Vapriikki Case: Design and Evaluation of an Interactive Mixed-Reality Museum Exhibit

Silvia Rubio Hernández

University of Tampere
Faculty of Natural Sciences
Software Development
M.Sc. thesis
Supervisor: Markku Turunen
December 2017

University of Tampere
Faculty of Natural Sciences
Software Development
Silvia Rubio Hernández: Vapriikki Case: Design and Evaluation of an Interactive Mixed-Reality Museum Exhibit
M.Sc. thesis, 78 pages, 3 index and 5 appendix pages
December 2017

Museums have the challenging job of educating while entertaining. In their attempt to achieve such, they have introduced the concept of “hands-on” exhibits, where visitors can interact first-hand with the phenomena being displayed. With the emergence of new technologies and devices, museums have exploited the opportunities hands-on exhibits can offer to further enhance interactivity. At the same time, researchers and museum workers have become aware of the importance of conducting evaluations to guide and assess the design and development process of these exhibits.

Traditionally, exhibit evaluations have utilized ethnographically-oriented methods and gathered all data manually. Though some studies have used log file analysis to explore interaction patterns, evaluations still heavily rely on the traditional methods. This thesis explores the design and evaluation processes of interactive exhibits by presenting a case study of the interactive replica of Abraham Edelcrantz’s shutter telegraph developed for the Vapriikki Museum in Tampere, Finland. Additionally, this work investigates the usage of a semi-automated log analysis in combination with qualitative methods to evaluate interactive museum exhibits.

The results of this thesis show that a semi-automated method can be used to separate and analyze sessions and conduct a longitudinal analysis on interactive exhibits. Furthermore, our findings indicate that the exhibit satisfied visitors with different interests because of its multiple appealing elements: problem-solving challenge, communication aspect and historical significance. Additionally, results show that interaction and collaboration in groups differs depending the age composition of group members. Finally, this thesis presents a set of design guidelines for interactive exhibits.

Key words and terms: Interactive Exhibits, Museum Exhibit Evaluation, Log File Analysis, Museum Exhibit Design

Acknowledgements

Writing this master’s thesis has been a long journey and I honestly cannot believe that it has come to an end. Throughout this process I received guidance and help from many people who I would like to thank.

First and foremost, I want to express my gratitude to Prof. Makku Turunen, for his guidance and patience during this thesis work and for encouraging me to *“just start writing”*.

I would also like to thank Dr. Jaakko Hakulinen for his valuable input and expertise, which were extremely helpful during the exhibit’s design and development. I am immensely grateful to Tuuli Keskinen who helped me in innumerable occasions throughout the evaluation and writing process. I also want to thank Pekka Kallioniemi for his kind comments and Sumita Sharma for her valuable advice and thorough remarks. To all my colleagues at TAUCHI, thank you for the support and trust.

This thesis work would not have been possible without the kind contribution from Outi Penninkangas, Niklas Nylund and all the Vapriikki staff, who aided me during the observations sessions and welcomed me in my uncountable visits to the museum.

Writing this thesis was slow and sometimes painful. I would like to thank all the wonderful friends I have made in Tampere, who kept supporting me along the way and cheering me up. Catalina, thank you for being a kind and awesome flatmate; Jobin, for your always-clever insights; Paula and Draupathy, for keeping me company. To all my loyal friends from Venezuela and to my family, thank you for always being there for me.

Finally, I would like to thank my dear Philipp, for your immeasurable support, kind words and endless patience. I could have done this without you, but oh how much harder it would have been!

Contents

1. Introduction

Museums, as culture propagators, have always played an important part in educating the public. Since the beginning of the 20th century, museums around the world started introducing simulations in their exhibitions. Inspired by this movement, the Exploratorium museum (in San Francisco, USA) opened their doors in 1969 with a new “hands-on” or interactive approach, which promoted learning through the exploration of physical objects or other phenomena [Caulton, 2006]. Even though in the beginning the hands-on movement did not involve the use of technologies, it was not long until museums started introducing the latest technologies emerging from the ICT (Information Communication Technologies) area. These technologies ranged from simple computer systems to more sophisticated ones such as tangible user interfaces and most recently augmented-, mixed- and virtual reality.

Designing museum exhibits can present challenges due to the – sometimes overwhelming – nature of museum environments where visitors stroll around only stopping at the most interesting spots [Falk et al., 1985]. For this reason, it is important to create an understandable exhibit that users can comprehend within a few seconds; it must be clear what kind of actions can be performed with the physical [Gibson, 1977; Norman, 1999] and non-physical parts of the exhibit [Allen, 2004]. Nonetheless, despite the designers’ efforts to make the design clear and provide a certain experience with the exhibit, visitors can end up using the exhibit differently from its conception. Considering these and other issues encountered when designing different exhibits, various studies aim to provide general guidelines that can be applied to the design of other interactive exhibits [Hinrichs et al., 2008; Marshall et al., 2016].

Museum exhibit designers and developers are increasingly aware of the close relationship between conducting evaluations throughout the development process and achieving successful exhibits. These evaluations can aid from the ideation stage until after the exhibit has been installed. Summative evaluations are conducted once the exhibit is in place to assess if it is achieving its purpose and, if necessary, amend possible flaws in the design [Caulton, 2006]. These evaluations often measure the success of an exhibit based on how attractive an exhibit is to users passing by (stopping power), how long users stop at or interact with an exhibit (dwell time) and if the exhibit conveys the intended message (communication power). Additionally, since exhibits are mostly visited in groups, some studies focus on understanding interaction patterns and other forms of collaboration in groups [Heath et al., 2005; Vom Lehn, et al., 2001]. In recent years, interactive exhibit evaluations also study how easy it is to use the exhibits (usability) and how it makes users feel (User Experience).

In order to obtain the metrics previously described, evaluations often use a combination of different data collection methods. Traditionally, museum exhibit studies have heavily relied on ethnographically-oriented, qualitative methods such as surveys, observations and interviews. These methods are effective to collect in-depth data of each visitor; however, they are costly on human resources. Thus, studies utilizing only these methods often gather data from a small set of visitors. Most recently, with the spread of computers and other technological devices used on interactive museum exhibits, interaction data is collected in the form of automated log files. Using log files provides means to obtain a larger set of data without using human resources. Nonetheless, log file data on its own can be "shallow" [Lazar et al., 2014] as it does not provide enough details and cannot be mapped to a specific user.

This thesis reviews the design of interactive exhibits and the methods utilized by other researchers and experts to evaluate them. It presents the design and evaluation process of an interactive replica of the shutter telegraph installed in the Vapriikki Museum. The focus of this work is the evaluation of interactive exhibits, more specifically, this thesis addresses the research question: *How to utilize a semi-automated log data analysis to evaluate interactive exhibits over a long period of time?* Furthermore, this work focuses on the combination of log data analysis with other ethnographic data collection methods to obtain a good level of understanding on the interaction with the exhibits. Finally, this thesis compiles the lessons learned from the design and evaluation of the Vapriikki optical telegraph exhibit and provides a set of guidelines aimed to help improve the design of this and future interactive exhibits.

1.1. Research motivation

Nowadays, interactive exhibits have a significant presence in museum exhibitions. With the constant development of ICT, it is likely that this kind of exhibits will increasingly find their way in museum spaces. Exhibit designers and developers are aware of the importance of evaluating museum exhibits. However, the evaluation methods used in many studies are still very similar to those used in the 60s as they rely on observations and interviews or other kinds of ethnographically-oriented studies. Such studies gather mostly qualitative data and are usually conducted for short periods of time because of the human resources they require.

Though many hands-on exhibits today make use of technologies, log data is rarely used or plays a smaller role in the evaluation of interactive exhibits. Log files offer the possibility of gathering large amounts of data that can be analyzed later to extract information depending on the goals of the study. Furthermore, log data analysis requires less resources and can be used to conduct evaluations for longer time periods. The challenge of using log data analysis for the evaluation of museum exhibits is presented in how to discern between different user sessions. Some studies explicitly require users to log the information using physical objects or cards [Hornecker and Stifter, 2006a; Marshall et al., 2016], however, such method relies on users to collect the log data. Therefore, it becomes relevant to investigate possible solutions to automatically log data and discern between user sessions in the analysis stage. Finally, it is key to explore the opportunities of combining a longitudinal log file analysis with traditional ethnographic methods with the goal to obtain a deeper understanding of the interactions with hands-on exhibits.

1.2. Research contribution

The contribution of this thesis work is a two-fold: 1) it suggests **a semi-automated method to separate user sessions on log file data** and combines it with traditional ethnographically-oriented research methods to obtain in-depth information for the evaluation. 2) it presents **a set of design guidelines for interactive exhibits** based on the findings of this empirical work and the results of related work. The results of this thesis can serve as basis for future studies that utilize log file analysis to separate user sessions such as museum exhibit or public display evaluations. Moreover, our proposed design guidelines aim to assist the decision-making process on the design of future interactive exhibits.

1.3. Thesis outline

This thesis consists of nine chapters. Chapter 2 provides an overview of hands-on exhibits, their history and design. Chapter 3 reviews the evaluation methods commonly used in museum exhibit studies as well as the metrics and different aspects evaluated on them. Chapter 4 describes the shutter telegraph technology replicated in the exhibit and explains the system design process and rationale behind the design choices. Later, Chapter 5 gives a detailed description of the exhibit installation, its components and the system architecture. Furthermore, the methodology and data collection methods are discussed in Chapter 6, followed by a thorough presentation of the findings on Chapter 7. Chapter 8 provides a discussion of the most interesting findings of this work and introduces a set of guidelines based on the results of this work. Finally, Chapter 9 summarizes all findings and future work.

2. Hands-on Museum Exhibits

A visit to the museum today is far from what it used to be in the last century. In their attempt to remain interesting to the public, museums have moved away from the old concept of “hands-off” exhibits, where visitors could only see and read facts about the objects displayed, to a more interactive and playful approach that entices their visitors. Caulton [2006] defines a hands-on or interactive exhibit as one whose goal is to educate by promoting users to physically explore the real objects or phenomena. He further clarifies that hands-on exhibits do not require technology to be interactive. For instance, an exhibit might encourage visitors to touch several different objects while being blind-folded and try to identify what they are.

Interactive exhibits are widely popular and spread among museums today, many of them employ various technologies such as computers, tablets, virtual- or augmented- reality or tangible elements. This chapter introduces the concept of interactive exhibits, reviews their history and explains some of the reasons for their popularity. Finally, it explores the different challenges, concepts and other aspects taken into consideration in the design of interactive exhibits.

2.1. History of interactive museum exhibits

The transition from “hands-off” to “hands-on” exhibits was not sudden. The process started in the early 20th century when museums all over the world – Deutsches Museum, Munich from 1925; Chicago Museum of Science and Industry from 1933, Palais de la Découverte, Paris from 1937, among other museums – started to introduce simulations as part of their exhibitions. The trend inspired the opening of the Exploratorium in 1969, which is considered the first institution to have a real hands-on approach [Allen, 2004; Caulton, 2006].

With the rapid development of ICT, museums have tried to incorporate – almost in parallel to the development process – the most current and novel artifacts to their exhibitions. It started with the introduction of basic computer exhibits and as new technologies emerged they also found their way to museum exhibitions.

The early 90s witnessed the initiative to move away from traditional human-computer interactions, using a desktop computer and mouse, towards interaction with real world objects which are augmented with the use of computers [Fitzmaurice, 1993; Wellner et al., 1993], a concept known today as Augmented Reality (AR). During this early stage, Rekimoto and Nagao [1995] foresaw the application of AR in the museum context. However, it was not until after the 2000s when implementations of augmented exhibits spread around museums [Wojciechowski, Walczak et al., 2004; Woods et al., 2004].

Inspired by augmented reality and ubiquitous computing, Ishii and Ullmer [1997] suggested the idea of interfaces where users could touch and manipulate real-world objects coupled to the virtual world. This idea evolved into the concept of Tangible User Interfaces (TUIs) where tangible elements, also referred as tangibles [Manches et al., 2009], are used as input in the interaction. Naturally, given the hands-on quality of this TUIs model, they were also introduced into the design of museum exhibits [Hall and Bannon, 2005; Horn et al., 2008; Rizzo and Garzotto, 2007].

Most recently, a definition of Mixed Interactive System (MIS) has been used to describe those interactive systems that have elements of augmented reality (AR), mixed reality (MR) and tangible user interfaces (TUIs) [Schmitt et al., 2010].

2.2. Popularity of interactive exhibits

The growing presence of interactive exhibits, especially those involving technologies, in museums is motivated by more than a desire to include novel technologies. Since the very beginning of this movement, the Exploratorium and other museums have studied the effects of such exhibits on visitors' experiences and behaviors.

One of the reasons why interactive exhibits have been successful is because they attract and hold users' attention for longer periods than traditional exhibits [Hornecker and Stifter, 2006a; Korn and Jones, 2000]. Hornecker and Stifter [2006a] further identified that traditional exhibits attract users with a specific interest (history, nostalgia or technology) whereas interactive exhibits are the only kind to attract visitors from all interest profiles.

Researchers have also studied hands-on exhibits as aid to the learning process. Allen [2004] explains that a certain level of interactivity in exhibits can help visitors recall more details about their experience. Furthermore, she describes the Exploratorium's work on interactive exhibits that promote learning and foster a "minds-on" effect, in other words, exhibits that encourage further thinking and allow for knowledge-seeking conversations to arise among its users. Falk et al. [2004] evaluated visitor learning after using various types of interactive exhibits. They concluded that interactive elements promote different levels of learning (knowledge and skills; perspective and awareness; motivations and interests; and social learning) both in short-term and long-term periods after the visits.

Besides attracting users and potentially enhancing learning, interactive exhibits can provide a better visitor experience. Studies comparing visitors opinions after using interactive and non-interactive versions of an exhibit show that visitors rate interactive exhibits as more enjoyable [Allen, 2004; Panagiotis et al., 2013].

2.3. Designing interactive museum exhibits

The museum environment presents several challenges to exhibit designers. In museums, visitors usually have a limited amount of time and a wide range of competing exhibit's to explore [Hornecker and Stifter, 2006b]. A key factor to attract users is to encourage them to start interaction. Lee et al. [2015] describe their approach when designing the exhibit *Trap it!*, they used prompt messages and cues as instructions to invite and guide users at the beginning of the interaction.

Once a visitor's attention is caught, the exhibit must successfully hold this attention. The first few seconds of interaction in museum environments are crucial [Hornecker and Stifter, 2006a], therefore, it is vital that the exhibit is understandable to the users to keep them from leaving. For this purpose, the *walk-up-and-use* paradigm is applied to the design of museum exhibits [Hinrichs et al., 2008; Lee et al., 2015]. Its basic principle is that instructions and interaction techniques should be simple for users to explore. Other factors can facilitate or hinder the first minutes of interaction, for instance, the position of controls has a great impact in users' interpretation of the exhibit. The design concept of *natural mapping* suggests placing controls in a way that correspond to what is being controlled [Norman, 2013].

In cases where physical objects are used, the design must make the *affordances* evident, in other words, it should be clear what can be done with the exhibit at a first glance [Gibson, 1977; Norman, 1999]. Failures to convey the correct *affordances* of the exhibit may result in *hidden affordances,* where users do not know that a certain action can be performed, or *false affordances*, where users mistakenly perceive that an object allows a certain action [Gaver, 1991].

Another factor playing against exhibit designers is a phenomenon known as museum fatigue, where visitors can no longer concentrate on exhibits after a certain period and only look for something particularly interesting [Falk et al., 1985]. To counteract this effect, Allen [2004] introduces the term of *Immediate Apprehendability*, which is the ability of an exhibit to convey its purpose and properties with no apparent cognitive load to its users. This concept is a bit broader than that of *affordances* since it can be applied to non-physical objects such as labels and instructions in the exhibit.

Though designers have a clear idea of how the interaction should go, there are factors that can alter the interaction from the way it was schemed. For instance, visitors might perform an action that was not anticipated in the design [Allen and Gutwill, 2004] or reveal answers of the exhibit's puzzles and challenges to the next user in line before they get their turn to interact [Heath et al., 2005].

Some exhibit evaluations contribute to the design process by providing guidelines based on their experiences. During the evaluation of *EMDialog*, an information visualization exhibit, Hinrichs et. al. [2008] identified guidelines applicable to other kinds of interactive exhibits:

- **Rewarding for short-term and long-term exploration:** the design should consider that users may take different amounts of time in the installation and aim to provide information or rewards to users regardless of the time they stay engaged.
- **Supporting collaborative information exploration:** the design and technologies used should support collaborative exploration for groups of visitors.
- **Making information exploration appealing:** making the interaction effortless and attractive can counter effect the performance aspect of interacting in such exhibits.

- **Supporting various exploration styles:** guiding those users who might need help and providing more freedom to more expert users.

Besides the challenges faced when planning any kind of interactive exhibit, designing interactive replicas requires extra considerations. While making design decisions for *REXband*, an interactive medieval band replica, the creators discuss the challenges of attempting to balance *authenticity*, *education* and *entertainment* [Wolf et al., 2007]. Their work summarizes the challenges as three paradoxes:

- **The Edutainment Paradox** (entertainment vs. education): refers to the goal of educating the audience while remaining entertaining.
- **The Disney Paradox** (entertainment vs. authenticity): explains the challenge of creating an entertaining exhibit without sacrificing the authenticity of the replica.
- **The Museum Paradox** (authenticity vs. education): specifies the difficulty of filtering the amount of information about the object being displayed yet providing a learning experience to visitors.

Hands-on exhibits have become part of museum environments and offer a major opportunity for visitor learning. Hence, the design of these exhibits must be carefully planned to fully exploit the advantages of interactivity. Design concepts and guidelines provide a good starting point in the design of exhibits, as they provide a source of learning from the experiences of others. Nonetheless, it is not sufficient to rely only on guidelines when designing a new exhibit, though many of their suggestions are applicable to similar exhibits, different elements such as the museum environment, audience and other factors can differ from exhibit to exhibit. Therefore, it is imperative to conduct evaluations on each exhibit to check if the design needs revision and find possible flaws and lessons that can aid exhibit designers in the future.

This thesis presents a study case of the design, implementation and evaluation of an interactive replica of an optical telegraph for the Vapriikki museum. This study uses exhibit evaluations as part of an iterative design process and aims to provide a comprehensive set of guidelines for similar interactive exhibits in the future. The following chapter addresses the concepts, methodology, metrics and other aspects to consider when conducting exhibit evaluations.

3. Evaluation of museum exhibits

Nowadays, the design and development of museum exhibits is closely tied to their evaluation. Different types of evaluations can be conducted throughout the various stages of development to assess and improve an exhibit. Planning evaluations requires careful consideration as it is important to choose the appropriate data collection methods and metrics depending on the goals of each evaluation. This chapter describes the types of museum exhibit evaluations and how they correspond to the different stages of development. Moreover, it explains the advantages and disadvantages of the data collection methods commonly used in museum exhibit studies. Finally, it explores the criteria and metrics widely used in museum studies to assess the success of exhibits.

3.1. Types of evaluation

Exhibit evaluations vary depending on the stage of the exhibit development. Caulton [2006] defines three types of evaluations: *front-end analysis*, *formative evaluations* and *summative evaluations*.

