



Proceedings of  
Umeå's 21<sup>st</sup> Student Conference in Computing Science  
USCCS 2017

S. Bensch, T. Hellström (editors)

UMINF 17.1  
ISSN-0348-0542

**Department of Computing Science**  
Umeå University



# Preface

The Umeå Student Conference in Computing Science (USCCS) is organized annually as part of a course given by the Computing Science department at Umeå University. The objective of the course is to give the students a practical introduction to independent research, scientific writing, and oral presentation.

A student who participates in the course first selects a topic and a research question that he or she is interested in. If the topic is accepted, the student outlines a paper and composes an annotated bibliography to give a survey of the research topic. The main work consists of conducting the actual research that answers the question asked, and convincingly and clearly reporting the results in a scientific paper. Another major part of the course is multiple internal peer review meetings in which groups of students read each others' papers and give feedback to the author. This process gives valuable training in both giving and receiving criticism in a constructive manner. Altogether, the students learn to formulate and develop their own ideas in a scientific manner, in a process involving internal peer reviewing of each other's work, and incremental development and refinement of a scientific paper.

Each scientific paper is submitted to USCCS through an on-line submission system, and receives a review written by members of the Computing Science department. Based on the review, the editors of the conference proceedings (the teachers of the course) issue a decision of preliminary acceptance of the paper to each author. If, after final revision, a paper is accepted, the student is given the opportunity to present the work at the conference. The review process and the conference format aims at mimicking realistic settings for publishing and participation at scientific conferences.

USCCS is the highlight of the course, and this year the conference received seventeen submissions (out of a possible eighteen), which were carefully reviewed by the reviewers listed on the following page.

We are very grateful to the reviewers who did an excellent job despite the very tight time frame and busy schedule. As a result of the reviewing process, all fifteen submissions were accepted for presentation at the conference. We would like to thank and congratulate all authors for their hard work and excellent final results that are presented during the conference.

We wish all participants of USCCS interesting exchange of ideas and stimulating discussions during the conference.

Umeå, 8 January 2017

Suna Bensch  
Thomas Hellström

## Organizing Committee

Suna Bensch  
Thomas Hellström

## With special thanks to the reviewers

Suna Bensch  
Johanna Björklund  
Jerry Eriksson  
Thomas Hellström  
Anna Jonsson  
Lili Jiang  
Kai-Florian Richter  
Ola Ringdahl



# Table of Contents

Gaining the E-consumers trust - The role of aesthetics .....	1
<i>Maja Andersson</i>	
Which actions force users to change their hand grip on their mobile devices? .....	19
<i>Tove Bäckström</i>	
Can mobile augmented reality enhance the user experience of internet shopping? .....	29
<i>Amanda Dahlin</i>	
Does haptic feedback effect users time to action in a form based interface? .....	45
<i>Simon Ekdahl</i>	
Haskell vs. JavaScript for Game Programming .....	55
<i>Christian Fischer</i>	
Usability of mobile interfaces with regards to left-handed use .....	65
<i>Filip Norman Golles</i>	
Replacing words with emojis and its effect on reading time .....	73
<i>Viktor Gustafsson</i>	
The effectiveness of searching for smartphone apps sorted by color .....	83
<i>Pontus Henstam</i>	
Third-Person Avatar Control in Virtual Reality: a Nausea Provocation Comparison .....	93
<i>Fredrik Johansson</i>	
Examining text quality in video encoded in different resolutions with OCR-software .....	101
<i>Simon Johansson</i>	
Evaluating efficiency of interactive notifications on mobile devices .....	111
<i>Joakim Ljunggren</i>	
Comparison Of Geometric Approach With Artificial Neural Network Approach For Inverse Kinematics Problem For AL5D Robotic Arm .....	125
<i>Vishnupradheep Raveendran</i>	
How Animations in a Mobile Web Application Impact User Interaction ..	139
<i>Klara Sporrang</i>	
Can highlighting screen borders be visual cues for swipe gestures? .....	159
<i>Mikael Wahlström</i>	

How correct is auto-correct when entering text on a smartphone? ..... 175  
    *Freja Wieslander*

**Author Index** ..... 183

# Gaining the E-consumers trust - The role of aesthetics

Maja Andersson

Department of Computing Science  
Umeå University, Sweden  
`maja.helena.andersson@gmail.com`

**Abstract.** Trust is a fundamental principle in order to form successful business relationships in the e-commerce context. A major part of the overall credibility judgement is based on the initial impression of a website. This study therefore examines the relationship between aesthetics and credibility in the e-commerce context. Two independent experiments with two groups of participants are conducted on 6 travel agency websites. One to assess aesthetics in four dimensions (simplicity, diversity, colorfulness, craftsmanship), and one to assess credibility. The material used in the study are an 18-item measure of perceived visual aesthetics and an 1-item measure of perceived trust. The findings of the study suggest that there is likely to exist a relationship between high level of aesthetics and high level of credibility and that colorfulness and craftsmanship are more prone to affect credibility than simplicity and diversity. The correlation between aesthetics and credibility however, can not be proven on a 95 percent confidence level, which means that the model includes too much uncertainty. To be able to construct a better model, a greater amount of stimuli would be necessary.

## 1 Introduction

The Internet provides fantastic opportunities for business to consumer relationships worldwide. Geography and time are no longer limiting since the Internet provides 24-hour stores where items can be ordered from all over the world. This means that a wider range of products and services are available at a more price competitive market than before. Due to the increasing competition vendors have to make sure their products and services are presented and accessible in a favorable manner to get potential consumers attention.

One important issue for vendors to be able to establish business relationships is gaining the trust of the consumers [1]. In the area of e-commerce, vendors can not depend on competent and reliable salesmen to build a relationship of trust with the consumer. Instead they depend on their reputation and their electronic storefront to give a reliable impression. In a Swedish study about online trust, Internet users acknowledged that their trust decision was based on an rather intuitive feeling of trustworthiness. During interviews participants stated that if a web site looked pleasant to them, i.e. was aesthetically appealing, they were

ready to trust it more easily [2]. A previous study also found that the initial impression of a company website affected the initial impression of the company itself [3].

The relationship between aesthetics and credibility online has previously been investigated in different contexts using recruitment agency websites [4] and information portals [5]. Both of these studies suggested that a connection between these two exists to some extent. For this study, the objective will therefore be to explore this relationship further in the e-commerce context. Since this environment involves money transfers it could affect the risk perception and make trust a more important factor than in previous studies concerning aesthetics. To examine the relationship, two independent experiments will be conducted to collect data from two groups of participants. The first experiment to assess the visual aesthetics of a set of e-commerce sites, and the second to assess the credibility of the same set of e-commerce sites. The assessment will be done using forms with statements about the site on which participants indicate their level of agreement on a scale. The results will be analyzed to determine the nature of correlation between aesthetics and credibility.

## 2 Background

The Swedish e-commerce industry has grown from 70 billion dollars to over 100 billion dollars in the last four years, presenting a 11 percent growth during 2016 alone<sup>1</sup>. Swedish e-consumers spent the largest amount of money in the travel industry and according to consumers' own estimations, 51 percent of their total e-commerce budget comprise travels. When it comes to choice of payment methods, 82 percent of the total e-commerce transactions during 2015 were made with debit/credit cards. Most consumers state they prefer this payment method due to convenience and safety. The consumers who prefer other type of payment methods, like invoices or services like PayPal, state only safety as their main reason. The fact that consumers feel the need for safety precautions indicates they associate online transactions with risk taking. Previous studies confirm this indication by stating that consumers perceive the risk as higher when purchasing products and services online than when purchasing in a physical store [6]. Their major concerns were identified as "security of online payments" and "privacy of personal information" [7]. Risk perception and trust are closely interrelated since trust describes the consumer's willingness to take a risk, and the level of trust is an indication of the amount of risk that the consumer is willing to take [8]. This means that in order to form successful business relationships and stimulate purchases, trust is a fundamental principle [1].

### 2.1 Trust online

Different factors affect the feeling of trust in an online environment. The perception of credibility has previously been classified in four categories; presumed

<sup>1</sup> <http://info.dibs.se/svenskehandel2016>, Swedish e-commerce 2016, DIBS annual report about Swedish e-commerce, accessed 2016-10-27

credibility, reputed credibility, earned credibility and surface credibility [9]. In the e-commerce context, presumed credibility could be described as a general assumption about a site based on e.g. the size and recognition of the brand without actual proof of credibility. Reputed credibility in contrast, is based on actual recommendations or proof of credibility from other customers. While reputed credibility is based on others experiences, earned credibility is based on the customer's own good experiences of shopping at a site. The last and the most interesting category for this study is surface credibility, since it is based on first impression and therefore concern aesthetics. Surface credibility is instant and perceived as fast as within 3.42 seconds [5]. A previous study which collected ratings and comments about different website's credibility found that 46.1 percent of the comments contained references to the appearance of the website [10]. This indicates that surface credibility plays a major role in the overall credibility perception of the site. To investigate the relationship further, we need to understand the human perception of websites in general and role of aesthetics in particular.

## **2.2 Website perception**

The human perception of websites has previously been described with three core constructs: content, usability, and aesthetics [11]. These factors interplay in different phases of website use and are all contributory factors when the consumer form an opinion of a website. An author in the field of human computer interaction (HCI) described the processing of an experience in three different levels [12]. The reflective level which is affected by individual differences, culture, previous experience and preferences. The behavioral level which is affected by the use and experience of something and the visceral level which is pre-consciousness and based on visual properties. In the e-commerce context, the reflective level could be described as the consumer's mental image of the e-commerce site or vendor. The behavioral level relates to the perception of the content and usability of the site, and the visceral level relates to the perception of aesthetics. Since the visceral level is pre-consciousness, people make reliable judgements whether they find a website appealing or not after only 50 ms [13]. Regardless of the quality of the content and the overall usability, an aesthetically unappealing website will likely produce a negative first impression with the user. The visceral experience could therefore be vital to ensure that user stay and engage the content at a cognitive level.

Compare this to the first encounter between two human beings, before any interaction like physical contact or conversation takes place, most people form an instant mental image of the other person. This mental image is based on several visually perceived factors like age, gender, body posture, clothing and the physical appearance of the other person. Even though this mental image is not proven to be valid it still influences how we feel about this other person and could potentially affect if and how we choose to interact.

To determine the constructs of a positive visceral online experience, a valid measure of web aesthetics needs to be defined.

### 2.3 Defining aesthetics

Aesthetics is often philosophically described a function of the properties of an object and characteristics of the perceiver. That is, an object consist of physical properties that influence a perceiver's experience, but this experience is also affected by properties at the perceiver's end. In the area of HCI, various studies to identify characteristics that influence perceived visual aesthetics have been conducted, where one managed to divide previous findings into 4 facets [14]. The facets jointly represent perceived visual aesthetics but can be distinguished from each other by different aspects;

**Simplicity** reflects the processing of the layout, such as clarity, orderliness, homogeneity, grouping, and balance.

**Diversity** reflects visual richness, dynamics, variety, creativity, and novelty.

**Colorfulness** reflects the use of individual colors and their composition.

**Craftsmanship** reflects the skillfulness in design and use of modern technologies.

To assess aesthetics in the area of HCI, other studies commonly use single-item measures of perceived aesthetics. However, the use of single-item measures assume that the measured construct is one-dimensional when most studies, like the previously mentioned, suggest that aesthetics is multidimensional. The previously mentioned study therefore validated an 18-item measure of perceived aesthetics which is called the VisAWI [14], and a subsequent shortened 4-item version [15]. In the VisAWI, one facet comprises 4 or 5 items where each item consists of a statement regarding different aspects of the design features of the stimuli on which the user indicates her level of agreement on scale.

To measure aesthetics in this study, the items in the VisAWI will be translated to Swedish and used according to the VisAWI manual<sup>2</sup> on different e-commerce sites.

### 2.4 Previous studies

Previous studies suggests that there exists a relationship between aesthetics and credibility online. The stimuli used in these studies were information portals [5] and recruitment agency websites [4].

In the study concerning information portals, 21 websites were used in two different versions. The first version was the original website and the other version was a website with the same content but "less aesthetic treatment" than the original. The subjects in the study, who were students, were told to assume they were taking a course in which they had been assigned to write a paper on a specific subject, and that the websites displayed where the results from a search engine. Subjects were shown each of the 42 images in sequence where they were asked to rate the websites presented to them on a scale from -7 (low credibility) to 7 (high credibility), based on their initial impression on the sites credibility. The

<sup>2</sup> <http://visawi.uid.com/>, VisAWI, accessed 2016-10-20

results showed that the original set of websites were given a rating of 1.05 while the "less aesthetic" set were given -0.55. Even though the value is statistically significant, it is still a small difference on a scale from -7 to 7. However, the result also showed that if a website's original version was compared to the "less aesthetic" version of the same site, the original site produce a higher credibility score in all cases but one. This suggests that the aesthetic treatment improved the perception of credibility. One limitation with this approach to investigate the relationship though is that what was considered as "less aesthetic treatment" was based on the researcher's own opinions. Another limitation is that it is impossible to tell from the data what constructs underlie these judgments since a single-item measure was used to assess aesthetics and credibility instead of a multi-dimensional.

In the study concerning recruitment agency, a similar approach as in the first study was used, that is, two versions of the same site with different level of "aesthetic treatment". This study though, tried to evaluate the different dimensions of aesthetics by altering only one dimension per website. The constructs used were balance, harmony, contrast and dominance. The participants were shown 13 pairs of stimuli, where only one construct per pair was to be evaluated, and told to chose the image they perceived as being more credible in each pair. 8 out of the 13 results were explicitly conclusive judgments on web credibility on the basis of aesthetics. However, harmony was inconclusive for 75 percent of the evaluated objects as well as 50 percent of informal balance and 33.3 percent of formal balance. This means that only contrast/dominance by contrast and dominance by size was conclusive in affecting the credibility judgement.

With this being said, this study will attend to the shortcomings of previous studies by not labeling stimuli as high or low aesthetics. Instead, aesthetics and credibility will be assessed separately through two different groups of participants where different dimensions of aesthetics will be addressed properly. The study will therefore continue the work of the latter study by investigating the relationship between the dimensions of aesthetics and credibility further and determine if certain constructs of aesthetics are more prone to affect credibility judgement than others.

### 3 Method

The intention of this study is to investigate the relationship between the dimensions of aesthetics and credibility further. To do this, two experiments were conducted to collect data from two different groups of participants. The first experiment were conducted to assess the visual aesthetics of a set of websites (Experiment 1), and the second to assess the credibility of the same websites (Experiment 2). SPSS statistics<sup>3</sup> was used to analyse the correlation of the two variables by comparing the mean ratings of aesthetics and credibility from each website.

---

<sup>3</sup> <http://www-03.ibm.com/software/products/sv/spss-statistics>, SPSS, accessed 2016-11-20

### 3.1 Stimuli

The stimuli was chosen within the same e-commerce sector so that a simple scenario could be constructed to get the subjects to make credibility judgments as grounded as possible. Knowing that the travel industry is the biggest sector in Swedish e-commerce, the flight industry seemed like an interesting sector.

To summarise the alternatives for consumers when looking for flights online there are airline websites, travel agency websites, or travel search engine websites. Travel agency websites gather information and sell tickets from many airlines while travel search engine websites only gathers information from travel agencies and airlines and redirect the user to a vendor. Since it is more convenient to use a travel search engine, many consumers choose to use these rather than visiting many websites to compare departures and prices. Examples of this type of websites are Skyscanner<sup>4</sup> and Momondo<sup>5</sup>. Consumers who visit Skyscanner and Momondo are somewhat familiar with these sites or names since they found their way to the websites. With this being said, they already have a relationship to the site on a reflective level based on presumptions, reputation or experience. When searching for a trip on these websites, the consumer will be presented with results supplied by different travel agencies or airlines. Upon choosing a departure, the consumer will be redirected to the vendors website. If the consumer had no prior relationship to this website however, it means their credibility judgement will be based solely on surface credibility when entering the site. Since investigating surface credibility is the main intention of this study, this real life occurrence seemed like a perfect scenario for the experiment. Therefore, 6 different travel agency websites were chosen as stimuli (See Figure 3.1). To be sure to further avoid biases based on presumptions and reputation, the websites were stripped from logos, names and "facebook likes" before screenshots were taken.

### 3.2 Participants

Experiment 1 included 20 participants, and Experiment 2 included 14 participants due to time constraints. The number of participants for Experiment 1 were chosen in accordance to the instructions in the VisAWI manual<sup>6</sup>. The two groups of participants are very similar, mainly consisting of young adults (median age 25 and 26) with a balanced representation of women and men as well as students and professionals. Both groups had a widespread behaviour of online shopping where every participant tried online shopping and around 90 percent identify as accustomed to online shopping. Around 95 percent had also bought airline tickets online before.

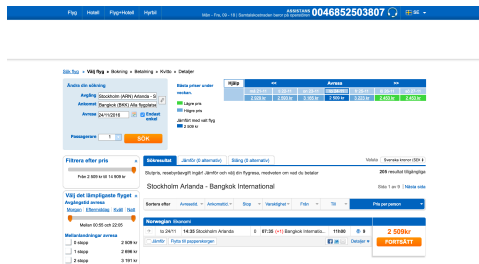
---

<sup>4</sup> <https://www.skyscanner.se/>, Skyscanner, accessed 2016-10-27

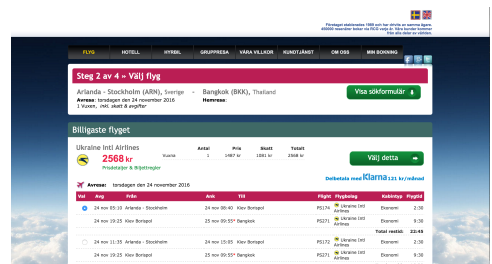
<sup>5</sup> <http://www.momondo.se/>, Momondo, accessed 2016-10-27

<sup>6</sup> <http://visawi.uid.com/>, VisAWI, accessed 2016-10-20

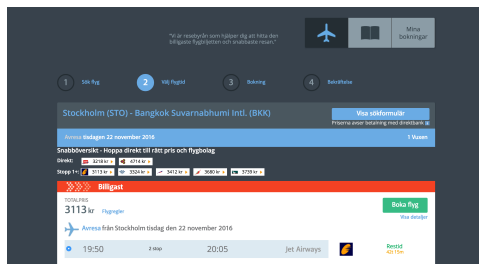




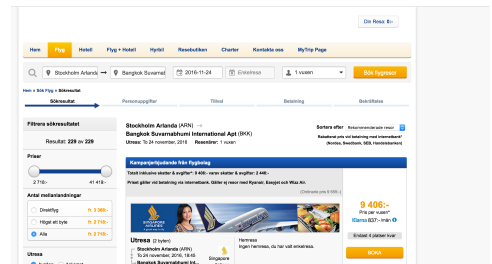
(a) stimuli 1



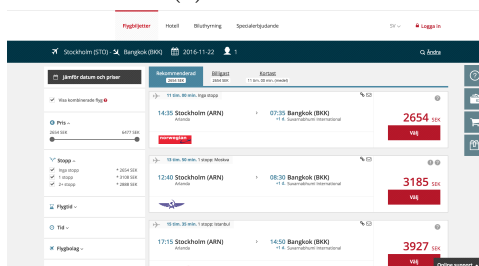
(b) stimuli 2



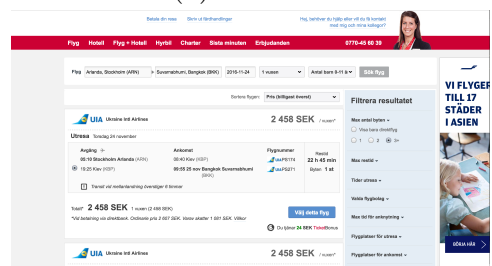
(c) stimuli 3



(d) stimuli 4



(e) stimuli 5



(f) stimuli 6

Figure 1: The 6 websites used in the study to evaluate aesthetics and credibility

### 3.3 Materials

The material used in experiment 1 was a translated version of the 18-item measure of perceived visual aesthetics, the VisAWI (See Table 1). For Experiment 2, an 1-item measure of trust was used (See Table 2). In addition to this, 10 standard questions (See Table 3) were used to collect demographic information and information about Internet shopping habits from both experiments.

### 3.4 Procedure

**Experiment 1** A testing procedure was carried out where the subjects were informed that their help was needed to evaluate the attractiveness of 6 travel agency websites. Demographic information and information about Internet shopping habits were firstly collected before the experiment started. After this, the participants were told that one image of a website at a time would be displayed to them. Based on their perception of each website, they were then told to answer a series of statements by indication their level of agreement to each statement on a seven-point Likert scale (ranging from 1 "strongly disagree" to 7 "strongly agree").

**Experiment 2** A testing procedure was carried out where the subjects were informed that their help was needed to evaluate the credibility of 6 travel agency sites. Demographic information and information about Internet shopping habits were collected before the experiment started. The participants were told to assume they were going for a vacation and looking for an airplane ticket on a search engine website. They were told that the websites that would be displayed to them were the result of a search, that is, different travel agency websites offering tickets. One website at a time was displayed to them, and based on their initial impression, they were told to indicate how comfortable they would feel about placing their trust in the website and making a purchase there. The participants were asked to indicate their answer on seven-point Likert scale (ranging from 1 "very uncomfortable" to 7 "very comfortable").

<b>The VisAWI</b>	
<b>Factor 1: Simplicity</b>	
1	The layout appears too dense. Webbsidan upplevs plottrig och informationst/"at.
2	The layout is easy to grasp. Webbsidans struktur /"ar l/"att att förstå sig på.
3	Everything goes together on this site. Webbsidans delar passar ihop.
4	The site appears patchy. Webbsidan upplevs osammanh/"angande.
5	The layout appears well structured. Webbsidan upplevs v/"alstrukturerad.
<b>Factor 2: Diversity</b>	
6	The layout is pleasantly varied. Webbsidan /"ar varierad på ett tilltalande vis.
7	The layout is inventive. Webbsidan /"ar kreativt utformad.
8	The design appears uninspired. Designen /"ar tråkig.
9	The layout appears dynamic. Webbsidans utformning upplevs som dynamisk (rörlig, för/"anderlig).
10	The design is uninteresting. Designen /"ar ointressant.
<b>Factor 3: Colorfulness</b>	
11	The color composition is attractive. Kompositionen av f/"arger på webbsidan /"ar attraktiv.
12	The colors do not match. F/"argerna passar inte ihop.
13	The choice of colors is botched. F/"argvalet /"ar dåligt utfört.
14	The colors are appealing. F/"argerna /"ar tilltalande.
<b>Factor 4: Craftsmanship</b>	
15	The layout appears professionally designed. Webbsidan upplevs professionellt designad.
16	The layout is not up-to-date. Webbsidan ser inte modern ut.
17	The site is designed with care. Webbsidan upplevs noggrant utformad.
18	The design of the site lacks a concept. Webbsidans design saknar ett genomgående koncept.

**Table 1.** The 18-item VisAWI of perceived visual aesthetics used in the study to evaluate aesthetics.

<b>Trust</b>	
1	How comfortable do you feel about placing your trust in this website and making a purchase here? Hur bekväm känner du dig med att förlita dig till den här webbsidan och genomföra ett köp här?

**Table 2.** 1-item measure of trust used in the study to evaluate credibility.

<b>Standard questions</b>	
<b>Demographics</b>	
1	Gender? Könsidentitet?
2	Age? Ålder?
3	Occupation? Yrke?
4	Current studies or educational background? Pågående studier eller studiebakgrund?
<b>Internet shopping habits</b>	
5	Have you tried Internet shopping before? Har du någon gång handlat varor eller tjänster på internet?
6	Are you used to shopping online? Är du van vid att handla på internet?
7	What type of product or services do you usually shop online? Vilken typ av varor eller tjänster handlar du oftast på internet?
8	What type of payment method do you prefer online? Vilket typ betalar du föredrar du vid ett internetköp?
9	Have you used the Internet to buy airplane tickets? Har du handlat flygbiljetter på internet någon gång?
10	Have you used websites which gathers information from different travel agencies and airlines, like skyscanner and momondo, to look for flight tickets? Har du använt webbsidor som samlar information från olika reseangörare och flygbolag, som skyscanner och momondo, för att leta flygbiljetter?

**Table 3.** Standard questions used in the study to collect demographic information and information about Internet shopping habits from the participants.

## 4 Result

### 4.1 Experiment 1

In experiment 1, evaluations of website aesthetics from 20 participants were collected. The mean values from the evaluation are displayed in Table 4 and visualised in Figure 1. The results show that Website 5 received the highest aesthetics rating (mean 4.9275) out of the 6 websites while Website 1 received the lowest aesthetics rating (mean 2.1231). Website 5 also received the highest rating for each facet while Website 1 received the lowest. Remaining webpage ratings varied through the different facets (See Figure 2).

### 4.2 Experiment 2

In Experiment 2, evaluations of website aesthetics from 14 participants were collected. The results show that Website 5 received the highest credibility rating (mean 4.9286) while Website 1 received the lowest credibility rating (mean 3.8571). The mean values from the evaluation are displayed in Table 5 and visualised in Figure 3.

### 4.3 Correlation

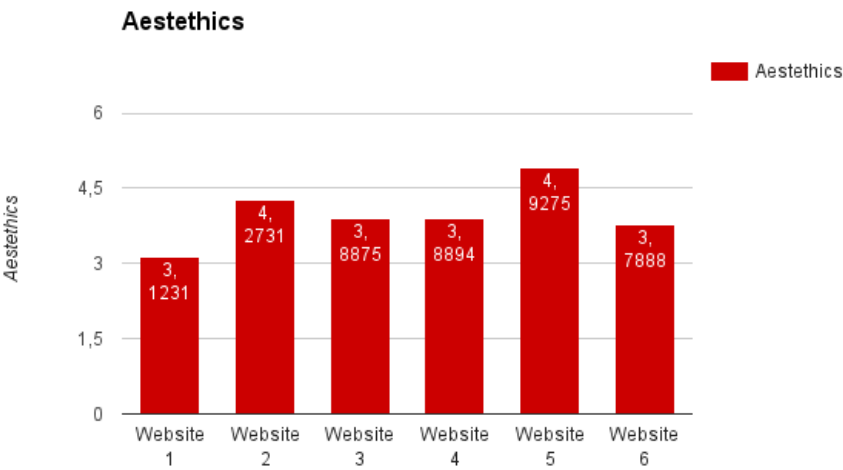
The result show some similarities between the evaluation of credibility and aesthetics since Website 5 and 1 received the highest respectively the lowest rating in both Experiment 1 and 2 (See Figure 4). To examine the relationship further a correlation analysis was performed with SPSS statistics. The null hypothesis in the test was that no correlation between the two variables existed,  $H_0 = 0$ , and the alternative hypothesis was that a correlation existed,  $H_1 \neq 0$ . The strongest correlation which can be received from the test is either -1 or 1 and the lowest is 0. A positive value close to 1 means that high values of one variable implies high values of the other, and a negative value close to -1 means that high values of one variable implies low values of the other. To be able to perform this analysis, the samples were assumed to follow a bivariate normal distribution [16].

The test result show a positive correlation of 0.785 which is a relatively high correlation (See Table 6). The significance level is above 0.05 (5 percent) though, which makes it impossible to reject the null hypothesis with a 95 percent confidence level.

Since aesthetics comprise simplicity, diversity, colorfulness and craftsmanship, a correlation analysis was also made between every facets of aesthetics and credibility since there might exist a statistical relationship between some facet and credibility. The significance level for both the simplicity-credibility test and the diversity-credibility test are above 0.05 (See Table 7 and Table 8), which makes the correlation statistically invalid on a 95 percent confidence level. However, the correlation between colorfulness-credibility and craftsmanship-credibility are high (0.830 and 0.851) and the significance level in both cases are below 0.05 (See Table 9 and Table 10), which makes the correlation statistically valid on a 95 percent confidence level.

Aesthetics	
Website 1	3.123
Website 2	4.273
Website 3	3.888
Website 4	3.889
Website 5	4.928
Website 6	3.789

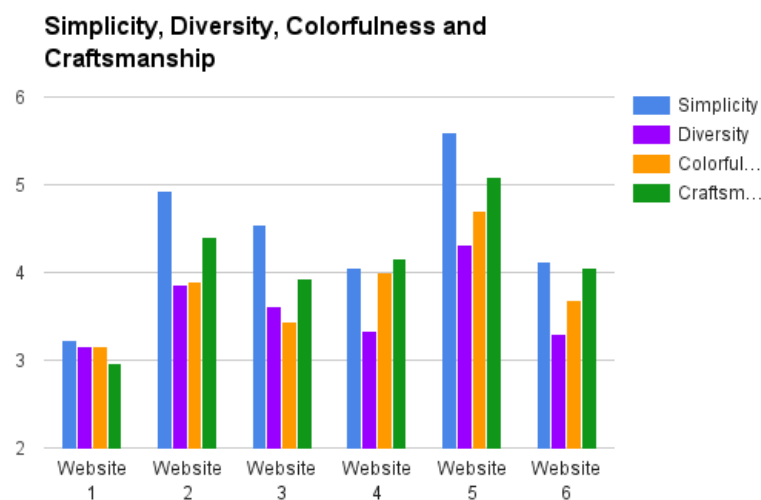
**Table 4.** The results displayed in the table are the mean ratings of aesthetics for each websites.



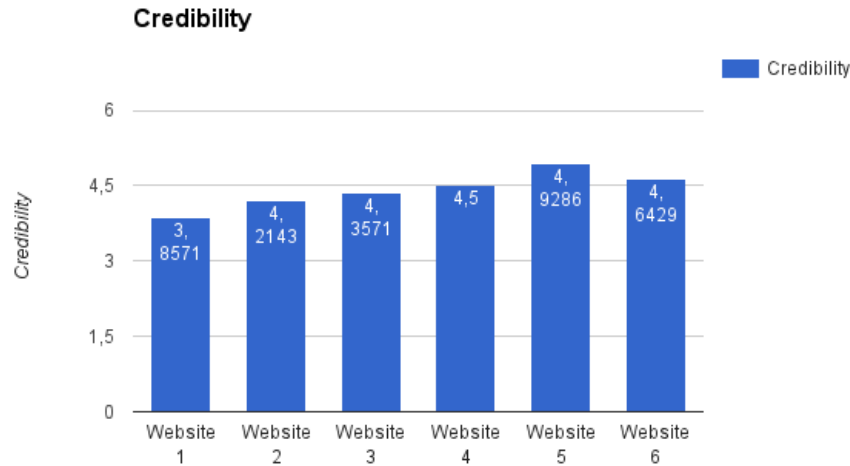
**Fig. 1.** The results displayed in the graph are the mean ratings of aesthetics for each website.

Credibility	
Website 1	3.857
Website 2	4.214
Website 3	4.357
Website 4	4.500
Website 5	4.929
Website 6	4.643

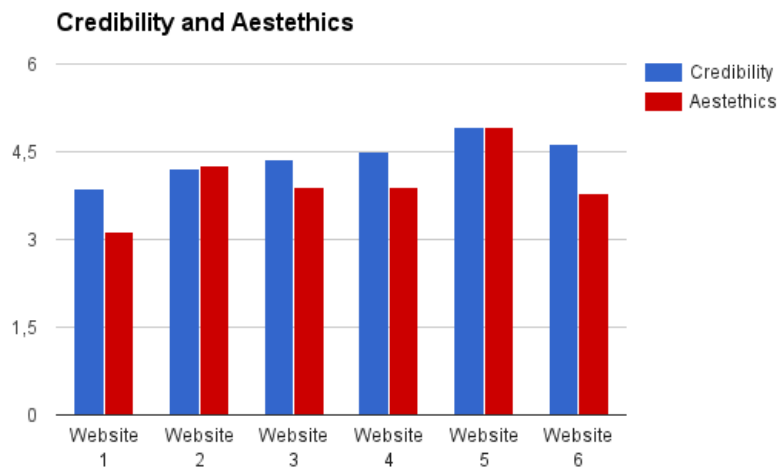
**Table 5.** The results displayed in the table are the mean ratings of credibility for each website.



**Fig. 2.** The results displayed in the graph are the mean ratings of each facets of aesthetics for each website.



**Fig. 3.** The results displayed in the graph are the mean ratings of credibility for each website.



**Fig. 4.** The results displayed in the graph are the mean ratings of credibility and aesthetics for each website.

Correlation		
		Credibility
Aesthetics	Correlation	0.785
	Significance level	0.064

**Table 6.** The results displayed in the table are from the correlation test between credibility and aesthetics and show relatively high correlation but a significance level above 5 percent which means that the model is uncertain.

Correlation		
		Credibility
Simplicity	Correlation	0.710
	Significance level	0.114

**Table 7.** The results displayed in the table are from the correlation test between credibility and simplicity and show relatively high correlation but a significance level way above 5 percent which means that the model is very uncertain.

Correlation		
		Credibility
Diversity	Correlation	0.591
	Significance level	0.216

**Table 8.** The results displayed in the table are from the correlation test between credibility and diversity and show medium correlation but a significance level way above 5 percent which means that the model is very uncertain.



Correlation		
		Credibility
Colorfulness	Correlation	0.830
	Significance level	0.041

**Table 9.** The results displayed in the table are from the correlation test between credibility and colorfulness and show a high correlation and a significance level below 5 percent which means that the model is valid.

Correlation		
		Credibility
Craftsmanship	Correlation	0.851
	Significance level	0.031

**Table 10.** The results displayed in the table are from the correlation test between credibility and craftsmanship and show a high correlation and a significance level below 5 percent which means that the model is valid.

## 5 Conclusion

Website 5 and Website 1 were the highest respectively lowest rated websites concerning both aesthetics and credibility. This means that the website that was considered the most aesthetically pleasing was also considered as more credible and the website that was considered the least aesthetically pleasing was also considered as less credible. This fact is interesting since it implies that the two variables are correlated to some extent. The correlation test gave a result of 0.785 which also suggest that a correlation exist. The correlation, however, could not be proven on a 95 percent confidence level since the significance level was too high. The problem is therefore that the model involves too much uncertainty to construct statistical proof of a correlation between aesthetics and credibility. The uncertainty of the model is connected to the low number of stimuli in the study and could only be solved if a greater amount of websites were included. This is a shortcoming of this study that should have been thought through before selecting the amount of stimuli to be used. A suggestion for further investigations would therefore be to select a high number of stimuli so that high uncertainties in the model can be avoided.

To investigate if any dimension of aesthetics were more prone to affect credibility judgements than the others, correlation tests between all four dimensions and credibility were performed. These results show that colorfulness and craftsmanship seem to be more correlated with credibility than simplicity and diversity. These results indicate that the use of individual colors and color composition as well as modern and skillful design seem to be the more important variables to achieve a higher degree of surface credibility.

The overall findings of the study suggest that there is likely to be a connection between high level of aesthetics and high level of credibility. It is important to emphasize that to be able to make a stronger statement about the relationship, a greater number of stimuli would have been necessary.

As a conclusion of the overall findings of the study, it could be advisable for e-commerce vendors to consider the fact that a positive visual experience could enhance consumers trust and stimulate purchases on their website. With this being said, it is important to know that other factor, such as desire, price and accessibility, also affect the willingness to buy.

## References

- [1] Corbitt, B., Thanasankit, T., Yi, H.: Trust and e-commerce: A study of consumers perceptions. *Electronic Commerce and Research Applications* **2**(3) (2003) 203–214
- [2] Karvonen, K., Cardholm, L., Karlsson, S.: Cultures of trust: A cross-cultural study on the formation of trust in an electronic environment. *Proceedings of the 3rd Nordic Workshop on Security* (2000) 12–13
- [3] Singh, S.N., Dalal, N.P.: Web home pages as advertisements. *Communications of the ACM* **42**(8) (1999) 91–98
- [4] Alsudani, F., Casey, M.: The effect of aesthetics on web credibility. *Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology* (2009) 512–519
- [5] Robins, D., Holmes, J.: Aesthetics and credibility in web site design. *Information Processing and Management* **44** (2008) 386–399
- [6] Tan, S.J.: Strategies for reducing consumers’ risk aversion in internet shopping. *Journal of Consumer Marketing* **16**(2) (1999) 163–180
- [7] Liebermann, Y., Stashevsky, S.: Perceived risk as barriers to internet and e-commerce usage. *Qualitative Market Research* (4) (2002) 291–300
- [8] Schoorman, F.D., Mayer, R.C., Davies, J.H.: An integrative model of organizational trust: past, present, and future. *Academy of Management Review* **32**(2) (2007) 344–354
- [9] Fogg, B.J.: *Persuasive Technology: Using Computers to Change What We Think and Do*. Elsevier Inc (2003)
- [10] Fogg, B., Soohoo, C., Danielsen, D., Marable, L., Stanford, J., Tauber, E.R.: How do users evaluate the credibility of web sites?: a study with over 2,500 participants. *Proceedings of the 2003 conference on Designing for user experiences* (2003) 1–15
- [11] Thielsch, M., Blotenberg, I., Jaron, R.: User evaluation of websites: From first impression to recommendation. *Interacting with Computers* **26**(1) (2014) 89
- [12] Norman, D.A.: *Emotional design: why we love (or hate) everyday things*. Basic Books (2004)
- [13] Lindgaard, G., Fernandes, G., Dudek, C., Brown, J.: Attention web designers: You have 50 milliseconds to make a good first impression! *Behaviour and Information Technology* **25**(2) (2006) 155–126
- [14] Moshagen, M., Thielsch, M.: Facets of visual aesthetics. *International Journal of Human-Computer Studies* **68**(10) (2010) 689–709

- [15] Moshagen, M., Thielsch, M.: A short version of the visual aesthetics of websites inventory. *Behaviour and Information Technology* **32**(12) (2013) 1305–1311
- [16] Alm, S., Britton, T.: *Stokastik - Sannolikhets teori och statistik teori med tillämpningar*. Liber (2008)



# Which actions force users to change their hand grip on their mobile devices?

Tove Bäckström

Department of Computing Science  
Umeå University, Sweden  
backstrom.tove@gmail.com

**Abstract.** The way users interact with a smartphone differ. Some users almost always interact with just one hand and some are used to use two hands. Some of the actions(tapping, swiping, zooming, dragging etc.) that the users perform can be more or less easy to perform. In recent years Apple and other popular smartphone manufacturers has launched larger phones than they did earlier, and this becomes more popular. The smartphone is not longer a small pocket-device, it is more like a mini-tablet. The larger smartphones can affect the way that users are interacting with a smartphone. Because the interface is the same as in smaller screens.

This research is about to analyze which actions that force users to change their hand-grip while interacting with a prototype with common gestures/actions on a smartphone. The analyze also include how the users are interacting, are they using the thumb or the index finger? How are they holding the smartphone when they perform different actions?

This analyze is performed with a user test where eight participants navigated through a prototype on a smartphone with screen size 4,7". All of the tests was video-recorded.

The majority of the participants were holding the smartphone in one hand. The action that the most users were changing their hand-grip were to tap in the upper left corner and type a text-input. Some of the participants was moving the hand to the upper part of the smartphone to reach the upper left corner. When typing a text-input the most usual was to interact was to use the both thumbs. In the discussion in section 4 there are some design proposals based on the results of the tests.

## 1 Introduction

Today's smartphones are launched in different screen sizes and some of the phones are too large to fit in a regular pocket. Humans have different hand sizes and their ability to perform different actions on the screen can differ. Some gestures can be done with one hand, but sometimes users have to take the other hand or change the one-handed grip to perform an action. It depends on where on the screen the clickable element is positioned. Some users always uses both hands if possible. If the user is carrying something in the other hand for example a bag, it can be hard to perform the task that the user wants without releasing

the bag. Although the large mobile screen affect the user experience in terms of reachability, the large mobile screen can improve the ability of users to see more content on the viewport and can prevent scrolling.

In many apps and operative systems for smartphones the clickable elements are positioned in the corners. This areas are harder to reach than the elements in the middle. The larger screens that are launched comes with the same interface as the smaller ones. For example the Iphone 7 and the Iphone 7 plus have the same interface as an Iphone SE and the screens are 5,5" for Iphone 7 plus and 4" for Iphone SE. Unlike the Ipad-mini that have specific apps that is custom-made for a larger screen. Therefore this is an interesting thing to investigate. Is this something that the designers should take into account when designing interfaces to smartphones that have different screen sizes?

This paper is about how users are interacting when reaching different areas on the screen. Because with smaller screens, users are usually interact with a smartphone with just one hand and what is happening when the screens are getting larger?

A qualitative analysis has been made about which actions the users are changing their hand grip, in which areas of the screen the users are changing their hand grip, and which fingers that are used to perform an action. An action is to swipe from one area to another or tapping in an area on the screen. The analysis is made with a user test, where the participants are given instructions and a prototype to navigate through. The prototype is made by inspiration of usual shortcuts interactions on iOS, Apple's operating system for Iphone. The user tests were done on Iphone 6/6S and the screen is 4,7".

## 1.1 Background

The most common way users hold their smartphone is with one hand [1] when possible. If the phone is in one hand, the most regular way to interact with the screen is to use the thumb[2]. Steven Hooper<sup>1</sup> observed users at distance to see how they held their mobile devices. The users were not identified and nor their mobile devices. He divided the way to hold a mobile device in three different ways: one handed, cradled, and two handed. The most common way to hold a mobile device was with one hand. He also found out that the users often change their hand grip. This is also influenced by the environment where the user is interacting with a mobile device. If the user is walking, bicycling or driving or if the user just laying down in a sofa.

In [3] motiongestures, tapping end swiping, while walking are examined. This is because smartphones are often used while the user is performing another task. The result of this research shows that motion gestures mean less time looking in the screen while walking, than tapping. This means that motion gestures may be better to use when the user is distracted by walking, driving or bicycling.

---

<sup>1</sup> How Do Users Really Hold Mobile Devices? Uxmatters, <http://www.uxmatters.com/mt/archives/2013/02/how-do-users-really-hold-mobile-devices.php>

To understand users touch behavior on large mobile screens an empirical experiment was conducted[4]. The experiment collected the users pattern of tilting the device when reaching areas on a mobile screen.

One-handed interaction with a large screen and the usage of thumb is a known problem. In a study[5] a virtual thumb of a real thumb, to address the target, called *extendedthumb*, are proposed. The *extendedthumb* have a longer moving distance and moves in the same direction as the real thumb. To select a target, a red cross is viewed on the target and when the real thumb is lifting up the target is selected. An empirical user evaluation of the *extendedthumb* was conducted. The result of the user evaluation said that *extendedthumb* achieved higher accuracy, perceived effectiveness and perceived ease of use than normal touch.

A study [6] where input accuracy and pressing pattern of one-handed thumb interaction has been investigated. The study collected accuracy rate and input offset. This study had three combinations of button sizes and twentifive different input locations on the touch screen. They found out that locations providing high impact rates were different depending on the size of the touch area of the button.

Users interact with a mobile device in many different ways. There are sensors and cameras that can enrich the user experience. But the most reliable and simple way to interact with a smartphone is still to touch the screen on the device[7]. There have also been a study[8] that investigate different interaction techniques. This study have been made on a website though. The study explored the relative effects of different interaction techniques on user engagement, interface assessment and behavior outcomes. They found out that liking and creditability of a website are based on a positive valuation of interface, followed by a cognitive.

## 2 Method

To understand how users interact when reaching different areas on a smartphone screen, a user test was made to analyze in which actions users change hand-grip while interact with a smartphone. The test was done with a prototype to a smartphone. The participants got some actions that they were going to perform. Each action is in one view, and when an action is performed, the next view with next action is showed on the smartphone. For details of actions see section 2.1.

As the test is lasting the hand-grip(one or two hands), area on the screen, and which finger that was used was analyzed. All the user tests was video-recorded. When the tests were analyzed, som design proposals were suggested.

### 2.1 Tasks

The interaction techniques that are tested is swiping and tapping. This is because it is common interaction techniques to use when interacting with a smartphone. The actions are inspired by common actions in the Apple Ios(operative system of Apples Iphone). Tapping is the most common interaction techniques and swiping is used when answering a call or get to shortcuts. The Iphone key-board is

also included in the test. Because it is also a common way to interact with a smartphone, when typing a text message or searching.

