study were given a survey to fill out which contained examples of both existing functions and functions that could possibly exist in the future. The participants were then asked to rate these different functions and add additional functions that would make them willing to consider implanting a NFC chip in their own body.

Results: 35 persons participated in the study and the results show that the majority of the participants preferred the functions that made their daily life easier such as opening and accessing doors or paying in stores which created a life where they did not need to carry around keys or a wallet etc. A lot of the answers and comments showed that some participants were worried about the security with the NFC ship.

Conclusion: The companies should start with improving the security with NFC to prevent the information on the NFC chips from being hacked and stolen. When security no longer is a problem they should start focusing on functions that ease the daily life of the user like payments, instead of high-tech functions like a smart fridge etc.

1 Introduction

Imagine if we could open all doors at work, at home or even our car with just our hand. Imagine if we could pay at the store, or if the doctors could see our medical records in case of need, by just reading a chip in the hand.

NFC stands for Near Field Communication and is a wireless technology that uses Radio Frequency Identification (RFID) technology and short range radio waves in order to provide communication between devices. This communication requires a distance shorter than 4-10 cm between the devices in order for it to work [1]. The possibility to use NFC technology with payments in the grocery store already exists today in 2016. We can pay by just touching the cashier
terminal with a NFC equipped credit card. NFC is also popular for identification systems at universities and at companies where the students or employees have cards that work as keys to open doors in order to gain access to different rooms.

Biohacking is a term that can be described as a practice of biological experimentation for a self-defined purpose using a variety of Do It Yourself (DIY) devices and techniques in a nontraditional way. The term was first used online in 2008 and describes a broad set of activities such as DNA manipulation and implanting of computer chips in the human body [2].

In 2016 there are at least 200 people in Sweden that use NFC in biohacking with implants of NFC chips in their hands (U. Swanson, Netlight, personal communication, April 28, 2016). These people use this new technology to do different things such as opening doors at work or at the gym, or storing a digital portfolio or a business card. One can wonder if this is enough, or does NFC allow us to create new ideas and usage areas.

1.1 Objective

The objective of this study was to identify the different usage areas that makes an individual willing to implant an NFC chip into the body.

People have already started with NFC implants in their bodies for different reasons, but what applications should the companies focus on in order to get this technology to be used by the general public? In this study the objective is to answer the following question:

- What different functions will make people implant an NFC chip in their body?

The results of this study were used to create a guideline for markets and companies in order to know where their focus should lie in the future development of NFC. After the objective was finished, the results were presented in a way that helped specify what usage areas the market and companies should focus on, in order to get users.

2 Theory

In order to understand the possibilities with NFC, it is important to understand the technology behind it. This section includes information about the technology and how it is used today and possibly tomorrow.

2.1 The technology

Near field communication (NFC) is one of the most promising and emerging technologies [3]. It is a wireless technology built on Radio Frequency Identification (RFID) [3]. NFC works through short range radio waves in order to communicate with another NFC device. Most of the new smart phones have NFC implemented. Since NFC tags can store data, the device (e.g. phone) scans
the data and then acts upon it. Hence NFC is built on RFID systems and creates a system that allows a two-way communication between devices instead of a one-way communication as earlier systems had [3].

The purpose of NFC is to make it easier and more convenient to exchange digital information, make transactions, and connect digital devices with a touch. Thanks to the NFC’s ability to both read and write information to devices, the future usage of NFC is believed to be bigger than e.g. standard smart cards [4].

NFC protocol exists of two statuses; active and passive mode. Tags and cards are always passive and therefore they do not need any power in order to work. In a passive mode only one side of the communication needs power in order to read and write information to the tag or card. Active mode is when both the initiating device and the target device generates their own radio fields in order to communicate e.g. NFC connection between two smart phones with their own power supplies [4].

2.2 The use of NFC today

The most common use of NFC today occurs in the passive mode, often presented in key card systems or in credit card payments. If we look at the past, shopping has been associated with cash or credit card payments, but later different possibilities to pay with a smart phone have been developed. Today we have the possibility to use our phone to pay for goods or to pay our bills [5]. The development of mobile payments has grown fast and over the last ten years we have been introduced to various different payment methods, such as Short Messaging Service (SMS). Even though these different methods provide good service, they are still not good enough or ideal if we view it from the traditional payment context [6]. A fresh way of paying is with mobile credit cards which is a contactless credit card payment using only a mobile phone with Near Field Communication technology (NFC). Through this technology we do not need a physical connection between the payment terminal reader and the consumer since the transaction can take place with just simple wave (a couple of centimeters from the terminal) or a simple touch [5].

Another usage area of NFC today is for Mobile Coupons. Instead of cutting out a piece of a newspaper, the user can now redeem their coupon through their NFC enabled smartphone [7]. The user does not need to carry a rewards card for every store, he or she can now load the coupons on their phone and the reward points go straight to the account without ever digging in the wallet for a card [3]. Functions like these show that NFC has the ability to change the market in a way that the coupons and commercials can be targeted towards a specific user depending on the users personal preferences, instead of a more general targeting such as in the newspapers [7].

Unlocking doors or computers is another way to use NFC technology. This function makes companies, universities, hotels and homes etc. more secure and more available for the user. This can be done through both a smartphone or NFC keycards. The technology allows users to open doors or even their computers by just waving their phone or card towards a reader [7].
Transportation is a necessity whether you live in Sweden or any other country. With NFC we have the possibility to pay for our rides just by waving our phone or card on the bus, at the train gates or at the platforms. Today, this is more often used with keycards, with NFC tags in them, instead of a smartphone [7].

Mobile gaming is an industry with tens of millions of users worldwide play games daily [8]. In 2011 a popular mobile game called Angry Birds implemented an NFC function where two devices could connect to each other by tapping the devices together while having the game running. When this was done, new levels were unlocked on the phones. Even though this game is not fully functioning on NFC, it gives the game developers an idea of what is possible to do with NFC in gaming. A Smart Home is created by different solutions for carrying out basic functions like opening doors, turning off alarms, or recognizing objects through sensors [9]. When all of these functions are combined the result becomes useful and exciting, this has created the idea and practice of Smart Homes. One way to achieve this is to strategically place NFC tags on different places around the house. When scanning these tags with your phone they will automatically activate a process, such as changing the temperature in the house, turning off and on lights etc. The user can choose what different functions the tags should activate to create the wanted Smart Home [10].

Smart Homes can include a varied amount and types of functions. For example, a smart refrigerator with NFC technology could detect if any compartments are open, tell the temperature and energy consumption etc. Different sensors collect this information and can then be sent to the user via a smartphone with NFC just by waving the phone in front of the fridge [11].

3 The use of NFC tomorrow

The future of NFC has the potential to be a pillar in our daily lives. Thanks to the possibility for NFC to connect with such a wide variety of devices, it creates a possibility for NFC usage to grow in an unlimited way [7]. Some functions are not really "future functions" anymore, because different companies already have them in development but they are not yet out in the open market.

Imagine if the fridge could, through NFC technology, keep track of what groceries the user has in the fridge. The fridge could then tell the user the amount of food in the fridge, and if it has expired or not. If NFC works together with Wi-Fi in this task, the fridge could then show shopping lists depending on the user’s preferences for food and drinks [7].

NFC technology can be used in healthcare for identification of patients in hospitals, which will be very useful since patients have different illnesses. If a doctor would confuse the illness of a patient with the wrong patient a fatal error could occur. With NFC we can easily protect patients from ever getting mixed up. Even if a person will faint or be unconscious when arriving to the hospital, the doctor could access the patients’ medical record to avoid further problems such as giving penicillin to a patient that is allergic to it, which could be fatal.
4 Safety

The era of the Internet and technology such as smartphones has taught us about security, and about the risks of being hacked and getting our information stolen. This is an aspect to process and discuss. There are different aspects to the safety issue and a big question is if the pros outweigh the cons in technology development.

4.1 Eavesdropping

Because of NFC being a wireless communication technology it creates a security problem with eavesdropping. NFC uses radio frequency waves in order to communicate and this allows an attacker to use an antenna to receive the same signals [12]. There are different factors that affects the possibility to eavesdrop and at what distance it is possible [12]. The only way to prevent these kind of attacks is within the application itself and not in the technology NFC. A possibility is to create a secure channel for NFC, which means that two NFC devices agree to a standard key protocol in order to establish a shared secret between the two devices. This secret is then used in order to provide confidentiality, integrity and authenticity of the transmitted data [12].

5 Method

First the identification of the future usage areas with NFC implants were evaluated and identified through research about NFC today and the future thoughts about the usage of NFC in general. The results of this constructed the foundation for developing the questions for the objective.

With the information gathered from the identification of the future usage areas the objective was researched. The different usage areas, that will make an individual willing to implant an NFC chip into the body instead of just using a chip in the wallet or such, were identified through a survey. Before the the real study took place, two pilot studies were made in order to evaluate and develop the form itself. In the first section of the finished survey, the users were asked to answer some basic questions about their age, gender and occupation.

In the second section the participants were asked to rate the usefulness of examples of NFC functions that are available today. The participants were asked to rate the usefulness of the functions on a one-to-five scale where number one represented "Not Useful" and number five represented "Very Useful". At the end of the second section, they were asked if there were any other NFC functions (that were not in the list) that they thought were useful.

The third section was very much like the first. The participants were asked to rate the usefulness of some future functions. At the end of this section they were asked to think freely and use their imagination to think of future usage areas with NFC that they thought could be useful.
In the fourth section in the survey for this study the participants were given a list of functions with NFC, some existing today in 2016 and some that can exist in the future. The participants were asked to mark which of these alternatives that together could create enough possibilities in order for the user to consider implanting an NFC chip in their hand. They were then asked if they would consider doing an NFC implant if their marked functions existed. At the end, they were given the possibility to add some functions that they would like to have in the future.

The fifth and last section of the survey contained some basic questions about if they knew what NFC was before this study, if they ever used NFC in some way, and how they used it.

5.1 Choice of participants and language

The participants for the study were chosen based upon their field of work or study. The goal was to have participants that already had some knowledge about what NFC is, how it works, and the risks and possibilities with the technology. Therefore the study first and foremost included students of the Interaction and Design program at Umeå University in Sweden. These participants study engineering with focus on interaction design with programming, user experience (UX) and user interface (UI), and the majority of these students are well up-to-date about technology development. The study was created in Swedish because the participants were all from Sweden. These choices were made in order to ease the study and to get more creative responses, compared to doing a survey with people not studying or working in this field or doing a study in another language than Swedish.

6 Results

There were 35 participants in this study and the average age was 25 years of age. There were 64.3 % men and 35.7 % women. As planned in the method there where 78.6 % students and 21.4 % working people. The participants were asked to rate existing functions of how important they thought they were and the result is illustrated below.
Fig. 1: Payments: As shown above, the participants rated payments as very useful. Only one test participant rated payments as "not useful", but 60 percent (21 participants) rated payments as "very useful".

Fig. 2: Coupons: The view on the possibility to get rewards and points to get discounts is quite widely spread as shown above.

Fig. 3: Access: The test participants rated Access as a useful function.
Fig. 4: Mobile Gaming: The usefulness in connecting phones in order to play together and to earn extra points.

Fig. 5: Smart Home - Example: book the laundry room with a NFC tag.

The third section of the survey included functions that are not available today for the users but could be in the future. The participants were asked to rate these different functions with respect to usefulness.

Fig. 6: Smart Fridge that keeps information about what groceries that are in it.
Fig. 7: Healthcare: could be used as identification of patients at hospitals and the doctor could get quick access to the patients medical records if needed.

Fig. 8: Verification and identification: for safety (Ex: creditcard + code + NFC chip)

The participants were in the fourth part of the survey given a list of functions. Some already existing today, and some that could exist in the future. They were asked to mark which of the alternatives that together creates enough possibilities for the user in order to consider implanting a NFC chip into their hand. The result of this can be seen in figure 9.

