

Master's Programme in Information Networks

# Improving usability of smart television application's search screen

---

Liisa Hyvönen

Copyright ©2021 Liisa Hyvönen

---

<b>Author</b>	Liisa Hyvönen	
<b>Title of thesis</b>	Improving usability of smart television application's search screen	
<b>Programme</b>	Information Networks	
<b>Thesis supervisor</b>	Prof. Marko Nieminen	
<b>Thesis advisor(s)</b>	Erno Paunonen	
<b>Collaborative partner</b>	Finnish Broadcasting Company	
<b>Date</b>	<b>Number of pages</b>	<b>Language</b>
22.11.2021	70 + 8	English

---

### **Abstract**

Televisions are a common household item and today most of them are smart televisions. Smart televisions have many similarities with older televisions such as remote controllers and a big screen while differences are time boundlessness and watching habits. One habit that comes from smart television is the common usage of search. Search is essential to find a specific title from a service. That is why it impacts the user experience of a smart television application. The aim of this research is to gain understanding how people use the search in smart television applications to help develop the applications more usable.

This research focuses on the usability and experiences of users, and it is conducted by usability testing with semi-structured interviews. Participants for this study were five people (3 male, 2 female) between ages 26–41. The gained data was transcribed and then analysed with a software called ATLAS.ti by coding the most relevant points out of data.

Based on the methods presented before, a lot of interesting points rose out of the data. Results did point out many different details that do affect the usability of a search. The goal was to find out which keyboard works best on-screen and based on the results, a square shaped alphabetical keyboard would work best with the numbers in a separate screen. It was found out that users have a need for Scandinavian letters and the possibility to move through edges of the screen. It was also found out that users expect the search to happen immediately while writing their search query. In addition, the search results should be shown as pictures to visualise the search.

In the future it Yleisradio could deploy the suggested search screen in their service. If one wanted to ensure reliability of current research, further research could be done with a wider sampling of user base.

---

**Keywords** Usability, Smart Television, Search, On-screen keyboard, User Experience

---

---

**Tekijä** Liisa Hyvönen

---

**Työn nimi** Älyteleviisiosovelluksen hakunäytön käytettävyyden kehittäminen

---

**Koulutusohjelma** Informaatioverkostot

---

**Vastuupettaja/valvoja** Prof. Marko Nieminen

---

**Työn ohjaaja(t)** Erno Paunonen

---

**Yhteistyötaho** Yleisradio

---

**Päivämäärä** 22.11.2021 **Sivumäärä** 70 + 8

**Kieli** Englanti

---

### Tiivistelmä

Televisiot ovat yleinen esine kotitalouksissa ja nykyään suurin osa näistä laitteista on älytelevisioita. Älytelevisioissa on monia yhteneväisyyksiä vanhempien televisioiden kanssa, kuten kaukosäätimet ja näyttöjen suuri koko, kun taas eroavaisuuksia ovat aikasidonaisuuden puuttuminen ja katselutottumukset. Yksi älytelevisioista johtuva tottumus on yleinen haun käyttö. Haku on välttämätön funktio, kun käyttäjä haluaa löytää jonkin tietyn nimikkeen palvelusta. Tästä johtuen haku vaikuttaa suuresti käyttäjäkokemukseen älyteleviisiosovelluksessa. Tämän tutkimuksen tavoitteena on hankkia ymmärrystä siitä, miten ihmiset käyttävät hakua älyteleviisiosovelluksissa, jotta näitä sovelluksia voitaisiin kehittää paremmiksi.

Tutkimuksen painopiste oli käytettävyydessä ja tutkimusmetodinä käytettiin käytettävyydestä puolistrukturoidulla haastattelulla. Tutkimukseen osallistui viisi ihmistä (3 miestä, 2 naista) ikäväliltä 26–41. Testeistä kerätty data litteroitiin ja analyysi tehtiin ATLAS.ti ohjelmiston avulla indeksoimalla tärkeimmät informaatiopisteet datasta.

Edellä kuvailtujen metodien avulla kerättiin paljon informaatiota käytettävyydestä. Tulokset kuvailevat useita yksityiskohtia, jotka vaikuttavat haun käytettävyyteen. Tavoitteena oli löytää, minkälainen näppäimistö toimii parhaiten näytöllä ja tuloksien perusteella neliön muotoinen aakkosjärjestetty näppäimistö toimisi parhaiten tähän tarkoitukseen. Käyttäjien toiveena oli, että numerot olisivat erillisellä näytöllä. Tuloksista huomattiin käyttäjillä olevan tarve käyttää ääkkösiä ja toiveena oli myös mahdollisuus kulkea näytön reunojen lävitse. Käyttäjät myös halusivat haun tapahtuvan välittömästi, kun he alkavat kirjoittaa hakusanaa. Hakutuloksien toivotaan tulevan näytölle kuvina, jotta pystytään visualisoimaan haku.

Tulevaisuudessa Yleisradio voisi ottaa käyttöön ehdotetun hakunäytön palvelussaan ja mikäli haluttaisiin varmistaa tulosten luotettavuus, voitaisiin tehdä jatkotutkimusta suuremmalla käyttäjäotannalla.

---

**Avainsanat** Käytettävyys, Älytelevisio, Haku, Näytönäppäimistö, Käyttökokemus

---

## Table of Contents

<b>1</b>	<b>INTRODUCTION</b> .....	<b>1</b>
1.1	FOCUS.....	2
<b>2</b>	<b>RESEARCH OBJECTIVE AND RESEARCH QUESTIONS</b> .....	<b>4</b>
<b>3</b>	<b>BACKGROUND CONCEPTS &amp; PREVIOUS RESEARCH</b> .....	<b>6</b>
3.1	USER EXPERIENCE AND USABILITY .....	6
3.2	SMART TELEVISIONS AND STREAMING SERVICES.....	9
3.3	SEARCH .....	18
3.3.1	<i>Search in smart television</i> .....	19
3.3.2	<i>On-screen keyboards in smart televisions</i> .....	20
<b>4</b>	<b>METHODOLOGY</b> .....	<b>24</b>
4.1	PROTOTYPING.....	25
4.2	USABILITY TESTS.....	31
4.3	DATA ANALYSIS.....	36
<b>5</b>	<b>RESULTS</b> .....	<b>40</b>
5.1	COMPARISON OF SEARCH SCREENS .....	40
5.2	USER EXPERIENCE.....	48
<b>6</b>	<b>CONCLUSIONS</b> .....	<b>53</b>
<b>7</b>	<b>DISCUSSION</b> .....	<b>56</b>
7.1	LIMITATIONS OF RESEARCH.....	57
7.2	FUTURE WORK .....	60
	<b>REFERENCES</b> .....	<b>62</b>
	<b>APPENDICES</b> .....	<b>I</b>
	APPENDIX 1: PROTOTYPE IN ORIGINAL LANGUAGE.....	I
	APPENDIX 2: SLIDES -PRESENTATION .....	II
	APPENDIX 3: INTERVIEW PLAN .....	VI

## 1 Introduction

All digital services are continuously developing, and smart televisions are not an exception. Televisions are one of the products that have been here for quite a long time but have changed since the beginning a lot. Nowadays most sold televisions are so-called smart televisions that have pre-installed applications, and they are usually connected to the internet. Smart televisions differ from traditional televisions by not being time-bound as people can use them whenever they want and watch almost whatever they want. According to Tefertiller and Sheehan (2019) this happens mostly web-based streaming or real time content delivery via internet.

The article made by Ficom published in 2020 '*Suomalaiset ostavat nyt äly-televisioita - myös nettiyhteyksiä päivitetään*' tells that approximately half of the households in Finland have a smart television and there is still a great interest and possibility for smart television market to grow in Finland. The expectation for this market to grow is because according to the article every fifth person is interested in getting smart television in the future (Ficom, 2020). It seems that all televisions in households might be smart televisions in the future as the need to access streaming services grows.

Yleisradio, or Yle, is a Finnish Broadcasting Company. According to Yle's own website, their purpose is to increase understanding of each other and the world. Yle aims to produce content and services for every Finn and succeeded quite well as according to a KMK-survey made in autumn 2020 it is stated that Yleisradio reaches 96% of the Finns weekly (Yleisradio, 2020). From Yleisradio's "This is Yle" -page it can be found that Yle has a lot of different services such as television channels, radio channels, news offices, online services and they even produce a lot of content in different languages. (Yleisradio, 2018)

The media is emerging all the time and that is why Yle needs to follow trends. Because Yleisradio has its own streaming service, called Yle Areena, they need to develop and adjust it to stay relevant for their users. One way to follow trends and stay relevant for users is by focusing on user experience of a product.

User Experience is everywhere. As a consumer, everyone experiences things and performing tasks with products might make people feel either satisfied or unsatisfied – either about ourselves or about the product. Sometimes it might feel like we as users are the reason the product is not working when other times, we tend to blame the product itself. According to Roto (2008) there are many different aspects that affect the experience with a product one of the most important ones being the context. When talking about smart

televisions, the context is important as they differ a lot from other digital devices. When compared to for example mobile, they are in a specific location such as living room.

A book by Rubin and Chisnell (2008), explains that one part of usability is called *usefulness* and that even if a system is other ways achieving the definition of usable product, if the product does not achieve its goals of a specific user, it will not be used even if it is given away for free. So, even while Yle is funded by Yle tax and therefore they are producing their services for free, it does not mean everyone is using their services. This might be because users might feel like they gain better value from somewhere else. All Yle's products have competitors from other companies are making similar services for people. If a user doesn't find a product useful for themselves, they might change to another service which they find more relevant for themselves. And when talking about streaming services, there is already a lot of competition in the field.

## 1.1 Focus

The reason for this research is a general need to develop Yle Areena more usable on smart televisions. In autumn 2020, Yle did research on smart television applications, and one theme that rose in all the four user tests was the common thoughts about search -feature and how it has been applied differently on each application they use. People quite often tended to remember which applications had the best but also the worst search function.

The aim in this research is to create a design guideline for a search screen on smart television applications, which could be used in developing the said applications. This will help Yleisradio to develop its streaming service to be better in the future. The search feature is a big part of current television usage as it is a way to find out if a specific content on each application. This way of using search differs a lot from the original use of televisions as in the past users just tried to search watchable content by switching from one television channel to another one. It seems that there is not much publicly available research about search -features in smart televisions and it is still a relatively new field. One reason for the lack of research might be that most of the research is done in commercial companies which do not publish their research and use it only for their own product development. Another reason might be that television is a device that is so common that people might not feel like it needs developing.

The whole project started by doing a quick analysis of the old transcriptions from user research done by Yleisradio in the autumn of 2020 and finding out if there were any specific features that did raise the emotions of users. From

this data, I found out that there is a need to develop the search feature in smart television applications.

After choosing the focus inside the application research began by searching up for former research with similar interests or relatively close to this one. Finding studies related to this helped to define the field and what has already been done. After this, I decided to brighten my knowledge of different ways of interacting with the search feature in smart television applications. This research will be focusing on studying qualitative data as that is the main source of gaining knowledge of users' experiences and expectations towards services.



## 2 Research objective and research questions

The research objective of this thesis is to investigate user experience in smart television streaming service applications. This thesis focuses on the search - features and especially the interaction between user and application. By researching the subject, the aim is to improve the user experience of said applications. This study seeks to create a better user experience based on data and to reach this objective research questions were created to clear the focus while doing the research. The starting point and the research objective of this thesis is following:

*RO: How to improve Yle Areena's search screen in smart television?*

This research objective is the base of where the whole project began and while the focus of research was formed, a decision to study deeper into the main objective about the usability and keyboards of search function was formed with two different questions. These questions go as following:

*RQ1: What things affect the usability of search screen in smart-tv applications?*

*RQ2: Which on-screen keyboard layout is most usable on smart-tv?*

As discussed in the introduction, smart televisions are a growing field and while televisions are devices aimed for everyone, it can be argued that the usability of this field is something that should matter both to users but also to designers. In this research the research method is by usability testing to study these research questions. Other studies with similar kind interests regarding to usability are using same methods to study the topic. Usability testing was used for example in study made by Augustine and Greene (2002) where the usability of a search in a library website was studied with end users. In that study the usability testing worked well and gave the researchers meaningful data. This and other similar research that have worked well to enlighten the understanding of users show that usability testing is a good method to understand this kind of research questions.

This research aims to answer these questions by investigating the topic with few different means. First of these methods was scoping the literature. This literature review began with investigating previous literature about user experience and usability. And after this learning about smart televisions and later digging into search and different keyboard layouts used in smart-tv applications. Literature review can be found on the chapter 3 of this paper. Based on existing literature, the methods for the qualitative study were decided and the focus of the interviews was formed. Part of this study was to

form and test out a prototype with the end users. Methodology of this research is opened in the chapter 4 of this paper. Methodology is divided into three parts, one of them being the prototyping and another one being the user testing. This leaves the last part being the data analysis phase where it is discussed how the analysis is made in this research.

The results that were found based on this kind of research questions can be found on the fifth chapter of this paper. The results are divided based on the research questions, first subchapter explaining results that were found about the first research question about the keyboards of a search while second subchapter focuses on the other usability attributes and aims to answer the question to the second research question. The results will be concluded in the sixth chapter, conclusions where the most important parts of the results are discussed and the answers for the research questions are explained. In the conclusion chapter a recommendation for the future is made based on the results. This thesis ends to conclusions chapter which is divided into three subcategories. First one compares the findings with previous research and in the second one the limitations of this research can be found. The third and last subcategory explains what kind of research could be done about this topic in the future.

### **3 Background concepts & previous research**

This chapter aims to establish the basis for the study. First, an overview of user experience and usability. Then literature aims to learn about smart televisions and usability regarding them. After this chapter, this paper will concentrate to search and later focusing search functions on smart television. In this part the paper also discusses the interaction with smart televisions and what kind of on-screen keyboards in televisions exists.

#### **3.1 User Experience and usability**

Experience regarding a service or product is often called User Experience, or sometimes in short UX. Regarding the experience of a product, International Organization for Standardization, or later referred as ISO and more specifically ISO 9241-210:2019 -standard describes user experience as “user’s perceptions and responses that result from the use and/or anticipated use of a system, product or service” which is a quite compact way to explain what user experience is (ISO, 2019). With this description it is easy to apply this to almost everything regardless of if it is a physical or digital product – or even a service.

Designing for UX is a part of product development but it might sometimes be overlooked by companies when developing a product or service. Sometimes the reason for this could be that the company does not have resources or is not interested in creating good experiences for users but sometimes the reason might be just an overall ignorance to the topic. Oftentimes taking UX into account is a crucial part of how the product, and therefore the company behind it, is perceived by the users. By creating good UX for services improves the attractiveness of a product and by that might gain more customers for a company that provides the service.

In the online book by Garrett (2010), it is explained that with the term user experience people often mean “the experience the product creates for the people who use it in the real world.” The main point of user experience is that the user is the key subject and experience consists of the user interacting with the product while this all happens inside of a context. By “the real world”, Garrett (2010) means that people get experience by interacting with it inside where it is supposed to be used or where the user wants to use it. For example, people might today use their mobile phones almost everywhere, but smart televisions on the other hand are usually utilized in people’s homes and more precisely in a single room i.e., the living room. So, the concept of usage differs from how the product is intended and wanted to be utilized. Often, users have an intention to do certain things with the product, such as order a pizza or

entertain themselves. Intentions with a product influence users' expectations about the product and how they end up experiencing the interaction with a product.

Garrett (2010) also states: "The biggest reason user experience should matter to you is that it matters to your users." Even though a user might not know themselves what user experience means, they will recognize desired as well as unsatisfactory feelings that arise from interacting with the product. And as Garret mentions it is the impression that the whole organization is created by interactions with the product they are utilizing. Also, it is explained that if users are not happy with a product, they might find a better alternative. In a philosophical point of view, Roto (2006) questions if a user experience is a sensation, perception, emotion, mental state, or an attitude. She also explains that UX can be either just a single key click experience or a multi-year experience with a product. There is no simple way to describe what is user experience.

Halvorson (2010) discusses in her paper that quite often it seems that designers often have the attitude that design implies everything but the content itself. This might cause problems as the content someone else creates might create a conflict between the design and the content. It will influence the experience users get from interacting with the product. That is why Halvorson suggests that designers should take content into account also as it affects the UX of the said product. In a context of smart-tv applications it can be discussed that programs on a streaming service affects the user experience. If the contents in a service are not found meaningful, the usability of a service does not matter. A paper by Sutcliffe (2002) describes attractiveness being something that makes the user stay on a site. The content of a site or service is part of attractiveness of it. After a user has made the decision to stay on certain service or a page starts the experience with interacting with the product itself.

To create usable products, we as designers need to do research to understand what kind of things users want from the product. Garrett reminds that designers are not designing for themselves, and therefore, it is important to understand who the users are and what kind of features and interactions they need and want from the product. Studying users is often quite complex, because some products have a diverse user base and therefore designers need to do a lot of research to understand the wide user base. That is why it is important to define who users of a product are before starting to study further. After defining who the users are, it is crucial to understand users. By researching the user needs, it is possible to strengthen our expectations about them but also to deny what we have expected users to think about a product. By user research, means studying users' needs and hopes by interviewing the users and testing the product with users. (Garrett, 2010)

In a book by Garrett (2010), he explains that the practice of aiming to design good user experiences is called user-centred design. The idea of user-centred design often is to keep users close while developing and consider them during every step of developing the product. The online book by Benyon (2019) has similar thoughts as Garrett about good user experiences and tells that UX designers should aim to design services that are enjoyable to use, that are useful and preferably also enhance the lives of people who use them.

With this kind of user-centred product development, companies usually try to create usable products. According to Garret (2010) by creating usable products, companies aim to attract more users for their products. In a way user-centered design is also a business decision during the product development. Focusing on the users while designing will take more time and cost more. If a company does not find advantages about designing for users, they might want to pass the usability part when designing a product if they don't see value in it. On the other hand, leaving user-centered design out of the design and planning phase of a product might save a company's resources, but also affect the product's usability in a negative way.

In a paper by Bevan (2009) it is compared and explained that if user experience includes all kinds of behaviour with a product, it could be presumed that it should include user's effectiveness and efficiency. Based on this statement it can be argued that technically usability is a part of user experience. Related to usability, Rubin and Chisnell (2008) describe usability by the following words: "what makes something usable is the absence of frustration in using it." One can assume that as consumers, most people might recognize this feeling. People don't often think much about things that are working as they should, but when a feature in a product frustrates, that is when people tend to comment how badly something is working.

There is also a longer definition of usability as it is quite a complex concept. Usability as a context can be divided into multiple smaller subcategories. Rubin and Chisnell explain these subcategories in their book. These subcategories are called: usefulness, efficiency, effectiveness, satisfaction, learnability, and accessibility. (Rubin & Chisnell. 2008) The existence of these subcategories might make feel like they are creating usability a complex and harder to understand for some. But for others these subcategories might also help understanding what kind of things are included into the context of usability.