During the ideation stage, a *front-end analysis* aims to comprehend how target visitors understand the potential exhibition. *Formative evaluations* are conducted during the mock-up phase to assess if users understand the elements and the concept of the exhibit. The aim of formative evaluations is to correct mistakes during the development of the exhibit, they can be conducted in a real context or in an isolated environment. Early evaluations can be especially useful in co-design processes and when there is a need to choose between prototypes of alternative designs [Schmitt et al., 2010]. Additionally, observations during the development process *"serve not only as a method to detect design flaws but are also needed to adjust system parameters"* [Wolf et al., 2007]. *Summative evaluations* aim to identify possible issues with the exhibit once it is installed. Their purpose is to evaluate the success of the exhibit, repair flaws that may remain, and to acquire knowledge for improving the design and evaluation of future exhibits.

3.2. Data collection methods

Museum exhibit studies utilize a combination of various data collection methods such as observations, interviews, surveys and automated log files. Early studies conducted on traditional exhibits relied only on ethnographic methods to collect their data. Recently, as computers and other technologies are used to provide interactivity in exhibits, the number of studies that utilize log files and other automated data collection methods to gather data has risen. The reasons to select a specific method, or a combination of them, depends on the objectives of each study. The following sections explain some of the advantages and disadvantages of each technique and how they were used in other museum exhibit studies.

3.2.1. Surveys

Surveys are used in several areas of research to collect large amounts of data for statistical analysis. Particularly, in the Human-Computer Interaction (HCI) field, questionnaires are regularly used to evaluate User Experience (UX) [Hassenzahl et al., 2015; Turunen et al., 2009].

The biggest benefit of utilizing surveys is that they do not require researchers to be present while being filled, facilitating the collection of large amounts of data [Lee et al., 2015]. Additionally, they are less obtrusive than interviews, which can be useful to obtain further insight from shy users who are not willing to be interviewed. On the down side, data collected using surveys can be “shallow” and highly depends on the researcher’s bias when designing the questionnaire.

3.2.2. Observations

Researchers rely on observations to evaluate all kinds of systems. Observations can be held in a lab-based environment or in the real context. In the case of museums exhibits, observations outside a controlled environment provide useful insights into different ethnographic aspects of the visitors such as their age, gender, whether it is a group, family group or singleton, among other factors.

A great advantage of observations is that researchers can write down interesting behaviors that happen only “in situ” and would not occur in a lab environment. For example, during the design process of *REXband*, an interactive medieval band replica, researchers conducted a user study at a public fair. They noticed that visitors were reluctant to try their installation because it was surrounded by “hands-off” exhibits [Wolf, Lee, and Borchers, 2007]. With such information, they re-evaluated their wishes about the location of the final exhibit installation inside the museum and considered how to indicate to users that it was an interactive exhibit.

One significant drawback of conducting observations is that it requires extensive human resources. Observations in museum studies can last as little as one day [Lee et al., 2015] or several weeks [Hinrichs et al., 2008]. In some cases, more than one researcher needs to be in place to be able to collect and observe the necessary amount of information needed for the analysis. For this reason, longitudinal analyses using this method are especially difficult to perform.

Moreover, if observation is the only method of data collection, it is not possible to fully understand the visitors' reasoning behind certain actions. There is also a risk that researchers might mistakenly draw conclusions about what is being observed. Consequently, observations on museum exhibits are usually held in combination with other methods such as interviews or surveys to obtain a deeper understanding of the usage and better evaluate the system at hand.

3.2.3. Interviews

While evaluating any kind of system, it is essential to obtain perceptions first-hand from the users. For this reason, most evaluations utilize interviews of some form during the design, development and post-deployment. Interviews work as a complement to other methods such as observations, surveys or log analysis. Interviewing users can validate findings obtained with other techniques and clarify the reasons for certain behaviors and patterns observed during usage or found in the log data. In other cases, interviews might contradict results from other methods and show users' true perceptions and preferences [Lazar et al., 2014].

A great advantage of interviews, specifically semi-structured and unstructured interviews, is that they give interviewers freedom to ask questions that may arise from a user's previous answer. This kind of exploration and first-hand user information cannot be acquired with any of the other techniques previously mentioned.

While interviews provide an excellent way to explore deeper into users' perceptions, they require extensive resources, first during the interview and then later during the analysis [Lazar et al., 2014]. Ergo, studies using this method are usually limited to a small number of interviewees [Grinter et al., 2002; Hornecker and Stifter, 2006b; Lee et al., 2015].

3.2.4. Log files

An early exploration of the evaluation of museum exhibits suggests that careful design in the data being logged can provide good answers for the evaluation with little extra work in the process [Heinecke, 1995]. Log files are an effective, unobtrusive and unbiased way to gather large amounts of data for later analysis. In contrast to observations and interviews, log file recording does not require human resources and can be performed over several days, months and even years to perform longitudinal analyses that would be impossible to achieve otherwise.

In museum exhibits, log data is often used to understand how visitors explore and use the content available on an exhibit [Boehner et al., 2005; Heinecke, 1995; Lee and Heller, 1997]. Moreover, it can be useful to compare diverse aspects of alternative design solutions. For example, Horn et al., [2012] compared the effect of two types of user interface – graphical vs. tangible – on the number, length and complexity of programs created by users of a programming exhibit at a science museum.

The biggest advantage of using log file data in studies is the large amount of data that can be collected with this method. Hornecker and Stifter [2006a] utilized log files to interpret data from several months and evaluate a media and communications exhibition. Their study revealed usability issues in several exhibits and helped identify which of the exhibits were more successful to attract and engage users. This information in combination with observation sessions resulted in improvements in the media exhibition.

Nonetheless, this large amount of data collection is not without a price, as log data tends to be “shallow” [Lazar et al., 2014] and contain less details than the information collected with other methods such as interviews. Another disadvantage of log data is that, when used alone, it is not possible to map the interaction to a certain user.

Log data analysis can also be used to quantify aspects of interaction that would be difficult or impossible to live-code while performing observations. Block et al. [2015] developed an algorithm to identify groups of users in the log data and quantified engagement in group interactions with a tabletop exhibit. Similarly, this thesis work uses log data analysis to count the number of codes constructed, the number of messages sent and switches moved, among other such metrics.

Though log file analysis has made its way in museum exhibits evaluations most of the results are still based on observations and interviews. It is important to understand the advantages of using log file analysis to evaluate exhibits over larger periods of time and with less resources. Nevertheless, the combination of qualitative and quantitative methods and data triangulation are necessary to support log analysis findings, clarify patterns and answer questions that may emerge from the data.

3.3. What is evaluated on museum exhibit studies?

It could be said that the general goal of museum exhibit evaluations is to assess if the exhibit is successful. Dean [1996] points out a series of questions that should be asked when evaluating the success of an exhibit. These questions address different aspects such as the exhibit's attractiveness, visitors' learning, visitors' experiences, among other factors. Studies have different criteria to assess whether an exhibit has achieved its purpose and there seems to be no consensus on how to measure an exhibit's effectiveness [Shettel, 2001]. The following sections describe a series of measurements and other criteria commonly used to evaluate museum exhibit studies.

3.3.1. Stopping power (Attractive power)

The first step towards having a successful exhibit is that it is attractive for users to stop at it, and in the case of hands-on exhibits, to start interacting with it. This characteristic is known as an exhibit's *stopping power* [Vom Lehn and Heath, 2005] or *attractive power*. The stopping power is represented as the percentage of potential users of an installation – visitors that pass the exhibition within a certain distance – that are drawn to start interacting with it. To measure the stopping power, researchers conduct observation sessions or analyze video recordings from the installation to calculate the number of potential users that interact with the system. In this work, we measure the stopping power during the observation sessions by tallying the number of visitors who pass the exhibit and those who stop to interact.

3.3.2. Total number of interactive sessions

Exhibit evaluations also consider the total number of interactive visitor sessions during a certain period. The number of sessions can be counted manually during observation sessions [Horn et al., 2008; Korn and Jones, 2000] or using log file data [Hornecker and Stifter, 2006a; Marshall et al., 2016].

Depending on the design of the exhibit, there are several ways to log data and record the number of sessions. Some studies count sessions using smartcards or other kinds of artifacts that users need to explicitly scan at the exhibit [Hornecker and Stifter, 2006a; Marshall et al., 2016], even though the action is optional, the incentive for users is that they receive additional information about the exhibit or a record of their own performance.

Logging the start and end of sessions using visitors' explicit actions can provide a concrete method to track users' movement across exhibitions and clearly discern different user sessions. However, they also present some limitations [Hornecker and Stifter, 2006a]. First, it requires users to scan the card when they enter the exhibit, some may forget to scan the card, leaving those sessions unrecorded. Second, the data is collected partially because only a portion of users are willing to put extra effort, especially if they need to buy additional artifacts such as smartcards, or use other physical objects while walking around the museum [Hornecker and Stifter, 2006a; Marshall et al., 2016].

Some exhibits may not be suitable to apply the methods described previously, therefore, it is important to find other options on how to distinguish different sessions. An alternative approach can be found in studies on public displays where researchers use idle times to separate sessions [Marshall et al., 2011; Peltonen et al., 2008]. Their strategy assumes a certain time gap where there are no interactions between one session and another. This thesis uses the same rationale to identify and calculate the number of user sessions. However, in the public display studies the separation was performed manually using video recordings whereas in this work the separation is performed automatically based on logged interactions.

3.3.3. Dwell time (holding time)

Studies often rely on *dwell times* – the time a user spends at an exhibit – to evaluate exhibits [Horn et al., 2012; Horn et al., 2008; Hornecker and Stifter, 2006b; Lee et al., 2015]. The reason for its importance is that it can be associated to the engagement with the exhibit and therefore studies link it to the exhibit's success [Horn et al., 2008; Lee et al., 2015].

Earlier work established an average dwell time of thirty seconds on museum exhibits [Cone and Kendall, 1978]. In the Exploratorium museum, studies were conducted to compare the dwell times among different exhibits. They identified exhibits which fostered an “active prolonged engagement” in users [Allen, 2004]. Horn [2012] cites evidence that the average dwell times of such exhibits was measured to be 3.3 minutes as compared to 1.1 minutes on traditional interactive exhibits. Sandifer [2003] investigated whether certain exhibit characteristics can contribute to longer dwell times. He measured and compared holding times on 61 exhibits, previously classified using four categories: *technologically novel*, *open-ended*, *user-centered* and *stimulates the senses*. His study asserts that *open-ended* exhibits – where visitors can freely perform different tasks or they can perform one task in several ways – and *technologically-novel* exhibits hold visitors’ attention longer than exhibits without these characteristics.

Though dwell times are widely used, they can be deceiving, as others affirm that interacting for longer periods does not necessarily indicate more engagement or reflect a better visitor experience [Shettel, 2001]. Instead, it can be due to difficulties encountered when using the installation [Heath et al., 2005].

As with stopping power, dwell times can be measured during observation sessions or can be calculated from logged data. When calculating dwell times from logged data, user entry and exit can be explicitly marked – e.g. using smartcards [Hornecker and Stifter, 2006a]. Nevertheless, this method relies entirely on users who may forget to scan the card when they leave the exhibit, leading to unrealistic session lengths. In this work, dwell times are calculated based on the sessions that derive from the semi-automated log file analysis. Additionally, dwell times are manually annotated during observation sessions to make comparisons with the automated results.

3.3.4. Communication power

A museum’s main goal is to act as a source of informal learning for its visitors. Hence, when exhibits are designed, they have a certain educational purpose or a message that they should deliver. The ability to deliver such message is known as an exhibit’s *communication power* [Vom Lehn and Heath, 2005].

Though some exhibits might succeed to attract (*stopping power*) and hold visitors (*dwell time*) for a certain time, they can still fail to educate visitors about the subject in matter. Dean [1996] argues that *"visitor numbers do not indicate whether anyone is taking away knowledge"*. Additionally, Shettel [2001] discusses that some of the measurements used to evaluate an exhibit's effectiveness simply fall short if the educational message is not delivered. He further exemplifies a case where users observed to interact at a certain exhibit for lengthy periods, could not answer later what the exhibit was about.

Studies commonly measure visitors' learning from exhibits by conducting surveys or interviews which assess their knowledge on the exhibit's topic prior to their visit and afterwards [Falk and Dierking, 2000; Panagiotis et al., 2013]. In this work, the communication power is measured subjectively by enquiring users about their learning in the interview and survey questions. Moreover, we aim to measure communication power objectively using the semi-automated log analysis (Section 6.5.3).

3.3.5. Group interactions and collaboration

Museum visits are social in nature as most museumgoers come with families or other forms of groups [McManus, 1994; Vom Lehn et al., 2001]. For this reason, some researchers evaluate their exhibits to assess if they are suitable for usage in groups [Horn et al., 2008]. Other studies try to identify patterns in group interactions [Hornecker and Stifter, 2006b; Lee et al., 2015] or establish comparisons between different groups and singletons [Fernández and Benlloch, 2000; Hinrichs et al., 2008; McManus, 1994].

Hornecker and Stifter [2006a] evaluated group interaction in several exhibits within a media exhibition, and their findings reveal that different factors in the exhibit, such as its size constraints and input controls, can affect the group sizes and the forms of interaction. The age composition of various groups can also affect the way interactions unfold. Sandifer [1997] found that children spend more time interacting with exhibits as compared to adults, similarly, McManus [1994] asserts that groups containing children are most likely to try interactive systems than other groups (adult groups, singletons, couples).

Groups can interact in various manners. Studies in different exhibits converge in collaboration or turn-taking as a common form of interaction

[Hinrichs et al., 2008; Hornecker and Stifter, 2006a; Lee et al., 2015]. In their evaluation of *Trap it!*, besides collaborative interactions, Lee et al., [2015] further identified two other types of interaction: guidance and competition. For example, some users would guide or explain the exhibit to others in the group. In addition, they observed a single occurrence of a competitive interaction among a pair of children. These various forms of collaboration also depend on the group type, for instance, family groups show an inherent teaching quality where parents usually guide and help children analyze and learn from the exhibit [Crowley et al., 2001; McManus, 1994].

A quantitative metric to identify an exhibit's suitability for group interactions is the number of passive and active users. Active users directly interact with the system while passive users remain as observers during interaction [Hornecker and Stifter, 2006a]. The rationale is that if an exhibit successfully allows for group interactions, more users will actively interact with it. This thesis work follows the same criterion and uses the number of active users as metric to assess whether the exhibit successfully supports group interactions and collaboration.

3.3.6. Usability

Nielsen [2012] describes usability as *"a quality attribute that assesses how easy user interfaces are to use (…)"*. Keeping exhibits usable is fundamental as an exhibit with usability issues could translate into an exhibit with low attendance [Hornecker and Stifter, 2006b].

Usability has five different components [Nielsen, 2012]: learnability, efficiency, memorability, errors and satisfaction. Though all components should be assessed, learnability – the ability to quickly grasp how to use the system for the first time – is probably the most relevant in the context of museum exhibits. If visitors require a long time to understand how to interact, they may lose interest and leave.

Evaluations during the exhibit's development [Schmitt et al., 2010] and after deployment [Hornecker and Stifter, 2006b] aim to identify and correct usability issues that could hinder the success of an otherwise valuable exhibit. As part of this study, observation sessions are used during the development and after deployment to identify possible usability issues on the exhibit.

3.3.7. User experience (UX)

In the past decade, system designers and developers have realized that a product's design goes beyond simple functionality and they have acknowledged the importance of how it makes users feel. Hassenzahl [2008] defines user experience as *"a momentary, primarily evaluative feeling (good-bad) while interacting with a product or service"*. UX evaluations now take part in the product and service design and development process, researchers and even companies invest time and effort to improve the experience of interactions.

In a similar way, museum environments have tried to innovate and variate their traditional style of displaying old objects and providing information in the form of long texts. Museums aim to provide a delightful experience where, after their visit, visitors feel that they have learned something new and had a pleasant time while doing so. Therefore, it could be said that museums have the important mission to *educate while entertaining*.

For these reasons, current museum exhibit evaluations also aim to assess the user experience in their installations [Gebhard and Karsten, 2009], compare the user experience between alternative prototypes [Schmitt et al., 2010] or measure whether interactive systems contribute to better experiences [Panagiotis et al., 2013]. In this study case, one of the goals of the exhibit is to entertain the visitors. Therefore, UX is used to measure the level of satisfaction in users and to assess if they found the exhibit fun and entertaining.

This chapter presented an overview of the complex process behind exhibit evaluations. These evaluations have become popular as part museum studies because their importance is widely known in the museum context. Museums can perform evaluations in different stages of the development process and utilize several data collection methods. However, the methodology used in the evaluations has not changed much over the years, with most studies still utilizing ethnographically-oriented methods and qualitative data to assess the exhibit for a short period of time.

Furthermore, this chapter reviewed several metrics and other factors used to evaluate the success of an exhibit: stopping power, total number of interactions, dwell time, communication power, interactions and collaboration, usability and User Experience. Nonetheless, it is relevant to note that the research presented on this topic mostly uses ethnographically-oriented methods to measure these metrics. Conversely, this thesis work explores the possibility of using log file data to automatically calculate some of the metrics – total number of interactions, dwell time and communication power – and validate the results with data obtained from surveys, interviews and observations. Moreover, this thesis investigates the use of formative and summative evaluations to guide the design process and assess the final installation of an interactive replica. Finally, this study utilizes log file data and a semi-automated log data analysis to perform a longitudinal analysis of the exhibit.

4. Vapriikki's optical telegraph design

Ever since hands-on exhibits were first introduced in the Exploratorium museum, they have become a desirable element in museum environments because of the advantages they offer over traditional exhibits (Section 2.2). Museum curators and designers increasingly aim to include interactive exhibits in new exhibitions, such was the case of the Rupriikki 2.0. exhibition in the Vapriikki museum, where a few interactive exhibits were planned to be part of the new exhibition.

The Vapriikki MIRACLE project was a collaboration between the Vapriikki Museum in Tampere and the Pervasive Interaction Research Group (PIRG) group from the Tampere Unit for Computer-Human Interaction (TAUCHI) research center in the University of Tampere. The goal of the project was to create an augmented-reality, interactive version of the shutter telegraph invented by Abraham Niklas Edelcrantz in the 1800s. This interactive replica was to become part of the exhibition Rupriikki 2.0. about the history of media and communications. The intention behind the installation was to demonstrate to users the operation of the telegraph through a hands-on experience. In this way, museumgoers could obtain a deeper understanding of how the communication worked with such apparatuses and enhance their visit beyond merely learning historical facts.

This chapter describes the design process of the optical telegraph replica developed during the Vapriikki MIRACLE project. First, it provides a foundation of the telegraph topic by describing the theory behind Edelcrantz's original design, followed by a description of the initial design of the replica and an overview of the two design phases in the project. Later, it shows a detailed explanation of each phase and the rationale behind the design choices taken to obtain the final exhibit installation.

4.1. Edelcrantz's optical telegraph

A crucial goal of the Vapriikki MIRACLE project was to educate users about the history and operation of the optical telegraph. The Vapriikki museum wanted to include an exhibit that would replicate the telegraph tower and its components and the codification principle used in the communication. Therefore, as part of the exhibit's design process, it was relevant to investigate the original design of Edelcrantz's shutter telegraph to give visitors a historically-accurate experience.

Helzmann and Pehrson [1994] describe the shutter telegraph from Abraham Edelcrantz. This optical telegraph model consisted of a tower with ten shutters: the top-middle shutter followed by a three-by-three shutter matrix below it, as shown in Figure 1.