After the participants marked their wanted functions they were asked if they would consider doing an NFC implant in their body given that their wanted functions existed.
Fig. 9: The alternatives that together create enough possibilities for the user in order to consider implanting a NFC chip into their hand.

Fig. 10: The results of the test participants’ answers to the question if they would consider implanting a NFC chip into their bodies given that their wanted functions existed.

At the end of this section, in the survey, the participants were asked if there were any other usage areas or functions that they would like to have, in order to
consider implanting a NFC chip into their hand. They were asked to think freely and to use their imagination. Examples of answers, translated from Swedish:

- "The functions that can make me leave my wallet and keys at home."
- "100 % guaranteed security, so that there is no possibility to hack information with an NFC reader."
- "Opening doors, paying etc"
- "Identification, for example when getting medical help, flying or such."
- "If I know that the chip can be implanted in my body in a safe way, and that I can be sure that these chips are not gonna be outdated in a near future."
- "Possibly for identification, in order to avoid carrying around driving licenses or other ID cards."
- "Unlock my phone and maybe measure my body temperature."
- "There are a lot of existing and future functions that are cool which would make a lot of daily tasks easier, but I care too much about my personal integrity to think this implant would be worth it. Maybe sometime in the future when we know more about how it works."
- "... to be honest I would never consider implanting a NFC chip in my hand today. As long as there are organization like NSA and CSEC. As long as there are no extreme laws to prevent these organizations to do surveillance on people I would never feel safe with this kind of implant..."

The last and finishing section of this survey included some questions about the participants' knowledge and usage of NFC today. There were 89.3 % that knew what NFC was before they took part in the study, 7.1 % did not know about NFC in advance and 3.6 % answered "maybe". The participants in the survey were asked if they had ever used NFC before and if they had they were asked to give an example. 35.7 % had never used NFC before, 7.1 % did not know if they had and 57.1 % had used NFC. Some examples of how the participants had used NFC before:

- "Connect phone to speaker system, booking the laundry room, open doors at work, SL-card (Train-card)"
- "Payed in the grocery store."
- "Sent pictures and contacts between phones."
- "To pare two phones for file transfer and to pare a phone and a speaker."

7 Discussion

From the results, it is quite clear that the majority of the test participants are willing (52.4%) or "maybe" willing (38.1%) to consider an NFC implant in their bodies. Only 9.5% answered that they would not consider implanting an NFC chip.

About the different examples of functions with NFC that were given in the survey it was three that stood out as less useful than the rest. Those three were coupons, mobile gaming and the smart fridge example. Payments, access,
health care and verification and identification were rated as very useful. This is confirmed by the comments and answers the participants wrote in text. The examples show that the majority of the answers contained ordinary functions such as opening doors, access to the phone, identification etc; functions that could make the daily life easier for the participants.

The author of this paper did an interview with Unn Swanström that has a NFC chip implanted in her hand. The interview gave information about how NFC was used today but also ideas of how it could be used in the future. One major opinion from Unn that corresponds to a lot of the answers, was the opinion about security. She was scared to store sensitive information on her NFC chip today with respect to the security and safety problems. In the comments from the test participants we can see answers such as ":.. I care to much about my integrity to think this implant would be worth it".

To further examine this topic, research could be done on other ways to use NFC, like in a bracelet or such, in order to reach the same usage areas and usability. It could also be interesting to look at the possibility to use fingerprints instead, but that would make this study too broad and it would create a lot of other questions such as safety issues etc. NFC could be used in various ways, but this study was focused on NFC as an implant in the hand because of its media exposure lately and the increasing number of people that are doing the implant. (U. Swanström, Netlight, personal communication, April 28, 2016)

8 Conclusion

Before the companies start to develop and market new functions, they should start to improve the security with NFC. This study shows that a lot of people are interested in NFC and its possible functions, but they are concerned about the safety of its use.

If there were no problem with the security and safety, the companies should focus on functions that could ease the daily life of the user. Functions like these could be the simple ones like accessing doors at work, home or even to the car. It could also be functions that could access the users’ medical records if needed or identification like a passport or a driving license. A way that NFC is used today and that people still need, are easy ways to pay in stores.
References

Ditching The Hamburger Menu
- An analysis of current menu design

Anton Forsberg
Department of Applied Physics and Electronics
Umeå University, Sweden
id12afg@cs.umu.se

Abstract. Menu design has become somewhat standardized and the use of what is called a hamburger menu is seen in apps and websites frequently, but some are calling for change. The menu layout has come under scrutiny in recent years for being inefficient and for losing user engagement. This study investigated the validity of these claims and tested the menu layout against another newly proposed layout, bottom navigation, to see how the two fared with experienced users. It also tried to determine which of the options the users preferred and to identify possible issues with the two layouts. The results show that while there is no immediate need to redesign the hamburger menu it is indeed very replaceable as long enough attention is paid to the choice of icons.

1 Introduction

The hamburger menu has gone from famous innovation to infamous user experience (UX) nightmare over the course of its lifetime. Most smartphone users see this many times over the course of a day, but few know it by name. In short, the hamburger menu (Figure 1) is an icon symbolized by three horizontal stripes found in the top left corner of most mobile web pages and apps. It is there to show us that there is off screen content available on click. The icon’s resemblance to a bun-meat-bun constructed hamburger has led to its playful yet deceiving name.

Fig. 1. The Hamburger Menu
The recognizable icon was first introduced in the Xerox Star operating system in 1981 and was intended to be a simple way to design menus with limited use of screen real estate [1,2]. The design stuck and has become the industry standard when designing app and website layouts for mobile platforms. In recent years the design has, however, received much criticism for its low discoverability and efficiency. It is supposedly responsible for lost user engagement and making simple tasks difficult for new users [3]. This study intends to find out what impact this menu design has on everyday users and if it can be done better.

1.1 Objective

The primary focus of this study lies in determining whether or not the hamburger menu layout has any negative affects on usability when compared to other layout designs. Furthermore, if this turns out to be the case, this study aims to provide suggestions for replacing this layout. More specifically this study will attempt to answer the following questions:

– Do experienced smartphone users experience issues with the hamburger menu compared to other common alternatives?
– When provided with another option, will users still prefer the hamburger menu?

2 Theory

The following subsections explain some of the terminology used as well as motivating the need for the study.

2.1 Mobile menu design

There are several different approaches to menu design when it comes to portable devices and smaller screens. In the early days of smartphones the most common layout in apps would be a springboard or launchpad (Figure 2) design. This type of layout acts as a navigation page where users can navigate to different sub-pages. Later the list menu was introduced, which is similar to the springboard layout but elements are displayed in a list layout with some hierarchical order [4].
While still in use widely on an operating system level, the layout has since been abandoned by most apps after Aza Raskin designed the *Side Drawer*. To access this menu he placed an icon with three horizontal stripes, the design we now know as the hamburger menu \[4,5\]. This design is now the most common in mobile apps and is also very present in mobile web pages. The design choice has lately been criticised however and several major companies have recently started to distance themselves from the icon and are figuring out different ways to navigate within an application \[3,6\].

### 2.2 The issues with the hamburger menu

The main issue with the hamburger menu lies in its low discoverability. The menu in itself offers no clue to what lies within it. For all the user knows it could be contain a list of dog species, friends birthdays or simply be empty. It is silent and secretive. An interface should be focused on
recognition rather than recall in the sense that users should immediately be aware of what action a button will perform rather than have to have this memorized from previous attempts. In this regard the hamburger menu offers only a simple clue, the three horizontal stripes that somewhat resemble a list. The hamburger menu has its contents hidden off screen in what is called an “off canvas” approach to menu design. According to Krug websites should be designed with a focus on scanning rather than reading. The idea is based on that users rarely read through entire pages but rather scan the page looking for a specific item. In this regard the hamburger menu, hiding its content, does not fare well. Furthermore, the minimalist implementation of the icon does generally not imply being clickable. It does in no way resemble a “real world button”. This, however, is outside the scope of this article and is more due to the advancement of minimalist design than to the use of this icon.

2.3 Usability

In 1995 Jakob Nielsen wrote down ten design heuristics which are nowadays considered paramount for good usability design. This study builds a major part of its motivation on the premise of these heuristics. However, some of the terminology used in this study may need further explanation. Discoverability is a term that aims to describe an item’s capacity to be found or discovered. A person can for instance research their own discoverability by performing an internet search for their own name and seeing what shows up. More specifically it refers to how likely a user is to find something in an app or in our case, a web page.

3 Method

To determine if the hamburger menu is replaceable by more efficient menu layouts and how it affects users compared to those layouts a test was performed. The test subjects were let to use a web page using either the hamburger menu or another type of navigation, namely bottom navigation. The users were then given a set of tasks to perform within the environment and were encouraged to share their thoughts during the process. After the test was complete the users were shown the web page with the other menu layout and were asked questions about which one they would prefer. This study also aimed to determine whether or not the hamburger menu was the preferred layout within a system if the user could choose another form of navigation.
3.1 The Environment

To get viable data a mock web page was designed and developed with two different menu layouts baked into it. One layout was the hamburger menu which was fixated in the top left corner and the other one was a bottom navigation menu with only icons on it (Figure 3). As the bottom navigation is now a proposed way of designing navigation within apps by Google [6] with limited menu items, this was chosen as the alternative layout. The hamburger menu had the items ordered in it with some hierarchy but the bottom navigation was, as per the design principles, without hierarchy or descriptive text alongside the icons.

Fig. 3. The application used to test the Users. Here it is showing both types of navigation, but the users were only subjected to either one of them. The hamburger menu (top) and bottom navigation (bottom).

The web page was created using the Angular framework [10] and the scaffolding tool Yeoman [11] was used to provide quick and easy set up and testing. The menus used contained five different items: profile, news, friends, events and settings.
3.2 The Testing

The ten test subjects were all chosen from Umeå University and they all had more than two years of experience using smartphones and apps. Furthermore they also considered themselves very comfortable using smartphone apps. The tests were carried out in the students’ native tongue Swedish.

Each user was presented with a web page using either of the two menu layouts, to mimic an A/B testing scenario [12]. To test what issues arose when the subjects would navigate within the environment they were given different assignments of varying difficulty. For each assignment the accompanying researcher would note each action the subjects took and they were also encouraged to explain what they were thinking as they were performing the tasks. This provided qualitative data for the study and was gathered to show possible flaws with either of the two layouts.

After the test was finished the test subjects were asked a few follow-up questions to determine if they had any issues with the navigation within the app. The users were also given the opportunity to familiarize themselves with the other menu layout and decide which one they were most comfortable with using.

The questions at the end of the test were as follows, (translated from Swedish):

– What was your experience navigating through the app menu?
– Did you encounter any difficulties while navigating within the application?
– Which of the two menu options do you feel is the easiest to use?
– Which of the two menu options is the most appealing to you?

4 Results

The users who were presented with the hamburger menu had no problems at all completing their assignments. All subjects completed their assignments without performing any mistakes. This was however not the case with the subjects who had the bottom navigation layout. 100% of these subjects navigated to the wrong item while performing a certain task.

The users had varying feedback on the different layouts but most commonly the users with the bottom navigation layout mentioned the icons being misleading or hard to interpret.

However, after being allowed to familiarize themselves with both layouts, all users considered the bottom navigation being the simplest layout
and a majority of users, (70%), also considered the bottom navigation the most appealing.

5 Discussion

The fact that the subject with the hamburger menu had no issues with navigating the application might seem overwhelming at first, but it needs to be put in perspective. Since all users were experienced smartphone users they were not expected to have any trouble. This layout is after all one of the most occurring menu layouts seen today and users should be somewhat familiar with it. The fact that the other subjects had trouble with the bottom navigation shows us something important however. Since this layout did not have the descriptive text in the menu items, the icons were supposed to give the users enough information alone. This is where the layout failed and it shows us the importance of having good icons with a strong symbolism to what action they perform when clicked. The users also mentioned this themselves, which only further emphasises the importance of a designers choice of icons. Another solution for the problem could be to have the text showing underneath the icons. This may provide some difficulties however, due to limited amount of space.