**Table 1.** List of the actions that the participants were performing during the user test.

Order of actions performed	Action
1	Tap on the icon in the right lower corner.
2	Swipe from bottom to the middle of the screen.
3	Tap on the icon in the right upper corner.
4	Swipe from the left to the right on the screen.
5	Tap on the icon in the left lower corner.
6	Swipe from the top to the middle of the screen.
7	Tap on the icon in the left upper corner.
6	Tap on the search-field and make a text-input.

## 2.2 Materials

The materials that are used in the evaluation is an Iphone 6/6S and a hi-fidely prototype made in the program Proto.io. The prototype have a white background with black/grey tapping areas or swiping areas. Because the participants were supposed to not be distracted from other functions or design elements. See figure 1. The tapping areas that is used is taken from the Proto.io Ios design-kit and also the swiping elements and search-field and keyboard. The Iphone 6/6S have the same screen size and the test were done on different phones but with same prototype and screen size. The video-recording was made with an Iphone 5S and the purpose was to analyze the tests after they have been made.

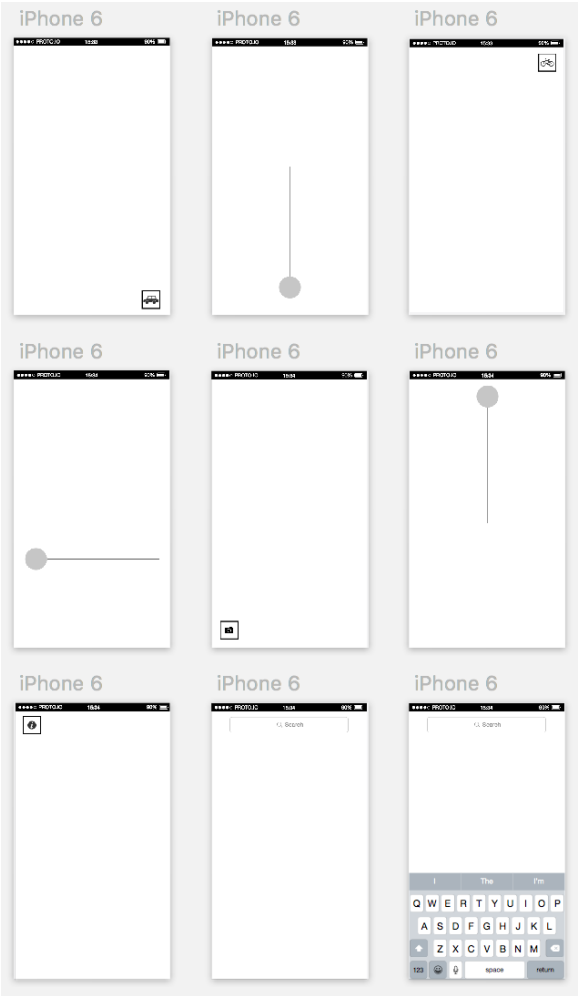
## 2.3 Participants

The evaluation was made with eight participants, four women and four men. The age was between 22-35. All of the participants were used to smartphones and use a smartphone daily. Three of the participants use an Iphone 6 daily and the rest uses a smaller or larger Iphone or Android smartphone. All of the participants are students or have been students at Umeå University.

## 2.4 Procedure

Before the test started the participants was given some information about how this test is used in this research and that they are anonymous. They were permitted to quit the test whenever they want it they are not comfortable. They were given instructions about the navigation. They were told that they were going to tap and swipe through a prototype with eight different views and that the last view were a search-view where they typed their name. The participants were told navigate through the prototype as they use to do. They were not told what the analysis was about. All of the user tests were video recorded, and all





**Fig. 1.** This is screenshots of the prototype that the participants navigated through. The rectangular icons were for tapping and the grey circular elements were for dragging along with the line (swiping). The search field was for tapping in the field and the make a text input.

the participants gave their permission. The environment where the user test was performed was not the same, three of the eight tests were done in a sofa, two while standing, and three at a table. This because it was hard to collect all the participants to the same place.

## 2.5 Evaluation

In the first case where the participants were tapping in the lower right corner five of eight participants were holding the smartphone in the right hand and tapped with the right thumb. Two of eight was holding the smartphone in both hands but used the thumb to tap on the icon. One participant were holding the smartphone in the left hand and tapped with the index finger. This showed that the majority started to hold the phone in one hand and that it was not a problem to reach the icon in the lower right corner.

The five participants that started the test with one hand used the right hand and the right thumb in the second case where they were swiping from bottom and up. And the other three participants did not change their hand-grip either. This shows that the swiping from bottom and up does not force the users to change their hand-grip.

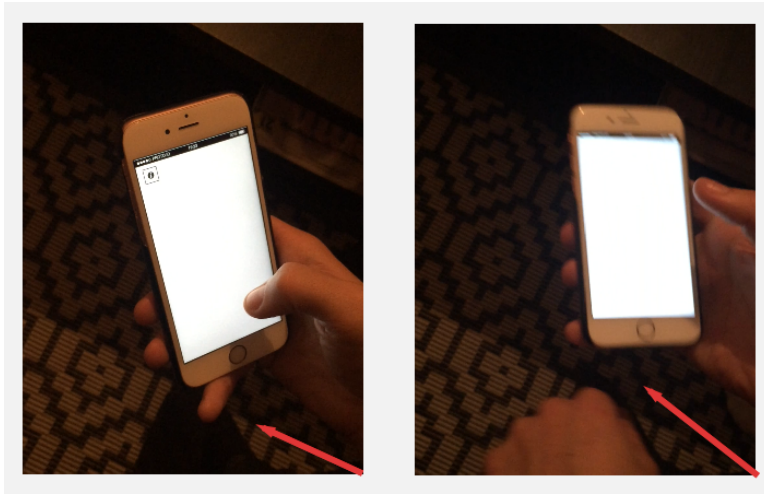
The next case was to tap on the icon in the upper right. There was two of the five one-handed participants that moved their hands, to reach the upper right corner. See figure 2. One of the two-handed participants changed the grip from holding the smartphone in two hands and used the thumb to use the index finger instead.

Swipe right made all but one participant to continue with the hand-grip they used last case. One participant changed the hand-grip by moving the hand down.

To tap the icon in the lower left was no problem for the two-handed participants, but one of the two-handed that earlier used the right thumb chose to use the left thumb in this case. Two of the one-handed participants changed their hand-grip to reach the lower left corner. One moved the hand down to the lower part of the smartphone and one participant supported the smartphone with the left hand, but still tapped with the right thumb.

The action to swipe from the top to the middle of the screen did not made someone to change their hand-grip. The case where the participants was tapping the icon in the left upper corner, made all but one of the one-handed participants to change their grip. Three of five moved their hands to the upper part of the smartphone and used their right thumbs. See figure 2. Only one took help with the left hand and used the left thumb to tap the icon in the upper left corner. All the participants that did the test with two hands did not change their hand-grip in this action.

To tap in the search-field did not made any participant to change their hand-grip, but two of the five one-handed did use both thumbs when they were texted their name on the keyboard. And two of the three two-handed used their both thumbs when they did a text input.



**Fig. 2.** Participant that moves the pinky finger to reach the upper left corner.

### 3 Result

This experiment shows that the majority of the participants did start by holding the smartphone in one hand, which was the most usual way to hold a smartphone.

The one-handed participants did almost all have trouble to reach the both upper corners without changing their hand-grip. Three of five had to change their hand-grip to reach the both upper corners and four of five changed when reaching the left upper corner. Two of the five one-handed participants chose to use both thumbs while writing their name on the keyboard. They did not have any trouble to reach the middle of the top of the screen. The action that made the most participants to change their hand-grip was to reach the upper left corner.

The two-handed participants did not have any trouble to reach the areas in the upper corners. But one chose to switch between using the thumb to the index finger in the case to tap the upper right corner. The two-handed participants did in most of the cases interact with one finger and had the other hand as a support. It was just in one case that the two-handed participant chose to use the left thumb instead of the right thumb. But in the text-input case two of three chose to use two fingers(both thumbs).

To summarize all participants, the two actions that had the most changes of hand-grip was to tap in the upper left corner and text-input. See figure 3. The most usual change when tapping in the left corner was not to take help from the other hand, but moving the hand that holds the smartphone. But when they typed on the smartphone keyboard, the most common change was to take help with the other hand.

Tap Lower Right	Swipe up	Tap Upper Right	Swipe Right	Tap Lower left	Swipe Down	Tap Upper Left	Tap Upper	Text Input	
Not changed	Not changed	Not changed	Not changed	Changed	Not changed	Changed	Not changed	Changed	2
Not changed	Not changed	Changed	Not changed	Changed	Not changed	Changed	Not changed	Not changed	4
Not changed	Not changed	Not changed	Not changed	Not changed	Not changed	Not changed	Not changed	Not changed	5
Not changed	Not changed	Not changed	Not changed	Not changed	Not changed	Changed	Not changed	Not changed	6
Not changed	Not changed	Changed	Changed	Not changed	Not changed	Changed	Not changed	Changed	8
Not changed	Not changed	Changed	Not changed	Not changed	Not changed	Not changed	Not changed	Changed	1
Not changed	Not changed	Not changed	Not changed	Not changed	Not changed	Not changed	Not changed	Not changed	3
Not changed	Not changed	Not changed	Not changed	Changed	Not changed	Not changed	Not changed	Changed	7

**Fig. 3.** A table that shows in which actions the participants changed their hand-grip while interaction with a smartphone. The five first line is the one-handed participants and the three last is the two-handed participants.

## 4 Discussion

The interesting thing to investigate was which actions that users were forced to change their hand-grip. It was hard to make the participants to navigate as natural as possible. The interaction differ of course in different situations. Therefore there was hard to investigate how the users interact when they for example is forced to use one hand. I did not come up with a smart solution to test how users interact with a smartphone if they are carrying something in the other hand. To get a more considerable result, the test would have to be longer and the participants would have to be more. The environment might be the same and somehow get the participants to start by holding the smartphone in one hand. Maybe tell the participants that the purpose is something else that distract them to think of how they are interacting. The participants were not told what the analysis was about. Only that they would interact with the prototype as they use to do.

One interesting thing about the result was that no one had to change the hand-grip to reach the middle in the top of the screen.

### 4.1 Design proposals

Many improvements have already been made in this area. The hamburger-menu is often placed in the upper left or right corner. But the IOS guidelines<sup>2</sup> do not use that type of navigation menu. Instead of having the main navigation in the upper corners they often have a navigation bar in the bottom of the screen. Which is better for the users user experience when they do not need to change their hand-grip. They also have a tool to reach the app-icons in the upper area of the screen.

One thing that I have thought about is that in many applications there are usually a back-button in the upper left corner. One proposal is to have a complement to tap in the upper left corner and it is to swipe on the screen from left to right. That is a reasonable gesture to do when you want to go back from where you came from. Maybe there are more clickable elements that can be replaced by gestures(like swiping) instead of clicking.

<sup>2</sup> iOS Human Interface Guidelines, <https://developer.apple.com/ios/human-interface-guidelines/overview/design-principles/>

Another way to think of it is to have different OS and apps for different screen sizes. Exactly like the Iphone and the Ipad. The Iphones today have three different screen sizes and that affect the user experience. A smartphone is something that is easy to carry unlike an Ipad that is larger. And the phones do not longer fit all users hands perfectly. I think that the screen size for smartphones is one important thing to think about when designing for smartphones.

## 4.2 Further work

It would be interesting to investigate how users interact with even larger smartphones. If they are interacting with just one hand or if they change the hand-grip more often or always use two hands. Maybe in a specific situation when the users are performing another task on the same time like walking or carrying something.

It would be interesting to know how it is best to interact with a mobile phone by ergonomic perspective. What is happening with our hands, fingers or neck when we interact with smartphones. How are we affected for a long time?

## References

- [1] Karlson, A.K., Bederson, B.B., Contreras-Vidal, J.L.: Understanding single-handed mobile device interaction. Technical report (2006)
- [2] Karlson, A.K., Bederson, B.B.: Thumbspace: Generalized one-handed input for touchscreen-based mobile devices. Proceeding INTERACT'07 Proceedings of the 11th IFIP TC 13 international conference on Human-computer interaction (2007) 324–338
- [3] Negulescu, M., Ruiz, J., Li, Y., Lank, E.: Tap, swipe, or move: Attentional demands for distracted smartphone input. Proceedings of the International Working Conference on Advanced Visual Interfaces (2012) 173–180
- [4] Chang, Y., L'Yi, S., Koh, K., Seo, J.: Understanding users' touch behavior on large mobile touch-screens and assisted targeting by tilting gesture. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (2015) 1499–1508
- [5] Jianwei, L., Dongsong, Z.: Extendedthumb: A target acquisition approach for one-handed interaction with touch-screen mobile phones. IEEE Transactions on Human-Machine Systems **Vol.45(3)** (2015) 362–370
- [6] Park, Y.S., Han, S.H.: One-handed thumb interaction of mobile devices from the input accuracy perspective. International Journal of industrial Ergonomics (40) (2010) 746–756
- [7] Rotzetter, D.: Gesture interaction techniques on cell phones: overview and taxonomy. Technical report, Department of Informatics University of Fribourg (2012)
- [8] Shyam, S.S., Saraswathi, B., Jeeyun, O., Qian, X., Haiyan, J.: User experience of on-screen interaction techniques: An experimental investigation of clicking, sliding, zooming, hovering, dragging, and flipping. Human-Computer Interaction **Vol.29(2)** (2014) 109–152



# Can mobile augmented reality enhance the user experience of internet shopping?

## A user study based on an interactive prototype toward augmented reality

Amanda Dahlin

Department of Computing Science  
Umeå University, Sweden  
dahlin-amanda@hotmail.com

**Abstract.** User tests have been conducted to investigate the hypothesis that augmented reality in mobile devices (MAR) improves the user experience and makes the user more positive toward an application. The participants were divided into two test groups. Both groups tested the same iPad application, with the difference that one included the MAR feature, and the other did not. The results showed a subtle indication that the concept of MAR might enhance the user experience. The participants were very positive toward this new innovative approach to buying clothes online, if the technology would be fully functioning, implemented using smart algorithms and 3D modeling projecting the garments onto the body and adjusting the fit depending on the customers' body.

## 1 Introduction

The technology of augmented reality is an advanced way of overlaying digital data on top of the physical world [1]. Augmented reality, or AR, can be described as a filter with digital information, otherwise unavailable to the user, added to the surrounding environment to add an extra dimension to the physical world. As technology in smartphones gets yet more advanced, with better cameras, GPS trackers, processing power and sensors, the possibilities for implementing AR technology in smartphones grows bigger. AR in smartphones enables it to be used on the go, namely mobile augmented reality (MAR) [2].

MAR has great potential since it presents information about the surrounding environment and adds digital information inaccessible to the naked eye, while moving around in the physical world. The technology has changed the way we interact and access information, enabling new types of user experiences [1].

The breakthrough of MAR in gaming has already come, but what if MAR combined with internet shopping might enhance the user experience? By enabling for customers to visualize in advance how garments will fit them personally might minimize the insecurity amongst customers and make them feel more confident shopping via internet.

Today the usage of mobile augmented reality is limited. There are a few apps, like Pokémon GO, SkyView, Ink Hunter, Augmented Car Finder, SnapShot Showroom<sup>1</sup>, utilizing the technique, although (except Pokémon GO) they are not yet very widespread. A lot of people got their eyes open for the technique as the MAR game Pokémon GO was launched this summer (2016), which got a lot of media attention.

The purpose of this study is to test if the concept of mobile augmented reality in the e-commerce clothing industry can improve the *user experience*. This will be done by using an interactive high fidelity prototype implemented as an iOS application for iPad and a so called A/B-testing methodology [3]. An A/B-test is conducted by testing two versions of a system or design, in this case with one version including the AR functionality, and one without. To measure the difference, the time spent using the app, the likelihood of a completed buy and a general grade given by the participants will be measured for objective results. To get more qualitative input a short semi structured interview will follow the test session. The participants will not be informed what is tested until after the interview.

## 2 Background

The clothing industry already has large revenue on the internet, which means people are willing to buy garments online even if they do not know if the clothes will fit. In the second quarter of 2016, the e-commerce sales was 8.10 percent of the total retail sales, which is about 32.5B USD in the US alone<sup>2</sup> and the numbers have been increasing (long term annual rate of almost 16.5 percent) every year.

If the customers would be able to visualize the garment on in advance using MAR, both customer satisfaction and user experience might increase. The technology of augmented reality has been around for a while, and has in recent years started to be integrated in mobile devices. There are some applications already utilizing this, but it has not yet been implemented in the clothing industry for e-commerce yet.

The shift from desktop to mobile, and now to mobile augmented reality, VR and ubiquitous computing integrated in everyday context, puts high demands on the interaction. It is difficult to design for a simple, intuitive interaction with a complex system and to do so it is good to develop standards for how to design interaction for MAR. The standards are few, or non-existing, since the technique still is quite new. Although, done right - the possibilities are great.

### 2.1 What is user experience?

User experience is a rather subjective concept and defines how a person experiences a system when it comes to instrumental (like usability) and non-

<sup>1</sup> Widder, B. Digital Trends. 2016. <http://www.digitaltrends.com/mobile/best-augmented-reality-apps/>

<sup>2</sup> Ycharts. 2016. US E-Commerce Sales as Percent of Retail Sales: 8.10 percent for Q2



instrumental (e.g. aesthetics, fun) elements [1]. It can be seen as a quality factor when talking about interactive systems, but since it differs between users it can be difficult to measure. Despite that, more and more people consider it to be a very important factor of system design [1]. User experience is today viewed as very important since it improves the product or service and makes it fit the target group better. A system that is easy to use and helps the user perform a task in a natural and easy way better withstand competition from other companies.

To be able to identify the user's needs research is a very important part of an User Experience (UX) Designers work. It is important to be well aware of who the people using it will be, what activities will be conducted, in what context it will be used and what technology will be involved. In other words, it is important that a PACT (People, Activities, Context, Technology) analysis [3] is made.

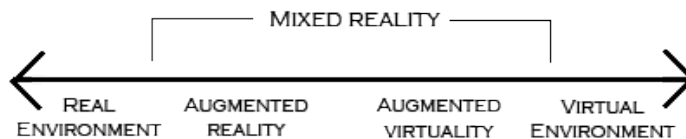
## 2.2 What is augmented reality?

Augmented reality is an enhancement of the physical environment surrounding us. It is a filter with digital information added to what we can percept with the naked eye. The technology enables interaction in real time with virtual elements applied on the real world superimposing our vision.

Digital information can be integrated in smart fabrics, head-mounted devices, or something as common as smartphones, and in the future possibly contact lenses [4]. To be able to call something augmented reality it must fulfill three properties [5]:

1. It should combine the real and the virtual
2. The augmentations should be interactive in real time
3. They should be registered in three dimensions

Examples of areas it can be used are for example tourism and geographical information, social activities, entertainment and games, shopping, maintenance etc. [6]. Information simply becomes more accessible using augmented reality, which lies within the mixed reality zone (see Figure 1).



**Fig. 1.** The virtuality continuum showing where augmented reality is in comparison to the real environment and virtual environment. Source: Image by Dahlin, inspired by [7]

## 2.3 Earlier work

Earlier studies on this subject have shown that mobile augmented reality might be a winning concept, especially when tested in the furniture industry. MAR,

in that field has been used to visualize furnitures in the customers home [8]. By using augmented reality customers can try different colors, placement and make customizations before purchasing an item. An example of where AR can be used is in museums. A study examined if a museum can be taken outdoors, notifying you when you pass something historical or worth paying extra attention to. This was only tested in beta and prototypes but is an example of how AR can enhance and be blended into people's everyday lives [9].

Some research also indicates that customers get influenced and affected in a positive way if a website uses AR or VR technique<sup>3</sup>. But many people also tend to be skeptical to new techniques such as AR, especially when it comes to important things that absolutely cannot go wrong. Although, once they try people seem to learn the concept of AR functionality rather quickly. To get visual feedback and assistance from augmented reality when performing a task seems very helpful and seems to improve the user experience [10].

A study from 2015 examined how augmented reality can be used in smart-phones to sell art. The goal was to focus on the user experience and if it can increase the art customer's satisfaction by letting the customers visualize how the art would look in its thought context [11]. The study mentions other examples of AR apps, like Contura, where the customers can project a fireplace to see how it would look like in their home. As a conclusion Botani says that the app has good usability since it helps the customers to visualize how a painting would look in their home [11]. A few years ago Co.Design wrote a post about concept use for AR in e-commerce, e.g. with a smart mirror projecting clothes on the customer's body<sup>4</sup>, but at that time the algorithms were not better than projecting an image like holding up the garment in front of you. Today the technology is better for that purpose though, since smarter algorithms can adjust fit better. E.g. Snapchat uses the technology of face recognition to add different filters to transform the users into different shapes. Synsam also use AR in some of their stores to enable their customers to try different glasses directly on a screen using a camera, but the service is not available on their website.

Yet another study from 2013 [6] investigated the usage of MAR in a shopping center, which showed that people have high expectations and demands on what the complex technique of MAR should be able to know and do and that it handles privacy and data in a safe way. Even though mobile devices just recently got powerful enough and sufficiently with sensors to be able to handle MAR people already have high expectations and visions for it, which is a good sign for the future.

---

<sup>3</sup> Woyke, Elizabeth. 2016. How Stores Will Use Augmented Reality to Make You Buy More Stuff. MIT Technology Review. <https://www.technologyreview.com/s/601664/how-stores-will-use-augmented-reality-to-make-you-buy-more-stuff/>

<sup>4</sup> Kuang, Cliff. 2009. Five To-Die-For Augmented Reality Shopping Apps. Co.Design. <https://www.fastcodesign.com/1313133/five-to-die-for-augmented-reality-shopping-apps>

Although some people are very positive about MAR systems, others tend to be very skeptical toward new technologies. How successful this new concept will be depends on the users and how accepting they are. Therefore it is important for designers to expose the gap between what is possible to achieve with technology and what the users want, need and expect [12] out of it, which is done by extensive user studies and UX work.

### 3 Methodology

To test the concept of mobile augmented reality as an aid in internet shopping, a native iOS application (developed in Objective-C) for iPad has been developed. Its aesthetics is similar to the online stores that exist today. It is interactive and the customers can go to different categories, scroll to view different garments, click on them to get more detailed information about that specific garment and add garments to their shopping cart. The only step that is not implemented in the process is the payment, but there is a button that says 'To payment' in the checkout which symbolizes a completed buy. The application is also not connected to a server or database, but fetches "dummy data" hardcoded locally.

The reason the concept was tested using a high fidelity (HiFi) prototype is because HiFi prototypes have shown to be best suited for testing mobile augmented reality (MAR) when it comes to evaluating concept, realism and feeling, compared to Lo-, or Medium-Fi [2]

A recent study also developed a set of guidelines for designing MAR apps, which was taken into consideration. These guidelines are [5]:

1. Clearly show what objects the application is referring to in the physical world to support learning.
2. Represent the internal state in so far that users may learn how the computer vision algorithm performs satisfactorily.
3. Inform users of actions needed to adequately maintain an environment suitable for augmenting.
4. Afford the relationship the device has to the trackable (e.g a marker).
5. Allow easy transition between known interaction paradigms and the augmented interaction paradigm of remediating content
6. 3D content should convey its meaning at natural viewing angles
7. Provide the same entry point for interaction even though the content is different
8. Humanoid characters that allow interaction should be responsive to the user

The test used the A/B-testing methodology [3] in combination with other user testing techniques. A/B-testing means that two different versions of the prototype were tested. One of them included the feature of MAR, simulating a projection of the selected garment onto the user by using the smartphone camera, and the other did not. A/B-testing is usually used to test whether one of two different designs is better than the other rather than to test if a specific feature adds value to the user experience. But since the purpose was to measure

the difference in user experienced an extra group was added as a control group. Also the time spent using the app and the likelihood of a completed buy will be noted for objective results. A buy was counted as completed if the user added clothes in the shopping bag, went to the checkout and clicked the button 'To payment'. The division into the two different groups were randomized.

During the testing session a key task-test [13], also known as scenario based user testing<sup>5</sup> was conducted. This methodology means that the test person is presented with a scenario to understand the context and are given some tasks which he/she should do. The scenario and tasks were combined with the technique think-aloud commentaries [3], which means that the test person explained what he/she did and thought during the test session.

The test was then followed up by a semistructured interview [3]. This part was important to capture details about the interaction and get to know more about the users' subjective view of the user experience. This was made by letting the user grade the level of usability and experienced level of "satisfaction" while using the application.

The questions asked after the test was the same for both test groups, except some extra questions was added for the MAR group specifically asking about that functionality. The interview gave both qualitative and quantitative data. The quantitative and statistically measurable data was given by questions using the Likert scale [14].

The questions was mainly about the user experience during the interaction. Afterward all the user tests were conducted the results of the two groups was compiled, compared and analyzed to see if there was a significant difference. Focus lied on the user experience and to see if the users seemed to prefer one version over the other.

### 3.1 Description of the prototype

The prototype is implemented to look like most e-commerce websites selling clothes on the web today. On the left side of the screen there is an overview of the categories and on the right side the start page is the category 'New in' showing a selection of the garments, see Figure 2. When the user clicks a specific garment he/she is redirected to a new page showing information about that specific garment, see Figure 3. On this page the user can choose to go back to the category, choose size, add the garment to the basket or, if testing the MAR version, check out how it would look using "augmented reality". In the augmented reality mode both the front and back camera can be used to get the best view of the garment on, see Figure 4, 5, 6, 7. In the MAR view it is possible to manipulate and adjust the garment's size, location and rotation to make it fit the best way. In the implementation of "real" MAR this will not be possible, since the application will detect the body and the garment will adjust itself accordingly.

<sup>5</sup> Norman,D., Nielsen, J. 2014. Turn User Goals into Task Scenarios for Usability Testing. <https://www.nngroup.com/articles/task-scenarios-usability-testing/>. Accessed 2016-09-25

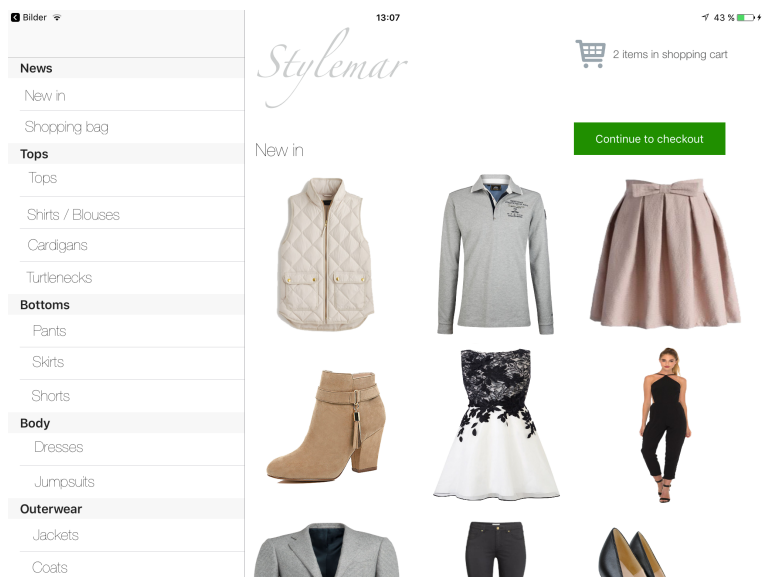


Fig. 2. Screenshot of home screen for prototype on iPad.

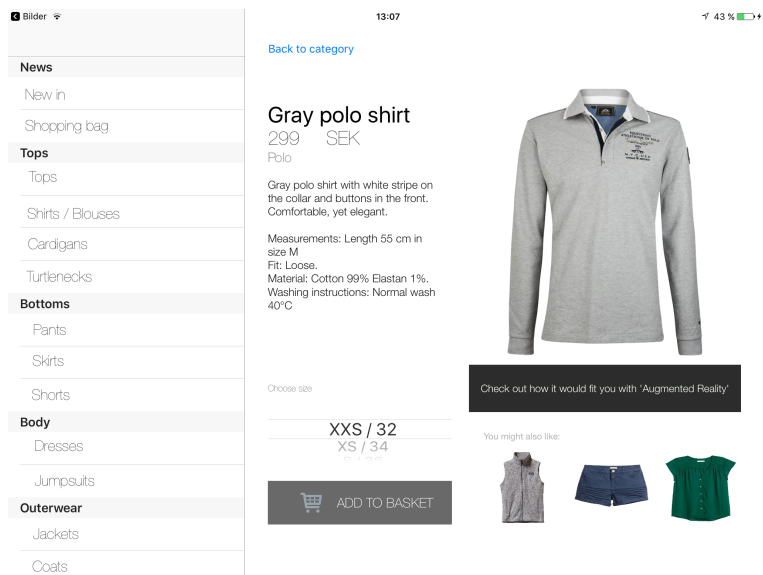


Fig. 3. Screenshot of garment detail view for prototype on iPad.



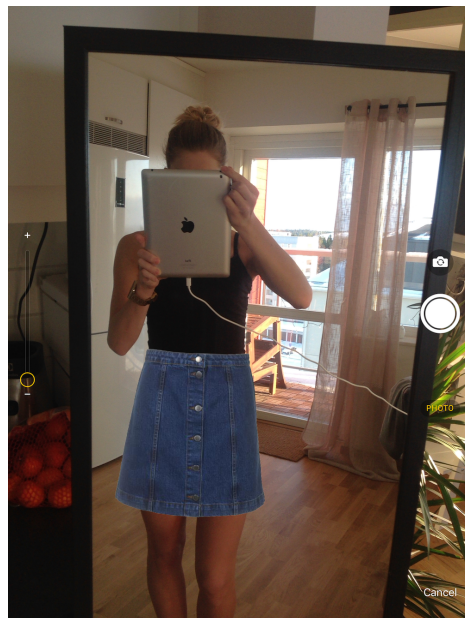
**Fig. 4.** Screenshot of beige vest projected onto a body. Example of the MAR concept using front camera



**Fig. 5.** Screenshot of gray blazer projected onto a body showing the augmented reality concept using back camera



**Fig. 6.** Screenshot of black dress projected onto a body showing the augmented reality concept using back camera



**Fig. 7.** Screenshot of jeans skirt projected onto a body showing the augmented reality concept using back camera

### 3.2 User tests

Since the user tests were about grading the experience of using the application a scenario and task based test was conducted. The participants had to get familiar with the app and interact with it to get an opinion. To get a similar experience for both groups the same tasks and scenario was used, except that testing the MAR feature was included in one of them. The interview questions were also about the usability and good and bad aspects about the prototype to not make it too obvious what was being tested.

Each test started by letting the participants know about their ethic rights, that the test is entirely voluntary, anonymous and that the functionality of the app is tested and not the knowledge or skills of the participants. They were also informed that they were free to quit the test session if they wanted without any questions asked about the reason of them doing that.

After getting to know about their ethic rights they were told about the context and general information about the setup of the user test and that it is a prototype and not a fully functioning app that is tested.

After some general questions about the participants background like age and earlier experience of online shopping (presented in the following section Participants), the tasks they were given to do were:

1. To scroll through the clothing categories as they wished and to explain what they saw and how they felt
2. To go back to the category 'New in' and tap on a garment they wanted to know more about
  - (a) If the user put the garment in shopping cart, they were asked to add another item
  - (b) If he/she went back to the category overview, they were asked to add two items to the shopping cart
  - (c) *If the MAR version was tested, they were asked to press the button 'Check out how it would fit you using Augmented Reality' if they did not do so by themselves*
3. Then they were asked to navigate to the shopping cart, look around and interact with the page as they wanted (if the user pressed the button 'Continue to checkout' and the button 'To payment' it was counted as a completed buy)

The time it took to complete the tasks above was also measured to see if participants using the MAR version spent longer time using the app in comparing to those testing it without the MAR functionality. The interview that followed the tasks included following questions:

- Would you use an app similar to this if it was out on the market?
- What was good about it?
- What was bad?
- Is there something that should have been done differently?
- What grade from 1-10 would you give the app in general, in comparison to similar apps and websites?



If the MAR version was tested the following questions were also included in the interview:

- What did you think about the augmented reality part?
- Is that something you would use if it would have been fully functioning adapting to your body?

## 4 Results

### 4.1 Participants

In total 7 user tests were conducted, 4 with the MAR functionality and 3 without. All of the participants were students in different fields like engineering, health care, law and order and police. Their experience with of internet shopping varied a little, but all of them felt rather confident interacting with technology. They were aged between 23 and 26 with an average of 24.14. For a summary of the results from the user tests see Table 1. They were divided into the two groups by random.

### 4.2 Test results

**Tech. conf.:** How confident the participant feels using technology where 1 is not at all and 7 is very confident.

**Exp. of e-com.:** How experienced the participant is buying things online where 1 is very seldom and 5 is very often.

App version	Gender	Age	Tech. conf. (1-7)	Exp. of e-com. (1-5)	Time (min)	Grade	Completed buy
w/ MAR	Female	23	6	4	12	8	No
w/ MAR	Female	26	7	2	7	9	Yes
w/ MAR	Female	26	6	3	6	8	Yes
w/ MAR	Female	23	6	3	7	8	Yes
w/o MAR	Male	23	6	4	3	7	Yes
w/o MAR	Male	24	4	2	5	7	No
w/o MAR	Female	24	5	4	4	8	Yes

**Table 1.** Table of results from user tests

### 4.3 Summary of interviews

In general the participants found the app very easy to use and to understand and that they could imagine themselves using the app. Many of them liked the minimalist design a lot and that the images of the garments were big. The

majority also thought it was good that the subcategories were visible the entire time. What many participants found needed improvement was the information about each garment in the overview. However, in this case it was not the usability of the application that was tested but the user experience and the participants opinions about the MAR feature.

#### 4.4 Comments about the MAR feature

Concerning the MAR functionality the participants thought it was very cool, innovative and something they had never seen before. A participant that rarely buys clothes online said the reason for that is because of insecurity in will fit but was very positive about this solution if it was made to automatically adjust the garments to the body in a relatively realistic way. Even though they liked the idea some test participants were skeptic they would really use it every time. However they found it would be helpful in making a decision between two colors or to help convince them if they were insecure about buying the garment or not. All of the participants said that they definitely would use the MAR functionality if it worked really well, adjusting the garment to their body automatically. Several participants said it would be more like a bonus feature than a necessity though, but that that it would be something that stood out against competitors and could attract more customers.

## 5 Discussion

### 5.1 Analysis of results

The results from the two test groups shows that the participants are somewhat similar when it comes to age, background and experience, which is positive for the analysis but does not give the same width to the study. From the received results the augmented reality feature does not seem to affect the rate of completed buys even though it differ a little in advantage to the MAR version. What differs is the time spent using the app, on average 8 minutes for the MAR group and 4 minutes for the control group. That the MAR participants spent twice as much time interacting with the app might be explained by them having another feature to examine. On the other hand, if customers stay in the "store" longer it can have a positive effect on the buying rate.

The app grade also differs. The average for the MAR version 8,5 for the MAR group and 7 for the other 2. This indicates a small positive effect on the user experience in the MAR version, but it is hard to draw any definitive conclusions since the number of participants are quite small and the prototype is not fully functioning regarding the augmented reality feature.

Since the participants were about the same young age they might be more susceptible for new technology and might be more experienced with online shopping. The participants in this user study might be the actual target group for this kind of functionality. To be able to visualize garments might also increase

the number of buys made online because they feel more confident it will fit them. Although, this is contradicted by the fact that there was not that big of a difference between the rate of completed buys between the two groups. But the limited implementation of the MAR in the prototype might also be the reason for that. From the interview it was found that the participants were very positive about the MAR in e-commerce, especially if the garments would automatically adjust to the body, using smart algorithms and 3D modeling of the garments projected. The positive feedback received during the interviews is a good sign that mobile augmented reality in the field of e-commerce might be the future.

Version	AVG age	AVG Time	AVG Tech. conf. (1-7)	AVG exp. e-com. (1-5)	AVG Grade	Buy rate
w/ MAR	24.5	8	6.25	3	8.25	75%
w/o MAR	23.67	4	5	3.33	7.33	66.7%
Total	24.14	6.29	5.71	3.14	7.86	71.4%

**Table 2.** Table showing compiled results for the two test groups and total

## 5.2 Subjectivity of user experience

User experience is a subjective measurement which makes it difficult to generalize what is good and what is not. It would be better to have conducted more user tests, but according to different experts in usability and user experience the number of five test participants finds almost as many issues in usability than recruiting more people <sup>6</sup>. But to measure the user experience for the statistic analysis in this study it would have been better to get more participants in both groups to be able to draw more credible conclusions.

## 6 Conclusion and future work

From the results received there is a slight indication that MAR in e-commerce might have a positive effect on the user experience. To continue to investigate if this hypothesis really is the case, it would be important to implement a better functioning prototype that projects a garment that adjusts depending on the shape of the the current user's body. The algorithm must be able utilize body recognition and 3D modeling to adjust the garments width, size, texture, shadowing of the garment, distance to the user and the angle the device is held. To confirm that the user experience really is enhanced using MAR it is also important to conduct more user tests, both qualitative and quantitative.

For future work it would also be interesting to see if the results differs between women and men, but it was not possible to draw any conclusions about this in

<sup>6</sup> Nielsen, J. Nielsen Norman Group. 2012. <https://www.nngroup.com/articles/how-many-test-users/>

this study, since only two participants were men and both of them were in the same test group.

## References

- [1] Irshad, S., Rohya Bt Awang Rambli, D.: User experience of mobile augmented reality: A review of studies. In: 3rd International Conference on User Science and Engineering (i-USer). (2014)
- [2] de Sa, M., F. Churchill, E.: Mobile augmented reality: Exploring design and prototyping techniques. In: Human-computer interaction with mobile devices and services: Proceedings of the 14th international conference, (MobileHCI '2012). (2012) 221–230
- [3] Benyon, D.: Designing Interactive Systems - A comprehensive guide to HCI, UX and interaction design. Volume 3. Pearson, Edinburgh, UK (2014)
- [4] Lukosch, S., Billingham, M., Alem, L., Kiyokawa, K.: Collaboration in augmented reality. Computer Supported Cooperative Work (CSCW) **24**(6) (2015) 515–522
- [5] Gjosaeter, T.: Interaction with mobile augmented reality: An exploratory study using design research to investigate mobile and handheld augmented reality. PhD thesis, University of Bergen (2015)
- [6] Olsson, T., Lagerstam, E., Kärkkäinen, T.: Expected user experience of mobile augmented reality services: a user study in the context of shopping centres. Springer-Verlag **17**(2) (2013)
- [7] Olmedo, H.: Virtuality continuum's state of the art. In: International Conference on Virtual and Augmented Reality in Education. Volume 25. (2013) 261–270
- [8] Khairnar, K., Khairnar, K., Mane, S., Chaudhari, R.: CRC: Furniture layout application based on marker detection and using augmented reality. International Research Journal of Engineering and Technology **2**(7) (2015) 12
- [9] Marshall, T.: Moving the museum outside its walls: an augmented reality mobile experience. Technical report, KTH Royal Institute of Technology, KTH Royal Institute of Technology, Stockholm, Sweden (2011)
- [10] Nilsson, S.: Augmentation in the Wild: User Centered Development and Evaluation of Augmented Reality Applications. PhD thesis, Linköping Studies in Science and Technology (2010)
- [11] Botani, R.: Art gallery web portfolio with web shop and app how to promote art through a web shop and app using augmented reality. Master's thesis, Sodertorns hogskola, Institutionen for Naturvetenskap, miljo och teknik (2015)
- [12] Dhir, A., Al-kahtani, M.: A case study on user experience (ux) evaluation of mobile augmented reality prototypes. Journal of Universal Computer Science **19**(8) (2013)
- [13] Krug, S.: Dont Make Me Think. Volume 2. New Riders Publishing (2005)
- [14] Munshi, J.: A method for constructing likert scales. Social Science Research Network (2014)



# Does haptic feedback effect users time to action in a form based interface?

Simon Ekdahl

Department of Computing Science  
Umeå University, Sweden  
`id13sel@cs.umu.se`

**Abstract.** In this study we compare two populations of feedback, visual and visual together with haptic feedback. We investigate if haptic feedback compared to visual feedback used in a mobile form-based interface can have any impact on how long a user will wait before taking second action and click again. In a user study consisting of ninety six participants divided in to two feedback groups, we measured how long it took for the test subjects to first click a button and then click the same button a second time when the interface did not respond within the expected time period. Results show that there is a significant difference in time to action between the two populations and that people who received haptic feedback on clicks in general did wait longer before taking second action.

## 1 Introduction

When a digital interface's response time is slow and users input do not result in immediate feedback, users tend to click the same button multiple times [1]. Multiple clicks from the user may cause a system to abort and perform a task all over again, or even starting multiple tasks along side each other. This is an unwanted error resulting in bad user experience.

Several software developers have throughout the years tried to tackle the problems occurring when a program needs time to perform a task before presenting a result for the user. According to [2] for delays lasting more than 1 second users should be provided with feedback to indicating that the computer is working on the problem, for example by changing the shape of the cursor.

Many of today's apps and software use visual feedback in the form of loading icons, like the one shown in Figure 1, to indicate that the user's input has been registered and that the task is being processed in the background.



Fig. 1: Activity indicator provided by Apple inc for iOS applications

While loading icons are common, studies like the one presented in [3] have shown that the presence of loading indicators may cause users to perceive a

software as being slow. With this in mind visual feedback, like loading icons might not be the ideal feedback to use in an interface if it can be avoided.

Other types of visual feedback common in user interfaces are small cues and color changes of elements upon clicks, letting the user know that their input have been registered by the software.

Devices like smart phones have the ability to provide more than just visual feedback to it's users. One type of non visual feedback is haptic feedback (sometimes referred as taptic [4] or tactile [5] feedback.) Haptic feedback is the use of sense of touch in a user interface design to provide physical information cues to an end user. Haptic feedback used in mobile phones generally means the use of vibrations from the device's vibration alert motor<sup>1</sup> or Taptic Engine<sup>2</sup>.

One use of haptic feedback in smart phones has been together with the touchscreen keyboard, when each letter press triggers a short vibration, letting the user know they have pressed a key.

In this study we investigate how long time users will wait before taking action and clicking a button a second time when a form based interface does not respond within the expected time period. The test is conducted by dividing test subjects in to two separate groups that will either receive visual feedback (cues indicating that a button have been pressed) or visual cues together with physical haptic feedback (a short vibration) upon button clicks. In the study we want to investigate if there is a significant difference in time to action between users receiving visual feedback compared to users receiving both visual and haptic feedback on input when navigating through an interface where loading icons are absent.

We believe that haptic feedback can be used to reinforce the message that an application has registered a user's input and thus make the user more convinced that the application is working on their request. We believe that this will make users prone to wait longer, than they normally would when receiving visual only feedback, before taking some action by re-interacting with the interface.

We hope that this study is a contribution to the area of feedback design and that it is both sought for and could make a difference in future user interface development.

## 2 Earlier Work

Study presented in [6] has shown that user satisfaction will decrease as system response time increase. What is considered to be an acceptable response time in a given situation is based on the assumption that people have established

---

<sup>1</sup> Vibration alert motor. <http://www.namiki.net/product/vibration/motor/tech.html> A vibration alert motor is a small motor often used in cell phones to alert the user of incoming calls using haptic feedback.

<sup>2</sup> The American tech company Apple have developed it's own physical feedback provider called the "Taptic Engine" which has been integrated in many of their late products, like the iPhone 7 <http://www.apple.com/iphone-7/>



expectations of response time based on their past experience [7]. In some situations, like pressing a button, users might expect a response within a tenth of a second. In these cases a delay lasting for a few seconds might be unsettling for the user and may result in errors caused by the user taking second action. Results from the study conducted in [8] claims that people are generally willing to wait longer for a website to load if they are provided with a loading bar. The study presented in [9] suggests that some types of visual feedback can distract a users attention from the passage of time while other studies, like the one in [10] suggest the opposite, that some types of feedback actually results in people paying closer attention to temporal information, making the wait seem longer. Just like the study in [3] showed that the use of loading icons may cause users to perceive a software as of being slow.