Usefulness as a part of usability is described as users' willingness to use the product in the first place. It also enables a user to achieve their goals while interacting with a product. If a user does not have any motivation to use a product, the product can be described as a useless product. Efficiency, the

second part of usability, means how fast the user's goals are accomplished. In another definition efficiency is described that it also includes the learning time of a said product (Frøkjær, et al. 2000). Rubin and Chisnell (2008) on the other hand argue effectiveness as being only something that happens with the product as intended. (Rubin & Chisnell. 2008)

Learnability by itself is also described as part of usability, and it can be linked together with effectiveness as to how much time it takes for a user to learn to use the product. Last subcategory of usability is called satisfaction which refers to the user's perceptions, opinions, and feelings about the product itself. Rubin & Chisnell say that often these subcategories are bond together as users tend to be more likely to perform well with a product that users find useful but also provides satisfaction. (Rubin & Chisnell, 2008)

In the paper by Winter et al. it is explained that there are other attributes too that are linked to usability and should be considered when developing a product. These attributes mentioned in the paper are called reliability and maintainability and they are often described as a part of usability even though not all papers about usability admit it's relation to usability. Maintainability is described as a quality attribute that helps developers to generate guidelines for a specific task but also it is described being a good basis for analysis and measurement. (Winter et al., 2008)

Rubin and Chisnell mentions that one thing which should be considered when developing a product is accessibility. They explain this term very broadly – by describing that accessibility means that a product is usable for people who have disabilities. It is also mentioned that when making a product more usable for people with disabilities or interacting with the product in special contexts it is often beneficial also for those who do not have disabilities. (Rubin & Chisnell, 2008) When designing a service such as smart televisions it is important to include all kinds of people, even the ones with disabilities. If designers remember to consider accessibility, it helps all users to interact with a service.

### **3.2 Smart Televisions and Streaming services**

Televisions have changed a lot within the last 15 years. They have changed by changing how people watch, what is being watched, but also what is the definition of television. One of the main causes that nudged the whole television industry towards a big change, was in 2007 when Netflix introduced their subscription based streaming service, where content was delivered in real time via the Internet (Hosh, 2020). Before this all content has been provided by media companies and watching television has been time bound meaning it happened in a specific moment, therefore a user needed to be present at



the right time when the media companies decided to broadcast their program. Also, one of the big differences between old and modern televisions is about the practices of looking.

In a paper by Lotz (2018) it is explained well that earlier the television mechanism had one single signal from one (media companies) to many (people watching televisions) when today the distribution happens through the internet which allows on-demand access and gives viewers a lot of options. A paper by Tefertiller & Sheehan (2019), describes this time as the “post-network” age of television where viewers have more choice to select what, when and where to view from plenty of options that each streaming service has.

Televisions have a lot of different features such as web browsers, different applications (like the ones in mobile devices), games, connections with phones, conference calling and timers. Even though smart televisions have a lot of different features, people mainly use the streaming services. Perakakis (2017) explains that users don’t like to interact with web browsers on televisions because they find the user experience of them so bad. Many websites don’t have responsive scaling suitable for smart televisions even when they are intended to be used in all different kinds of devices. This might be since smart televisions aren’t yet seen as mainstream for the and developers don’t see the value for building responsive websites as televisions have their own requirements when compared to computers and mobile devices. One of the big limitations in building web sites suitable for the smart televisions is the clumsiness of a remote controller which is used as an interaction device most of the time.

One of the big features television applications do have is so called multiplatform television. Multiplatform television is described in a paper by Shin and Shim (2017) as a possibility to use the same services with different devices. For example, a person can start watching a series first on their smart television but then leave the house, while continuing watching the same series on the train with their smartphone. Shin and Shim (2017) also explain that users want multiplatform interaction to be flexible and smooth to transfer from one device to another. Responsibility to create smooth transfer is often on streaming service developers and many of them are already doing it well as their services work well on multiple platforms and change between them works smoothly.

While televisions have changed from time-based watching to mainly streaming services it gives users freedom when and what to watch. These streaming services often have wide databases of content. The big amount of content makes finding what a user wants to watch hard. And sometimes finding wanted content needs a lot of decisions from users, and they need to

browse through big chunks of content before finding what they want to watch.

According to a report by Ericsson ConsumerLab (2017) when using streaming services people tend to spend 24 minutes on average daily browsing for content they want to watch. By browsing Ericsson (2017) means going through available content and checking if there are something worth watching. Some users have decided to go around this problem of browsing by repetitively watching the same series or movies repeatedly. This way of watching could be compared to re-runs. This kind of behaviour is explained in a paper by Gilbert (2019) where she explains that this kind of viewing sometimes happens to overcome boredom or to relieve loneliness while the viewer reacts to the content mostly passively by doing something else at the same time.

Browsing for content seems to be a big step while using televisions nowadays and research made by McNally and Diederich (2019) goes deeper into the topic and explains that people have different intentions while browsing for content. For some television users this content browsing means to search what to watch when kids are sleeping, others want background noise while working and sometimes people want to just focus and watch a thriller. There are different needs for content as well as different situations people are using television. McNally & Diederich (2019) also explain that one way of browsing for content is by searching it on their mobile phone in advance and then watching with a television when it is the right time, but for many the browsing happens with the end device which in this case means smart television. One could assume that the reason behind why people are browsing in advance with their mobile phones might be because of poor usability of smart televisions.

When users make these choices and browse through available content using the smart television applications, it provides these applications a lot of data. Applications often gather information about what kind of content this specific user is interested in watching as well as more comprehensive idea what titles are trending. With this collected data, the application makers can and try to provide better customization and suggestions for users. For example, according to Lawrence (2015) Netflix does this determination of user's preferences in two different ways. One way of finding out user's preferences is by judging users' preferences from interactions with the system and another way is to strictly ask what a user prefers (Lawrence, 2015).

By serving this kind of suitable content for users, services aim to create a good user experience. This algorithmic customization of content that is shown to the users can be called Algorithmic experiences or AX, as Shin et al (2020)



describe in their paper. It is explained that the algorithm aims to filter information and identify groups with similar tastes to target users. And after this grouping it combines the ratings inside each group to make recommendations for said user (Shin et al., 2020). Recommendations are also mentioned in the paper by Lamkhede & Das (2019) where it is explained that the recommendation system should work in a way that it helps users find something they want to watch from the streaming service. With good working recommendation system might lead to users avoiding using the search of the application if they tend to follow the suggestions.

Even though streaming services often aim to serve as engaging and try to create as good user experience as possible for their users, still according to Brasel and Gips (2011), users tend to engage with other devices while watching television. This affects how users are dividing their attention between the devices. When people need to divide their attention by focusing on multiple tasks it complicates executing the tasks. That means that the tasks should be relatively easy to execute even if the user was multitasking, otherwise users might get frustrated or in the worst-case situation -fail to do the task(s).

Study by Neate et al. (2017) explains that oftentimes the reason behind people using multiple displays is because of a need to understand a part about a program they are already watching or because the program they are watching is not stimulating enough to have their full attention. This is something that smart television designers should keep in mind while developing applications as it might affect how their products are seen and even lead users to try out competitor's services. When designing the streaming services, considering this kind of user behaviour might help to create more usable and engaging service.

A lot has changed with the interaction and behaviours related to televisions but there still are features that remain the same in modern televisions compared to older ones. One of these similarities is the remote controller as the main interaction device. One matter that should be considered while studying smart televisions is the fact that there is not much standardisation in remote controller design so each device producer can have a differently working and looking remote controller. When each device has a different layout, it complicates learnability and changing between devices might frustrate the users.

Remote controllers were not originally designed to write text, as their original purpose was to browse through linear television (which usually could be done with number -buttons as the channels were assigned to their own numbers.) Modern controllers usually consist of arrow keys (up, left, down, right) and

they often have an ok/enter button in the middle of arrows. Enter is most times used to confirm the usage of a certain interaction that can be seen on the screen. Because remote controllers are the main interaction device between the user and television, it affects the user experience with the device considerably. Nowadays users need to write things with the remote controller while searching for content and this either happens with the arrow keys of the controller to interact with on-screen keyboard or with the T9 keyboard on the remote controller.



**Figure 1.** Modern remote Controller. Arrow keys that are often used for moving inside smart applications can be found in the middle of the controller. T9 keyboard is located on the upper part of the controller.

Remote controllers could have been designed differently if writing was their original and main purpose. Now that it is one of the many ways to use television, users sometimes find writing with remotes slow and frustrating. Ingrosso et al. (2015) talks about problems with the remote controller in smart television, one of them being issues with text-input while using T9 -keyboard, but there were other problems identified such as problems with system response lag, shutting down systems by accident and other problems with the remote controller itself. Because of these problems, participants of their study felt like remote control is not a suitable input device when compared to other digital devices such as smartphones and computers.

A study by Jang & Yi (2019) discusses the problems with remote controller usability by discussing that usability of controllers is something that users often overcome by learning how to use it and getting used to using it. After learning to use the remote controller the usability problems did decrease as the users did not mention usability problems of a remote controller later. Based on this the biggest usability problem in remote controllers might be the learnability and how time consuming it might be for the users.

Even though there are a lot of different studies about user experience in general, it seems that the number of studies regarding user experience in smart televisions are relatively hard to find. In a paper by Jang & Yi (2018) it is explained that they searched for other UX studies about smart televisions written in English and the total number of papers published between 2005 and 2016 was 14. This gives perspective of how hard it is to find related studies regarding a topic.

Even though there are not many studies about user experiences in smart television applications, it seems like experience with using remote controllers is repeatedly mentioned in papers about the topic. In a paper by Tanabian and Tanabian (2015) the remote controller is also mentioned in discussion about what kinds of things affect the user experience of a smart television device. Remote controllers' bad usability is likewise mentioned in the paper by Jang et al. (2016) where it is described as one of the worst user experience problems of smart televisions. It is explained that even if the simple design of a remote makes it aesthetically appealing, Tanabian and Tanabian (2015) suggest that full keyboard on a remote controller would overcome some of the problems with a remote controller.

Even though studies of usability in smart tv were limited, remote controllers were not the only major problem with usability. Study shows that there were also problems with inconsistency and language in smart television applications, which both influenced user experience negatively (Miesler et al. 2014). In the same study also response times of applications were pointed out in with problems in feedback for users.



**Figure 2.** Apple TV Remote Controller. Apple has decided to minimize the number of buttons by creating a simple remote controller. (Picture by Yarema, Unsplash (2021)).

For example, Apple has tried to overcome these problems in remote controller usage by designing their remote controller in a simpler way when compared to regular remote controllers. (Figure 2) One of the biggest differences between a traditional remote-control layout and Apple TV's remote is that in Apple's version has a touchpad, which can be used for movement to the alternative of moving with arrow keys. By reducing the number of buttons on a remote, it most likely reduces the time needed for learning to use it, which could lead to more positive user experience while using the remote.

It was easy to find studies about remote controller usability problems, but this rose an interest to study further if developers had tried to overcome this problem. A paper by Yang et al. (2016), explains that there has been research about voice-controlled televisions and gesture controls in televisions. Popovici and Vatavu (2018) also studied the usability of gesture controls and concluded it would be a suitable way of communication with smart TV with the exception that the biggest limitation of their study was that participants were aged between 22 and 28. This problem can be pointed out with research by Guérit et al (2019) where it is explained that gesture control works well on participants of young age but they suggest this occurs because they have a deeper understanding of proposed gestures. It was found in their study that older users over 58 struggled with using the gesture controls (Guérit et al., 2019). As of the nature of the televisions, they are devices for everyone, which

means interaction should be designed for all ages and this might be the reason behind why gesture control hasn't been adapted more widely as a main control in smart televisions.

There are few different ways to do search in smart televisions. Different television brands and different applications use on-screen keyboards that have different layouts in them, but there are also other ways to interact with the television. One way to interact with television is voice. According to a paper by Rao et al. (2017) one reason voice control is increasingly gaining users in televisions is because the smart-home applications are becoming common and therefore people are used to using their voice with interaction. They also explain in their paper that voice recognition still has some problems, but if these are fixed and voice control would become more efficient, using voice recognition would be a pleasurable experience for the user.

Yang, Liu and Lo (2016) studied using mobile phones as a remote controller on smart televisions. They describe that using mobile phones worked well in their prototype and it offered a user-friendly experience to the smart television users. Based on this study it could be expected that television developers would have implemented mobile phone usage to their televisions. By studying this topic further, a lot of different applications for this kind of control could be found for phones.

Mobile phones as remote controllers were also mentioned in a study by García, et al. (2016) In this research it was studied if mobile phones would be suitable for that option. They found out that when people use phones as remote controllers, users often try to focus on looking at the television screen instead of glancing at the phone. Ouyang and Zhou (2019) studied with usability testing that using a mobile phone as a controller had more problems in success rating compared to remote controllers. It was suggested that this might be due to mobile phones did not have haptic feedback and it requires users to change their focus between the smart television and mobile phone. Even though Ouyan and Zhou (2019) argue against mobile phones as a remote controller, the research made by Sun, & al. (2015) studied the same subject with different methods. Their study was measured with a System Usability Score (SUS). According to the scores of SUS researchers argue that touch gestures on mobile devices were preferred over hard system keys and users felt satisfied with the tested system (Sun et al. 2015).

According to a report by Grand View Research (2021) in 2020 the smart television market was 268.9 million units, and they state that it is expected to expand to 1.18 billion units by 2028 (Grand View Research, 2021). Based on the market size is so huge, one could assume that the television business would have reckoned the accessibility in their devices. There has been some



research about accessibility in smart televisions. A paper by Vatavu (2021) argues accessibility problems in smart televisions. In that paper it has been explained how much this has been studied and it is described that there are still a lot of things to address before one could describe smart televisions as accessible. There were a lot of problems in accessibility in smart television development as only 4,23% of papers do address topics about accessibility in television (Vatavu, 2021).

One way of creating more accessible services is by focusing on creating usable products for those who are novice and need the most aid. Dou & al. (2019) discusses in their study smart televisions still having quite bad user experience to this day. They list the most common usability issues based on their research to be the following: Search functionality, mobile phone connection and television shopping. One of the biggest reasons for these giving bad user experiences was the inconvenience of input methods. Study did focus on Chinese elderly using smart televisions, but the problems are relevant as smart televisions are designed for a wide range of people. In all design work it is easy to focus on the experts as most of the time designers and developers are experts themselves. When someone acts as an expert, it is easy to forget how to create systems that are usable for the novices and people with limitations. In a study by Rice and Alm (2008) it was stated that there is a lot of work needed to design interfaces that support the skills and abilities of an aging population. In this research it was advised to do more research about the older adults using smart televisions to create more usable interfaces.

Regarding the usability, one way of avoiding the negative emotions of a user towards the product is by optimizing the product to work efficiently. One way of doing this is by optimizing the application's performance. McKay (2017), Netflix employee, explains in his blog that Netflix chose to use React to build their streaming service application in smart televisions to optimize the performance of said application (McKay, 2017). React is a popular JavaScript library that is used for building user interfaces on the web, but it is also suitable for television applications. Smart television applications and the web have other similarities such as big screen size and a huge user base where people's expertise varies a lot. All these aspects influence how the interaction should be done and how it will affect the user experience.

In overall there doesn't seem to be much standardization regarding smart televisions, which might be the reason why there are so many problems with usability of them. Awale and Murano (2020) discuss in their research about design problems with Apple TV through heuristic evaluation and suggest that if designers have been following user interface design guidelines, as well as the already available standards, the Apple TV would be more usable. They argue in their paper that following ISO standards and Shneiderman's Eight

Golden Rules of Interface Design (Shneiderman, 2018) guidelines would have helped to overcome the problems presented in the study. Awale and Murano (2020) suggest that no one else has evaluated Apple TV's usability problems and because no other similar studies couldn't be found it might be that similar kind of problems could be found on other television applications as well, but no one had addressed them.

### **3.3 Search**

To understand search functions in smart televisions it is a requirement to first understand search in general. Search engines are quite familiar for the consumers that use the internet daily and even the youngest of internet users know how to use search engines. According to 99Content's statistics (2021) 93% of all web traffic came through search engines in 2019. (99Content, 2021). Knowledge of how to use search engines seems nowadays like a basic knowledge for those who consume media. Oftentimes in different browsers the default home page when opening an internet browser is Google. This might even give internet users the feeling that Google, the biggest search engine in the world, is the "home page of the internet" and the place where net browsing for many people starts.

Even though people use search engines in their daily lives a lot, not many understand how this piece of technology practically works. Halavais (2018) explains that search engines commonly work with "keywords" that are used to search from digital text that is in websites. Paper by Kumar et al. (2018) goes deeper into this and tells search engines are using robots that are commonly called crawlers. These crawlers are explained in Kumar et al. 's (2018) paper to work by following hyperlinks and then processing the data they find and creating the index according to found data. So, all in all, searches try to find certain keywords from huge amounts of information by crawling through the net. This all happens in a very short time which is a feature that people expect internet searches to have.

Halavais (2018) also explains in his book that there are differences between search engines. Some of them do the search by seeking to index a specific portion of data rather than doing it to the entire web. Even though search engines themselves are an enormous market these days, search functions are also common in different websites that do not mainly focus on search but have the search as one of the site's many functionalities. These searches that are found in other websites, usually focus on finding a specific thing inside their site.

Halavais (2018) describes that search on the internet can be used to find information, text, pictures or even people. Also, sometimes the search is aimed

at a specific kind of information and for example in this study Google Scholar has been used because of it is a search engine aimed to find academic texts. Research by Dudel et al. (2007) tell that one important quality of search is that it works fast. It was also argued in the same research that while search engine works fast, the quality of search results can't be guaranteed.

Currently one thing that should be considered when discussing search engines is how largely they affect what people can find. A paper by Epstein et al. (2017) talks about a search engine manipulation effect, which influences what kind of digital content is shown to the web users. Companies can build crawlers that behave in a certain way to either concentrate or avoid things that developer wants. Search engines can have political influence as they can be built biased in the first place.

### **3.3.1 Search in smart television**

Interacting with search in smart televisions is quite different when compared to PC or phone interaction as the main interaction device is the remote controller. Some remotes have the possibility to move a cursor by waving the remote itself towards the screen and instead moving with buttons it can moved in a similar way as a computer cursor on the screen. Some smart television users want to use an additional keyboard and mouse, but the applications should be designed so that they are usable with the remote controller that comes with the television as not all users have possibilities for additional devices.

In a study by Alam et al. (2019) it is discussed that even though smart televisions need more interaction compared to linear television, content searching is problematic. The reason for its problem might be because it needs many clicks. It is argued in the paper that developers should rather focus on building the interfaces in a way that it avoids unnecessary clicks when searching. They describe that this should be done in a way that suggestions would be emphasized, and content should be shown to user based on users' interests. Even though many smart television developers have improved their ways of how their content is being served to users there is still a need that if a user wants to find a certain title, application providers need to have the possibility to easily search for this content.

A study made by Lamkhede and Das (2019) points out that using on-screen keyboards are hard and slow to use and that is why they recommend that searches should work with so-called "instant search" which means that while writing, all keystrokes should provide instant results. This same point rose from the usability tests conducted by Yleisradio's research team in autumn 2020. In those usability tests people were interviewed about thoughts



towards search features in smart television applications. Smart television users seem to have interest in getting instant response from televisions to reduce time and nerves. Lamkhede and Das (2019) compare instant search to auto complete and query -like suggestions used on a mobile as it can guide users with typing, but also allows them to notice and correct possible mistakes faster.