The fact that most users do consider the bottom navigation to be the easiest to use of the two layouts illustrates how the issues they had with it are of little importance after repeated use. The fact that a majority of users considered the same layout to be the most appealing could either be due to the ease of use it provides or it is simply the novelty of the design.

6 Conclusion

The study can determine that experienced users do not have any issues using the hamburger menu, meaning that there is no immediate need to replace it. The results however indicate that the hamburger menu can be successfully replaced in many applications by implementing a bottom navigation. Not only was this way of navigating considered easier to use by all subjects, it was also considered as more appealing by a majority of the respondents. However, usage of this layout requires a limited amount of navigation options as it is not recommended to use more than 5[6].

It can also be concluded that first time users are heavily reliant on the usage of icons and how well they represent their performed actions. This shows that if a designer opts to use icons without explanatory text
within a menu design then the user experience of the navigation is heavily reliant on the chosen icons. With this information the bottom navigation can be used as an on-screen alternative to the hamburger menu in certain applications, as long as icons and explanatory text are considered.

References

Interaction limitations on mobile devices related to screen size

David Bergvik
Department of Applied Physics and Electronics
Umeå University, Sweden
dabe0052@cs.umu.se

Abstract. With the current trend of mobile devices increasing in size, this paper tries to identify possible issues related to screen size when performing one-handed interactions on mobile devices. Furthermore, it evaluates if there are constraints that arise when users’ reach becomes insufficient for the desired interaction. The interactions tested were based on placements recommended by Android’s design principles, with focus on one-handed interactions. A testing environment was created and installed on Android devices with different screen sizes, and the effective time between interactions could be measured during sessions of qualitative testing followed by a post test interview. The results showed that the effective time when interacting with a larger mobile device is significantly longer. Furthermore, it indicates that as screens become larger, the physical reach of the user becomes more relevant. This paper can serve as a baseline for further investigations on the effects of screensize within the spectrum of one-handed interactions.

Keywords: Screen-size, one-handed interaction, mobile devices, interaction limitations

1 Introduction

Personal digital devices are as of this point a big part of everyday life, with more than 4 billion mobile subscriptions in 2014 [1]. Mobile phones has emerged as a major communication and interaction resource between each other and our environment. With a steep development curve, mobile phones are getting bigger and faster by the minute. But with quick changes and new trends within mobile development, principles that used to be a key part tends to be forgotten.

In 2007, Steve Jobs and Apple introduced the first generation iPhone [2]. With a screen-size of 3.5 inches, the company claimed that it is the optimal screen-size for a handheld smartphone. Jobs described how a larger screen will have physical size that is too big for a human hand. Regardless, the rest of the phone industry started to produce bigger phones each year. Apple themselves also had to keep up with this new trend, and left their 3.5 inch screen size. The question that
remains is: Was Steve Jobs right back in 2007? Does the demand for bigger and faster, make the industry forget about the interaction abilities necessary for a handheld device. Is it all up to the manufacturer, or is it the software that need to adapt to these changes?

2 Objective

The objective of this study is to identify the interactive issues caused by the physical size of the device. The main area for this will be one-handed interaction sequences with smartphones, where the testing interfaces are guided by industry standard principles [3]. It will also try to evaluate how the size and screen size of the device will affect the ability for the user to interact correctly. This study will also try estimate possible differences in effective time when interacting with bigger devices, and establish when the devices prevents certain interactions. The questions that will be used to evaluate this are:

– Are there interactions that gets affected by the size of the mobile interface?
– How is the effective time of common interactions affected by the size of the interfaces?
– How does hand-size (reach) of the user effect the one-handed interaction?

To answer these questions a testing software will be developed. Using the software, qualitative testing will be done with subjects with experience of mobile devices. After analyzing the results of the tests, possible solutions to these problems will be presented.

3 Theory

In 2015, more than 84 percent of all mobile devices run some version of the Android operating system [4]. Even though that also includes tablets and watches, it is enough of a majority for basing this study on. It also comes with an open source base for creating free applications. With this operating system (OS) comes design principles [3] necessary for creating an Android application as well as hardware buttons (see Figure 1). These buttons will be harder to include in testing software seeing as they also interact outside of the created application.

Fig. 1: Navigation bar of a nexus 4, made by LG and Google.
3.1 Core design concepts

When mobile developers create a new application, they have to obey some of the rules and concepts that have been constructed by the OS owners [5] [6]. The owners can reject an application that does not fully follow these concepts, in order to achieve a higher standard and convergence in terms of design within their application-pool. This study will be conducted on Android devices and will therefore focus on the design principles of that OS.

Android Screen sizes In today’s market, a vast variety of different mobile devices exists with different screen-sizes [7]. In order to help mobile developers with creating applications that work with all these devices, Android recommends designing for four different versions in terms of user interface (UI) [8]:

- Small: Screens below 3" (Wearables and Phones)
- Normal: Screens between 3" to 4.5" (Phones)
- Large: Screens between 4.5" to 7" (Phones and tablets)
- Extra large: Screens larger than 7" (tablet)

Android Key Elements Android applications have some navigation elements and components that classify as OS standards [9], and as such their position on screen will be key to find interaction combinations that are executed in a majority of applications. Back navigation is used to navigate back to a previous view, and is to be located in the top left corner of the screen (see Figure 2a). In the main view, a menu button is generally used to display a side menu, also displayed in the top left corner of the display (see Figure 2b). The menu-button will be replaced by the back-button when the user navigates to deeper parts of the application.

In the top right corner, a lot of different buttons are displayed. Usually a search-button (see Figure 2a), a button to display a right-side menu or drawer or a navigation to go to a view deeper in the application. As an overlaying element in the lower right corner, Google presents a floating "add"- or new-button (see Figure 2b) [10]. This element is considered in this study for its position always on top of the rest of content in that view, usually located in the lower right corner.

3.2 Default hand placements

How a mobile phone is held does at first glance seem trivial, but depending on how the hand is positioned, the users reach will be limited by the length of the finger and the hand placement used when interacting [11]. This forces the user to change the phones placement in order to perform certain interactions. The position of hardware buttons on the phone is primarily the source of the phones hand-placement, seeing as some of the hardware-buttons usually have to be pressed in order to unlock and start interacting with the phone. In this study the user is allowed to use any starting hand placement, as the tests concerns interactions only within the digital interface.
3.3 One hand usage

In a field study done by Hoober [12], observing 1333 people in public locations, 49% of the people that was using a mobile device was only using it with one hand.
This correlates with what Karlson et al. [13] shows in their study, that using a single hand to interact with a mobile device is common practice for users that are traveling while interacting with the device. Furthermore they explain how users want to be able to interact with one hand only, but that some mobile devices, especially touchscreen interfaces are not designed for this type of interaction. Karlson's study did not focus solely on touchscreen devices, but determined that the hardest interaction to execute for a right-handed user is from NorthWest <-> SouthEast (see Figure 2).

![Figure 2: Visualization of the hardest interaction on a mobile held with users right hand](image)

### 3.4 User benefits of different screen-sizes

In 2011, a study conducted in South Korea researched the effects that screen size has on perceived mobility and enjoyment [14]. The authors Ki Joon Kim, S. Shyam Sundar, Eunil Park explain how they with a video-based test evaluate these parameters based on three different screen sizes (3.5", 5.7", 9.7"). Their results conclude that a smaller screen gives the user a greater perceived mobility while a bigger screen offers the user more enjoyment.

### 4 Method

By examining the typical elements in an Android application interface, a qualitative user study was conducted. The study consisted of two parts; a user test followed up by an interview. With this approach, the results consisted of measurable test data and in-depth responses to that data. The tests were performed on ten subjects that will be randomly picked students from Umeå University, that used a smartphone on a regular basis.
4.1 Interactive testing software

A testing software is created using Android Studio to be used on different Android-devices. The goal of this part is to establish statistical support for possible differences in execution-time of identical tasks with varying screen sizes. Since a major factor in this type of testing is the subject’s hand and finger size, this must also be addressed in the tests in order to collect viable results. The testing process consists of two parts;

Setup

*Credentials* The testsubject is presented with a series of screens, where personal credentials are to be answered. This data will be correlated with the result of the rest of the test. The testsubject will state their name (see Figure 3a) and which hand the device will be held in during the test (see Figure 3b). Selecting which hand will be used, is a critical piece of data when analyzing the current design principles that mobile applications are to follow.

![User interface of credentials input](image)

(a) Input field for name  
(b) Hand selection for tests

*Fig. 3: User interface of credentials input*

*Measurement of reach* In order for the results to be consistent over different subjects over different hand and finger sizes, the subjects reach capabilities over the screen is measured. To accurately measure reach, the hand should be flat with the thumb pointing upwards. The device is then placed on top of the hand (on the index-finger), with the usb-connector against the first joint of the thumb. The reach is then measured by interacting with the slider on the interface of the device (see Figure 4a), sliding it as far as possible towards the top of the screen.
The setup is after those steps complete. A screen is displayed that briefly explains the test, and encourages the test subject to start the test when ready (see Figure 4b).

**Test: Sequence of interactions** The subject will undergo a series of tests where the objective is to perform a sequence of different interactions, which will be measured by miss-clicks and time to completion. When the user is ready and starts the test from the initial layout (see Figure 5a) buttons will appear one at a time, and will disappear on press. The test will focus on the outline of the screen, so buttons will be placed along the outline in corners (see Figure 5c), and will always be followed by a button in the center (see Figure 5b). With this approach each sequence can be singled out without bias to placement of previous buttons. When the test is complete (see Figure 5d) the score can be sent to a server where all test results are saved. All results can be viewed thru that server via a web-client.

4.2 **Interview**

To follow up on the user’s thoughts and experiences when undergoing the tests, an interview will be conducted. The interview consists of a few standard questions for all subjects, and a more subjective part where the focus is on the question why and how? Why did certain situations occur during the tests? Why did you perform certain actions? How would you have wanted to interact instead? The results from the testings software and the interview will be compared and analyzed in order to find patterns and possibly a higher bound for screen-size.
Fig. 5: Testing interface with buttons

(depending on hand/thumb size). This higher bound will mark where the screen-size makes interactions awkward or hard to perform, and will be calculated by finding long execution-times and comparing that with the post-test interviewee.

**Limitations** These test will measure execution-time over different devices and screensizes, however it will not cover how hand-placement will impact the results. The phones that will be used are unlocked when handed to the test subject, and they will be told that no hardware-buttons will be used for the tests. Since the software do not recognize if one hand is used, the tests will be supervised to verify this.

5 Result

5.1 Test: Sequence of interactions

The results from these tests are conducted on three (3) different devices within the large-spectrum (that was mentioned earlier) [8]:

- Samsung Galaxy Core: 4.5 inch (Core)
- LG Nexus 5X: 5.2 inch (5X)
- LG G4: 5.5 inch (G4)

All of the test subjects chose to use their right hand for the tests, and the results are divided and presented according to the specifications in the Method-section:

**Total time over different screen sizes** The effective time for the test subjects to complete all sequences over all three devices are summerized (see Figure 6)
The results show that performing interactions with the Core is faster than the other two devices. Furthermore, the majority of test subjects had a faster execution-time with the 5X inch screen compared to the device with the largest screen (G4). The abnormality in this result is test-subject 2 performing the test with the 5X almost double the time compared to the median time. This occurrence is from the test-subject almost dropping the device due to the strain from reaching towards the next button.