When it comes to feedback, adding physical cues to visual has shown to increase task performance significantly beyond that found in a visual-only system [5]. Results from the Study [10] made in 2008 showed that haptic feedback upon key strokes on a touchscreen keyboard resulted in significantly improved finger-based text entry, bringing it close to the performance of a physical keyboard. The study presented in [11] where a tilting interface was used, showed that the use of haptic feedback could improve users performance in terms of accuracy and speed up to 22 percent. Many of today's software have embraced both visual and some times auditory feedback when navigation through a digital environment, but they often tend to skip physical or haptic feedback even though research indicates that in can enhance task performance.

One market that has embraced the multi-sensing approach in a satisfying way is Console Gaming, where they use vibrating controllers in order to enhance gaming experience and provide different kinds of information to their players<sup>3</sup>.

All though haptic feedback can enhance task performance in terms of accuracy and speed, not many studies have been made on haptic feedback and it's effect on users clicking behavior or time to action.

## 3 Method

### 3.1 Test Application

We developed an iOS application specifically for this study. The application consists of a multi-page reply form with different questions and answers on each page. Test subjects were then provided with one out of two sets of feedback upon clicks in the interface. The main purpose of a test have been to track the time it takes for a test subject to click a button a first time and then click the same button a second time, when the first click did not result in showing the next page in the test sequence. The study was performed on an Apple iPhone 7, 128GB Smart Phone.

---

<sup>3</sup> Xbox Wire staff (2013), Xbox wire staff explains how vibration in the Xbox One controller creates a rich and immersive gaming experience. <http://news.xbox.com/2013/06/06/xbox-one-controller-feature/>

### 3.2 Form layout

The reply form consists of five different pages containing general questions about age, gender, occupation, housing and mobile operating system. Each of the five pages contains of a set of radio buttons representing different answers corresponding to the question on the page and a "next button" leading to the following page in the test sequence. Figure 2 illustrates the third page in a test sequence where the button containing the text "Work" has been selected.

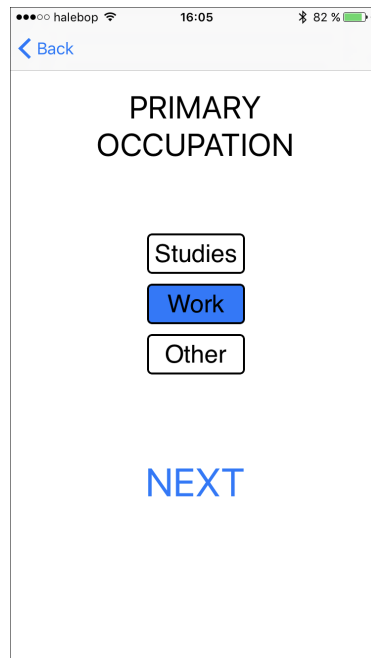


Fig. 2: Visual appearance of page 3. Buttons with a blue background is representing the selected choice

Page 1, 2, 3 and 4 appears to the test subject after a fixed response time set to 0.2 seconds. A delay of 0.2 - 1.0 seconds will make the users notice the delay and thus make them feel the application is working on the command [2]. The fifth page in the sequence will need two inputs from test subject to load.

Tapping the next button the first time on page 4 triggers a timer to start measuring time. When the test subject clicks the next button a second time on page 4, the timer stops and page 5 is shown. The application keeps track of what type of feedback is provided through out the test and the time it takes for a test subject to click the next button a second time on page 4.

The flowchart for a test sequence is shown in Figure 3.

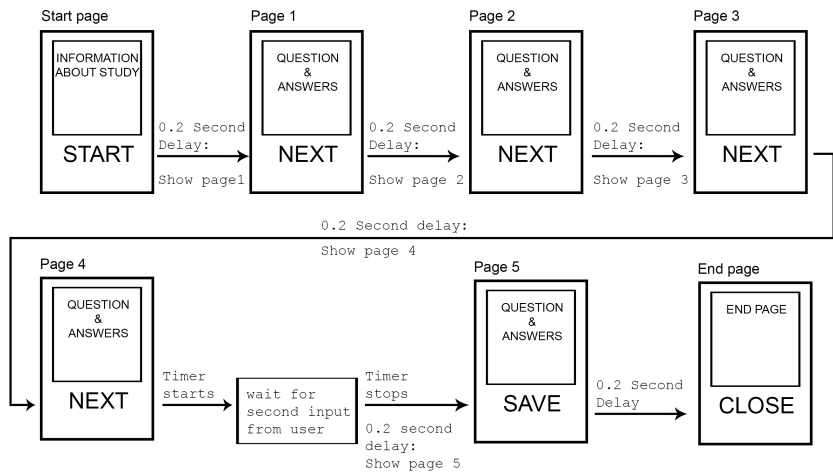


Fig. 3: Flow chart of the form navigation and events

### 3.3 Two sets of feedback

During a test sequence, the application provides one out of two different sets of feedback to the test subject upon button clicks:

- 1) Test subjects will receive only visual feedback to indicates that a button was tapped, before loading the next page
- 2) Test subjects will receive both visual and haptic feedback<sup>4</sup>. when a button is tapped before loading the next page.

The visual feedback provided by the "next button" is inspired by iOS design guidelines for UIButton<sup>5</sup>. The button has by default the appearance called "default" with a semi-dark blue text color. When a test subject taps the "next button" the button text flashes to a "highlighted" state, changing the text color to a much brighter blue tone , and then goes back to the default state as the test subject's finger leaves the touch screen of the device. The visual differences when a button is default or tapped is illustrated in Figure 4 and Figure 5

NEXT

Fig. 4: Visual appearance of an interface button. State: Default

NEXT

Fig. 5: Visual appearance of an interface button. State: On tap

<sup>4</sup> A short burst of physical feedback provided by the iPhone 7, called using swift 3.0 function `AudioServicesPlaySystemSound(1520)`  
<sup>5</sup> Apple developer, <https://developer.apple.com/reference/uikit/uibutton>

### 3.4 User Study

The test group in our user study consisted of ninety six people in the ages 18-56 years. The main part of the participants were chosen by randomly addressing groups of two people or more whom at the time of the study remained in Umeå University's premises. The people willing to contribute were handed a smart phone with instructions of how to perform the survey. All tests were performed on the same iPhone 7 provided by us and the majority of the tests were done at Umeå University campus in Sweden. Each participant were asked to fill out the reply form once by simply answering the questions on each of the five pages. Every other participant received one of the two specified sets of feedback through out the whole test sequence, and was not aware that other participants could receive a different kind of feedback. The pages in the form looked the same for both groups regardless of the provided feedback. This in order to give all participants the same visual conditions when filling out the form. Each test person filled out the form once and thus only performed one time to action test.

### 3.5 Data assumptions and methodology

We can assume that all samples are independent since participants performed the tests individually and therefor did not effect each other. All test were performed in a non controlled environment in order to come as close to a real life situation as possible.

Before performing the hypothesis test we first want to verify the assumption of normality for each of the two populations. Prior to this, we also want to remove some of the outliers<sup>6</sup>. Since our study was performed in a non controlled environment test subjects might have been exposed to distracting elements affecting their time to complete the survey. This might result in some unusual large differences in the test data. The outliers can be identified by using a box plot, like the one seen in Figure 6. The main reason for removing outlier is to get a more normally distributed data set. Given the assumptions that our data is normally distributed and independent, we can perform a Two sample T-test. With a Two Sample T-test we examine whether the the means of our two independent populations are equal which corresponds to the null hypothesis ( $H_0$ ). We have chosen a 99 percent confidence interval. If the Two sample T-test generates a significant result, we can reject the null hypothesis and assume with 99 percent confidence that the difference between the two population means should not be considered as random.

To perform the statistic analysis we used Minitab Express<sup>7</sup>, a statistical software which also have been used to generate the figures linked to our results.

<sup>6</sup> An outlier is an unusually large or small observation. Outliers may cause a disproportionate effect on statistical results [12], such as the the mean value, which can result in misleading interpretations. <http://blog.minitab.com/blog/michelle-paret/how-to-identify-outliers-and-get-rid-of-them>

<sup>7</sup> Minitab Express <http://support.minitab.com/en-us/minitab-express/1/help-and-how-to/basic-statistics/inference/how-to/two-samples/2-sample-t/interpret-the-results/all-statistics-and-graphs/>

## 4 Results

After removing four of the outliers we had 44 samples in our visual feedback population and 48 samples in our haptic feedback population. We have tested our data for normal distribution using the Anderson-Darling test, which gives us a P-values greater than 0.05 for both populations (Table 1.)

Table 1: Anderson-Darling test

Feedback	P-value
Haptic	0.5368
Visual	0.5869

Therefor we can not reject null hypothesis ( $H_0$ ) meaning we can assume both populations to follow a normal distribution. Given this we will use a Two sample T-test to examine whether the the means of our two independent populations (Table 2) are equal to each other. Our confidence interval is set to 99 percent and the null hypothesis ( $H_0$ ) that we want to reject can be described as  $H_0 : \mu_h(t) = \mu_v(t)$ . Equal variances are not assumed for this analysis.

Results from the Two Sample T-test is presented as a box plot in Figure 6, showing the difference between the two populations. The Two Sample T-test resulted in a P-value of  $P = 0.0003$ . Since  $P < 0.01$  we can reject the null hypothesis ( $H_0$ ) which indicates that there is a statistically significant difference between the two samples.

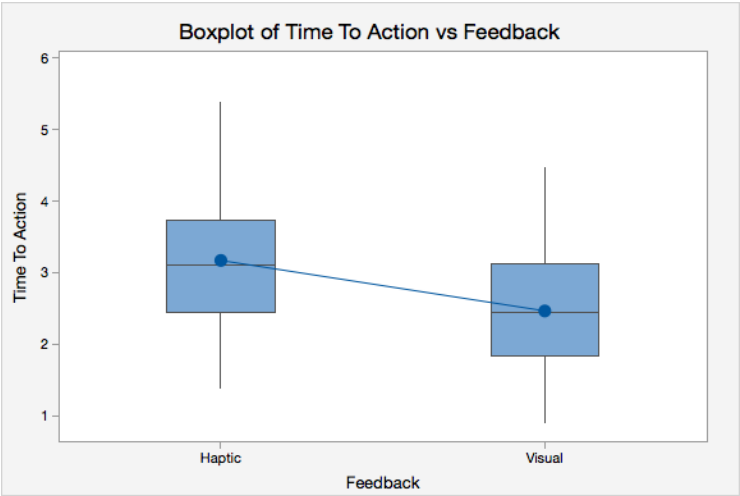


Fig. 6: Box plot ot Time to action comparison between the visual feedback and the haptic feedback populations

Table 2: Time to action

Feedback	N	Mean	StDev
Haptic	48	3,1692	0,9669
Visual	44	2,4625	0,8290

## 5 Discussion

Results from the study show that there is a significant difference in time to action between the two populations depending on the type of feedback that was received upon button clicks in the interface. We believe that difference in time indicates that people who receive haptic feedback together with visual feedback are more convinced that the software did register their input and thus will wait longer before taking second action. The mean difference in time to action between the two populations is approximately 0.71 seconds. This might not seem like a big difference at first glance. But when taking the mean values of time to action between the two populations into consideration, 0.71 seconds is in fact a 28,7 percent difference in time.

The reason for removing some of the extreme outliers before performing the statistical analysis have been to get more normally distributed data sets. When a data set is relatively small, just a few extreme values can offset the normal distribution and affect the mean value excessively. However the Two Sample T-test is considered to be robust against non-normality<sup>8</sup>, meaning that even though normality is an underlying assumption for the test, it should work for non normal data almost as well as if the data were normally distributed.

When performing the user study, the main purpose of the test was not told beforehand to the test subjects, since that information could affect their performance and clicking behaviour. We do not believe that the answers from the form questions have had any impacts on the results of this study and are therefor not taken in to consideration.

### 5.1 Drawbacks and Limitations

One of the drawbacks from the study presented is the use of one specific type of haptic feedback. A vibration alarm motor or taptic engine have the ability to provide many different types of haptic feedback to a user and it is up to the developer to chose which feedback to implement. For the case of this study we chose a subtle burst of haptic feedback using the Swift 3.0 command "AudioServicesPlaySystemSound(1520)" upon button clicks. With this in mind the results of a similar user study might have a different outcome if other types of haptic feedback is used.

<sup>8</sup> Two Sample T-Test robustness and the impact of non normality <http://support.minitab.com/en-us/minitab/17/Assistant\Two\Sample\t.pdf>

## 5.2 Future Work

In the future, similar user studies can be performed in different kind of interfaces or with different kind of haptic feedback. The result of our study only show that there is a significant difference in time to action when users are provided with a small burst of haptic feedback upon clicks in a task driven environment. Our results does not investigate user experience or how haptic feedback effects how an interface is perceived by the user. An interesting future study would be to investigate how haptic feedback of different magnitude or vibration-pattern would effect users clicking behavior and overall experience.

## 6 Conclusion

Given that our Anderson-Darling test showed positive results for both populations when testing for normal distribution we could assume normality and were able to continue with the Two sample T-test analysis. The results showed a statistically significant difference between the two samples leading to the conclusion that test subject receiving haptic feedback are generally waiting longer before taking second action compaired to test subjects receiving visual only feedback.

Considering that the time difference is in fact less than a second leads us to the conclusion that haptic feedback might not be the ideal feedback to use instead of visual feedback, like loading icons, but might have a beneficial role to play when it comes to other areas of interface design.

Even though the study shows that there is a significant difference between the two populations we should also examine the effect size of the statistical difference. This can be done using Cohen's  $d$ <sup>9</sup>. Cohen's  $d$  gives us a value of  $d = 1.03$  which indicates a large effect size [13] leading to the conclusion that the time difference between the two types of feedback should be considered as meaningful and therefor is something that developers should account for in practice.

## 7 Acknowledgements

The Author would like to thank all the peer reviewers who have assisted with valuable input on the content and structure of this paper.

## References

- [1] Hutchinson, H.B., Bederson, B.B., Druin, A.: The evolution of the international children's digital library searching and browsing interface. In: Proceedings of the 2006 conference on Interaction design and children, ACM (2006) 105–112

---

<sup>9</sup> Cohen's  $d$  - research the effect size of a statitital difference <http://rpsychologist.com/d3/cohend/>

- [2] Nielsen, J.: Usability Engineering. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1993)
- [3] Branaghan, R.J., Sanchez, C.A.: Feedback preferences and impressions of waiting. *Human Factors: The Journal of the Human Factors and Ergonomics Society* **51**(4) (2009) 528–538
- [4] Carlson, J.: Apple watch. (2015)
- [5] Prewett, M.S., Elliott, L.R., Walvoord, A.G., Coover, M.D.: A meta-analysis of vibrotactile and visual information displays for improving task performance. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* (2012)
- [6] Shneiderman, B.: Response time and display rate in human performance with computers. *ACM Computing Surveys (CSUR)* **16**(3) (1984) 265–285
- [7] Miller, R.B.: Response time in man-computer conversational transactions. In: *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, ACM (1968) 267–277
- [8] Nah, F.: A study on tolerable waiting time: how long are web users willing to wait. (2004)
- [9] Hornik, J., Zakay, D.: Psychological time: The case of time and consumer behaviour. *Time & Society* **5**(3) (1996) 385–397
- [10] Hoggan, E., Brewster, S.A., Johnston, J.: Investigating the effectiveness of tactile feedback for mobile touchscreens. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (2008) 1573–1582
- [11] Poupyrev, I., Maruyama, S., Rekimoto, J.: Ambient touch: designing tactile interfaces for handheld devices. In: *Proceedings of the 15th annual ACM symposium on User interface software and technology*, ACM (2002) 51–60
- [12] Osborne, J.W., Overbay, A.: The power of outliers (and why researchers should always check for them)
- [13] Cohen, J.: Statistical power analysis. *Current directions in psychological science* **1**(3) (1992) 98–101



# Haskell vs. JavaScript for Game Programming

Christian Fischer

Department of Computing Science  
Umeå University, Sweden  
`christian@chfi.se`

**Abstract.** The aim of this paper is to evaluate which paradigm is better suited for programming video games: Procedural programming, which is most commonly used, or functional programming.

To answer this question a simple video game is implemented in each language, and the resulting code bases are evaluated quantitatively by counting the number of identifiers in both codebases, as well as the sizes of the implementations in lines of code and file sizes. The Haskell implementation was considerably bigger than the JavaScript one, but it also used more library functions.

No concrete conclusions are made, and suggestions for future research include using other languages and minimizing differences in libraries used by the implementations.

## 1 Introduction

Many video games are written using C++<sup>1</sup>, C#<sup>2</sup>, Java<sup>3</sup>, or JavaScript<sup>4</sup> (JS), depending on the platform (e.g. web games generally require different tools than console games) and size of the game. All these languages are procedural and object-oriented [1][2][3][4]. Haskell [5] is a purely functional programming language, a type of language that has seen very little use in this domain.

The purpose of this paper is to explore whether Haskell, and the functional paradigm, may be a better language and tool for game programming than more commonly used tools, i.e. procedural and object-oriented languages. By “good” we mean “high-level”, as defined by Alan Perlis [6]: “A programming language is low level when its programs require attention to the irrelevant.”

According to this definition, a high-level language (or library) should allow the developer to focus more on what the game should be like as seen and experienced by the player, rather than describe what should happen “below the hood”, such as memory being allocated, ensuring objects are moved at the correct speed with respect to current framerate, etc. Instead those things should be handled, automatically, by the language or library.

---

<sup>1</sup> E.g. Unreal Engine 4-based games <https://www.unrealengine.com/unreal-engine-4>

<sup>2</sup> E.g. Unity-based games <https://unity3d.com/unity>

<sup>3</sup> E.g. LWJGL-based games, such as Minecraft <https://www.lwjgll.org/>

<sup>4</sup> Web-based games, Unity-based games

Historically, it has been difficult to use a purely functional language such as Haskell for interactive applications such as video games, as, by definition, a pure language cannot actually perform input and output, as IO is a side-effect, something not allowed in purely functional languages! Haskell solved the problem of IO using monads[7]. With the introduction of a programming paradigm known as Functional Reactive Programming (FRP), implemented in several libraries in Haskell, writing real-time interactive applications has become even better, by making time-flow explicit [8].

Haskell and JS differ enormously, both in syntax and semantics. Thus programs written in either language require very different code design approaches<sup>5</sup>. While most games have been written in procedural languages such as JS, it is possible that a vastly different language, such as Haskell, may prove to be more appropriate to the problem domain, or at least that some parts of the problem may be better solved using another approach. As games continue to grow in complexity, it becomes more and more important to use appropriate tools to build them, as that may save both time and money.

To test the hypothesis that functional programming and FRP may be more appropriate for game programming than procedural programming, a simple action game was implemented twice, once in Haskell and once in JS. The implementations were functionally equivalent. The differences in code between implementations were evaluated using several quantitative measures; the use of different types of identifiers (names of functions and variables), and code size in lines of code and file sizes. These measures were used to examine which language of the two is more suited to game programming, or if there are any specific parts of game programming that one language excels at.

JS was used as the procedural language, as it is a very commonly used language, and one in which it is easy to develop interactive graphical applications — e.g. most interactive web pages on the Internet. While C or C++ would have been a more idiomatic choice given the area, the author has relatively little experience in those language, especially in this area; thus JS was a natural choice. JS is a multi-paradigm language, and indeed supports functional programming concepts such as first-class functions, however those features were deemphasized, and greater focus placed on the procedural features of the language.

## 1.1 Preliminaries and Background

While games have been written in Haskell and other functional languages, there has been little work in comparing the use of procedural languages to that of functional ones, when it comes to writing video games. There is also little work in comparing statically typed languages to dynamic languages.

The biggest difference between functional and procedural languages is that programming in functional languages has the programmer describe what a pro-

<sup>5</sup> As can be seen in this side-by-side comparison of a parser implementation:

<http://infrared-clearance.weebly.com/top-down-operator-precedence---in-haskell.html>

program *is*, while procedural languages have the programmer describe what a program *does*. This difference makes it easier to implement many higher-level algorithms in functional languages, while it has been the case that procedural languages generally execute faster, and allow for more detailed control over the performance of applications written in them.

The video games industry largely uses procedural languages<sup>6</sup>, as can be seen with Unreal Engine 4 and Unity, two of the most widely used game engines as of writing, and can be developed for using C++ and C#, respectively.

Video games also make large use of mutable state, i.e. video game can be defined as a loop [9], where player input goes in, updates the current game state, and the state is displayed to the player. When the game state is large, and the game loop needs to run as quickly as possible using as little space as possible, is where low-level languages have shone, as languages with manual memory management allow for direct control over memory usage, and thus potentially more efficient memory usage, than e.g. languages with garbage collection. As available memory increased, garbage collected languages started to find a place in game programming, e.g. Java. As compilers become better at optimizing, the performance differences between languages become smaller<sup>7</sup>, and as computers become faster, the impact of performance differences shrink.

Haskell is a purely functional language, meaning that functions can only transform data given as inputs, and there is no implicit state, nor can any side effects be produced. As a consequence of this, Haskell has referential transparency, which means that if a function is given some  $X$  as input and returns  $Y$  as output, it *always* returns  $Y$ , and *does nothing else*. In procedural languages such as C or JS, this is not necessarily the case. For example in C, the input to a function is often a pointer, and the function changes the memory pointed to by that pointer, while returning a status code signaling whether the function succeeded or not. Indeed, a function in C can theoretically write to arbitrary places in memory (excepting operating system-level security) — not so in Haskell.

Objects in JS are basically dictionaries of strings to values, i.e. key-value maps. These allow for dynamic addition and removal of properties of objects. On the other hand Algebraic Data Types (ADTs) are types that are constructed by combining existing types in various ways. ADTs make it easy to create representations of data on the type-level, and thus help ensure program correctness, at the cost of the ease of expressiveness that JS-style objects allow.

Another aspect of the functional purity of Haskell is that it is immutable by default, meaning variables cannot change throughout the running of the program. Instead of changing the elements of a list by applying a function to its elements, one creates a new list consisting of the elements resulting from calling the function on the elements — that is, the old list remains unaltered.

---

<sup>6</sup> Though exceptions exist, e.g. <http://all-things-andy-gavin.com/2011/02/04/making-crash-bandicoot-part-3/>

<sup>7</sup> As shown in informal benchmarks such as <http://lambda.jstolarek.com/2013/04/haskell-as-fast-as-c-a-case-study/>

Haskell is also a “lazy” language. This means that values are only evaluated as needed, rather than as soon as they are encountered. For example, the following snippet of Haskell code returns the list containing the first 50 positive integers, even though the `[1..]` syntax generates an infinite list containing *all* positive integers. The function `take n l` returns a list containing the first  $n$  elements from a list  $l$ ; no more than the first  $n$  elements of  $l$  need to be evaluated. Laziness makes it easy to compose functions, increasing opportunities for code reuse [10].

```
first50 :: [Integer]
first50 = take 50 [1..]
```

Functional Reactive Programming (FRP) is a programming paradigm that makes time-flow explicit, by providing the programmer with “behaviors” and “events”. Behaviors are time-varying values, to which functions can be applied to to get new behaviors. E.g. the position of the mouse cursor on screen can be described by a behavior containing a pair of numbers, *mouse* :: *Behavior*( $x, y$ ). We can apply any function that works on pairs of numbers to *mouse*, and get a new behavior. “Events” are values that happen at a point in time, e.g. the presses of a button. Functions can be applied to events as well; one could apply a function to transform a stream of keystrokes to a stream of numbers.

FRP makes it possible to describe interactive applications by composing smaller functions [11]. For example, the position of something controlled by the arrow keys can be described by mapping a function that takes keystrokes to numbers — the velocity of the object — and then taking the integral of that event stream.

FRP is a programming paradigm on its own, and has been implemented in Haskell as libraries — it is not a feature inherent to the language, though Haskell supports tools and abstractions that make it possible to express the FRP paradigm quite naturally in the language [11].

Haskell and JavaScript are two very different programming languages, and require different approaches to program design. They do not even share any superficial features such as syntax; these differences will likely manifest themselves throughout the implementation process of any application.

Specifically to implementations of a game, Haskell’s type system and ADTs leads to a different representation of game state, compared to objects used in JS.

Referential transparency, and functional purity more generally, likely to leads to more code reuse in the Haskell version [10]. On the other hand, those features make it more difficult to represent changing state, as well as input/output — both of which are vital in a video game. JS has no such problems.

## 2 Method

To evaluate the differences of Haskell and JavaScript with respect to implementing a video game, a clone of the game Curve Fever<sup>8</sup> was written in each language. The game implementations were functionally identical.

The evaluation was done using several quantitative and qualitative measures, similar to those used in [12], see below. Quantitatively, the number of unique functions were counted, as well as code size.

Curve Fever is a fairly simple game, yet complex enough to highlight various differences between the languages.

### 2.1 The game

Figure 1 shows a typical screenshot from a game of Curve Fever.

The game consists of three player controlled curves. A curve has a head and tail; the head is that which is directly controlled by the player, and is what collisions are checked against. The tail of the curve is static, once placed it does not change, and all tails behave identically.

Each player's curve has a unique color. All players move with the same speed and turn with the same speed/radius.

The game board is two dimensional and of finite size. If a player's curve runs into a wall, the player loses.

### 2.2 Haskell

The Haskell implementation will use the Simple DirectMedia Layer (SDL)<sup>9</sup> library for input and output, which is commonly used in game programming in many languages. The FRP implementation used is the Netwire<sup>10</sup> library.

### 2.3 Javascript

The Javascript version will use no additional libraries. A HTML5 canvas will be used for graphics output, and the Javascript/HTML5 built-in keyboard event system will be used for input.

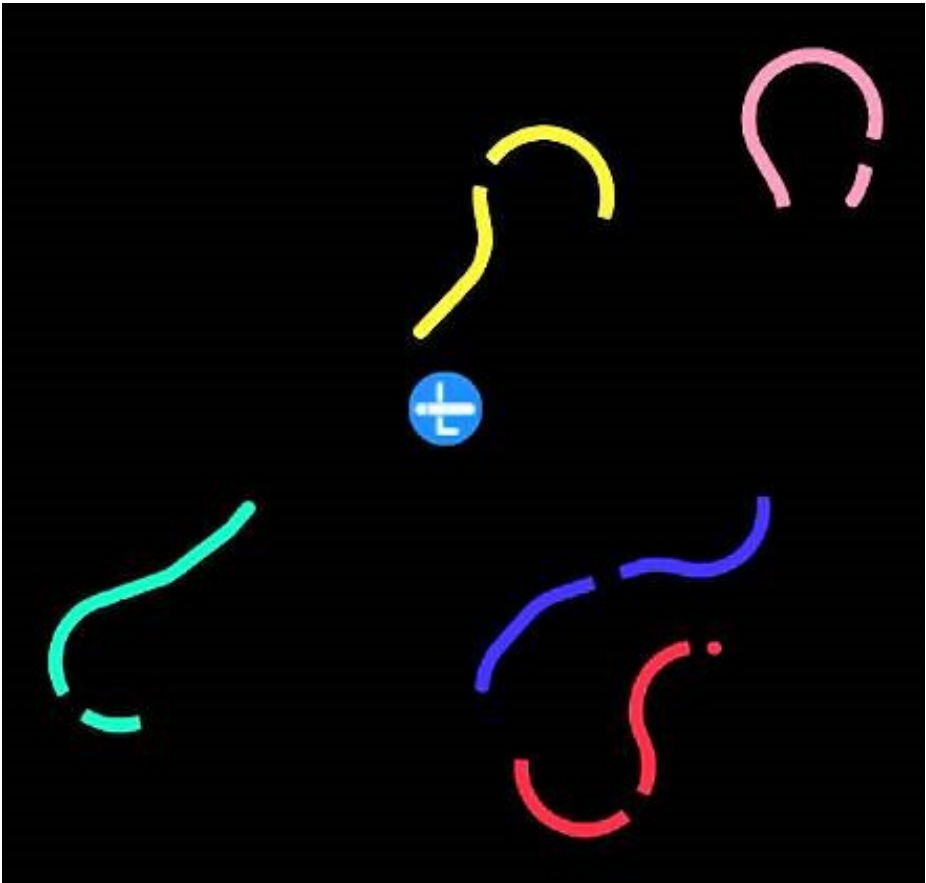
### 2.4 Evaluation

The two implementations were evaluated by analyzing the two codebases. The number of lines of source code are counted, as a general measure of code efficiency and expressiveness, to evaluate which language is better suited for game programming. Similar metrics have been used in previous comparisons [12].

<sup>8</sup> <http://curvefever.com/play.html>

<sup>9</sup> <https://www.libsdl.org>

<sup>10</sup> <https://hackage.haskell.org/package/netwire>



**Fig. 1.** Screenshot of Curve Fever

By “better suited” we mean, as mentioned in the Introduction, that a language which requires less attention to things irrelevant to the game is better than one that requires more attention in that area. Thus, performance is disregarded, and any memory-managed language is “better” than any language with manual memory management.

Using this definition of a “good” language, we can define some quantitative measures. If a language has many library functions which can be used by the developer, then it is better than one that requires the developer to manually write those functions, as that requires the developer to pay attention to something “irrelevant”. Similarly, code reuse can be considered good, as that implies that the developer only has to solve a general problem once, and then can focus on more important things.

More generally, each unique variable or identifier could be considered as a “unit of attention”, and it could thus be argued that more variables are worse, by this metric, than fewer. Similarly, a larger code base (counted in lines of code) can be taken to assume that more attention was required.

By counting the number of defined functions, and use of library and built-in functions, the level of code reuse was estimated. A high ratio of use of library functions to use custom functions implies that the libraries used contained many tools that could be used by the application, i.e. code that could be reused without modification.

The Haskell code was analyzed using the Haskell Source Extensions<sup>11</sup>, a Haskell parser that can be used from Haskell itself. The JS code was analyzed using Esprima<sup>12</sup>, a JavaScript parser meant for code analysis.

The analysis consisted of recursing on the right hand side of each declaration in the code, e.g. the bodies of functions, or values of variables, until an identifier, a name, was reached. Each identifier was added to a key-value map, with the identifier as key and number of times it occurred in the code as the value. The resulting lists of identifiers were then manually analyzed and put into five categories, as seen in the next section.

## 3 Results

Quantitative results are presented, followed by some implementation details. The code for the Haskell implementation can be found at <https://github.com/chfi/sc-haskell>, that for the JS version at <https://github.com/chfi/sc-js>.

### 3.1 Haskell

The Haskell implementation consists of 7 modules, with a total of 339 lines of source code (excluding blank lines and comments). The identifiers were placed into six categories: Var, Const, Func, Lib, and Type.

<sup>11</sup> <https://github.com/haskell-suite/haskell-src-exts>

<sup>12</sup> <https://github.com/jquery/esprima>

Var included all temporary variables as well as arguments to functions. Const included all custom written constants, such as screen size. Func consisted of all custom written functions. Lib contained all functions built in to Haskell or imported from SDL or any other library. Type contained the uses of custom type constructors.

Table 1 shows the distribution of the various identifier types.

Identifier Type	Total	Unique
Var	140	47
Const	8	8
Func	42	29
Lib	184	98
Type	15	5
Total	389	187

**Table 1.** Haskell identifier categories

The Haskell implementation is 12937 bytes large, where 6.1 KB are directly related to the game loop and curves, and the rest are boilerplate related to the input/output libraries used.

### 3.2 JavaScript

The JS implementation consists of a total of 140 source lines of code, and 4677 bytes, of which the vast majority are game specific, with very little boilerplate.

The identifiers in the code were placed into five categories: Var, Lib, Const, Func, and Class. All categories except Class were the same as those for the Haskell implementation, with Class containing custom written classes, being analogous to Type. The Lit category is not represented due to literals not being identified. Table 2 shows the total counts of each category of identifier.

Identifier type	Total	Unique
Var	232	34
Lib	57	29
Const	19	5
Func	31	12
Class	15	4
Total	360	87

**Table 2.** JavaScript identifiers

## 4 Discussion

The Haskell implementation consists of more than double the number of lines of code, and ended up three times larger than the JS implementation in terms of



file size. Much of this was due to the additional code required to do relatively simple things such as drawing to the screen and interpreting user input, as well as the Haskell version requiring the import of libraries.

The Haskell implementation also used many more library functions. While several were used in the SDL boilerplate, most were used in the game and curve logic itself. This fits with the expectation that functional programming encourages code reuse [10].

The fact that the Haskell version has many more unique identifiers is, again, likely due to use of higher-order functions and reuse of code. It is also true that the Haskell standard library is quite large compared to the JS one.

The larger use of local variables in the JS version may be due to the use of so-called “point-free style” in the Haskell version, which means that functions are written by composing other functions, without explicitly naming their arguments. However, it could also be due to Haskell discouraging the use of local state. Finally, it could be an artifact in the tools used to analyze the two codebases.

Neither version required any data structure more complex than records and lists, so not much can be said about the appropriateness for either language in describing the data used in games.

In short, while the Haskell codebase is larger in size, the higher level of code reusability and use of library functions could be seen as evidence that Haskell is better suited for game programming than JavaScript. The lower count of variables support this claim, as well. Even so, there is not enough evidence to make any definite conclusions; these are only two codebases, written by a single programmer, and bias is unavoidable.

## 4.1 Future work

Using libraries with a more similar API would decrease the variation due to boilerplate, and lead to clearer results. Using a language such as PureScript, which is similar to Haskell but compiles to JavaScript, would decrease these differences and make it easier to focus on the higher level semantics of the languages.

A more complex game, making use of more complicated data structures, could be the subject of future research. That may make it possible to take a more concrete stance on which language or paradigm is more suitable for describing the structure of data used by video games.

Another approach would be to let many programmers repeat this or a similar task, and so hopefully reduce both experimenter bias as well as minimize noise due to earlier programmer experience and approaches to problem solving.

The most striking difference between the identifier tables (Tables 1 and 2) is the much larger use of library functions in the Haskell implementation, and the much larger use of local variables in the JavaScript code. It is also worth pointing out that the JS implementation uses nearly as many identifiers as the Haskell one, despite being much smaller in terms of code size. If the JS codebase was bigger, it is likely that it would use even more identifiers than the Haskell version.

It could be argued that this use of more local variables makes it more difficult to jump into functions and understand what their purpose is, as a programmer.

## 5 Conclusion

It is difficult to draw any concrete conclusions. The implementation tools and libraries were very different, with the boilerplate-heavy low-level SDL library, and large standard library, on the Haskell side, and the small and easy-to-use HTML5 canvas API, and minimal standard library, on the JavaScript side. These differences introduce a large amount of noise to the data. However, it seems possible that Haskell, and functional programming in general, may deserve a better position in the pantheon of languages used for game programming, than they currently have.

## References

- [1] Standard C++ Foundation: C++ big picture issues. <https://isocpp.org/wiki/faq/big-picture> (2017) Accessed: 2017-01-03.
- [2] Hejlsberg, A., Wiltamuth, S., Golde, P.: C# Language Specification. Addison-Wesley Longman Publishing Co., Inc. (2003)
- [3] Gosling, J., Joy, B., Steele, G., Bracha, G., Buckley, A.: The Java® Language Specification. Oracle America (2014)
- [4] European Computer Manufacturers Association: EcmaScript 2016 language specification (2016)
- [5] Hudak, P., Hughes, J., Peyton Jones, S., Wadler, P.: A history of haskell: being lazy with class, ACM 12–1
- [6] Perlis, A.J.: Epigrams on programming. SIGPLAN Notices **17**(9) (1982) 7–13
- [7] Jones, S.P.: Tackling the awkward squad. In: Engineering theories of software construction, Press (2001) 47–96
- [8] Courtney, A., Nilsson, H., Peterson, J.: The yampa arcade. Haskell '03, ACM 7–18
- [9] Nystrom, R.: Game Programming Patterns. 1st edn. Genever Benning
- [10] Hughes, J.: Why functional programming matters. The computer journal **32**(2) 98–107
- [11] Hudak, P., Courtney, A., Nilsson, H., Peterson, J.: Arrows, robots, and functional reactive programming. Springer 159–187
- [12] Harrison, R., Smaraweera, L., Dobie, M., Lewis, P.: Comparing programming paradigms: an evaluation of functional and object-oriented programs. Software Engineering Journal **11**(4) 247–254

# Usability of mobile interfaces with regards to left-handed use

Filip Norman Golles

Department of Computing Science  
Umeå University, Sweden  
`id12fgn@cs.umu.se`

**Abstract.** User experience and interaction design for mobile user interfaces have come a long way in the last few years. However, not a lot of thought has been given to the different preferences and ability of the individual user. A majority of people are right-handed, and prefers to use mobile devices with their right hand, and mobile interfaces are designed to cater to this majority. The aim of this study was to investigate how efficient and comfortable commonly used design elements are when used with the left hand. User tests were conducted and the results of the participants who preferred their left hand were compared to a right-handed control group. The participants own opinion on comfort and usability of the user interfaces was collected and analysed in addition to the data collected in the user tests. In this study, there was no clear correlation between the users preferred hand and the efficiency of usage along the horizontal plane of the mobile device. There was however a significant difference in performance between the top part and the bottom part of the device, showing greater efficiency towards the bottom. The study also shows that efficiency and user experience does not necessarily coincide, with most users not preferring the most efficient way of interaction, but seemed to pick their favorite based on other factors.

## 1 Introduction

During the last decade, technology has become something more than just pure functionality. Designers and consumers alike have noticed the importance of technology not only aiding the user in solving whatever tasks that might need solving, but also the efficiency and comfort when using the technology itself [1]. A lot of improvement has been made in the area of mobile interface design to ensure that people with different experiences, preferences and ability can fully utilize the functionality of the interface in the most comfortable and efficient way possible. However, not a lot of thought has been given to how the usage differs between left-handed and right-handed people. The vast majority of users and developers are right-handed (studies suggest up to nine in ten people are right-handed [2]), and/or prefers interacting with the right hand, and this reflects in the user interfaces of mobile applications. This is very noticeable in menus and other navigation. The popular slide-out menu is placed in the top-left corner

in a vast majority of modern mobile applications, without any regard to the preferred hand of the user. It is also noticeable during testing of applications during development. Handedness of the users is often not investigated at all, or simply disregarded as outliers (due to the relatively low number of left-handed people). The aim of this study is to test and evaluate common mobile user interface (UI) features with regards to left-handed use.

## 1.1 Earlier work

There is quite a bit of research done in the area of user experience (UX) and UI design, and a lot of it is focused on adapting UIs to all kinds of people and preferences. However, to the authors knowledge, no research has been done regarding the difference between right- and left-handed use.

UX has been a subject of discussion for quite some time. Only in recent years have the scientific community started to agree on a more strict definition [3] of the area, and how it should be evaluated in an empirical manner [4]. The authors conclude that several factors need to be taken into account when evaluating UX, and both recorded tests and interviews or surveys must be conducted in order to get as accurate a picture as possible.

There is some research and proof that handedness has an effect on the use of touch screen interfaces [5]. However, this research is focused on much larger displays, and does not take neither the navigation aspect of the interface, nor one-handed usage into consideration. Research also shows that handedness has an effect on which design patterns (DPs) were more efficient and preferred by the users in the context of mobile learning [6]. While this research does not focus on mobile learning, DPs are used in UI design, and have a large impact on the overall structure of the application and its navigation.

While not specific to mobile interfaces, it is shown that the identification of objects differ between right- and left-handed people [7]. This study found that the mental representations involved when drawing and identifying objects contained a sense of direction which influenced how well participants performed in the tests, especially concerning objects with a clearly defined front and back, such as a car or an elephant. The direction participants preferred was related to their handedness, their non-dominant hand being the "forward" direction. For example, right-handed people would draw an elephant with the trunk to the left and the tail to the right.

## 2 Method

In order to conduct this research, a specialized but simple application was developed, containing the design elements in focus, and the placement of these in the overall interface. No functionality was implemented, as it might draw the users away from the relevant design elements. The design elements in focus were different types of menus, in particular the slide-out menu ("hamburger menu"), slide-down/slide-up menu, as well as a circular menu("carousel menu"). Menus

and navigation are the most interesting parts of a user interface in this regard, since it is where most of the active interaction takes place.

The overall aim of the study is to pinpoint which of the navigation options described above are more efficient when used with the left hand and right hand respectively. To determine the efficiency, the users hand and finger movements were analyzed. When navigating directly on a touch screen, parts of the screen will unavoidably be covered by the users hand and fingers. The size and position of the covered area were considered in the analysis, as well as the speed and accuracy of the users during the test.

Both left- and right-handed users were needed in order to be able to come to a conclusion. While subjective, it is possible to reliably evaluate UX [4]. Since this research focuses on the more measurable parts of UX, rather than the individual feelings of the user, video taped observations were used to come to a conclusion. However, a small post-test survey was conducted in order to get the users own input on the usability and the overall comfortability of the navigation. Additionally, the tests will contain tasks engineered to produce measurable results, specifically the time required to complete a task, and the number of user errors. The tests were carried out on an iPhone 6/6S (4.7" screen).

An application (Fig. 1) was developed for the purpose of testing the relevant aspects of UX described above: speed, user errors, as well as the area of the screen covered by the users hand while performing the tasks. Each of the navigation options in question were placed on different "pages" of the application, to avoid any unwanted interaction between the design elements themselves. No functionality outside of the menus was added, in order to cause as little distraction as possible for the users. The different ways of navigation, with exception for the circular menu, are implemented in different positions of the screen: the slide-out menu in the top-right, top-left, bottom-right and bottom-left corners, and the slide-down/up menu in the top-center and bottom-center positions. The circular menu is always placed in the top-center position.

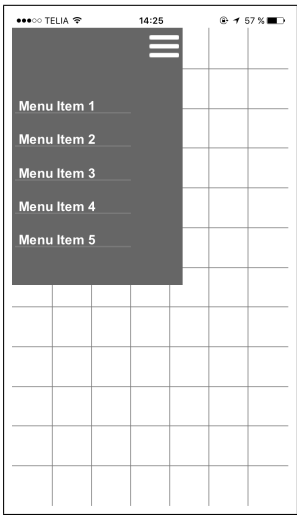
### **3 Test Procedure**

#### **3.1 Participants**

Eleven test users were divided into two groups: those who preferred to use their mobile devices with the left hand, and those who preferred the right hand. The two groups consisted of six and five users respectively. All of the users claimed to have a lot of experience with mobile devices and were in roughly the same age group, ages twenty to thirty. No other variables were considered when dividing the users into groups.

#### **3.2 User Tests**

During the tests the users performed a simple task, opening the menu and selecting a specific menu item, for each of the different menu configurations with



**Fig. 1.** The application. This image shows the "hamburger" menu opened in the top-left position.

their preferred hand. The tests were video taped in their entirety for later analysis. Participants were also required to answer a quick survey about which of the positions they preferred for the slide-out menu and the slide-down/up menu, and which menu they preferred overall, regardless of position.

**3.3 Data Analysis**

The video taped user tests were analysed, the tasks performed in the tests were timed, and any user errors made were noted. All individual test results were rounded to the nearest whole second. The average time for each of the menu/position configurations were taken for each of the test groups, as well as the number of errors made. The survey results were paired with the individual users test results.

**4 Results**

The user tests did show some differences between the two groups. However, due to the small population of users, the numerical results of the actual user tests can be argued to not carry much significance. The post-test survey did however produce more interesting results, especially when paired with the results of the tests.

During the slide-out ("hamburger") menu section of the test, neither group made any user errors, and their speed was similar across all the different screen positions (Table 1). Both groups performed the fastest in the bottom-right position of the screen, with an average of 2.2 and 2.3 seconds for the right-handed group (RG) and the left-handed groups (LG) respectively. Both groups produced

the slowest results in the top-left and top-right positions, with an average of 2.8 seconds in both positions for LG, while RG had a much larger variance with an average of 5.6 and 3.6 seconds for the top-left and top-right positions respectively. The post-test survey did show some differences between the groups. Three out of five right-handers prefer the top-right positions, while the left-handers prefer either the bottom-right or bottom-left positions (two out of six in both cases).