Even though it was found that many studies about remote controllers being a bad way to interact with a television. It seems that many television makers have decided to continue developing their products with a remote controller. Research made by Téllez et al. (2016) compared remote control usage, voice control, touch gesture, hand gesture and tactile buttons and conclusion seemed to be that between these interaction practices the remote controller still seems to be the best way to interact with smart television keyboards (Téllez, Montoya & Trefftz. 2016). In a study by Miesler et al (2014) it was discussed that one of the biggest problems of user experience in smart tv is inputting text with a remote controller. (Miesler et al. 2014) Because from a streaming service point of view there was no possibility to develop a new way to interact with smart television, focusing on the keyboard and text input itself was a natural continuum.

### **3.3.2 On-screen keyboards in smart televisions**

On-screen keyboards are quite popular nowadays as smart televisions do gain more ground in people's everyday life. While on-screen keyboards get more popular, Perrinet et al. (2011) tells that there are multiple different keyboard types used in smart television applications and there is no standardisation for how keyboards should be done on screen. It seems that finding studies about usability of on-screen keyboards remains a challenge to this day and therefore some of the papers referred to are mostly timed between 2010-2016 as no newer studies were published. This might be since research is often done by companies that are developing their own systems and they want to keep information they find in studies confidential so that their competitors do not get access to the same information.

Barrero et al. (2014) states that still there is not much user testing done about the keyboards in smart televisions and most predictions about which keyboard suits the best are created according to Fitt's prediction model(s) or other theoretical analysis models. While this is model which is based on algorithms it does not measure user experience or usability in any other way than efficiency. Barrero et al.'s research was measured with the impressions of ease of use, user satisfaction and speed of use in their studies. In this study the aim was to compare different keyboard methods.

According to Barrero et al. (2014) one of the most popular keyboards even for on-screen is the keyboard called QWERTY. This keyboard type is the most common keyboard used in both phones and computers, but also some television applications use this. Even being one of the most common keyboards, the QWERTY keyboard has a major weakness when it is used in smart televisions. This weakness comes from that QWERTY was originally designed to be used with two hands and might make it lose its accuracy when used in other contexts with only one cursor-like selector.

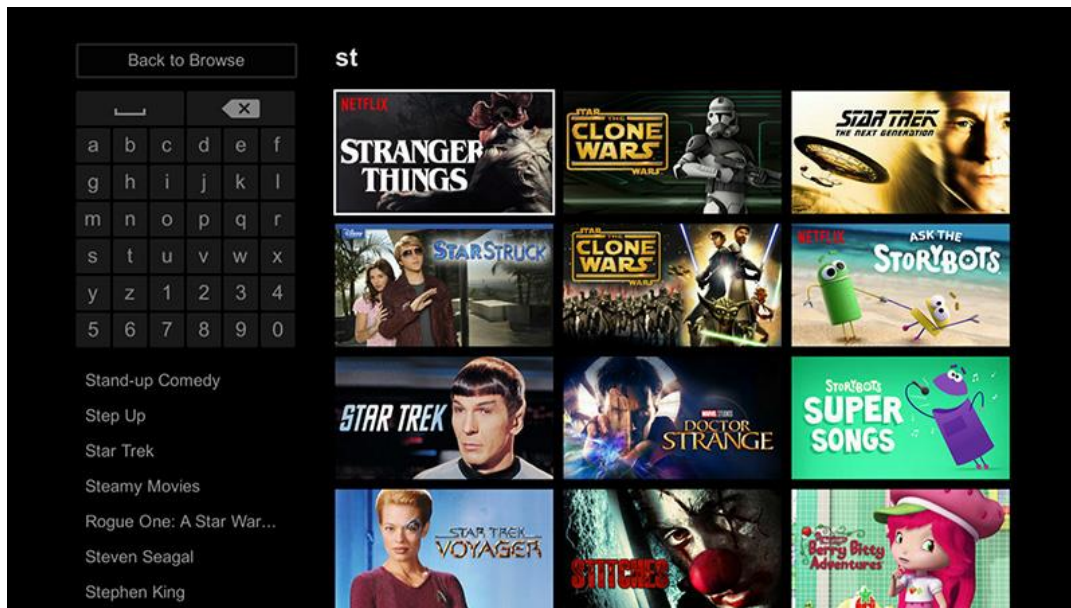
The second keyboard type stated in the paper is an alphabetic keyboard that is square shaped. This keyboard is discussed also in the paper by Perrinet (2011) where he explains it being the easiest one to use as each person who knows the alphabet can use it without wasting time on the learning process. One of the keyboards stated in Barrero et al.'s (2014) research is the one that imitates the old mobile phone input by using the number keys with a multitap method which is sometimes also referred as T9 keyboard.

In study by Barrero et al (2011) they tested multiple different keyboards speed with end users. One of the tested keyboards was so called genetic keyboard which was based on a genetic algorithm that did generate keyboard layout in aim to improve efficiency of users. The research shows QWERTY being slowest keyboard on-screen. In the study it was found that the alphabetic square was a bit faster and the multitap method would be the fastest method. Based on this, the multitap method seems to be the best way of inputting things to a smart television. But because application developers can't affect the physical appearance of the device it was decided to gather more information about search and input of on-screen keyboards. Genetic keyboard was described being almost as efficient as alphabetic or qwerty based on this study, but it was argued getting better when users learn using it.

Research made by Sinharay et al. (2013) compares QWERTY keyboard to so-called Hierarchical layout which has letters in alphabetical order. This study is based on the biometric responses in users' brains while using the system. Hierarchical keyboard differs a bit from other alphabetical keyboards as in it the symbols are organized into groups of four, the first group of symbols being ABCD. Screenshot of this keyboard can be found on the Appendix 2. Sinharay et al.'s (2013) research shows that this Hierarchical layout requires less brainwork than QWERTY but also it seconds the idea that QWERTY is not a reasonable way for an on-screen keyboard for smart televisions.

Insights 2020 -study made and published by Audience Project shows that in 2020 Netflix was the most popular streaming service in Finland with 69% of the attendees of this research answered they are using it. Yle Arena is right after it with 65% users which makes it the second most popular streaming

service in Finland. Based on this knowledge of which streaming service is the most popular, it is possible to create an assumption that most common keyboard type that is used in Finland is the alphabetic square (6x6) because that is the keyboard which is used by Netflix. Yle Areena's current keyboard is alphabetical too, but it is in a shape of 3X15 which makes the distances between letters longer.



**Figure 3.** Netflix's search screen. There is a keyboard on the left side of the screen, under it there are recommendations on how to fill the user's query and on the right side of the screen all the search results as pictures. (Tivo, 2021)

Use tests conducted by Yleisradio in autumn 2020 had repetitiveness in the mentions of Netflix's keyboard when interviewed about which services had good or bad search features. Based on these four interviews I did follow; it seems like that part of the reason Netflix's keyboard is so liked is because it's search page starts to give suggestions about what the user is searching for after the user has only written a few characters of what they are aiming to search for. Based on the assumption and correlation between statistics Netflix seems to have the most popular keyboard in Finland in smart televisions. Because of this, it was decided to pick Netflix as a layout as model that would for the discussed in the interviews. Netflix's search screen has its characters ordered in a box of 6x6 where they have all letters in alphabetical order and after them the numbers.

This literature review gave a good insight that ways to interact with search in streaming services are not regulated at all and companies seem to be doing it in different ways. Also, there are on-screen methods for input in search but

also ways to use different input methods. In this research, it was decided to focus on the on-screen keyboards as it is the primary way of interacting with the search. Smart televisions come with a remote controller that has arrow keys that can be used in navigating inside the applications.

## 4 Methodology

This section introduces the methodology used in this study.

As explained in the book by Rubin & Chisnell (2008), the hypothesis must be first formulated before conducting any usability tests. The research begun with forming the research questions. With the research questions it was easy to plan what kind of things should be studied and how. Based on the first research question, one of the focuses was to learn if it was possible to point out if any keyboard type would be more usable than other kinds of keyboards. The other main goal for this research based on the research questions was to study what kind of features affect the usability of search page in smart televisions.

Because the focus was to study usability, which is mostly composed of qualitative aspects, the decision to study by doing qualitative research was done even though some of the studies found and discussed in the literature were focusing on efficiency and brain signals. But because the usability in a more user-friendly aspect was decided the aim of this research was to focus on the experience and emotional segment of usability.

Decision to do semi-structured interviews was made as it was found most suitable for this study. Semi-structured interviews were found the best method as the topics and part of the questions could be planned, but if a participant chooses to discuss about another part of search screen than planned the questions could be adjusted based on it. Semi-structured interview was divided into three parts first one being the interview part, then continuing to usability testing in participants own smart television and finally studying what kind of emotions do rise from the prototype.

In a book by Rubin & Chisnell (2008) it is explained that one reason for doing usability testing is to eliminate design problems and therefore frustration that users might have while using the product in the future. Sometimes companies themselves might have interest towards usability because of improving profitability but in this study the sole reason for these tests is to understand users and their expectations better and therefore to eliminate design problems before they happen.

The decision to focus on the alphabetical keyboard was made based on the popularity of that keyboard and a prototype was made based on this. However, it was decided to learn about emotions and thoughts towards other keyboards and therefore this study does not exclude other keyboard layouts. Focus was on the prototype, but few other keyboards were included in Google Slides that were shown during the interview for the participants. By choosing

to discuss and include some of the keyboard on the slides rather than testing all of them, this study contains the test narrow enough even while studying about multiple keyboards and emotions rising about them. It would have been interesting to learn about other kinds of input types used with smart televisions to get a more complete idea of how people use search features, but because this study has limited resources and needed to focus on the application side it mainly includes on-screen keyboards.

Resources for this project were limited so there was a need focus on doing usability tests for a small number of participants. In the end the number of tests was 5 participants, but in qualitative research, it is more important to focus on getting quality data rather than getting it in large quantities. The usability tests were planned so that it started with a semi-structured interview to find out more information of users' background and with an introduction about that the users themselves are not tested but the products are. In the end of the interview different keyboards are discussed with a help of screenshots shown from the presentation. After interviewing, different streaming services are discussed and shown and lastly a short discussion about the prototype and what kind of emotions do rise about it. After testing users had a free word to comment on the search feature in general.

A paper by Garret (2010) discusses that there are few different ways to do user research and the chosen method depends on how deeply one wants to study the users. He explains that for example surveys, interviews and focus groups are most suitable for gathering general attitudes of users, but on the other hand user tests and field studies are best to gather knowledge about interaction with a product or user behavior related to it. The aim of this study was to combine these by creating an interview in the beginning of participant interaction and then continuing into usability testing. By this the objective is to study both user behavior as well as general attitudes towards smart televisions and more specifically search in them.

## **4.1 Prototyping**

After studying literature about on-screen keyboards and search in smart television applications, an empirical study about what kind of keyboards different streaming services is using in smart televisions was done with asking people that did own televisions and searching information on different streaming services websites. This was done to gain understanding of the field. Prototype for this study was planned based on the literature review about on-screen keyboards and combining this information with what kind of keyboards are currently on the market. Most of the current streaming services



use different versions of alphabetically ordered keyboards and with this information the decision to create a prototype in alphabetical order was made.

As technology develops it is important that products are feasible and usable, and one way of achieving this is by prototyping. Prototyping is a way to find design flaws before they happen in real-life situations by creating a version of a product that is being developed. According to Kim (2019), prototypes are used to test out if designs work as designers have intended them to. Prototyping is a phase in product development that can be executed at the very beginning of development, and it can be done with low cost.

According to Lauff et al. (2019) prototype refer to both physical and digital representations of a design, regardless of its fidelity. They also explain that the prototype usually serves to answer a question or test and assumption. In this study, the reason for a prototype because of the aim was to learn more about how smart television users perceive the search screen. Another intention for prototype was to give potential users a more real-life touch about what kind of things could be expected from a final product. One important reason for this prototype was to compare it to old Yle Areena and other streaming services' smart television applications to understand the differences and what kind of emotions do rise from each search layout.

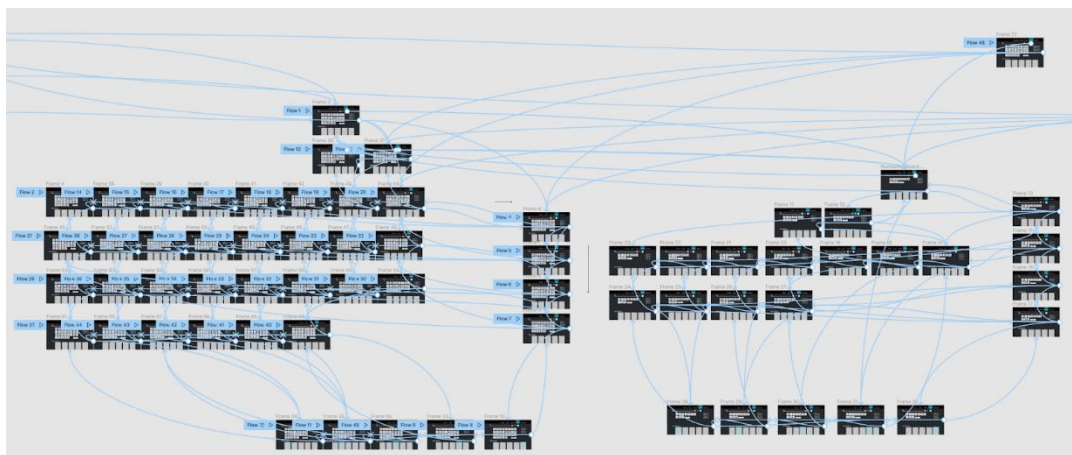
Prototyping can happen during the whole cycle of development of a product. Sometimes it is used at the very beginning, when developers need to evaluate the idea for the product but other times it is used to ideate with the customers for new functionalities or even for acquiring new customers. Prototyping can be used to understand market fields and even to find out about a competitor's product's flaws. With prototyping it is possible to understand people's user experience with products that are being developed before the product is ready or to compare different versions of a product. According to Camburn et al. (2017) there are few different objectives for design prototyping and based on the objective options, this study has an objective of exploration. Exploration is described as the process of seeking out new design concepts. (Camburn et al. 2017) In this study, the aim is to test out different version of search screens and learn the differences between them.

If usability testing of a product happens only at the phase where the product has already been published, tests might unveil problems that require a lot of resources to fix, which could have assessed at former phases if they were tested with prototypes. Creating a prototype in the beginning of a product development will help to understand user needs and what kind of expectations they have of such a product. The cheapest way to do prototyping is by doing the simplest low-fidelity prototype with just a pen by sketching on

paper. In this case there is need to create a new, more usable Yle Arena for smart televisions, which raises the need for a prototype.

The tool that was chosen to create a prototype for this study is called Figma. Figma is a prototyping tool that is commonly used for wireframing and graphic design. Figma has a good version control, and it can be used for remote testing of a prototype because it is web-based. In addition to these features the reason for choosing to use Figma was that researcher was already familiar with using it due to former projects. Familiarity with a used tool did make it easier to plan and create the prototype.

Because Figma is originally created for computer usage, there was a need to improvise a bit on how to create the search screen so that it simulates how people use smart televisions with their remote controllers. This led to the need to study a bit more about what is possible with Figma's tools. A possibility was found to create connections between screens so that the movement happens with any key of the keyboard. It was decided to do movement so that it happened with arrow keys that a physical keyboard has and to encourage interview participants to imagine that these arrows are arrows on remote. As for an ok -button to (to confirm a click) it was decided to use enter -button as it was closest to ok and would be easy to understand by the users. This button was only used for changing between the alphabetic and numeric keyboards.



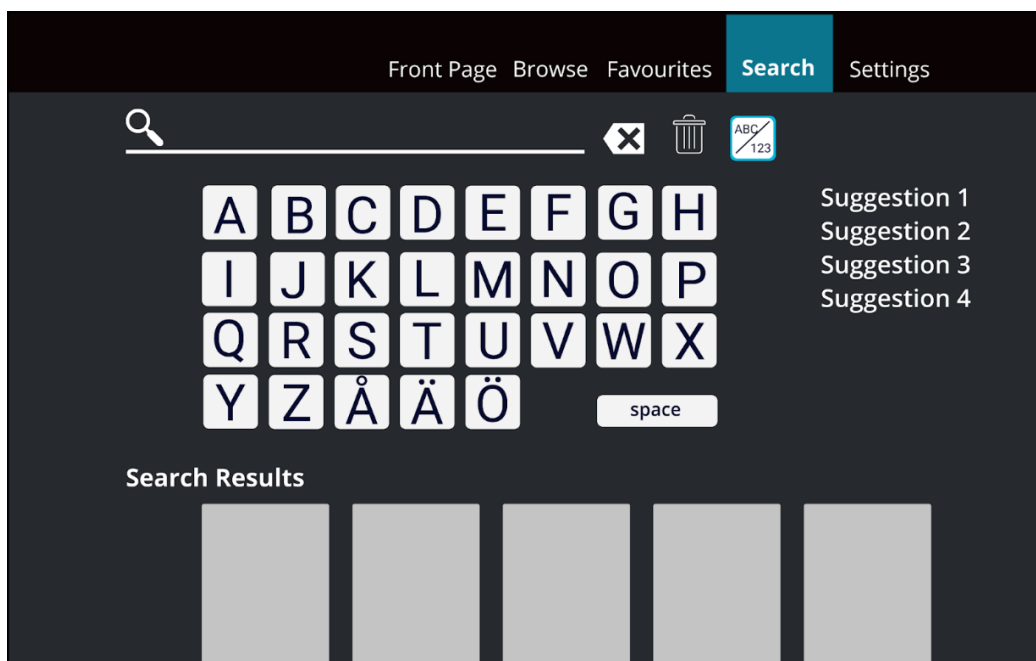
**Figure 4.** Screenshot of screen relations of the prototype.

To make the movement work on the prototype there was a need to create a total of 70 screens that worked in a way that the selection was moving to each button on the screen. Because of the need to focus on the search screen, the decision to exclude the menu -buttons in the movement was done to avoid irrelevant data and unnecessary errors and interactions with the users.



Other keyboards were discussed with other methods explained in the next chapter. With this prototype the focus was on how the keyboard and its functionalities should look and work as a completely new design often rises different kinds of ideas and thoughts in the users.

Current Yle Areena does have an additional “search” -button and one goal with testing was to find out if removing it raises any questions or emotions in users. It seems that most competitors streaming services don't have the search -button anymore and their system starts to do search queries right away when a user has written some characters, usually by three characters it starts searching. Prototype also has search results and suggestions in separate spaces as compared to current Areena which has only the search results but no suggestions. This is a functionality that seemed to be common in many other streaming services and aim was to learn if people use the suggestions and what things they expect in suggestions. This feature is often used to fill up a user's keyword to create a more efficient user experience.