**Combined time over each sequence** Following are the results of effective time in each section (between each button press), combined over all devices and test subjects (see Figure 7a) and combined over all test using the Core device (see Figure 7b). Due to the abnormality in one of the tests, that result is not included in this chart, but will be included as an effect of the test itself. The reason is for the results of the cart to reflect general successful tests measuring effective time.

These results highlights three major spikes in the time-data, sections 2, 4 and 6. We can establish that the button placed in the top opposite corner of the hand used is the hardest one to reach, followed by bottom opposite corner and top corresponding corner (see Figure 8).

**Reach over total time** All subjects measured their reach on the device before conducting the sequence tests, in order to determine if reach has an impact on the effective time when interacting with mobile devices (see Figure 9). The scale on which the reach is measured is abstract, and does not transfer between the three test devices. Because of that, the results of this part is only measured on a device basis.
(a) Combined time for each section of the test over all 3 devices

(b) Combined time for each section using Core (4.5 inch)

Fig. 7: Charts with combined results for reach test and total test time

Fig. 8: All buttons within the test spectrum, time-spikes are marked with respective sequence number

(a) Total time with reach on Core (4.5 inch)  (b) Total time with reach on 5X (5.2 inch)  (c) Total time with reach on G4 (5.5 inch)

Fig. 9: Charts with combined results for reach test and total test time

Looking the results obtained from the Core (see Figure 9a) it is not possible to determine any obvious pattern, as the data is scattered over the entire reach spectrum. After examining the results from 5X and G4 (Fig 9b and 9c respectively) there is a resemblance to a pattern that the effective time is decreased when the reach increases.
5.2 Interviews

The interviews that were performed after each test resulted in 80

6 Discussion

This study requires the test subjects to be somewhat equal, even if factors like reach are taken into consideration. It relies on that all subjects have the same skillset when handling a mobile device. This is not easy to determine, and from watching the tests it became easier and easier to spot the subjects who perform these kind of interactions on big screens on a regular basis. They have certain techniques that allows them to reach further than with a steady grip. This did however not change the fact that the effective time still was slowed down because of the large screen size. And even in this controlled test environment there was one subject who almost dropped the device while trying to interact, and many subjects that came to a stall when they had to change the positioning of the device within the hand itself. When out "in the wild" (in the real world) it becomes a much bigger problem to not have a firm grip of your device. The reason for this is all due to the physical size, that forces a shift in hand placement. By observing the test subjects performing these tests, it became clear that when they were not able to reach a button from the default hand placement, they came to a stall while readjusting the device within their hand.

When going deeper into the test results (see Figure 7a), it is very clear that the time that is lost on the larger devices originates from interactions within the South and NorthWest region of the screen (see Figure 2 for details).This correlates well with the study by Karlson et alciteOneHandedUse that was mentioned earlier. All test subjects managed to complete the tests, which shows that the screens used in this study did not prevent any interactions. It did however force many subjects to stop interacting with the screen and change their hand placement before the interaction was possible. From the results from the reach test, the data is not as clearly stated. There is a resemblance of pattern for the 5X and G4, but for the Core there is no pattern at all (see Figure 9). What this shows is that when the screensize of the mobile device is increasing, the reach and thumbsize becomes a more significant factor. This test group does however only consist of ten subjects, which by its small size makes it harder to recognize a deviation from an actual finding.

The tests performed in this study are limited to simple interactions such as pressing buttons, as this is the way that users interact with the elements and components mentioned earlier in Section 3.1. It does however not include more complex interactions such as swiping or dragging objects. In order to go further into the issues that this article is focusing on, test with those kinds of element should be considered in order to get a deeper understanding of the users behaviour and potential issues that arise. The number of devices that were tested could also increase. Because of technical issues, the test environment was not installed on a forth 4.0 inch device. This size is not within the large spectrum , but would be a good benchmark to compare with. After observing how the Core
was used, it showed that not a single test subject had any issues interacting with a device with a screen size of 4.5 inch. There where no adjustment of hand placement or awkward positioning. As we can see from the Core results (see Figure 7b) the biggest total time difference between the sections are 2.5 seconds for 9 tests. If we compare that with the overall results that have almost 12 seconds over 27 tests, this is a significant difference that shows how the issues with interactions that are affected by screen size actually get enlarged. These findings should affect the mobile software developers in a way that they will have to take this into consideration when developing for a wide range of devices. It should also be a part of the design process when manufacturers creates new devices, in order to meet as many needs of their customers as possible.

7 Conclusion

Using a larger device with one hand is not an easy task nowadays. Its becoming harder and harder for users to interact with their mobile device in a way that they have before. From the data obtained in this study, it is certain that the screen size of mobile devices do affect interactions with it. Not only is it more time consuming, it also creates a physical strain on the hand. From the results (see Figure 6) it is clear that the screensize do affect the time it takes to interact with it. It also point to that even a minor enlargement (from 5.2 to 5.5 inch) changes the effective time when interacting.

Interactions located at positions 2, 4 and 6 (see Figure 8) are directly affected by the screensize of the mobile device, and this should be accounted when designing mobile applications that supports different screen sizes. The users reach when interacting with mobile devices becomes a more significant factor when the screen becomes larger, as the user will be forced to change hand postioning in order to perform said interaction. Mobile devices have over the years become larger, both as a part of our daily lives and also as the physical size keeps growing, and the effects of this are clearly shown in this study. Steve Jobs stated that a 3.5 inch screen is optimal for handheld interactions, and with the results from this study this statement still stands strong.

References

Evaluating the potential and limitations of Google Glass

Joakim Zakrisson
Department of Applied Physics and Electronics
Umeå University, Sweden
zakrisson88@gmail.com

Abstract. Google Glass was originally predicted to become a very popular device amongst the general public. This turned out to not be the case as Google has since January 2015 halted it’s production of the Google Glass device. It was found that Google Glass had issues with its current devices, such as hardware limitations, no clear usage area, high price tag and users experience they are clunky. Although Google Glass has also been used together with experiments within the fields of telemedicine and teletoxicology where they might become useful. But first the device would need to resolve some of its issues. To evaluate the potential and limitations of Google Glass different methods were used, such as theoretical examination and a survey. The theoretical examination was looking into the background of the Google Glass, which area of interest the device might have and what kind of limitations the device had. And the survey was sent to persons who are generally interested in technology which improves the chances that they have used a wearable device and is more likely to leave relevant feedback to the survey. The survey had open questions about wearable devices in general where the person could leave feedback for devices they had used. Also the survey examined the pro’s and con’s with each wearable device and was investigating what made a person not be interested in the Google Glass. The data gathered from the theoretical examination and the survey was used together to examine why Google Glass did not become as popular as other wearable devices. It was concluded that Google Glass did have limitations with the current devices which had a negative effect on the user experience, also people do prefer to wear a wrist-worn device instead of glasses. Furthermore it was concluded that Google Glass might have usage areas within different work-related fields such as telemedicine.
1 Introduction

Google Glass is originally developed by Google X, which is a semi-secret research and development facility created by Google. Google Glass is a popular device within the category of smart glasses. The original intended purpose of the smart glasses was to be a hands-free display of information available to most smartphone users. According to Cellan-Jones [1] Google announced as of January 2015 that they would stop producing the Google Glass prototype but is still committed to keep developing the product where project Glass has been “graduated” out of Google X.

Fig. 1. Forecast of sales for wearable devices during 2014-2016 [2]

Looking at other wearable devices it is something that has increased in popularity during the last few years. The total amount of unit sales of wearable devices reached 36 million units during the year of 2014. Smart glasses did account for 1 million units and smart watches 4 million units. During 2015 the total amount of wearable devices being sold was predicted to 72 million units with smart glasses at 2 million and smart watches at 26 million units [2]. Wearable sales is really taking off and it is something that is becoming very popular and attractive to the general public. With sales for wearable devices increasing this
much especially smart watches, why is smart glasses not increasing in popularity the same way?

1.1 Objective

The purpose of this research is to identify what is making wearable devices attractive and why Google Glass has not increased in popularity the same way other wearable devices (e.g. smart watches) have. During this research the following questions was taken into consideration:

- Are there any hardware or software limitations that users have experienced with Google Glass which might restrict the device’s popularity?
- What functionality is a user generally looking for at a wearable device, and can such functionality be added or any other improvements be made for Google Glass to make them more attractive to the general public?

Theoretical examination was made to examine these questions and also a survey was made with purpose to figure out what is making a person become interested in a wearable device and to analyse what could possibly make the person more interested in Google Glass.

2 Theory

Smart watches has increased in popularity the last few years with the release of devices like Apple Watch. During the second quarter of 2014 Samsung stood for 73.6 percent of the global market share of smart watches with 0.7 million devices sold [3]. However, during the same period in 2015 Apple had 75.5 percent of the global market share for smart watches with 4 million devices sold [3]. This data shows how popularity has increased for smart watches between 2014-2015 and how much Apple is dominating the market since Apple Watch was released.

Smart glasses did not become as popular as predicted; In late 2013 Google Glass was predicted to by 2016 become a mainstream product and sell million of devices [4]. This however did not turn out to be true since Google halted it’s production of Google Glass as of January 15, 2015 [1]. Also while Google Glass together with other smart glasses where still being produced and sold in 2014 they only accounted for 1 million devices sold while smart watches sold by 4 million devices [2].

In a scientific paper about the importance of well designed applications it is stated that an important aspect with wearable devices is not just the aesthetic value of the device itself but also the design of the applications [5]. They also conclude that well designed applications with good functionality will make a person more likely to enjoy the wearable even though it does not have as good aesthetic value as other devices.
2.1 Background

In an article by D’Onfro [6] which is about the first release of Google Glass devices it was stated that Google Glass was announced as the first smart glass device offered for personal use back in 2013. D’Onfro continues, that in 2013 a handful of people could get their hands on Google Glass where they could pick up a pair of the device at designated Glass "Basecamps" located in San Francisco, New York, Los Angeles and London. Furthermore D’Onfro states that the device had a price tag of 1500 USD by then and the first employees responsible for working at those "Basecamps" were hired without knowing what they would be doing because Google Glass had been a secret project. She continues that during training those employees had to practice by pitching random Google products. Also in a scientific paper about who might be interested in Google Glass it is stated that Google started distributing a limited number of Google Glass to selected people such as developers and researchers in 2014 [7]. They also state that Google then announced they would start selling Google Glass to the general public during the fall of 2014.

Google themselves stated in January 15, 2015 that the Google Glass project is moved out of Google x because Glass at work has been growing and they are seeing big developments with Glass in the workplace [8]. Google also state that Google Glass will instead have their own team at Google and is now focusing to build for the future where they eventually will release new versions of Google Glass. According to Google the last pair of Google Glass explorer edition could be obtained until the 19th of January 2015, thereafter Google’s explorers program was closed.

2.2 Area of interest

According to a research article about clinical and surgical applications of smart glasses [9], there is currently great interest in using Google Glass together with health information technology within health care to improve patient care, decrease cost, and increase efficiency of health care professionals. The article also states that hands-free web connected devices have great utility to health care professionals by providing access to patient records for viewing or input. Furthermore they continue by stating that studies show that smart glasses have great potential to reduce the time doctors spend managing Electric Health Record systems and data entry where the front facing camera also has potential to help with surgeries, improve documentation of clinical data and analyse diagnostic tests. Furthermore the article states that additionally the camera can be used with telemedicine. The article also shows that smart glasses still have flaws and before they can be recommended for general medical use improvements to the hardware and specialized software need to be developed.

Google Glass have also been examined for use in telemedicine as stated in an article where a study was made where two paramedic teams triaged the same 20 patients [10]. Telemedicine is considered the use of telecommunication and information technologies to provide clinical health care at a distance. In
this study one team was using telemedicine together with Google Glass which they used to communicate with an off-site physician disaster expert. The other team was using no telemedicine and was acting as they would normally do. The conclusion of this study showed that there were no particular increase in triage accuracy when paramedics were evaluating disaster victims using telemedicine with Google Glass than paramedics who did not use the device. Furthermore the study that the team of paramedics who used telemedicine required more time to evaluate and triage the patients than the team who did not. The study also states that this might change in the future since Google Glass at this time have some hardware limitations which might had influence on the result and also the paramedics were unfamiliar with this kind of technology.