Group	Top-left	Top-right	Bottom-left	Bottom-right
Left-handed group	2.8s	2.8s	2.6s	2.3s
Right-handed group	5.6s	3.6s	2.8s	2.2s

**Table 1.** The average time spent on the tasks in the slide-out ("hamburger") menu section of the user tests.

For the slide-down/up menu, the two groups performed in a similar way (Table 2). Both groups had the fastest average times in the bottom position, 2.6 seconds for the LG, and 1.6 seconds for the RG. The average times for the top position was 3.2 and 3.6 for the LG and RG respectively. As with the previous section of the test, the RG produced a much larger variance in their results. The survey showed that the bottom position was preferred by most of the participants, with four out of six in the LG and three out of five in the RG preferring the bottom position.

Group	Top Bottom	
	Top	Bottom
Left-handed group	3.2s	2.6s
Right-handed group	3.6s	1.6s

**Table 2.** The average time spent on the tasks in the slide-down/up menu section of the user tests.

Both groups performed the slowest in the test with the carousel menu. The average time for the groups was 6.0 seconds for the LG, and 4.8 seconds for the RG. There were a couple of user errors made by participants in the LG. Removing those outliers brings the average time to 4.0 for the LG.

When pairing the individual test results (Table 3) with the corresponding survey results, we see that efficiency (speed) does not necessarily coincide with the users preferred menu/position configuration.

Additionally, the survey asked the participants which menu they preferred overall, regardless of position. Four out of six participants in the LG preferred the slide-out ("hamburger") menu, and the remaining two preferred the slide-down/up menu. All participants in the RG preferred the slide-out ("hamburger") menu.

Comparing the results of the RG and the LG (Table 3) for each of the test cases using the student's t-test (Table 4) shows that five out of the seven menu/position configurations carry no statistical significance between the two

Preferred hand	H-TL	H-TR	H-BL	H-BR	S-T	S-B	C
Left	3s	3s	2s	<b>2s</b>	4s	<b>3s</b>	4s
Left	<b>4s</b>	4s	3s	4s	<b>4s</b>	2s	5s
Right	4s	<b>3s</b>	4s	2s	<b>5s</b>	2s	4s
Left	2s	<b>2s</b>	2s	1s	2s	<b>2s</b>	4s
Right	7s	2s	2s	<b>2s</b>	2s	<b>2s</b>	8s
Right	8s	<b>5s</b>	3s	3s	7s	<b>2s</b>	5s
Right	2s	2s	<b>3s</b>	2s	2s	<b>1s</b>	3s
Left	2s	3s	<b>2s</b>	3s	3s	<b>4s</b>	3s
Left	3s	3s	4s	<b>3s</b>	<b>3s</b>	3s	10s
Left	3s	2s	<b>3s</b>	1s	3s	<b>2s</b>	10s
Right	7s	<b>6s</b>	2s	2s	<b>2s</b>	1s	4s

**Table 3.** Individual test results for the slide-out menu (H), the slide-down/up menu (S) and the carousel menu (C) for the positions top-left (TL), top-right (TR), bottom-left (BL), bottom-right (BR), top-center (T), and bottom-center (B). Times in **bold** are the users preferred position for the given menu, H and S.

groups. The configurations that did show significance were the "hamburger"-menu in the top-left position, and the slide-up menu.

Menu/Position	P-Value	Significant
H-TL	0.0293	<b>Yes</b>
H-TR	0.3676	<b>No</b>
H-BL	0.7957	<b>No</b>
H-BR	0.8220	<b>No</b>
S-T	0.6718	<b>No</b>
S-B	0.0349	<b>Yes</b>
C	0.4790	<b>No</b>

**Table 4.** T-test results for the slide-out menu (H), the slide-down/up menu (S) and the carousel menu (C) for the positions top-left (TL), top-right (TR), bottom-left (BL), bottom-right (BR), top-center (T), and bottom-center (B). P-values lower than 0.05 (95% confidence interval) shows statistical significance.

## 5 Conclusion

Looking at the data collected during the user tests, there are a few things we can conclude. Both groups had the fastest speeds in the bottom positions of both the "hamburger"-menu and the slide-down/up menu, likely because of the shorter distance to the hand of the user. Regarding the horizontal axis, left to right, there is no convincing evidence that the users preferred hand has any significant impact on the preferred positioning of the UI elements.

The results of the t-tests showed that only two out of seven menu/position configurations carried statistical significance between the two groups. Considering the low number of participants, six in LG and five in RG, and the fact



that most of the tests showed no significance, we cannot draw any conclusions regarding the efficiency of the groups based on numerical results alone.

Looking at the individual test results and pairing them with the answers to the survey, we can conclude that there is no obvious correlation between the efficiency of the menu and position configuration and the users preferred configuration. This result suggests that a users perceived efficiency and comfort might not have that much to do with the actual speed and efficiency of the interaction, but rather of a variety of other factors in combination with efficiency.

Both groups performed the slowest in the carousel menu section. It is also the only menu where participants felt confused and made any user errors. The author concludes this is in part because of the placement of the menu. The carousel was always placed in the top-center position of the screen and, as proved by the other sections of the test, usage in the top part of the screen was not as efficient as the other parts of the screen. Another factor for the poor performance might be the way the specific menu functions. It is drastically different from the other more "classic" menus, and this may likely have lead to confusion in some of the participants.

## **6 Discussion**

The aim of the study was to evaluate common mobile UI components with regards to left-handed use. During the user tests no significant correlation between handedness and efficiency (speed and number of user errors) were found. However, if the post-test survey is weighed in with the numerical results, there were some more interesting results.

The study showed that designing to satisfy both left- and right-handed people might not be such a big issue, at least not one that requires explicit thought and effort during the design process. During the user tests, when considering the users own opinions on which navigation option and position they preferred, it became clear that efficiency might not have as big of an impact on UX and overall usability of an application. A majority of test users did not prefer the configuration in which they performed most efficiently. It is likely that a users preferred configuration is more influenced by a combination of factors, such as familiarity, aesthetics, context and efficiency, rather than by efficiency alone. This does however require further research in order to come to a conclusion.

For any future research, other variables other than just the preferred hand of the user should be considered. The findings in this study suggests that handedness, if at all relevant, is only a small part of a users perceived experience when using software on mobile devices. Familiarity and context should be investigated further. Aesthetics may also influence a users perceived experience [8]. This study only separated the participants by their preferred hand. All participants were in the same age group, and claimed that they had a lot of experience in using mobile devices. It is very likely that there are several other variables that must be considered in order to get a more complete picture of the problem, such as hand

size, size of the device/screen, experience with mobile devices/smart phones, and age.

## References

- [1] Hassenzahl, M., Tractinsky, N.: User experience-a research agenda. *Behaviour & information technology* **25**(2) (2006) 91–97
- [2] Hardyck, C., Petrinovich, L.F.: Left-handedness. *Psychological bulletin* **84**(3) (1977) 385
- [3] Law, E.L.C., Roto, V., Hassenzahl, M., Vermeeren, A.P., Kort, J.: Understanding, scoping and defining user experience: a survey approach. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2009) 719–728
- [4] Arhippainen, L., Tähti, M.: Empirical evaluation of user experience in two adaptive mobile application prototypes. In: *Proceedings of the 2nd International Conference on Mobile and Ubiquitous Multimedia*. Number 11, Linköping University Electronic Press (MUM 2003) 27–34
- [5] Kang, H.Y., Shin, G.: Effects of user handedness on hand usage pattern when using a desktop touchscreen display. *Ergonomics Society Conference 2013* (2013) 231–234
- [6] Al-Samarraie, H., Ahmad, Y.: Use of design patterns according to hand dominance in a mobile user interface. *Journal of Educational Computing Research* (2016) 769–793
- [7] Viggiano, M.P., Vannucci, M.: Drawing and identifying objects in relation to semantic category and handedness. *Neuropsychologia* **40**(8) (2002) 1482–1487
- [8] Tractinsky, N., Katz, A.S., Ikar, D.: What is beautiful is usable. *Interacting with computers* **13**(2) (2000) 127–145

# Replacing words with emojis and its effect on reading time

Viktor Gustafsson

Department of Computing Science  
Umeå University, Sweden  
`id12vgn@cs.umu.se`

**Abstract.** Emojis has grown in popularity and recent versions of operating systems has introduced new ways of using them. One behavior that has emerged is the use of emojis as a replacement of words. In this study we investigate how replacing words with emojis effect reading time. This paper studies this by conducting tests where users read texts with and without emojis while we record the time it took for the subjects to read the texts. We did a statistical analysis of the collected reading time data and found that it takes significantly longer to read a text with some words replaced by emojis than plain text. From the result we make the recommendation that emojis should be used to enhance tone or emotion to texts but not to replace words.

## 1 Introduction

Emojis has recently grown in popularity<sup>1</sup>, invaded pop culture and has become a worldwide phenomena. Emojis are small pictograms or image versions of emoticons that could represent facial expressions, objects, situations and more. Emojis are often used in text based communication to enhance the meaning of a message, add emotion [1] to a sentence or to replace words [2]. Not to be confused with emoticon which is older and a typographic display that by combining keyboard characters could represent different facial expressions. The most widely adopted standard of emojis are one governed by the Unicode Consortium<sup>2</sup> who decides what emojis should be part of the standard and creates design guidelines of how the emojis should look.

The recent rise in popularity of emojis are thanks to Apple, Microsoft, Google and other software vendors adopting the unicode standard of emojis into their operating systems (iOS, macOS, Windows and Android), so they are now a commodity in text based communication, social media and marketing campaigns. Also new features with iOS 10 make it even easier to use emojis. One of the new features introduced was that autocorrect now suggest emojis together with the suggested words.

---

<sup>1</sup> Dimson, T. (2015). *Emojineering Part 1: Machine Learning for Emoji Trends*. Instagram Engineering Blog <http://instagram-engineering.tumblr.com/post/117889701472/emojineering-part-1-machine-learning-for-emoji>

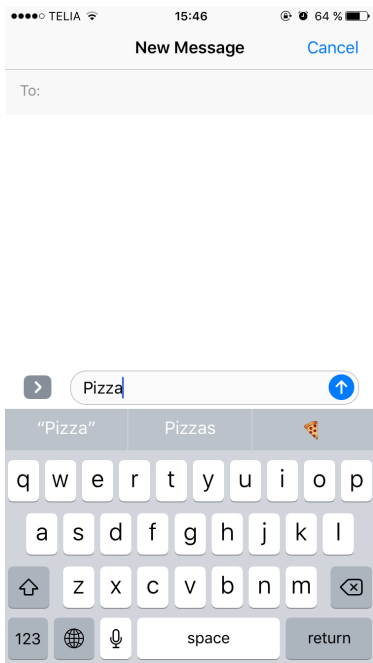
<sup>2</sup> Unicode Emoji <http://www.unicode.org/emoji/>

With the rise in popularity and easier access to use emojis we think it is important to investigate their affect on reading times. Therefore this study aims to answer the following question: *Does replacing words with emojis affect the reading time of texts?*

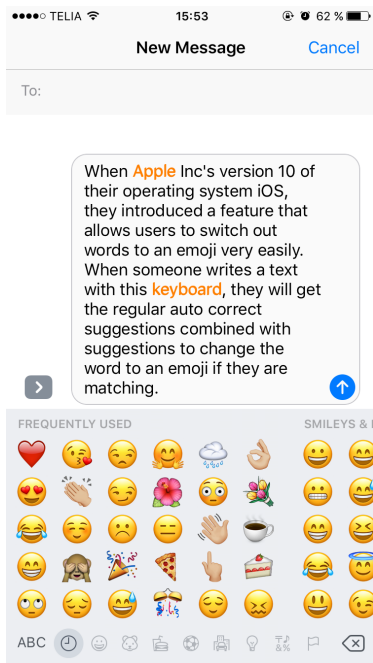
2 Background

In Apple Inc’s version 10 of their mobile operating system iOS, they introduced features that will make it easier to replace words with emojis (see Figure 2 and 2). One of the new features is that the auto correct functionality of the built-in keyboard in this version suggests emojis when writing words that correspond to that emoji (see Figure 2). Another feature is that words available to be replaced by emojis are highlighted so that tapping on that word replaces the word to its corresponding emoji (see Figure 2).

When a text message was written with the built-in keyboard, they got the regular auto correct suggestions combined with suggestions to replace the word with matching emojis, so when the word "pizza" was written, the keyboard suggested to replace the word with the slice of pizza emoji.



**Fig. 1.** Image showing the iOS 10 auto correct when typing the word 'pizza'



**Fig. 2.** Image showing the iOS 10 feature to replace words with emojis

## 2.1 Earlier work

Reading time is one of the aspects of the broader term readability which is a quite well studied area. Research done in a Windows environment [3] show that there is a difference in reading time between different fonts. This is backed up by more recent research [4] where they studied reading time, reading efficiency, perceived legibility and perceived attractiveness of fonts and conclude that there is a significant difference in reading time between different fonts.

With emojis being a relatively new phenomenon most articles exploring emojis are pretty new and not that many. There is previous work exploring the use of emojis in text messaging [2] that concludes that many people use emojis to enhance the tone or emotion in messaging.

As the findings of research suggests [1], emojis could be interpreted differently by different people. This might be an issue since the user might have different expectations of what emojis will be displayed instead of a certain word. This aspect is something this paper does not investigate but something that is important to acknowledge when doing research involving emojis.

In [5] the authors conclude that if a word was removed from a sentence it was easy for people to predict what the missing word was. Our study will investigate if the read time is affected by using emojis instead of words. It is interesting to have previous work showing that the human brain is very capable of predicting missing words of certain word classes, since if we remove a word and replace it with an emoji, the person interpreting that message might be predicting the missing word instead of analyzing the emoji.

## 3 Methodology

A user study was conducted with the aim to investigate our research question. This was done by developing a test application and inviting participants to read texts with and without some words replaced by emojis and record the reading time.

We prepared 8 shorter texts for our test participants to read. The texts were all selected from news articles written in English, had a length of 48 to 63 words and only consisted of full sentences. Each text had two versions, one were the original text (*without* emojis) and the other one were the text where some of the words had been replaced by emojis. The method of selecting what words was replaced by emojis was done by placing the texts on a phone running iOS 10 and letting the built in keyboard suggest what words should be replaced by an emoji. Every emoji only replaced one word. It was concluded by [6] that emojis that represent facial expressions and feelings could be interpreted differently in higher rate than other emojis. Therefore we avoided texts where facial expressions and feelings were mentioned.

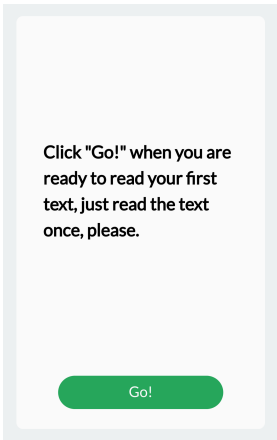
Previous research investigating the test environment while testing mobile applications [7, 8] suggests that there is no significant difference in regards to performance by doing field tests compared to lab testing. The one thing Nielsen

Text number	Word count	Number of emojis in alternative version
1	55	4
2	62	4
3	52	6
4	61	5
5	48	5
6	49	3
7	63	3
8	45	3

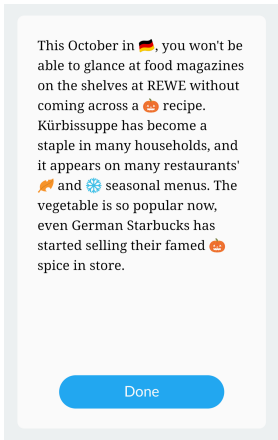
**Table 1.** Information about the texts used in the study

et al. [8] and Kaikkonen et al. [7] find is that field testing is more time consuming but could give a more detailed result while doing qualitative research.

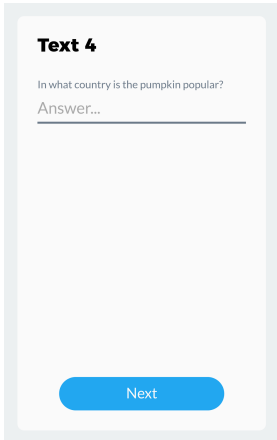
We conducted our study as a field test at the campus of Umeå University with the motivation that it should not affect the performance compared to lab testing. The number of people who volunteered for the study was 22 (12 males and 10 females), they had an average age of 22.09 (standard deviation: 1.77) and all of the test subjects had normal or corrected to normal vision. In our study we had two test groups, participants was divided randomly in one of the two groups. The first group read the original version (*without* emojis) of text 1, 3, 5, 7 and the alternative version (*with* emojis) of text 2, 4, 6 and 8. Group two read the original version (*without* emojis) of text 2, 4, 6, 8 and the alternative version (*with* emojis) of text 1, 3, 5, 7.



**Fig. 3.** The prompt given to users before they read each text



**Fig. 4.** The interface given to users when they read a text



**Fig. 5.** Image showing the form each user got after each text

Similar the method Mustonen et al. [9] used in 2004, the test were conducted by letting test subjects read each text and when they had read it through, they pressed a button to record the time. After each text they got a questionnaire to validate that they had read it through. In our study, the participant was

first handed information and instructions about the test through a pre-written page in the test application, this was followed by a form where the participants filled in basic information about age, occupation, gender and English proficiency. After that, the subjects got instructions about the procedure and pressed a button to confirm that they were ready to start reading the texts (see Figure 3). When they felt like they had read each text through, they pressed a button to record the reading time (see Figure 4). After reading each text the subject got a question about the text they had just read to validate that they read the text (see Figure 5). After answering the question they got the next text and the text-question loop went like this for all the 8 texts. The study was done on a smartphone with a 5.5" screen that had a 1440 x 2560 pixel resolution and a pixel density of 534 ppi (pixels per inch).

## 4 Result

After the study we collected the reading time of all the participants. Since the texts had slightly different lengths (see Table 1) we had to calculate a comparable average reading speed for each reading. We used the measurement WPS (Words Per Second) as the comparable unit of reading speed. This worked since one emoji replaced one word, so the number of words did not change when a text had one word replaced with an emoji. We could therefore compare the result of different texts with each other. How WPS is calculated can be seen in Equation 1.

We wanted to find out if there was a significant difference between the WPS of texts read with emojis versus texts without. We performed ordinary least squares regression analysis on the data. The regression analysis tries to find a model and determines how well that model fits with the given dataset. In our case the model compared WPS against text with or without emojis.

$$\text{WPS} = \frac{\text{Words in text } n}{\text{Reading time (in seconds) of text } n \text{ by participant } m} \quad (1)$$

Group	Mean	Std. Err.	[95% Conf. Interval]	T-Value	P-Value	VIF
WPS difference if text had words replaced by emojis	-0.3970	0.0845	[-0.5640,-0.2300]	-4.70	0.000	1.00

**Table 2.** Result of ordinary least squares regression on the data

The results from the ordinary least squares regression (seen in Table 2) is telling us that the number of words the user read per second is predicted to decrease with 0.3970 seconds (with a 95% confidence interval between 0.2300 and 0.5940 second decrease) if the test subject was reading a text with words replaced by emojis.

## 5 Discussion

This study tells us that it takes significantly longer time to read a text with words replaced by emojis than without. This answers our original question *Does replacing words with emojis affect the reading time of texts?* with a clear *yes*. When we started our research we had an idea that replacing emojis could decrease the reading time, since we ourselves felt that we stopped to interpret emojis while reading. The previous research we found on emojis mostly focused on the interpretation on emojis and how people was using them, not so much of how they had affect to our everyday life. Our research gives a strong incentive and proper arguments on that replacing words with emojis makes people spend more time reading than they have to.

Even if the age span of our participants was limited, the age span was some of the most frequent emoji users in 2015<sup>3</sup>. This could mean that they were more used to emojis than people of other ages and should have had less problems interpreting them compared to older people. We see our result as an interesting aspect to think about when developing new means of communication. Even if emojis are used to enhance the tone or emotion in messaging [2] it may be smart to think twice about developing software that makes their users increase the time it takes for the recipient to read their messages. We therefore suggest that *using emojis should be limited to enhancing the tone or emotion in messaging and not to replace words*.

### 5.1 Limitations

Due to a limited time frame to plan our study we conducted the study in our surroundings. This meant that volunteers for the study was in a limited age group and all of them were students. From our result we could not draw any conclusions about why it takes longer time to read texts with emojis. This study do not say anything about reading time of texts may decrease once people get more accustomed to reading texts with emojis in them.

One aspect that we ignored in the texts was the interpretation of the emojis. We did this since [1] suggests that emojis are often interpreted differently by different people, this might have caused confusion to some since one emoji could have been interpreted as a certain word but that did not match with the rest of the sentence and hence, confused the reader. This aspect is also relevant to the choice of emoji replacing method. We used the built in function in iOS 10 to replace words with emojis in the text, and somewhere someone at Apple have made the decision of what words should be replaced by what emoji, introducing another level of complexity to the problem that emojis are interpreted differently by different people.

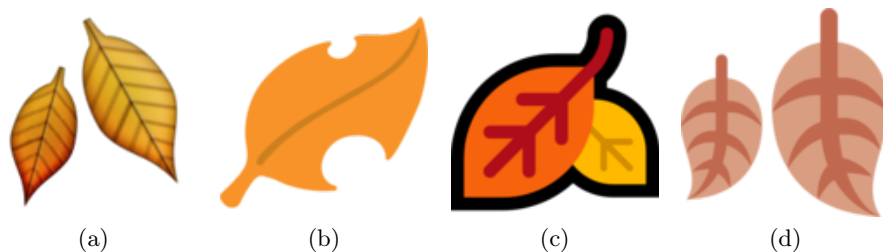
We did not take into account what mobile operating system the participants was accustomed to. The fact that emojis look differently in different software or

---

<sup>3</sup> Emogi Research Team (2015). *2015 Emoji Report* [http://emogi.com/documents/Emoji\\_Report\\_2015.pdf](http://emogi.com/documents/Emoji_Report_2015.pdf)



operating systems as shown in Figure 6 could mean that participants may have had different experience with the style of emojis used. The issue was limited by the fact that all participants performed the test on the same phone using the same style of emojis.



**Fig. 6.** Different versions of the Fallen Leaf emoji designed by (a) Apple, (b) Google, (c) Microsoft and (d) Twitter

## 6 Conclusion and Future Work

Our study finds that replacing words with emojis in texts increase the reading time. We found this by conducting 22 user studies with a custom application that measured the reading time when people read 8 texts with and without some words replaced by emojis.

We did not however investigate what is causing the increase in reading time. This would be something to look into. It would also be interesting to investigate if emojis have any effect on reading comprehension since our research just used reading comprehension to validate that they had read a text through. It would also be interesting to see if people perceive that they are reading texts slower when they read text with emojis in them.

As concluded in [5] it is often easy for people with enough knowledge of a language to anticipate words and fill missing words after they have been removed from sentences. The study also concludes that the rate of being able to fill in missing words differs for different word classes. It could therefore be interesting to investigate how users read texts with emojis, are they looking at the emoji and interpreting it or are they able or trying to fill in the missing word?

## 7 Acknowledgements

I would like to thank Suna Bensch and Thomas Hellström for being great teachers and giving me a great introduction to scientific writing with their course *Student Conference in Computing Science* at Umeå University. I would also like to thank my anonymous reviewer who provided great and constructive feedback, it surely made my paper better.

Many thanks to my peer review group and course mates Simon Johansson, Simon Ekdahl, Jan Nyhlen and Pia Nolen for great feedback, interesting and

giving discussions and support. I would also like to give a big thank you to the Interaction Design class of 2012 who provide the motivation and support that pushes me to become a better student and person.

## References

- [1] Miller, H., Thebault-Spieker, J., Chang, S., Johnson, I., Terveen, L., Hecht, B.: “blissfully happy” or “ready to fight”: Varying interpretations of emoji. ICWSM’16 (2016)
- [2] Cramer, H., de Juan, P., Tetreault, J.: Sender-intended functions of emojis in us messaging. In: Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services, ACM (2016) 504–509
- [3] Tullis, T.S., Boynton, J.L., Hersh, H.: Readability of fonts in the windows environment. In: Conference Companion on Human Factors in Computing Systems. CHI ’95, New York, NY, USA, ACM (1995) 127–128
- [4] Bernard, M., Lida, B., Riley, S., Hackler, T., Janzen, K.: A comparison of popular online fonts: Which size and type is best. Usability news **4**(1) (2002)
- [5] Van Berkum, J.J., Brown, C.M., Zwitserlood, P., Kooijman, V., Hagoort, P.: Anticipating upcoming words in discourse: evidence from erps and reading times. Journal of Experimental Psychology: Learning, Memory, and Cognition **31**(3) (2005) 443
- [6] Tigwell, G.W., Flatla, D.R.: Oh that’s what you meant!: reducing emoji misunderstanding. In: Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, ACM (2016) 859–866
- [7] Kaikkonen, A., Kekäläinen, A., Cankar, M., Kallio, T., Kankainen, A.: Usability testing of mobile applications: A comparison between laboratory and field testing. Journal of Usability studies **1**(1) (2005) 4–16
- [8] Nielsen, C.M., Overgaard, M., Pedersen, M.B., Stage, J., Stenild, S.: It’s worth the hassle!: the added value of evaluating the usability of mobile systems in the field. In: Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles, ACM (2006) 272–280
- [9] Mustonen, T., Olkkonen, M., Hakkinen, J.: Examining mobile phone text legibility while walking. In: CHI’04 extended abstracts on Human factors in computing systems, ACM (2004) 1243–1246



# The effectiveness of searching for smartphone apps sorted by color

Pontus Henstam

Department of Computing Science  
Umeå University, Sweden  
`pontushenstam@gmail.com`

**Abstract.** This study examines whether it is more efficient to find apps sorted by color than sorted alphabetically, in a smartphone app library. A test was conducted on eight participants. The test was based on letting people search and find specific apps, as quick as they could, in two different prototypes by using a smartphone. Both prototypes were based on an app library where you search for apps by scrolling up and down, with the difference that one of the prototypes had the apps sorted by color, and the other one had the apps sorted alphabetically. The results showed that a majority of the participants found most of the apps faster when they used the prototype with apps sorted by color. We can conclude that people can more effectively, or equally good, find apps that are sorted by color than apps sorted alphabetically.

## 1 Introduction

It can be frustrating for people while searching for apps in their smartphones because the apps can take some time to find, especially if there are many apps to search from. Apps in smartphones are often, especially in Android phones, automatically sorted in alphabetical order without any other options. This could be a reason for why people sometimes have to struggle while searching for their apps.

This study investigates whether it is more efficient to find apps sorted by color than sorted in alphabetical order. If it is more efficient to search for apps that are sorted by color, then sorting them that way could be a useful solution to add in smartphones.

The average number of installed apps in smartphones are 96 according to statistics from 2014 [1].

How people manually choose to arrange their apps on their smartphone desktops has been researched earlier [2]. The results showed that most people liked to arrange them so they could reach the most-used apps quickly. It was also found that some people liked to arrange their apps in color scale, however in this case, due to aesthetic reasons. How effective it was to find the apps sorted in color scale were not tested or discussed.

Research [3] of how people like to organize their files on their personal computers, found that visual search for files based on their location was preferred

over a text-based search. People liked to put their files in specific folders or regions, rather than to search for them by their names or keywords.

It is said, regarding visual search, that color differences support efficient visual search, and that there is a long history of research that points that color is one of the best ways to make something visually "pop out" from its surroundings [4].

It is common to ask what visual characteristics makes an object pop out the most. This question does not have a definite answer though because it always depends on the intensity of the particular feature and the context [5]. For example, in research [6], color was compared to orientation, to see what made an object pop out more. The result showed that it depended on of the size and the color saturation of the object, but also of its degree of difference from surrounding colors.

Studies [7] [8] has shown that there are two important factors when determining whether a target successfully pops out from its surrounding non-targets. Those factors are the degree of visual difference between the target and the non-targets, and the degree of visual difference between the non-targets. For example, if a word in a piece of text is highlighted with a red color, and the rest of the words are black and white, the red word would successfully pop out. But if the rest of the words would be in many different colors, the red word would pop out less successful.

## 2 Method

To test whether apps sorted by color are more efficient to find than apps sorted alphabetically we conducted a test. The test was based on letting people search and find specific apps, as quick as they could, in two different prototypes by using a smartphone. Both prototypes were based on an app library where you search for apps by scrolling up and down, with the difference that one of the prototypes had the apps sorted by color, and the other one had the apps sorted alphabetically.

Both of the prototypes contained 96 apps since that is what an average smartphone contains [1]. The apps that the prototypes contained was a sample of apps from an ordinary citizen's smartphone. The 20 apps that we chose for the participants to search for was apps that we considered not especially known by people, so the test would not be biased.

The test was performed on a smartphone from the brand OnePlus. Model name was OnePlus 3. We choose this phone because it was a regular smartphone. Almost any smartphone would work though.

The test was performed on totally eight different participants. Each participant searched for 20 different apps, one app at a time. Each app was searched for using the two prototypes separately. So totally each test contained 40 searches.

The test was conducted on four females and four males of different ages and occupations, to get an unbiased result.

The 20 apps that each participant searched for was randomly ordered for each test. The order in which the prototypes were used to search for each app was also randomized. We did this so that the test would not be biased.

We chose 20 apps for each participant to search for because it was the highest number we could make time for without making the test too time-consuming for the participants. Each test took between seven to 15 minutes.

## 2.1 Materials

We needed the following materials for the performance of the test on the participants.

- One high-fidelity prototype of a smartphone app library with apps sorted by color.
- One high-fidelity prototype of a smartphone app library with apps sorted alphabetically.
- A smartphone for running the prototypes, and for recording the participants' performance on the screen during the tests.
- A computer showing images of apps presented with logotype and name for the participants to search for.

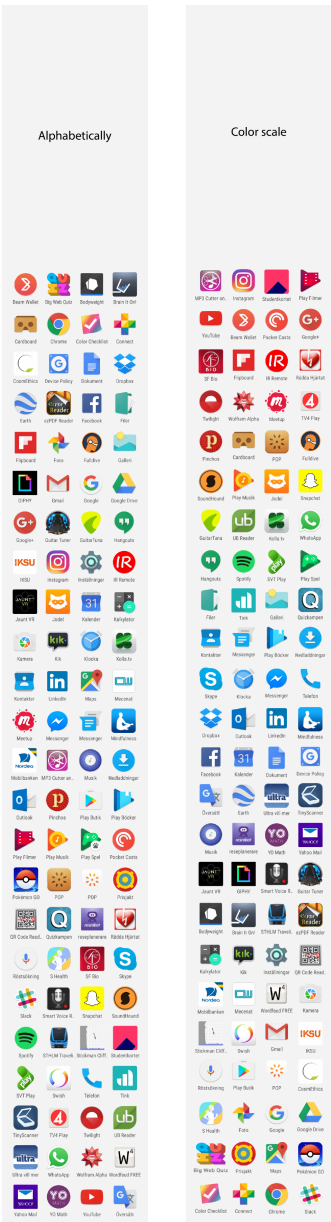
## 2.2 Procedure

Before each test, we told each participant that the test was anonymous and that they could cancel the test at any time. The participant were also asked following questions, to see if their answers could be reasons for their results and to determine if they were fit for the test.

- What is your age?
- What is your occupation?
- Do you know the alphabet?
- Are you color blind?
- Do you have any vision problems that you believe could affect your performance?
- What phone do you use today?
- How is your phone's app library sorted today?
- How often do you search for apps in your phone's apps library?

We then explained how the test would be performed. Each participant got time to look at an image explaining in what order the apps in the color ordered prototype would be in, so they would not have to spend time trying to understand the layout during the searches. We also explained what determined where an app would be placed in the prototypes so they would understand where to look for it.

The prototype with the apps sorted by color had the apps first ordered by a common color scale, from the top. It started with pink apps that shifted to red



**Fig. 1.** A figure showing the two prototypes. To the left is the prototype with the apps sorted alphabetically. To the right is the prototype with the apps sorted by color.



apps, to yellow apps, to green apps, to turquoise apps, to blue apps, to purple apps. Beneath the color scale, there was a section of all the black apps followed by a section of all the white apps and finally a section of all the multi-colored apps. What determined where an app would be placed was which color of the app's icon that was the most dominant. If an app's icon had several equally dominant colors, we placed it in the multi-colored section. We chose this arrangement because there was no standard for such a sorting, and the apps had to be sorted in some way.

Before, and during, each search the participants was shown an image of the icon and the name of the app that was to be searched for, so they would know what to look for.

Each test was recorded using an application that was running on the phone the participants used during the test. The application recorded everything that happened on the screen, where the participant clicked, and also what was said during the test, all to help us measure the participant's performance. When the tests were finished, we went through the recorded clips and measured how long it took for each participant to find the apps in the two different prototypes.

To make it easier for us to determine when a participant began a search, we had added an empty section in the top of the prototypes, just above the apps, where every search by the participants began. When the participant was in this section, it was clear that the searching had not begun because no apps could be seen on the screen from that view. When the participant then started scrolling down to search for apps, we could easily determine at what moment the search was started.

To make it easier for us to understand at what moment the participants had found the app, we asked them to click three times on it when they found it. By urging them to do this, it was a lot easier for us to distinguish between conscious and unconscious clicks on the screen when we went through the recordings.

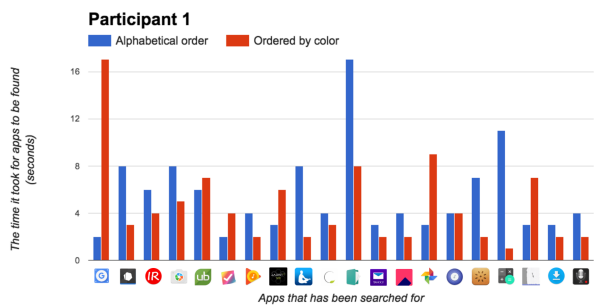
After each test, we asked the participants what they thought about using the two different prototypes to search for apps, and which prototype they preferred to use. We asked these questions to get a better understanding of the results.

### **3 Result**

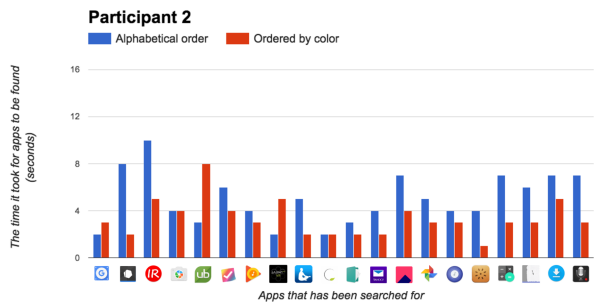
The results showed that six of the eight participants found most of the apps faster when they used the prototype with apps sorted by color. The result of these six participants varied from 12 apps more to only one app more that was found more quickly, than when the prototype with apps sorted alphabetically was used. The remaining two of the eight participants found most apps faster when they used the prototype with apps sorted alphabetically. However, for both of these two participants, it was a marginal difference of only one app more that was found faster than when the prototype with apps sorted by color was used.

3.1 Data Collection

Following Graphs shows how the participants performed at the test. The graphs compare the time it took for each app to be found, using the two different prototypes, to see which prototype was the most efficient for each participant. The red bars represent the time, using the color ordered prototype. The blue bars represent the time, using the alphabetically ordered prototype.



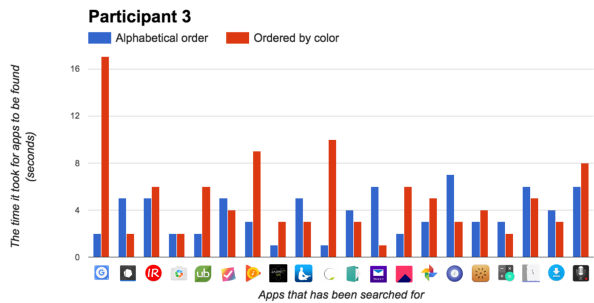
**Fig. 2.** A graph showing the performance of participant 1. 13 of the apps were found quicker when the participant used the color ordered prototype. Six of the apps were found quicker when the participant used the alphabetically ordered prototype. One of the apps was found equally quick.



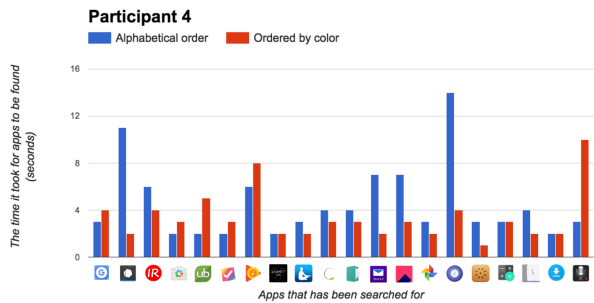
**Fig. 3.** A graph showing the performance of participant 2. 15 of the apps were found quicker when the participant used the color ordered prototype. Three of the apps were found quicker when the participant used the alphabetically ordered prototype. Two of the apps were found equally quick.

4 Discussion

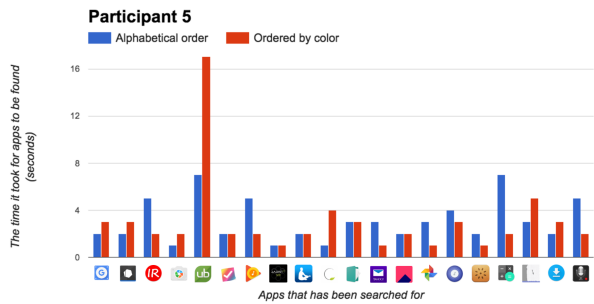
One thing that could affect the data is, of course, the participants' energy level during the test. For example, if a participant felt tired during the test one of the



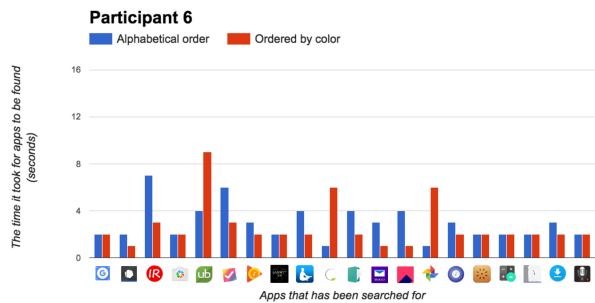
**Fig. 4.** A graph showing the performance of participant 3. Nine of the apps were found quicker when the participant used the color ordered prototype. 10 of the apps were found quicker when the participant used the alphabetically ordered prototype. One of the apps was found equally quick.



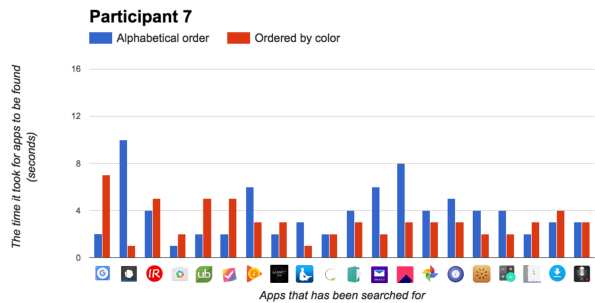
**Fig. 5.** A graph showing the performance of participant 4. 11 of the apps were found quicker when the participant used the color ordered prototype. Six of the apps were found quicker when the participant used the alphabetically ordered prototype. Three of the apps were found equally quick.



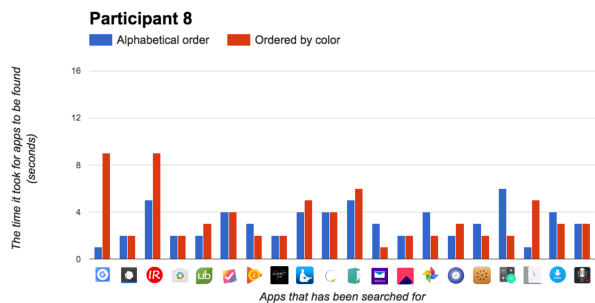
**Fig. 6.** A graph showing the performance of participant 5. Eight of the apps were found quicker when the participant used the color ordered prototype. Seven of the apps were found quicker when the participant used the alphabetically ordered prototype. Five of the apps were found equally quick.



**Fig. 7.** A graph showing the performance of participant 6. 10 of the apps were found quicker when the participant used the color ordered prototype. Three of the apps were found quicker when the participant used the alphabetically ordered prototype. Seven of the apps were found equally quick.



**Fig. 8.** A graph showing the performance of participant 7. 10 of the apps were found quicker when the participant used the color ordered prototype. Eight of the apps were found quicker when the participant used the alphabetically ordered prototype. Two of the apps were found equally quick.



**Fig. 9.** A graph showing the performance of participant 7. Six of the apps were found quicker when the participant used the color ordered prototype. Seven of the apps were found quicker when the participant used the alphabetically ordered prototype. Seven of the apps were found equally quick.

prototypes could probably be more effective because of that. The same goes for participants that might have felt extra alert during the test.

Something that could affect the result while searching for some apps is the apps positions in the prototypes. For example, if an apps name starts with a letter that is at the very beginning of the alphabet, it would also be placed at the top of the prototype with apps sorted alphabetically. The participants would then only have to scroll down just a little bit in the prototype to get to the app, which of course could save some time. Though, the same goes for the apps that were placed in the top of the prototype with apps ordered by color.

Another thing that could affect the result was that the arrangement of the apps that were sorted by color was new to the user, unlike the alphabetical arrangement of apps. Even though the participants got time to study an image explaining the order of the colored sorted apps, the alphabetically ordered apps should have been more familiar to them.

#### 4.1 Limitations

The tests were only conducted in calm environments for all of the participants. Smartphones are of course not only used in calm environments, so to conduct the tests in different environments would have been interesting. If for example, the tests would be conducted on a bumpy bus, the result might have turned out differently.

Since the number of participants was limited to only eight, we got a limited result.

## 5 Conclusion

From this study, we can conclude that some people more effectively find apps that are sorted by color than apps sorted alphabetically. We can also conclude that some people perform equally good using the two different sorting methods. We can not conclude that people can significantly more effectively find apps that are sorted in alphabetical order, than apps sorted by color, though. We have to keep in mind though that the tests were performed only on eight participants, and that the result might be different under different circumstances. Therefore further studies are necessary before any general conclusions can be made.

## References

- [1] Baeza-Yates, R., Jiang, D., Silvestri, F., Harrison, B.: Predicting the next app that you are going to use. In: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, ACM (2015) 285–294
- [2] Böhmer, M., Krüger, A.: A study on icon arrangement by smartphone users. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM (2013) 2137–2146

- [3] Barreau, D., Nardi, B.A.: Finding and reminding: file organization from the desktop. *ACM SigChi Bulletin* **27**(3) (1995) 39–43
- [4] Pashler, H.: *Attention*. Psychology Press (2016)
- [5] Ware, C.: *Information visualization: perception for design*. Elsevier (2012)
- [6] Callaghan, T.C.: Interference and dominance in texture segregation: Hue, geometric form, and line orientation. *Perception & psychophysics* **46**(4) (1989) 299–311
- [7] Duncan, J., Humphreys, G.W.: Visual search and stimulus similarity. *Psychological review* **96**(3) (1989) 433
- [8] Quinlan, P.T., Humphreys, G.W.: Visual search for targets defined by combinations of color, shape, and size: An examination of the task constraints on feature and conjunction searches. *Perception & psychophysics* **41**(5) (1987) 455–472

# Third-Person Avatar Control in Virtual Reality: a Nausea Provocation Comparison

Fredrik Johansson

Department of Applied Physics and Electronics  
Umeå University, Sweden  
`c04fjn@cs.umu.se`

**Abstract.** With regard to nausea provocation, a more suitable option for controlling an avatar in a 3D-world from a third-person view using a virtual reality head mounted display (VR headset) was sought. The new control mechanism removed all movement of the camera except the rotation created by the VR headset and added new ways of moving the camera and the avatar. The conclusion was that the new control mechanism was harder to use but less nausea provoking. This paper also contains future work of improving the user experience of the new control mechanism, to make it easier for the user to use and more intuitive.