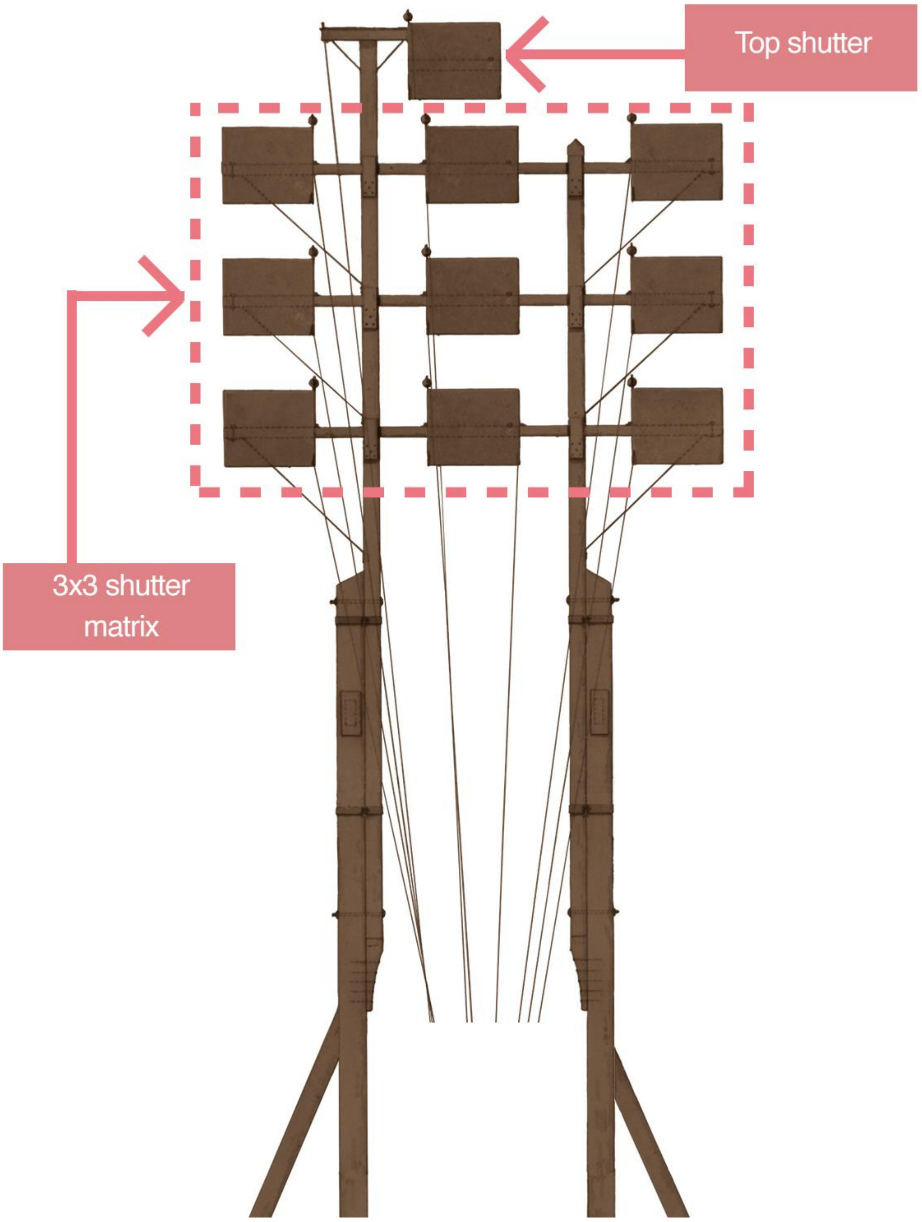


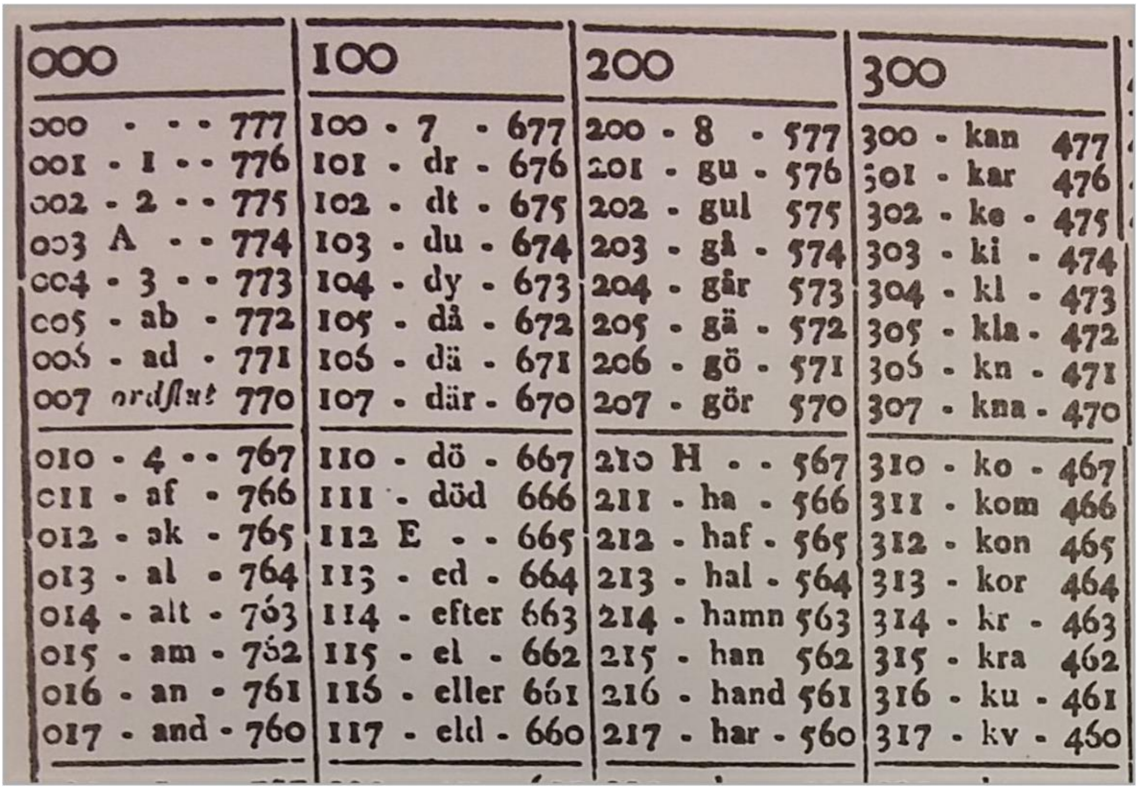
Figure 1. *Telegraph tower overview with all shutters visible.*

Each shutter had two possible states: hidden or visible. Showing or hiding different shutters could create a total of 1024 possible combinations. Each one of these combinations represented a different code or signal in a communication codebook present in every tower. Many consecutive codes were needed to send one entire message to the next tower.

The communication between cities required several telegraph towers to be built within sight of each other in some sort of "communication line". A message – comprised of many individual signals – was sent from an originating tower and was continuously repeated by all consecutive towers until it reached the destination tower.

4.1.1. Codification principle

Every telegraph tower had a codebook containing several codes or signals that could be formed by changing the states of the shutters. Figure 2 shows an excerpt of a codebook used in the telegraph towers. As it can be seen, a single code generally represented a small combination of (1-4) characters.



000			100			200			300		
000		777	100	7	677	200	8	577	300	kan	477
001	1	776	101	dr	676	201	gu	576	301	kar	476
002	2	775	102	dt	675	202	gul	575	302	ke	475
003	A	774	103	du	674	203	gå	574	303	ki	474
004	3	773	104	dy	673	204	går	573	304	kl	473
005	ab	772	105	då	672	205	gä	572	305	kla	472
006	ad	771	106	dä	671	206	gö	571	306	kn	471
007	ordſlut	770	107	där	670	207	gör	570	307	kna	470
010	4	767	110	dö	667	210	H	567	310	ko	467
011	af	766	111	död	666	211	ha	566	311	kom	466
012	ak	765	112	E	665	212	haf	565	312	kon	465
013	al	764	113	ed	664	213	hal	564	313	kor	464
014	alt	763	114	efter	663	214	hamn	563	314	kr	463
015	am	762	115	el	662	215	han	562	315	kra	462
016	an	761	116	eller	661	216	hand	561	316	ku	461
017	and	760	117	eld	660	217	har	560	317	kv	460

Figure 2. *Excerpt from an original codebook (adapted from* [*Tahvanainen, 1994]).*

In order to send a full message, multiple signals needed to be used to produce a final, concatenated message. For instance, following the codebook from Figure 2, if a tower wanted to send the message "*handed 3*" it needed to emit the signals: 216, 113, 000 (space) and 004.

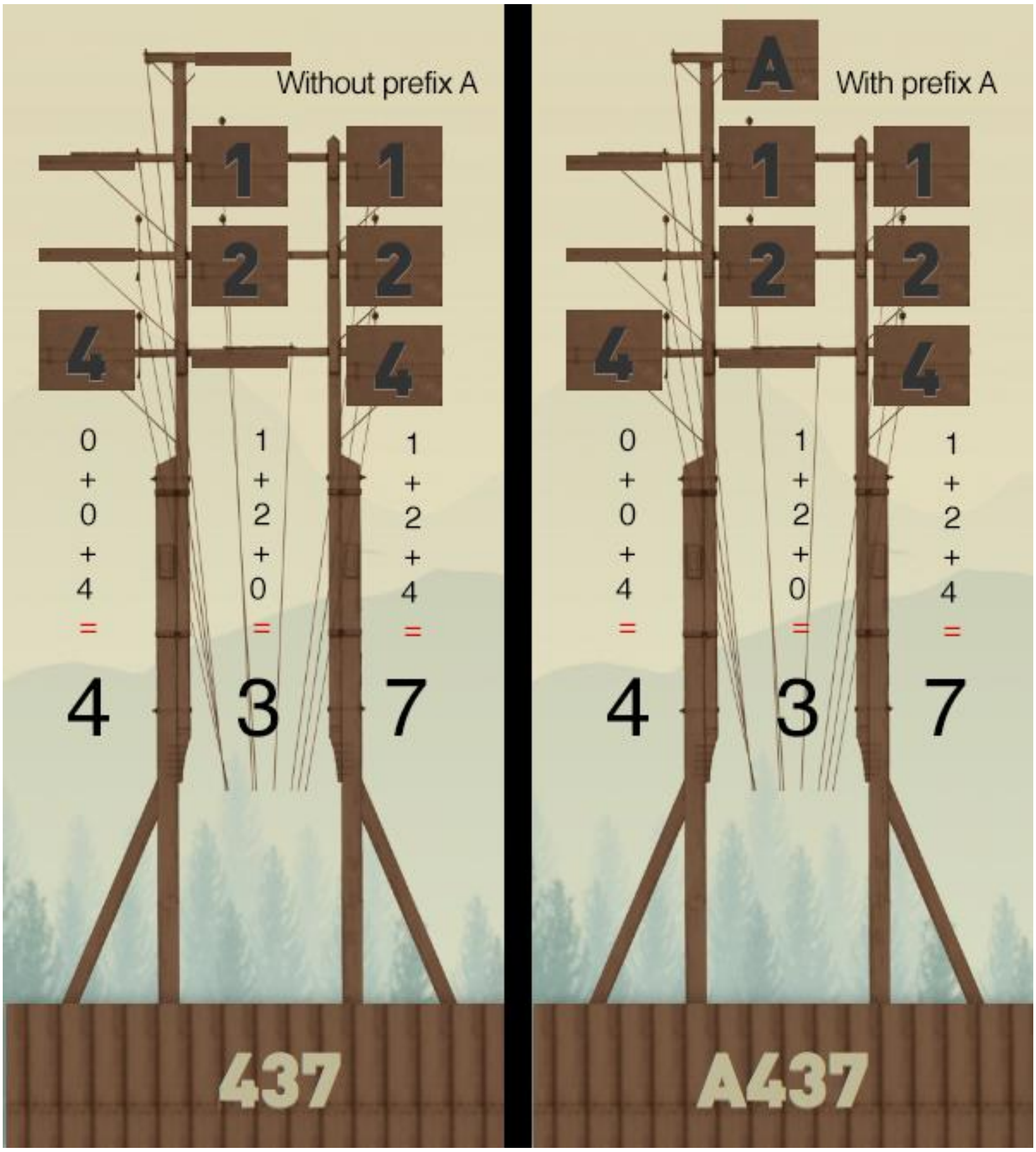


Figure 3. *Example of codification principle and calculations.*

The codification of a single signal consisted of a three-octal number, which could be prefixed by the letter A. The octal numbers were formed by adding the values of each shutter column in the matrix. Hidden shutters had a value of 0. Visible shutters, on the other hand, had different values depending on their row numbers. Shutters in the first row had a value of 1, shutters in the second row had a value of 2 and those in the third row had a value of 4 [Holzmann and Pehrson, 1994].

Finally, the prefix A depended on the status of the top flag. When the top flag was hidden then the signal consisted only of the three-octal number.

Conversely, if the top flag was visible then the prefix A would be added to beginning of the signal. Codes with the A prefix and their non-prefixed counterpart represented different messages (letter combinations). Figure 3 depicts how the shutters looked when emitting the code signals 437 and A437. Additionally, it explains the sum calculations behind each matrix column.

4.2. Initial design of Vapriikki's optical telegraph replica

The main goal behind the design of the replica was to provide an entertaining simulation of the communication while giving a realistic overview of the telegraph's codification mechanism. In order to achieve this goal, the design focused on two key elements.

The first element was the communication with the chatbot. The chatbot was an artificially-intelligent component which would reply appropriately to the messages sent by users. The idea was to immerse users into the experience of having a conversation with the inventor of the shutter telegraph Abraham Edelcrantz. Additionally, a set of pre-defined user messages were created in such a way that the purpose of the conversation would be to talk about Edelcrantz's life and the history of the telegraph. Thus, the interaction in the exhibition would provide some historical facts in a different, more engaging manner.

The second, and perhaps most important element, was the simulation of the codification mechanism. Users would be able to send and receive message codes similarly to the way it was performed in the original telegraph, using a codebook. Ideally, the replica would use the original codebook (Figure 2), however, the original version was complex and would require a great amount of time for users to send a single message. Demanding such a long time from visitors in favor of historical authenticity would arguably take away part of the entertainment factor, evidencing the Disney Paradox (Section 2.3). Therefore, two different codebooks were used in the replica: one that would favor entertainment (user codebook) and one that would favor authenticity (chatbot codebook).

The first codebook version is the user codebook, Table 1 shows the English version of such codebook (Finnish version in Appendix 2). As can be seen, it is a considerably simplified version of the original codebook. It contains a total of 21 codes, each representing a full pre-defined message. The simplified user codebook has two main advantages: the first is that it facilitates the interaction since users only require one code to send a message, the second is that the predefined messages are designed to lead users to converse about the history of the telegraph and the life of its inventor.

Code	**Message**
A001	MESSAGE RECEIVED
A002	DO YOU COPY?
A004	UNDERSTOOD
A005	REPEAT MESSAGE
A007	END COMMUNICATION
A010	WHO AM I TALKING TO?
A020	TELL ME MORE
A077	YES
A700	NO
A777	MAYBE
A301	WHO?
A302	WHEN?
A304	WHERE?
A311	WHY?
A312	HOW?
A313	WHAT?
A401	FIRST
A402	SECOND
A403	THIRD
A701	FORMER
A702	LATTER

Table 1. *User codebook.*

The second codebook, is the chatbot codebook (Edelcrantz's codebook) with a total of 142 codes (Appendix 1). This version of the codebook resembles the original codebook logic and allows freedom in message generation. Table 2 shows an extract of the chatbot codebook, it can be noted that a code represents a combination of characters, generally 1 to 6 characters. These character combinations were tailored based on the message pool available in the chatbot. Unlike the user codebook, the same version is used for communication in Finnish and English.

Using two different codebooks allows users to interact with the system and construct the pre-set codes easily while still getting the responses from the chatbot in a format that follows the historical facts more accurately. Additionally, because the user codebook considerably reduces the amount of time required to send a single message, users can send few or many messages, enabling short-term and long-term explorations (Section 2.3).

CODE	MESSAGE	CODE	MESSAGE
000	0	**011**	8
001	1	**012**	9
002	2	**013**	A
003	3	**014**	AA
004	4	**015**	AI
005	5	**016**	AL
006	6	**017**	AN
007	7	**020**	AAN
010	SPACE	**021**	AR

Table 2. *Extract of the chatbot codebook.*

4.3. Design process overview

The design period lasted approximately eight months (Feb 2015 -Oct 2015) and consisted of two main phases: the prototype phase and the final exhibit phase (Figure 4). The goal of the first phase was to implement a prototype to test the design concept and improve the design of the final installation. Subsequently, a second phase consisted of revising the prototype design based on the information and feedback gathered during its evaluation to obtain the final exhibit design.

The co-design process involved staff from the Vapriikki Museum (graphic designers, hardware experts, curators, etc.) together with people from TAUCHI's PIRG group (researchers and a research assistant). The museum staff's responsibilities concerned the design of the exhibit stand, hardware installations and managing the resources available within the exhibition area. The PIRG group on the other hand, handled the system's design and the implementation of the digital replica. My contribution was the implementation of the web-based application, the graphical user interface design, and contributions to the system architecture. At the same time, the chatbot service was designed and implemented separately by two of the researchers in the project, therefore, a detailed explanation of its design is out of the scope of this thesis.

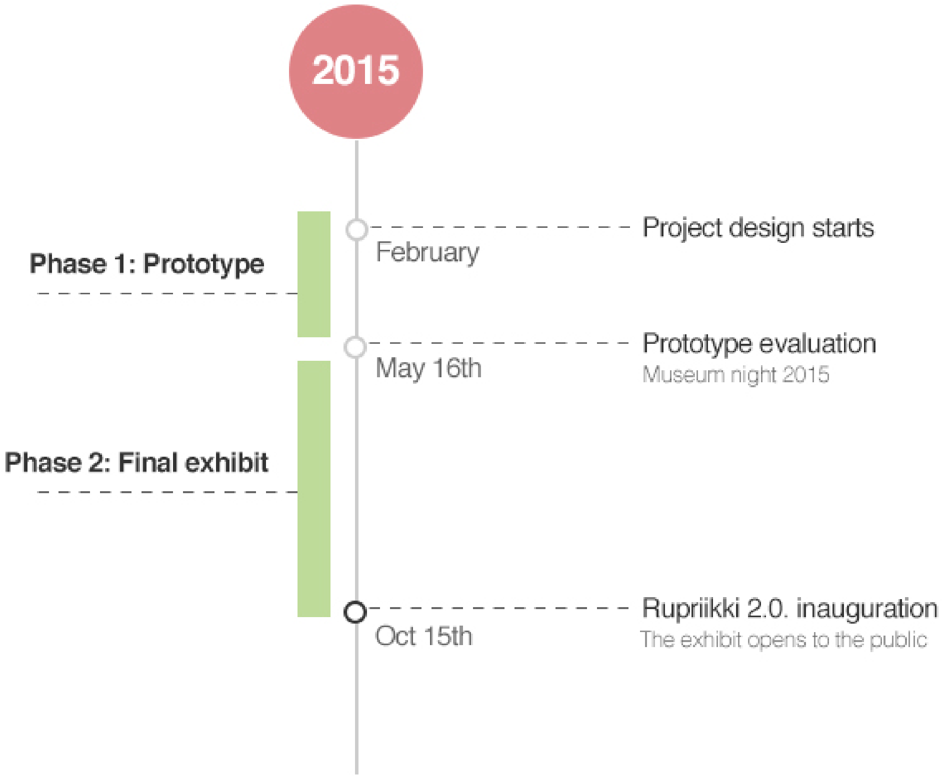


Figure 4. *Design process timeline.*

4.4. Phase 1: Prototype

In the first prototype, the basic interaction set up was presented as two towers. The local tower where users could send messages, and the remote tower, where users would receive messages from a virtual version of the inventor of the telegraph. The virtual version of Edelcrantz, the chatbot, was an artificially-intelligent agent designed to converse with users and provide historical facts about the telegraph and its inventor. This first demo was only available in Finnish.