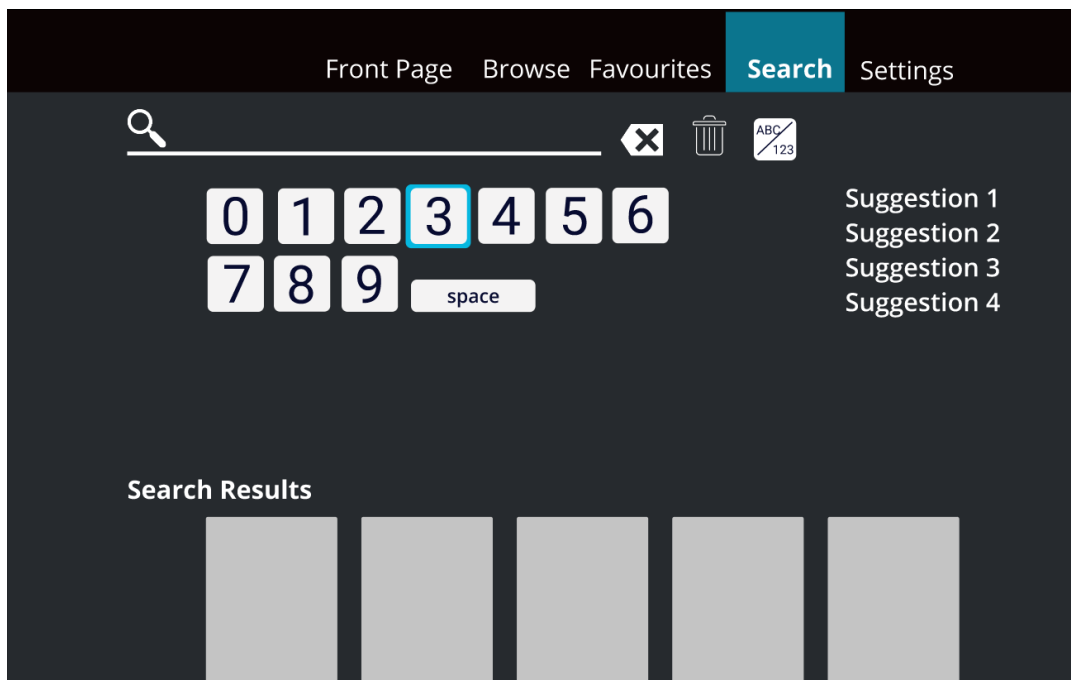


**Figure 5.** Screenshot of prototype translated into English. In this screen a user could move the selection, which is highlighted in turquoise color. Due to limited time this prototype didn't have the possibility to access the menu button. Selection of this screen is in the ABC/123 button. Prototype in the original language can be found as Appendix 1.

When compared to the older Areena, this prototype is using icons instead of buttons with texts “*tyhjennä*” (Eng. empty) and “*poista*” (Eng. erase). One of the goals was to find out if people prefer one over another. Two of these icons that were used as the buttons were from the Noun Project, and they are all

under Creative Commons license. Search -icon is made by Barudak Lier, ID and trash can is made by iconlabs, IN. The icon for backspace/erase -button is from Iconsdb and is made by Iconic.

One big difference between prototype and present Arena is the number - keyboard on the prototype is created as a different screen while in the current design all numbers are presented after the letters. This difference was made because of researcher's curiosity about if this kind of design raises any thoughts in users and if they prefer to have numbers in the same screen or not. Decision to use pictures in the search results came from researcher's instinct that Arena's current search page is not engaging as it only has the search results as a text query on top of each other. That way of presenting gives a dull feeling and as the service is intended mainly for entertainment the emotion from a service is important.



**Figure 6.** Screenshot of prototype's number screen translated into English. In this screen a user could move the selection, which is highlighted in turquoise color. Selection of this screen is in the number 3-button.

As prototype was mainly based on Yle Arena old version and Yle's graphical guidelines it was still important to follow the common guidelines of usability heuristics. In a paper by Sutcliffe (2002) he explains that these heuristics could be used to evaluate usability and they are often used to judge the quality of the interface by experts. Most known list of heuristics is Ten usability heuristics by Jakob Nielsen (2020) which is one of the most known papers in the field. It is a paper gives designers general principles for interaction design (Nielsen, 2020). Because this research was focusing on the future of Yle

Areena, heuristic evaluation by itself did not feel as an effective way to develop the service, but it was used as a guideline for creating this prototype. The list of heuristics that were included on the paper worked as a helpful reminder when developing the prototype.

In addition to these screens about numbered keyboard and keyboard with letters there was three other screens created to help the discussion with users so that they would better understand how the result could look like. These screens were not interactive, but they did include a proposition of search query, search results and search suggestions on the screen after a user has written 2-4 letters. One of these screens did have “Mu” as a search query to help discuss if any emotions do rise about the immediate search.

Even though in the prototype a user did not have possibility to move away from the search page, heuristic guideline number 1 was followed by showing the visibility of system status on the menu. This heuristic was considered also an important part of the movement on the screen as the selector on the screen was found important as it shows users what is going on the screen. The second heuristic, “Match between system and the real world” can be seen on the prototype from what kind of words are used on the screen but also, designing prototype was done with avoiding unnecessary text so that most of the User Interface of the search menu could be understood easily without any understanding of language. In the prototype this was done with indicating functions with common icons that are often used on digital services. The Backspace/delete icon does originate from computer keyboards where backspace is used to delete one character at a time, and I used the icon of it to the same function on this prototype. Trash icon and magnifying glass are often used as a delete all and icon for searching for things inside a system.

When creating the prototype also other guidelines were followed. These guidelines were created by Perakakis and Ghinea (2017) and they helped to understand what kind of things could be expected from smart television applications. Even when their paper was written with the web in mind, many things are still applicable to consider while designing an application. It was mentioned that D-pad navigation is preferred in televisions as there is no input device that could be used for pointing (compared to computer mouse). D-Pad refers to the arrow keys found on the remote controller but because of technical limitations this was done with arrow keys found on the keyboard of the computer. Also, the clear focus of selected navigation items was highlighted in the study, and they suggested doing this by stronger colouring. This highlight color was done by using the tan from Yle’s colour guides. In a paper it was explained that highlighting helps users to focus on what is relevant on the screen. (Perakakis & Ghinea. 2017)

The usage of icons can also be justified with heuristic number 3 - consistency and standards. This heuristic also considers a certain usage of words on the interface. This is important because if the language that is used is reminiscent of other similar products (in this case smart television applications) it is considerably easier for users to learn using a new product thanks to consistency. One important matter is that things should stay in one place when the user is interacting with the product. This was applied in a way that when a user changes between alphabetical and numeric keyboards they are positioned in same keyboard grid organized similarly. Heuristic number 8; “Aesthetic and minimalist design” was pursued in this design. It can be observed on the prototype in a way that it mainly focuses on the essentials, and it doesn’t have any unnecessary elements to distract the users.

Some of the heuristics are almost impossible to consider in a prototype that is only focusing on a certain part of a system, as in this case the search screen. Because the prototype did not have any real functionalities it would have been hard to go through all heuristics, but in the future when the final product of Yle Arena will be developed hopefully designers will use these heuristics as guidelines when developing the whole service.

## **4.2 Usability tests**

Steve Krug explains in his book *Älä Pakota Minua Ajattelemaan* (original Eng. *Don’t Make Me Think*, 2006) that the most optimal number of usability tests is four. (Krug, 2006) He argues this because the greater number gives more notes and analysing them will take a lot more time, but seldom any more information is found, rather than makes researchers focus on small details, those that rarely are the focus of a research. Even though he advises to use only four test subjects I decided against his recommendations and conducted tests to five different users because of another research that is made by Nielsen and Landauer (1993) where they recommend using five either heuristic evaluators or test users to find most usability problems. This amount of tests is argued because of the cost efficiency for finding problems, but also because usability testing should be done as an iterative process throughout the design phase.

Mainly due to the ongoing COVID-19 pandemic, all usability tests in this research were done with remote testing. In this case remote testing meant that tester and test subject were not located in the same place and a connection to communicate between these two was established. In this research connection did mean using different video conference software that were used to connect and communicate with the participants in real time while interviewing and interviewer suggesting the tasks for the participant.

Because tests were done remotely, there was a need for a way to communicate between tester and the participant of the test. It might have been easiest if one specific tool was chosen for the communication. But because different video conferencing tools have grown familiar for researcher, it was decided to give users a list of programs that were allowed to use for testing. The limitations for these video conferencing tools were that some features were required. All these services had the following features: video and voice call, possibility to share screen (both for the researcher and for participant) and to avoid unnecessary workload and easy use recording of the meeting. In the end there was a total of 3 different services used for this and they were the following: Microsoft Teams, Zoom and Google Meets. Users seemed to prefer the kind of service they were familiar with and almost everyone mentioned they used the service of their choice at work.

There are arguments against doing testing remotely such as some might say it might affect the quality of data collected from the testing as then users are in locations of their own choice which can't be controlled as in the location chosen by the researcher. There is always more possibility for distractions when doing the test remotely. Also, one big variable in these tests was that each participant was using their own devices such as laptop, microphone, and television. This could have caused a problem if one of the devices didn't work as intended.

On the other hand, there are arguments in favor of remote testing even in the situations without pandemic. Some of these are mentioned in a book by Norman and Kirakowski (2018) where they justify remote testing with testing in the right context of actual use of a system. Because of the laid-down nature of usage of televisions, it could be argued that test users don't get the same feeling when tested in laboratory settings and remote testing might get more valuable data. Aziz (2019) explains that field studies are often used when matters related to daily living and preserving how natural something is. In these television settings it could be more natural for the users and feel closer to real life -situation when tested at participant's own home. This way it might improve the usefulness of data collected in the research. Remote testing lets participants use their own devices, which they usually already are familiar with instead of forcing people to use non-familiar devices which might lead to unnecessary learning experience during the usability testing.

A paper by Gray et al. has a good list of ideas that should be considered in while planning a remote testing. In this study it was pursued to follow this advice. The paper considered providing the test users technical information in advance and this was done already at contacting phase where it was asked if participants had the suitable devices and if they were familiar with using the devices for it. This is important because it helps avoid technical

difficulties during the tests. Gray et al (2020) also explain that it is important to consider different kinds of distractions. The test scheduling was planned so that researcher's live-in partner would be out of house while hosting the tests to avoid the biggest distractions on her part. To avoid big distractions also on the participants' side it was asked if they had a calm enough situation and scheduled the tests in a way that was most suitable. Two participants did have kids and if kids would have been present, it could cause a lot of distraction.

Even though the biggest reason for remote testing was the ongoing pandemic, it can also be justified as a good method with the possibility to test with people that are not located near the researcher. In this test this can be seen with a wide diversity of locations on the participants, the furthest participant was all the way in North Savo while closest one could be found in Uusimaa. Another important advantage of remote testing is described in study by Gray et al. (2020) as participants have mentioned that they feel more comfortable speaking about a personal topic in a familiar space but also that in a video meeting they feel more connected with the interviewer. The personal connection between a participant and interviewer helps participants to relax and to speak about a topic more openly.

Participants of the usability tests for this study were a group of five that were able-bodied. Accessibility was not considered in this study in any other way than considering the colors of prototype that if someone was colorblind it would not affect the results. This was due to smallness of a research. These participants were people that researcher had the opportunity to contact easily because of limitations in both resources and time for this study. Focusing on the people that were easy to contact was mainly because tests were held in the beginning of autumn 2021, when many people either are still in summer holidays or just returning to work after holidays and might not be as interested in participating in that time if a completely stranger asked about participating in a study.

People that were reached out for tests were people that all use smart television and are familiar with using apps in them. These five people were between ages of 26 and 41 and there was both men and women to gain more diversity (3 men and 2 women). Sadly, people that were reached out to were homogenic in a way that they all had academic degrees. The data could have more variety if there were people with more differences in their backgrounds. Focusing on people that had previous experience in using technology was conscious decision to avoid technical issues with the combination devices. This way the technical problems wouldn't consume too much time in the remote testing situation.



All participants were speaking Finnish as their mother language and were in Finland. This was a conscious choice as Yle Areena is meant for Finnish people. I did not have any contacts nor proficiency to plan or conduct the tests in Swedish and therefore the focus was only on the Finnish speaking citizens. By doing all the tests in Finnish also ensures that it was easier to communicate results to Yle and results or quotes don't need any translation when discussed with other researchers.

Oftentimes it is advised to conduct a pilot test before the actual testing to understand if the planned test is giving the wanted answers. In a paper by Lowe (2019) it is explained that the pilot test is used also to evaluate if the intended audience does accept and understand the study. The decision was made not to have any of the usability tests as pilot study as such. The need for pilot studies has been emphasized especially when doing mixed methods studies (Williams-McBean, 2019) but as there was a decision made to focus only on the qualitative approach it didn't feel as necessary.

Instead of pilot study, after each test, the collected data was read through and based on findings, questions and tasks were adjusted accordingly to gain better knowledge from next test. This adjustment was done based on which topics seemed to give the most rich and relevant data compared to those that seemed to make users lost in thought or feel confused. This decision was made because the resources for the whole research were limited but iterative way of studying might help to understand the end users.

It is important to plan the tests beforehand and a book about interviewing discusses that field guide is a document which describes what kind of things happen in the interview. The field guide is based on what kind of research goals there are for the interviews. (Portigal, 2021) Based on the guide on how to plan interviews the structuring of the plan began. Because there was no pilot study it was important to estimate how long it takes participants to answer each question so that there would be an estimate how long the interview would take so that it can be informed to participants in advance. The structure of interview was created flexible so that questions discussed in there could be changed or dropped out if the time or need to understand requested it.

Structure of the tests were as follows: Interview began with a short introduction about who is the researcher and what is the goal of the study. Participants were also informed for which organization the research is done and this was followed by what kind of things are being collected from the test and how the data is going to be stored and used. After the introduction participants were asked if they had any questions at that point and answered them if

needed. After the introduction the recording was started and the interview itself started.

The test began with easy questions relevant to the topic to warm up the participant to harder ones. This was done with words that worked as a discussion starter so that participants were asked what comes to mind out of each word. After the warmup discussion the interview went deeper into the topic by showing different layouts of keyboards, which were discussed through emotions and which one's participants did prefer or wanted to avoid and why. Both the words and different keyboards were presented in a Google Slides -presentation that was shown to the participants by screen sharing. This Slides -presentation can be found as an Appendix 2.

After discussing through these keyboards that were included in the presentation, the participants were asked to turn on their television and turn camera towards it to show what was happening on the screen. While using their television it was asked to answer a few questions about which streaming services they are using and to show how they use search and if they are enjoying using the search. After showing 1-3 different streaming services users owned, participants were asked to use Yle Areena's current search page and tell their thoughts about it.

After discussing the current Yle Areena the participant was advised to turn their camera back to their own face. Interviewees were given a link to the Figma prototype was shared and discussion about search continued. Users were asked to share their screen to show how they were using the prototype so that it was easier for researcher to follow and advice what to do with the prototype. In this phase there was questions and aim to get some comparison between the prototype and current Areena as well as comparison between prototype and other streaming services. One reason for the interviewing done in a semi-structured way because all participants had a bit different streaming services and because it is most efficient way to gain information about experiences. The interview plan that can be found as an Appendix 3.

Oftentimes, when a service is familiar for users, people might prefer it over others even if it wasn't the most suitable for the usage. The aim of this research was to overcome this bias of familiarity by cropping the pictures on the slide show so that they did not have logos shown.

Reason behind the interviews was to understand experience and therefore it was important to form questions in a way that participants talk about themselves rather than talking too much in general. In the paper by Kuniavsky (2012) it was recommended avoiding questions like "Is this useful feature" and rather ask if the participant finds the feature useful for themselves. With



this kind of guidelines, the questions were shaped and planned for the interviews.

For the testing part of the study, where participants were interacting with their own device and prototype, it was chosen to use a method called Think Aloud, or TA in short. According to a paper by Eccles and Arsal (2017) TA is a method where study participants talk about what they are doing while doing the tasks. It is also discussed in the same paper that TA -method should be used with interview methods to gain better understanding of mental strategies of a user. Users were advised about Think Aloud method when the interaction with their own smart television started but they were reminded to continue it while interacting with a prototype.

### **4.3 Data Analysis**

By following the methods presented above a rich and broad dataset was gained from the tests. To understand and benefit as well as possible it is necessary to analyze the data. But before analyzing it is important to convert the data to proper format. The raw data that was reserved from testing which was in video format after the tests. In the video, either the participant or their television was shown. The analysis of found data began by transcribing all tests from word to word and by doing notes in the transcription if there were things that should be necessary for the analysis part - such things were emotions, faces or gestures participants did while interacting with the device. These were things that happened during interviews but couldn't be marked as something they said. After writing down each transcription the original recordings were watched once more while reading transcriptions so that it was made sure there was no information loss between these formats.

After the data was converted to suitable format, the analysis began by importing all transcriptions to the ATLAS.ti project. ATLAS.ti is a software for qualitative data analysis and the decision to use it for this project was made partly because of supervisor's recommendation and because of the practical reason that as an Aalto student there was a possibility to have access to this program. I had no experience in using the program and I tried to gain knowledge on how to use it for analysis before using real data by trying it out and by watching tutorials from ATLAS.ti's YouTube channel.

Taylor-Powell and Renner (2003) have created a useful guideline on how to do qualitative data analysis and as all data gathered for this research were qualitative, it was an easy decision to follow this guideline. After all the data was uploaded to the software, it was important to consider the main objectives of this research. By recalling the objective, the borders of the research were formed, and they help researcher to remember how and what things

should be picked up from the dataset and if it was possible to answer the questions, formed in the beginning of this project.

Based on research questions a decision to focus on the questions and topic was made, as Taylor-Powell & Rennel (2003) guides to do as alternative to focusing on individuals or groups. After this I as researcher started going through each transcription and marking all important information points. According to Taylor-Powell & Rennel (2003) this is called categorizing information and it is used to identify themes and patterns in the data. This way of pointing out important categories from data is also called thematic analysis. Evans and Lewis (2017) explain that it is specifically useful if the goal of a research is to discover participants making meaning out of their experience. One of the reasons for this study is to understand the user experience, and researcher did not have any personal experience regarding the devices, this way of analyzing the data suits the needs of a study well.

Marking the important information points can be either called coding or indexing, the idea is still the same - to point out what things are important in the raw data. This indexing often happens by tagging data that is relevant to a particular theme or question. In a paper by Elliot (2018) it is discussed that one way of going through the data is by first finding all data points that answer a specific question, rather than going chronological order, and pointing out everything interesting. The data analysis in this research follows this guideline as I went through all the data multiple times with different questions or themes in mind.

First theme that was indexed from data was pointing out things that participants mentioned regarding the usability of search in smart television. After this one big theme was pointing out things that were leading to which keyboard might be the most liked. Indexing this theme was done with multiple codes as there was a need to create a main code and code for each keyboard type. In addition to these main categories mentioned, some other categories were created as well. These categories were things like negative and positive emotions towards applications and they were created to help understanding users' emotions.

Most of the categories were formed based on familiarizing with the data a few times by reading through it and before starting the coding part. But in addition to these codes set beforehand some of the categories did rise from the data, and they are called emergent codes. In a paper by Elliot (2018) it is described that these categories often require more editing to make them more consistent with other codes. In this data it was relatively easy to make sense of these emergent codes and when looking back at them afterwards some of these codes should have existed in the beginning of coding and it was rather

researchers' ignorance and lack of experience about indexing that codes were not originally created. For example, a code for comments regarding to remote controllers were created as well as general usability index about smart-tv apps which did not affect the usability itself.

Coding of the data was rather time consuming as it required going through the data over multiple times until there were enough codes and data points related to them. Linneberg and Korsgaard (2019) argue that it is helpful to do coding in multiple cycles as it might help the researcher to consider in the following cycles the code types that are more descriptive than in the first cycle. This way it was possible to find more significance from the data.

The focus was to identify the patterns inside and between the categories defined in the indexing stage. This identifying of patterns was following the guidelines by Taylor-Powell and Rennel (2003) where they suggest organizing data to find out relationships and larger connections in the data. Even though most of the work was done with ATLAS.ti, which is the software for analyzing, I felt the need to export one of the bigger categories as a table so that I could do the smaller categorization inside it easier. This was the category "*what things affect the usability of search screen in smart-tv applications*". And aim was to categories the subcodes inside it.