## 1 Introduction

Today, games played in a 3D-environment are widely spread across the gaming industry. In many of these games, there is often some type of character, animal or another object the players needs to control or interact with.

Through many years of evolution, different types of control mechanisms have surfaced, one being the third-person view and the different control mechanisms which come with it. This kind of view is widely used for games where the main point is not to get the player immersed and have they feel like they are in the game, but get the player to better experience and be aware of their surrounding, and also to let the player see how the avatar interacts with its environment and other entities within the world [1].

Lately, a technique called virtual reality (VR) has reached the general public and made it possible for anyone to experience gaming from within this world and no longer only be able to look at it through a screen, like looking through a window. This new technique and experience early adopted the use of control techniques which were common for regular non-VR games, this resulted in many games being controlled using a first-person view, and not so many being controlled through a third-person view because the increased nausea provocation which a regular third-person control provides.

In regards to nausea provocation, this paper explores and compares an alternative approach of controlling an avatar in a computer generated 3D-world using a virtual reality head mounted display while using a third-person perspective.

## 2 Background

One of the biggest concerns with virtual reality today is that it makes users more or less nauseous. The strength of nausea differs from user to user and also between different ways of controlling the avatar [2]. The more natural way of moving and looking around in the 3D-environment and the less movement involved, the better the experience the player will have. A simulation where the user cannot move, and only look around is the least nausea provoking, and a simulation where the camera moves in high speed, jumps and do quick maneuvers are very likely to cause nausea [3].

Today, most avatar's controlled in a 3D-environment with a virtual reality headset are controlled through a first-person view approach because having the avatar one-to-one mapped to the actual person controlling it, looking through the eyes of the avatar, is the most natural way of interaction and the least nausea provoking. This type of control works well for a 3D-environment where the avatar represents a human being, or when the field of view does not need to be bigger than that of a human in real life or the user does not need to have increased situational awareness. But there are many cases where the user wants to be able to see more of the environment and its surroundings than what the user can see while using a first-person view approach. For example, if there are many big objects close to the avatar, and the information about what is happening and going on in the surrounding environment is important, then only being able to see what the avatar sees does not give the user sufficient information. In cases like this, a third-person camera is commonly used.

### 2.1 Avatar and the Avatarial Camera

An avatar is described by R. Klevjer as a fictional body through which the player can act as an agent in the fictional world it inhabits, and give the player realistic agency [4]. He also writes that there exists two avatar types, one which lives in the 2D world and one that lives in the 3D-world [4]. This paper will only be on avatars in a 3D-world, and about the viewpoint from which the player observes the avatar and the 3D-world the avatar inhabits. The viewpoint used in a third-person view approach is called the avatarial camera, which brings the experience closer to that of cinema [4]. However, Klevjer writes "The goal of visual realism in avatar-based 3D is not to imitate cinema or to make cinema interactive, but to give the player realistic agency within the game world." [4]. By this he means that the avatarial camera is not to make the player's experience to be equal to the experience of cinema but to have the same viewpoint but still have full control and realism in a 3D-world.

### 2.2 Regular Control Mechanism

Usually a control mechanism for controlling an avatar in a 3D-environment consists of three components: how the position of the avatarial camera relates to



the avatar, how the user controls the avatarial camera in the 3D-space and how the avatar moves according to the user's move- and rotate-inputs.

Using the regular control mechanism, the avatarial camera is located at a certain distance behind the avatar and always has its neutral direction facing the avatar's back. The avatarial camera follows the avatar one-to-one, this means that if the user uses provided inputs for rotating and moving the avatar, the avatarial camera rotates and moves as well. A rough real world comparison would be to have a video camera strapped at the end of a stick pointing along down the stick, and then have the stick strapped to your body pointing back and about  $45^\circ$  up in the sky (see Figure 1).

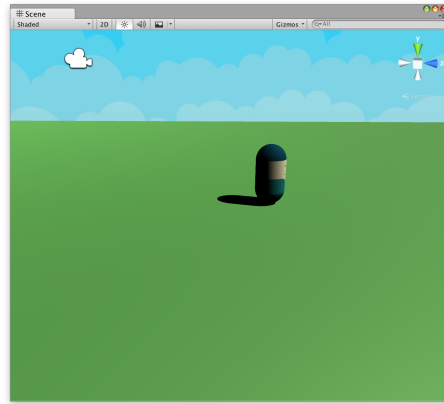


Fig. 1: Third-person view, demonstration

### 3 The Proposed Control Mechanism

The proposed means of controlling an avatar using a third-person avatarial camera in a VR world is that, instead of having the avatarial camera rotating around the avatar as the avatar rotates, the avatarial camera will be locked in place with exception for rotation around itself for the user to look around the environment mapped to the movement of the user's head.

As shown in Figure 2, the avatarial camera in 2a has, together with the avatar, rotated  $90^\circ$  anticlockwise such that the blue forward-arrow is pointing in the same direction as the avatarial camera, while the avatarial camera in 2b has not moved, but now sees the avatar from its side.

Usually, in third-person controlled environments, the avatar's direction of movement is determined by the direction of the avatar, but with the proposed control mechanism the avatar can rotate without having the avatarial camera move, so the avatar will instead turn to look in the direction of the mouse pointer's position. Today the movement keys (usually there are four of them) moves the avatar relative to its facing, that is: forward, backwards and sidestep left or right, but with my control, the keys for movements would move the avatar

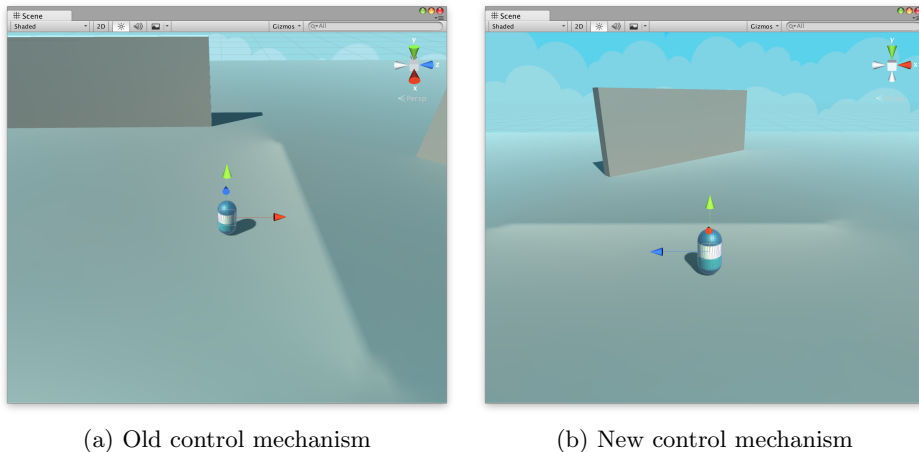


Fig. 2: The proposed new control mechanism compared to the old control mechanism, showing how the avatarial camera is positioned in relation to the avatar after the avatar is rotated  $90^\circ$  anticlockwise.

up, down, left or right relative to the screen, independent of the rotation and facing of the avatar. So even if the avatar is facing to the left of the screen and the user presses the up key, the avatar would move sideways to its right (that is up the screen) still facing left of the screen.

Because the avatarial camera does not move by following the avatar, the user controlling the avatar would not be able to see the avatar if they move the avatar too far away from the avatarial camera or around a corner, out of line of sight for the avatarial camera and the player. To solve this, something which is already widely used in virtual reality for movements of the avatarial camera is used, and that is teleportation [5]. By using teleportation the user selects an area in the virtual world and the avatarial camera jumps there, or rather instantly teleports to that location.

With the combination of an avatarial camera not dependent on the avatar, teleportation, and the improved way of moving the avatar in relation to the screens up, down, left and right instead of the direction the avatar is facing, it is possible to make it playable and really enjoyable in a third-person controlled virtual reality.

## 4 Method

To explore the possibility of designing a control mechanism for VR that is less nausea provoking than a regular third-person view approach, the game engine Unity together with the virtual reality headset Oculus Rift was used.

The process consisted of creating an avatar, a very simple game world, an avatarial camera, and two different mechanisms to control the avatar and the

avatarial camera in the game world. The process also involved giving the test subjects a purpose and task while they explore the different control mechanisms. The different control mechanisms were developed using the programming language C#.

#### 4.1 The Avatar

To represent the avatar, a capsular shape was used together with the texture shown in Figure 3a. It was given a couple of Unity Components<sup>1</sup>: Box Collider for detecting collision, a Mesh Renderer to render the texture, and a Movement Script for movement of the avatar. A Rigidbody was also added to the list of components for the avatar, the Rigidbody component made it possible to use the built-in physical engine in Unity to add forces to the avatar for realistic movement and collision. To not have the avatar be knocked over and fall down, the rotation axis for X-axis and Z-axis were locked in the Rigidbody component so the avatar only could rotate around its Y-axis.

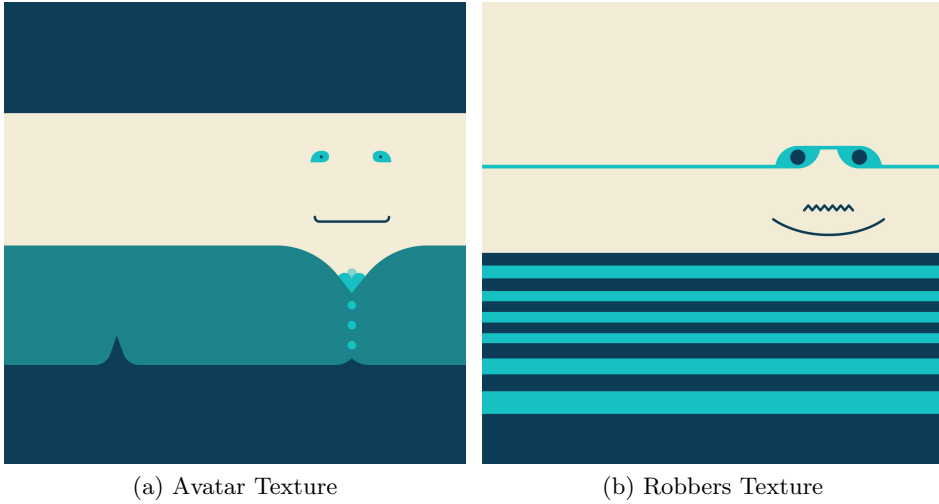


Fig. 3: The textures used on the capsular shaped game objects to represent the avatar and robber's.

#### 4.2 The Environment

The environment consists of cubes in different shapes and sizes. To represent a floor, a flat cube was added with a grass texture. The floor had brown-colored walls, about three times the height of the avatar and twenty times the width. Both the walls and the floor had Box Collider components added to them, so

<sup>1</sup> <https://docs.unity3d.com/Manual/Components.html>

not the avatar and the robbers would fall through the floor and run through the walls.

### 4.3 The Gameplay

To give the test subjects something to do so they could test the control mechanism, robber's were added to the game (see 3b for texture) as well as the possibility to fire bullets were added to the avatar. Except for a different texture then the avatar, the robbers got a very simple script for movement, they always turned towards the player avatar and moved straight towards it at a slow speed. The robbers were destroyed and removed from the world after they got hit by three bullets shot by the player. Also, a robber spawner-script were added to the environment which spawned a robber every third second.

### 4.4 The Old Control Mechanism

The avatar was added as a parent to the avatarial camera, this made the avatarial camera rotate and follow the avatar. A movement script was added to the avatar so it would rotate according to the controller or mouse movements.

### 4.5 The New Control Mechanism

The avatar needed two control mechanisms, one for rotating and one for moving. A script written in C# was assigned to the avatar. This script was run once every tick emitted from the game engine. For each tick, the script did two things. First, using the input values from a physical controller or keyboard a directional velocity was calculated and applied as a force on the avatar which moved in the desired direction. Second, get the position of the cursor and calculate the end point of an invisible line which originates from the avatarial camera, goes through the mouse pointer and ends up somewhere in the game world, then rotate the avatar towards that end point, be it on the floor or on a wall. At that point, which moves according to mouse movement, a blue, circular, thin disc was placed to show the user where they were pointing.

To rotate the avatar another script was added to the avatarial camera. This script used the same functionality as the movement script for tracking where in the game world the mouse pointer was pointing and when the user used the teleport button, the avatarial camera got moved to that position and turned facing the avatar.

### 4.6 Evaluation

After implementation, an evaluation of the hypothesis was conducted by measuring the nausea intensity of the test subjects.

For best result when measuring nausea, the frequency, intensity, and duration can be measured [6], but in this evaluation, only the intensity was measured since

the frequency has to be measured over a long period of time, days or even weeks, and the evaluation only last for maximum half an hour. Measuring the duration of nausea takes a long time, and often the patient gets back to the evaluator and reports at regular intervals, and for this time was not sufficient [6].

Since nausea cannot be observed by an external observer, the test subject reported the level of nausea using a visual analog scale (VAS) [6]. After finishing the test, the test subjects made marks on a 100 mm long line with 'no nausea' at the bottom and the 'worst nausea I have ever felt' at the top of the scale.

No questions were asked regarding prior experience with VR or games in general because according to Bendiksen [7] it does not have a noticeable impact on the nausea provocation.

Each test subject tested only one of the control mechanisms to avoid nausea from the first test to influence the second.

The goal of the test subject was to, for 5 minutes try to wipe out as many robbers as possible, this was mostly done to give the test subject something to do while experiencing the control mechanism.

15 people participated in the test, 7 of them got to play the game using the old control mechanism, and the other 8 the new.

## 5 Result

The result from the evaluation was as follows.

ID	Level	ID	Level
A	4	H	0
B	0	I	0
C	3	J	0
D	2	K	2
E	0	L	0
F	1	M	1
G	7	N	0
		O	0
Mean	2.429	Mean	0.375

Table 1: Showing test subjects A to O and their reported nausea level. The left most table shows the nausea provoked by the old control mechanism and the right most the new.

An unpaired t-test is derived from the calculated t-value of  $t = 2.2184$  together with 13 degrees of freedom and an alpha level of 0.05, which gives a standard error difference of 0.926. The two-tailed P value equals 0.045 ( $p < 0.05$ ) which shows that the difference is considered to be statistically significant.

## 6 Discussion

Even though this new way of control mechanism was proven in this small sample to be less nausea provoking, the users found it harder to use than the old control mechanism, and it was even more noticeable on those test subjects who had played similar games earlier, be it in VR or not. But at the same time many saw the potential in this type of control mechanism and found it entertaining, and looking at the scoring of the nausea level for this new control mechanism, most subjects did not feel any kind of nausea and some even said that they could go on forever. Which shows that motion and rapid movements of the avatars camera are big concerns regarding VR.

During the evaluation, a couple of things were noticed, which was not shown during developing. For instance, when teleporting, most users selected a spot where they wanted to get vision, where they wanted to see, and not where the camera should go, which were the case in the new control mechanism, so in a future development, the teleportation technique needs to reflect this problem.

Another thing which was noted was that with the implementation used during testing, the blue circular disc used for showing the user where they were pointing, did not only move according to the position of the mouse but also while the user was moving its head for rotating the avatars camera. This could easily have been solved by not using the current position of the mouse to highlight the teleportation spot but use the difference in mouse movement instead.

Maybe the old control mechanism is more intuitive, but doing more tests and listen to the subjects and implement changes which response to the user's expectations will lead to a more intuitive control mechanism.

## References

- [1] Black, D.: Why can i see my avatar? embodied visual engagement in the third-person video game. *Games and Culture* (2015) 1555412015589175
- [2] Shaw, L.A., Wünsche, B.C., Lutteroth, C., Marks, S., Callies, R.: Challenges in virtual reality exergame design. (2015)
- [3] Team, O.: Introduction to best practices
- [4] Klevjer, R.: What is the avatar. Fiction and embodiment in avatar-based singleplayer computer games (2006)
- [5] Buckley, S.: Why 'teleportation' makes sense in virtual reality
- [6] Del Favero, A., Tonato, M., Roila, F.: Issues in the measurement of nausea. *The British journal of cancer. Supplement* **19** (1992) S69
- [7] Bendiksen, S.H., Jørgensen, L.: Balancing the virtual reality experience: A study of human reactions with the oculus rift dk2. (2016)

# Examining text quality in video encoded in different resolutions with OCR-software

Simon Johansson

Department of Computing Science  
Umeå University, Sweden  
simonljohansson92@gmail.com

**Abstract.** Text quality in a video is critical to the user’s perceived video quality, and when video is consumed through smaller screens and access to bandwidth through mobile networks is limited text quality may deteriorate with video quality. In a test with four varying high definition videos, we investigated the relationship between video quality and text quality of different font sizes using Optical Character Recognition software. The results from the tests showed that there’s a bicubic relationship between video quality and text quality, and that the smaller font sizes we tested are affected the most when video quality deteriorates.

## 1 Introduction

Consuming video through your smartphone through various streaming services is getting more and more common <sup>1</sup>. For majority of the online video streaming services [1], the video data YouTube transmit to the user’s device go through a process to limit the amount of transmitted data to ensure the delivery of the video. This process may degrade the quality of the video stream by some factor depending on user’s available bandwidth, and any text content in the video may suffer in quality. Large text within a frame may still have sufficient text quality to be interpreted, but smaller font sizes may become unintelligible. Findings of [2] suggest that text quality contributes to the overall perceived video quality.

This study has been conducted to find the relationship between video quality and text quality of different font sizes, and will do so by examining four font sizes using OCR-software in various videos encoded with the H.264 codec.

This study is helpful for video producers or creators for them to consider the text quality in the video when the video is encoded in different resolutions. This is also key to making their content more accessible to a wider audience where access to bandwidth is limited. This study may also help target any problems that current compression techniques have with text.

---

<sup>1</sup> <https://www.emarketer.com/Article/Smartphones-Continue-Drive-Mobile-Video-Consumption/1013389>

## 1.1 Background

YouTube is one of the most popular video streaming websites, and the website makes up about 20-35% of all Internet traffic [3]. YouTube supports High Definition (HD) videos in the MP4 container format, which get encoded with the H.264 codec when they are uploaded [4]. H.264 is a successor to H.262, MPEG-2 Video, H.263, and MPEG-4 with a set of new features that increases the compression efficiency relative to its predecessors. The largest difference to previous standards is the increase of flexibility and adaptability of the standard, and use 60% less bit rate to achieve the same reproduction quality [5].

Video bit rate, often used with the unit kbps (kilo-bytes per second) or kb/s, and the codec is what reveal the image quality of each frame in the video [6].

There are two types of techniques for compressing data; lossless and lossy compression. Lossless compression techniques will lose no information, such as the reconstructed data will be identical to the original data. Lossy compression techniques will suffer from information loss, so the encoded data set will not be identical to the original data, but similar [7]. The compression algorithm used in H.264 is regarded as lossy, i.e., the decoded signals of the video will lose information and not be identical to the original video [8].

Optical Character Recognition (OCR) is a system that can retrieve full alphanumeric characters, printed or handwritten, from a document. [9]. OCR software are commonly used for automatic postal processing, script recognition, banking, security (i.e. passport authentication) and language identification [10]. The authors of [11] reviewed how OCR software can extract text from a document. The process they presented involves three main steps which try to extract the string of characters by a character-by-character basis:

1. Extracting the individual character images
2. Analyze the character images and determine its shape
3. Contextual processing

There are OCR-software that use an additional post-processing algorithm using a dictionary or neural network to improve the accuracy [12], such as Google's Tesseract, an open-source OCR-system [13]. The authors of [14] has conducted annual tests since 1990 of the accuracy of different OCR-software, where Tesseract participated with the name "HP Labs OCR" [13]. Tesseract performed well; it was the in top three of best overall accuracy on binary images [14]. Due to the accuracy and availability of Tesseract, it will be the OCR-system used when gathering test results.

## 1.2 Earlier work

The authors of [2] evaluate the effects of degraded text quality on perceived video quality in different video encodings. Their study did not solely study the effect video encoding has on text quality, and the font size was not one of their variable parameters in their test.



The authors in [14] used a method to find the OCR software's accuracy by calculating  $\frac{n - \#errors}{n}$  where  $n$  is the number of characters and  $\#errors$  are the number of errors. Errors are determined by the number of substitutions and/or deletions required to reach the target string.

In the domain of retrieving text from a source of degraded quality, the author of [15] determined OCR software's ability to correctly reproduce text from a degraded quality document of text. They targeted different aspects that may make character recognition more difficult, and these areas of difficulty are used in the discussion when analyzing the results.

## 2 Methodology

This section describes the method used to gather data to be able to find the relationship between the video quality and text quality. The method is based on the OCR-system Tesseract's ability to retrieve text from video frames, and how well it performs is used as the indicator of text quality.

### 2.1 Test material

The test material for this test is gathered on the same criteria of temporal motion as the research of [6] and [16]. The set of videos is sourced from the Red sample material <sup>2</sup>.

Content	Temporal motion	Length
Portrait A	Low	17 s
Portrait B	Low	9 s
Underwater	Medium	15s
Helicopter	High	7 s

**Table 1.** The contents of the selected test material, their perceived temporal motion and length

Each video in the test material 1 is provided with an overlay of the pangram <sup>3</sup> "The quick brown fox jumps over the lazy dog" in 4 different font-sizes; 60, 36, 24 and 16 pt. The font used for the pangram is sans-serif Arial, one of the most common fonts used [17], with black font color with a white background. The white background is important to differentiate the text color with the background of the video. The videos are resized from its original resolution to 1080p (1920 by 1080 px), and limited to a bitrate of 224 kbps, similar to [18].

### 2.2 Test procedure

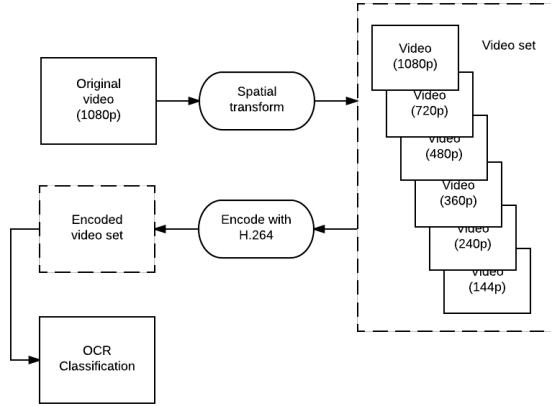
The following process is applied to each video in the Test material section.

<sup>2</sup> <http://www.red.com/sample-r3d-files>

<sup>3</sup> <http://clagnut.com/blog/2380/>

1. The video is encoded with H.264 using the FFMPEG software <sup>4</sup>.
2. Each frame is gathered from the video.
3. Tesseract interprets the text content for each video frame in the encoded video file.
4. The string produced by Tesseract of each frame is compared with the target string in the video. The method of comparison is that of a normalized Levenshtein string distance, similar to that of the method used in [14]. The string distance metric determines the similarity of the target string and the string provided by Tesseract on a scale of 0 - 100%.

The process described is repeated for the resolutions 720p, 480p, 360p, 240p and 144p. The average result of each video frame retrieved from step 4 and the standard deviation is presented in the Result section.



**Fig. 1.** System overview every video are processed through.

The entire system is described in figure 1.

### 3 Result

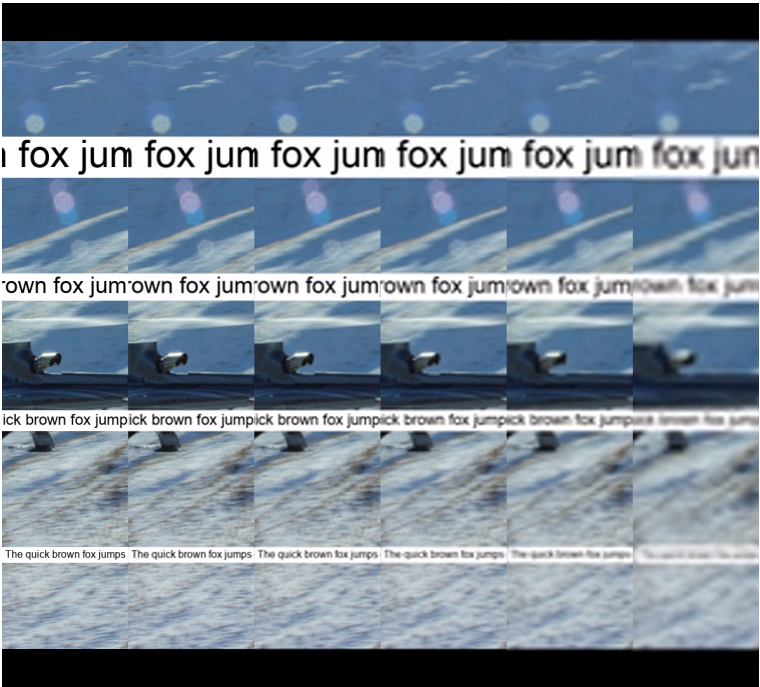
The result of the method applied to each video is presented in tables 4, 5, 6, 7, 8 and 9. Screenshot of the first frame of the video "Helicopter" is provided in figure 2. A montage of screenshots of the first frame in different resolutions of the video "Helicopter" is provided in 3

By comparison of box plot in figure 4 and 5, the OCR-system could reliably and accurately retrieve the text in 1080p for all font sizes, while for the font sizes 24 and 16 pt in resolution 720p the accuracy is decreased by  $\tilde{11}\%$  and  $\tilde{35}\%$  respectively compared to the corresponding font sizes in 1080p.

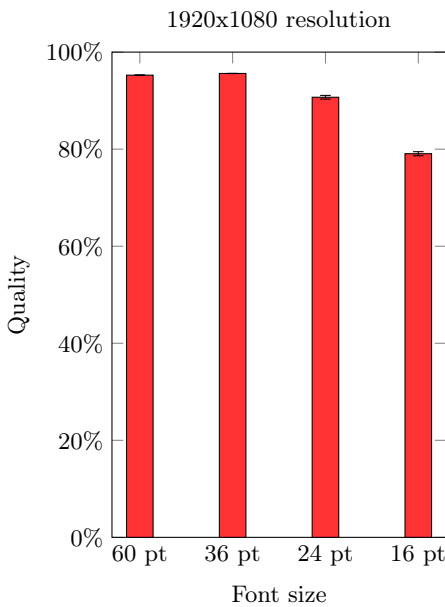
<sup>4</sup> <https://www.ffmpeg.org>



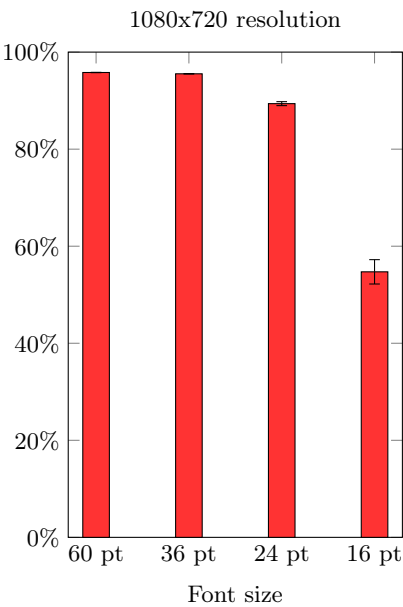
**Fig. 2.** Screenshot of the first frame of the video "Helicopter" in 144p resolution.



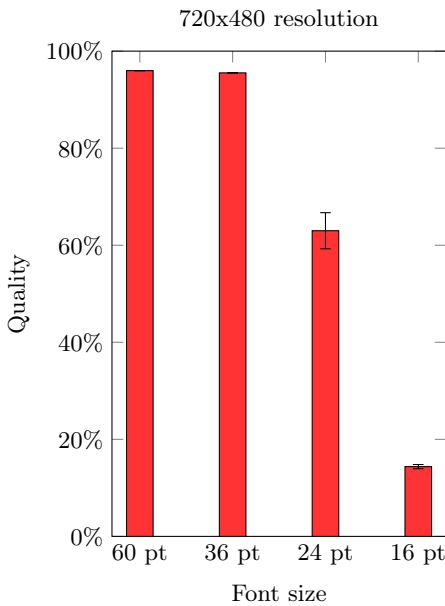
**Fig. 3.** Screenshot montage of the first frame of the video "Helicopter" in descending resolution from left to right.



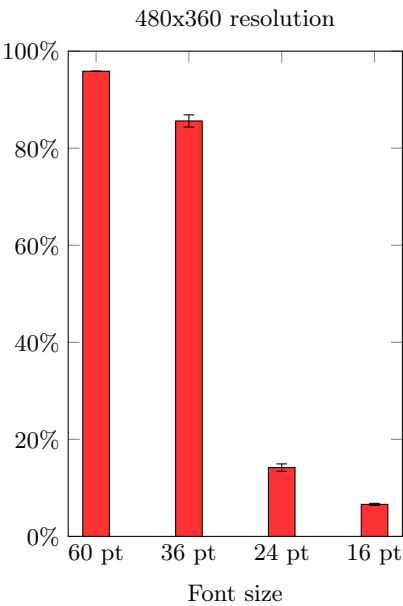
**Fig. 4.** Average text quality of all videos per font size in 1080p resolution.



**Fig. 5.** Average text quality of all videos per font size in 720p resolution.

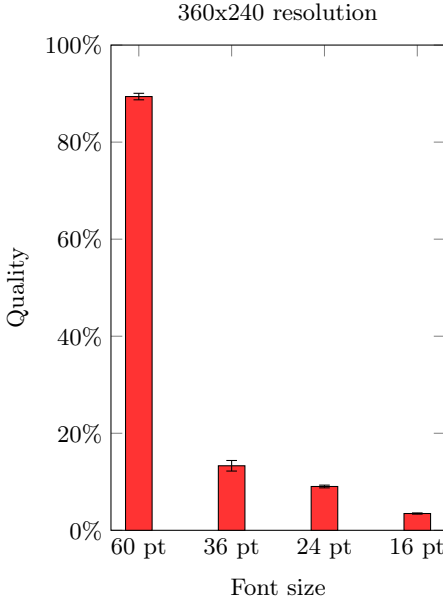


**Fig. 6.** Average text quality of all videos per font size in 480p resolution.

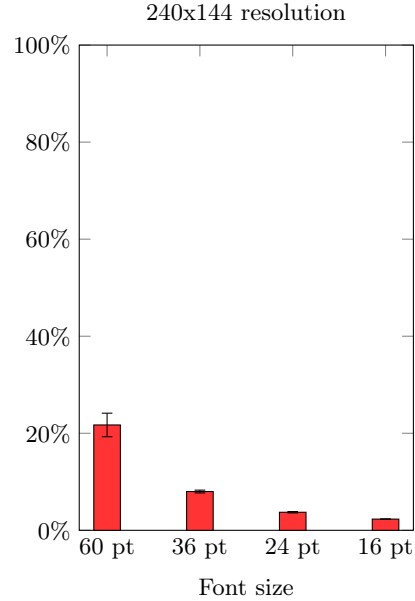


**Fig. 7.** Average text quality of all videos per font size in 360p resolution.

For figure 6 and 7 the OCR-system Tesseract can accurately retrieve font sizes 60 and 36 pt, similar to 4 and 5, but there's a notable drop off in text quality for font sizes 24 and 16 pt. Comparing figures 7 and 6 there's a  $\tilde{61}\%$  difference in quality for 24 pt, and a  $\tilde{20}\%$  difference for 16 pt.



**Fig. 8.** Average text quality of all videos per font size in 240p resolution.



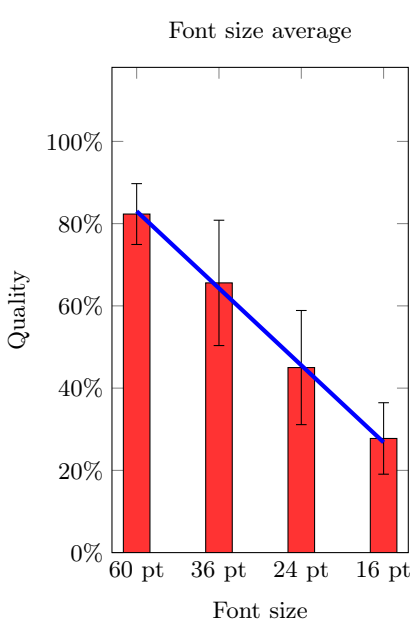
**Fig. 9.** Average text quality of all videos per font size in 144p resolution.

For resolution 240p in figure 8, the text quality of 60 pt is still high with only a drop of  $\tilde{8}\%$  in quality compared to 480p, but with a higher error compared to all higher resolutions in figures 4, 5, 6 and 7. Figure 9 and resolution 144p has the most prominent drop in text quality across all font sizes, which all are below 16%.

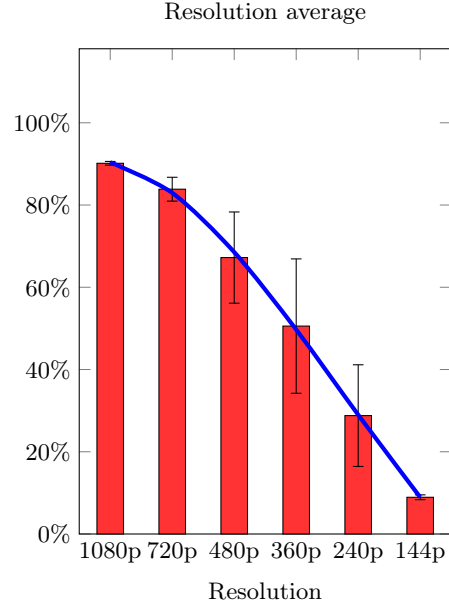
Comparing all font sizes and their average quality across all different resolutions, as seen in figure 11, we can see that the bigger font size 60 pt retains the best text quality across all resolutions. The relationship between font size and average quality appear to follow a linear function across all tested resolutions. Comparing all resolutions and the average text quality across all font sizes the highest resolution 1080p maintain the best text quality across all font sizes, respectively the lowest resolution 144p keep the worst text quality. The relationship between resolution and quality appear to follow a bicubic function.

## 4 Discussion

From the results we can deduce that there is a relationship between video quality and text quality of the font sizes we tested. The relationship between font



**Fig. 10.** Average text quality per font size, blue line emphasizes the linear relationship.



**Fig. 11.** Average text quality of all font sizes per resolution. Blue line fitted from data and emphasize a bicubic relationship.

size and text quality is linear for the tested font sizes. This may suggest that proportionately to the size of the font it will retain text quality across a larger spectrum of different resolutions and video compression.

Looking at the average text quality of all font sizes, figure 10, the resolution 1080p produces consistently high-quality text, while for the lowest tested resolution 144p the text quality is consistently bad. This figure also suggests that the relationship between video quality and text quality is not linear, but follow a bicubic function.

Examining figure 3 we can see that the text in lower resolutions, inhibit some of the characteristics as test material of the study [15], with blurring text and loss of text detail. This is most prominent in lower resolutions, especially 144p, as seen in figure 2.

Due to the low variance of the results in figure 4, 5, 6, 7, 8 and 9 we can conclude that the contents of the video do affect the text quality, but not by much.

#### 4.1 Limitations

We can not conclude if the linear relationship between font size and text quality holds for font sizes higher or lower than the tested font sizes, since if the linear relationship is true for larger font sizes the text quality for font size 96 pt would

be greater than 100%. Most likely there's an upper and lower limit where the text quality will be equal or close to 100% above the upper limit respectively equal or close to 0% lower than the lower limit.

## 4.2 Future work

Further work is needed to verify the relationship between font size and text quality for additional font sizes, both within the span of tested font sizes and outside. Additional further work is needed to be able to say that OCR-software's ability to retrieve the text of document correlates to the subjective text quality of a document. More results on text quality can be gathered in any future tests from different font types and different bit rates. The objective method suggested for this study could be validated through a subjective study assessing how text quality affects readability.

## 5 Conclusion

From the results we can conclude that there is a bicubic relationship of video quality to text quality of the font sizes we tested; when video quality deteriorates so does text quality of the tested font sizes. Larger font sizes retain text quality better for a larger span of video resolutions, and smaller font sizes are affected the most by resolution reduction and the application of the H.264 codec. Higher resolutions, such as 1080p and 720p, provide good text quality for all tested font sizes. To ensure good text quality in a video of different resolutions we suggest using a font size of 60 pt or larger.

## References

- [1] Sullivan, G.J., Wiegand, T.: Video compression-from concepts to the h. 264/avc standard. *Proceedings of the IEEE* **93**(1) (2005) 18–31
- [2] Knoche, H.O., McCarthy, J.D., Sasse, M.A.: Reading the fine print: the effect of text legibility on perceived video quality in mobile tv. In: *Proceedings of the 14th ACM international conference on Multimedia*, ACM (2006) 727–730
- [3] Finamore, A., Mellia, M., Munafò, M.M., Torres, R., Rao, S.G.: Youtube everywhere: Impact of device and infrastructure synergies on user experience. In: *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, ACM (2011) 345–360
- [4] Garapati, N.: Quality estimation of youtube video service. (2010)
- [5] Wiegand, T., Schwarz, H., Joch, A., Kossentini, F., Sullivan, G.J.: Rate-constrained coder control and comparison of video coding standards. *IEEE Transactions on circuits and systems for video technology* **13**(7) (2003) 688–703

- [6] Jumisko-Pyykkö, S., Häkkinen, J.: Evaluation of subjective video quality of mobile devices. In: Proceedings of the 13th annual ACM international conference on Multimedia, ACM (2005) 535–538
- [7] Pujar, J.H., Kadlaskar, L.M.: A new lossless method of image compression and decompression using huffman coding techniques. *Journal of Theoretical & Applied Information Technology* **15** (2010)
- [8] Richardson, I.E.: H. 264 and MPEG-4 video compression: video coding for next-generation multimedia. John Wiley & Sons (2004)
- [9] Charles, P.K., Harish, V., Swathi, M., Deepthi, C.: A review on the various techniques used for optical character recognition. *International Journal of Engineering Research and Applications* **2**(1) (2012) 659–662
- [10] Chopra, S.A., Ghadge, A.A., Padwal, O.A., Punjabi, K.S., Gurjar, G.S.: Optical character recognition. *image* **3**(1) (2014)
- [11] Srihari, S.N., Lam, S.W.: Character recognition. In: *Formal Description Techniques, IV (FORTE'91)*. (1992) 165–184
- [12] Bassil, Y., Alwani, M.: Ocr post-processing error correction algorithm using google online spelling suggestion. *arXiv preprint arXiv:1204.0191* (2012)
- [13] Smith, R.: An overview of the tesseract ocr engine. In: *Proc. Ninth Int. Conference on Document Analysis and Recognition (ICDAR)*. (2007) 629–633
- [14] Rice, S.V., Jenkins, F.R., Nartker, T.A.: The fifth annual test of ocr accuracy. Technical report, Information Science Research Institute (1996)
- [15] Hartley, R.T., Crumpton, K.: Quality of ocr for degraded text images. *arXiv preprint cs/9902009* (1999)
- [16] Winkler, S., Faller, C.: Maximizing audiovisual quality at low bitrates. In: *Proc. Intl. Workshop on Video Proc. and Quality Metrics for Consumer Electronics*. Number LCAV-CONF-2005-026 in 1 (2005)
- [17] Bernard, M., Liao, C.H., Mills, M.: The effects of font type and size on the legibility and reading time of online text by older adults. In: *CHI'01 extended abstracts on Human factors in computing systems*, ACM (2001) 175–176
- [18] Knoche, H., McCarthy, J.D., Sasse, M.A.: Can small be beautiful?: assessing image resolution requirements for mobile tv. In: *Proceedings of the 13th annual ACM international conference on Multimedia*, ACM (2005) 829–838



# Evaluating efficiency of interactive notifications on mobile devices

Joakim Ljungren

Department of Computing Science  
Umeå University, Sweden  
`id12jln@cs.umu.se`

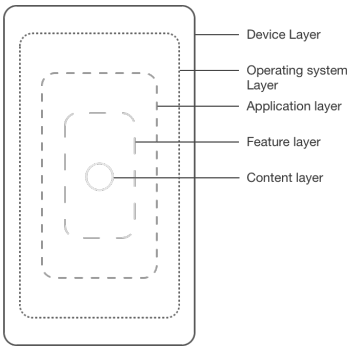
**Abstract.** This study evaluates the efficiency as a part of usability of using interactive notifications on mobile devices for conversational applications. A quantitative user study was conducted in the field with measurements of task completion time. Test users tried two methods of user interaction with notifications: To open the application and respond (conventional method) or to respond inside the notification directly (interactive notification method). The results showed that there was no significant difference in efficiency between the methods. We argue that the potential of a dynamic interface like this is not negligible. We cannot claim that either method is better nor worse in terms of usability, we can only say that the evaluated type of interactive notifications is not more or less efficient than the conventional approach. More research is necessary to find if the user experience can improve from this type of interaction model.

## 1 Introduction

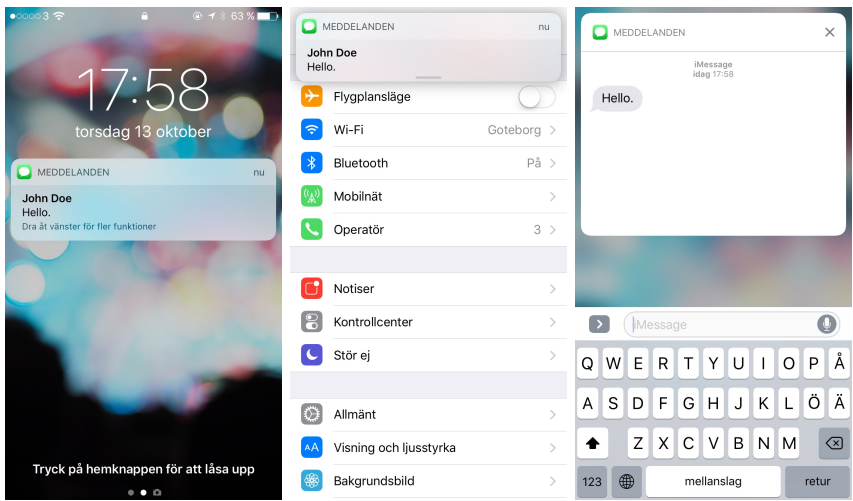
Notifications on modern mobile devices are issued through the operating system. In some cases, this information is easy for the user to directly consume from the notification. However, if some kind of user response is sought, the conventional approach is that the application from which the notification came is opened. The application user interface is opened for instance by tapping the section of the screen with the notification. The operating system interface, where the notification is managed, therefore serves as a link to the application.

A user interface (UI) of a specific application can in a way be seen as a partial layer of the mobile interface as a whole. The parent layer in this perspective would be the operating system (OS) layer where the notifications appear, see Figure 1. In the conventional approach of giving response to a notification, it is necessary to open the application UI to interact with the application functionality. However, there are some examples of when the application functionality transcends the OS layer. This alternative approach does not require the application UI to be opened and the user interaction remains within the OS layer. This enables the user to directly interact with the application from outside the application layer [1]. The latest versions of popular mobile platforms like iOS

and Android use this approach, and it will be referred to as interactive notifications in this article, see Figure 2. Interactive notifications is also an example of a more dynamic interface, which is another term that will be explained in the Background chapter.



**Fig. 1.** The layers of a mobile user interface. Inspired by a figure of mobile infrastructure [2].



**Fig. 2.** iOS 10 example of interactive notifications with possibility to respond to a message by expanding the notification card rather than opening the application.

This paper will try to decide if using interactive notifications regarding messaging applications is more efficient compared to the traditional approach of opening the application UI. What efficiency in this perspective means will also follow in the next chapter. It is expected that the method of interactive notifications is more efficient since it is something that is getting more common on mobile platforms and intuitively should be a more efficient option. On the other

hand, it could be so that the user rather needs the greater context of the application UI to make quick decisions. There is a possibility that the full occupation of the users attention-span on the screen is necessary for the user to process the specific content properly and being able to respond quickly. That several notifications are available to expand in the same view could be distracting and thus the option of responding inside interactive notifications ends up less efficient. The method of opening the application UI could prove to give a better clarity of what system the user is interacting with and in that case maybe represent the faster interaction process compared to be using interactive notifications. We try to answer the following: *Which is the most efficient method for interacting with notifications when responding to messages on a mobile device: the conventional approach through the application layer or the interactive notifications approach through the OS layer?*

We do this by conducting a quantitative user test, comparing the two methods when interacting with messaging notifications. The tests was performed with a mobile device on two groups of users and the task completion time of each test was measured to compare the efficiency of the two methods. The result shows if there is a more efficient interaction pattern between the two methods. The conclusion from the results with respect to surrounding aspects will then tell whether the approach is a better alternative or not compared to the conventional approach. The ambition is also for this research to serve as a good indicator of what role efficiency plays as part of the user experience on a mobile interface.