The demo station, presented in Figure 5, was set on a table in the museum and it comprised the following elements:

1. One laptop showing a detailed remote tower view (Figure 6). This view showed the remote tower including some hints to help users crack the coding principle. Additionally, it showed the decoded meaning of every incoming code and the full message as it was being concatenated.

2. A tablet with a virtual representation of the local tower view that visitors interacted with. Such interaction was performed via the touch screen. Users needed to tap on the tower shutter they wished to hide or show. Figure 7 shows the local tower view, which contains the local tower, the code message produced and a chat history with the conversation.

3. Two paper sheets for the two different codebooks. Each one indicating their purpose: transmission or reception.

4. A projection of the simple remote tower view (no hints given) on a fabric wall. The projection was placed at approximately 3 meters from the demo station table. The purpose was to give an illusion of a tower located at a far distance, simulating how optical telegraphs were used in the past.

Another important characteristic of this first prototype was the customization of its views, meaning that the different codification hints could be hidden or made visible before users started interaction. Those hints were:

- Numbers on each shutter: indicating their value (1, 2, 4 or A) depending on the shutter position.
- The tower code: result of adding shutter values.
- The decoded code: value in the codebook for the tower code.
- The decoded message (only in remote tower view): message being formed.

The purpose of these customizable views was to try different configurations with visitors and assess the ideal amount of aid that should be given to users to make the exhibition challenging yet entertaining.

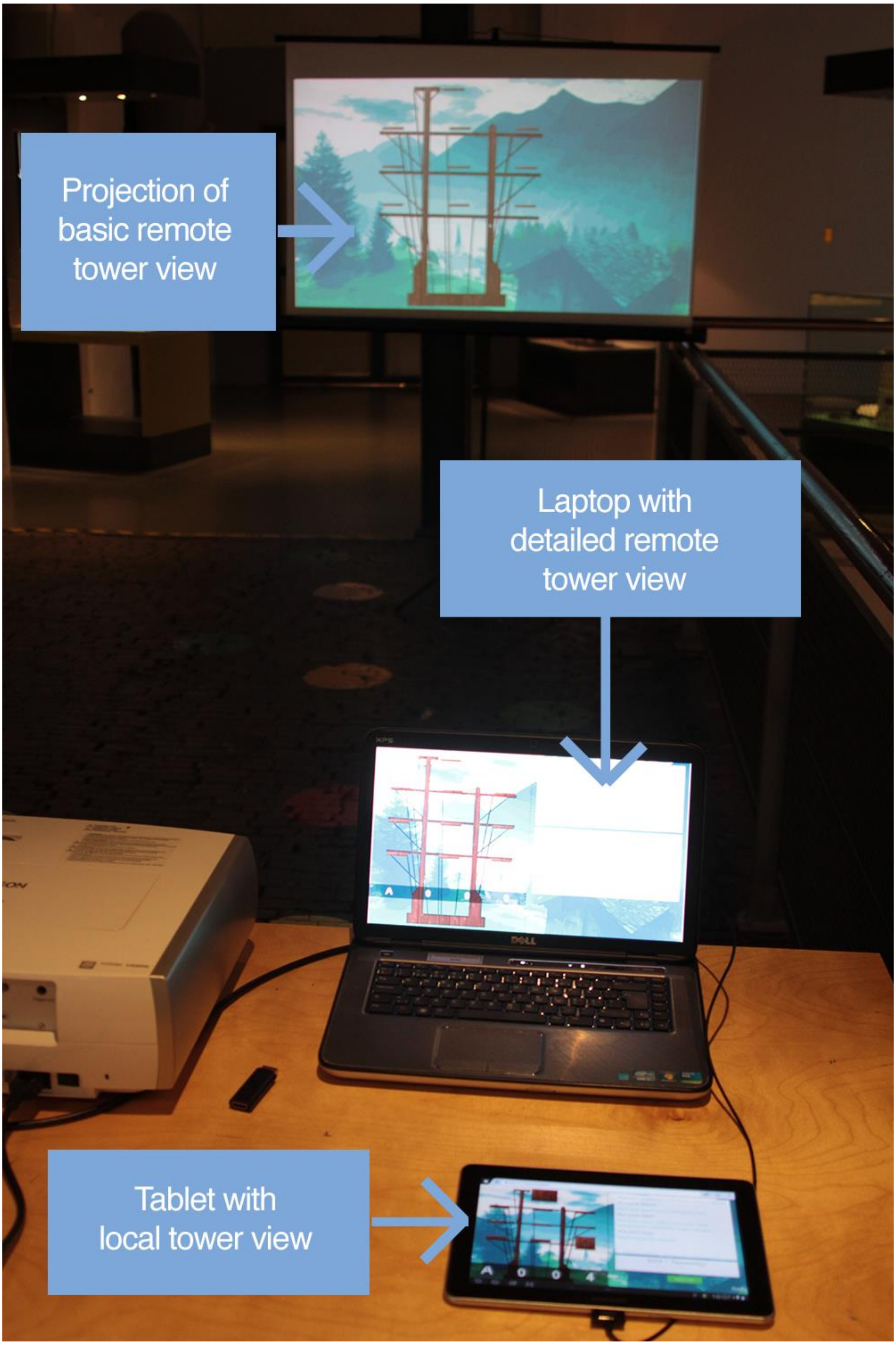
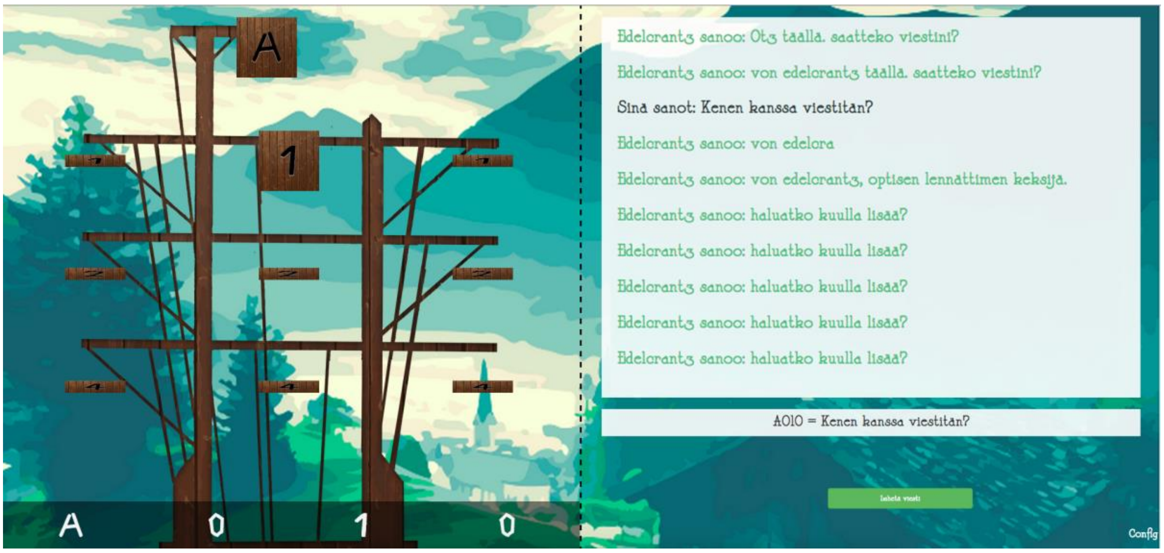


Figure 5. *Demo station for the first prototype.*

340 = le

olen von edelcrantz, optisen

Figure 6. *First prototype: Detailed remote tower view. A part of a message while being formed translated: "I am von edelcrantz, the optical's".*



Edelcrantz sanoo: Otz täällä. saatteko viestini?

Edelcrantz sanoo: von edelcrantz täällä. saatteko viestini?

Sinä sanot: Kenen kanssa viestitän?

Edelcrantz sanoo: von edelcra

Edelcrantz sanoo: von edelcrantz, optisen lennättimen keksijä.

Edelcrantz sanoo: haluatko kuulla lisää?

Edelcrantz sanoo: haluatko kuulla lisää?

Edelcrantz sanoo: haluatko kuulla lisää?

Edelcrantz sanoo: haluatko kuulla lisää?

Edelcrantz sanoo: haluatko kuulla lisää?

A010 = Kenen kanssa viestitän?

Figure 7. *First prototype: Local tower view. The local message A010 reads: "Kenen kanssa viestitän?" (Who am I talking to?). On the top right, a chat section shows a conversation between the chatbot and the user.*

4.4.1. Prototype evaluation

A formative evaluation using a prototype was conducted during the museum night (Figure 4) at the Vapriikki museum. Three different data collection methods were used in the evaluation: observation, surveys and interviews. The focus of this evaluation was to assess how the telegraph operation could be explained to users and whether they found the idea of the replica compelling. Two main factors were evaluated from the demo: 1) the amount of information (hints) given to users to help them understand the codification mechanism and 2) the content richness of the chatbot conversations, in other words, that the vocabulary and responses from the chatbot provided an interesting conversation to the users.

There were two researchers and a research assistant on site to assist users and collect data about the system and its usage. Users filled surveys and were interviewed after using the prototype. In addition, researchers observed and took notes while users interacted with the system. The observation session had a duration of approximately 5 hours. During that time, a total of 12 users filled the survey questionnaire, of which 11 also participated in the interview with one of the researchers.

Relevant findings from this first phase were:

- Most participants enjoyed using the system and considered the idea to be interesting.
- Users had difficulties and needed time to figure out the codification logic, even when all possible hints were given.
- The chatbot responses were a bit monotonous and visitors expressed their wish for more extensive conversation possibilities.
- Users were confused about the two different codebooks on the table, namely, the user codebook and the remote tower codebook.
- The interaction points in the tablet were not clear at first, otherwise put, users did not know that they could click on the tower shutters to show and hide them.

4.5. Phase 2: Final design of the Vapriikki optical telegraph

The final design of the replica preserved the same two key elements as the initial design: the communication and the codification principle (Section 4.2). Furthermore, based on the findings of the prototype evaluation, some revisions were made to the initial design. The design of the final installation (Figure 13) contains the following changes:

1. **Simplified remote tower representation:** As described in Section 4.4, the first prototype had two representations of the remote tower: a laptop view with all codification hints and a distant projection. Given space constraints in the exhibition space, the projection could not be placed at a distant location from the exhibit's stand. Therefore, both representations of the remote tower were combined into one, a remote tower projected on a veneer plate on top of the local tower view. This combined remote tower

view contains three codification hints: the shutter numbers, the tower code and the decoded message.

2. **All hints visible in both views:** The formative evaluation revealed that despite all the hints given, it was still challenging to understand the coding principle. Therefore, hints were included on the local tower (number on shutters, tower code and decoded code) and on the remote tower (numbers on shutters, tower code and decoded message). The purpose of these hints was to make facilitate the understanding of the coding principle and increase the communication power of the exhibit (Section 3.3.4).

3. **Switches as interaction input:** In the first prototype the state of the local tower shutters was controlled by tapping on the tablet touch screen. However, the evaluation revealed that users did not realize they could tap on the shutters. Therefore, the design of the final replica uses switches as interaction input. There were two main reasons for the change: 1) the affordances [Gibson, 1977; Norman, 1999] of switches are evident and make the first minutes of interaction easier, 2) the tangible characteristic could offer a fun factor to the exhibit.

4. **Hidden chatbot codebook:** The first prototype used two different codebooks that were presented as two paper sheets. Having to check two codebooks confused users, hence, it was decided to abstract the level of difficulty and show only the codebook necessary to send messages (user codebook). Hiding the second codebook would presumably reduce the cognitive load on users, however, advanced users could still understand the logic when looking at the hints presented in the remote tower view.

5. **Extended chatbot vocabulary:** During the prototype evaluation, users expressed that the communication was basic and requested more communication possibilities. Thus, the chatbot vocabulary was extended with more conversation topics and answers. The intention behind it was to elicit longer conversations and increase the exhibit’s dwell time (Section 3.3.3).

This chapter explained the process and design choices that led to the final installation of the shutter telegraph replica. The design of the installation was part of a collaborative design process between the Vapriikki staff and the PIRG group. The process consisted of two phases: a prototype phase and a final exhibit phase. In the first phase, a prototype was created based on the original concept of Edelcrantz's shutter telegraph, trying to maintain as much of the historical accuracy as possible. In the second phase, based on the results of the prototype evaluation, some revisions were made to obtain the final exhibit design. This iterative design process and the use of a formative evaluation was key to assess the appropriate level of challenge in the exhibit and it revealed other opportunities for improvements such as the use of switches for input interaction and the simplification of the remote tower view. The following chapter thoroughly describes the final installation and all its components.

5. Vapriikki's optical telegraph: system description

After the process and design choices described in Chapter **Error! Reference source not found.**, the final exhibit installation was implemented for the inauguration of the exhibit Rupriikki 2.0. The final exhibit consists of an optical telegraph tower printed on a veneer plate. It has two main sections: a local tower where users send messages, and a remote tower that communicates back. This chapter offers a thorough description of the final exhibit, starting with an overview of the installation and an in-depth explanation of the tower views and interaction points of the system. Lastly, this chapter provides an outline of the system architecture and illustrates the relationship between the different components of the system.

5.1. Installation overview

The installation is presented as an optical telegraph tower printed on a veneer plate. However, in practice, it simulates two towers, a local tower where users send messages to a remote tower and a remote tower that communicates back. Figure 8 shows a schema of the final installation and its different components, the real installation is shown in Figure 13.

The upper part of the veneer plate is augmented using a projector and it represents the remote tower section. The purpose of this tower is to display the incoming communication from the remote interlocutor. The communication is shown as moving shutters that turn to create different kinds of message codes, in the same manner as the original telegraph would. Additionally, each shutter contains a numerical value – used for codification purposes – to help users deduce the logic behind the message codes they receive.

The lower part of the veneer installation represents the local tower section; it comprises a built-in monitor, some switches and buttons utilized for user interaction. The monitor shows a virtual representation of the local tower, which can be controlled by using multiple switches to change the states (ON or OFF) of each shutter. In addition, the screen displays some additional information such as a biography from the telegraph's inventor and the local tower's codebook. Moreover, a chat history is presented on the screen where users can see the messages they have sent and received, allowing them to keep track of the conversation.

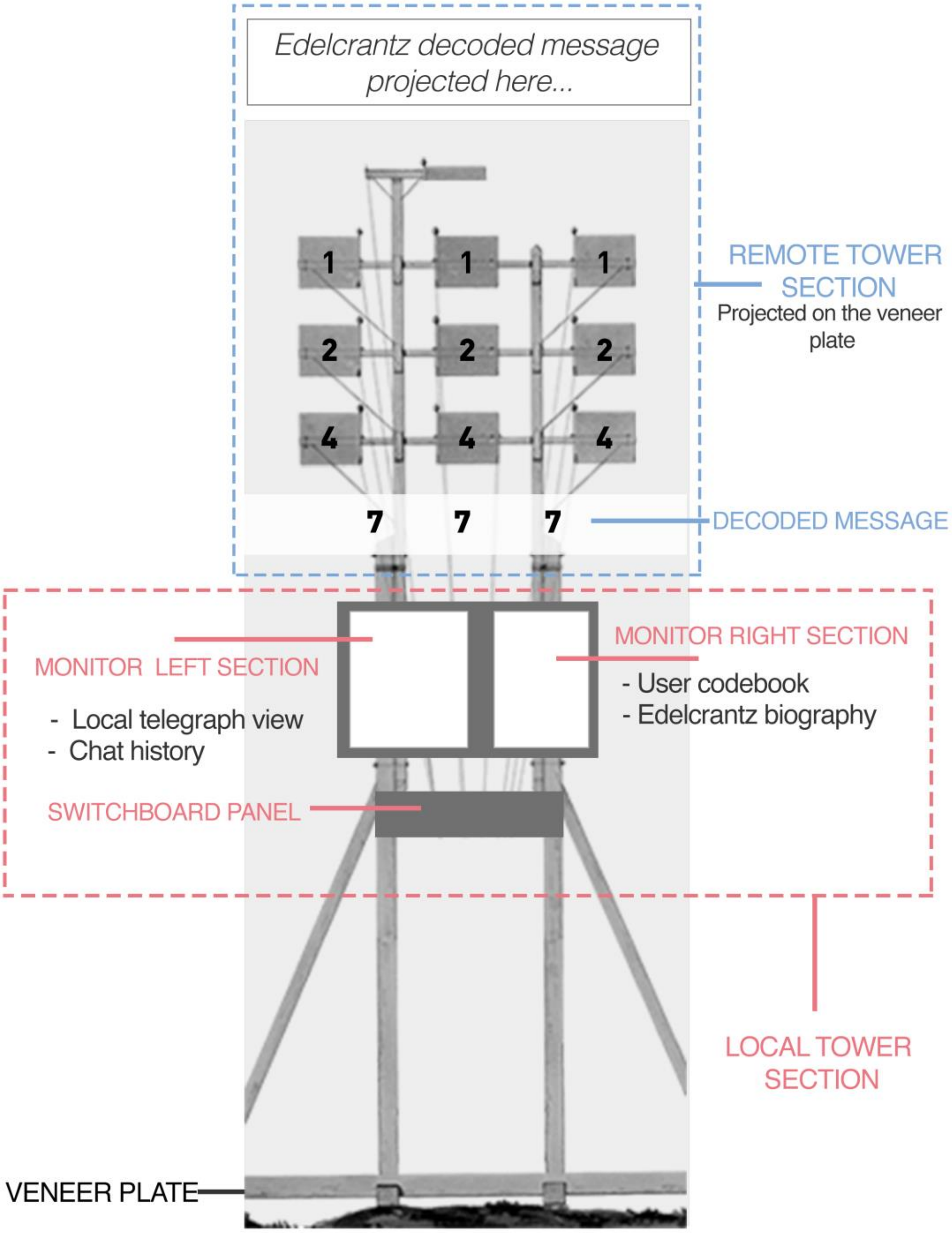


Figure 8. *Schema of the installation.*

5.2. Design and views

As previously explained, the exhibit representation of the optical tower serves the purpose of two different towers or views: the remote tower and the local tower. The remote tower view is projected on the top section of the veneer plate while the local tower view is shown in the built-in screen (placed at a height of 1.80 m) in the installation.

5.2.1. Remote tower view

The remote tower view is projected on top of the veneer pane. Users do not control the remote tower; instead, it shows the messages as they are sent by the entity on the other side of the communication. This entity is a virtual representation of Edelcrantz, the inventor of the optical telegraph.

The view runs on an Internet browser's window and it comprises three main elements: the tower shutters, the tower code and the decoded message panel, as shown in Figure 9. There are ten shutters in the remote tower view, which turn based on the message codes sent by the chatbot. Additionally, these shutters are augmented with a numerical value – used in the codification – with the purpose of aiding users to understand the codification concept.