I did this categorization in Google Sheets table where categories were color coded and smaller new categories were created. Even though this was done with a different software than original analysis, the qualitative analysis guidelines were kept in mind to gain meaningful data. This whole analysis could have been done with ATLAS.ti, but after going through the same data multiple times I had problems perceiving the data and felt like giving the data some space would make it easier to understand. The change to Google Sheets seemed to help as the subcodes were easier to find and understand.

After all data finally seemed to be ready, interpreting the collected and categorized data began. The final part of the data analysis was by abductive reasoning. Abductive reasoning as explained in a paper by Tavory and Timmermans (2020) is a form of reasoning through which researchers can relate unexpected or surprising observations to other observations by specifying the grounds of a plausible relationship. With abductive reasoning I intended to connect the findings of the data with theory and existing literature about the topic.

- ◇ Codes (25)
  - ◇ Apple-TV (One-Line) {39-0}
  - ◇ Apple-TV Box (6X6) {12-0}
  - ◇ Areena (ABC 3x15) {69-0}
  - ◇ Browsing {24-0}
  - ◇ Disney+ {5-0}
  - ◇ Elisa {11-0}
  - ◇ Wouldn't pay {1-0}
  - ◇ figmaproto {123-0}
  - ◇ Hierarchical ABC {22-0}
  - ◇ Most-Hated {7-0}
  - ◇ Most-Liked {13-0}
  - ◇ Negative {185-0}
  - ◇ Netflix (ABC + 123 6x6) {69-0}
  - ◇ Other inputs {9-0}
  - ◇ Positive {195-0}
  - ◇ QWERTY {28-0}
  - ◇ Remote Controller {18-0}
  - ◇ Search {46-0}
  - ◇ T9 {19-0}
  - ◇ TV + Phone {11-0}
  - ◇ Usability of Smart-Tv apps in general {15-0}
  - ◇ What things affect the usability of search screen in smart-tv applications? {269-0}
  - ◇ Which on-screen keyboard layout is most usable on smart-tv? {140-0}
  - ◇ yle {52-0}
  - ◇ Youtube (ABC 4x7) {81-0}

**Figure 7.** List of codes found on the ATLAS.ti software

One of the biggest inconveniences with the data was that the quantity of it was so large that analyzing it took a lot more time than I had anticipated beforehand. All transcriptions were between 7-12 pages long and based on this data a total of 3 code groups and a total of 25 codes were created. Codes in the analysis were created based on research questions in mind to learn if this data could answer them. To answer the research question about if one keyboard is better the codes were created based on the different services and input types.

These codes did have a lot of points of interest, the smallest codes having 1-10 interesting points while the two biggest codes did reach a total of 140 and 269 data points. The biggest code was aimed to answer the second research question with its datapoints but was divided into smaller groups such as “*suggestions*”, “*search results*”, “*other comments*” and “*write before search*”. These subcodes were created under this bigger code to understand the issue better.

Finding relevant information from a dataset this wide took a lot of time, but as seen in the results and later in conclusions, it was possible to answer the research questions. During the whole progress of analyzing the aim was to strive for understanding of the topic from the perspective of users. Results of this study are presented in the next chapter.

## 5 Results

Purpose of this research was to find if a specific on-screen keyboard is found more usable than other keyboards and to learn what kind of things are the major factors that affect the usability of a search screen in smart televisions. The quotes included in this chapter are from the interview transcriptions and they are translated from Finnish to English. The content and tone of the quotes is kept as close to the original comments as possible.

This chapter presents the results of the study. These results were gained based on the methods of usability testing, interviewing and data analysis. This chapter is divided into two main sections. The first main section of this chapter presents findings about what kind of keyboard is found most usable in smart television and aims to serve answer for the first research question. The first category is divided into smaller sections as it discusses the keyboards based on the codes presented in the data analysis phase. The second main section focuses on the user experience and what kinds of features affect the usability of search. It is not divided into smaller sections but aims to offer the answers to the second research question.

### 5.1 Comparison of search screens

One of the main goals of this research was to comprehend which of current on-screen keyboards is the most liked and to understand the concept better I wanted to clarify why or what kind of features make one service better than others. The features that do not link into any of the keyboards themselves are presented in the next chapter.

Some of the keyboards were discussed only with pictures shown from a presentation (Appendix 2) which was made beforehand and other keyboards were discussed more thoroughly when users were subscribed to them. This might have affected some keyboards with more emphasis both in negative and positive comments because when people were using the keyboard, they discussed the emotions that came from it entirely compared to only looking at a picture.

**The most disliked keyboards** were discussed with the users and, but all participants had problems choosing only one and all of them chose 2 keyboards that they disliked from the keyboards shown from the slides. The keyboard called hierarchical keyboard was the most disliked keyboard as all participants said that was one of their dislikes. As for the other keyboards that were disliked, the second hated keyboard was T9, the old mobile phone-like way of writing with a total of 3 dislike votes. There were also two other

keyboards that were disliked both with one vote: one being Apple's one-lined keyboard and another one QWERTY.

**The hierarchical keyboard** was shown to all participants. The dislike towards hierarchical keyboards can be seen from the negative emotions that arose about the keyboard. The discussion about the hierarchical keyboard was about how strange it was, and one participant even said that they would be having nightmares after seeing it. All participants seemed to understand how this keyboard would be used but few mentioned that choosing the area of alphabets felt like an unnecessary phase while writing and people wanted to avoid that.

**The T9-keyboard** was shown to all participants, and one user told she was using it with her device. The comments about T9 were mostly sceptical whether it would be good in use. There was one mention about the nostalgia from that kind of keyboard but described that the younger generation might not know how to use it. Another user said it using T9 would be "*kind of functioning*" but said that the problem might be that if the connection between television and remote was even slightly problematic it might not take all button pushes while writing. This was something that could be seen also in the test where the other user had this kind of remote and she had problems with writing. Out of frustration she sums up the attitude towards the system: "*This happens because of the devices of course, but it sucks!*"<sup>1</sup>

**Apple's one-lined alphabetical keyboard** got mostly negative comments and one participant mentioned it being the worst of the discussed keyboards, because it would be frustrating to write with so long distances between the letters. It was interesting to see that two different participants did not recognize the shown picture to be a keyboard at first and the first commented: "*Wait a minute, is this a keyboard?*"<sup>2</sup> when finally understanding what was on the screen. Neither of these participants did like this one-lined keyboard.

Also, other participants discussed how weird or hard it would be to find the letters as they were arranged in just one line. One user suggested that maybe the search function in the service would be so efficient that the function would know to suggest right things after just writing one letter but sounded doubtful while suggesting this. The one participant that used a T9 keyboard as main device commented that this one-lined keyboard could be better than T9 if there was no lag (which she later mentioned having with another

---

<sup>1</sup> Original quote. "Tämä on laitteiden syytä tietysti, mutta aivan perseestä!"

<sup>2</sup> Original quote. "Siis hetkonen, onks tää näppäimistö?"



keyboard on her device). What was interesting is that one person had an Apple-TV which had this kind of keyboard as a default. He mentioned that this might not be the most efficient, but it was clear. He also stated that this was one of his favorite keyboards, but he mentioned that this might be because he referred to himself as an Apple fanboy – as he had a lot of different devices from apple.

Because one of the participants had on-hand experience with this kind of keyboard it was interesting to hear also that even though he claimed to like this one-lined keyboard, he admitted often using phone to write the search. He suggested that maybe Apple had this kind of possibility just because the keyboard has so long distances between the letters. Also, his experience with the suggestions that are coming from this service were frustrating as it might give suggestions such as program ‘Game of Thrones’ even when the system does not have this in their catalog. As a researcher this discussion about one-lined alphabetical keyboard gave quite the idea that the one-lined is not the keyboard people wish to use even if they had a positive attitude towards it as it was on a familiar service.

**QWERTY keyboard** is one of the most common ones to this day and it was interesting to uncover the results about it. QWERTY was the first keyboard shown in the interviews and all participants did recognise the layout. One person said it was their favourite and another one mentioned that usually it would’ve been his favourite but this specific one “felt wrong”, he described it was because the buttons were tilted in a way that lines were not straight. One user who did prefer QWERTY over other keyboards admitted that it might not be the best (for on-screen), but they liked it because it felt traditional.

Discussion about QWERTY pointed out that it is seen as the best option if it is a physical version, but one participant mentioned he would prefer alphabetical order in on-screen keyboard. The participant that was using a T9 wondered if it might have been better if it just took all button clicks efficiently. In the discussion only one mentioned that they have used QWERTY as an on-screen keyboard on a bigger screen. He explained that it was used in PlayStation -service and input for it is used with a touchpad which is integrated on the controller. It was described being useful and easy to use. Keyboard from PlayStations was also mentioned in another interview where a participant explained that one of his favourite features in on-screen keyboards is to have a possibility to get through the edges and back on the other side of the keyboard.

**Disney+** service was discussed only shortly as it was not included in the presentation in the beginning. Only comment regarding its usability was that one participant mentioned it’s search function to work badly but as they did not have the subscription to this service by the time, they could not explain



further by showing what was bad about the search. There was also another keyboard that was only discussed in one interview as one of the participants realized during the testing that they could change the default keyboard in their television. He was happy about the finding as the original default was the one-lined keyboard and it changed to a Grid -shaped alphabetical keyboard. In the end he concluded that this was his favorite as it had a 'perfect-ratio' in how the letters were arranged. It was surprising that he did not know about the possibility of this keyboard and found it accidentally.

**The favorites** in the keyboards were also discussed in the interviews and there were once again challenges to answer which one was the best and most participants chose 2 best options again. One user mentioned the one-lined Apple's keyboard in the beginning of the test but changed the favorite later to being the Apple's Grid -shaped keyboard when learning about it.

Two users said QWERTY was their favorite as it felt familiar for them. One another mentioned that usually QWERTY works well, but when used on-screen it loses its efficiency and he'd prefer Alphabetical ones. The keyboard from YouTube got total of 2 votes for being the best keyboard and Areena's present version got 2 votes for being the best keyboard at this point. Later in the discussion one of the participants pointed out that they would prefer the prototype even when not asked about it.

**YouTube's search** was discussed a lot in the interviews because it was in the presentation slides but also because 4 out of 5 said that they have used or use it with their smart televisions. Most comments concerning YouTube's search's usability were positive. Participants seemed to like that there is a possibility to move up and down to reduce the distance between characters. People mentioned YouTube's keyboard being clear and compact. Users that had used this service seemed to mostly like this and one participant mentioned YouTube having a suitable User Interface for a 6-year-old.

Suggestions were mentioned being one of the YouTubes best feature related to search screen and one user said that YouTube's algorithm has learned well how he uses their service while another user describes that YouTube even gives the suggestions before she is ready as she expects that she would need to write more before getting the result. On the other hand, in another test, it was mentioned that it is expected that search happens quite fast after starting to write the query. People seemed to appreciate the suggestions as even though YouTube's keyboard was liked, people still understood that inputting with a remote is slow and suggestions were sometimes used.

The conversations about YouTube were comprehensive and many comments were not related strictly to the keyboard itself and one user for example

mentioned that she always uses the search because the recommendations that YouTube has are random in her point of view. When showing the screenshot of YouTube's search screen there was discussion whether an on-screen keyboard should be QWERTY or Alphabetical as one user mentioned that YouTube's search is in other respects good, but he would've preferred QWERTY keyboard. To oppose this another person said that they find the right letters easier from the alphabetical as it is easier to remember how alphabetical is arranged.

YouTube's search caused a bit of confusion in one of the participants as he mentioned that it was weird when the whole keyboard is in uppercase letters but when writing, outcome is written in lowercase letters. Users mentioned that they do not use the old search queries even when the service shows them, and users tell to always write a new search. One user commented how he always tries to go through the edges of a keyboard and that he was a bit disappointed that YouTube didn't have a possibility to go through edges to get to the other side. Still, he was content with how YouTube handled the edges as user could get to the other functionalities from going over the edges of a keyboard. Even though YouTube had a lot of positive comments, the user with most modern television said that still he often uses his phone to write the search query.

**Yle Areena's** search was described as almost the same as YouTube's by one participant as she commented when shown a screenshot of Yle Areena. But based on the results people had a lot of different thoughts and emotions risen from Areena compared to YouTube. Some users did recognize Areena's search screen, and these users mentioned have used it in the past. The language was also something that was discussed in some interviews as few people really liked that the service was in Finnish, and one was so curious about the language options that he inspected if in Areena there would be possibility to change the language to Swedish and English. He was a bit disappointed that there was English missing from the language options.

The user who had a T9 as a main interaction said that she would prefer using regular Areena, instead of writing with a remote's keyboard. The discussion regarding the Areena's keyboard itself was about if number of rows was suitable and surprisingly one participant said they would feel like Areena is clearer with fewer rows (compared to Netflix) as the movement doesn't need as much cognitive work. Opposing opinions were in 3/5 interviews where it was mentioned that they would prefer shorter rows so that there would be more rows. In the discussion this was because few of the participants felt like the distance between the letters was too long in the present keyboard. One specifically mentioned that going through edges would help this situation

and was again baffled why streaming services do not include it to their applications.

Search in Areena was commented also with discussion if it was necessary to have the letters in such big font size. Same participant also commented if the uppercase letters were the best option by saying “...it confuses that all are in capital letters. I am so used that it gives me a bit of a silly/simple feeling”<sup>3</sup> Discussion about the letters will be discussed more thoroughly in the Usability chapter.

One user liked that the current results in Areena that were written as it could fit more results but said that he would still prefer the pictures as it feels more tangible. Another participant was eager to study what can be done with the Areena’s search and tried to search things by misspelling the query and searching with actor’s name. He got disappointed as Areena only worked when writing right or if query was missing the last letters. Also, by doing this, he found a bug as Yle Areena continued to give result: “No search results: Dwayne” even when he had used search after this.

The biggest problem in current Areena seemed to be the need for user to push a “search” -button before it started giving the results. The negative comments about additional “search”-button were mentioned in 4/5 tests and the one that did not comment on the usability of that function mentioned that she would expect the search to happen when pushing “OK” -button. This difference with other participants might be because of she had mostly experience with T9 keyboard. Another user did sum this problem with using the search -button well by explaining that if she needs to use a button, she writes the whole thing before being sure it searches.

**Netflix** was another one from the present streaming services that was discussed in all the tests quite a lot as all the participants had Netflix subscription. About the Netflix people liked how fast and efficient their search was, it sometimes even gives the right result after writing two letters. 3/5 users mentioned that they sometimes search content by actor’s name and all of them said that they liked this feature. 4/5 people talked about how they like that the search starts immediately when writing the search query but two people mentioned in the interviews that sometimes suggestions in the search screen are not found in Netflix and this frustrates them. One of users questioned this feature “Why does it give the suggestion, why It doesn’t just show that it can’t be found...?”<sup>4</sup>

---

<sup>3</sup> Original. ”...se vaa hämmentää, ku kaikki on isolla. Ku tottunu, et tulee vää semmonen yksinkertanen fiilis”

<sup>4</sup> Original: ”Miks se sit ees ehdottaa et miks ei vois tulla suora tekstiä, että ei löydy valikoimista tai jotain?”

Frustration can be noticed towards Netflix's search screen as two of the users had problems how slowly Netflix was working on their devices while writing. Yet there was another user that commented how fast the Netflix was with the search. The reason behind this kind of big differences might be explained with the hardware as all users were using their own devices and there was differences between them.

One interesting characteristic that was only discussed about Netflix compared to others was that people discussed that they use it when they do not know what to watch and often use it with the intention of browsing for content instead of using search. In 3 interviews it was mentioned that seldom they know what to watch when opening a Netflix. When people don't know what to watch they often were using categories.

One user explained that if after writing 3 characters, he can't see the right result in the results nor the suggestions then he is confident that the content he looks for can't be found on Netflix. There was not similar discussion when talking about other keyboards and same user mentioned that he is more patient while using YouTube for example.

Keyboard of the Netflix was found relatively easy to use even though one of the users mentioned that it might confuse users when numbers were on the same rows after letters. In one another interview also there was discussion about the location of the letters and one participant wondered if the numbers could be located separately from letters. It was interesting to hear this kind of discussion about Netflix if the preference was to have letters and numbers separated, the layout of a prototype can be argued with this kind of commentary. One small problem with Netflix's keyboard was that it was missing Scandinavian letters – in one interview there was discussion about how people can search Scandinavian content which uses Scandinavian letters.

**Figma prototype** was introduced to all users in the end of the user testing. A lot of interesting ideas rose while testing and discussion about the prototype. My theory is that this might be because the prototype was new to all participants, and they were looking at it with more critical eyes. Two participants questioned how simple the prototype looked and this might be because they did not understand that prototype is not meant to be the final design even though it was told before showing the prototype. They were expecting it to work and look more like a final product, but it might have been better because then these participants were giving suggestions and react to it as a final product. Rest of the users were rather curious about the design proposition and all participants gave interesting insights about how the prototype could be developed further. About the prototype people were discussing more

about the detail rather than search screen entirely, which was interesting and helpful. The discussion about prototype gave understanding what kind of attributes are affecting the experience with search in smart televisions.

Users did recognize that the design was following Areena's current graphical appearance, and my assumption is that it might have helped them to react to the prototype as a real product. A lot of insights were given on how it could be developed but also mentioned things that they preferred in prototype compared to other streaming services. The discussion about the number of rows appeared in multiple interviews and people were in consensus about that there is still need for making alphabet squarer shaped – people either compared the keyboard to Netflix's 6x6 shaped keyboard or to YouTube's 4x7 - keyboard when discussing about getting more rows and getting rid of some of the columns. This was interesting as the keyboard in the prototype was close to YouTube's square (being 4x8). The difference in number of letters can be explained with Scandinavian letters.

Users said that they would prefer even more square shaped keyboard, but all the participants seemed happy with the feature that numbers were separated from the alphabet and one even said that it would make it easier to communicate with the television as the numbers were not taking space from the screen. The keyboard was found clearer than Areena's current design, and few users mentioned that often when they need numbers, they are searching for a year and do not have a need to use numbers and letters at the same time.

Because the prototype was not fully functioning it raised a lot of speculation in the users about how the suggestions work. One of the users wondered that if the suggestions work so that they try to fill what the user is writing, he will see himself in using this keyboard gladly. In another test it was discussed that the location of the suggestions should be relatively close to the search bar and keyboards, and he compared this to Apple's way of having the suggestion between search bar and keyboard. It was also mentioned that this way the suggestions wouldn't take as much space on the screen. In one test it also came up that the number of suggestions should be more, for example when searching with actor's name if user does not remember the surname. He also was a bit disappointed as one of the suggestions was in Swedish and he mentioned that he wouldn't be happy if in the real-life situation, the artificial intelligence did not know that he wouldn't be watching *Moomin valley* in Swedish.

The discussion about icons that were used in prototype was rich. All users did understand what would happen from the icon buttons, but one user said he would prefer the Areena's current way of written function buttons. He mentioned that as written they were clearer to him. There was discussion on the layout of the screen and most of the users seemed happy about how functions

were located but in 2 of the tests there was discussion about moving the functions a bit. One suggestion was about location of the space bar should be after the keyboard (just like in physical one) when another user said that back-space should be closer to the rest of the buttons.