## 2 Background

It is important to perceive different associations of the word efficiency as well as its counterparts to properly grasp the content of this paper. The term efficiency itself literally means the accomplishment of a task with minimum expenditure of time and effort, according to the dictionary<sup>1</sup>. This definition is accurate for the context of this research. However, in the perspective of usability, such as in this study, efficiency can also be seen as one aspect out of three that affect the usage of an interface. This definition is more detailed and described with reference to usability. The other subcategories are then effectiveness and satisfaction [3]:

**Effectiveness** as a factor of usability means the accuracy and level of completion with which users reach a goal. Indicators of this are solution quality and error rates. The measure of effectiveness is therefore the outcome of the user interaction with a system.

**Satisfaction** is the part of usability where the users subjective experience is addressed. It is the qualitative factor covering what attitude users has towards using something and where their preference lies.

**Efficiency** is closely related to effectiveness. In the usability perspective, efficiency can be referred to as the relation between how well a goal is achieved

---

<sup>1</sup> [http://www.dictionary.com/browse/efficiency?](http://www.dictionary.com/browse/efficiency) A definition of the term efficiency, accessed 2016-10-18

and the resources it took to get there. This relation means that the task has to be fulfilled at some level to be measured, and the measurement is about the resources expended to complete it. To see that a task is completed and that the measurement of the efficiency is legitimate, the task completion (effectiveness) has to be validated. Hence, it is important to control the effectiveness of tasks to be able to measure the efficiency as a part of usability. Task completion time and learning time are both indicators of efficiency. Thus, efficiency measurements are in terms of time.

Some claim that learning time is a separate measurement that belongs to the attribute learnability, and that users should at least have some experience of a task before the efficiency can be measured [4, 5]. Another view is that it is rather the change of efficiency that indicates learnability. The same author shows that time between actions could be a measure of efficiency, but that it is always dependant on the context. The most common measurement of efficiency is task completion time [6].

The significance of this study comes from comparing the usability of two methods of interacting with mobile interfaces and with notifications in particular. Usability can as explained be seen to be dependant on three factors: efficiency, effectiveness and satisfaction. These conditions should be individually evaluated when measuring usability and assumptions about their correlation should be avoided. Thus, efficiency can normally not prove as a single factor of the complete usability and it can neither tell anything about the other two factors. It only serves as its own part of usability [3].

There are other definitions of usability. Nielsen and Norman assert that it is in fact dependant on five and not three components: learnability, memorability, satisfaction, errors and efficiency [4]. The different definitions cover up the same entirety, perhaps in different detail. There seems to an agreement of what efficiency as a single factor represents in this context.

It can be argued that multiple UI:s on your mobile device become less and less relevant and there is a lot of third-party content that could be used in a more unified manner. We could be moving towards having one dynamic interface instead of one for each application [7]. Especially since the biggest technology conglomerates are no longer competing with one functionality, they want to engage in all of the users activity [8]. Clear examples of this could be Google Assistant<sup>3</sup>, Google Now<sup>2</sup>, Siri<sup>4</sup>, Alexa<sup>6</sup>, Cortana<sup>5</sup> and other artificially intelligent assistants or multi-tasking environments. This abstraction raises questions of how the user experience will change and if it will be all for the better. This is regardless of what type of interface that will rule in the future, like touch screens, voice control (which all of the mentioned AI systems use), gesture control, or something else. It is still a question about general user experience. Our study concerns this uncertainty by questioning if users really gain something from the

unified and dynamic type of interface and doing so relevant to what methods that are used today.

Since the real-world context of interacting with notifications on a mobile device is often done in stressful environments, on many different locations and in many contrasting situations, one could argue that the user tests of this research should be effected in a similar fashion. This is usually called field testing and it is different from laboratory testing, where the testing environment is sealed off from the surrounding context. Both methods have their pros and cons, so what method to use depends on the research objective [5]. There is research that acknowledges that field testing is better when it comes to usability studies on mobile devices because it was proven to find more problems with the system [9]. In spite of this, according to another evaluation, conducting a field test instead of a laboratory test when evaluating usability on mobile devices takes up more time and does not guarantee substantially different results [10]. However, it seems as the more similar the laboratory environment is to the actual user environment, the smaller difference it is between the results of a field-type test and laboratory-type test [10]. A field test has an advantage because of how it is in a mobile context, but the methodology also lacks control of the test participants [5]. This challenge could be managed by having short tests, so that the attention of the test users could be focused for a short period of time.

The most common type of user studies are of the qualitative type and they often address the whole usability rather than one aspect of it [9, 10]. When doing quantitative measurements like efficiency testing, the possible fact that field testing gives better usability results in terms of number of problems detected is less relevant. On the other hand, considering the similarity of testing environment, the circumstances for the tests should be adapted to a typical environment in a viable extent. A user test of efficiency could thereby be argued to be most fitted to the field, but in a static section without too much interference.

### 3 Method

When conducting this type of quantitative user study, it is important to secure that the results can be statistically reliable. The number of users in the testing group therefore has to be of an amount that gives a trustworthy sample. According to a well regarded source, a good number of test users when measuring

---

<sup>2</sup> <https://www.google.com/search/about/learn-more/now/> Google Now, a more unified interface for accessing all information, accessed 2016-10-24

<sup>3</sup> <https://assistant.google.com/> Google's AI, probably soon to be inseparable from Google Now, accessed 2016-10-24

<sup>4</sup> <http://www.apple.com/se/ios/siri/> Siri, an AI interface for Apple devices, accessed 2016-10-24

<sup>5</sup> <https://support.microsoft.com/sv-se/help/17214/windows-10-what-is-Cortana> Cortana, Microsoft's counterpart of an AI assistant, accessed 2016-10-24

<sup>6</sup> <https://developer.amazon.com/alexa> Alexa, Amazon's intelligent assistant, accessed 2016-10-24

quantitative usability metrics like efficiency is 20 [11]. This is per design, so in our case, per approach of notification interaction. The total number of test users was for this reason 40.

We use task completion time as a measurement of efficiency in this research. Consequently, the time it takes for a user to complete a task. For tasks that are habitual or routine to the user, task completion time may also be used as an indicator of overall usability. In this study, the user interaction will not be obvious to the users and assumptions about the overall usability cannot be made [3]. The evaluation is therefore about efficiency exclusively.

### 3.1 Investigation

The test of this research is a comparison of the efficiency between using interactive notifications as opposed to opening the application UI and responding to a message there.

The evaluation method isolates efficiency a single aspect of the usability. Efficiency is explicitly related to the hypothesis that interactive notifications saves time and it is therefore most relevant to the investigation. Due to time-constraints, the evaluation does not test the effectiveness or satisfaction of interactive notifications. Also because it is seen more important to properly evaluate the most relevant aspect of the usability than assuming too much of the prospects.

The investigation was narrowed down from the wide concept of notifications, by focusing solely on conversational applications. One reason is that interactive notifications could be especially beneficial for more quickly being able to reply to messages. Another part of it is how the trend of artificial intelligence (AI) and many service applications are moving towards having more of a conversational UI [2]. It is because of this even more relevant to reflect upon how user interaction is handled in critical moments of demanding attention to a conversation.

With reference to earlier research about usability testing on mobile devices, the tests were executed in a public environment but at a secluded area with privacy between the test user and the test instructor [9, 10].

A potential impact on the testing could be internet connectivity, since retrieving data to the application and notification requires internet connection. To minimize this influence, all tests were conducted in the same area on the same wireless network. It may still have fluctuated, but then it means the same random impact on all tests and this is acceptable. It is simply a partial factor that should be included for a mobile context [5]. The same goes for hardware and software performance. It could for example require more processing and thereby take longer to open the application UI than to use interactive notifications, which also is a partial factor that should not be outlined in the evaluation of efficiency.

### 3.2 Equipment

The hardware used for the tests was an Apple iPhone 6<sup>7</sup>. The operating system running on this phone was iOS 10<sup>8</sup> and the conversational application used

to test the notification interaction was Apple's standard messaging application Messages<sup>9</sup>. The messages in the test was sent from a computer and the time intervals was measured with a stopwatch.

The interaction design of the used operating system and application used in the tests needed to be ensured as a mobile UI with a sufficient user experience. It seems to follow usability guidelines of mobile UI design [12].

The hardware and software were easy to retrieve for the testing procedure. In order to save time, the tests were performed with already existing materials. The tools and resources used for the study was chosen out of suitability, accessibility and time-constraints for the research.

### 3.3 Test procedure

Forty randomly selected test participants were divided into two equally sized groups, where each group tried one of the two methods in interacting with mobile notifications. Group 1 tried the approach of interactive notifications. Group 2 tried the conventional approach of opening the application UI. All the instructions and text in the test was in Swedish, by reason of testing in Sweden and to then avoid confusion about language. The test contents has in this paper been translated to English.

Each test user was first asked about their smartphone usage and notification habits to see if they are familiar with notifications on a mobile device. Age and gender were documented to determine a diverse and representational crowd. They were ensured that it was an anonymous test and that it was something about the mobile device that was tested and not in fact themselves. The test users were evaluated separately. They were informed about what their task in the test was and also instructed of how they should interact with the notification before the test began. The test user instructions was as follows in table 1.

The participants received a notification with a text message including a question. They had got instructions about their task, which was to open the text message with the corresponding method and to reply to the question.

The question was meant to be of easy character that required no severe thought process to reply to. The aim was also that each answer would be of about the same length. The message was simple:

*"Hello, thank you for participating. This is a test that examines something about mobile devices. How old are you?".* After expanding the notification or opening the application such as in Figure 2, the different interfaces contrast as in Figure 3.

It was validated that the user had effectively completed the appointed task if they performed the right interaction and sent a logical reply to the message as

<sup>7</sup> <http://www.apple.com/se/shop/buy-iphone/iphone6s> Apple iPhone homepage, accessed 2016-10-27

<sup>8</sup> <https://support.apple.com/en-us/HT201287> Apple Messages application support page, accessed 2016-10-27

<sup>9</sup> <http://www.apple.com/se/ios/ios-10/> iOS 10, the iPhone operating system, accessed 2016-10-27

Introduction	"Try to pretend that this is your own mobile phone and that you are in your everyday situation. I will not help or answer questions during the test, please just focus on the task. You will get a notification with a text message that contains a question. Your only task is to open the message and answer the question. The phone has no password."
Instructions Group 1	"I want you to open the message by swiping the notification to the left and pressing show" (interactive notifications)
Instructions Group 2	"I want you to open the message by swiping the notification to the right."

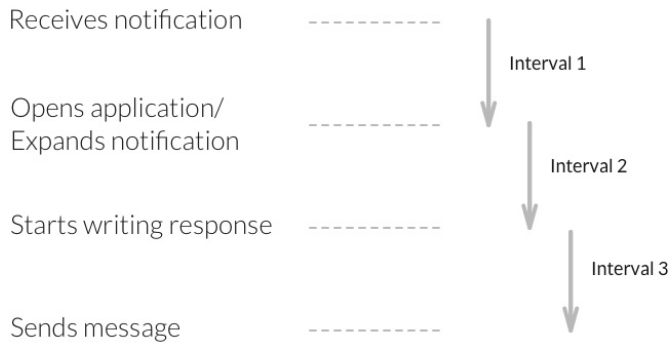
Table 1. User instructions before test.



Fig. 3. The interfaces of the two compared approaches to replying to a message from opening/expanding the notification. The application UI to the left and the Interactive notification to the right.



shown in Figure 3. Both groups got the same task and the same text message. The only difference was in how to open the notification and reply to the message. Each person tried only one of the methods of the notification interaction so that they wouldn't know the question. The time it took them to complete their assignment was measured in three intervals, see Figure 4. The periods of time were manually measured with a stopwatch. The margin of human error from the measurements is negligible since it was the same for all cases and should for that reason not have a noticeable impact on the comparison.



**Fig. 4.** The measured time intervals between user inputs.

The total task completion time is the sum of the three measured intervals. It is however also relevant to see if there is a difference between the interaction methods in how long it takes for the user to start writing after the message is received, regardless of for how long the users will type. The sum of the first two intervals can be seen as the task completion time but with typing speed as a redundancy. In the same way, each interaction interval could be compared individually.

## 4 Results

The distribution of test users were 48.8% men and 51.2% women, in ages varying between 18 and 36 years. All participants were regular smartphone users and they were used to receiving notifications. The different smartphone users were 30% Android-users<sup>10</sup> and 70% iOS users in their daily life. Each method of interacting with notifications was tested with 20 participants. The total task completion time was compared between the groups, see Figure 5. Each separate time interval of the interactions was also compared, see Figure 6 to 8. Group 1 used interactive notifications and Group 2 used the conventional approach.

A statistical analysis of the data was done to see whether the difference between the total task completion time of the two methods is statistically significant. A two-sample t-test was carried out with a confidence interval of 95% and

thus a significance level of 0.05. The result was that there is no significant difference. Analyzing each measured interval for the methods gave the same result. Thus, any combinations of the intervals would also give the same result. Hence, it is not valid to say that either of the evaluated methods is more efficient than the other in interacting with messaging notifications. In fact, the difference was not noticeable for any reasonable significance level (even up to 0.2).

Descriptive Statistics: Total

Group	N	Mean	StDev	SE Mean
1	20	13,138	4,728	1,057
2	20	12,9670	3,3615	0,7517

**Fig. 5.** The mean, standard deviation and standard error (of mean) of the total task completion time by Group.

Descriptive Statistics: Interval 1

Group	N	Mean	StDev	SE Mean
1	20	3,3995	1,9653	0,4394
2	20	3,2665	1,4834	0,3317

**Fig. 6.** The mean, standard deviation and standard error (of mean) of time in interval 1 by Group, see Figure 4.

Descriptive Statistics: Interval 2

Group	N	Mean	StDev	SE Mean
1	20	5,5725	2,3763	0,5314
2	20	5,4015	1,9827	0,4433

**Fig. 7.** The mean, standard deviation and standard error (of mean) of time in interval 2 by Group, see Figure 4.

It was also documented what operating system and hardware device that each test user was familiar with. There were only two occurring cases of earlier mentioned Android and iOS. What OS the test users was used to proved to have no considerable effect on the results, because the type of OS that the test users had did not have a statistically significant effect on the task completion time in either of the groups (tested with significance levels up to 0.1), see Figure 9.

<sup>10</sup> [https://www.android.com/intl/sv\\_se/](https://www.android.com/intl/sv_se/) Android, an open source operating system, accessed 2016-11-21

**Descriptive Statistics: Interval 3**

Group	N	Mean	StDev	SE Mean
1	20	4,2160	2,4082	0,5385
2	20	4,2990	2,4497	0,5478

**Fig. 8.** The mean, standard deviation and standard error (of mean) of time in interval 3 by Group, see Figure 4.

**Descriptive Statistics: Total**

Smartphone user	N	Mean	StDev	SE Mean
android	12	14,4558	3,2735	0,9450
iphone	28	12,4511	4,2515	0,8035

**Fig. 9.** The mean, standard deviation and standard error (of mean) of the total task completion time by smartphone user.

## 5 Discussion

Since there is no significant difference in efficiency between the evaluated methods of interacting with notifications in conversations, we are entitled to dispute the purpose of using interactive notifications at all. The question is if the apparent lack of improvement with the methodology of interactive notifications lies with the idea of dynamic interfaces in general, or if the used software in the study is an inadequate representation of the goal with interactive notifications.

We argue that there is more to be done in terms of creating a good enough dynamic interface of notifications. The version of interactive notifications used in the tests (iOS 10) is still too similar to the conventional method of actually opening separate UI:s. A dynamic interface will not be preferable unless it utilizes a completely different interface methodology. The dynamic interface needs to take advantage of its benefits in a greater extent.

One claim could now be that conversational UI:s are shown to be badly suited for being integrated in a dynamic interface. We believe that the problem is not in fact with the suitability, but with the current integration. Conversational UI:s and dynamic interfaces could go hand in hand. But maybe the interface has to be completely different to achieve the objective in usability. It could be that voice control is a better suited alternative as an interaction method for dynamic interfaces, which many previously mentioned commercial AI systems seems to indicate, see section Background.

### 5.1 Limitations

This study was a general investigation about the efficiency of interactive notifications, in the perspective of usability. It was not an evaluation of templates in monetary design. It can be argued that involving familiar mobile platforms and

applications in the tests invites biased views of brands and design. Due to previous user habits and earlier phone usage, this could have been disadvantageous if it had potential to affect the test results. However, the tests does not cover the satisfaction and attitude towards the content and neither does the test evaluate the specific design. It only measures the efficiency of two typical approaches to notifications. Beforehand, all user test participants got clear instructions about the simple interaction with the notification that was included in the task. Each user got only one task, which meant only one navigation step and one typing input. In this way, the risks with using established software and hardware were minimized. It was also proved that this limitation did not crucially affect the outcome, see Figure 9.

The benefits of choosing widely used versions of mobile software for conducting the tests were that they are well implemented and adapted to this very use case. Furthermore, how interactive notifications already work today is the most important thing to evaluate, since the design might be a stepping stone of how these methods will evolve in the future. It then needs to be questioned if the usability is prioritized when moving towards a more unified UI, and the efficiency attribute is a part of this.

## 5.2 Future work

More studies need to be conducted in the comparison of dynamic interfaces versus the conventional pattern of multiple UI:s to be sure of how the user experience could change, this for mobile devices. Implementing new solutions without a profound reason is not in line with sustainable technical development.

This research was solely about efficiency and how the potential can be seen in terms of time. More aspects of usability need to be evaluated before an overall statement of the usability in dynamic interfaces can be made, especially the qualitative type of testing. Even if there is no difference in efficiency between the methods, as shown, it could still be preferable to use according to other aspects of the user experience.

Relevant to the same research topic, it will also be important to further look at how dynamic interfaces should be shaped and if completely divergent types of interfaces give a better user experience, like comparing voice interaction with touch interaction.

## 5.3 Conclusion

The potential of a unified and dynamic interface is still present, but in the perspective of using interactive notifications, the evaluated case might not be a step forward. In any case, we cannot claim that the concept of a dynamic user interface is going to improve the general user experience, we can only make confirmations about usability conditions in terms of efficiency.

The ambition is for this research to serve as a good indicator of what role efficiency plays as part of the user experience on a mobile interface. It can give insight in how dynamic interfaces should be addressed in design.

Interactive notifications for messaging on mobile devices today is no more or less efficient than the conventional approach of notifications. Further extensive research is required to see if dynamic interfaces are advantageous in comparison to the current alternative methods.

## 6 Acknowledgements

We would like to thank the test participants of the study and the peer reviewers throughout the work of this research. A special appreciation goes to the anonymous reviewer of the editorial board at the student conference for valuable feedback before the final revision of this paper.

## References

- [1] Adams, P.: The end of apps as we know them. <https://blog.intercom.com/the-end-of-apps-as-we-know-them/> (2014)
- [2] Connolly, E.: Messaging apps are just getting started. <https://blog.intercom.com/messaging-apps-just-getting-started/> (2015)
- [3] Frøkjær, E., Hertzum, M., Hornbæk, K.: Measuring usability: are effectiveness, efficiency, and satisfaction really correlated? In: Proceedings of the SIGCHI conference on Human Factors in Computing Systems, ACM (2000) 345–352
- [4] Nielsen, J.: Introduction to usability. <https://www.nngroup.com/articles/usability-101-introduction-to-usability/> (2012)
- [5] Zhang, D., Adipat, B.: Challenges, methodologies, and issues in the usability testing of mobile applications. *International Journal of Human-Computer Interaction* **18**(3) (2005) 293–308
- [6] Hornbæk, K.: Current practice in measuring usability: Challenges to usability studies and research. *International journal of human-computer studies* **64**(2) (2006) 79–102
- [7] Vinh, K.: What is a card? <https://www.subtraction.com/2014/08/26/what-is-a-card/> (2014)
- [8] Wilson, F.: App constellations. <http://avc.com/2014/05/app-constellations/> (2014)
- [9] Nielsen, C.M., Overgaard, M., Pedersen, M.B., Stage, J., Stenild, S.: It's worth the hassle!: the added value of evaluating the usability of mobile systems in the field. In: Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles, ACM (2006) 272–280
- [10] Kaikkonen, A., Kekäläinen, A., Cankar, M., Kallio, T., Kankainen, A.: Usability testing of mobile applications: A comparison between laboratory and field testing. *Journal of Usability studies* **1**(1) (2005) 4–16
- [11] Nielsen, J.: Quantitative studies: How many users to test? <https://www.nngroup.com/articles/quantitative-studies-how-many-users/> (2006)
- [12] Gong, J., Tarasewich, P.: Guidelines for handheld mobile device interface design. In: Proceedings of DSI 2004 Annual Meeting. (2004) 3751–3756



# Comparison Of Geometric Approach With Artificial Neural Network Approach For Inverse Kinematics Problem For AL5D Robotic Arm

Vishnupradheep Raveendran

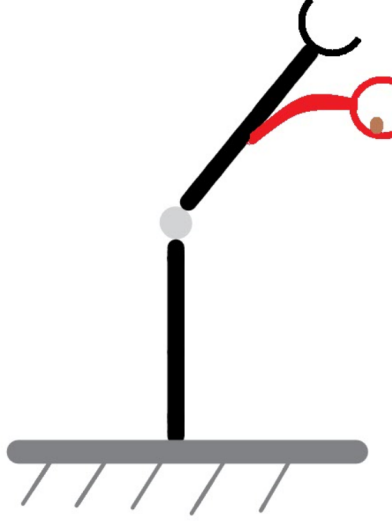
Department of Computing Science  
Umeå University, Sweden  
`mrc12pvn@cs.umu.se`

**Abstract.** One of the main problems in robotics is to find the inverse kinematics for the robot manipulator. When the robotic arm is floppy like AL5D robotic arm, the traditional way of mathematically modelled geometric solutions for inverse kinematics problem leads to inaccurate solutions. This gives rise to the research question, is it possible to achieve more accurate estimation of robot's end-effector position for floppy robotic manipulator like AL5D, using Artificial Neural Network approach than geometric approach for inverse kinematics problem. Result from the experiment, concludes that the Artificial Neural Network gives more accurate estimation of robot's end-effector position for floppy robotic manipulator like AL5D than geometric approach for inverse kinematics problem.

## 1 Introduction

A robotic arm consists of several rigid bodies called links that are connected by various joints which move relative to each other. The last links of a robot, which are designed to interact with the environment, are called end-effector. The degrees-of-freedom (DOF) of robotic arms, are defined by how many links can rotate or move along a straight line around the fixed axis.

One of the main problems in robotics is to find the inverse kinematics for the robot manipulator. Inverse kinematics refers to the use of the kinematic equations of the robot to determine the joint angles that provide a desired position of the end-effector [6]. One of the most general problems in robotics is the higher degrees of freedom manipulators. When the number of manipulators degrees of freedom increases, the traditional way of mathematically modelled geometric solutions for inverse kinematics problem leads to inaccurate or no physical solutions [5]. Therefore alternative methods like Artificial Neural Network (ANN) are widely used for inverse kinematics and modelling in robotics [2, 13-15], which is independent of the number of degrees of freedom. For example, ANN is used in both 2DOF planar manipulator and 3DOF planar manipulator [2] to solve inverse kinematics problem. Complexity arises even further when the robotic arm is floppy. Floppy robotic arm is the arm which has a not-so rigid links, which



**Fig. 1.** Floppy robotic arm which shows, 2 links (black color), a joint (grey colored circle) and a flexibility of the link 2 (red color).

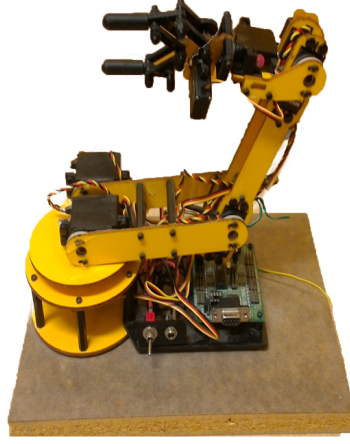
can bend due to its flexibility. For an example, link 2 (red color) in Figure 1 is bended slightly due to its flexibility. Some robotic arms like AL5D (Figure 2), are usually floppy compared to the Industrial robotic arms, which make the arm unstable. Instability is caused by a drop in the angle of each link. This gives rise to the research question, is it possible to achieve more accurate estimation of robot's end-effector position for floppy robotic manipulator like AL5D, using ANN approach than geometric approach for inverse kinematics problem.

The proposed approach consists of inverse Kinematic solution by inverse geometric model and for ANN consists of multi-layered perceptron (MLP) with a back propagation training algorithm. ANN are trained with end-effector position and desired joint angles for particular end-effector position[5]. MLP is used because joint angles are non-linear. The end-effector position accuracy comparison for these methods are made by taking the difference between the actual position attained by the robot end-effector and the desired position of the robot end-effector. Similar comparisons have been made for 6DOF complex offset wrist industrial robotic arm [3], but here this paper differs by using ANN for floppy robotic manipulator.

## 2 Background and Method

Robot kinematics is the study of the motion of robots [5]. Robot kinematics are mainly of two types : forward kinematics and inverse kinematics. In forward kinematics, the length of each link and the angle of each joint are given and





**Fig. 2.** AL5D robotic arm.

the position of the robot's end-effector is calculated [5]. In inverse Kinematics, the position of the end-effector is given and the angle of every joint should be calculated.

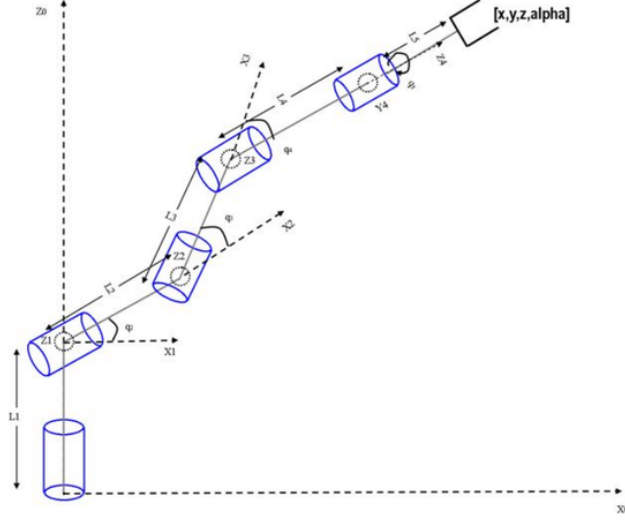
In the Figure 3, a joint is illustrated by a point inside the circle where the frames are attached, ( $x, y$  and  $z$  axis) and links are illustrated by using lines and cylinders. The " $x, y, z$  and  $\alpha$ " notation at the last semi cylinder of Figure 3 illustrate position and orientation of the end-effector respectively.

## 2.1 Forward kinematics

The forward kinematics problem is concerned with the relationship between the individual joints of the robot manipulator and the position and orientation of the end-effector [6]. The idea of forward kinematic is to keep track of rotation and translation (function that moves every point a constant distance in a specified direction[8]) by moving from the base frame ( $x_0, z_0$  axis in Figure 3) to the end-effector frame ( $y_4, z_4$  axis in Figure 3), which calculates the position and orientation of end-effector.

To keep track of, one needs to attach a coordinate frame ( $x_i, y_i$  and  $z_i$  axis in Figure 3) to each link. A commonly used convention for attaching coordinate frames in robotics is the Denavit-Hartenberg, also known as DH convention. Based on [6], if DH convention is to be used, the following rules must be followed at all times when attaching coordinate frames:

- $z_i$  axis should be along the actuation line of the current link
- $x_i$  (current link  $x$  axis) should intersect with  $z_{i-1}$  (previous link  $z$  axis)
- $x_i$  (current link  $x$  axis) should be orthogonal to  $z_{i-1}$  (previous link  $z$  axis)



**Fig. 3.** Configuration of AL5D robotic arm, blue colored cylinder represent links, black colored dotted circle and black colored dotted line represent coordinate frame.

For an example in Figure 3,  $z_2$  coordinate frame of link-2 is attached along the actuation line, which is the circle in link-2 (in 3D view  $z$ -axis projects out of the circle).  $x_2$  coordinate frame of link-2 is attached in the place where, both  $x_2$  intersect with  $z_1$  and  $x_2$  orthogonal to  $z_1$  as shown in Figure 3.

It can be observed that after attaching coordinate frames, according to [6] four DH parameters ( $d, \theta, a, \alpha$ ) should be assigned for each link to keep track of rotation and translation from the base frame to the end-effector frame. An example for calculating DH parameters for the second link in Figure 3 is as follows:

- $a_2$  : the link length  $l_2$  is the distance between the current origin  $o_2$  and the action axis of the previous frame  $z_1$
- $\alpha_2$  : the link twist is the constant angle between the unit vectors defining the previous and current action axis,  $z_1$  and  $z_2$
- $d_2$  : the link offset is the constant distance from previous origin  $o_1$  to current  $x_2$  axis
- $\theta_2$  : the link angle which is the constant angle between the unit vectors  $x_1$  and  $x_2$ .

Figure 4, illustrate the derived DH parameters for all the links of AL5D robotic arm.

The next step is to derive the homogeneous transformation (transformation between two frames can be defined by translation and rotation) for each link, with respect to the previous link. This can be done using the equation

$$A_i = Rot_{z,\theta} \cdot Trans_{z,d_i} \cdot Trans_{x,a_i} \cdot Rot_{x,\alpha}$$

where,

- $Rot_{z,\theta}$  is the rotation of current frame  $z$ -axis by current joint angle theta,
- $Trans_z, d_i$  is the translation of current frame  $z$ -axis by the current link offset  $d$ ,
- $Trans_x, a_i$  is the translation of current frame  $x$ -axis by the current link length  $a$ ,
- $Rot_{x,\alpha}$  is the rotation of current frame  $x$ -axis by the current link twist angle  $\alpha$  and
- $\cdot$  is the dot product among  $Rot_{z,\theta}$ ,  $Trans_z, d_i$ ,  $Trans_x, a_i$  and  $Rot_{x,\alpha}$ , gives the position and orientation of the end-effector ( $A_i$ )

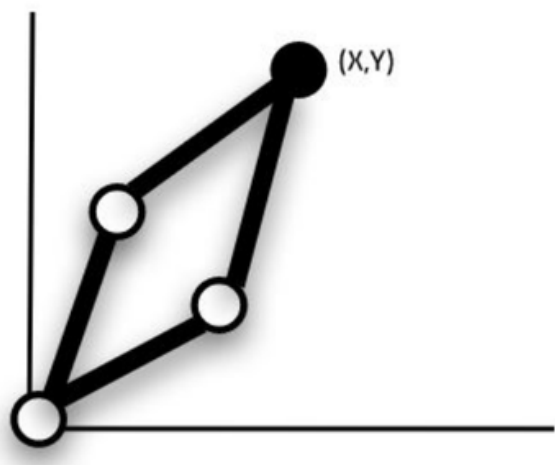
Link	Joint variable ( <i>Theta</i> )	Link Offset ( <i>d</i> )	Link length ( <i>a</i> )	Link Twist ( <i>Alpha</i> )
1	$q_1$	L1	0	$\text{Pi}/2$
2	$q_2 - \text{Pi}/2$	0	L2	0
3	$q_3$	0	L3	0
4	$q_4 - \text{Pi}/2$	0	0	$-\text{pi}/2$
5	$q_5$	L4+L5	0	0

**Fig. 4.** DH table for AL5D robotic arm, for 5 links.

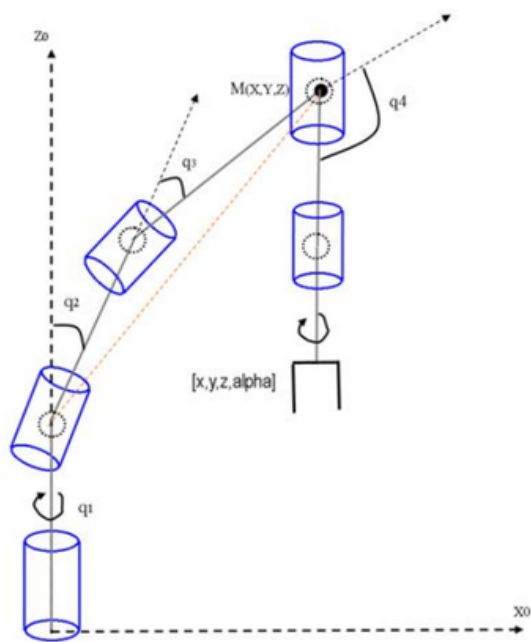
## 2.2 Geometric solution for Inverse Kinematics

The inverse kinematics is defined as the procedure of finding joint angles, from the position and orientation of end-effector. Inverse kinematics is calculated based on pure geometry, but unlike forward kinematics, the geometric solution for inverse kinematics gives many solutions, except in the case where the robot manipulator configuration has one degree of freedom. As seen in Figure 5, the two degrees of freedom manipulator's (two-link elbow robot) end-effector position  $(x, y)$  can be reached through two solutions : elbow up solution and elbow down solution. Therefore the number of solution increases as the degrees of freedom increases, as in AL5D robotic arm.

From the number of solutions, one solution is used according to the AL5D robotic arm configuration, for inverse kinematic geometric approach as shown in Figure 5. The general idea of the geometric approach is to solve for joint angles by projecting the manipulator onto  $x_{i-1}, y_{i-1}$  plane and solving a simple trigonometry problem [6]. According to Figure 6, to solve for  $\theta_1$ , the arm (end-effector position) is projected onto the  $x_0, y_0$  plane which gives,  $\theta_1 = \tan^{-1} (y_0, x_0)$ . Similarly  $\theta_2, \theta_3, \theta_4$  and  $\theta_5$  are solved, which finally gives the joint angle for all links.



**Fig. 5.** Two degrees of freedom manipulator, with 2 links and 2 joints (black colored circle).

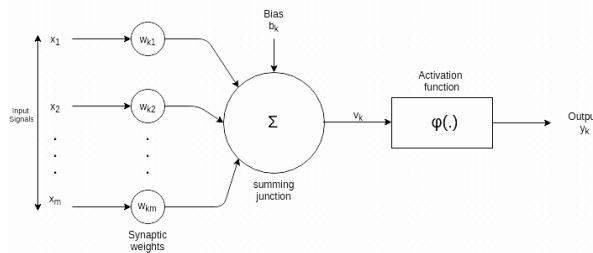


**Fig. 6.** Configuration of AL5D robotic arm for solving Inverse kinematics, blue colored cylinder represent links and black colored dotted circle represent joints .

### 2.3 Artificial Neural Network

Neural Networks are an adaptive statistical model based on an analogy with the structure of the brain [9]. They are adaptive in that they can learn to estimate the parameters by getting the set of training data. The goal of this type of network is to create a model that correctly maps the input to the output using historical data so that the model can then be used to produce the output when the desired output is unknown [5]. The main advantage of neural networks is that it is massively parallel and can act as both linear and non-linear learning.

Neural Networks are built from simple units called neurons [9]. Summing junction in Figure 7 represents neuron or unit. These units are interlinked by a set of synapses (Figure 7, input signals), each of which is characterized by a weight or strength of its own [10] called synaptic weight (Figure 7). Synaptic weight decides the strength of the input signals (larger the weight, more stronger the connections). The input and their weights are linearly combined by a summing junction (Figure 7), by summing all the input signals multiplied by their weights. Linearly combined inputs and their weights are sent to the activation function (Figure 7), which limits the amplitude of the input of a neuron. The activation function is also referred to as a squashing function as it squashes the permissible amplitude range of the output signal to some finite value [10]. For an example, activation function reduces the output to 1 if the result of summing junction is greater than some threshold and -1 otherwise. Therefore if the output gives the value -1, the neuron considered as the negative feedback, calculates the error by taking the difference between actual value (which is -1 here) and the targeted value (which is 1 here). Then the weights are updated by, the current weight of the inputs multiplied by the error. The same process is repeated until the output function gives the value 1 and this process is called learning.

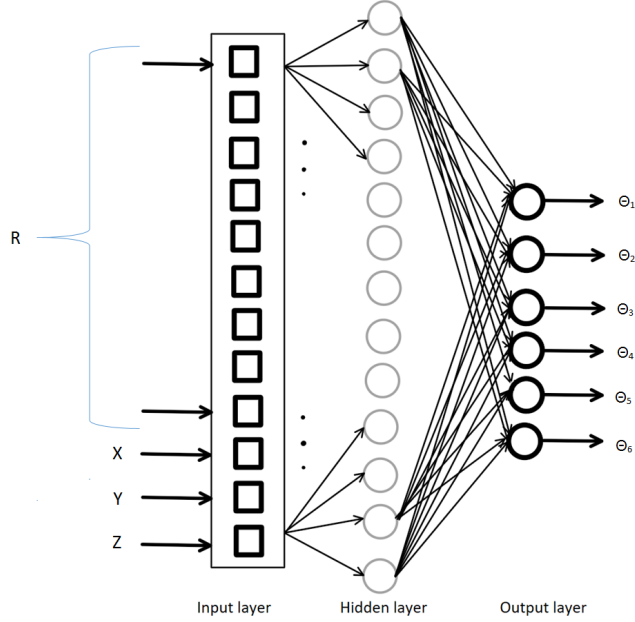


**Fig. 7.** Model of a simple Neural Network .

As most of the problem in this world are analog (continuous), the binary threshold function fails. linear activation model is used to over come this problem, where a curve is fitted in a appropriate place among the data. Being a continuous function, one of the biggest advantages of the linear activation function over the unit step function is that it is differentiable [11]. This property allows to define a cost function that we can minimize in order to update the weights [11]. In the case of the linear activation function, we can define the cost function

as the sum of squared errors (SSE), which is similar to the cost function that is minimized in ordinary least squares (OLS) linear regression [11]. In order to minimize the SSE cost function, a gradient descent optimization algorithm is used, to find the local minimum of linear systems [11]. That is the gradient descent optimization algorithm, find weights with low cost so that the output from the neural network approximates for all training inputs. Detailed explanation about gradient descent optimization algorithm is explained here [12].

Multilayer perceptron or multilayer ANN (Figure 8) is the non-linear Neural Network, used in this paper as the joint angles are non-linear. Multilayer perceptron are similar to the linear or single layer ANN. In multilayer perceptron the error is first computed at the last node of the ANN (in Figure 7, output layer). Then the errors that are calculated at the output layer are propagated to the hidden layer (Figure 8) and again the errors are calculated in the hidden layer. Further errors from the hidden layer are back propagated to the input layer and again the error is calculated. After that the SSE is minimized.



**Fig. 8.** Neural network structure for AL5D Robotic Arm .

### 3 Implementation

Implementation is done by using AL5D robotic arm and MATLAB. Neural Network tool box in MATLAB is used for implementing ANN.

### 3.1 Geometric solution

The end-effector of AL5D robotic arm which is in initial position, moved to a new position by moving links of the robot manually - tele operated, Figure 9, column 1. For moved position, the end-effector's position and orientation are calculated by forward kinematics (section 2.1) and is named as actual position, Figure 9, column 2. Given the end-effector position and orientation from the forward kinematics, the joint angles are calculated by geometric solution for inverse kinematics (section 2.2), Figure 9, column 3 (named as JIA).

Tele-operated joint angles	Forward kinematics for tele-operated joint angle  <i>(Actual position)</i>	Joint angles from Inverse kinematics, giving <i>Actual position</i> as the input  <i>(JIA)</i>	Forward kinematics for <i>JIA</i>  <i>(Desired position 1)</i>	Percentage difference between <i>Actual position</i> and <i>Desired position 1</i>	Joint angles from ANN, giving <i>Actual position</i> as the input	Forward kinematics for the joint angles from ANN  <i>(Desired position 2)</i>	Percentage difference between <i>Actual position</i> and <i>Desired position 2</i>
link 1= 70°	X= -8.0523  Y= -22.123  Z= 9.0980	link 1= 70°	X= -7.7764  Y= -21.36  Z= 16.29	X= 3.486 %  Y= 3.485 %  Z= 56.67 %	link 1= 70°	X= -7.8248  Y= -21.943  Z= 9.1791	X= 2.874 %  Y= 0.818 %  Z= 0.887 %
link 2= 32°		link 2= 16°			link 2= 32°		
link 3= 28°		link 3= 33°			link 3= 28°		
link 4= 88°		link 4= 72°			link 4= 88°		
link 5= 0°		link 5= 0°			link 5= 0°		

**Fig. 9.** Comparison of both geometrical solution and ANN for a single position .

After calculating the joint angles of that new position, the robot is moved back to the initial position manually.

The end-effector which is in initial position is made to move by performing tele-operation on the resultant joint angles obtained from the geometric solution of that new position. The resultant end-effector position from tele-operation is calculated from the forward kinematics and is named as desired position 1, Figure 9, column 4.

### 3.2 Artificial Neural Network

The structure of ANN for AL5D robotic arm, to solve inverse kinematics is shown in Figure 8. The training data set is composed of 9000 input-output parameters. That is, 18 input-output parameters are taken 500 times, for every 5 degree joint angles. The 18 input-output parameters are taken by moving the end-effector of AL5D robotic arm to many new points by tele-operation. For every tele-operated new position, the end-effector's position and orientation are calculated by forward kinematics (section 2.1).

The input vectors of the ANN have 12 parameters. 9 parameters are from rotational matrix (Orientation - Figure 8, shown as 'R' in the input layer) of the homogenous transformation (section 2.1) and 3 parameters are from position vector (X,Y,Z). The output vector or target vector have 6 parameters (Figure 8, output layer). After several trial and errors, one hidden layer with 13 hidden nodes was decided as the hidden layer architecture for the ANN (Figure 8, hidden layer). The input vectors and corresponding target vectors from the training data set are trained by adjusting the weights. Training is terminated when SSE is minimized to 0.03. Weights are initially assigned randomly and are adjusted using gradient descent learning algorithm ('traingd' - as a training function in matlab).

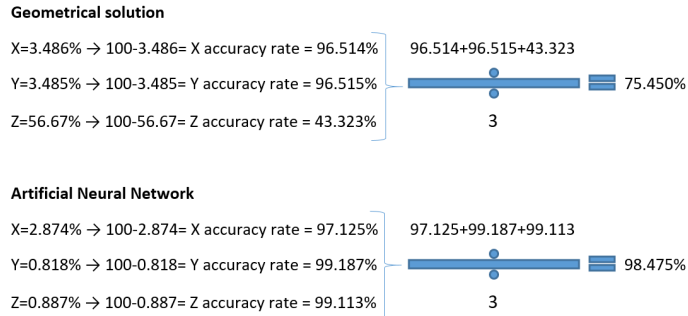
The same end-effector new position which is used in geometric solution are given as an input (x,y,z) to the ANN (Figure 8). ANN gives the output in the form of joint angles, Figure 9, column 6.

The end-effector which is in initial position are made to move by performing tele-operation on the resultant joint angles obtained from the output of an ANN. The resultant end-effector position from tele-operation is calculated from the forward kinematics and is named as desired position 2, Figure 9, column 7.

## 4 Result

The end-effector position accuracy between geometric approach and ANN are made by comparing, the percentage difference between the geometric solution actual position and the desired position 1 (Figure 9, column 5), with the percentage difference between the ANN actual position and the desired position 2 (Figure 9, column 8).

Result is obtained by subtracting the percentage difference (Figure 9, column 5 and Figure 9, column 8) with the value 100, which gives the accuracy rate of X,Y,Z of both the methods (Figure 10).



**Fig. 10.** Accuracy rate of both geometrical solution and ANN for a single position .

The average of, accuracy rate of X,Y,Z are calculated (Figure 10), which gives the accuracy rate of both the methods. Figure 10 shows accuracy rate of geometrical



solution (75.450%) and accuracy rate of ANN (98.475%) for a single position. Similarly the average accuracy rate of both the methods are calculated for 85 positions.