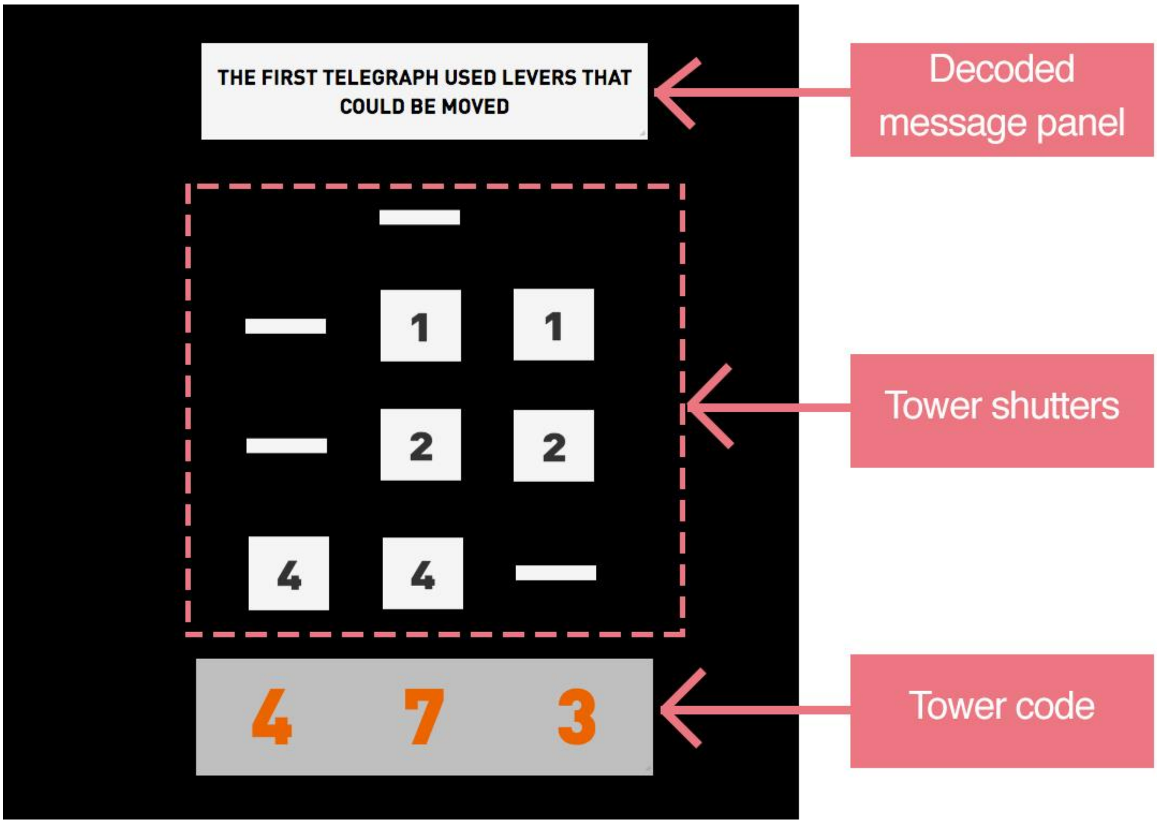

THE FIRST TELEGRAPH USED LEVERS THAT COULD BE MOVED

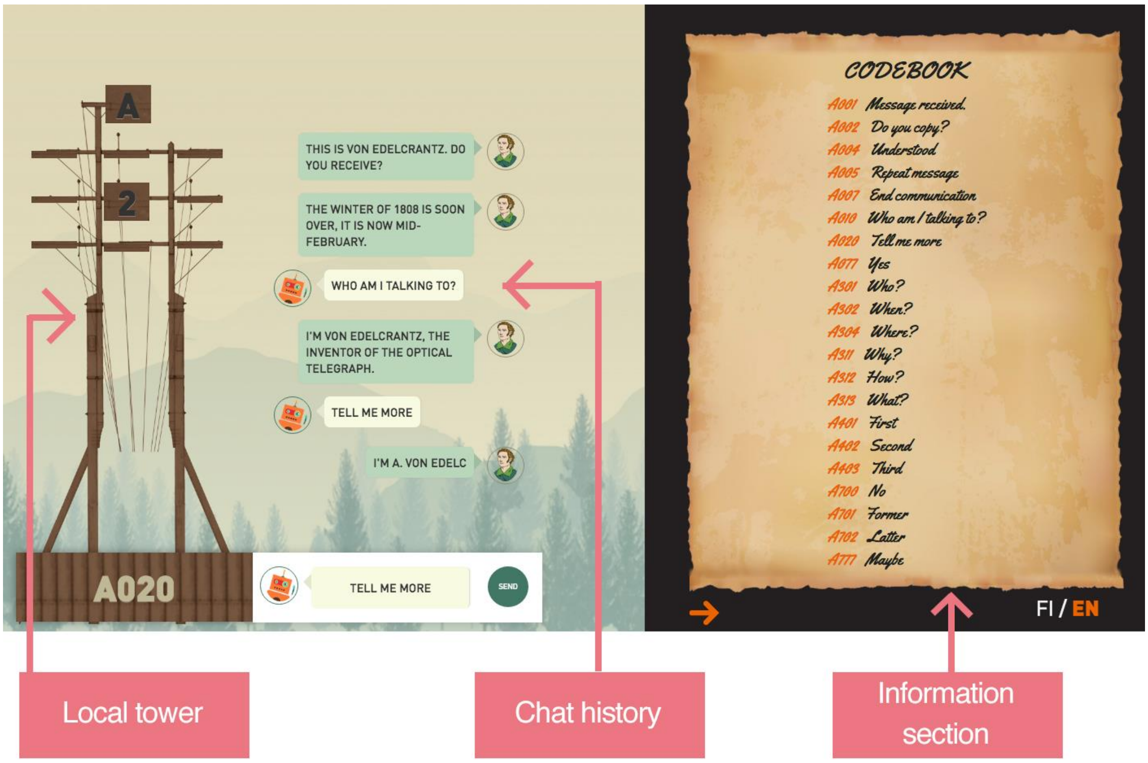
Figure 9. *Elements in the remote tower view.*

Moreover, the codes being sent at a certain moment appear at the bottom of this view. Finally, the decoded message is shown in a panel on top of the view (tower code), the message is formed dynamically based on the message codes coming from the chatbot.

The shutters and panels are presented as simple white and grey shapes on a black background. The reason for such simplified view and its color contrast is that it can be projected against the veneer plate to achieve an augmented version of the tower, as shown in Figure 13.

5.2.2. Local tower view

The local tower view comprises several elements such as a virtual local tower, a chat history and an information section, as seen in Figure 10. This view also runs on an Internet browser window. However, the veneer plate was cut out in such a form that the built-in screen appears to the user as two separate screens, as shown in Figure 13.



THIS IS VON EDELCRANTZ. DO YOU RECEIVE?

THE WINTER OF 1808 IS SOON OVER, IT IS NOW MID-FEBRUARY.

WHO AM I TALKING TO?

I'M VON EDELCRANTZ, THE INVENTOR OF THE OPTICAL TELEGRAPH.

TELL ME MORE

I'M A. VON EDELC

TELL ME MORE

CODEBOOK

A001	Message received.
A002	Do you copy?
A004	Understood
A005	Repeat message
A007	End communication
A010	Who am I talking to?
A020	Tell me more
A077	Yes
A301	Who?
A302	When?
A304	Where?
A311	Why?
A312	How?
A313	What?
A401	First
A402	Second
A403	Third
A700	No
A701	Former
A702	Latter
A777	Maybe

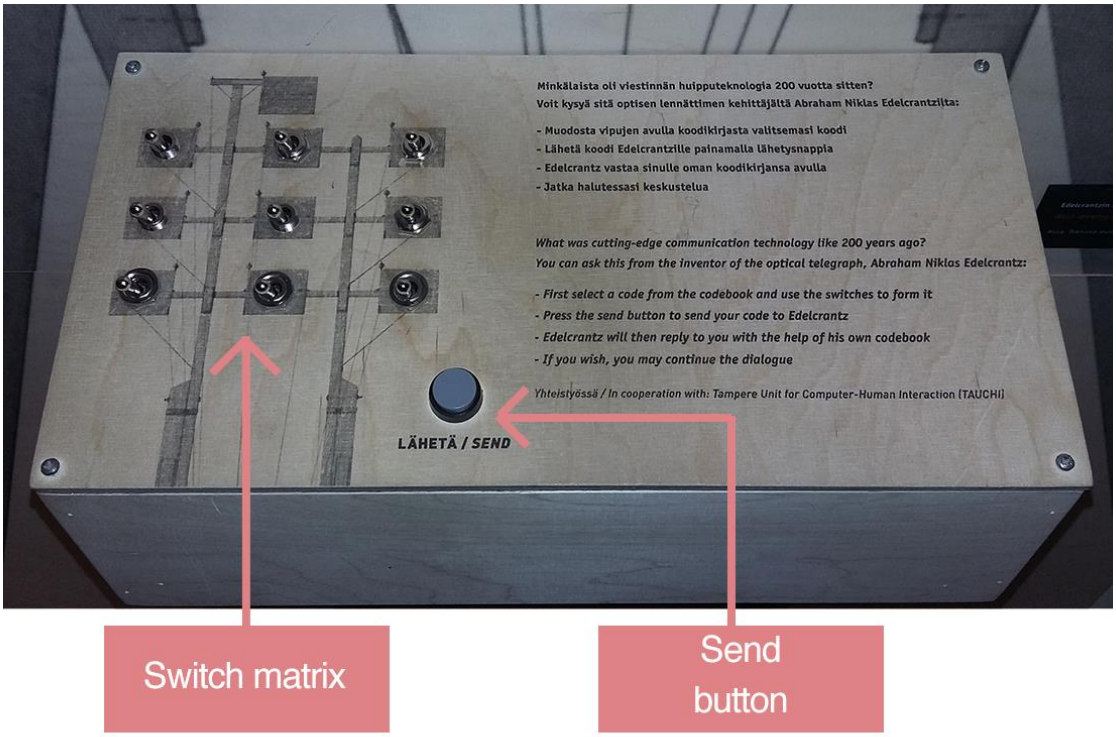
Figure 10. *Elements in local tower view.*

The digital representation of the optical telegraph tower has ten different shutters. Users can control the state of nine of these shutters by toggling the switches in the switchboard panel (Figure 11). The chat history is located to the right of the virtual tower; it shows the latest messages received from the remote tower and those sent by the user.

Another component of the local tower view is the information section. This section appears in the smaller, rightmost screen of the installation (see Figure 13). The information section contains the user codebook, language options and a brief biography of the telegraph's inventor.

5.2.3. Interaction input

User interaction is registered via switches and buttons placed in the installation. A panel with nine switches and a send button allows users to change the state of the local tower's shutters and send messages, respectively.



Minkälaista oli viestinnän huipputeknologia 200 vuotta sitten?
Voit kysyä sitä optisen lennättimen kehittäjältä Abraham Niklas Edelcrantzilta:

- **Muodosta vipujen avulla koodikirjasta valitsemasi koodi**
- **Lähetä koodi Edelcrantzille painamalla lähetysnappia**
- **Edelcrantz vastaa sinulle oman koodikirjansa avulla**
- **Jatka halutessasi keskustelua**

What was cutting-edge communication technology like 200 years ago?
You can ask this from the inventor of the optical telegraph, Abraham Niklas Edelcrantz:

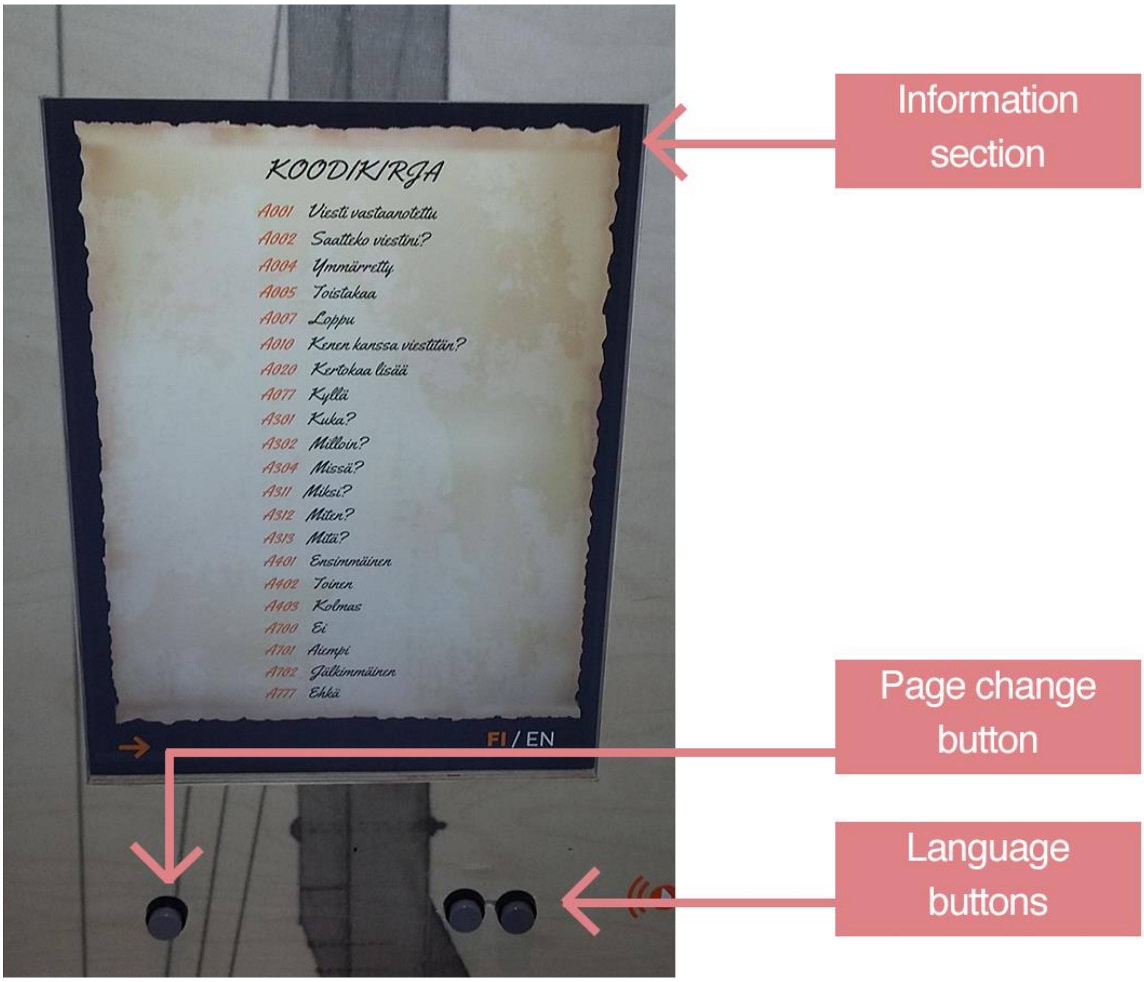
- *First select a code from the codebook and use the switches to form it*
- *Press the send button to send your code to Edelcrantz*
- *Edelcrantz will then reply to you with the help of his own codebook*
- *If you wish, you may continue the dialogue*

Yhteistyössä / In cooperation with: Tampere Unit for Computer-Human Interaction (TAUCHI)

Figure 11. *Switchboard overview.*

Nine switches in the switchboard panel are located as a 3x3 matrix, as seen in Figure 11. Following the concept of *natural mapping* [Norman, 2013], the switch positions within the panel correspond to the shutter positions in the local tower's digital representation (Figure 10). Hence, facilitating the understanding of the strategy behind code generation. A switch can have two states ON or OFF. When the switch state is ON its corresponding shutter is visible. Conversely, when a switch is OFF the corresponding shutter will be invisible. These switch actions also represent a comprehensive mapping to the objects actions, moving a switch up (ON) makes a shutter visible while moving the switch down (OFF) hides the shutter. Users can send the messages using a send button also located in the switchboard.

Besides the input devices described previously, Figure 12 shows three buttons located directly below the information section's screen. Two of these buttons are used to set the system's language to Finnish or English. The other button changes between the two information pages, namely the biography and codebook pages. The biography page contains a biography of Abraham Edelcrantz and it is the default page while no interaction occurs. The codebook page shows a list of the messages that users can send along with their respective codes and it is automatically displayed when a user starts interacting with the switches.



KOODIKIRJA

A001	Viesti vastaanotettu
A002	Saatteko viestini?
A004	Ymmärretty
A005	Toistakaa
A007	Loppu
A010	Kenen kanssa viestitän?
A020	Kertokaa lisää
A077	Kyllä
A301	Kuka?
A302	Milloin?
A304	Missä?
A311	Miksi?
A312	Miten?
A313	Mitä?
A401	Ensimmäinen
A402	Toinen
A403	Kolmas
A700	Ei
A701	Aiempi
A702	Jälkimmäinen
A777	Ehkä

Figure 12. *Extra buttons overview.*

Figure 13. *Exhibit's final installation. The right side of the interactive setup displays historical information about the optical telegraph.*

5.3. System architecture

The system comprises three main components:

1. The core module: A Node.js, HTTP server that hosts the web application. This module handles the switch interaction from users and the communication with the chatbot module.
2. The chatbot module: A Node.js module that synchronizes the communication between the chatbot service and the core module.

3. The chatbot service: A Python-based artificially-intelligent agent that generates language responses according to the messages received from the user. In addition, the service invites museumgoers to interact with the system by constantly greeting them while there is no one using the exhibition.

The relationship between the different architecture components can be seen in Figure 14.

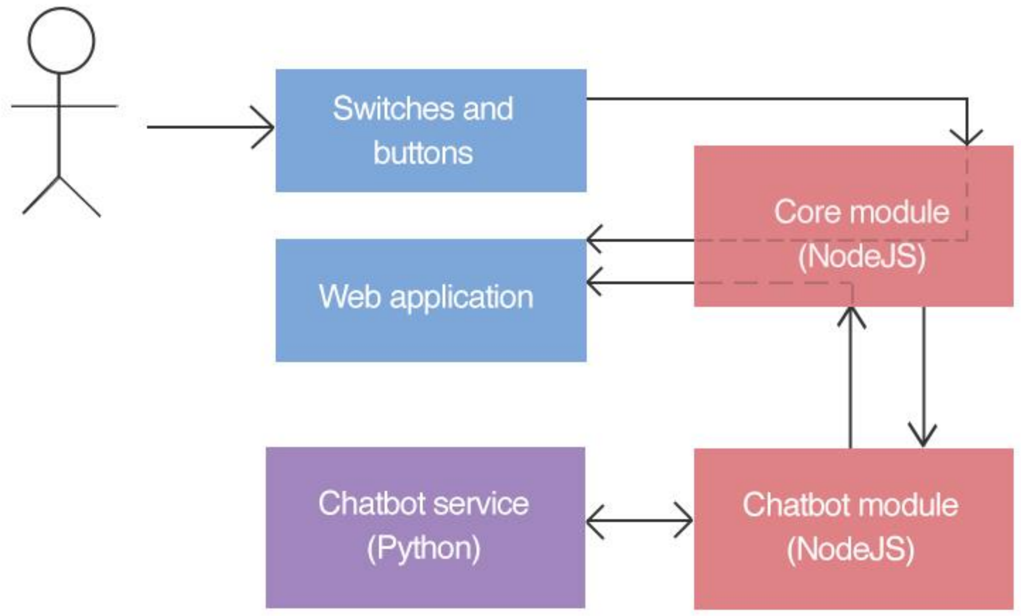


Figure 14. *System architecture diagram.*

This chapter presented a detailed description of the final exhibit installation as implemented and delivered for the Rupriikki 2.0. inauguration. Firstly, it provided an overview of the installation. Secondly, it presented a detailed description of the different tower views and the interaction input used in the installation. The chapter concluded with an outline of the system architecture and how the different components communicate with each other. The results of this thesis work are based on the evaluation of the final installation described on this chapter, more information regarding the evaluation can be found on Chapter 6.