In two different tests it was mentioned that users would prefer lowercase text in the keyboard rather than uppercase, which is currently used. This was surprising as in the rest of the tests no one mentioned anything about what kind of casing should be in the letters of keyboard. Another discussion was held in regarding the font of the keyboard as users seemed to prefer if the keyboard size itself was smaller compared to screen because they felt like it was taking too much space on the screen.

One interesting remark about the results in the prototype was that people would prefer search results as a picture but even though many did find the DVD -shaped search result pictures nostalgic, users would prefer the ones with same aspect ratio as televisions screen as those fit better. One of the users said in the end of the interview that they would prefer the prototype over all keyboards he had used and seen. This might be because he liked the possibility to go through edges of keyboard to the other side but even for him this keyboard was not perfect. In overall people seemed to be more critical about the prototype, but also many were curious about how Yle will develop their services based on these tests.

There were many different services discussed during the interviews but at least as many left out of the study. All services had their pros and cons, but one comment sums it well as comment regarding the keyboards was *“I don’t even know which one would be actually good”*<sup>5</sup> Even though after studying this topic there is no easy answers to one keyboard being better than others.

## 5.2 User Experience

As discussed in a literature review, user experience is a complex context and there are many details in each service that affect the overall experience with a service. A lot of different things were found out about the usability that was tested with users, some of the things affecting usability were new information while some of the things that arose were expected based on the literature found about the subject.

One of the most obvious reasons why people had negative experience when using search on smart televisions was because of the remote controller. Many users did mention how clumsy and hard it is as an input device. In one test it was described that remote controller was good enough in the old, regular

---

<sup>5</sup> Original. “En ees tiiä mikä ois oikeesti hyvä”



television but for streaming services it is not found useful. Even though one user said that his remote feels better than the common ones because it had a touch swipe, he still often avoided writing with it by using his phone instead of the remote controller. There were also other problems regarding the remote controller as one user described remote controller being the “one stick that is always missing” and that she can’t see what happens in which button now that she had problems with sight.

One of the users suggested that could it be possible for application developers to use the colored buttons to for example have a faster change between keyboards (when having letters and numbers as separate ones). It was also mentioned that in his own television there was a possibility to use these buttons when writing. The one user that was regularly using their phone to for search, mentioned that even before his Apple-TV he was using his phone with Samsung’s television application. He suggested that television developers are aware that it is hard to write with remotes and that is why they have given users an option to use more familiar device.

Because this research is done in the application developers’ point of view, the usability of remote controllers isn’t that much of a concern as it can’t be changed. Therefore, it is important to focus on the aspects that can be developed to create more positive experiences interacting with the service.

Search is an interaction that all users say that they are using while fetching content in smart television. Search happens with an on-screen keyboard and many services do these keyboards differently as there are no standards regarding it. The common idea from all interviews was that square shaped keyboards were preferred as they had the least distance between characters. Another point regarding the keyboard was that as that people don’t use numbers often while searching and because of that they don’t feel the need for having numbers in the same screen as letters they felt like numbers would make more sense when on a separate screen.

Positive attitude towards the numbers on a separate screen was common and one participant said that it confuses users if numbers and letters were on the same grid. One of the participants mentioned that he searches a lot of things that require both letters and numbers from YouTube, but he mentioned that in other services he would see separated numbers and letters as positive feature. When on YouTube he would prefer them in the same screen. Also, in one of the interviews a participant mentioned that special characters are only needed in smart television if there is a need for logging in to the service and therefore, they could be separated from the letters too.



What was surprising to learn from the interviews was the need to go through edges of the screen to the other side of the screen. This function seemed to be even more important if there were no other functions on the sides, but all participants discussed that they would use or at least try going through edges if needed. It was interesting to hear about this as it seems from my experience that no streaming service has this possibility currently, but one user mentioned that he had learned this from old PlayStation he used to own.

By the geographical point of view, the Scandinavian letters were discussed in 4 tests and some of the users seemed passionate about this as they immediately noticed if the example pictures were missing Scandinavian letters. The capitalization of letters on the keyboard was discussed in two different interviews where two participants narrated that they did prefer lowercase letters to uppercase ones. One of them said it was because it made him feel simple while another one explained that it would be easier to distinguish the letters from each other when lowercased.

From the interviews it still was not completely effortless to distinguish which keyboard is the best as 3 participants were suggesting the alphabetical order while rest, 2 participants were on QWERTY's side. In the end of one interview, the other QWERTY supporter called out for box shaped alphabetical being the best keyboard for on-screen usage. Both QWERTY users also discussed about how the shorter distance between letters was better – which opposes the idea of QWERTY which is not square shaped.

Search screen has often other kinds of functions besides the keyboard itself. In the interviews the other functions were discussed through. All participants did understand buttons used on services, the ones that had icons and those that had the function written in them. One participant was passionate about the importance of the delete and backspace functions as she explained that she often does mistakes while writing and has a need for these to fix the writing. Few users said they really preferred the functions with icons while one participant opposed this with preferring the ones with text. Locations of the function -buttons seemed to rise differing thoughts as some people preferred the ones next to search bar while others would like functions to be on the side. Location of the spacebar on the other hand was discussed to be after the keyboard based on the feedback from the users.

All users seemed to expect search to start immediately when they are writing, and four users mentioned they do not want to click “search” button exclusively. The comment from one participant sums well the overall attitude towards this *“It would be horrible if it was needed to push some “search” –*

*button separately*”<sup>6</sup> While another user said that if it was necessary to use “search” -button, she would probably write the whole name of a movie or video she was looking for.

Users expect search results to be shown as pictures even when the number of results is more limited than. One participant mentioned that he would add the names of results next or under the pictures to make them clearer when sometimes the pictures from shows don’t always have their name on. Some users did mention DVD -shaped pictures being nostalgic as they remember DVD’s and cassettes, but two participants mentioned that they would prefer pictures with same aspect ratio as televisions themselves.

There was also a lot of discussion about how search results give users bad experience when the technology does not work as they expect it to work. Sometimes results are bad, and one user mentioned that sometimes she thinks that she finds specific content by browsing faster than by using the search. There was also discussion about how search results show other results than the ones that answer the search query. It was discussed that if you are aware of this problem, it might not annoy as much but usually this seems to annoy users a lot. Search results were also liked. One user mentioned how it is scary even how well the results sometimes are even after writing one letter. Also, it was praised how one service had learned to know what kind of content the user wants to find.

When using searching, some users mentioned that they might use the already used searches again if they felt the need to find specific thing multiple times, but other two mentioned that they would never use the same search word again with streaming services. Four participants said they use and would like to use search to find movies or series with actor’s name. There were few reasons people liked to do this as some said they do this when being a fan of one actor, while others said they might have seen a trailer and forgot name of the movie but remember it can be found on a streaming service and saw a certain actor on the trailer.

One user mentioned that they either use old searches or suggestions while searching. Suggestions are the text that is not a search result, but they are found on the search screen, and they often work so that they try to complete what user is trying to write. Suggestions were both liked and disliked. Users have high expectations about artificial intelligence which is expected to understand what kind of content they users are trying to find. Users also liked that there are suggestions as they seem to help to avoid writing title completely and help filling the search. One user said that he uses suggestions as

---

<sup>6</sup> Original. “Se ois ihan kauhee jos tarvis erikseen painaa jotain ”Hae” -nappia”

backup if results are not showing the searched result and then he mentions turning to suggestions to find wanted title.

Regarding suggestions, one user mentioned that he has different expectations based on which streaming service he is using. He said that he is more patient with YouTube because of the amount of content while he gets frustrated more easily towards Netflix, which has more controlled library size. Same user also explained that sometimes the suggestions are way too aggressive as he mentioned watching once a video about a topic he don't usually watch and then all suggestions on that streaming service change to being about that once watched topic. Worst issue with suggestions seemed to be that some streaming services tend to suggest things that can't be found on said services. This and the annoyance towards this were discussed in a total of 3 different interviews as it rose a lot of negative emotions.

Suggestions are given after user starts to write their search query and there was a lot of discussion about how much users do expect they need to write before a service starts giving right results or suggestions. There was a difference in how patient users were but in overall it was expected to give right results after writing one word the latest. Some participants were expecting to have right search results already after 2-4 letters, but no one said that they would write the title completely. One user mentioned that streaming services start on giving the results already before they feel being ready for right results. On the other hand, one participant mentioned that if after 5 letters he can't see what he wants in the results, he is certain that it can't be found from that service.

Users do expect the services to either correct their misspelling or to search so efficiently that they find right results even with misspelled search queries. One user suggested that there could be more search options, such as filters for whole series but this does not seem too major issue as nothing similar was mentioned in other interviews and few participants said that they often browse through different categories when they are not sure what to watch.

In overall all participants knew how search works, but they all had negative experiences with searching on smart television and get easily annoyed when search does not work as intended. One user sums the results well with her comment about search in smart televisions *"It is not too good in any service... they all work same way..."*<sup>7</sup>

---

<sup>7</sup> Original. "Ei oo oikein hyvä missään palvelussa se... Jotenki toimii kaikissa samalla tavalla..."

## 6 Conclusions

The aim for this study was to gain knowledge about how search could be improved in smart television applications. This objective was divided into two sectors, one of them focusing on study of keyboards in smart televisions (RQ1) while another one was aiming to learn what kind of things affect the usability of a smart televisions search (RQ2). The research was done with qualitative method of interviewing users who were familiar with using smart televisions.

To answer RQ1, all different keyboards were identified from the gained data and studied which one of the presented keyboards had the most satisfied users. All services seemed to have their pros and cons, but the best keyboard type was quite clear based on the answers, the most liked keyboard for on-screen is keyboard that is in an alphabetical order. Even though based on the literature hierarchical keyboard was stated to be one of the fastest ways of interacting, hierarchical keyboard did not attract the users at all in this test. Interesting detail is that two participants preferred QWERTY over alphabetical were both working on IT-field and my assumption as a researcher is that this can affect their attitudes as they are so used to working with QWERTY keyboard and therefore prefer it over other options

It seemed to be important for the users that a keyboard has the least distance between characters, meaning its layout should be square shaped. This finding can be verified with Nielsen's (1993) Heuristic number 7 which guides designers to aim for efficiency of use when designing a product. When looking at the commonly used keyboards, it seems like many service providers already knew that square shaped alphabetical keyboard was the most desired keyboard as many streaming services available in Finland are currently using square shaped on-screen keyboard. This notice was based on what came up from the interviews.

Usability of a keyboard did not only focus on the layout, but it seemed that users were disappointed if the Scandinavian letters were missing from the keyboards. The absence of them was surprisingly common on the international streaming services. There was also one big thing learned from the keyboards as on a traditional streaming service, users do not have a need for numbers being on the screen at the same time as letters. Users feel like numbers are taking unnecessary space and it would help writing efficiently if numbers were on a separate writing screen from the letters.

On-screen keyboards have different functions besides the letters and numbers, and these buttons should be easily accessible and according to user feedback should be located near the keyboard. Especially the spacebar was

something that people seem to expect to be found right under the letters. Other functions were expected to be close so that they are easy to use.

Coherent answer to RQ2 is far more complex than to RQ1 as it focused more on the experience and emotion of a user. To understand emotions of a user the analysis of data needed to be done subtly and needed many repetitive times of reading through the raw data. Some features were asked directly while others did naturally rise from the interviews while discussing in general. Researcher intended the questions so that they wouldn't be leading but rather general to gain understanding about different usability problems.

What was not surprising was the usability problems with using the remote controller, but the number of negative emotions towards remote controller was still surprisingly big. One thing that was found out about the usability of search screen needs to be addressed as an accidental research discovery. Prototype was designed in a way that there was a possibility to move through the edges of the screen. This functionality that all users were happy to use they would use if a service did implement this.

This feature was not studied beforehand and must be addressed as a lucky mistake. Researcher thought feature like this was a default in the smart televisions. But this kind of wrong assumption could have avoided if there was a possibility for researcher to try out smart television applications herself beforehand. But if she had learned there was no possibility to go through edges it could have affected the prototype proposition. Going through the edges was feature that only one user mentioned having experience with it while using with television screen, but all users seemed intrigued by the possibility of it.

As expected, users did not like the idea of pushing search -button exclusively and they expect search to happen instantly when they have started writing. Search should happen fast, and it is expected that between 2 letters and one word the search should already know what user wants to find. If search requires more writing, people assume that the searched program cannot be found on the used service.

While searching for programs, people often do not know what specific program they want to watch and therefore they sometimes might search with a person's name, either an actor's or director's name. As writing with on-screen keyboard is clumsy, people expect suggestions to work as a helpful hand and aid with spelling errors as well as to avoid writing for too long. There was a lot of negative experiences regarding to suggestions as in some streaming service's suggestions give users suggestions that are not found on said service.

Results from search are sometimes disliked as when the service is missing a searched program, services tried to give other options to watch instead of telling user that the searched program cannot be found. This raised negative feelings towards the search, but when the search was working as meant, users were content with it. Because search results were appearing fast, users were happy about how well the artificial intelligence knew them, but also in some of the tests the doubtfulness was present when discussing about artificial intelligence if even knew too much about them.

One detail about the results was that there is a need to show the results visually. When the results were shown visually, users felt they were more attractive. Also, the result pictures when shown as horizontal were preferred compared to ones in vertical.

## 7 Discussion

The main goal of this project was to study usability of search in smart tv applications. The need for this kind of study is existing as not similar studies were found. It is possible that companies have done research, but they decide to keep them secret because of huge competition in the market.

As this study does aim to answer the research questions and on the researchers point of view it does success this goal.

One of the important things that was pointed out from the data gained in this research was how the search should work immediately while users start writing the search query. This is something that also was suggested in the Lamkhede and Das (2019) as it confirms the importance of this attribute.

Barrero et al. (2014) does suggest T9 as a good option to interact with television search but as this study shows, people are not too excited about using it nowadays. It might be because it's usability problems such as slowness and clumsiness or the problem might be that users do not see it as a modern way of interacting with a television. One problem that was discussed in one interview was the response time being slow. This is something that also was discussed in the Miesler et al.'s (2014) research as it is one of the things that greatly affect the usability of a smart television and one of the big problems in a remote controller.

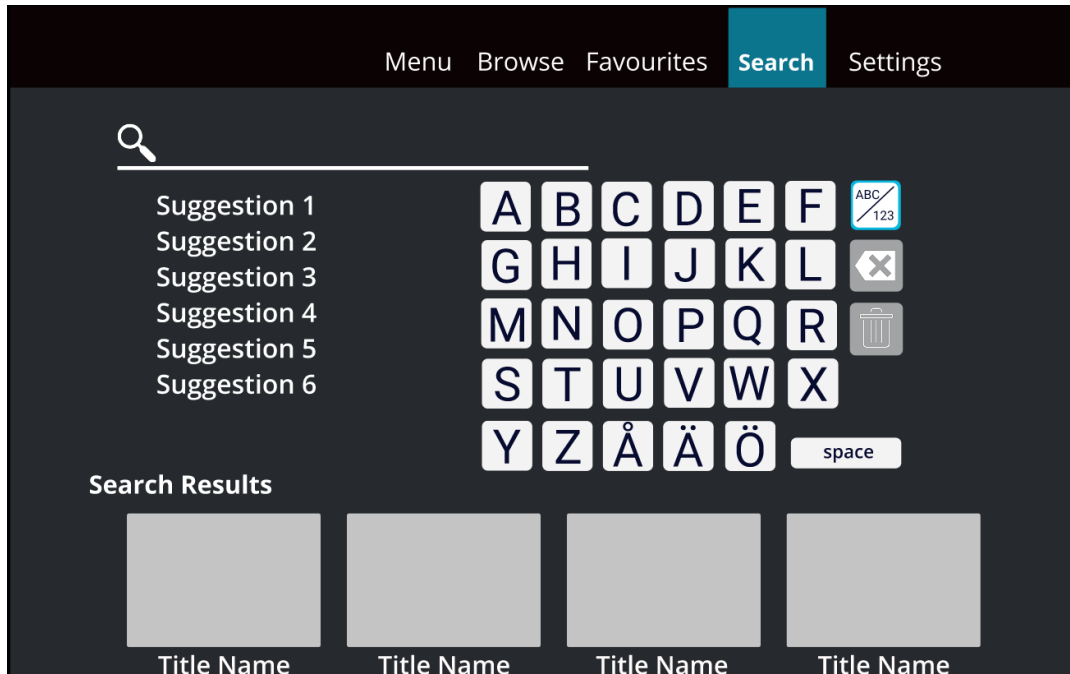
The study by Sinharay et al. (2013) did propose that based on the brain signals the Hierarchical keyboard would have worked best with users but it might be either that their study did not consider what kind of user experience it does give, or the screenshot shown to the users in this study might have been too low resolution and that way affected the experience towards that screen.

Because of all participants were familiar with using technical devices it was expected that familiarity with technology would reduce the frustration with remote controller as suggested in the paper by Jang & Yi (2019) where it was described that after learning to use remote controller, the problems were greatly reduced. All users mention and have negative experiences with a remote controller even while they were already familiar with the device. The problems with remotes were already reckoned in the literature view but the severity of this problem was rather surprising.

Now would be a good time to look back to the Research Objective of this study, which was presented in the second chapter of this paper. Research Objective was "*How to improve Yle Areena's search screen in smart television?*"



And while research questions already were answered, the way of this study aims to answer this objective is by proposing a new prototype that could be used for further development.



**Figure 8.** Proposition for a new prototype that could be used for further usability testing.

A proposition for a new prototype was created based on what were the most crucial parts that affect the usability of a search screen in smart television applications. This prototype was done by adjusting the prototype already tested in this study and changing it based on the findings from interviews. In the future there could be further study with a prototype like this or Yleisradio could even fit it to newer design of Areena.

## 7.1 Limitations of Research

This study provides new information regarding the on-screen keyboards and search of smart television applications. Findings of this study confirm that key factors affecting usability of search in smart televisions. However, like in all research there are limitations that apply that should be addressed when discussing the research. These limitations that apply to this research should resonate to the future research.

The concept of reliability is described in a paper by Golafshani (2003) as a concept that is used for evaluating research. To ensure reliability of a qualitative research, it is necessary to take trustworthiness into consideration to

ensure the scientific relevance of a study. Scott and Morrison (2006) define reliability as how repeatable a study is. There are different classes of reliability estimates that are defined in Scott and Morrison's paper. One way of verifying reliability of a research is by re-testing the same research arrangement and comparing the results. If results remain similar, it can be assumed that research remains reliable. There is also discussion on a paper about internal consistency reliability, which can be confirmed without further research as it is possible to prove that there is consistency between items in a test. (Scott, Morrison. 2006).

What could have been done to enhance the validity of the research is to use triangulation. According to Noble and Heale (2019), triangulation is a method, which is used to increase both the credibility and validity of research. Carter et al (2014) explains that there are multiple ways to do triangulation. It can be done with few different ways such as triangulation of data sources, investigator triangulation or method triangulation. In research that has so limited resources, triangulation often must be left out as it makes the research more time-consuming. In this study there was to methods to gaining data such as literature review and usability testing.

Literature found about this topic is little in numbers and some of the papers can be counted as outdated. The studies formerly published about the topic were focusing more on the efficiency side as other literature did study the usability with electromagnetic signals in users' brain which makes the approach a lot different. My guess for the reason there is little research published about it is because companies might keep the studies secret if they did test on their own products to avoid helping competitors. The lack of literature creates a need to study this topic more and hopefully in future there will be more studies available about usability in smart televisions.