From the result obtained (85 positions), geometric solution gives an average of 63% accuracy rate, in comparison with the ANN which gives about an average of 88% accuracy rate.

The average ANN error of the both 2DOF and 3DOF planar manipulator (example which was mentioned in the introduction) is 5%, when compared to ANN's 12% average error of 5DOF AL5D robotic arm. The 7% difference in error is due to higher DOF.

Higher DOF (5 DOF) in AL5D robotic arm, increases the average error percentage as compared to 2 or 3DOF. This is because (as mentioned in section 2) the number of solution for inverse kinematics increases with higher DOF. As ANN for AL5D robotic arm is trained with all possible solution (all possible joint angles with respect to the particular input) for inverse kinematics, 15% error is caused and it's called overfitting problem. Overfitting problem arises when a model (linear or non-linear) learns both the actual and noise data in the training data. Overfitting impacts negatively on the model performance or generalization ability on new data.

On the other hand, ANN for 6DOF complex offset wrist industrial robotic arm (section 1) gives 8.78 root mean square (RMS) error of joint angles [3], while ANN for 5DOF AL5D robotic arm gives 7.56 RMS error of joint angles (result of ANN for AL5D robotic arm is better than the result of ANN for complex offset wrist industrial robotic arm). RMS is calculated by taking the square root of, mean square error (MSE) of joint angles. MSE is calculated by taking the average of the square of the joint angle errors - difference between tele-operated joint angles (Figure 9, column 1) and joint angles from ANN (Figure 9, column 6).

## 5 Conclusion

So from the result obtained it is concluded that, ANN is more accurate than geometric solution in calculating Inverse kinematic for floppy AL5D robotic arm. The reasons behind why geometrical solution gives only 63% accuracy rate compared to 88% of neural networks are,

1. AL5D robotic arm has 5DOF. This higher degrees of freedom leads to inaccurate solutions as the complexity of solving geometrical solution increases.
2. Geometrical solution for inverse kinematics method, assumes that the configuration for the robotic arm is rigid (Figure 1, link 2 black color) without taking the floppiness (Figure 1, link 2 red color) into account, as geometric solution for inverse kinematics is meant for the rigid robotic arm like industrial robotic arm - a manipulator particularly for manufacturing and production industry to perform hard job. But in ANN as the joint angles and their corresponding end-effector position are fed for learning (which

automatically includes floppiness of the robotic arm), gives more accurate output than geometric solution.

Disadvantages of using ANN for solving inverse kinematics are,

1. Need of high number of data for training.
2. Overfitting problem

Future work is aimed at reducing Overfitting problem. Overfitting problem can be reduced by cross-validation - which will improve ANN's accuracy rate for inverse kinematics.

## References

- [1] Eimei OYAMA and Nak Young Chong, Arvin Agah, Taro MAEDA and Susumu TACHI. Inverse kinematics learning by modular architecture neural networks with performance prediction networks, Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference, pp. 1006 - 1012 vol.1, May 2001.
- [2] Guez, Ahmad. Solution to the inverse kinematics problem in robotics by neural networks, Neural Networks, 1988., IEEE International Conference, pp.617-624 vol.2, 24-27 July 1988.
- [3] Bingul, Z., H.M.; Oysu, C. Comparison of inverse kinematics solutions using neural network for 6R robot manipulator with offset. Computational Intelligence Methods and Applications, 2005 ICSC Congress.
- [4] Ali T. Hasana, N. Ismaila, A.M.S. Hamoudab, Ishak Arisc, M.H. Marhabanc, H.M.A.A. Al-Assadia. Artificial neural network-based kinematics Jacobian solution for serial manipulator passing through singular configurations., Journal Advances in Engineering Software, Volume 41, 359-367 February 2010.
- [5] Bassam Daya, Shadi Khawandi, Mohamed Akoum. Applying Neural Network Architecture for Inverse Kinematics Problem in Robotics.
- [6] Book, Robot Modeling and Control by Mark W. Spong, Seth Hutchinson, M. Vidyasagar .
- [7] Use of Artificial Neural Networks for the development of an Inverse Kinematic Solution and Visual Identification of Singularity Zone(s) Luv Aggarwal, Kush Aggarwal, Ruth Jill Urbanic.
- [8] Osgood, William F and Graustein, William C. (1921). Plane and solid analytic geometry. The Macmillan Company. p. 330.
- [9] Tech report, Neural Networks, a SAGE University paper
- [10] Neural Networks, a comprehensive foundation, Second Edition, Simon Haykin
- [11] Article, Artificial Neurons and Single-Layer Neural Networks, How Machine Learning Algorithms Work Part 1, written by Sebastian Raschka on March 14, 2015
- [12] Book, Deep Learning by Ian Goodfellow, Yoshua Bengio and Aaron Courville , Chapter 4

- [13] R. Koker, C. Oz, T. Cakar, and H. Ekiz, "A study of neural network based inverse kinematics solution for a three-joint robot," *Robotics and Autonomous Systems*, Vol. 49, pp. 227-234, 2004.
- [14] L. Wei, H. Wang, and Y. Li, "A new solution for inverse kinematics of manipulator based on neural network," *Machine Learning and Cybernetics*, Vol. 2, pp. 1201-1203, 2003.
- [15] J. Guo and V. Cherkassky, "A solution to the inverse kinematic problem in robotics using neural network processing", *International Joint conference on Neural Networks*, Vol. 2, pp. 229-304, 1989.



# How Animations in a Mobile Web Application Impact User Interaction

Klara Sporrang

Department of Computing Science  
Umeå University, Sweden  
klsp0002@umu.se

**Abstract.** This study investigates if the use of animations in a web application can improve user interaction in terms of speed to perform specific tasks. Animations are often used in applications today but there is lack of proof that animations improve user interaction. This study is based on user tests of two identical mobile web applications with the only difference that one application contained animations and the other one did not. Five types of animations were tested and the result showed that animations could improve user interaction in some of the test areas.

## 1 Introduction

The use of animations is widely spread in web applications today. Some developers use animations to improve user interaction and to catch user attention, whereas others just use animations to make an application more "fun". Animations should only be used if the animation improves user interaction and understanding of the application [1].

Research indicates difficulties in proving that animations have a positive impact on user interaction [2]. This is a problem since animations are widely spreading throughout applications today. It seems crucial to be able to certify that the developer of an application have selected proper animations in order to improve user interaction since it's known that some animations can distract the user [1]. This could mean that animations not suited for a specific application or task could lower the user interaction.

This paper contains a study on how animations in a web application affect the user interaction in terms of speed to perform tasks and navigate through an application. Furthermore, this paper also presents general information about animation in graphical interfaces, how animations are used, an introduction of different types of animations, and what advantages and disadvantages animations have. The resources used are a combination of papers with similar studies, papers on animation technology and guidelines on how to use animations on the web.

## 2 Hypothesis

The aim of this study is to investigate if animations in a mobile web application can improve the users' speed to perform specific tasks. The question asked is: *Can animations in a mobile web application interface improve user interaction?*

This study focuses on user interaction in terms of speed to perform different tasks. The hypothesis is that simple animations, directly mapped to the user interface and the specific tasks to be performed, can improve the users' speed to perform the tasks.

## 3 Background

It is very common that applications contains animations today. Most people might not notice animations, but supposedly they help users to understand the application and how to interact with it.

There exists guidelines on how animations should be used in a web page or an application [1]. One other common resource of animation technology guidelines is Google Material Design<sup>1</sup>. Google encourage developers and designers to use natural motion to improve user interaction in applications. The guidelines in [1] together with Google's Material Design guidelines form the basis for this study as well as the basis in the creation of the user tests. This paper is not evaluating Google's Material Design guidelines specifically, it is rather an evaluation of earlier work guided by Material Design together with other animation guidelines.

To make sure that the animations that are used in an application improves the user interaction, the existing guidelines which are available on the web, have to be followed and evaluated. This section presents earlier work regarding animation technology, and an introduction to Google Material Design.

### 3.1 Earlier Work on Animation Technology

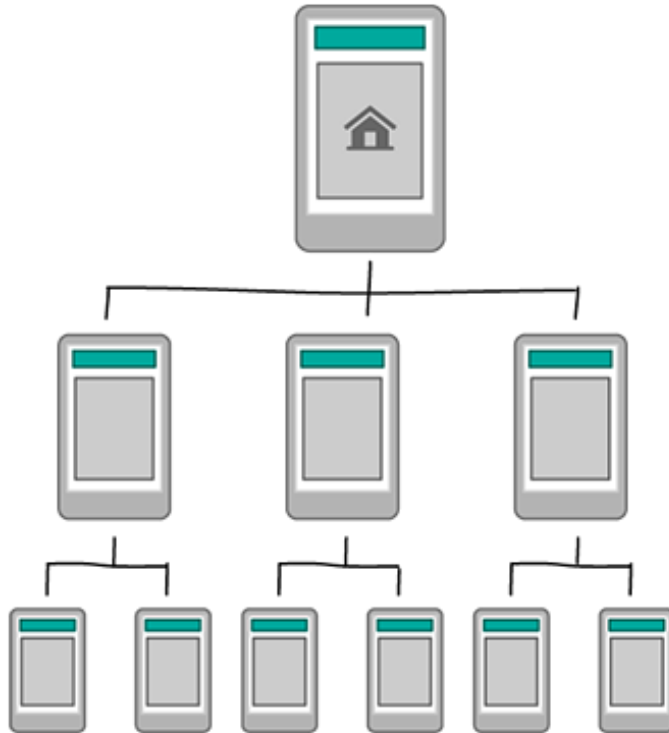
The author in [3] claims that users are more likely to be affected negatively by animations when they are perceived as a secondary stimulus, while performing an information seeking task. Animations perceived as secondary stimulus are animations which are not connected to the main interaction of an application or to the current task. In [3] secondary stimulus is referred to animated online advertising. The author also claims that a user is less affected by animations as a secondary stimulus if the task is more complex [3], which could be a result from the user being more focused on a more difficult task and therefore tend to be less distracted. With all this in mind one could propose that animations which are included in the main interaction of an application (primary stimulus), and which are similar to the task, could improve the user interaction and speed.

An important aspect to consider when using animation in a graphical interface is to only use it when it can provide better interaction and understanding of

---

<sup>1</sup> <https://material.google.com>, Google Material Design, accessed 2016-09-21

the application [1]. Animations can be used as cues to emphasize the structure of an application [4] (see Fig. 1-2). A good example of this is a swipe animation between views which provides a cue that the views are at the same level in the application hierarchy (see Fig. 3). To show that a view is the child of another view a "slide over" animation can be used (see Fig. 4).



**Fig. 1.** Application structure: Hierarchical view.<sup>2</sup>

Animations can also give hints on how an element should be interacted with or how this element can behave [4], for example by showing with an animation which direction the element can be moved.

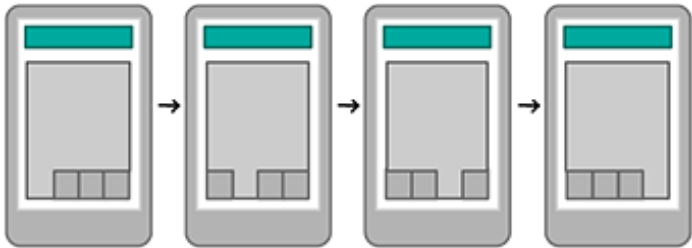
The order and relation between elements in an animation is also an important aspect to consider. Animations can provide cues for how elements are related to

<sup>2</sup> <http://www.uxbooth.com/articles/designing-for-mobile-part-1-information-architecture/>, UX Booth website, accessed 2016-10-20

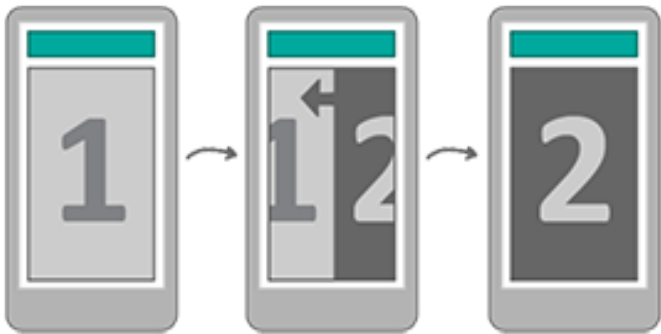
<sup>3</sup> <http://www.uxbooth.com/articles/designing-for-mobile-part-1-information-architecture/>, UX Booth website, accessed 2016-10-20

<sup>4</sup> <http://www.uxbooth.com/articles/designing-for-mobile-part-2-interaction-design/>, UX Booth website, accessed 2016-10-20

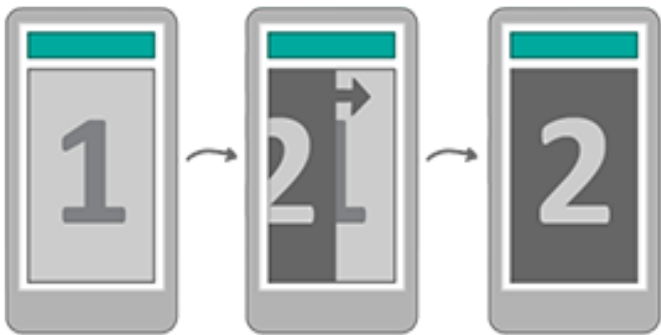
<sup>5</sup> <http://www.uxbooth.com/articles/designing-for-mobile-part-2-interaction-design/>, UX Booth website, accessed 2016-10-20



**Fig. 2.** Application structure: Tabbed view.<sup>3</sup>



**Fig. 3.** Transition: Swipe animation.<sup>4</sup>



**Fig. 4.** Transition: Slide over animation.<sup>5</sup>



each other and make these relations more salient [4]. The position and organization of an animated group of elements should follow the Gestalt law of proximity [1] which refers to that related items should be animated together. This means that when an animation consists of more than one element, these elements has to be animated together with the same animation pattern. The animation could otherwise cause confusion.

Further, animations can also be used to show concepts of change [4]. An animation which is correctly positioned and organized can be used to add understanding and clarity after a specific interaction is performed [1]. This is often used in applications and web sites to show that a button is clicked or that a new view is opened.

It is easy to draw a user's attention to an animation since people are easily distracted by movements. Hence, it is important to only use animations where it is relevant to the given task and where it supports the preferable interaction to perform the task [1]. An animation should not have irrelevant movements since this can distract and confuse the user [4][5]. If the animation is done right it can with advantage be used to grab the users' attention, and guide the attention to specific locations on the screen. One of the most efficient ways of doing this is to use an animation which suddenly appears on the screen, such as a pop-up or a notification banner. Another attention taker is an animation which moves over the screen [4].

The duration of an animation is also something that should be taken under consideration. It is important that an animation is not too complex or too fast [1][4], but it should not be too long either. A too short animation duration can be confusing for the user, and a too long duration can make the user bored and cause the user to loose focus [1].

### 3.2 Google Material Design

In 2014 Google developed Material Design, which is a design language for the web. Material Design provides guidelines regarding animation and motion. It also provides design and layout guidelines. The name Material Design comes from the belief that real material such as paper can be used as a metaphor in web design to bring meaning to the user interface and increase understanding of the application or website.

Google claims that motion can provide meaning to an application or a website if it is done right. There is many use cases where motion can improve a user interface. One case is to use motion to change the users' focus when a new view is opened or when an object is changed within a view. Motion can also be used to show the hierarchy of an application, which refers to how the views are ordered and placed and also how elements are related to each other. An example of this could be a small square which grows and covers the main view when it is clicked. This animation implies that the square literally lies on top of the main view. Google also recommends to use animation and motion as a hint on how to interact with the user interface. Motion can also be used to keep the user informed when background processes are running, for example by showing an

animated loading circle or similar. Animations can also be used to bring focus to objects which needs the user's attention.<sup>6</sup>

## 4 Definition of Different Types of Animations

Five areas of animation technology were defined to simplify the creation of the user tests. The areas were defined based on the information in Section 3. Each defined area is connecting the guidelines provided by Google (see 3.2) with the other guidelines presented in 3.1. Thus, all areas can be found in both the Material Design guidelines as well as in the other guidelines studied. These connections were made to prove that the Material Design guidelines are connected to other guidelines. These five areas formed the basis of the development of the test applications. The five areas are presented below.

**Hierarchy and Structure.** The first defined area is animations used to show application structure and hierarchy. This area can be connected with the study in [1] and [4] which concludes that animations can be used show the relation between views and emphasize structure (see Fig. 1-4).

**Change and transformation.** The second area is animations to show when a object changes within a user interface. This is directly connected to the study in [1] and [4], where it is concluded that animations can be used to show concepts of change as well as transforming the users' attention to specific locations.

**Anticipation.** The third defined area is animations used as a hint on how to interact with the user interface. For example, an animation which gives a hint on how a object can or will behave. This is connected to the study in [4] where it is concluded that animations can provide instructive cues on how certain objects will, and can behave. In [1] it is concluded that animations can be used to support the preferable interaction to perform a specific task.

**Action and feedback.** The fourth area is animations used to show that an element has been interacted with. This type of animations are used to distract and/or inform the user that action has been taken and that background processes are running. This area is connected to [1] where it is concluded that animations should be used to provide understanding of the application and that animations should be connected to a given task or interaction. It is also connected to [4] and [5] where it is claimed that animations should not have irrelevant movements, which implies that a good animation is related to the source of interaction.

**Attention.** The fifth area is animations used to bring focus to a specific object which needs the users' attention. This is connected to [4] where it is said that animations are efficient to use when the users attention needs to be switched to a certain object or a certain area of the screen.

---

<sup>6</sup> <https://material.google.com>, Google Material Design, accessed 2016-10-13

## 5 Testing Method

To evaluate whether animations can improve user interaction and speed, a series of tests were conducted.

Two simple mobile web applications were created and used to perform tests. The applications were identical with the only exception that one contained animations and the other one did not. Five different tasks were created for the test subjects to perform. Each task was testing a specific type of animation. The tasks were based on the five areas presented in Section 4.

### 5.1 Testing Tools and Frameworks

The two test applications were built in HTML and JavaScript. In addition a few of libraries and frameworks were added. The frameworks used were jQuery<sup>7</sup>, jQuery Mobile<sup>8</sup>, Angular Material<sup>9</sup> and Materialize<sup>10</sup>.

jQuery is a JavaScript library which is created to simplify the JavaScript syntax. It is used to modify the HTML, DOM and CSS in a web page.

jQuery Mobile is a JavaScript framework which is optimized for touch based devices. It is based on basic jQuery but it has extended methods such as swipe and long press event handlers, which enables the development of touch optimized applications for the web.

Angular Material is a UI framework which is based on Material Design and which can be used in web applications built in Angular JS. Angular JS is a Model View Controller (MVC) framework for web applications developed by Google.

Materialize is a CSS and JavaScript framework based on Google Material Design. It contains a bunch of CSS components and JavaScript methods which are easy to use and which simplifies the implementation of Material Design in a web application.

### 5.2 Tests

The two test applications were based on the different types of animations defined in Section 4. The areas to be tested were:

- Hierarchy and structure
- Change and transformation
- Anticipation
- Action and feedback
- Attention

---

<sup>7</sup> <https://jquery.com/>, jQuery homepage, accessed 2016-10-12

<sup>8</sup> <https://jquerymobile.com/>, jQuery Mobile homepage, accessed 2016-10-12

<sup>9</sup> <https://material.angularjs.org/latest/>, Angular Material homepage, accessed 2016-10-24

<sup>10</sup> <http://materializecss.com/>, Materialize homepage, accessed 2016-10-12

The test applications were distributed to the test subjects based on the current number of completed test rounds. The distribution was made using a script that determined which test application that currently had the least number of completed test rounds. The application with the least amount of completed test rounds was presented to the test subject. A complete test round refers to a test round where all five tasks were completed.

The tests were conducted such that each task was presented to the test subject by showing a cover page with information before the start of each task (see Fig. 5, 10, 14, 16, 18). When the test subject had understood the task, a button was pressed to start the test for the specific task.

The test subjects performed the tests without any monitoring and the test results were saved in a database after each task was completed.

The two test applications were developed to be supported in the most common mobile web browsers. The two applications were tested on Google Chrome for Android, Safari for iOS as well as on the default web browser on a Samsung Galaxy S6 device.

### 5.3 Measurements

Each test application resulted in four different kinds of data for each task: navigation time to complete the task, number of clicks to complete the task, number of incorrect clicks and task completion rate i.e. could the user perform the task? The data was collected using a timer and a click counter which were integrated in the application, but not visible for the test subjects.

To evaluate the number of clicks used to complete each task, the click counter was used. Each task had a minimum number of clicks required to complete the task. The number of clicks were later compared between each test subject.

To measure the number of "incorrect clicks" (i.e. clicks not relevant to the task), the number of clicks performed on each task was compared to the minimum number of clicks required to complete the task.

Each task had a time frame of 60 seconds which the user had to stay within for the task to be considered as completed. If the user could not complete the task before the time ran out, the task was considered as not completed.

To measure the time spent to complete a task the timer was used. The timer started when the test subject had read the given instructions and pressed "start test". The timer stopped when the task was completed or if the maximal time frame were exceeded. Some tasks had a second information page which appeared half through the task to give further information on how to complete the task. In these cases the timer also was stopped as long as the second information page was open.

### 5.4 Test Subjects

The test applications were sent to 82 people in various ages. The author knew the test subjects personally. The test participants were automatically divided

into two equally sized groups based on the current amount of completed test rounds. One group performed the tasks using the application which contained animations. The other group performed the same tasks using the application which had no animations.

The test persons were not forced to perform the test and therefore did not all 82 persons participate.

The test persons were between 19 to 56 years old. The various ages was not considered as something that would change the test result. Since the test application was not created for a specific user group, the various ages of the test participants was considered as not relevant.

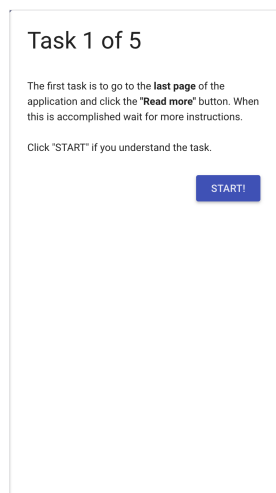
## 6 Test Tasks

Five tasks were developed to investigate the five areas defined in Section 4.

### 6.1 Task 1

The first task tested the area **Hierarchy and structure**. In this task the test subject was asked to navigate to the last of three pages lying in the same hierarchical level (see Fig. 5), and then down to the child page of the last page (see Fig. 6-7). When the child page was found, the test subject was asked to navigate back to the first page again (see Fig. 8). In the animated version the transition when the child page was opened was animated with a slide over animation (see Fig. 9).

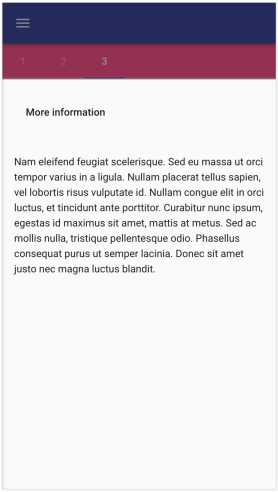
This task was measured by tracking the time spent to find the child page as well as navigating back to the first page. The number of clicks was also measured.



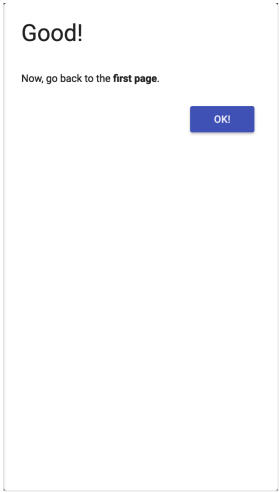
**Fig. 5.** Task 1 cover page.



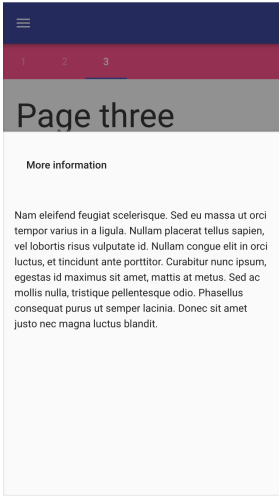
**Fig. 6.** Task 1 application. Last page with "Read more" button.



**Fig. 7.** Task 1 application. "Read more" section opened.



**Fig. 8.** Task 1 application. Second info page.

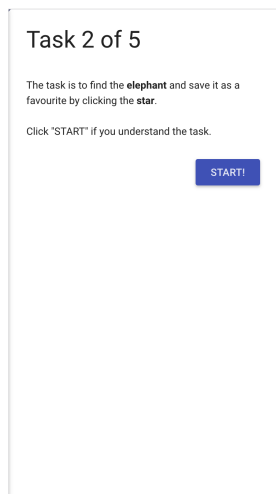


**Fig. 9.** Task 1 application. Closing the "Read more" section.

## 6.2 Task 2

The second task tested **Change and transformation**. In this task the test subject was asked to choose a specific item and save it as a favorite by clicking on a star next to the element (see Fig. 10). When the star was clicked, the item was saved in a side menu and the user was then asked to find the item in the side menu (see Fig. 11 - 13). In the animated version of the application the star was animated with a transition to the side menu.

This task was measured by counting the number of clicks to find the item in the side menu after the item first was chosen as a favorite in the main view. The time spent to find the item after first choosing it and the number of clicks was also measured.



**Fig. 10.** Task 2 cover page.

## 6.3 Task 3

The third task tested the area **Anticipation**. When this test started a view with a carousel of images were shown. The test subject was asked to find a specific image in the carousel (see Fig. 14). The only way to interact with the carousel was to swipe right or left (see Fig. 15). In the animated version the carousel was animated with a small spin when the view was first loaded. The number of clicks was measured as well as the time spent to complete the task.

## 6.4 Task 4

The fourth task was testing the area **Action and feedback**. The test subject was asked to click on a button in the first view of the application (see Fig. 16). The only thing that had to be done to complete the task was to wait for seven





Fig. 11. Task 2 application.

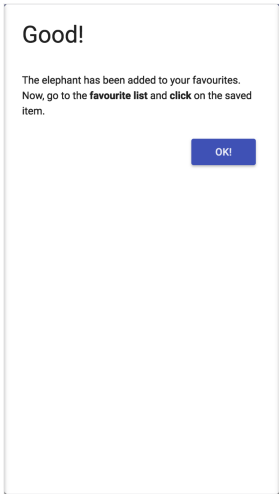


Fig. 12. Task 2 application. Second info page.

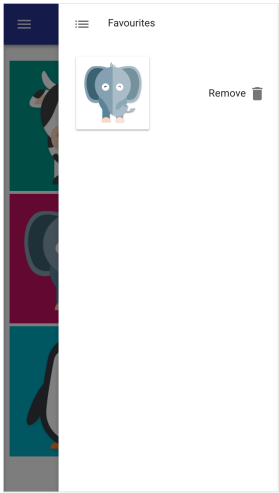


Fig. 13. Task 2 application. Side menu.

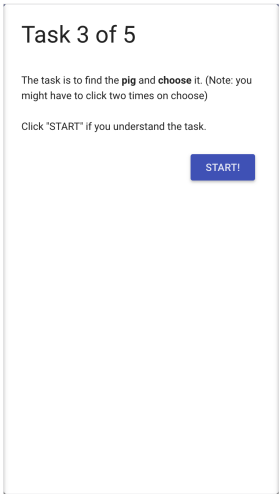
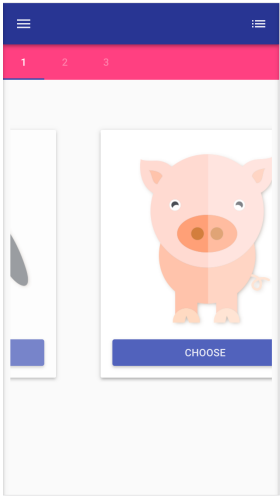


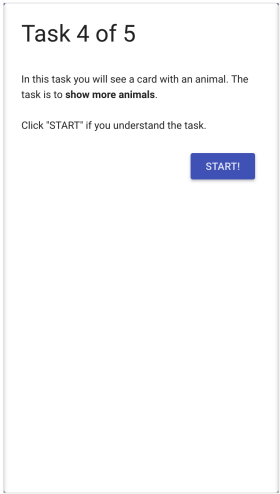
Fig. 14. Task 3 cover page.



**Fig. 15.** Task 3 application. Sliding sideways.

seconds after the button was clicked. The delay of seven seconds was added to simulate a loading time when an action is processed. In the animated version an animated loading circle was shown as soon as the button was clicked (see Fig. 17).

The aim of this task was to see if the user waited for the next page to load without clicking the button again. Here the number of clicks was counted as well as the time spent to complete the task.



**Fig. 16.** Task 4 cover page.

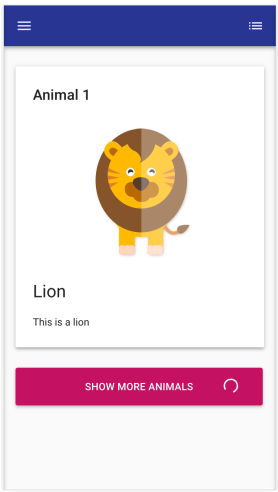


Fig. 17. Task 4 application. Loading button active.

6.5 Task 5

The fifth task was testing **Attention**. Here the test subject was asked to collect ten items in a list (see Fig. 18). When the tenth item was chosen a "done" button was shown (see Fig 19). In the animated version the button had a flashing animation where the button switched background color between white and blue. The application without animations had only a white button without a flashing animation.

In this task the time from when the tenth item was chosen until the "done" button was clicked was measured.

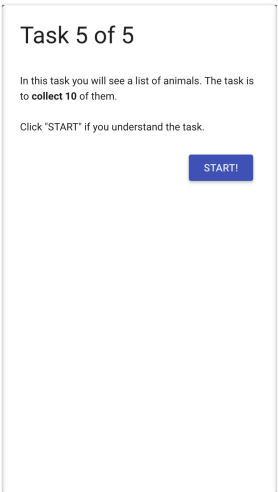
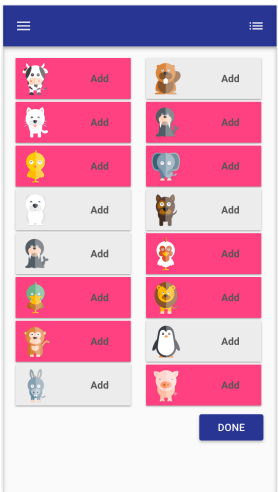


Fig. 18. Task 5 cover page.



**Fig. 19.** Task 5 application. Ten items chosen.

## 7 Result

The tests resulted in 71 data points. The data points were not distributed equally over the two test applications since not all test rounds were completed. The test rounds performed on the application containing animations were all completed. Though, the test rounds performed on the application without animations had a completion rate of 55.6 percent (i.e. completed test rounds). Since the test applications were distributed to the test subjects based on the number of completed test rounds, there was in total a greater number of started test rounds on the application without animations than on the application containing animations (see Table 1). Application 1 (containing animations) had an equal or less amount

App nr	Started test rounds	Completed	Incompleted	Completion rate (%)
App. 1	25	25	0	100
App. 2	45	25	20	55.6

**Table 1.** Test rounds.

of clicks on each task compared to Application 2 (without animations) as well as a lower total time on each task. Though, there was one exception, the total time on task 2 was lower on application 2 than on application 1 (see Table 2 - 3).

In Table 4 a comparison of the number of clicks between the two applications is presented. To be noticed is that the number of clicks in both applications is pretty equal. The clearest exception is the number of clicks performed on task 4. The difference between the two applications is 2 clicks (application 1) against 20 clicks (application 2) which means a ratio of 10 times more clicks performed

Task	Nr of clicks	Diff min. clicks	Total time (s)
Task 1	14	7	12.1
Task 2	5	1	7.1
Task 3	2	0	6.1
Task 4	2	1	9.1
Task 5	1	0	1.6

**Table 2.** Application 1 (containing animations), average result data.

Task	Nr of clicks	Diff min. clicks	Total time (s)
Task 1	14	7	13.7
Task 2	5	1	5.9
Task 3	2	0	6.2
Task 4	20	19	17.1
Task 5	2	1	3.7

**Table 3.** Application 2 (without animations), average result data.

on the application without animations compared to the application containing animations.

Task	App 1	App 2	Diff (App2 - App1)	Ratio (App2/App1)
Task 1	14	14	0	1
Task 2	5	5	0	1
Task 3	2	2	0	1
Task 4	2	20	18	10
Task 5	1	2	1	2

**Table 4.** Average number of clicks (comparison).

In Table 5 a comparison of the average time spent on each task between the two applications is presented. The clearest difference is the time spent on task 5. The time spent to complete task 5 was more than double in application 2 compared to the time spent on the same task in application 1. The time spent to complete task 4 was also almost doubled in application 2 compared to application 1.

## 8 Discussion

The data presented in the result section implies that animations in an application improves the user interaction in most cases. The tasks where the user interaction was improved the most was task 4 and task 5. These tasks were testing **Action and feedback** and **Attention**.

Many test participants had problems to complete task 4 within a reasonable time frame and with few clicks. This task should only take about seven seconds to complete since this was the loading time that was simulated when the button was clicked. This task should also require only one click since this was the only interaction needed (to click the "show more animals" button) to start the loading

Task	App 1 (s)	App 2 (s)	Diff (App2 - App1) (s)	Ratio (App2/App1)
Task 1	12.1	13.7	1.6	1.13
Task 2	7.1	5.9	-1.2	0.83
Task 3	6.1	6.2	0.1	1.02
Task 4	9.1	17.1	8	1.88
Task 5	1.6	3.7	2.1	2.31

**Table 5.** Average time (comparison).

simulation. As the result shown, the test subjects which tested the application without animations clicked the button 20 times compared to 2 times for the users who tested the animated application. This result implies the importance of informing the user that the application is responding to a certain interaction.

The result also shows that it can be useful to use animations to draw the users' attention which was done in task 5. Even though the time did not differ more than two seconds between the applications, the time spent to finish the task was still more than double in application 2 compared to application 1. This result can be interpreted as that the animation helped the user to find the button faster.

### 8.1 Does Animations Improve User Interaction?

Over all, the result implies that animations can help to improve user interaction in terms of speed to perform specific tasks. Since it was only two results that stood out from the others when comparing the results from the two applications, it is hard to claim that all five types of animations had a positive impact on user interaction. Though, there is no proof that animations lowered the user interaction.

## 9 Future Work

If this study were to be conducted again a few areas could be improved. This study is a source which can be used advantageously as a base in further studies. The test applications which were used in this study should be evaluated and further developed and optimized if they were to be used in a future study. A suggestion is to perform the user tests in a supervised and controlled test environment. One suggestion is that the tests can be supervised by a test manager in a controlled environment. This could lead to a more reliable result and hidden problems regarding the specific tasks could easier be found.

Another thing to consider in future work is to choose the test subjects more carefully. Different applications can have different users and it would therefore be advantageous to perform the test with test subjects that fits the specific user group of the application.

## References

- [1] Weir, G.R., Heeps, S.: Getting the message across: ten principles for web animation. In: 7th IASTED International Conference on Internet and Multimedia and Applications (IMSA 2003). (2003) 121–126
- [2] Morrison, J.B., Tversky, B., Betrancourt, M.: Animation: Does it facilitate learning. Technical Report SS-00-04, Department of Psychology, Stanford, CA (2000)
- [3] Zhang, P.: The effects of animation on information seeking performance on the world wide web: Securing attention or interfering with primary tasks? *Journal of the AIS* **1**(1) (2000) 1–28
- [4] De Koning, B.B., Tabbers, H.K., Rikers, R.M., Paas, F.: Towards a framework for attention cueing in instructional animations: Guidelines for research and design. *Educational Psychology Review* **21**(2) (2009) 113–140
- [5] Galitz, W.O.: The essential guide to user interface design: an introduction to GUI design principles and techniques. John Wiley & Sons (2007)



# Can highlighting screen borders be visual cues for swipe gestures?

Mikael Wahlström

Department of Computing Science  
Umeå University, Sweden  
`oi12mwm@cs.umu.se`

**Abstract.** In this study, transparently highlighted borders of the screen was investigated to work as visual cues for swipe gestures on touch screens. It was analyzed by A/B testing two versions of a mobile mock-up application: one had and the other had no highlighted borders where it was possible to swipe. 20 students from Umeå University participated in this study and their task was to swipe through views containing owl images until a view with a punk rock owl was found. Touch movement tracking and screen recording were used to collect data from this test and after task completion some questions about the test and mobile usage habits were answered by the participants. The highlighted borders was considered to help if test users exposed to them were faster and did less errors when swiping to find the punk rock owl than users not exposed to them. The test results showed no proof that highlighted borders could or could not work as visual cues for swipe gestures. This was believed to mainly depend on too low visibility of the highlighted borders, an insufficient sample size and an unclear task.

## 1 Introduction

With the development of touch screens, a new way to interact with devices arose: the gestural interfaces. Directly by using fingers people could navigate and manipulate content in devices. This is an interface smart phone and tablet users interact with daily and the use of touch gestures in such devices are even increasing [1]. Furthermore, devices get equipped with more sensors and features, expanding the possibilities of interacting with them [2]. Alongside these new techniques, the interaction may become more complex and, naturally, it would put a higher demand on designers to explain and guide the user in how to interact with devices. However, it is often up to the user to find out how and when gestures can be used, and there are no generally standardized design principles for gestures developed [3]. Norman and Nielsen [4] even claimed that gestural interfaces are a step backward in usability and that this partly depended on the inconsistency between designs and the lack of hinting available gestures.

Current visual cues of gestures might often: require that the user touches the screen before getting the cue, minimize or hide the actual content the user wants to see, use rather advanced and time-consuming animations/tutorials to explain

the gestures. Therefore, we explore another alternative to these, using light in order to give hints about gestures. More specifically, investigating the use of transparent highlighted borders of the screen as visual cues for swipe gestures. Transparent highlighted borders of the screen could work as visual cues without the need to minimize or fully cover the actual content the user wants to see, it is rather easily implemented for designers, and it might not be time-consuming for the users to understand and learn.

It may also be possible to successively decrease the visibility of the highlighted borders alongside the user's increased usage of the swipe functionality, until they are not longer visible and the user has learned where it is possible to swipe. This could make it work as a minimalistic tutorial for the user.

## 2 Background

One commonly used touch screen gesture is swiping (i.e., flicking). It is used by dragging one finger in a horizontal or vertical direction on the screen. It is frequently used to browse/navigate in devices but its increasingly being adapted to do more actions. Swiping is often used to: handle content (e.g., delete, hide, share, move, like); retrieve menus, get controls to manipulate specific sections or contents; and even mark emails as read [5]. This might possibly depend on a strive to help users to focus on particular important content but still have quick access to these rather important controls. However, these actions often have the disadvantage of not showing their abilities to be used by swiping, and sometimes they are only accessible through swiping [5].

The same concept could be applied to basic navigation between views. Some applications use both icons and swiping to navigate in them, but without having any clear appropriate way to notify the users of the swipe functionality, which may lead users to press the icon instead of swiping.

Still, there exist visual solutions for indicating when a swipe gesture is possible. These can be divided into three categories: animations, instructions and indications [3].

**Animations** are meant by one or several dynamical elements that gradually change their appearance/style. They can for example be used when need be or at initial start up of applications. Some benefits with them are that they often are clear and informative (e.g., an animated hand showing that you can swipe). They can also tell the user what the gestures will do by animating the action conducted alongside a visualization of the gesture. Some disadvantages of using animations are that they may be rather difficult to implement and there are no guidelines for designing them. There is also an issue with determining when they should be used [3]. In some applications, animations are also used to hint actions accessed through gestures when touching specific content. For example, the swipe-gesture for starting a chat with a friend in the mobile application *Snapchat* has been hinted with a bounce of that field to the right when tapped on, showing a chat icon to

the left and at the same time the direction to swipe to get to the chat. Unfortunately approaches like these often occupies gestures, such as tapping in this case, that otherwise could be used for something other than hinting gestures.

**Instructions** are mainly meant by text-based tutorials. Which have similar usage scenarios as animations. One benefit of them is that they can be clear and expressive. Although, they can have the disadvantage of interrupting if they pop up in front of the actual content and if they on the other hand always are visible they take up space [3]. Another problem that may appear if they are presented at startup of a new application (or after updates) is that the user has to remember what was written later on. They also need to have time to read them at that moment, otherwise they will probably skip them and then possibly never get the instructions since it might only be available once to avoid irritating user that does not want to read it.

**Indications** usually stands for hinting of more content with the use of discrete design solutions. The visual cue for swiping tested in this study, using highlighted borders of the screen, belongs to this category. Three other examples of indications for swipe gestures are presented in Figure 1. The first of these is using dots underneath the content that can be swiped; the second is a zoomed out view to partially show there are content vertically or horizontally accessible by swiping; and the third is tabs showing available pages, which most often are accessible through both pressing the tab and swiping, giving the user both a possible action and an indication of a gesture.



**Fig. 1.** Three indications of a swiping possibility: three dots underneath, partially visible pages on the sides, and tabs on the top of the page

## 2.1 Earlier work

An earlier investigation [3] of how to give users visual cues of gestures resulted in a touch sensing interface. It tried to display a user's possibility of gestures through symbols around the finger or fingers used to interact through the screen. While this certainly gives the users indications of what gestures they can do, it

requires the user to press the screen in order to get them. Which consequently leads to one unnecessary action to get the visual cues. Getting a cue directly without having to touch the screen can be considered to be a faster way of hinting what gestures the user can do.

In another study [6] touch force/pressure was investigated to work as an additional parameter for gestures. This could also be a possible solution for how to give hints of gestures: e.g., a light pressure could be used to give hints instead of doing actions. Touch pressure sensing is something that will probably be in most mobile devices in a few years so it might be a valid solution soon. This does however still require the user to touch the screen, possibly slowing down the time it takes to be informed of possible gesture actions.

A way of overcoming the lack of visual hinting in gestural interfaces has been investigated [7] to be solved using eye-tracking. This by giving more visual information if the user looks at an element it can interact with. While this would solve the previous issue with having to touch the screen to get the hint, it is not yet considered robust and reliable enough [7]. Even if it were, the transition phase for it to be common in a majority of the mobile devices on the market is probably not realizable in a near future.

Another investigation [8] for ways to give users visual cues of gestures is pre-touch sensing of fingers on touch screens. In this study they managed to interact with the mobile device by hover over the screen with fingers. The article is not really about visual cues for gestures, but indirectly this technique opens up a lot of new possibilities to give visual cues to prompt gestures. Although, one disadvantage this may have is that fingers would block parts of the screen, possibly hiding the visual cues. Also, as with eye-tracking, this technology is not likely to be common in a majority of mobile devices in a near future.

In contrast to these solutions, transparent highlighted borders of the screen is something that does not require touch pressure, is easily integrated today and does not fully cover content on the screen.

### 3 Method

An A/B-test (i.e., split test) was conducted to test the transparently highlighted borders as visual cues. Two versions of a mobile mock-up application were created for the test using the software *Pirate Studio*. Both were identical except that one did and the other did not have highlighted borders of the screen where it was possible to swipe to see another view.