6. Data collection and evaluation

The evaluation of the final exhibit was conducted in a period of eight months, the evaluation consisted of three evaluation days and utilized qualitative and quantitative methods to assess different aspects of the exhibit. This chapter describes the evaluation process of the final exhibit and provides a detailed explanation of the methodology used for collecting and analyzing data. The chapter begins with an overview of the evaluation period, illustrating a timeline showing when the evaluation days occurred and when the log data was collected. Subsequently, this chapter describes the context in which the observation sessions were held and the data collected. Later, it specifies the survey questionnaire and metrics used to assess it, followed by the interview structure and methods used to analyze the interview data. The chapter concludes with a detailed explanation of the data collected as log entries, the metrics derived from such data and the process to validate and perform statistical analyses on the log file data.

6.1. Evaluation overview

The evaluation of the final exhibit was conducted over a period of approximately eight months (May 2016 – Dec 2016). Throughout this time, a total of three evaluation sessions were conducted, each session used three data collections methods: surveys, observations and interviews. Furthermore, the evaluation utilized automated log files that were collected along five months (August 2016 - December 2016) with the intention of facilitating a longitudinal analysis of the exhibit. Figure 15 depicts a timeline of the evaluation phase, specifying the dates of the evaluation days and the five-month period of automated log data utilized in this study.

6.2. Observation

Observations are widely used in museum exhibit evaluations as they provide a good source of information regarding users' interactions (Section 3.2.2). In this study, we conducted three observation sessions as part of the evaluation days (Figure 15). The first observation session was held during the annual Tampere museum night (May 2016), a special event when the museum receives a high number of visitors. Two more observations were conducted over a weekend in October 2016. The length of such sessions varied from 4-6 hours (Table 3), summing up to approximately 14 hours of observations. It is worth noting that all observation sessions were unobtrusive, thus, visitors did not know they were being observed because the observers sat within 1-3 meters away from them. The purpose of keeping users unaware of the evaluation was to prevent them from interacting longer than they normally would [Block et al., 2015]. Hence, visitors were only approached for further questions after they had finished interacting and left the installation.

Date	Duration	Number of observers
May 16th, 2016	~ 6 hours	1
Oct 21st, 2016	~ 4 hours	2
Oct 22nd, 2016	~ 4 hours	2

Table 3. *Observation sessions details.*

The objectives of the observation were to obtain information about the visitors, identify interaction patterns, understand possible usability issues as well as providing further context and validation to the log, survey and interview data. An observation form (Appendix 4) was used to annotate several details about users such as their age group, gender, group type (family or non-family groups), among other information. Additionally, some aspects concerning the interaction with the exhibit were also recorded, such as user's spontaneous comments, their attitude while interacting and their possible understanding of the exhibit. Since the observation was intended to be unobtrusive, the age groups and gender were subjectively collected based on the perceptions of the observers. Nonetheless, this information was also enquired from those users who filled in the survey questionnaire.

In order to assess the exhibit's support for group interactions and collaboration, observers annotated the number of *active* and *passive* users on each group [Hornecker and Stifter, 2006a]. Users were categorized as *active* when they directly manipulated the physical elements (switches and buttons) in the installation. On the other hand, users that remained as observers only, commenting or collaborating without touching the switches were considered *passive*. Additionally, during the data analysis, group sessions where all members in the group were *active* users were categorized as *fully active groups*.

Another important goal of the observation sessions was to measure the *stopping power* of the exhibit (Section 3.3.1) and how noticeable it was to visitors passing by. For that purpose, the weekend observation (Oct 22 – Oct 23, 2016) was conducted by two observers. One of the observers was solely in charge of tallying every person who passed in front of the installation, every person who noticed the installation and every person who started interacting with the system – at least one button press or switch move were considered as interaction.

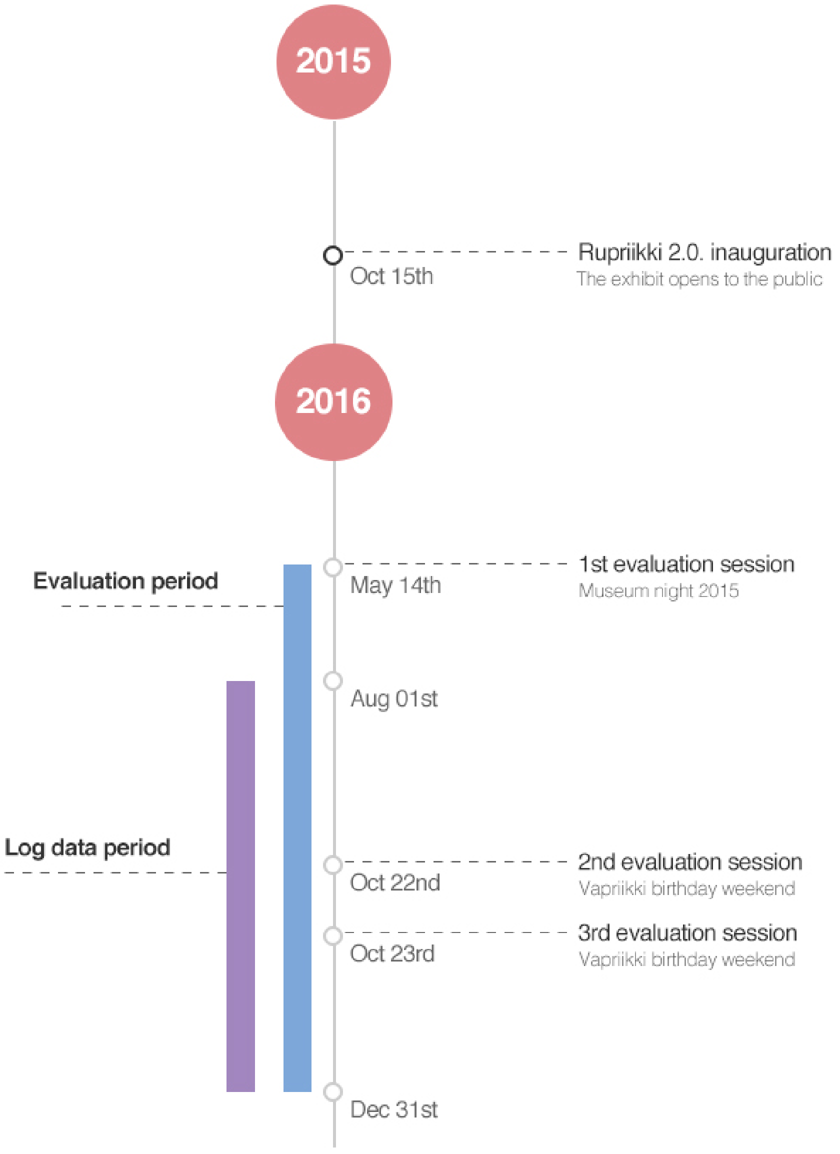


Figure 15. *Evaluation timeline.*

Finally, a key motivation for the observation sessions was to allow the validation of the log file analysis, which would enable a longitudinal analysis on the exhibit. For this reason, the observers carefully annotated the start and end times of each interaction session. This information was crucial when designing the session separation algorithm as it helped identify a constant time gap value between sessions (Section 6.5.1). Moreover, these notes were used as a “source of truth” to validate the results yielded by the session separation algorithm.

6.3. Survey

After users interacted with the exhibit, those who used it for periods longer than one minute were asked to fill in a survey questionnaire (Appendix 3). This survey was conducted only during the evaluation days, the reason for this decision was that it allowed mapping between the answers given in the survey questionnaire and the behaviors observed during interaction. Additionally, this provided a source of reliability that all users filling in the survey had indeed interacted with the exhibit.

The survey questionnaire was available in two languages (Finnish and English), it comprised a series of statements regarding the experience of using the telegraph as well as some other background information concerning the users’ interests (Table 4). Each statement was assessed using a five-point Likert scale (1= Totally disagree, 3= Neither agree nor disagree and 5 = Totally agree). Finally, the overall impression of the system was rated using a smiley scale (Figure 16) – a scale commonly used in children studies [Davies and Brember, 1994] and to assess customer satisfaction [Pousette et al., 2014]. A total of 54 visitors responded the questionnaire.

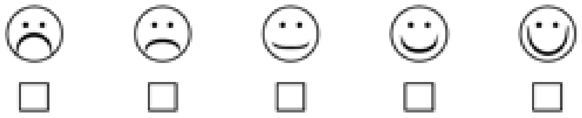


Figure 16. *Smiley scale*

6.4. Interviews

During the evaluation days (Figure 15), a semi-structured interview was conducted. The interview was performed after visitors used the system and filled out the survey form. The aim was to gather the visitors’ impressions of the telegraph installation, evaluate the quality of communication with Edelcrantz’s chatbot, assess the difficulty level of the coding principle and obtain suggestions for improvement. Table 5 shows a list of the structured interview questions, in some cases users were enquired further to clarify or obtain more information about their previous answers.

Summary	Full question
First impression	The first impression about the telegraph was interesting.
Clarity of operation	The idea of the telegraph was clear.
Irritability	The telegraph was irritating.
Fun	Communicating with the telegraph was fun.
Willingness to repeat use	I would like to communicate with the telegraph again.
Ease of communication	Sending the messages was easy.
Coding principle difficulty	Figuring out the coding principle was difficult.
Telegraph's operation learning	The application helped me to understand the principle of the optical telegraph.
Telegraph's history learning	The application taught me about the telegraph's history.
Appeal to interactive installations	These kinds of applications would increase my interest towards museum visits.
Recommendation	I would like to tell my friend about the telegraph.
Social media	I would like to share my coding experience with my friends in social media (e.g., Facebook, Twitter or Instagram).
User's technical interests	I am interested in technique.
User's gaming interests	I am interested in playing.
User's problem-solving interests	I am interested in problem-solving tasks.
User's history interests	I am interested in history.
Overall score	How much did you like the telegraph installation as a whole?

Table 4. *Survey statements.*

A total of 34 interviews were held during three evaluation days. In the case of multiple users, a single interview was conducted with all members of the group together. All interview answers were analyzed using content analysis. The answers were coded and grouped into different categories using emergent coding [Lazar et al., 2014]. Subsequently, a quantitative approach was used to count the number of instances of a certain category. Thus, making it possible to calculate the percentage of respondents that gave a certain answer (Section 7.3).

Summary	Full question
Best part of telegraph	What was the best part of the telegraph? Why?
Worst part of telegraph	What was the worst part of the telegraph installation? Why?
Understanding of telegraph operation	Did you understand the idea of how the telegraph operates?
Coding-principle difficulty	Was it easy/difficult to solve the coding principle?
Chatbot communication quality	Did you get the answers you were expecting from Edelcrantz? / If not, why?
Pleasant/ unpleasant	Did you find anything pleasantly or unpleasantly surprising while interacting with the application?
Nice/ fun/ annoying/ difficult	Was there something especially nice/fun/annoying/difficult?
Suggested changes	Would you make any changes to the application?
More comments	Any other comments or suggestions you would like to add?

Table 5. *Interview questions.*

6.5. Log data

Log files were used to store the interaction information from the installation for several months (Figure 15). The analysis conducted for this thesis work uses the data from five of these months (August 2016 - December, 2016). The dates when the museum was closed such as Mondays and holidays were excluded from the data analysis. Furthermore, some additional days where the system presented technical issues were also omitted. Hence, over the five months a total of 123 days was analyzed.

The log files contain log entries for every user interaction with the exhibit such as switch moves, button presses and the messages they sent; as well as the chatbot responses. All log entries contain a timestamp. Additionally, information regarding the user presence was also logged, more specifically when a user entered or left the switch panel area. The user presence information was recorded using an ultrasonic proximity sensor that detected when users stood in front of the switch panel at a distance of one meter or less.

6.5.1. Session separation algorithm

To obtain basic metrics such as the number of sessions and dwell times, it was necessary to separate and analyze interaction sessions individually. To discern between two different sessions, it was assumed that there is a certain time gap where there are no user input interactions between one session and another. Such strategy was previously used in public display studies [Marshall et al., 2011; Peltonen et al., 2008]. Though the principle behind the session separation is the same, the public display studies used videos to manually identify different user sessions while in this work the process was automated using a session separation algorithm.

Since the session separation was based on a time gap value, it was key to identify a constant that would yield the correct number of sessions. Therefore, we extracted the value from the manual annotations performed during the observation sessions. As the observation sessions were held on different dates and during different times of the days, it was assumed that this value could be generalizable for a longitudinal analysis of the optical telegraph exhibit. The constant value used to define the time gap or idle time between two sessions was 35 seconds.

The session separation algorithm works as follows:

1. Filters the log events to include only user input interactions, in other words, only switch moves and button presses.
2. For every user input interaction, checks the time gap between the current input interaction and the next one.
3. If there is a time gap of less than 35 seconds, adds both input interactions to the same user session. Otherwise, creates a new session, where the latter input interaction is the first user input of the new session.

Two different methods were used to validate the accuracy of the algorithm. The first method compared the results of the session separation algorithm with a human annotator. For this end, two log files from random dates were automatically annotated and manually annotated, the results from both sources were compared to confirm that the algorithm would yield the same results as those obtained when the log data was manually annotated by an individual. The second method was a three-way comparison between the human (manual) annotation, the automatic annotation and the notes taken during the observation. This three-way comparison was performed on a log file from one of the evaluation days (Oct 23rd).

6.5.2. Metrics extracted

Once sessions were successfully separated, additional information could be extracted from each of them to acquire further understanding of the session interactions. Table 6 shows the different metrics derived from each of the sessions.

Metric	Description
Session dwell time	Time between the first and last interaction with the physical input (switches/ buttons)
Total switch moves	Number of times users moved the switches during one session.
Total *different* codes formed	Number of *unique* codes users formed during one session (repeated codes were omitted).
Total messages sent	Number of messages sent within one session.
Total "send" attempts	Number of times a user pressed the send button without having formed a code.

Table 6. *Metrics derived from automatically separated sessions.*

Finally, using all the session details previously described, a daily average of each variable was calculated. Consequently, a list of all daily summaries was stored on a CSV file for further statistical analysis using MS Excel and SPSS.

6.5.3. Session classification

Existing literature suggests that open-ended exhibits – where visitors can freely perform different tasks or they can perform one task in several ways – hold visitors' attention for a longer time [Sandifer, 2003]. However, this open-endedness cannot be immediately appreciated by users and they can only take advantage of this quality when they interact for longer periods [Sandifer, 2003]. Taking this into consideration, it was important to analyze the visitor behavior in short and long sessions separately. Therefore, sessions were classified in two different categories: *engaging* and *unengaging*.

Sessions were categorized as *engaging* if the user(s) had managed to form at least three *different* message codes during the session, while sessions that did not meet this criterion were categorized as *unengaging*. This criterion was based on the assumption that users who formed at least three *different* messages had very likely understood the coding principle and therefore were more engaged with the exhibit, whereas users who did not manage to construct enough codes would be easily bored and leave the exhibit.

It is relevant to emphasize that only *different* messages formed were considered in this analysis, in other words, when users constructed the same code several times it was counted only the first time it was constructed. For instance, if a user formed the codes: “A777”, “A001”, “A777” only two *different* messages would be counted. The intention behind such restriction was to avoid false positives, namely, sessions categorized as engaging in cases where users “luckily” constructed a single message code several times. In this way, engaging sessions could also be utilized to measure the exhibit’s communication power (Section 3.3.4), as we consider it unlikely to construct three *different* codes by chance.

This chapter explained the methodology used for the evaluation of the Vapriikki optical telegraph. There were three evaluation days, each day using three qualitative methods: unobtrusive observations, surveys and interviews. These ethnographically-oriented methods were used to gather information while users interacted with the system and to understand user impressions of the system after their usage. Furthermore, five months of automated log files were utilized to perform a quantitative, longitudinal analysis of the exhibit. The methodology used in this study investigates the use of a semi-automated log data analysis combined with qualitative data to obtain meaningful information of user interactions with the exhibit, which could be used to guide the design of other similar exhibits. The results of the data analysis are presented in the following chapter.

7. Evaluation results

This chapter presents the results obtained from the evaluation (Figure 15) and analysis described on Chapter 6. Firstly, it presents statistics derived from the observation sessions such as the exhibit's stopping power, users' age and gender distribution, as well as some qualitative aspects of the individual interactions and group collaborations. Secondly, it summarizes the survey results with emphasis on users' interests and impressions of the system. Thirdly, the chapter shows interview findings regarding users' opinions after using the telegraph and their recommendations for improvements. Lastly, the chapter concludes by reviewing the results from the log data analysis such as the validity of the automated session separation and session details extracted from the analysis.

7.1. Observation results

As described in Section 6.2, three observation sessions were conducted – in May and October 2016 – to gather information about visitors' behavior and experiences with the telegraph installation. This section presents the results from those observations, starting with the exhibit's stopping power, followed by a description of the users' gender and age distributions, and the user groups classification (single users, family groups, non-family groups). Finally, this section highlights some interesting user interactions and establishes a comparison of the interaction and collaboration across different group types (family groups vs. non-family groups).

7.1.1. Stopping power

During two of the evaluation days (Oct 21st and Oct 22nd, 2016), museumgoers who passed and interacted with the installation were observed and tallied. A total of 267 passersby were tallied amid these two evaluation days, of which 49% noticed the telegraph and 27% started interacting with the telegraph installation.

7.1.2. Gender and age distributions

In all three observation sessions, an observer noted down the gender and age groups of all users that interacted with the system. From a total of 178 observed users, 44% were male and 56% female. The age distribution of 176 of the users is depicted in Figure 17, the largest number of users corresponds to adults with ages ranging from 19-29 (29%) and 30-59 (29%), respectively. Followed by teenagers from 12-18 years old (21%) and children under 12 years old (20%). Finally, only one percent (1%) of the users were 60 years and over.

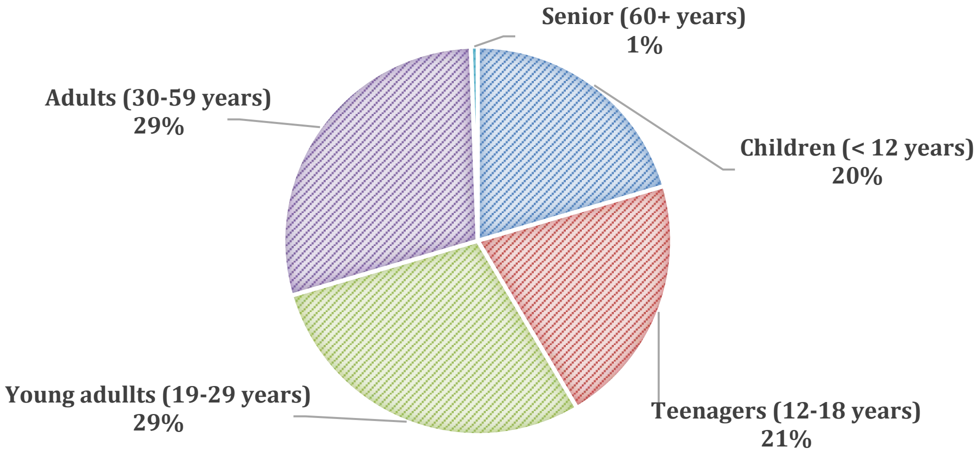


Figure 17. *Users' age distribution.*

7.1.3. User groups classification

The observers classified users into three different groups: single users, family groups or non-family groups. Groups were categorized as a family when there was at least one child or teenager and one parent or adult with them. Non-family groups were comprised by people in roughly the same age groups. These groups included couples, groups of children and groups of adults, also referred as peer groups. As can be seen in Figure 18, a clear majority of users interacted with the system in groups, either in family groups (36%) or non-family groups (36%) while a smaller percentage interacted as singletons (28%).