Google Glass have also been examined within the area of teletoxicology. Teletoxicology is like telemedicine but with focus on chemicals and to counsel persons who might have been exposed to chemicals at a distance. A study was conducted where 19 consults were attempted by a medical toxicology consult staff [11]. Google Glass were used at this study and provided real-time video communication with the patient and medical staff. The same study was also made without the use of Google Glass. The study concluded that in the case where teletoxicology were used together with Google Glass 89 percent of the cases were considered successful according to a supervisory consultant whereas without the use of teletoxicology 59 percent of the cases were seen as successful. This study shows that Google Glass do have potential within this area of use.

Health and wellness areas will see the biggest early jump in market numbers since sensors in wearable devices can be cross-purposed for everything from fitness to technology for aging which is stated in an article by Ranck [12]. Ranck also concludes that many customers are unsatisfied with medical devices that look like medical devices and prefer design aesthetics of a company like Apple. Also according to Ranck wearable technologies developed for military use are also making their way into the sports and outdoor markets.

2.3 Google Glass limitations

Some limitations and issues have been encountered with the Google Glass device. Described below are some limitations that was encountered with a study where Google Glass was used as an experiment together with telemedicine [10].

- The need for internet connectivity with dependence on either Wi-Fi network or Bluetooth connection with a smart phone.
- Short battery life which further decreased during video streaming.
- Microphone position which also picked up lots of background noise and was not always responsive to verbal commands
- The screen was very bad outdoors, the contrast made the screen very hard to see in outdoor light.

Furthermore the study concluded that there were also software problems which occurred. Furthermore the study concluded that there were a significant lag in
live-video broadcasting with a delay of more than 20 seconds between the disaster drill site and the off-site physician, although audio communications was almost instantaneous. They continue by the software also had no lock-functionality to prevent the possibility of inadvertently halting the streaming. The study also states that the device was unable to opt-out the frequent OS-updates which complicates appropriate technology selection and research planning.

Another study also states similar issues with the Google Glass device. During this study Google Glass were researched within use of teletoxicology [11], here a total of 19 consults were made where Google Glass provided real-time video communication with the patient and medical staff. The persons using the device experienced hardware limitations such as video lag and interruptions which occurred during video-feed. One of the consulsants also experienced audio-lag and in another case the consulant were unusable to use the device because of both video and audio lag occurred.

In a research paper where Google Glass was used together with video-streaming it was concluded that the device also has temperature problems [13]. The results from this research paper shows that Google Glass has a threshold at 55 degree celsius from where it will ask the user to stop the current activity for the Glasses to cool down. They continue by stating that after exceeding this threshold the glasses will operate at a lower frequency for the CPU, also lower the voltage. This will result in less power drainage and slower temperature change but at the cost of user experience. Furthermore to test these temperature problems the authors used the Google Glass to live stream a 5 minute video via bluetooth and Wi-fi. Their results from this test shows that the temperature for the glasses rises significantly the first 3 minutes of live-streaming and then gradually stabilizes. The article further states: "Our results suggest that Google Glasses do not seem to comply to user expectations in terms of good user experience for the streaming applications. Streaming applications are not suitable for longer duration of use on Glass with its current hardware configuration as it leads to high power consumption and temperature rise. High power consumption leads to low battery life and high temperature which can be dangerous to user skin" The same research paper also does web browsing experiments together with the Google Glass where 16 different web pages are measured to check the power and heat consumption of Google Glass [13]. The research concluded that for example wsj.com needs 3 redirects before going to the main mobile webpage of the site. This would increase the page load time by 2-3 seconds on the Google Glass. They also state that apart from time, redirects also leads to wastage of power on the glasses. The article also concludes: "Extra time to load and extra power dissipation due to redirections is not good for the quality of user experience on wearables. We think that the number of redirections should be minimized and better be avoided by better webpage design for such sites." From the websites tested youtube.com was the most optimal for the Google Glass. It uses efficient network protocols and image formats to deliver the content faster. Ted.com could similary to youtube convert all .jpg images to .webp to reduce the power consumption on the Google Glass by 12 percent with no change to the user experience.
3 Method

To fulfill this research a survey was made to find out why smart phone users also want to use a wearable device and how they use it to complement their smart phone. The survey was also used to find out what is making a wearable device interesting for the user and what kind of functionality they are looking for in such device. Furthermore the survey was used to figure out which properties are most important and what kind of functionality or characteristics is the smart phone users looking for in a wearable device. Also the survey was used to identify why a user would not be interested in Google Glass.

Similar studies or research articles were used as a source to help answer this. Furthermore, other research papers will be used with focus on smart glasses to find out why they did not become as attractive to the general public as other wearable devices e.g. smart watches and why Google Glass did not become as popular as people first had predicted.

3.1 Survey

The survey was used to identify what functions a smart-phone user is generally looking for in a wearable device. The survey was also examining what the user did find lacking with Google Glass and possibly what functionality or other characteristics could change or be added to make them more attractive. Another objective with this survey was to identify why the person is currently using or not using a wearable device and what was the persons reasoning for this.

The survey was split into two different parts. The first part consists of questions about Google Glass where the survey will try to gather pros and cons of people who did use them. Furthermore some general questions about Google Glass was asked, for example if the person would be interested in owning their own pair of Google Glass and what the person would be willing to pay for the device.

The other part is about other wearable devices. Also here the general opinion about the wearable devices a person has used was asked including pro’s and con’s. Furthermore the survey was identifying why a person would prefer another wearable device instead of Google Glass. It did also examine which wearable device are most appealing to the person and which property in such device is most important.

The aim was for the survey to reach 10-20 persons who were interested in technology and preferably have used some kind of wearable device to be able to share their general opinions about them.

3.2 Survey target group

The target group for the survey was persons who currently own a smart phone and are within the age of 20-30 years old. The persons within the target group was university students who were currently studying within the field of technology or computer science. This specific target group was selected mainly because
a relatively small range of age indicates that the persons have similar experience with smart phone and wearable technology. And persons studying within the field of technology would more likely be interested in wearable devices and would therefore be more likely to have used some wearable device themselves. Also persons who currently own a smart phone was mandatory because a wearable device is often used as a complement to the user’s smart phone.

3.3 Theoretical examination

To complement the survey, theoretical examination was done to review scientific articles with focus on wearable devices or smart glasses. The purpose was to evaluate what current wearable devices are being used for and why smart glasses, such as Google Glass, did not take off the way it was predicted to. To identify this, theoretical examination was made for other wearable devices such as smart watches to analyse what characteristics were making such device popular. Furthermore, Google Glass current usage areas and limitations with the current version of Google Glass was examined.

4 Results

4.1 Theoretical examination

In this section conclusions gathered from theoretical examination will be shown. Below is a list of some hardware related limitations experienced by the users in the experiments with Telemedicine [10], Teletoxicology [11] and video-streaming [13].

- Dependence on either Wi-Fi network or Bluetooth connection.
- Short battery life.
- Microphone not always responsive and picking up background noise.
- The screen was very bad outdoors, the contrast made the screen very hard to see in outdoor light.
- Overheating issues.
- Lag while video- or voice streaming.

The hardware related limitations are not the only problems which users did experience with Google Glass. In the video-streaming experiment they also identified that due to websites being poorly optimized for devices such as Google Glass some websites drains more power than needed looking from a user-experience perspective [13]. The experiment show that for example Ted.com could lower the power consumption by loading the website by 12 percent only if they did change all images on the site from .jpg to .webp. It’s also important that websites use efficient network protocols to become more efficient. This shows that limitations experienced with the Google Glass could be altered by providing optimized content, like applications and websites.
4.2 Survey

In this section data gathered from the survey is shown.

**Fig. 2.** Data from survey: Whether a person has used Google Glass

**Fig. 3.** Data from survey: Whether a person would be interested in Google Glass
**Fig. 4.** Data from survey: How much the person would be willing to pay for Google Glass

**Fig. 5.** Data from survey: Which wearable devices has the person used
**Fig. 6.** Data from survey: Weighted average for what score each wearable device had. The scores range from 1-5 where 5 was given to the most interesting device.

**Fig. 7.** Data from survey: Weighted average for which property is most important. The scores range from 1-4 where 4 was given to the most important property.
Furthermore the survey had some open questions where the most relevant answers will be shown below:

For the 36 percent of the persons taking the survey who are not interested in owning a pair of Google Glass there was an follow-up question as of why not. The results are listed below:

- Limitations, such as battery time and overheating issues.
- No clear usage area.
- Technology is not fascinating.
- The device being too expensive.
- They feel clunky.

Also the survey had another question for people who have used a wearable device such as smart glasses, smart watches or wrist-worn activity trackers. In this question they could state their general opinion about such device. The results are listed below:

Smart glasses:

- Easy to obtain information.
- No clear usage area.
- Can get tiresome to use after a while.

Smart watches

- Easy to access information from smart phone directly on the wrist.
- No clear usage area.
- Too small which cause irritation.
- Too much focus on fitness applications.

Wrist-worn activity trackers

- Help motivate a more active lifestyle due to result logging and being able to compete with old results.
- Relatively low price tag.
- Easy to access information from smart phone directly on the wrist.
- Clear usage area
- Can get tiresome to use after a while.

There was also a general question about what would make the person taking the survey prefer another wearable device instead of Google Glass. The results are listed below:

- A lower price-tag
- Prefer wearing a wristband over glasses
- A clear usage area
- Better hardware specifications
- Appearance
- Wrist worn devices is more subtle and is easier to wear
5 Discussion

From the survey and theoretical examination it can be concluded that Google Glass had some limitations, especially hardware related ones. These limitations were experienced by users when the device was being used in different experiments such as within telemedicine and teletoxicology. Such limitations include problems with poor battery time and the device overheating. Also other limitations such as application lag while video- or voice-streaming, microphone picking up background noise and the screen having a bad contrast which made it hard to see, especially in outdoor light.

Other than hardware related issues it is also challenging to find a clear usage area for the Google Glass device. This was shown as a result in the survey where the persons who took the survey did find wearable devices helpful since they can easily provide the user information without having to pick up their smart phone. But other than that many stated that the Google Glass had no clear area of use which made them not as interesting as smart watches or activity trackers. Persons who took the survey did find that activity trackers had a very clear usage area which was to motivate the user to a more active lifestyle where smart watches was like Google Glass, a device to help provide the user information without having to look at their smart phone. Although people did prefer wearing a smart watch compared to smart glasses because a wrist-worn device is more subtle and can easily be brought anywhere by the user. This would mean that even though the smart glasses and smart watches had similar functionality most people would prefer wearing the watch.

Another problem with the Google Glass is the high price tag they initially had. As stated in an article by Business Insider [6] the Google Glass had an initial price tag of 1500 USD which is equal to 12100 SEK [14]. However, from the survey we can gather the following results for what a person would be interested to pay for Google Glass: 42.86 percent would be willing to pay between 0-2500 SEK, 50.00 percent would be willing to pay between 2500-5000 SEK and 7.14 percent would be willing to pay between 5000-7500 SEK. No person who took the survey would be willing to pay more than that. This clearly shows that persons are not willing to pay as much as 12100 SEK for the Google Glass device but rather somewhere between 0-5000 SEK.

From the text above it can be concluded that Google Glass has issues and is in need of some refining before the device will become popular to the general public. It seems Google would need to add functionality other devices do not have, lower the price tag and fix the devices hardware limitations before the glasses would become interesting to the general public.