#### 3.1 Participants

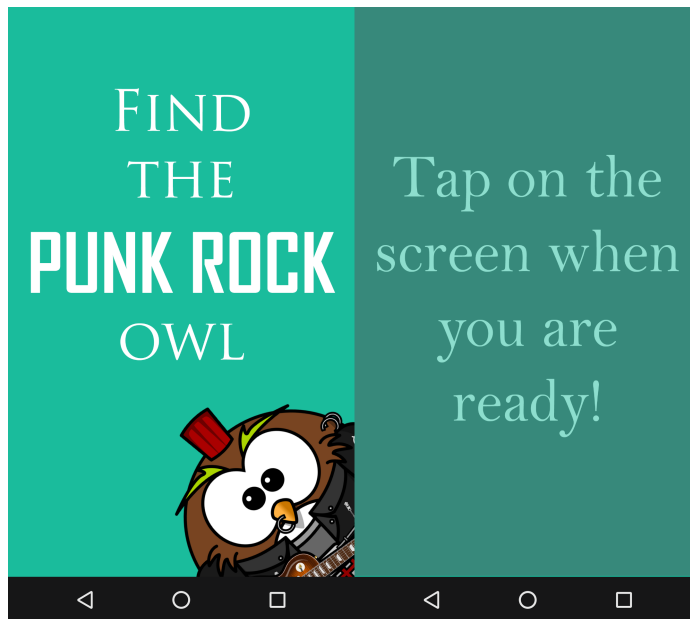
20 students from Umeå University were recruited for the study by asking random students to participate in testing a prototype of a mobile application. The participants were randomly divided into two equal sized groups named highlighted and regular. Gender were not believed to give different outcome so the groups were not divided equally between gender. They were neither divided based on

age since they all were believed to be in a rather similar/small range of age considering they were students.

### 3.2 Design of the application

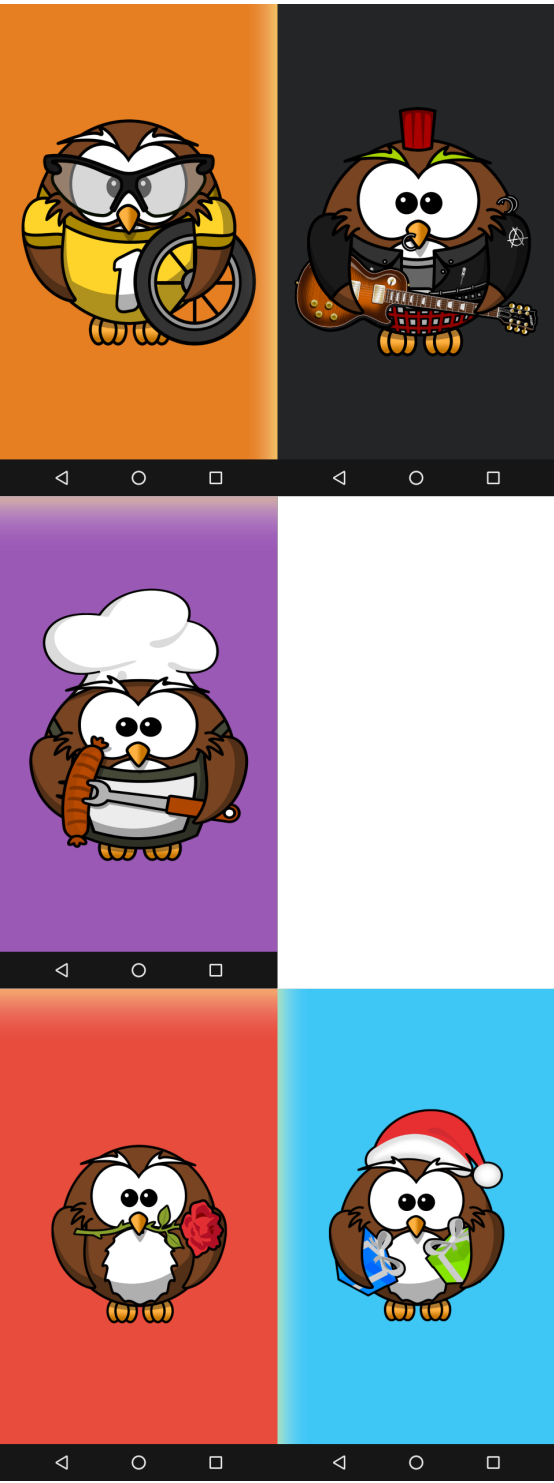
At the first view of the application the user gets told to “tap the screen when ready”. After that screen had been tapped a second view was shown describing the task, “find the punk rock owl” (see Figure 2). When that description view was shown for 5 seconds a new view got displayed where the actual test started and the test user was supposed to swipe until finding the punk rock owl. The second view was to the left of the first, and the third and fourth view above their earlier views, and then the fifth view was to the right of the fourth view (see Figure 3 for an illustrative description of this navigation layout and the looks of the views). When the test user got to the punk rock owl, the background of that view got enlightened after 0.5 seconds, and when it had been visible for 2 seconds, one new last view was shown saying “good job” (see Figure 4).

The path used to navigate through the owl views was chosen to be illogical (e.g., not always having the new view to the right of the current) to rule out the possibility that users succeeded to swipe out of habits of interacting with applications, not because of the highlighted borders.

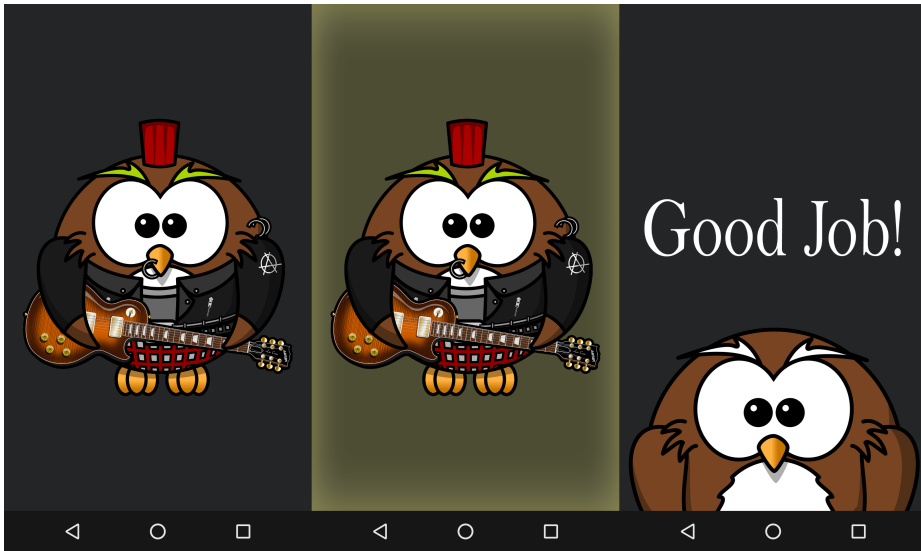


**Fig. 2.** First and second page of the application. Here the user taps the screen to start the test, sees the task description, and after that the test starts.

The yellow color for the highlighted borders was RGB(255, 255, 135) and had an opacity of 50 percent to make it see-through. The highlighted borders faded



**Fig. 3.** Third to seventh pages of the application, third being the Santa owl with blue background and seventh being the punk rock owl with black background. These are the views which the actual test data is collected from. The navigation layout is represented by the views placement in relation to each other. They are navigated by swiping one finger away from the lit edges.



**Fig. 4.** Seventh and eighth pages of the application, showing how the background gets lit up after the seventh view has been visible 0.5 seconds to show that the punk rock owl is found and 1.5 seconds later the eighth view, with the “Good Job” text, gets showed, ending the test.

in to the background making the visible width of them approximately  $1/15$  of the total screen width for the horizontal and  $1/25$  of the total screen height for the vertical borders of the screen. The first owl view had a blue color RGB(62, 198, 245), the second owl view had a red color RGB(231, 76, 60), the third owl view had a purple color RGB(155, 89, 182), and the fourth owl view had an orange color RGB(230, 126, 34).

The views had different background colors since it was thought to make it easier to distinguish between views both for the user and while analyzing the test data. It was also applied to test the highlighted borders without limiting them to one color context.

### 3.3 Test procedure

Before the participants started the test, they got informed that the screen and touch movement was recorded but that it was an anonymous test. The participants also got told that they can not do something wrong, it is the application that is tested, not them.

After that, the investigator started the recording, told the task: “find the punk rock owl”, and handed over a mobile phone with the prototype displayed on it with the help of the mobile application *Pixate*. The phone was a black *LG Google Nexus 5* with the brightness set to approximately 20 percent of its maximum ability. There was no communication between the investigator and test user until the task was completed.

Data was collected by recording the screen with user touch movements visible on the screen at the same time. This was carried out using the application *AZ Screen Recorder*. After task completion, the test users answered a few questions (see Table 1) about the test and their mobile usage habits.

Questions	Answer alternatives
What is your gender?	Male, Female
What is your age?	(Digits)
I consider myself to be an experienced mobile user.	Strongly disagree to strongly agree (scale of 1-5)
I am custom with swiping/flicking to navigate, do actions and handle content in mobile phones.	Strongly disagree to strongly agree (scale of 1-5)
Out of these 5, which are the ones you frequently use on your mobile phone:	Snapchat, Instagram, Tinder, Any arbitrary photo viewer, Any arbitrary e-book reader.
I think it was easy to find the owl.	Strongly disagree to strongly agree (scale of 1-5)
During the test, did you notice that occasionally one of the screen borders was lit (brighter/colored)?	Yes, No (no further questions asked).
Did the lit borders help you find the punk rock owl?	Yes, No (no further questions asked), I do not know.
Already at the first view I realized that I would navigate myself onwards by swiping to get towards the lit border.	Strongly disagree to strongly agree (scale of 1-5)

**Table 1.** Questions answered by the participants and their answer alternatives after they conducted the test.

Lastly, if the participant wanted an explanation of the test they got told about the highlighted borders or lack of highlighted borders (depending on test group) and that they were tested to work as visual cues for the swipe gesture.

### 3.4 Method for analysis

*OpenShot Video Editor* was used to view, clip and analyze the recorded test data. The highlighted borders was considered to help the users to find the punk rock owl if the users exposed to them spent less time completing the task and did less error swipes (i.e., swipes in the wrong direction) during the task than the users without highlighted borders.

Other than analyzing the total time it took and amount of swipe errors made from first owl view to the punk rock owl view which would give an overall efficiency check, the time and errors on the first to second view were analyzed as the essential part for *intuitively* understanding the highlighted borders as visual cues for swiping. This because then the users would not have experienced the correlation between highlighted borders and swiping yet. The results from navigating through the rest of the views was considered to potentially be biased since



by then users might have learned the correlation between highlighted borders and swiping, making it something taught and not intuitively understood. That part was thereby analyzed with the interest of seeing if users understood the purpose of the highlighted borders after encountering them.

## 4 Results

### 4.1 Participants

The lit group had 3 females and 7 males, and the unlit had 6 females and 4 males. In the lit group they had mean (M) age of 22.6 years with a standard deviation (SD) of 2.547 and the range 20-27 years. In the unlit group they had an mean age of 21.7 years with a SD of 1.703 and the range 20-24 years.

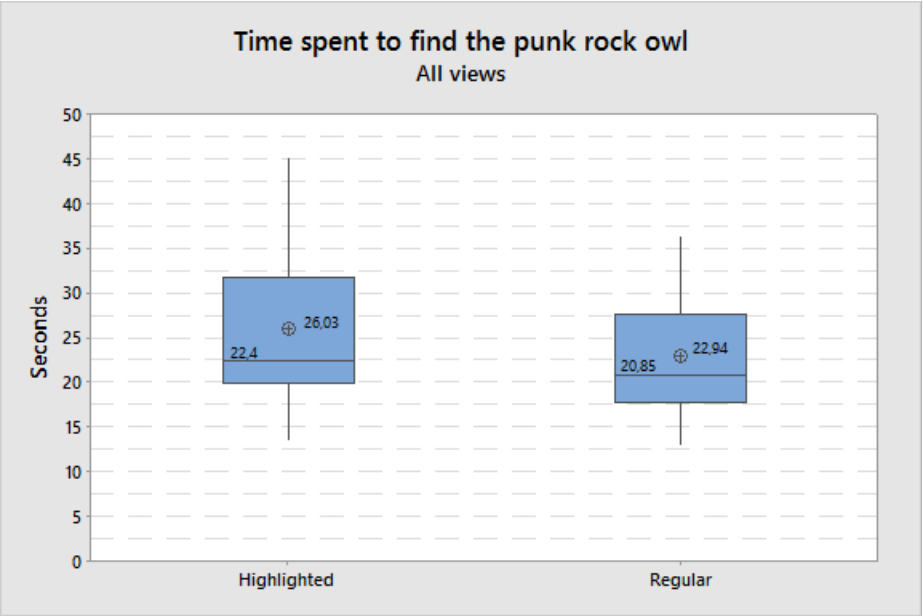
The users own estimation of their experience of mobile usage (lit: M=4.3, SD=0.675, range=3-5; unlit: M=4.2, SD=1.033, range=2-5) and their confidence in their knowledge and habit of swiping (lit: M=3.9, SD=0.994, range=3-5; unlit: M=4.5, SD=0.707, range=3-5) was similar in both groups. There were not any remarkable difference between the groups' frequent usage of applications that often require swiping to navigate (lit: Snapchat=8/10, Instagram=7/10, Tinder=0/10, arbitrary photo viewer=4/10, arbitrary e-book reader=2/10; unlit: Snapchat=8/10, Instagram=9/10, Tinder=1/10, arbitrary photo viewer=5/10, arbitrary e-book reader=2/10).

### 4.2 Test

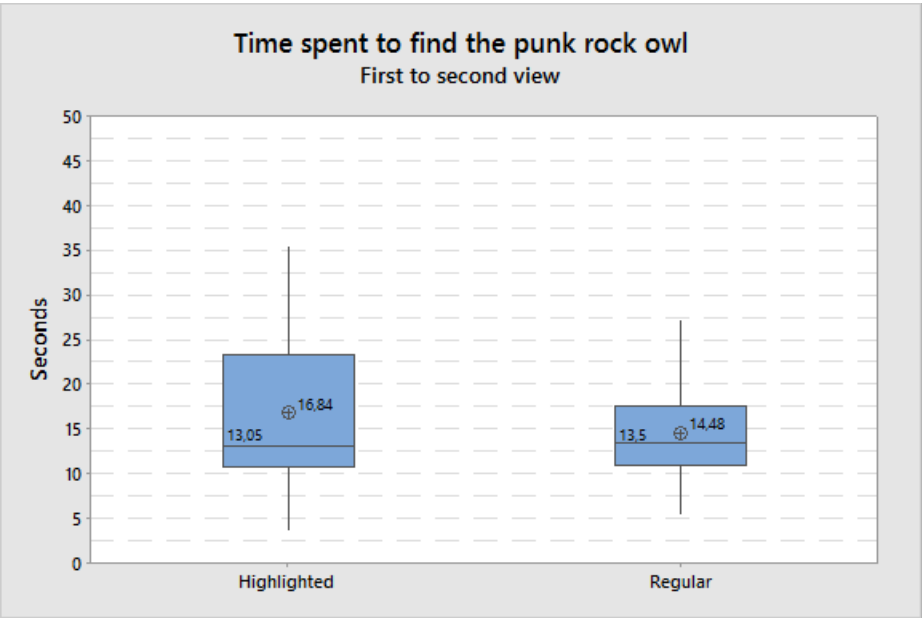
The test results for both groups regarding the time participants spent in order to find the punk rock owl for all views is presented in Figure 5, from first to second view is presented in Figure 6, and from second to fifth (last) view is presented in Figure 7. The equivalent results for amount of swipe mistakes made is presented in Figure 8 for all view, in Figure 9 for first to second view, and in Figure 10 for second to fifth view.

According to Anderson-Darling normality tests, no sample set had high probability of having normal distribution. Thereby, parametric Student's t-tests were not conducted due to possible unreliability of their results. When visually comparing two coherent samples from the groups they did not show any similar distributions among each other either, so nonparametric Mann-Whitney U-tests were neither conducted due to possible unreliability of their results.

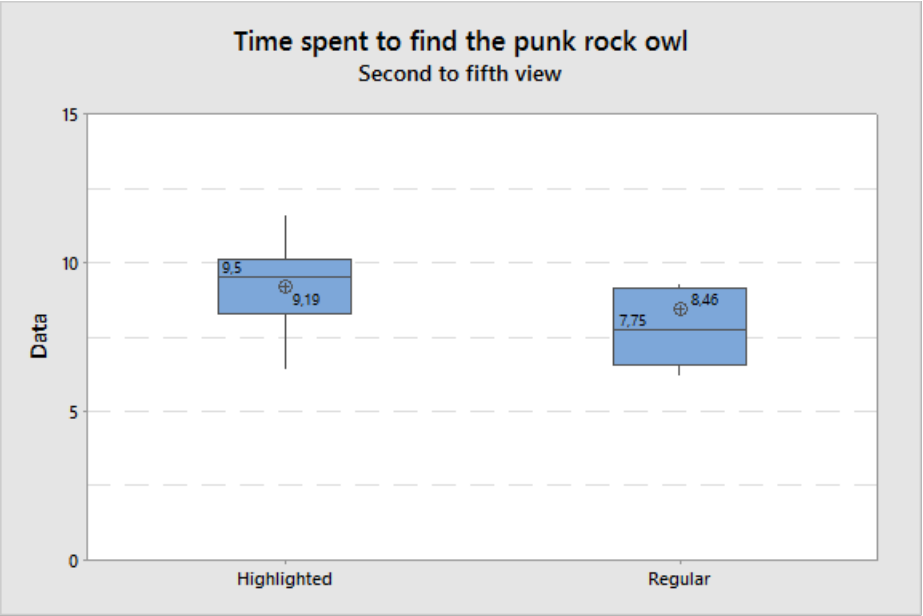
In the group with highlighted (i.e., lit) borders, five out of the ten test users reported that they saw these borders. Out of these five, regarding the question if the lit borders helped them to find the punk rock owl view: one answered that they did, one answered that they did not, and three answered that they did not know. Out of the four that either said that the lit borders helped or did not now if it helped them, two of them strongly disagreed and two disagreed upon the statement that the lit borders helped them to navigate from the first to the second view. There was no remarkable difference between the two groups regarding how easy they thought it was to find the punk rock owl.



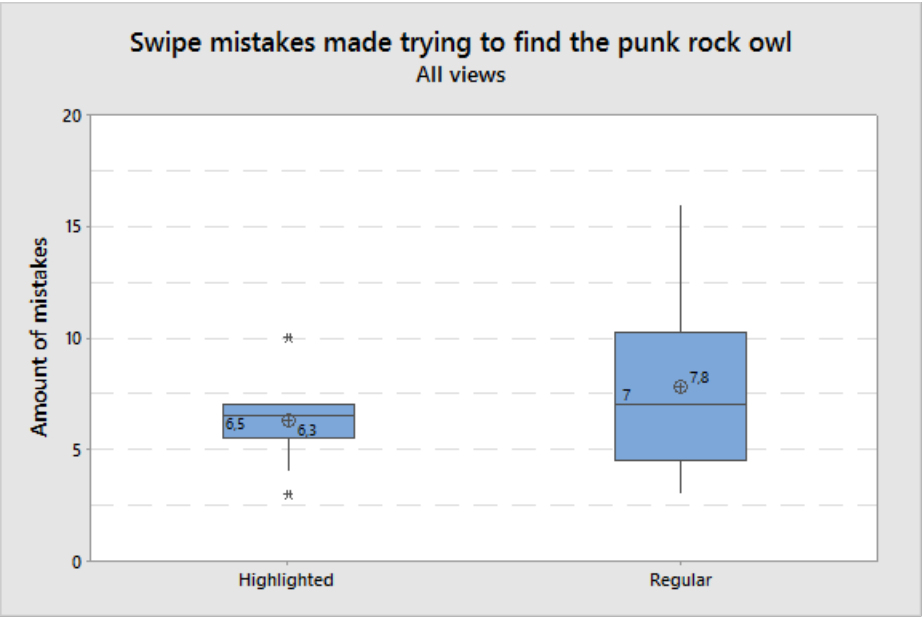
**Fig. 5.** Boxplots of the time participants spent in order to find the punk rock owl with highlighted and regular borders for all views.



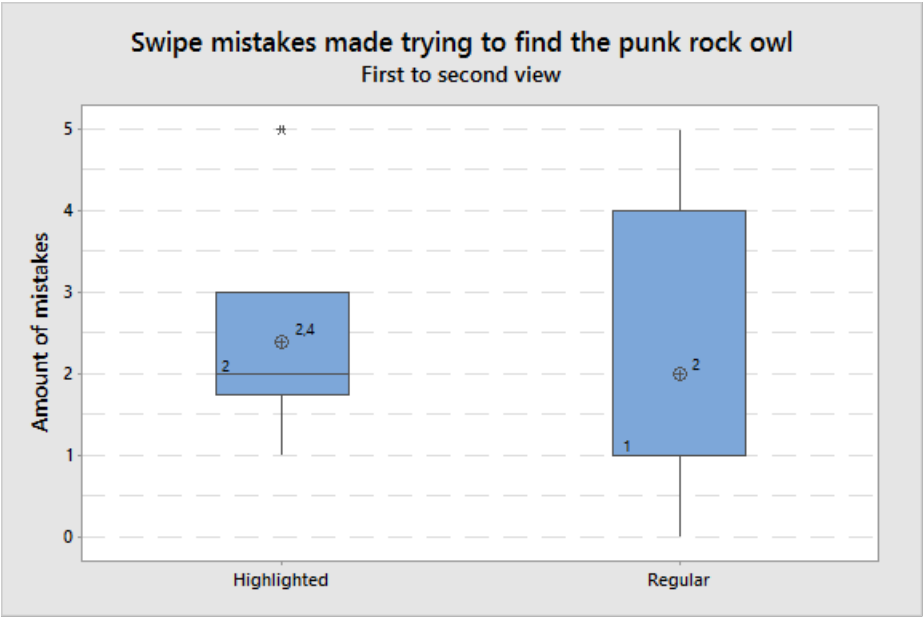
**Fig. 6.** Boxplots of the time participants spent in order to find the punk rock owl with highlighted and regular borders from first to second view.



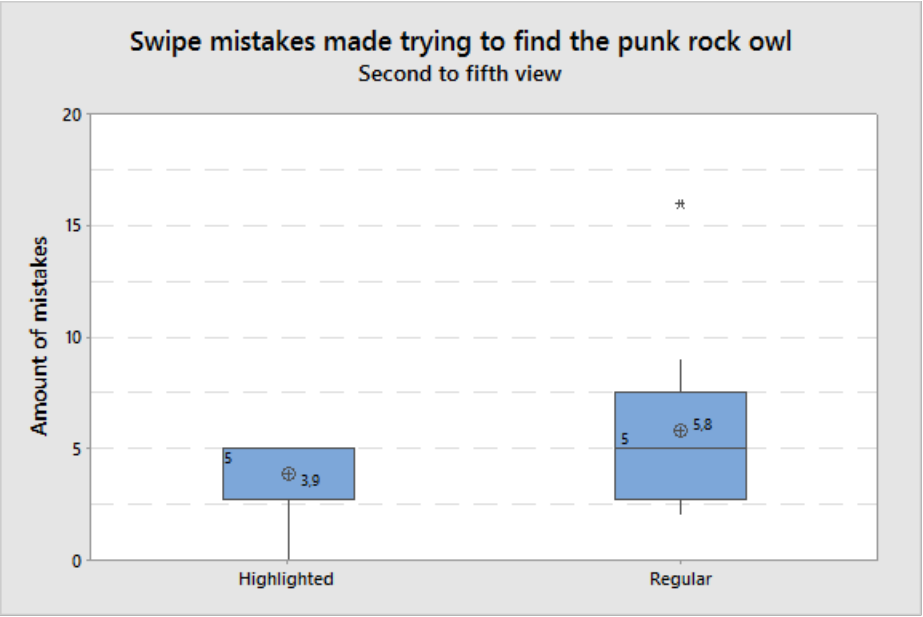
**Fig. 7.** Boxplots of the time participants spent in order to find the punk rock owl with highlighted and regular borders from second to fifth (the punk rock owl) view.



**Fig. 8.** Boxplots of the amount of swipe mistakes/errors made by the participants before finding the punk rock owl with highlighted and regular borders for all views.



**Fig. 9.** Boxplots of the amount of swipe mistakes/errors made by the participants before finding the punk rock owl with highlighted and regular borders for first to second view.



**Fig. 10.** Boxplots of the amount of swipe mistakes/errors made by the participants before finding the punk rock owl with highlighted and regular borders for second to fifth (the punk rock owl) view.

## 5 Discussion

Since no significance tests could be made, no scientific evidence could be extracted. Reasons behind this might partly depend on these three factors: 20 participants were an insufficient sample size, there was an uncontrolled widely varying willingness to swipe in order to explore the application which gave large variations in time spent and amount of errors, and also that it was a somewhat confusing task so some test users might have thought that the punk rock owl was going to appear by itself if they waited and not that they should interact with the interface to find it.

Furthermore, a strong doubt about the validity of the test arose from the fact that, assuming that they fully understood the question, only five out of ten users in the group with highlighted borders reported that they saw the lit borders. Therefore, the five users who reported that they did not see the lit borders was probably only affected by them on a subconscious level, which was not the intention of this test.

### 5.1 Future work

To further test if highlighted borders of the screen could work as intuitive visual cues of swipe gestures, a strong advice is to conduct a more extensive study using more participants to get a more accurate result. Another strong advice is to have better visibility of the highlighted edges. If increasing contrast or luminance of the borders do not seem to be sufficient, pulsing or blinking borders might be considered to draw even more attention. Eye-tracking could also be used in future studies to see more accurately if the participants notice the light, rather than asking what the users recall after the test. One more rather important aspect is to come up with a task that does not give such varying delays before the users start swiping, but still without directly mentioning the highlighting or even swiping since that is believed not to be representative in a natural scenario when someone uses an application.

It can also be beneficial to consider beginning with only using one background color to make the highlighting have a consistent intensity and contrast throughout all views. Finally, if a future study is focusing on the intuitive understanding of this visual cue, narrowing down the test by only using two views to swipe between would decrease the time it takes to test and analyze the result yet still contain the vital part for researching the intuitive understanding of the cue.

## 6 Conclusion

This study does not prove that highlighted borders could work as intuitive visual cues for swiping. This is believed to depend on the fact that the test results and method used for of this study was insufficient and unreliable. Highlighted borders of the screen can therefore not be confirmed or rejected as visual cues for swipe gestures until a more extensive study is conducted.

## References

- [1] Han, S.H., Park, W.: Intuitive multi-touch gestures for mobile web browsers. *Interact. Comput.* **25**(5) (2013) 335–350
- [2] Widgor, D., Fletcher, J., Morrision, G.: Designing user interfaces for multi-touch and gesture devices. In: *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, New York, USA, ACM (2009) 2755–2758
- [3] Lundgren, S., Hjulstrom, M.: Alchemy: Dynamic gesture hinting for mobile devices. In: *MindTrek '11 Proceedings of the 15th International Academic MindTrek Conference*, New York, USA, ACM (2011) 53–60 [http://www.cse.chalmers.se/research/group/idc/ituniv/courses/11/mc/gesture\\_hinting\\_camera\\_ready.pdf](http://www.cse.chalmers.se/research/group/idc/ituniv/courses/11/mc/gesture_hinting_camera_ready.pdf), accessed 2016-09-20.
- [4] Norman, D.A., Nielsen, J.: Gestural interfaces: A step backwards in usability. *Interactions* **17**(5) 46–49
- [5] Bank, C., Zuberi, W.: *Mobile UI Design Patterns*. UXPin (2014)
- [6] Heo, S., Lee, G.: Force gestures: Augmenting touch screen gestures with normal and tangential forces. In: *UIST '11 Proceedings of the 24th annual ACM symposium on User interface software and technology*, New York, USA, ACM (2011) 621–626
- [7] Sirui, L.: Attentive gestural user interface for touch screens. Master's thesis, The University of Hong Kong (September 2013) <http://hub.hku.hk/bitstream/10722/192863/1/FullText.pdf>, accessed 2016-09-28.
- [8] Hinckley, K., Heo, S., Pahud, M., Holz, C., Benko, H., Sellen, A., Banks, R., O'Hara, K., Smyth, G., Buxton, B.: Pre-touch sensing for mobile interaction. In: *CHI '16 Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, New York, USA, ACM (2016) 2869–2881





# How correct is auto-correct when entering text on a smartphone?

Freja Wieslander

Department of Computing Science  
Umeå University, Sweden  
`id12fwr@cs.umu.se`

**Abstract.** This article investigates if auto-correct is obstructing more than the user misspell words when not having auto-correct. It was performed by a user test comprising 13 participants that had to enter a text including ten independent phrases twice. The phrases were selected from MacKenzie and Soukoreff's phrase set [1] that was produced for evaluating text entry techniques. The phrases were translated into Swedish because of the fact that the test was conducted in Sweden to simulate a real life condition. The collected data was evaluated into two rates that were based on Steven John Castellucci's *Character Wise Error Rate (CWER)* [2], where one was based on the text with auto-correct and the other one without. The auto-correct proved to have a significant effect of the participants' CWER ( $P < 0.05$ ). With auto-correct turned on, the users' CWER will be lowered by 2.88 percent with the standard error 1.08 and confidence interval (-5.13, -0.63) percent.

## 1 Introduction

Frustrations based on misinterpretations by auto-correct are a general problem of smartphone users. Jillian Madison wrote [3] that she created a website where people could upload their misspelled and embarrassing sentences confused by auto-correct when sent the sentence to their friends and families. The website got 500 new uploads each day, which made her realize how much frustration is lying behind the topic. Therefore she wrote the book, "Damn you auto-correct!" [3], to summarize the problem and show how wrong it could be in some situations. The intensity of the frustration caused by auto-correct makes one believe auto-correct is obstructing more than it helps. Does it really help smartphone users more than they need to correct the auto-correct? On the other hand, today all smartphones include auto-correct, which indicates advantages of the helping tool in some way. This study will find out if auto-correct is obstructing more than the user misspells words when not having auto-correct. A user test was made, where participants had to enter a text twice, including ten independent phrases translated into Swedish. The handset was an Android running by the time of writing the latest version operation system, 7.1. and was documented with screen recording that saved entered text. Each participant got two *Character Wise Error Rate (CWER)* [2],

one with auto-correct and one without. Finally, all participants CWER:s got analyzed by Minitab 17<sup>1</sup> to conclude the result.

## 1.1 Background

Text entry is a pervasive activity on smartphones, and is required even in the most basic tasks. To just mention few, entering a password, writing a short message or searching for contents. Even if the touch screens have enabled new intuitive interaction techniques they still continue to present challenges for the users, especially for text entry. One of them is the phenomena called *Fat Finger Problem* [4] which refer to the situation when the users' finger may press multiple buttons at the same time, which will result in an error input. Another phenomenon called *Perceived Input Point Model* [5] which means that the user believe they press the right spot but touch a location offset from the intended target due to the difficulty in positioning themselves accurately with their fingertips. Christian Holz and Patrick Baudisch [5] came to the conclusion in their research that the incorrect touch is not caused by the fat finger problem. It is more likely a question of how good the software analyzes different users and finger postures. A third challenge in entering correct input on a smartphone is the user's ability to type correct letters of the word. They may be uncertain about spelling, which leads to misinterpretation of the input. This mentioned phenomena, that raised some of the challenges in entering text, will be facilitated by the helping tool auto-correct. Auto-correct is a software function that analyzes your input and tries to correct mistakes and present to you a better version of the text. Just as Jillian Madison [3] wrote in her book, it is hard to find information that explains how auto-correct really works. The mobile phone industry and software developers keep the trade secret for themselves and avoid discussing it. Anyhow, there is some basic information about how it works. When a user enters a word, the auto-correct checks the letters against a dictionary, which is built in the software. If the letters do not match an established word in the dictionary, it will offer similar or alternative words to the one the user is trying to write. Jillian Madison [3] even explained that the smartphones include a learning function that adds new words due to the user's patterns of use.

## 1.2 Related work

There is little previous work done in this subject where auto-correct does not help the user. Because companies have their algorithms as secret there is a possibility that there are quite a few studies on this topic but may not been made public.

However, there are a few studies about the input interaction of text entry in different devices. Katie A. Siek et al. [4] investigated by a usability study that there is no major differences between older and younger users when interacting with personal digital assistance's. B. Chaparro et al. [6] made a research of differences in input between three keyboard styles: Acer netbook keyboard and

---

<sup>1</sup> Minitab 17 is a statistical software program for effectively analyze data.

an iPad with a soft keyboard that was evaluated in both landscape and portrait orientation. This study did not test auto-correct but the differences in the raw input where the size of the keyboard and tactile feedback was studied. The study was made with a user test where the participants entered a series of phrases in each of the three keyboards. Similarly to this article, the participants were told to enter as accurately and quickly as possible without correcting their mistakes. There was no significant difference between the two orientations of the iPad but the physical keyboard was proven to have a faster typing speed. However, there were some differences between the physical keyboard and the iPad in portrait for input error.

There are some researches that want to improve the text entry by working around the issues of the virtual keyboard. Mayank Goel et al. [7] propose a text entry system, called ContextType, that adapts by switching by different keyboard models depending on the users' hand posture. The hand postures were classified in four; two thumbs, left thumb, right thumb and index finger. ContextType were evaluated and resulted to reduced the total text entry error rate with 20.6 percent. As Mayank Goel et al. [7] conclude, that by letting our mobile devices be aware of their users the experience can be improved. Patrick Baudisch and Gerry Chu [8] argue that we need to use the devices backside when using very small screens to not end up in the *Fat Finger Problem* [4]. They created a 2.4" prototype device and made a user study where the participants performed a pointing task using a back-of device interface. Although the technique was not used for text input in the study, it could solve the problem for text as well.

The related work of Panos Sakkos et al. [9] where a study was conducted to evaluate a dynamic keyboard. The keyboard alternates its key size dependent on the probability for the next character. The keyboard was tested with a predefined dataset and was shown to have a precision of roughly 23 percent.

Daryl Weir et al. [10] made a research of two algorithms for auto-correct, GPType and ForceType, where the first algorithm was Gaussian Process regression for evaluating the probability for the next character and word. ForceType uses pressure sensitivity so the user can define the uncertainty themselves. The conclusion was the GPType had a significant improvement from not having auto-correct, specifically seven percentage in decrease in character error rate.

A related work of Xiaojun Bi et al. [11] where a study to test if auto-correct can be used in conjunction with auto-complete. The study used an automated testing method with pre-defined inputs to measure the accuracy, similarly to the study of Panos Sakkos et al. [9].

## 2 Method

The research was performed with a user test. Participant wrote a number of predefined short text messages twice, with auto-correct and without auto-correct. The data was then processed, analyzed and summarized to a conclusion.

## 2.1 User Test

In order to accomplish the user test, 13 users with an average age of 23 participated. All of them was Swedish students and were required to have experience of text entry on a smartphone including a touch screen.

The participants were first informed about the process of the test. They were permitted to quit anytime during the test if desired. They were also informed about the anonymity of the study. That focus was on the interaction of the text entry and not the individual's performance. The test was documented by screen recording and was therefore required acceptance from participants.

The participants were given ten predefined text phrases from MacKenzie and Soukoreff's phrase set [1] that was produced for evaluating text entry techniques. The phrases were moderated in length which will make it easier for the user to remember while entering. The phrases were translated into Swedish because of the fact that the test was conducted in Sweden to simulate a real life condition. It also includes an everyday language that reflects our way of expressing ourselves when typing messages as in real life situation. Though it is not academic language, all the words are represented in the Swedish dictionary [12]. The phrases presented for the user test is shown in Table 1.

Phrases
Jag försöker hitta en parkeringsplats.
Min klocka föll i vattnet!
Var lämnade jag mina glasögon?
Rävar är väldigt smarta djur.
Huvud, axlar, knän och tår.
Kan vi spela kort ikväll?
Det var första gången hon försökte simma.
Mina fingrar är väldigt kalla.
Varför frågar du konstiga frågor?
Jag vill ha två eller tre koppar kaffe.

Table 1: Phrases presented in the user test that was translated into Swedish.

The phrases was presented on a smartphone as a message in the application Facebook Messenger<sup>2</sup>. They where instructed to reconstruct the same phrases as answers within the application. Participants had to type as accurately and quickly as possible to force them to not be too careful when entering the text and simulate real environment. The users were not able to press the backspace key, in the interest of identifying all errors that appear while typing. Participants were informed to leave any mistakes uncorrected.

The test was conducted in the way that all the participants used the same smartphone to rule out differences in the auto-correct algorithm. The handset was an Android running by the time of writing the latest version operation

<sup>2</sup> Facebook Messenger is a software application that provides instant text and voice communication. The application gives the opportunity to view a longer message at the same time as entering a message.

system, 7.1. More advanced versions of auto-correct also learns from the users input to give personalised corrections [3]. This could in turn have an effect on the results. To have the same conditions for all the participants this was solved by clearing the data storage of the keyboard after each participant in the test.

2.2 Result Analysis

Evaluating text entry includes a lot of parameters that can be considered. The two main variables are entry speed and accuracy. This study focused on accuracy since the interest with text entry combined with auto-correct is the correctness towards the wanted output. The correctness is measured by errors in words and characters as defined by Steven John Castellucci’s [2] method. He explains that it is more accurate to evaluate which characters that are misplaced, missing or redundant than to just focus on the number of the incorrect sequence of characters or incorrect words. Steven suggests an equation to calculate and get a rate of how correct a phrase/text turned out to be. This rate was calculated by the character wise error divided by the maximum number of the characters of the presented text versus the entered text. Steven named this rate as *Character Wise Error Rate (CWER)*[2]. Two CWER was collected from each participant, one based on the auto-correct and one without.

3 Result

The summarised data can be seen in table 2, by only taking the average the CWER is considerably higher without auto-correct. The auto-correct proved to have a significant effect of the participants’ CWER ( $P < 0.05$ ). An ordinary least squares (OLS) regression gave the effect of auto-correct with an effect constant of -2.88 percent with confidence interval of ( -5.13, -0.63) percent as calculated by Minitab. The baseline CWER with auto-correct off is 5.0 percent for the data with the calculated confidence interval of (3.07, 7.04) percent.

	CWER with auto-correct		CWER without auto-correct	
Start with auto-correction	st.dev	average	st.dev	average
0	1.55	3.53	2.54	7.3
1	1.31	2.82	3.49	5.76

Table 2: The summarised response.

3.1 Result Analysis

To draw a conclusion from the user test an analysis was made of the collected data. To interpret how auto-correct is affecting the resulting text of the users’ CWER, an ordinary least squares (OLS) regression model was made where

CWER is modeled with three predictors. The three predictors are categorical and are as follows; participant with values in the range of A-M, auto-correct 1/0, start with auto-correct 1/0. By the OLS regression model, the percentage of the variation in the data is explained to 75.49, by the coefficient of determination ( $R^2$ ). One of the participants was identified as an unusual observation, which means one of the participants' CWER differs from the other by a great amount and therefore has a big effect on the resulting regression. It is therefore a good idea to remove them from the dataset. Outliers could mislead the result and are therefore important to identify why the outliers appear. The participant who reported to be an outlier had newly made nails, which both bothered the participant self and was showing to have difficulties carrying out the task. The participants' data was removed from the collection. A new OLS regression was made by the remaining data which resulted in a new unusual observation. It identified a participant that have had problems with her own smartphone for the last two weeks, which has forced her to learn a new typing method by only writing on a specific part of the screen. During the test she often used the same work-around-pattern on the test handset which could be the cause of her unusual results. A new OLS regression model was made from the remaining data with ( $R^2$ ) = 80.5 percent, which therefore means that the regression has a high degree of explanation to the the whole variation in the data. The two main predictors of the regression model are participant itself and auto-correct, on or off. The participant stands for 57.8 percent of the variation while auto-correct stands for 17.5 percent of the variation. Which explains the large explained variance by the regression. Although the participant stands for a big portion of the variation, auto-correct still has a significant effect ( $P < 0.05$ ). With auto-correct turned on the users CWER will be lowered by 2.88 with the standard error 1.08 and confidence interval (-5.13, -0.63). The big percentage of the variation that is explained by the participant holds importance, where the variation is bigger between participant-to-participant than the auto-correct setting.

## 4 Discussion

This study let's all the participants perform the user test on the same smartphone, which can be discussed. The argumentation to justify the decision is to make sure the users are using the same algorithm of auto-correction. This is required to draw a final conclusion of the result. It is also justified by the fact that all the users should have the same conditions. So that the test can be presented in exactly the same way for all participants. Another argument is the anonymity aspect. If the test were performed on the participants' own smartphones, it would be hard to make them feel that they are completely anonymous. Since the test required screen recording, it would require that the participant has that application or function installed on their smartphones. They would probably not feel comfortable with having their personal screen recorded. On the other hand, each participant is used to its own smartphone and could therefore, be affected by the unaccustomed to write on someone else's smartphone. The unaccustomed

factors may include the user interface, screen size and settings for their own auto-correction. The settings for physical feedback as sound and vibrate when a key was pressed could also be a relevant factor.

However, auto-correction includes a learning function that adds new words based on the user's patterns of use. In a future research it may be interesting to investigate with more iterations than made in this study. This study focused on one iteration ensured by resetting the cached data input from keyboard between the participants. That was made to ensure that participants could not affect each other. If the participants were entering the same text more than once, would the auto-correction help the user more or collect misspelled words to the user's own dictionary that would result in more error output of the text? A daily use of auto-correction includes daily iterations of entering text which is a more natural situation. The challenge though, is to create a user test where the user is entering the same text more than once without learning the given text.

## 5 Conclusion

Text entering on a smartphone can be challenging based on numbers of problems outlined in the article. The helping tool auto-correct is created to correct the users' mistakes and present a better output automatically. Our findings showed the auto-correction to work. The baseline CWER with auto-correct off is 5.0 percent for the data with the confidence interval of (3.07, 7.04) percent. The auto-correct have a significant effect of the participants' CWER ( $P < 0.05$ ). With auto-correct turned on, the users' CWER will be lowered by 2.88 with the standard error 1.08 and confidence interval (-5.13, -0.63) percent.

## References

- [1] MacKenzie, I.S., Soukoreff, R.W.: Phrase sets for evaluating text entry techniques. In: CHI'03 extended abstracts on Human factors in computing systems, ACM (2003) 754–755
- [2] Castellucci, S.J.: Optimizing Human Performance in Mobile Text Entry. PhD thesis, YORK UNIVERSITY TORONTO (2014)
- [3] Madison, J.: Damn you, autocorrect! Random House (2012)
- [4] Siek, K.A., Rogers, Y., Connelly, K.H.: Fat finger worries: how older and younger users physically interact with pdas. In: IFIP Conference on Human-Computer Interaction, Springer (2005) 267–280
- [5] Holz, C., Baudisch, P.: The generalized perceived input point model and how to double touch accuracy by extracting fingerprints. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM (2010) 581–590
- [6] Chaparro, B., Nguyen, B., Phan, M., Smith, A., Teves, J.: Keyboard performance: ipad versus netbook. Usability News **12**(2) (2010) 1–9

- [7] Goel, M., Jansen, A., Mandel, T., Patel, S.N., Wobbrock, J.O.: Contexttype: using hand posture information to improve mobile touch screen text entry. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM (2013) 2795–2798
- [8] Baudisch, P., Chu, G.: Back-of-device interaction allows creating very small touch devices. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM (2009) 1923–1932
- [9] Sakkos, P., Kotsakos, D., Katakis, I., Gunopulos, D.: Anima: Adaptive personalized software keyboard. arXiv preprint arXiv:1501.05696 (2015)
- [10] Weir, D., Pohl, H., Rogers, S., Vertanen, K., Kristensson, P.O.: Uncertain text entry on mobile devices. In: Proceedings of the SIGCHI conference on human factors in computing systems, ACM (2014) 2307–2316
- [11] Bi, X., Ouyang, T., Zhai, S.: Both complete and correct?: multi-objective optimization of touchscreen keyboard. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM (2014) 2297–2306
- [12] Ordlista, S.A.: Sök i ordlistan. Hämtad från (2016)



## Author Index

Andersson, Maja, 1

Bäckström, Tove, 19

Dahlin, Amanda, 29

Ekdahl, Simon, 45

Fischer, Christian, 55

Golles, Filip Norman, 65

Gustafsson, Viktor, 73

Henstam, Pontus, 83

Johansson, Fredrik, 93

Johansson, Simon, 101

Ljungren, Joakim, 111

Raveendran, Vishnupradheep, 125

Sporrong, Klara, 139

Wahlström, Mikael, 159

Wieslander, Freja, 175