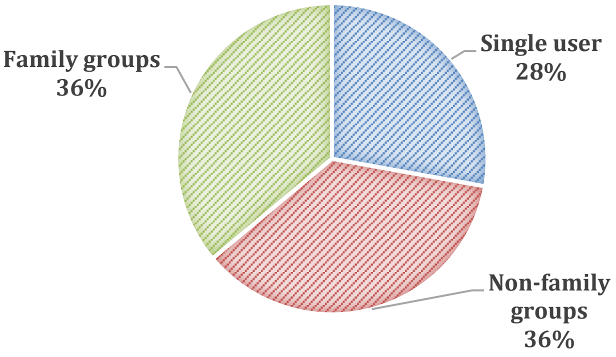


Figure 18. *User groups classification.*

7.1.4. User interactions

As part of the observation sessions, observers took notes of any interesting behaviors, spontaneous comments, or happenings occurring while users interacted with the exhibit. Some highlights from these interactions are presented below:

Positive reactions: Users repeatedly smiled while interacting with the exhibit, chuckling and showing signs of enjoyment when communicating with the chatbot. In two opportunities, users took pictures of their conversations on the screen before leaving the exhibit. Additionally, some users came back after some time to interact once more.

Interaction with tangibles: Children showed especial excitement with the switches, running towards them and moving them eagerly. Similarly, users would press the send button many times during their interactions.

Screen touching: Many users, most of them children, tried to press some elements of the screen which represented the physical buttons in the installation (see Figure 11 and Figure 12). After one or more attempts, they realized that they needed to press one of the physical buttons instead.

Chatbot: On several occasions, the chatbot did not understand or gave incoherent answers to users. This situation was observed mostly when the user had started interaction right after someone else left the exhibit. A couple of users reacted in different ways, one of them repeating the message *"I don't understand"*, another one verbalizing *"What don't you understand?"*.

Interaction with remote tower: The remote tower projection was hardly noticed by the users once they were interacting with the system. A couple of times, users looked up after their first switch moves to check if it moved the plates in the projection, one user commented: *"Ah this comes here!"* (while pointing at the remote tower message and the chatbot message). Additionally, the projection caught the attention of passersby as this was the first element they would notice while walking around the exhibition.

7.1.5. Group interactions and collaboration

As previously stated, over 72% of the users interacted with the system as part of a group. To measure the level of participation and collaboration in groups, the number of *active* and *passive* users on each group were tallied during observations. Furthermore, during the analysis, groups where all users were active at some point of the session were categorized as *fully active groups* (Section 6.2). Relevant findings regarding the participation and collaboration in groups can be summarized in the following points:

Participation (active users vs. passive users): A record of the activity of 150 users who interacted in groups reveals that 71% of the users manipulated the physical elements (switches and buttons) at some point of their session while 29% of them remained as passive users without taking control of the physical elements. Moreover, interaction activity data from family groups (23 groups) and non-family groups (26 groups) were used to categorize *fully active groups* – groups where all members interacted as *active* users. The categorization shows that 22% of family-groups were *fully active* while 54% of non-family groups were *fully active.*

In general, groups of two people were commonly seen moving the switches. Groups of three, four and up to five people were observed interacting. However, in such cases only a subset of the group (two to three of the users) remained active.

Collaboration in non-family groups: The most common form of collaboration between two or more active users in non-family groups was switching places and taking turns to use the switches and send messages. Nevertheless, there were also occasions where two users simultaneously moved the switches while interacting. Moreover, collaboration among group members was observed in other ways, for example, one person deciding which messages to send while another person interacted with the switches. Furthermore, users would often have discussions to try to understand the coding principle and how to use the installation in general.

Collaboration in family groups: In family groups parents mainly remained as passive users (78% of the sessions) and left the operation with the switches to the children. In the rest of the cases, where parents and children were active users, there were different ways of collaboration. Parents and children moved switches at the same time or the child pressed the send button and the parent moved the switches. Most of the times, the interaction also evoked some form of conversation or explanation between parents and children, where parents acted as guides [Lee et al., 2015].

7.2. Survey results

As part of the evaluation, a post-usage survey was conducted. The survey statements focused on gathering information about users' interests and their impressions of the system (Table 4). In the three evaluation days, a total of 54 users filled in the survey questionnaire. This section presents the results of the survey data analysis.

7.2.1. Users' impressions

Users' responses regarding their impressions and experiences with the system are presented as a boxplot graph on Figure 19. The figure shows the median and distribution of data for each of the answers. Users' impressions of the system can be summarized as follows:

Good impression of the system: Users had a good impression of the system, which is evidenced in the high general score of the installation (median = 4). Additionally, the respondents expressed that the communication was fun (median = 4.5) and showed disagreement when asked if the telegraph was annoying (median = 2).

"Right" level of difficulty: Respondents found the idea behind the telegraph clear (median = 4). Moreover, users' ratings for the level of difficulty ranged in the middle, as can be seen in the answers regarding message emission difficulty (median = 3) and the code-solving difficulty (median = 2).

Worthy of recommendation: Participants stated that they would wish to communicate with the telegraph again (median = 4) and that they would recommend the installation to friends (median = 4).

An aid for learning: Furthermore, they considered that the application helped them understand the telegraph's operation (median = 4) and history (median = 4). Additionally, visitors strongly agreed with the statement that installations such as the telegraph raise their interest in museum visits (median = 5).

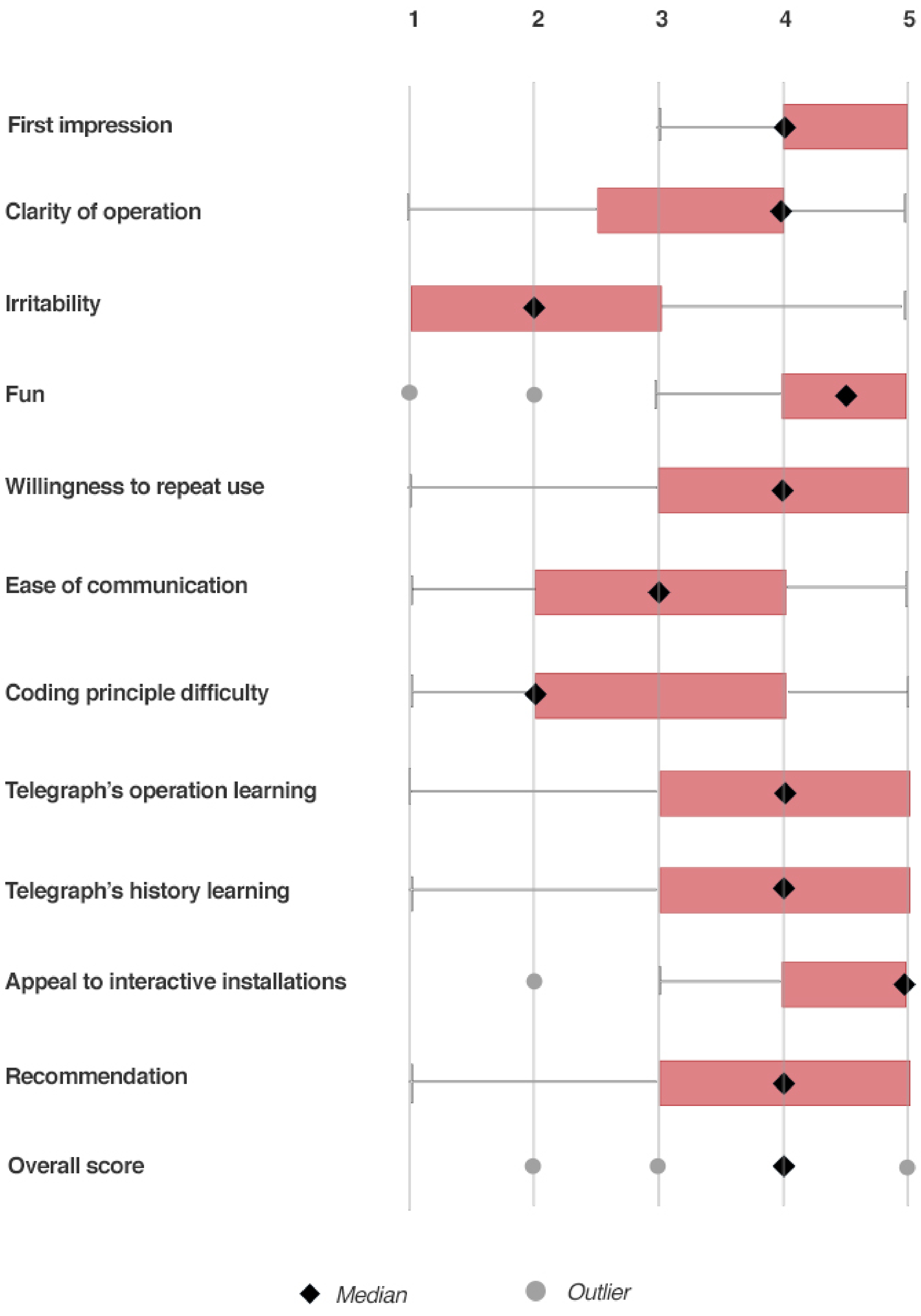


Figure 19. *Survey responses (Users' impressions and experiences).*

Furthermore, correlation analyses using Spearman's Rho revealed various statistically significant relationships between some of the responses to the survey. A moderate relation was found between the clarity of the telegraph idea and the overall score ($r = .514$ $p = .000$). Similarly, the fun factor had a slight influence on the final rating of the system ($r = .372$ $p = .006$) and a moderate influence in the visitors' willingness to use the installation again ($r = .492$ $p = .000$).

7.2.2. Users' interests

Participants were enquired about their personal interests as part of the survey, their answers are depicted as a boxplot graph on Figure 20 which shows the median and distribution of data. Using Spearman's Rho correlation, the relationship between users' interests and their impression of the system were investigated. The results show that two interest profiles may influence users' overall rating of the system: users with technical interests ($r = .436$ $p = .001$) and users with problem-solving interests ($r = .471$ $p = .000$). Conversely, there were no significant results in the relation between the general score of the installation and users' history and gaming interests.

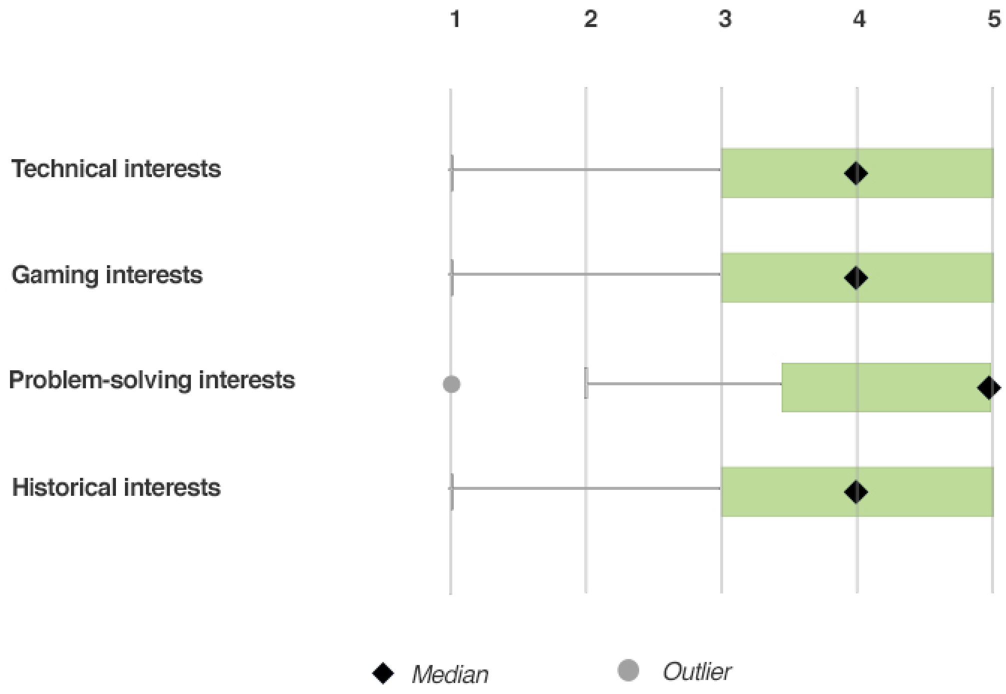


Figure 20. *Survey responses (Users' interests).*

7.2.3. Main drive for interaction

Users also stated their main drive for interaction with the system. As can be seen in Figure 21, 41% of the participants selected ‘solving the code’ as their main interest for interaction, 30% were interested both in communicating with the chatbot and solving the code while 22% preferred only the communication aspect with the chatbot. One of the respondents reported ‘other’ interests in the system, namely, tinkering with the switches.

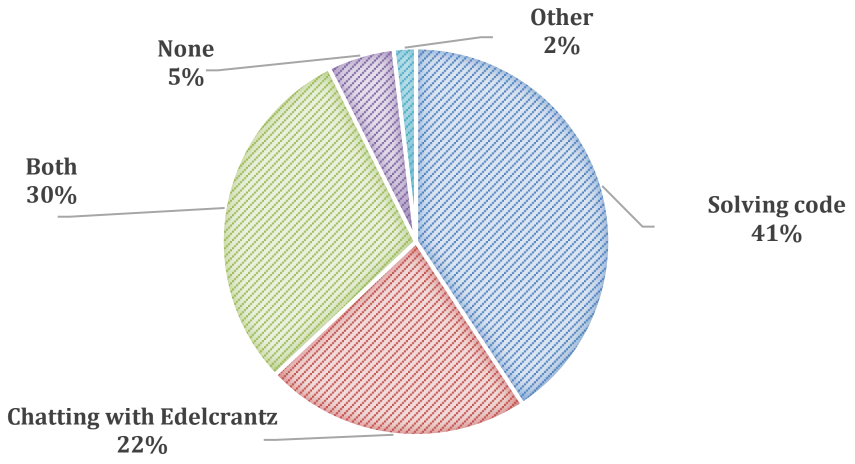


Figure 21. *Main interest during interaction.*

7.3. Interview results

As part of the three observation sessions, users were prompted to answer a post-usage interview about the telegraph installation (Section 6.4). In total, 34 groups and single users agreed to participate in the interview. The notes from these interviews reveal that users enjoyed interacting with the installation and that the majority understood the logic behind Edelcrantz’s codification system and how the telegraph worked. Different findings from the interview answers can be summarized as follows:

Best part of the telegraph: Visitors expressed that the best part of the telegraph was solving the coding principle as conveyed in one of the users’ responses, *"of course cracking how it works and that you feel clever".* Closely following was the communication factor, one of the respondents considered the best part to be *“the feeling that somebody is answering, the communication”*. Other respondents pointed out the historical factor and the system operation as their favorite part while others noted the interactivity, collaboration, the interesting and fun factors as well as its uniqueness. Based on these answers, three main user interest profiles were identified: **problem-solvers, communicators and history-enthusiasts.**

Worst part of the telegraph: When enquired about the worst part of the telegraph, the first item pointed out by users was the period before understanding how the messages could be sent (coding principle). One of the respondents mentioned *“the time before understanding how it works. It is frustrating"*. Another aspect frequently mentioned was the poor quality of the conversation with the chatbot, as stated by one of them, *“the conversation was a little bit stupid. The guy didn't answer anything other than I don't understand”*. Some other factors worth noting were the limited amount of grammar in the codebook, slow chatbot responses and the feeling of uncertainty about the other end of the conversation.

Coding principle difficulty: Over 81% of the participants claimed to have understood how the coding principle worked. Most users considered that solving the coding principle presented an easy (45%) or medium (36%) level of difficulty. In some cases, visitors showed preference for a certain challenge level *"It would be less interesting if there were clear instructions"* and *“(…) it is more fun to figure it out”*. Though most users seem to have understood the coding principle, some of them did not grasp the concept of how the telegraph was used. When enquired if they understood the telegraph operation one visitor commented: *“Not sure. I recognized how to send messages but not the whole idea”*, another interviewee responded: *“Yes, the code but not more”*.

Quality of chatbot conversations: One key goal of the interview was to obtain information regarding the quality of the conversation between visitors and the chatbot. More than a third of the respondents (38%) conveyed that they did not get the answers they were expecting from the Edelcrantz bot. There were various reasons behind the visitors’ dissatisfaction such as the bot not understanding messages, the answers being too simple, incoherent or slow.

Surprises during interaction: Users were further questioned about any pleasant and unpleasant surprises during the interaction. In terms of pleasant surprises users indicated having somebody talking back, the message projection in the veneer plate (remote tower) and the technology itself. A user exclaimed that *"it was surprising to have something like that in the 1800s!"* Regarding the unpleasant side of interaction, one user mentioned the limited codebook options, *“the only choices were questions”*. Additionally, a couple of users indicated that the chatbot did not understand their messages and that the start of interaction was unclear. Finally, another visitor expressed that the speed of Edelcrantz’s responses was slow, *“the way he talked, please faster!”*

Nice and annoying elements: Participants also mentioned some nice and annoying elements of the system. Among the nice factors were the interaction with switches, the hands-on experience, the communication, problem-solving quality and historical parts of the system. In regards to the annoying elements, visitors manifested void conversations. A visitor expressed *“the irritating part was when the conversation was not going anywhere”*. Other factors considered annoying were the limited codebook, the difficulty to see the projection in the remote tower (located too high) and the bot not being able to understand users’ questions.

Switches appeal: Playing with the switches was considered one of the fun and pleasant elements in the exhibit. When further enquired about their opinions on the interaction with the switches, one user conveyed they were *“cool, sort of old technology”*. Moreover, some users expressed their preference for the “old-fashioned”. One interviewee stated *"old-fashioned feels nice, instead of just a screen”* while another one mentioned *"I like it (…) more fun than just a touchscreen, you can really feel feedback"*.

Suggestions for improvement: 48% of the visitors responded that they would make changes to the installation. Most of these respondents were in favor of having more options in the codebook while others indicated they would prefer faster answers from Edelcrantz. Another suggestion from a couple of the participants was changing the direction of communication. *“Something that triggers interesting conversations. Instead of me asking questions. I could be answering; it would be probably easier”* was the proposition from one participant. Other – nearly marginal – suggestions include lowering the difficulty level, better understanding from the chatbot side, adding gaming elements, communicating with other humans and making it more appealing to children.

7.4. Log file results

Automated log files were collected (from August to December 2016) and analyzed as part of this study (Section 6.5). This section describes relevant findings regarding the log file analysis, it shows the results of the validation of the automated session separation and presents the results and metrics extracted from each session.

7.4.1. Automated session separation

Before addressing the results of the log file analysis, it is important to emphasize some of the findings related to the validity of the session separation results. The log file analysis started with a process to discern different user sessions automatically, an algorithm was used to separate and analyze each of the sessions. To validate the algorithm, the results from the automated analysis were compared with manual human annotations and the notes from the observation session held on one of the evaluation days (Section 6.5.1).