One important thing to address is that this study relies on usability testing of five participants that were rather homogenous. That is why the generalization of a study is uncertain and, in the future, should be considered when studying the results. In the future, sampling of participants should be done more efficiently.

The chosen research strategy puts one important limitation to this research. As qualitative research focuses on people and their experiences, it is important to determine if participants are being genuine and telling the truth and there is always a possibility that interviewees were unintentionally or intentionally trying to please interviewer. Especially in this kind of situation where the interviewer and interviewee know each other in advance it is possible that interviewees try to please the interviewer.

Nevertheless, all interviewees seemed to talk openly about their experiences and feelings that rose related to smart televisions search. Both negative and positive feelings were expressed in the interviews and there were some critical, insightful comments which might indicate that these answers were representing their experiences and opinions in real life.

Researchers try to stay unbiased while studying a topic, but it is crucial to remember that everyone has their own history and expectations towards their own research, and it affects different sections of the study. Biases in researchers also have an influence on how semi-structured interviews are constructed as it leads how questions are formed on the position which will affect the results gained from the interviews. It also can affect how the results are interpreted. I aimed to stay critical to my own thoughts and choices in this research but the most efficient way of avoiding this kind of bias would be with investigator triangulation where multiple researchers could question each other's opinions and attitudes on the study.

There was a lot of variation in the devices used in this study which might affect the data gathered, but as research aims to represent wide range of users, it is possible there is variation in them too. One of the participants had Apple-TV and one was using Elisa Viihde -box and they both had their own User Interfaces even inside the application. For example, on Apple-TV Yle Areena looked like an original one but it's search page had Apple's own keyboard. Elisa Viihde on the other hand did have the whole User Interface based on their graphical look and the search happened with a remote controller that did have T9 keyboard instead of clicking on the screen.

It can be argued if usability evaluation is robust enough to be counted as science. Scientifically of usability research was challenged by Greenberg and Buxton (2008) where they argue against it. One of their arguments against usability evaluation as a scientific method is the lack of replication and hypothesis testing that isn't risky enough. They suggest that designers often try to create the testing situation favourable to the newest technique. This argument of designers being too shy with their hypothesis testing might sometimes be true, but I tried to avoid being too favourably for own prototype as there was sole curiosity regarding the topic.

As the logos from slide show was cropped out, it led three different participants to guess if a certain keyboard was from a specific service, they were familiar with. One of the participants mentioned that he liked the keyboard from Apple television but decided to comment also that he was a big fan of Apple in a way that he was admitting that there might have been a bias in his comments. This was interesting insight from the participant himself and

even though other participants did not discuss anything similar it is possible that they had similar feelings towards the most familiar application.

It is important to understand that this kind of bias happens and one idea that could overcome the bias by creating all the slides so that they would not present any current streaming services. With this way the discussion would focus on the keyboard and search screen and its elements rather than which service it is from. This would take more time than finding pictures of the different screens but would make people focus on the screen itself rather than the service or the resolution of a picture.

Even though Think Aloud method was applied to this research, when looking afterwards it did not work as well as desired. Think Aloud method requires active participation for the interviewee to push the users to remember to think aloud. Especially it seemed like users forgot to talk while they were interacting with their own devices but the discussion about the prototype gave more rich data. This might be because users needed to think more while the system was not familiar for them, and they were discussing more critically about the prototype.

External validity of a research is also affected by that during the data analysis phase there was only one researcher who analysing the interviews. The former experiences and perspective could have affected how researcher did the categorizing and analysis of a found data. What to be noted is that researchers experience with smart televisions is very limited as she does not own a smart television herself. Therefore, the planning and analysis part has not been affected by personal experience and rather by curiosity to learn about the topic.

## **7.2 Future Work**

According to the results discussed formerly there are many things that affect the usability of an on-screen keyboard and the experiences with search but there was still many details and possibilities left undiscovered and the interest to learn more deeply about the topic might lead into further research.

Based on the findings it one could create another prototype and to do more usability testing. One possibility could be by focusing on details that were not as clear from this research and focus on comparing and challenging the findings so far. This could be done for example with A/B testing. One of the things that should be studied would be learning more about how users feel if the keyboard was in lowercase letters, like 2 users suggested. It might be that in this study other participants did not comment on the casing of letters even

though they preferred uppercase or maybe lowercase would be better, but this this is something that can't be claimed without further studying.

The future testing should also focus on people that are not as familiar with using technologies as this participant group was narrow and focused on people that were already more familiar with using technologies. As Yle's services are aimed at all Finns it is important that the service is tested on people with different backgrounds. By studying the topic with wider range of methodology and with multiple researchers, it could be possible to verify the validity of a research and to make new observations as well.

## References

- 99 Content. (2021). *Search Engine Statistics*. 99 Firms. Retrieved May 5, 2021 from, <https://99firms.com/blog/search-engine-statistics/#gref>
- Aziz, H. A. (2017). Comparison between Field Research and Controlled Laboratory Research. *Archives of Clinical and Biomedical Research*, *01*(02), 101–104. <https://doi.org/10.26502/acbr.50170011>
- Alam, I., Khusro, S., & Khan, M. (2019). Usability Barriers in Smart TV User Interfaces: A Review and Recommendations. *2019 International Conference on Frontiers of Information Technology (FIT)*. Published. <https://doi.org/10.1109/fit47737.2019.00069>
- Augustine, S., & Greene, C. (2002). Discovering How Students Search a Library Web Site: A Usability Case Study. *College & Research Libraries*, *63*(4), 354–365. <https://doi.org/10.5860/crl.63.4.354>
- Awale, B., & Murano, P. (2020). A Preliminary Usability and Universal Design Evaluation of a Television App User Interface. *Baltic Journal of Modern Computing*, *8*(3). <https://doi.org/10.22364/bjmc.2020.8.3.03>
- Audience Project. (2020). *Insights 2020*. Retrieved May 5, 2021 from, [https://www.audienceproject.com/wp-content/uploads/audienceproject\\_study\\_tv\\_streaming\\_2020.pdf](https://www.audienceproject.com/wp-content/uploads/audienceproject_study_tv_streaming_2020.pdf)
- Barrero, A., Melendi, D., Pañeda, X. G., García, R., & Cabrero, S. (2014). An Empirical Investigation Into Text Input Methods for Interactive Digital Television Applications. *International Journal of Human-Computer Interaction*, *30*(4), 321–341. <https://doi.org/10.1080/10447318.2013.858461>
- Benyon, D. (2019). *Designing User Experience: A guide to HCI, UX and interaction design* (4th ed.). Pearson.
- Bevan, N. (2009). What is the difference between the purpose of usability and user experience evaluation methods?. Proceedings of the Workshop UXEM.
- Brasel, S. A., & Gips, J. (2011). Media Multitasking Behavior: Concurrent Television and Computer Usage. *Cyberpsychology, Behavior, and Social Networking*, *14*(9), 527–534. <https://doi.org/10.1089/cyber.2010.0350>
- Camburn, B., Viswanathan, V., Linsey, J., Anderson, D., Jensen, D., Crawford, R., Otto, K., & Wood, K. (2017). *Design prototyping methods: state of*

*the art in strategies, techniques, and guidelines*. Design Science, 3.  
<https://doi.org/10.1017/dsj.2017.10>

Carter, N., Bryant-Lukosius, D., DiCenso, A., Blythe, J., & Neville, A. J. (2014). *The Use of Triangulation in Qualitative Research*. Oncology Nursing Forum, 41(5), 545–547. <https://doi.org/10.1188/14.onf.545-547>

Dou, J., Qin, J., Wang, Q., & Zhao, Q. (2018). Identification of usability problems and requirements of elderly Chinese users for smart TV interactions. *Behaviour & Information Technology*, 38(7), 664–677.  
<https://doi.org/10.1080/0144929x.2018.1551423>

Dudek, D., Mastora, A., & Landoni, M. (2007). *Is Google the answer? A study into usability of search engines*. Library Review, 56(3), 224–233.  
<https://doi.org/10.1108/00242530710736000>

Eccles, D. W., & Arsal, G. (2017). The think aloud method: what is it and how do I use it? *Qualitative Research in Sport, Exercise and Health*, 9(4), 514–531. <https://doi.org/10.1080/2159676x.2017.1331501>

Elliott, V. (2018). Thinking about the Coding Process in Qualitative Data Analysis. *The Qualitative Report*. Published.  
<https://doi.org/10.46743/2160-3715/2018.3560>

Epstein, R., Robertson, R. E., Lazer, D., & Wilson, C. (2017). *Suppressing the Search Engine Manipulation Effect (SEME)*. Proceedings of the ACM on Human-Computer Interaction, 1(CSCW), 1–22.  
<https://doi.org/10.1145/3134677>

Ericsson ConsumerLab. (2017). *ConsumerLab report on TV and Media 2017*. Telefonaktiebolaget LM Ericsson. Retrieved June 28, 2021, from <https://www.ericsson.com/en/reports-and-papers/consumerlab/reports/tv-and-media-2017>

Evans, C., & Lewis, J. (2018). *Analysing Semi-Structured Interviews Using Thematic Analysis: Exploring Voluntary Civic Participation Among Adults*. SAGE Research Methods Datasets. Published.  
<https://doi.org/10.4135/9781526439284>

Ficom. (2020). *Suomalaiset ostavat nyt älytelevisioita - myös nettiyhteyksiä päivitetään*. Retrieved May 10, 2021 from, <https://www.ficom.fi/ajankohtaista/jaseniston-tiedotteet/suomalaiset-ostavat-nyt-alytelevisioita-myos-nettiyhteyksiä-paivitetaan/>



Frøkjær, E., Hertzum, M., & Hornbæk, K. (2000). Measuring usability. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '00*. Published. <https://doi.org/10.1145/332040.332455>

García, E., Torres, J. C., & de Castro, C. (2017). Usable Control System for Interaction with Ubiquitous Television. *Communications in Computer and Information Science*, 149–160. [https://doi.org/10.1007/978-3-319-63321-3\\_11](https://doi.org/10.1007/978-3-319-63321-3_11)

Garrett, J. J. (2010). *The Elements of User Experience: User-Centered Design for the Web and Beyond (2nd Edition) (2nd ed.)*. New Riders.

Gatlan, S. (2015, October 5). *Apple's search screen* [Screenshot]. Softpedia. Retrieved July 25, 2021 from, <https://news.softpedia.com/news/new-apple-tv-api-will-bring-universal-search-to-all-content-providers-493638.shtml>

Gilbert, A. (2019). Push, Pull, Rerun: Television Reruns and Streaming Media. *Television & New Media*, 20(7), 686–701. <https://doi.org/10.1177/1527476419842418>

Golafshani, N. (2015). Understanding Reliability and Validity in Qualitative Research. *The Qualitative Report*. Published. <https://doi.org/10.46743/2160-3715/2003.1870>

Grand View Research. (2021). *Smart TV Market Size, Share & Trends Report Smart TV Market Size, Share & Trends Analysis Report By Resolution (4K UHD TV, HDTV, Full HD TV, 8K TV), By Screen Size, By Screen Type (Flat, Curved), By Region, And Segment Forecasts, 2021 - 2028*. Retrieved August, 20, 2021 from, <https://www.grandviewresearch.com/industry-analysis/smart-tv-industry>

Gray, L., Wong-Wylie, G., Rempel, G., & Cook, K. (2020). Expanding Qualitative Research Interviewing Strategies: Zoom Video Communications. *The Qualitative Report*. Published. <https://doi.org/10.46743/2160-3715/2020.4212>

Greenberg, S., & Buxton, B. (2008). Usability evaluation considered harmful (some of the time). *Proceeding of the Twenty-Sixth Annual CHI Conference on Human Factors in Computing Systems - CHI '08*. Published. <https://doi.org/10.1145/1357054.1357074>

Guérit, R., Cierro, A., Vanderdonckt, J., & Pérez-Medina, J. L. (2019). Gesture Elicitation and Usability Testing for an Armband Interacting with

Netflix and Spotify. *Advances in Intelligent Systems and Computing*, 625–637. [https://doi.org/10.1007/978-3-030-11890-7\\_60](https://doi.org/10.1007/978-3-030-11890-7_60)

Halavais, A. (2017). *Search Engine Society (Digital Media and Society) (2nd ed.)*. Polity.

Halvorson, K. (2010). Intentional communication. *Interactions*, 17(3), 75–77. <https://doi.org/10.1145/1744161.1744178>

Hosh, W. L. (2020). *Netflix*. Britannica. Retrieved June 15, 2021, from <https://www.britannica.com/topic/Netflix-Inc>

Hvannberg, E. T., Law, E. L. C., & Lárusdóttir, M. K. (2007). Heuristic evaluation: Comparing ways of finding and reporting usability problems. *Interacting with Computers*, 19(2), 225–240. <https://doi.org/10.1016/j.intcom.2006.10.001>

International Organization for Standardization. (2019). *Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems*. ISO 9241-210. Retrieved July 15, 2021, from <https://www.iso.org/home.html>

Ingrosso, A., Volpi, V., Opromolla, A., Sciarretta, E., & Medaglia, C. M. (2015). UX and Usability on Smart TV: A Case Study on a T-commerce Application. *Lecture Notes in Computer Science*, 312–323. [https://doi.org/10.1007/978-3-319-20895-4\\_29](https://doi.org/10.1007/978-3-319-20895-4_29)

Jang, J., & Yi, M. Y. (2019). Determining and validating smart TV UX factors: A multiple-study approach. *International Journal of Human-Computer Studies*, 130, 58–72. <https://doi.org/10.1016/j.ijhcs.2019.05.001>

Jang, J., Zhao, D., Hong, W., Park, Y., & Yi, M. Y. (2016). Uncovering the Underlying Factors of Smart TV UX over Time. *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video*. Published. <https://doi.org/10.1145/2932206.2932207>

Kim, D. Y. (2019). A Design Methodology Using Prototyping Based on the Digital-Physical Models in the Architectural Design Process. *Sustainability*, 11(16), 4416. <https://doi.org/10.3390/su11164416>

Kumar, M., Bindal, A., Gautam, R., & Bhatia, R. (2018). Keyword query based focused Web crawler. *Procedia Computer Science*, 125, 584–590. <https://doi.org/10.1016/j.procs.2017.12.075>

Kuniavsky, M. (2012). *Observing the User Experience*. Morgan Kaufmann Publishers. Published. <https://doi.org/10.1016/c2010-0-64844-9>

- Krug, S. (2006). *Älä Pakota Minua Ajattellemaan* (V-P, Ketola. Trans.) New Riders Pub.
- Lauff, C., Menold, J., & Wood, K. L. (2019). Prototyping Canvas: Design Tool for Planning Purposeful Prototypes. *Proceedings of the Design Society: International Conference on Engineering Design*, 1(1), 1563–1572. <https://doi.org/10.1017/dsi.2019.162>
- Lamkhede, S., & Das, S. (2019). Challenges in Search on Streaming Services. *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. Published. <https://doi.org/10.1145/3331184.3331440>
- Lawrence, E. (2015). Everything is a Recommendation: Netflix, Altgenres and the Construction of Taste. *Knowledge Organization*, 42(5), 358–364
- Lotz, A. D. (2018). Evolution or revolution? Television in transformation. *Critical Studies in Television: The International Journal of Television Studies*, 13(4), 491–494. <https://doi.org/10.1177/1749602018796757>
- McKay, I. (2017). *Crafting a high-performance TV user interface using React*. Medium. Retrieved July 15, 2021 from, <https://netflixtechblog.com/crafting-a-high-performance-tv-user-interface-using-react-3350e5a6ad3b>
- McNally, J., & Harrington Diederich, E. (2019). Browsing for Content Across Pay TV and On Demand Video Options. *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video*. Published. <https://doi.org/10.1145/3317697.3323353>
- Miesler, L., Gehring, B., Hannich, F., & Wüthrich, A. (2014). User Experience of Video-on-Demand Applications for smart TVs: A Case Study. *Lecture Notes in Computer Science*, 412–422. [https://doi.org/10.1007/978-3-319-07638-6\\_40](https://doi.org/10.1007/978-3-319-07638-6_40)
- MJ Tube. (2018, October 23). *Google keyboard* [Screenshot]. YouTube. Retrieved October 19, 2021 from, [https://www.youtube.com/watch?v=\\_Hr1lh058Xc](https://www.youtube.com/watch?v=_Hr1lh058Xc)
- Neate, T., Jones, M., & Evans, M. (2017). Cross-device media: a review of second screening and multi-device television. *Personal and Ubiquitous Computing*, 21(2), 391–405. <https://doi.org/10.1007/s00779-017-1016-2>

Nielsen, J. (1994, updated 2020). *10 Usability Heuristics for User Interface Design*. Nielsen Norman Group. Retrieved June 4, 2021, from <https://www.nngroup.com/articles/ten-usability-heuristics/>

Nielsen, J., & Landauer, T. K. (1993). A mathematical model of the finding of usability problems. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '93*. Published. <https://doi.org/10.1145/169059.169166>

Noble, H., & Heale, R. (2019). Triangulation in research, with examples. *Evidence Based Nursing*, 22(3), 67–68. <https://doi.org/10.1136/ebnurs-2019-103145>

Norman, K., & Kirakowski, J. (Eds.). (2018). *The wiley handbook of human computer interaction set*. John Wiley & Sons, Incorporated.

Lowe, N. K. (2019). What Is a Pilot Study? *Journal of Obstetric, Gynecologic & Neonatal Nursing*, 48(2), 117–118. <https://doi.org/10.1016/j.jogn.2019.01.005>

Ouyang, X., & Zhou, J. (2018). How to Help Older Adults Move the Focus on a Smart TV? Exploring the Effects of Arrow Hints and Element Size Consistency. *International Journal of Human–Computer Interaction*, 35(15), 1420–1436. <https://doi.org/10.1080/10447318.2018.1534346>

Perakakis, E., & Ghinea, G. (2017). Smart Enough for the Web? A Responsive Web Design Approach to Enhancing the User Web Browsing Experience on Smart TVs. *IEEE Transactions on Human-Machine Systems*, 47(6), 860–872. <https://doi.org/10.1109/thms.2017.2726821>

Perrinet, J., Paneda, X., Cabrero, S., Melendi, D., Garcia, R., & Garcia, V. (2011). Evaluation of Virtual Keyboards for Interactive Digital Television Applications. *International Journal of Human-Computer Interaction*, 99999(1), 1. <https://doi.org/10.1080/10447318.2011.555305>

Popovici, I., & Vatavu, R. (2018). Perceived Usability, Desirability, and Workload of Mid-Air Gesture Control for Smart TVs. *RoCHI*.

Portigal, S. (2021). *Interviewing Users: How to Uncover Compelling Insights* (2015–04-21). Rosenfeld Media.

Rao, J., Ture, F., He, H., Jojic, O., & Lin, J. (2017). Talking to Your TV. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. Published. <https://doi.org/10.1145/3132847.3132893>

Rice, M., & Alm, N. (2008). Designing new interfaces for digital interactive television usable by older adults. *Computers in Entertainment*, 6(1), 1–20. <https://doi.org/10.1145/1350843.1350849>

Roto, V. (2006). *User Experience Building Blocks*. Nokia Research Group.