6 Conclusion

The questions stated by the objective in section 1.1 is taken into consideration and will try to be answered. To begin with the limitations with the Google Glass device is being looked at. As seen in the results the Google Glass device
had some limitations which were both hardware and software related. Many users experienced issues like the device becoming overheated after a some time of use. The battery time was very short. The device was always dependent on either Wi-Fi network or Bluetooth connection. The screen had poor contrast which made it hard to see clearly, especially in outdoor light. Other than that it had lag related issues which was experienced together with video- or voice streaming and the functionality offered by the device makes it hard to find a clear usage area and users would prefer wearing a wrist-worn device.

To continue the functionality a user is looking for in a wearable device is being looked at. And was there any improvements that can be made for Google Glass to make them as attractive to the general public as other wearable devices. It can be concluded that users mainly wants to use a wearable device to easily obtain information without having to reach out for their smart phone. Also while smart watches has no clear usage area besides providing the user information from their smart phone, much like the Google Glass, it was concluded that people would still prefer wearing a wrist worn device instead of having to wear a pair of glasses. Data from the survey shows that most people could see a clear usage area for activity trackers which is mainly used to help motivate the user to a more active lifestyle. People also do prefer to wear a more subtle device which both activity trackers and smart watches are compared to smart glasses, this also makes them more neat to wear in their everyday life.

It is challenging to conclude what kind of functionality could be added to Google Glass to make them more attractive but to start with it would need to resolve the limitations the users have experienced with the device. To lower the price tag would most likely benefit the popularity of Google Glass, which was seen from the survey that people were not willing to pay as much for the device as it was initially priced at. To make smart glasses as attractive to the general public as smart watches or activity trackers could still prove to be challenging as people do prefer to wear wrist-worn devices. Perhaps smart glasses should not have its usage area within the general public but rather be used within work-related fields which it has already been experimented with. It may become very popular within fields such as telemedicine where the biggest issue was the hardware limitations which can always be resolved.
References


507
Design of brain-computer interface to ease everyday life for people with severe motor impairments: A metanalysis

Ellinor Rönn
Department of Applied Physics and Electronics
Umeå University, Sweden
elra0017@student.umu.se

Abstract. For people who are fully motor impaired, brain-computer interface (BCI) may provide the only means of communication. However, there is a gap between the BCI performance and the users’ desires for a meaningful communication. The goal of this study was to develop guidelines of how to design user interfaces of BCI applications that meet the users’ desires. This metanalysis evaluated scientific articles with focus on users’ desires and present design solutions in order to identify design solutions that meet the users’ desires. The results show that the users want a spelling speed of at least 15-19 characters per minute (cpm) and an accuracy of 90 %. The accepted training time to be able to use the BCI independently is approximately five sessions. Many of the studied designs achieved the desired training time and accuracy (max = 100 %). However, only a few reached the desired speed (max = 20.6 cpm). Because of the trade-off between accuracy and speed, none of the designs reached the desired performance in both accuracy and speed. Both P300 and SSVEP was found to be suitable with a performance close to the users’ desires. The spellers, at least the P300 based spellers, can be improved with predictive text entry to achieve better efficiency and user satisfaction. The trade-off between speed and accuracy should be carefully considered when designing BCI spellers. Furthermore, to design good BCIs for users with severe motor impairments it is also important to consider differences in e.g. vision, cognitive functions in working memory, etc.

1 Introduction

Individuals with severe motor impairments often lack the means of communications, such as speech, gestures and voluntary facial expressions. Happ studied in 2000 how voicelessness affects both the patient and their family members and loved ones [1]. In this study Happ showed that the inability to speak made the patients terrified and frustrated, while the patients’ family members frequently expressed regret and anxiety over the voicelessness. Happ further showed that even for family members it was nearly impossible to interpret the voiceless patient’s feelings, experiences and desires. This indicates the difficulties for the
patients to express their opinions and decisions regarding clinical treatment or their living and biological wills.

Brain-computer interface (BCI) has been described as a possible tool to enable these patients to communicate and control their environment [2,3]. BCI is a technology that enables humans to control devices through their mind, independent of muscle movements [4]. A BCI translates the users brain activity into commands, usually by measuring the user’s electroencephalogram (EEG) acquired from the scalp [5]. There has been a lot of work in research and development of BCI, both in health care and entertainment. For example, there exists BCI technology that enables people with motor impairments to spell messages [6–9], operate a cursor on a screen [7] and control environmental features [10].

1.1 Objective

This study examines previous papers and existing BCIs from a usability aspect in order to find “good” designs from the user’s point of view. In this study, a “good” design is measured by Rubin and Chisnell’s [11] description of usability. That is usefulness, efficiency, effectiveness, learnability and satisfaction.

The goal of this metanalysis is to use these attributes to develop guidelines for how to design user interfaces of BCI applications to ease everyday life for people with severe motor disabilities. Since it has been shown that communicating is one of the most important things locked-in patients want from a BCI system [3], this study focuses on this specific task. Questions to answer are:

– What do BCI users with motor impairments want in a BCI system?
– How should BCIs be designed to meet the users’ desires?

To answer these questions, previous studies will be analysed in order to identify which properties motor impaired people consider most important in a BCI system. Furthermore, to analyse and compare present design solutions, their advantages and disadvantages.

2 Theory

Brain-computer interface (BCI) is a technology that enables humans to control devices through their mind. In 1999, in an international meeting between different research groups, BCI was defined as “a communication system that does not depend on the brain’s normal output pathways of peripheral nerves and muscles” [4]. This is usually accomplished by measuring the user’s electroencephalogram (EEG) acquired from the scalp, and translate the response into commands [5]. There are several commercially available BCI devices with a variety of sensor interpretations, for example the Emotiv EPOC headset and Neurosky MindWave headset, which both have shown to be suitable for use in BCI research [12].
2.1 BCI stimuli

BCIs use different forms of stimuli for communication. Three regularly used forms are P300, steady-state visually evoked potentials (SSVEP), and the imagined movement paradigm (also known as motor imagery, event-related resynchronisation or event-related synchronisation) [13].

**P300** is an event-related potential, manifested as a positive response in the EEG around 300 ms post stimulus [8, 9, 14]. The user focuses their attention on a rare stimulus among many irrelevant stimuli [9]. An example is a matrix of characters where the rows and columns flash in a random manner, and the user focuses on the character they wish to select [8, 9, 13, 15].

**SSVEP** is a periodic response evoked by a visual stimulus that flashes at a constant rate, evoking a response in the user’s EEG that matches the frequency of the flash rate [13, 14]. By using multiple stimuli flashing in different rates, a number of choices can be made [13, 14]. An example of SSVEP is to use four light emitting diodes as stimuli for navigating up, down, left and right. SSVEP based BCI spellers usually use a decision-tree method or moving the cursor position to select the target [14].

**Imagined movement** is controlled by the user imagining moving a certain body part, such as a hand or foot. This will induce sensory motor rhythms over the sensory-motor cortex that can be detected and mapped onto different commands [14, 16].

2.2 BCI for motor disabled people

Lately there has been a lot of research in the field of BCI and how it can be used to help people with motor disabilities in activities of everyday living. For example through spelling programs, prosthesis and robotic arms that can be controlled by the use of BCI. People can suffer from lack of motor function and severe levels of paralysis for many reasons, e.g. amyotrophic lateral sclerosis (ALS), brain injury and stroke. In the late stage of the neuromuscular disease ALS, the patients have no possible way to communicate due to loss of speech and motor functions. BCI has been described as a possible tool for ALS patients, to enable them to control and communicate with their surrounding [2, 3]. In 2011 Huggins, Wren and Gruis [3] showed that being able to communicate with other people is one of the most important things ALS-patients want in a BCI-system.

It has been shown that for people who are partially motor impaired, it is in many cases more effective with devices that detect voluntary motions such as head movements, eye gazing, or eye blink, than using BCIs [6]. In another study, however, it was reported that painting with BCI was less exhausting than using eye tracking [17]. For fully motor impaired people however, BCI may offer the only means of communication [6, 8, 18, 19].
2.3 Character by character versus predictive text entry

In a predictive text entry (PTE) system the user is provided with examples after writing the first letter in a word. For example, after spelling "y" the alternative "your" can appear with the number "1" in a list of possible words, and the user can spell the word "your" by focusing on "1" instead of spelling the whole word character by character [9]. Since a predictive speller may enable the user to write more text with fewer selections it could enhance communication for people using a P300 BCI. However, in a study by Ryan et al. [8] it is suggested that predictive spellers in non-BCI context increases cognitive demand, which can lead to a lower accuracy.

3 Methods

Since there were no BCI devices available for this study, it was conducted by analysing previous articles. When selecting articles, search terms used were: "brain-computer interface(s)" or "BCI" in combination with "motor impairments", "motor disabilities", "design(s)", "speller(s)", "user-centred (design)" or "usability". The reference list of the retrieved articles were further looked into to identify additional relevant articles. From the search result, all relevant articles that had been peer-reviewed was selected. However, due to the time limitation of the study, it was not possible to analyse all of the articles. The articles were analysed in random order until the time limit was reached. One selected article was excluded from the study due to numerical errors in their results.

The information extracted from the articles was training time, output characters per minute (OCM), accuracy, information about the test subjects, user evaluation and any other relevant information. Some of the articles did not provide the result of OCM, which was needed in order to compare the users’ desires to the performance of the studied BCIs. When no OCM was provided, it was estimated by the time it takes to enter one character. If this was not given, it was estimated by the given formula used to calculate the information transfer rate, which depends on the typing speed. With a given time for each selection and an accuracy above 50%, the correctly entered characters per minute was estimated with the formula

\[ OCM = \frac{60(2P - 1)}{T} \]

where \( P \) is the accuracy and \( T \) is the time needed to spell one character. For \( P \leq 0.5 \), \( OCM = 0 \). This is based on the formulas Park et al. [21], and Yin et al. [25] used in their studies.

3.1 Metrics

The different designs was measured by Rubin and Chisnell’s [11] description of usability:
Usefulness - The degree to which a product enables a user to achieve their goals and the user’s willingness to use it at all. This was measured by self-reported opinions in the design evaluation.

Efficiency - The quickness with which the user can accomplish their goals accurately and completely. This was measured by the selection rate, i.e. the number of commands/characters per minute (cpm) when writing a message. Several BCIs used predictive text entry, which means that one selection results in more than one character. Therefore the output characters per minute (OCM), was also used. OCM refers to total number of characters spelled, i.e. “your” is interpreted as four characters in both character-by-character and PTE, in contrast to the selection rate that interpreted “your” as two characters when selecting “y” followed by the predicted word “your”.

Effectiveness - The extent to which the product behaves the way the user expects it to, and the ease with which the user can use it to do what they intend. This was measured by the the accuracy, that is the number of successfully completed commands divided by the total number of commands.

Learnability - The user’s ability to operate the system to a defined level of competence after a predetermined amount of training time. This was measured by the training time needed to be able to operate the BCI independently.

Satisfaction - The user’s perceived comfort and acceptability while using the product. This was measured by evaluating the users’ self-reported satisfaction.

To evaluate which of these properties is most important from a user’s perspective, ratings from earlier studies were collected and compiled. Usefulness however, was not further investigated when evaluating the users’ desires since it is assumed that the users want a useful system. Other important aspects mentioned in the articles were also investigated.

4 Results

The study identified 2 articles regarding what BCI users with motor impairments would want in a BCI. Furthermore, it identified 11 articles that presents one or more design solutions with data from user studies.

4.1 What users want in a BCI

Huggins et al. 2011 Huggins et al. [3] conducted a survey to evaluate what ALS patients would want in a BCI. 63 participants with ALS responded to a telephone survey, with 61 participants answering BCI questions. Huggins et al. showed that the highest rated BCI design feature is ‘Accuracy of BCI operation’, followed by ‘simplicity of BCI setup’, ‘Functions the BCI provides’ and ‘reliability of BCI standby mode’. Appearance was the least important feature.