Rubin, J., & Chisnell, D. (2008). *Handbook of Usability Testing* (2nd Second edition) [E-book]. John Wiley Sons Inc.

Scott, D., & Morrison, M. (2006). *Key Ideas in Educational Research (Continuum Research Methods Series)* (1st ed.). Continuum.

Shin, D. H., & Shim, Y. (2017). Mapping user experience of multiplatform services: the quality factors in multiplatform television. *International Journal of Mobile Communications*, 15(5), 554. <https://doi.org/10.1504/ijmc.2017.086368>

Shin, D., Zhong, B., & Biocca, F. A. (2020). Beyond user experience: What constitutes algorithmic experiences? *International Journal of Information Management*, 52, 102061. <https://doi.org/10.1016/j.ijinfomgt.2019.102061>

Sinharay, A., Chatterjee, D., & Sinha, A. (2013). Evaluation of Different On-screen Keyboard Layouts Using EEG Signals. *2013 IEEE International Conference on Systems, Man, and Cybernetics*. Published. <https://doi.org/10.1109/smc.2013.88>

Linneberg, M. S., & Korsgaard, S. (2019). Coding qualitative data: a synthesis guiding the novice. *Qualitative Research Journal*, 19(3), 259–270. <https://doi.org/10.1108/qrj-12-2018-0012>

Sun, J., Li, Y., Wang, L., Li, X., Ma, X., Xu, J., & Chen, G. (2015). Controlling Smart TVs Using Touch Gestures on Mobile Devices. *2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom)*. Published. <https://doi.org/10.1109/uic-atc-scalcom-cbdcom-iop.2015.222>

Sutcliffe, A. (2002). Assessing the Reliability of Heuristic Evaluation for Website Attractiveness and Usability. *Proceedings of the 35th Hawaii International Conference on System Sciences*. Published. <https://doi.org/10.1109/HICSS.2002.994098>

Tanabian, S., & Tanabian, M. (2013). Optimizing Usability on Video Streaming Devices and Smart TV's. *Communications in Computer and*

*Information Science*, 164–168. [https://doi.org/10.1007/978-3-642-39473-7\\_34](https://doi.org/10.1007/978-3-642-39473-7_34)

Taylor-Powell, E., & Renner, M. (2003). *Analyzing Qualitative Data*. University of Wisconsin-Extension.

Tavory, I., & Timmermans, S. (2020). *Abductive Analysis and Grounded Theory*. The SAGE Handbook of Current Developments in Grounded Theory. Published. <https://doi.org/10.4135/9781526485656>

Télez, J. F., Montoya, J. C., & Trefftz, H. (2016). Design and usability evaluation of a multi-input interface in an idTV environment. *Applications and Usability of Interactive TV*. Published. [https://doi.org/10.1007/978-3-319-63321-3\\_10](https://doi.org/10.1007/978-3-319-63321-3_10)

Tefertiller, A., & Sheehan, K. (2019). TV in the Streaming Age: Motivations, Behaviors, and Satisfaction of Post-Network Television. *Journal of Broadcasting & Electronic Media*, 63(4), 595–616. <https://doi.org/10.1080/08838151.2019.1698233>

TiVo. (2021). *Netflix's search screen* [Screenshot]. Retrieved August 9, 2021 from, <https://www.tivo.com/support/how-to/watch-netflix>.

Vatavu, R. D. (2021). Accessibility of Interactive Television and Media Experiences: Users with Disabilities Have Been Little Voiced at IMX and TVX. ACM International Conference on Interactive Media Experiences. Published. <https://doi.org/10.1145/3452918.3465485>

Winter S., Wagner S., Deissenboeck F. (2008) A Comprehensive Model of Usability. In: Gulliksen J., Harning M.B., Palanque P., van der Veer G.C., Wesson J. (eds) *Engineering Interactive Systems*. EHCI 2007. Lecture Notes in Computer Science, vol 4940. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-540-92698-6\\_7](https://doi.org/10.1007/978-3-540-92698-6_7)

Williams-McBean, C. T. (2019). The Value of a Qualitative Pilot Study in a Multi-Phase Mixed Methods Research. *The Qualitative Report*, 24(5), 1055–1064. <https://doi.org/10.46743/2160-3715/2019.3833>

Yang, S. H., Liu, X. W., & Lo, Y. C. (2016). A design framework for smart TV: Case study of the TaipeiTech smart TV system. *2016 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*. Published. <https://doi.org/10.1109/icce-tw.2016.7521053>

Yerema, J. (2021, May 22). *Apple TV's Remote Controller* [Photograph]. Unsplash. Retrieved August 19, 2021, from, <https://unsplash.com/photos/dTa1LJAJvhI>

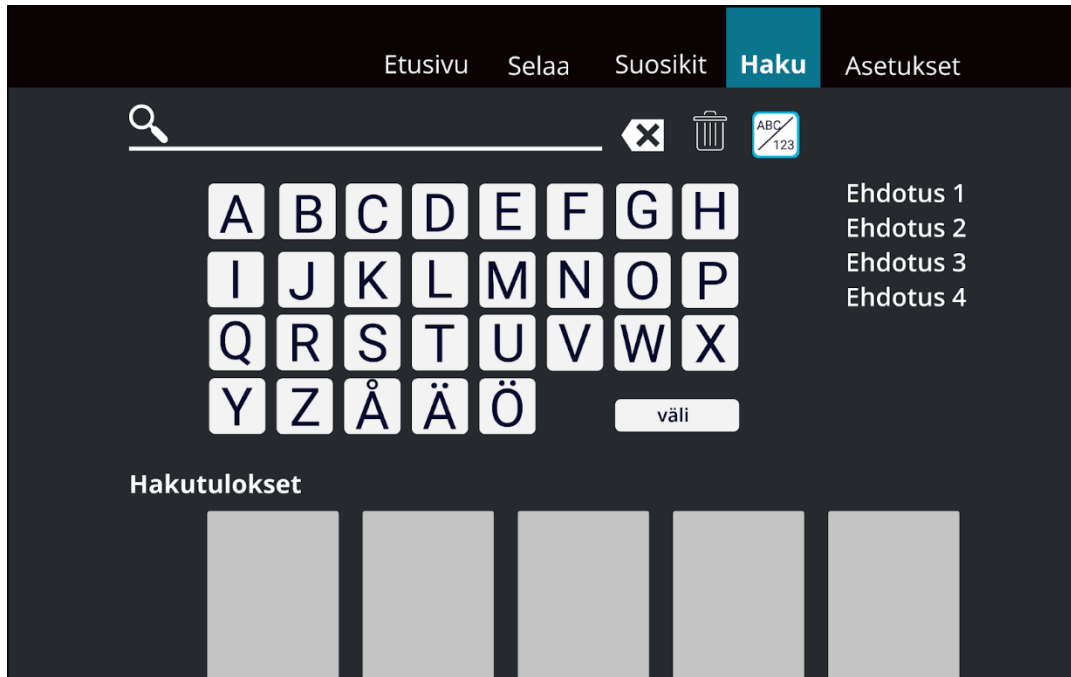
Yleisradio. (2018). *This is Yle*. Retrieved June 5, 2021 from, <https://yle.fi/aihe/artikkeli/about-yle/this-is-yle>

Yleisradio (2020). *Yle is perceived to be more important and reliable, and more people are using Yle's services extensively*. Retrieved June 5, 2021 from, <https://yle.fi/aihe/artikkeli/2020/12/18/yle-is-perceived-to-be-more-important-and-reliable-and-more-people-are-using>



## Appendices

### Appendix 1: Prototype in original language



## Appendix 2: Slides -presentation

Televisio

Striimauspalvelut

Disney+ HBO  
Netflix Yle Areena YouTube  
Twitch Ruutu Cmore  
Viaplay Discovery+

Selailu

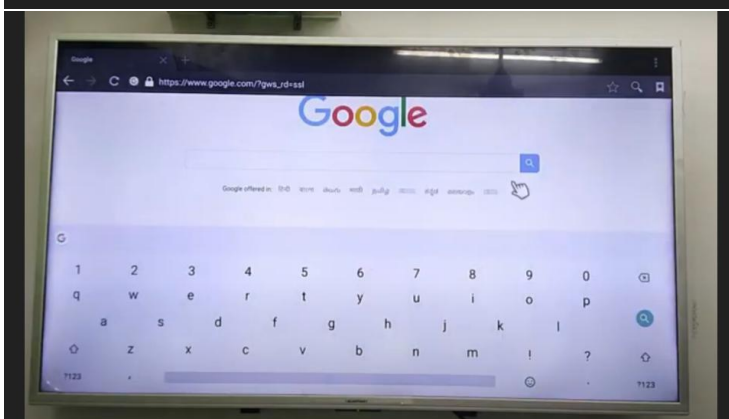
(eng. Browsing)

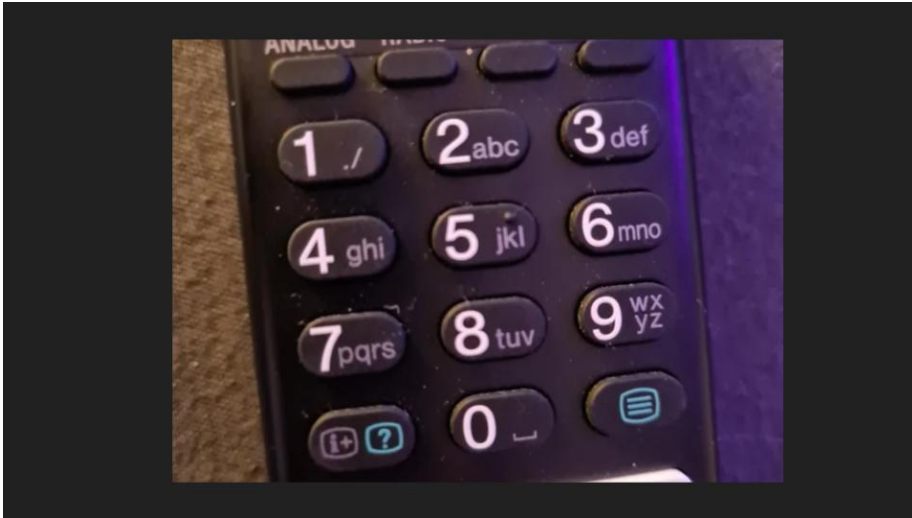
# Television käyttö puhelimella

casting

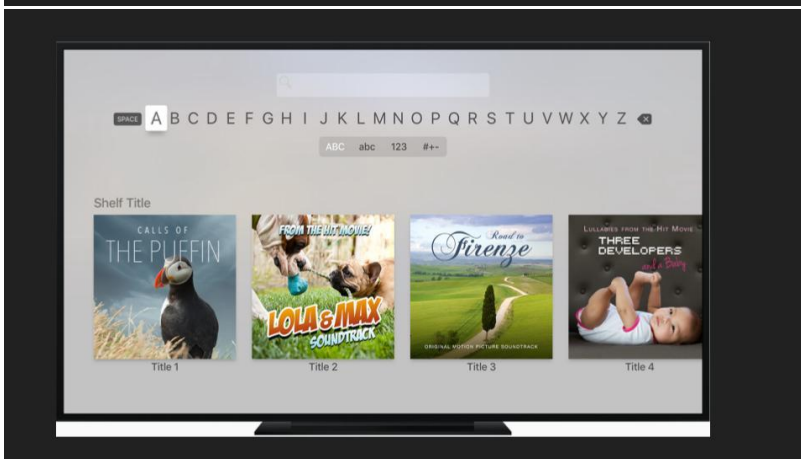
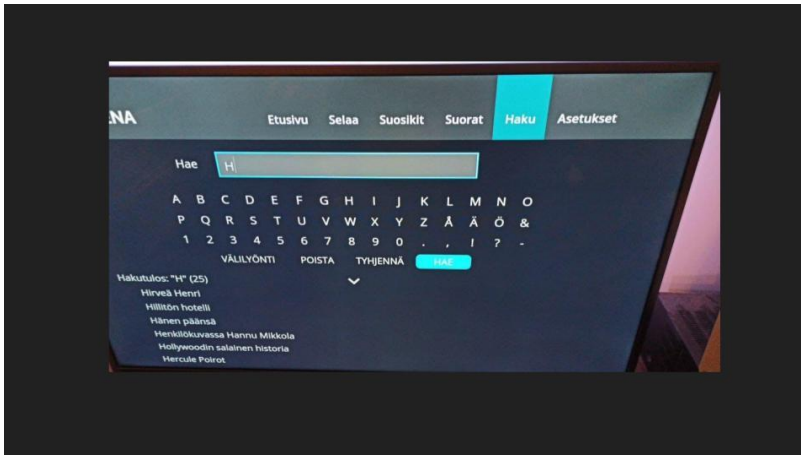
Kaukosäädin

Haku

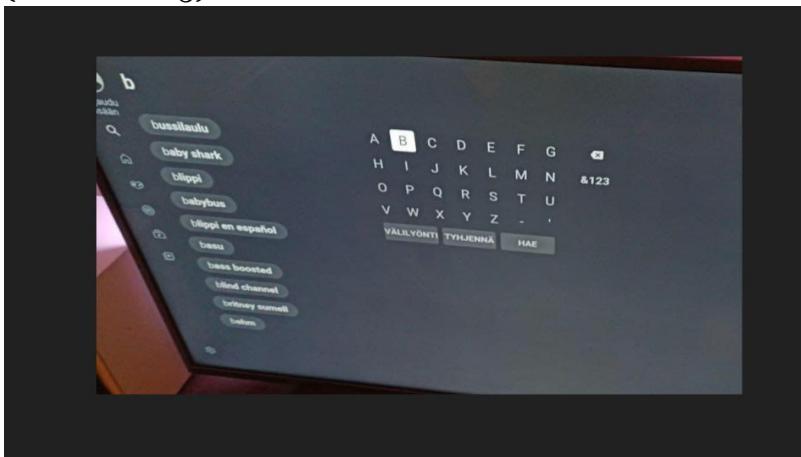




(Sinharay et al. 2013).



(Gatlan. 2015).



## Appendix 3: Interview Plan

### Intro – alkukeskustelu + tekninen säätö (~5min)

” Moikka, varmaan tunnetaankin jo ennestään, mutta mä oon Liisa ja opiskelen Informaatioverkostojen Maasteriohjelmassa Aallossa ja (oon erikoistunut käyttäjälähtöiseen suunnitteluun (User-centered design)). Teen tällä hetkellä mun diplomi- eli lopputyötä Yleisradiolle. Ylellä pyritään kehittämään palveluita paremmiksi asiakkaiden tarpeiden mukaan. Tää mun dippa liittyy älytelkkareihin ja erityisesti älytelkkareiden hakuun.

Haluun vielä muistuttaa, että en siis oo mikään devaaja tai designeri, joka tekisi näitä juttuja, joten mä en pahota mieltäni mistään mitä sanotte. Enkä oikeestaan innostukkaan erityisesti mistään vaan tätä tutkitaan tutkiumisen vuoksi. En myöskään hae mitään tiettyjä vastauksia vaan haluan rehellisiä kommentteja asioihin, joten turha yrittää miellyttää ja pyrkiä johonkin tiettyyn suuntaan lähtemään ”

**Huom varmista vielä uudestaan, että se varmasti oli ok, että tää tallennetaan! Kertaa:**

- Datan säilytys! (data ei tunnistettavissa, mulla access)
- + **aloita tallennus**

### Alotellaan eka helpoilla kysymyksillä

- Minkä merkinen telkkari sulla/teillä on?
- Käytätkö sä paljon telkkaria?
  - *Esim. monta kertaa päivässä, päivittäin, viikottain, kuukausittain jne?*

### ”Näytän tässä joitakin sanoja ja juttuja niin keskustellaan niistä” (5-10 min)

*Presentation avaus + näytönjako päälle!*

- Televisio
    - Mitä se tarkoittaa sun mielestä?
    - Mikä merkitys/rooli?
    - Miten käytetään?
    - Millanen telkkari teillä on?
  - Striimauspalvelut
    - Käytetäänkö niitä?
  - Striimauspalvelut: eri mestat
    - Onks tuttuja?
    - Onks vieraita?
    - Lemppareita?
  - Selailu/Browsing
    - Tuleeko tehtyä striimauspalveluissa?
  - Kaukosäädin
    - haluuks näyttää millanen teillä on?
    - käytätkö sitä vai jotain muuta pääosin?
      - millanen systeemi se sun mielestä on?
    - onks kiva käyttää?
    - käytätkö jotain muuta, esimerkiks just konsoliohjainta?
- Haku
- ihan suora sanahaku,
  - kumpaa enemmän selailua vai hakua?

**Tämän jälkeen erilaisia näppäimistö vaihtoehtoja, joista keskustellaan, mistä tykätään, mitä inhotaan? Onko tuttuja? Millaset fiilikset jokaisen kohalta. Sitten oma näytönjako pois.**

**Vaihdetaan telkkarin puolelle, käännä kamera siten, et se näkyy jotenkin taustalta, kunhan sun puhe edelleen kuuluu.**

Yks mitä mä pyydän että sä teet, niin **pyri kertomaan kaikki mitä sä ajattelet ja teet** koska mä en pysty näkemään kunnolla esim mitä painat - enkä tietenkään arvaamaan mitä sun päässä liikkuu. Ja yritä ettei mennä liian nopeasti, jotta mä kerkeän myös näkemään mitä siellä näytöllä tapahtuu.

**Ajatellaan, että sulla ois nyt tunti aikaa katsoa jotain ohjelmaa, minkä alustan nyt avaisit,**

- Miksi?
- voitko näyttää miten tekisit sen?

**Mitä jos sulla olis joku vaikka kahden kaverin kanssa sovittuna joku eläkuvalta,**

- mistä katsoisit,
- mitä katsoisit? Miten etsit sen?

Jos ei päätynyt jo käyttämään hakua edellisessä:

**\*\*Käytä HAKUA?! \*\***

**Miten etsisit? Käytätkö hakua.**

**Sitten kehoitetaan menemään youtubeen jos ei jo siellä**

käytätkö sitä?  
miten etsit?  
käytätkö hakua?  
mitä tykkäät hausta?

**Kuinka monta merkkiä odotat, että tarvitsee kirjoittaa ennen kuin löydät haluamasi?**

**Areena**

Millanen fiilis?  
miten etsit?  
käytätkö hakua?  
mitä tykkäät hausta?



## **Areenan jälkeen ”tulevaisuuden” Areena eli figmaproto**

<https://www.figma.com/proto/2LJONYEKeDenLiiDIuxGLf/Untitled?node-id=27%3A9&starting-point-node-id=27%3A9&scaling=scale-down>

Millanen fiilis verrattuna nykyiseen Areenaan?  
Entä verrattuna muihin olemassaoleviin palveluihin?

### **Tää linkki käyttäjille!**

#### **---- muistutus haastattelun ajaksi ---**

Tavotteena:

- vertailla eri vaihtoehtoja
    - nykyinen Areena
    - joku kilpailija
      - Netflix (itelle tunnukset tähän?)
      - tai Youtube
    - Ite tehty Figma prototyyppi
      - vrt nykyiseen
- saada irti ihmisten mielipiteitä, fiiliksiä jne  
mikä on ihmisten mielestä paras  
EI sitä montako klikkausta menee tms..  
Muistuta Think Aloudista jos käyttäjät hiljenee kesken!

**QWERTY vs ABCD vs T9** mielipide?