The result in their study shows that the users want a typing speed at 15-19 letters per minute (see Table 1). Furthermore they want an accuracy at minimum
That is the minimum percentage of the time the BCI has to recognize your commands correctly. Acceptable training time was 2-5 sessions. Another important finding is that the common occurrence of visual and auditory impairments should be considered during BCI design.

**Huggins et al. 2015** Huggins et al. [22] conducted a survey to establish what BCI users with spinal cord injury would want in a BCI. 40 subjects participated in the study. The results showed that the patients want a spelling speed at 20-24 letters per minute, and an accuracy at minimum 90%. Furthermore, the patients would accept a training time at 6-10 sessions to be able to use the BCI independently.

<table>
<thead>
<tr>
<th>Article</th>
<th>Subjects</th>
<th>OCM</th>
<th>Accuracy (%)</th>
<th>Training sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huggins et al. 2011</td>
<td>63</td>
<td>15-19</td>
<td>90</td>
<td>2-5</td>
</tr>
<tr>
<td>Huggins et al. 2015</td>
<td>40</td>
<td>20-24</td>
<td>90</td>
<td>6-10</td>
</tr>
</tbody>
</table>

### 4.2 Present design solutions

**Kaufmann et al. 2012** Kaufmann et al. [9] compared character-by-character paradigm versus predictive text entry (PTE). 19 healthy students, naive to BCI, participated in the study. The participants were instructed to spell German sentences with a P300 based BCI with a $6 \times 6$ matrix that flashed row/column wise. Selecting a predictive word was performed, like selecting a character, by counting each highlighting of the desired word to spell. The participants also reported a forced choice questionnaire.

Their results (see Table 2) showed that the output characters per minute (OCM) was significantly higher in the PTE condition ($M = 20.6$ characters per minute, cpm) than the character-by-character condition ($M = 12.0$ cpm). In the character-by-character condition the performance ranged from 100% accuracy to 76%, with an average of 91.2%. This was not significantly higher than the PTE condition that ranged from 100% to 74%. All participants managed to handle the system on their own, without expert support, from the moment the electrodes were mounted. The questionnaire revealed that, on a scale on 1-4, the participants estimated in average 3.84 that they would be able to manage the BCI application on their own in the future and 3.74 that they could explain it to others. The questionnaires also revealed that all participants would prefer to use PTE instead of spelling character by character, when exposed to BCI-based spelling in the future. One participant reported higher effort in selecting complete words compared to single character.
Table 2. Results of Kaufmann et al. 2012 [9]

<table>
<thead>
<tr>
<th>Condition</th>
<th>OCM</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTE</td>
<td>20.6</td>
<td>76-100</td>
</tr>
<tr>
<td>CBC</td>
<td>12.0</td>
<td>74-100 (M = 91.2)</td>
</tr>
</tbody>
</table>

Ryan et al. 2010 Ryan et al. [8] conducted a study where 24 able-bodied adults were instructed to accurately copy a sentence with a P300 based BCI. All the participants were naive to BCI use and they had no uncorrected visual impairments or any known cognitive impairments. The participants completed two sessions (one session in PTE condition and one character-by-character condition) on separate days within the period of one week. Both sessions started with a calibration and there was no stated training for the participants before the sessions. In both conditions the keyboard consisted of an 8 x 9 matrix, with the characters A-Z, 0-9, and other keyboard commands.

The results (see Table 3) did not show any significant difference in selections per minute between the character-by-character condition (M = 3.76) and the PTE condition (M = 3.71). However, the PTE condition reached an average OCM of 5.28, which is significantly higher than the 3.76 selections/minute in the character-by-character condition. The accuracy was higher in the character-by-character condition (M = 89.80%) than the PTE condition (M = 84.88%).

Table 3. Results of Ryan et al. 2010 [8]

<table>
<thead>
<tr>
<th>Condition</th>
<th>OCM</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTE</td>
<td>5.28</td>
<td>84.88</td>
</tr>
<tr>
<td>CBC</td>
<td>3.76</td>
<td>89.80</td>
</tr>
</tbody>
</table>

Pan et al. 2013 Pan et al. [23] conducted an experiment to compare two different P300 based BCI spellers. The first interface used a 6 x 6 matrix of characters whereas the second had a region based interface, with six circles, each containing six characters/actions, arranged in a circle. 12 healthy subjects participated in this experiment. Ten of the participants were naive to BCI, the other two had limited prior experience from the system development. There was no stated training for the participants before the sessions. The results (see Table 4) shows that both conditions reached the same average speed, however, the region based interface reached a higher accuracy. Region based condition showed a higher accuracy (M = 93.49%) and OCM (M = 2.42 cpm) than single character (M = 89.32%, 2.42 cpm). The region based speller showed a greater increase of accuracy for each trial.
Ahi et al. 2011 Ahi et al. [24] evaluated two different P300 based PTE spellers, both with a $5 \times 6$ matrix of characters. The difference between them was the order of the characters. One had the letters arranged in alphabetic order whereas the other used a modified order based on the words in the dictionary. 14 able-bodied subjects, with little or no previous experience of BCI, participated in their study. The subjects spelled 15 four-letter words, once in each condition. The five first in each test were training words. The results (see Table 5) shows that the accuracy is increasing with the number of trials for each condition. With the conventional order and typing character-by-character, 5 trials per character was also best regarding OCM (M = 2.85). The highest OCM was achieved with 2 trials per character for both conventional order with PTE (M = 12.96) and Modified order with PTE (M = 19.70).

Table 5. Results of Ahi et al. 2011 [24]. The results from 3 trials was was excluded since they did not count the time between two consecutive selections

<table>
<thead>
<tr>
<th>Number of trials</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional without PTE</td>
<td>Accuracy (%)</td>
<td>8.57</td>
<td>29.29</td>
<td>62.86</td>
</tr>
<tr>
<td></td>
<td>OCM</td>
<td>0</td>
<td>0</td>
<td>2.00</td>
</tr>
<tr>
<td>Conventional with PTE</td>
<td>Accuracy (%)</td>
<td>89.29</td>
<td>67.14</td>
<td>95.00</td>
</tr>
<tr>
<td></td>
<td>OCM</td>
<td>0</td>
<td>12.96</td>
<td>11.40</td>
</tr>
<tr>
<td>Modified with PTE</td>
<td>Accuracy (%)</td>
<td>53.57</td>
<td>87.14</td>
<td>95.00</td>
</tr>
<tr>
<td></td>
<td>OCM</td>
<td>4.92</td>
<td>19.70</td>
<td>11.40</td>
</tr>
</tbody>
</table>

Park et al. 2012 Park et al. [21] compared a P300 speller with the frequently used random flashing pattern, to P300 spellers with flashing pattern decided from a partially observable Markov decision process (POMDP). Ten able-bodied students participated in the study. It was not clear whether they had any previous experience with BCIs. The participants did not get any more training than the calibration phase, where 20 trials randomly assigned a target letter. In their study, the subjects used a $6 \times 6$ matrix to spell “MACHINE LEARNING”. Their result (see Table 6) shows that both average accuracy and spelling speed was highest for the improved POMDP (96.88%, 4.74 cpm), followed by POMDP (92.50%, 4.31 cpm), and last random flashing pattern (85.63%, 3.18 cpm).
Table 6. Results of Park et al. 2012 [21]
<table>
<thead>
<tr>
<th>Flashing pattern</th>
<th>OCM</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>3.18</td>
<td>85.63</td>
</tr>
<tr>
<td>POMDP</td>
<td>4.31</td>
<td>92.50</td>
</tr>
<tr>
<td>Improved POMDP</td>
<td>4.74</td>
<td>96.88</td>
</tr>
</tbody>
</table>

Yin et al. 2015 Yin et al. [25] designed a SSVEP based speller with a 6 × 6 matrix of characters. Their goal was to enhance the performance of the speller, compared to previous SSVEP based BCIs, in terms of the number of selectable items on the interface, accuracy and speed. They used two different interfaces, one with fixed approach (the speller provides result once the optimal stimulus time is met) and one using a dynamic approach (the stimuli stop and give spelling result once the SSVEP response reaches a threshold). 11 healthy subjects participated in their study. The participant did not get more training than the offline session (approximately 50 min), where the proposed approached were validated. Four of the participants had previous experience of SSVEP based BCIs. The spelling speed using dynamic optimization (M = 16.98 cpm) was significantly faster than the fixed optimization (M = 12.27 cpm).

Table 7. Results of Chen et al. 2014 [14]
<table>
<thead>
<tr>
<th>Stimulus (s)</th>
<th>OCM</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>17.94</td>
<td>84.1</td>
</tr>
<tr>
<td>3</td>
<td>14.67</td>
<td>90.2</td>
</tr>
</tbody>
</table>

D’Albis et al. 2012 D’Albis et al. [16] suggested a user interface for BCI spelling based on imagined movement with PTE. Three able-bodied subjects, naive to BCI, participated in the study. They did not get any more training than the offline session, that was used to train the algorithm. Two of the subjects used three classes (left hand, right hand, and both feet motor imagery), whereas one
subject had an additional class, namely both hands. All of the subjects used both three and four classes in the offline session. In the online test, the participants were instructed to perform seven repetitions of the phrase “what a wonderful day”. The result (see Table 8) shows that the subject using 4 classes performed better in both accuracy (M = 80%) and speed (M = 3 cpm).

Table 8. Results of D’Albis et al. 2012 [16]

<table>
<thead>
<tr>
<th># classes</th>
<th>OCM</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>3</td>
<td>80.20</td>
</tr>
<tr>
<td>3</td>
<td>2.35</td>
<td>70.15</td>
</tr>
</tbody>
</table>

Perdikis et al. 2014 Perdikis et al. [27] conducted an evaluation of a motor imagery hybrid speller, using both EEG and electromyographic signals. The authors compared three different interfaces, using either context awareness (CA), hybrid based ‘undo’ error handling, or both. In the study 16 subjects participated, of which six were severely disabled and ten were able-bodied. With all functions enabled, there was a completion rate of 100% for both user groups. Across all conditions and subjects the average accuracy was 94.2% (98.2% for the severely disabled test subjects). The results of the comparison (See Table 9) show that the CA condition reached the highest average speed of 1.76 cpm, but the lowest accuracy of 81.8%. CA + hybrid reached the highest average accuracy of 100%. 24 severely disabled subjects participated in the training phase, where approximately 50% reached 70% accuracy within nine sessions. None of the six motor disabled subjects had undergone more than five training sessions.

Table 9. Results of Perdikis et al. 2013 [27]

<table>
<thead>
<tr>
<th>Condition</th>
<th>OCM</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA</td>
<td>1.76</td>
<td>81.8</td>
</tr>
<tr>
<td>Hybrid</td>
<td>1.34</td>
<td>96.90</td>
</tr>
<tr>
<td>Hybrid + CA</td>
<td>1.68</td>
<td>100</td>
</tr>
</tbody>
</table>

Xu et al. 2014 Xu et al. [28] developed a parallel-BCI speller of four independent P300+SSVEP-B (P300 plus SSVEP blocking) with different flashing frequencies. Each sub-speller had a 3 × 3 matrix of characters. The interface had no support for error handling. 11 healthy subjects participated in the study. One subject had previous experience of a P300+SSVEP-B-based BCI system, the others were naive to BCI. The only training the subjects got before the test was in the offline session, where they were required to pay attention to a specified character and count silently how many times it was extinguished. The subjects
spelled a short, predefined message, once with five trials for each character selection and once with only three trials for each selection. The results (see Table 10) show, like Ahi et al. [24], that more trials provides a higher accuracy whereas the OCM was higher with three trials (M = 11.45) compared to five trials (M = 8.80).

<table>
<thead>
<tr>
<th># trials</th>
<th>OCM</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>11.45</td>
<td>87.8</td>
</tr>
<tr>
<td>5</td>
<td>8.80</td>
<td>93.9</td>
</tr>
</tbody>
</table>

**Table 10. Results of Xu et al. 2014 [28]**

Yin et al. 2013 Yin et al. [29] also evaluated hybrid BCI spellers based on P300 and SSVEP. 12 healthy subjects participated in the study, where they were required to input 14 symbols in random order. The results (see Table 11) show that the hybrid speller performed much better in both Accuracy (M = 93.85%) and OCM (M = 10.95 cpm), than the P300 (68.65% and 4.73 cpm respectively.

<table>
<thead>
<tr>
<th>Condition</th>
<th>OCM</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P300</td>
<td>4.73</td>
<td>68.65</td>
</tr>
<tr>
<td>Hybrid</td>
<td>10.95</td>
<td>93.85</td>
</tr>
</tbody>
</table>

**Table 11. Results of Yin et al. 2013 [29]**

5 Discussion

The results show that the users with motor impairments see accuracy as the most important feature in a BCI, it should translate at least 90% of the commands correctly. Furthermore they want the BCI to provide a spelling speed of at least 15-19 characters per minute (cpm), however, a large group of people want a spelling speed of 20-24 cpm. Many of the solutions meet the users’ desires in accuracy, whereas only a few of the studied BCIs meet the users’ desires when it comes to efficiency.

The design solutions that best meet the users’ desires were P300 using Predictive text entry (PTE) by Kaufmann et al. [9], and Ahi et al. [24] using modified order and 2 trials per character. Yin et al. [25] developed a speller using dynamic optimization that reached the users’ desires in speed, however the average accuracy was not stated in this study, which makes it difficult to say whether this is a good approach. The SSVEP speller by Chen et al. [14] was also relatively
close to the users’ desires, however, when using 2 s stimuli the OCM reached the
users’ desires whereas the accuracy was below the accepted limit. With the 3
s stimuli the accuracy reached the accepted limit, however, this time extension
forced the OCM below the user’s desires.

This trade-off between accuracy and speed is common when designing BCIs.
However, in the study of Park et al. it was shown that the choice of flashing
pattern can improve both accuracy and speed, which suggests that the flashing
pattern is an important aspect to consider in order to achieve both high speed
and high accuracy. This is probably because the higher accuracy is achieved
without changing the required time to make an action. This leads to fewer errors
and thereby a higher speed since the user does not have to delete the mistake
and type the correct character.

The time windows for stimuli causes a trade-off between accuracy and speed.
The longer stimuli time the higher accuracy, at the cost of speed. The number of
trials per selection causes a similar trade-off between accuracy and speed. More
trials per selection provides a higher accuracy at the cost of speed. Since spelling
speed showed the largest gap between the users’ desires and the performance of
the BCIs, it might be a good idea to choose the shortest time window and lowest
number of trials that provides an acceptable accuracy.

5.1 P300, SSVEP or imagined movement

Two out of five proposed P300 spellers could provide performance close to the
users’ desires. This was only achieved with PTE, however, one of the P300 based
spellers with PTE did not show satisfactory results in accuracy or OCM. The
subjects were able to use the P300 spellers independently immediately or after
one short training session. This suggests that the learnability matches the users’
desires to be able to use system independently within 5 training sessions.

Two out of two studied articles using SSVEP based spellers could provide
performance close to the users’ wishes. One study used a dynamic optimization,
which could increase or decrease the spelling speed dynamically. This approach
resulted in a higher OCM than the fixed optimization, however no specific num-
bers for the accuracy was given. This makes it difficult to analyse the results in
a good way. I feel that this could be a good approach to achieve a higher OCM
if the algorithm to decide how to adjust the spelling speed is well implemented.
The trade-off between speed and accuracy should be well considered to provide
both satisfactory accuracy and speed. The studied SSVEP based speller with the
highest OCM (M = 17.94) used a 5 × 9 matrix and 2 s stimulus time. However,
this resulted in an accuracy below the desired limit.

Both spellers based on imagined movement were far from reaching the users’
desires regarding spelling speed, with the highest speed of 3 cpm. The designs
with the highest OCM, did not reach the desired accuracy either. With a hybrid,
using both EEG and electromyographic signals, the accuracy (M = 96.90) was
higher than any of the other studies could achieve. Combining this with context
awareness, improved the accuracy to 100%. However, this only resulted in a
spelling speed of 1.34 cpm and 1.68 cpm respectively. In one study, only around
50% of the 24 severely motor disabled subjects that participated in the test session managed to reach a 70% accuracy after within nine training sessions. This means that 50% of the subjects did not reach a satisfying accuracy after the accepted training time. Based on these findings, BCI spellers using imagined movement can not meet the users’ desires.

5.2 Character by character versus predictive text entry

Three articles compared P300 based BCI spellers with predictive text entry (PTE), to character-by-character spellers. The results show that PTE provides a higher spelling speed in all cases. One study suggests that the PTE does not reach as high accuracy as the character-by-character speller, whereas another study suggests the opposite. In the study of Ryan et al. [8] in 2010, the accuracy for character-by-character was much lower than the best results from Ahi et al. 2011 [24]. However, the results from the study by Ryan et al. were not satisfactory for the character-by-character condition either. This could indicate that the low results was caused by some other factor, for example the large amount of possible actions. Kaufmann et al. [9] found no significant difference in accuracy between the two conditions.

Furthermore, the subjects in one study stated that they would prefer a speller with PTE if they would be exposed to BCI-based spelling in the future. Based on these findings I would argue for using PTE in at least P300 based BCI spellers. For SSVEP and imagined movements, however, it is not possible to draw the conclusion whether PTE or character-by-character is most suitable, from the results in this study. This is because none of the studied articles used SSVEP together with PTE, and the only study that used imagined movement with PTE did not compare the results to character-by-character. This means that the bad performance in the speller using imagined movement with PTE could as well be caused by some other parameter.

5.3 Limitations

The study did not differ between different causes of the impairment and focused on people with intact cognitive abilities. However, some deceases cause impairments even in the cognitive abilities. For example, ALS patients often show cognitive impairment in working memory, attention response inhibition, naming and other functions [13]. These differences need to be accounted for when designing BCIs for motor impaired people.

This study focuses on letter based alphabets and takes little consideration to how the BCI application would work to spell messages with a different kind of alphabet, such as the Chinese alphabet with a lot more characters. This means that the solutions found here is not necessary applicable for all languages.

Furthermore, the time limitations of this study caused a limitation on the number of studied articles. The results could be more accurate if a larger number of articles were studied for P300, SSVEP and imagined movement combined with PTE, dynamic optimization, region based versus matrix based, etc. Moreover,
the results could have been affected by the mixture of subjects, with different motor impairments or none at all. In one study, the subjects with motor impairments reached a higher average accuracy (98.2%) than the able-bodied subjects. However, only 50% of the motor disabled subjects could reach an accuracy of 70% within the given training time.

6 Conclusions

Both P300 and SSVEP was found to be suitable with a performance close to the users’ desires. Imagined movement however, did not perform good in effectiveness, efficiency or training time. Combined with electromyographic signals however, the accuracy was very satisfying. P300 based spellers could be improved with predictive text entry to achieve better efficiency and user satisfaction. A well chosen flashing pattern can also improve the P300 speller by increasing both accuracy and speed. When selecting number of trials and stimulus time, it is important to consider the trade-off between speed and accuracy in order to provide satisfactory results in both. Since spelling speed showed the largest gap between the users’ desires and the performance of the BCIs, the time window and number of trials could be set to the shortest time window and lowest number of trials that provides an acceptable accuracy. Dynamic optimization could provide good results in both accuracy and speed, by adjusting the time required to type one character, if this trade-off is carefully considered. Furthermore, when designing a BCI for motor impaired people, it is important to account for differences in vision, cognitive functions in working memory, etc.

References

Limitations and Possibilities of Using Eye-Tracking Technology with Smart Glasses

Anton Bäckström
Department of Applied Physics and Electronics
Umeå University, Sweden
anba0126@student.umu.se

Abstract. As wearables are expected to become the next generation of mobile devices, head-worn displays are becoming more affordable and available for commercial use. Although the emergence of Google Glass, the culmination of smart glasses as a successful device is yet to happen. Smart glasses, such as Google Glass, offer a lot of potential advantages in front other mobile devices, with ambition of providing a hands-free experience and always being accessible. In practice however, it is not as convenient. By being limited to the environment and the set of voice commands, furthermore the size and placement of the capacitive touch sensor, limits the range of possible inputs. Additionally, the placement can be physically demanding while doing regular tasks under an extended period of time. This paper examines the potential advantages, challenges and future possibilities of integrating eye-tracking technology with smart glasses, in order to achieve a more convenient, natural and hands-free experience, but also to consider the social aspects, based on a literature review with experts in this field of study, and a supervised user study to evaluate the social aspects of using such device. While potential advantages were found in eye-tracking, such as providing an increased bandwidth from the brain and the computer, and the human eye to be distinctly faster and naturally controlled, challenges were found in the difficulty of interpreting the user’s intentions, due to the eyes are always “on” and there are no natural way of making selections with them, which can easily lead to a gruelling experience of not being able to look anywhere without triggering another command. More sensors could be added in order to surpass the problem, by combining information from the eye and other lightweight inputs to help understand the user’s intentions and give the right response. Findings from the user study indicated that the conceptual solution of integrating eye-tracking with smart glasses, could be helpful and more socially accepted in way of being more discrete while interacting, but it does also introduce the temptation of using it discreetly, resulting in the feeling of being interfered while communicating with people who are wearing such device, knowing; they can always start and stop interacting with the device, without being fully able to notice it.
Analysis of text based information in virtual reality

Isak Dunér Lundberg
Department of Applied Physics and Electronics
Umeå University, Sweden
dv12ilr@cs.umu.se

Abstract. Virtual reality (VR) is a technology with a long history dating back to the 1960’s. It is a technology that let humans get immersed in virtual worlds and experience things that would not other wise be possible. Today the field of VR have gained traction with the release of VR consumer products such as the Oculus Rift and HTC Vive. This time around many of the previous limitations of older hardware have been addressed and improved for a better user experience. Regardless of the use case for the VR devices there will almost always be a need for relaying information to the end user. A solution for this problem is to integrate text elements into the virtual world, but one have to consider the aspects of the users feeling of immersion and comfort when developing such interfaces.

The goal of this paper was to perform an analysis of placement of text on a two dimensional plain in a VR user interface (UI). The paper considered problems regarding readability of the text and how the text was integrated within the virtual world as well as the UI. The research question posed was as follows: To identify differences in text display methods in the interface of virtual reality applications.

What is the important differences that can arise when looking at the UI from an perspective of user comfort and the ability for the user to understand the relayed information? Understanding the information is regarded as how easy the text is to read without too much effort from the user.

To limit the scope of the paper, the focus was solely on the positioning of the text. This paper did not consider researching different fonts, sizes or colors applied to the text. Nor evaluating differences with backgrounds and environments where the user could be placed. Which possibly could have some effect on the users ability to read the text presented to them.

To evaluate the research question an heuristic evaluation method was used and applied on three different types of text integration that is common in current applications. The types asses was fixed text, the text is fixed to the users view, which means no matter where the user looks the text will follow. Partially fixed text, text locked to the environment in 3D space. The text will therefor only obstruct a partial part of the world and can be moved out of view if desired. Finally integrated text in the world as in text that is a part of the world objects not just text projected on top of the screen.
The result showed only minor differences between the display methods where the most obstructive and potentially immersion breaking method was using fixed text. None of the methods had any major differences in readability and performed about the same. The conclusion was that what affected the readability the most was how close the text was to the user and when the text was too far away they blurred into small pixels. This suggests that the screen resolution even with the new hardware today, is of vital concern for the ability to read text. In other words the different display methods considered in this paper had no major effect on the readability.