The results of the comparisons show that the manual annotation and automatic annotation yielded the same session-separation results in 90.70% of the sessions. However, when comparing the automatic and manual results to the observation notes, it is evidenced that neither a manual annotator nor the algorithm could successfully identify the same sessions as those found in the observation notes. More specifically, the algorithm clustered different sessions as one. This discrepancy can be attributed to the high influx of visitors during the evaluation day, which made the time gap between sessions marginal. Therefore, it can be said that the algorithm can successfully identify sessions on most of the logged days with a regular visitor influx, where the time between one visitor session and the next can be clearly recognized.

It is relevant to note that another form of validation was planned for the session-separation algorithm. An ultrasonic sensor was used to detect when there were users standing at one meter or less from the installation. However, the device was not accurate and therefore its data could not be used to complement the session-separation algorithm.

7.4.2. Session information and classification

During the five-month period (August 2016 – December 2016) analyzed for this study, a total of 2697 interactive sessions were recognized with an average of 22 sessions per day and a median of 19 sessions per day. The average dwell time of the exhibit was 1:27 minutes.

In order to analyze short and long sessions separately, they were classified into two categories: *engaging* and *unengaging* (Section 6.5.3). Sessions were considered *engaging* when users constructed three or more *different* codes. The result of the **semi-automated analysis classified 53.99% as engaging and 46.01% as unengaging** with averages of 18 and 14 sessions per day, respectively. The differences between engaging and unengaging sessions are specified in Table 7, all average values were calculated as weighted averages where the weight depends on the number of sessions per day.

	Engaging sessions	**Unengaging sessions**
Interaction time	2:21 mins	0:23 mins
***Different* codes formed**	6.72	0.96
Messages sent	7.91	1.30
Send attempts	11.16	2.62
Switch moves	60.91	11.43

Table 7. *Engaging Sessions vs. Unengaging Sessions.*

The average interaction time for engaging sessions was considerably higher than unengaging ones with 2:21 and 0:23 minutes, respectively. Engaging sessions had a higher number of codes formed, with an average of 6.72 codes per session while unengaging sessions had an average of 0.96 codes formed. An average of 7.91 messages were sent on engaging sessions whereas unengaging sessions had an average of 1.30 messages. Similarly, engaging sessions had a higher number of switch moves and failed attempts to send messages.

During the manual annotations of random log files and the semi-automated log file analysis, it was observed that some users formed many more codes than they sent. Otherwise stated, users moved the switches until successfully constructing one of the messages in the codebook but would not press the “send” button afterwards. Such behavior in users could be attributed to the fact that those users were mostly interested in the codification principle and how to create messages rather than on the communication aspect of the interaction. However, as evidenced when comparing the total number of messages sent (Table 7) and the *different* codes formed, on average, users did not show this behavior.

7.4.3. Codebook messages used

The log analysis revealed that some of the messages in the codebook were much more used than others, as seen on Figure 22. The six messages most used during communication – *'Ehkä' (Maybe)*, *'Viesti vastaanotettu' (Message received)*, *'Kenen kanssa viestitän?' (Who am I talking to?)*, *'Kertokaa lisää' (Tell me more)*, *'Ymmärtetty' (Understood)* and *'Saatteko viestini?'* (*Do you copy?)* – represent 57.4% of the messages. On the other hand, three of the messages in the codebook – *'Miten?' (How?)*, *'Aiempi' (Former)* and *'Missä?' (Where?)* – were never sent during the five-month period analyzed. The six most-used messages described previously are consistent in the interactions in both languages. However, in English all messages from the codebook but one, '*Second'*, were sent at some point.

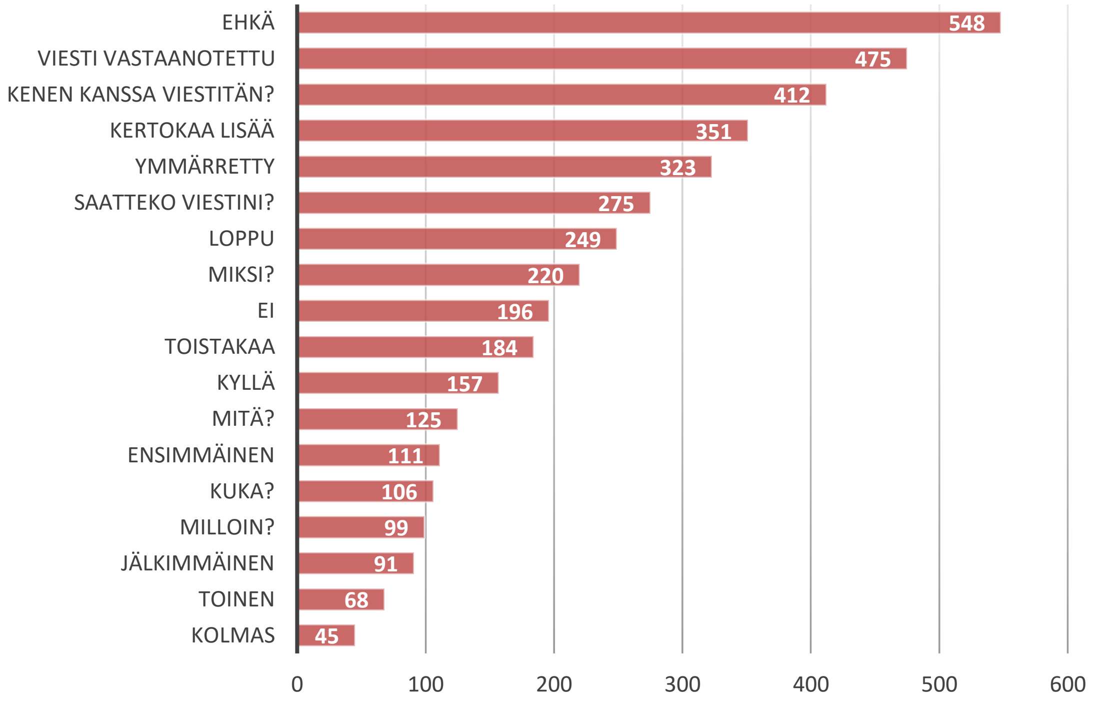


Figure 22. *Count of different messages sent (Finnish).*

This chapter reviewed the findings after the evaluation of the Vapriikki's optical telegraph exhibit. The results of three observation sessions show that most users interacted with the exhibits in groups and that a majority were adults. Observations also revealed differences between the interaction and collaboration of family and non-family groups, with non-family groups being more active in the interaction. Furthermore, survey results indicate that users liked the exhibit and that users' interests may have influenced their impressions of the installation. Interview data shows that users enjoyed the problem-solving challenge in the exhibit and that the communication was sometimes considered clumsy, additionally, interviews revealed a special appeal in the interaction with switches. Moreover, the chapter presented findings from the log data analysis such as the validity of the automated session separation and the metrics extracted from each session. These metrics enabled the identification of two different kinds of sessions: *engaging* and *unengaging*.

The results presented on this chapter provide an interesting insight on the strengths and weaknesses of the optical telegraph exhibit and bring light to different aspects of interaction with hands-on exhibits that can be used to guide future exhibit designs. Furthermore, the results and methodology used for the log data analysis is relevant to other studies that require session separation to perform evaluations over long periods of time. The opportunities previously mentioned are further explored on the following chapter, which combines the results from the different data collection methods of this study with the results of previous related work to analyze the most relevant findings of this work.

8. Discussion and conclusions

The Vapriikki MIRACLE project was a collaboration between the Vapriikki museum and the PIRG research group from the University of Tampere. During this project, an interactive replica of Edelcrantz's optical telegraph was designed and implemented. Furthermore, as part of this thesis work, a systematic evaluation of the exhibit (Section 6.1) was conducted utilizing a set of qualitative and quantitative methods. This chapter discusses the results of the exhibit evaluation and how they relate to previous work. The limitations of this thesis work are also presented in this chapter. To conclude, this chapter offers suggestions for future work and a set of guidelines for the design and evaluation of interactive exhibits.

8.1. Exhibit's appeal

Overall, the exhibit was well received among visitors, subjective measures from survey data indicate that users liked the installation since the exhibit was highly rated – a median score of four out of five. People seemed to have a pleasant time while interacting with the exhibit, they frequently smiled and showed signs of victory when they understood the coding principle. In two occasions users even took pictures of their conversations on the screen. Users also had positive comments about the exhibit. One of them commented: *"It was nice, very nice that I can try for myself how it works. Good in the museums"*, another one mentioned: *"It was nice to interact with such things, you learn faster. Not just reading but doing"*. Moreover, a visitor enquired after the interview if the installation was available online because he wished to continue playing.

The problem-solving challenge was one of the most attractive and strongest aspects of the exhibit, as evidenced in the survey data results which show that 41% of the participants' main interest was the problem-solving challenge. In addition to this, there was a statistically-significant, positive relationship between the overall rating of the system and the users' technical and problem-solving interests. Moreover, subjective impressions from the interview show that the challenge's difficulty level was appropriate, with 81% of users rating it as easy or medium. Some answers further reveal visitors' preference for a challenge level *"It would be less interesting if there were clear instructions"* and *"(…) it is more fun to figure it out"*.

An interesting factor is that the exhibit was given a high rating despite frequent complaints about the communication quality with the chatbot agent. An explanation to this could be found in the different interest profiles of visitors, as interviews revealed three main profiles in the users: **problem-solvers, communicators and history-enthusiasts**. Therefore, we infer that because of the different dimensions in the interaction: problem-solving challenge, communication with the artificial agent and historical aspect; users may have overseen the flaws in the communication aspect and still enjoyed the overall interaction with the exhibit.

8.2. Stopping power

It can be said that the telegraph replica has a moderately successful stopping power (27%), as compared with other studies [Sandifer, 2003]. This could be explained by a series of factors:

1. The exhibit was located right at the entrance of the exhibition, which might have favored its visibility and affect the attracting power [Korn and Jones, 2000].
2. A large projection of the remote tower was visible and caught people's attention when entering the exhibition.
3. The use of prompt messages inviting users to start interacting [Lee et al., 2015].
4. The switches acted as an inviting element for interaction, especially among children, who would run towards the switches to start moving them.

8.3. Dwell time

Quantitative measures extracted from the log data – presented on Section 7.4.2 – indicate that the exhibit's dwell time (1:27 minutes) was among the average and consistent with previous findings in other museum exhibit studies [Horn et al., 2008; Hornecker and Stifter, 2006a; Lee et al., 2015; Sandifer, 2003]. However, a deeper analysis of different sessions showed that the dwell time of sessions where users formed at least three *different* message codes – categorized as *engaging* sessions – is significantly higher (2:23 minutes) in comparison to the exhibit's general dwell time.

The difference in holding times could be due to the simple fact that forming and sending more messages requires more time. Nevertheless, it can also be attributed to the fact that users who formed at least three *different* messages were more likely to have understood the coding principle and therefore became more engaged with the exhibit. This assumption can be further supported by the fact that users on engaging sessions constructed an average of 6.27 *different* codes per session, a number much higher than the threshold defined to categorize sessions as engaging (3 codes formed). Hence, it can be interpreted that users in these longer sessions had indeed a good understanding of the coding principle and were engaged in the interaction.

Other attributes in the exhibit may have contributed to eliciting longer sessions. Sandifer [2003] observed that exhibits which give freedom to users to complete one task in several ways (open-ended quality) hold visitors attention for longer periods of time. In the case of the telegraph installation, users could communicate freely with the chatbot using a variety of messages with no strict order on how these messages could be sent. This ability to support conversations of various lengths suggests that the exhibit was also successful in rewarding short-term and long-term interactions [Hinrichs et al., 2008].

8.4. Communication power

There are some indicators that visitors understood the principle of how the telegraph worked and learned about its history. Subjective user impressions collected in the survey show that users considered the exhibit helped them learn about the telegraph's operation and its history (Figure 19). Additionally, 81% percent of the interviewees manifested that they understood how the codification principle worked.

Quantitative metrics extracted from the log file analysis provide extra evidence that the exhibit was successful to deliver the learning message, in 54% of the sessions users constructed more than three *different* codes (Section 7.4.2) with an average of over 6 *different* codes formed per session. This metric strongly suggests that users in these sessions understood the coding principle as it is unlikely that users could construct multiple codes by chance. Nonetheless, the fact that the exhibit possibly failed to deliver the learning message in the remaining sessions (46%) evidences that there is room for improvement in regards to the exhibit's communication power. Therefore, this study recommends that some form of aid is provided to novice users to help the learning process and increase the number of sessions with a high number of codes constructed (Section 8.9).

8.5. Group interactions and collaboration

A clear majority of users interacting with the exhibit were part of a group, coinciding with results from previous museum studies [Ciolfi and Bannon, 2003; Hornecker and Stifter, 2006b; Lee et al., 2015]. Quantitative measures obtained from the observation sessions reveal that 71% of group members actively manipulated the physical elements in the installation, suggesting that the exhibit was supportive of group interactions and collaboration [Horn et al., 2008].

8.5.1. Family groups vs. non-family groups

When taking a closer look at the interaction and collaboration behaviors in groups, observations showed that family and non-family groups differed from each other in the number of active members participating in the interaction and how they collaborated with each other.

Non-family groups were likely to have all their members interacting with the system. Moreover, non-family groups were notably more active than family groups, showing a *fully-active* behavior – with all members active – twice as often as family groups (Section 7.1.5). One explanation to family groups being more passive than non-family groups could be that children tend to be *active* users more frequently than adults [Sandifer, 1997]. Therefore, as family groups usually contain children, the adult members in the group would remain passive and watch the children interact.

Collaboration was also different among both group types. Non-family groups presented a turn-taking behavior, also found in previous studies [Hinrichs et al., 2008; Hornecker and Stifter, 2006a; Lee et al., 2015]. On the other hand, family groups were more likely to have discussions where parents guided and explained to the children how the exhibit worked, presenting a guiding pattern instead [Lee et al., 2015]. Active family group members collaborated in more cohesive ways; they were more likely to share the physical elements, moving the same switches together or parents moving switches and children pressing the send button.

Differences in collaboration behaviors depending on the group type could be attributed the age differences between group members. Non-family groups generally consisted of people within the same age group whereas family groups comprised people of different ages, mostly adults and children. Therefore, in non-family groups visitors were all ‘peers’, equally eager to interact and discuss to understand how the exhibit worked. On the other hand, in family groups parents were more willing to act as guides and explain how the exhibit worked and leave the first-hand interaction to their children, a quality of family groups also identified in other studies [Crowley et al., 2001; McManus, 1994].

8.6. Effect of preceding interactions

Existing literature explains how previous visitors using an exhibit can negatively affect future users' interactions [Heath et al., 2005]. This phenomenon seemed to affect user interactions with the telegraph exhibit in two different ways: the quality of communication with the chatbot and the initial position of the switches.

The chatbot communication was hindered in situations where visitors started interacting immediately after the previous user left the exhibit, the situation was particularly evident during the museum night observation when the museum was crowded. The chatbot was designed with the assumption that there would be some idle time between different sessions. Therefore, it was not able to interpret these immediate transitions and instead continued the previous conversations. This translated into many situations where the chatbot gave incoherent answers to the visitors' messages or just answered *"I don't understand".*

Preceding users also affected the initial interaction with the switches. Tangible objects such as the switches, are limited by their physical constraints, they cannot move independently and, unlike virtual objects, do not provide a trace of their previous states [Manches et al., 2009]. Thus, the starting positions of the switches always depend on the previous users and impede the same entry-point for all users, which could confuse newcomers. During the interviews, two visitors pointed out that the positions of the switches had been moved, otherwise stated, previous users had left some switches OFF and others ON. Presumably, these users expected all switches to be in an initial state with all switches OFF or ON, displaying all shutters in the local virtual tower as hidden or visible, respectively.

The matters described previously can hinder the usability of the exhibit as users may not be aware of what is happening when they start interaction, affecting the "visibility of the system" [Nielsen, 1994]. The issue was evident when some users did not clearly understand what is the status of the conversation or the switches positions when they started interacting. When referring to the chatbot conversations one user commented: *“Maybe we entered the conversion in the middle of someone else's”* while another one stated: *“At first I didn't realize that the conversation had been going on”*. Additionally, when discussing the difficulty of the coding principle an interviewee pointed out: *"You just have to try a couple of times; the switches were turned on in the beginning”*. One solution to improve the quality of conversations with the chatbot on such crowded situations is adding a reset button to the installation. A reset button allows users to explicitly start a new conversation and guarantees interaction from an initial state.

8.7. Interaction with switches

One very important element in the design of the telegraph exhibit was the tangible input with switches and buttons. As described in the previous section, the initial position of the switches caused confusion in a few of the users. However, despite the potential issues presented by the switches, they appeared to add substantial value to the user experience of the exhibit. Visitors repeatedly expressed positive attitudes towards the switch interaction. For instance, an interviewee reported to enjoy *“the sound of the switches when you move them”,* while another one mentioned: *“flickering the switches was fun”*. Additionally, the switches seemed to have an extra appeal to adult users given its “old-fashioned” quality, as expressed by one of the users, *“I like the switches, it is nice to have them these days”*. Some interviewees even conveyed their preference over newer technologies, one of then mentioned: *"old-fashioned feels nice, instead of just a screen”* while another one expressed: *“I like it. More fun than just a touchscreen, you can really feel feedback"*. Therefore, it can be inferred that in the case of the optical telegraph exhibit, the switches contributed more to positive experiences than they hindered interactions.

8.8. Design shortcomings

The evaluations exposed two possible flaws in the design of the exhibit. These shortcomings can be summarized as follows:

Perceived buttons on screen: Users tried to press the language and send buttons on the screen. There are two factors that could have contributed to this *false affordance* [Gaver, 1991]: the first is that certain elements in the graphical user interface appeared to be clickable buttons, as expressed by one of the interviewees, *“It was confusing that the send looked like a button* *on the screen".* The second factor could be that nowadays many installations have touch screens and users immediately expect screens to be touch-sensitive.

Limited conversations: As presented in Figure 22, an examination of the messages sent by users revealed that more than half of the messages available in the codebook were hardly used. Nevertheless, many users expressed that they would like to have more message options in the codebook to communicate better (Section 7.3). This could suggest that the least-used messages did not lead to interesting conversations or that users would simply like to have the illusion of broader communication possibilities.

Although both issues may have an explanation in users' own perceptions and expectations of the system, there are possible to solutions to reduce their impact on user interactions. Regarding the false affordance of the screen, the problem could be potentially reduced by changing the visual design to eliminate the button appearance of the elements on the graphical user interface. As for the chatbot limitations, the chatbot responses could be further improved and the user codebook could be expanded in order to provide broader conversation possibilities.

8.9. Design guidelines for interactive exhibits

The evaluation of the Vapriikki optical telegraph replica revealed positive and negative aspects about the exhibit's design. The results of this study, in combination with knowledge from previous work, can provide valuable insights for the design of future interactive exhibits. The following is a compilation of lessons learned, presented as a set of guidelines for the design of interactive exhibits:

1. **Design for multiple interests:** Museumgoers have different interest profiles that influence the exhibits they visit [Hornecker and Stifter, 2006b]. Designing an exhibit that appeals to the interest of different personas is key to attracting a broader spectrum of users and holding their attention for longer times, hence, enhancing the exhibit's attractive power and dwell times. In the Vapriikki optical telegraph evaluation, we identified three different profiles: **problem-solvers, communicators and history-enthusiasts**. As evidenced in the interview answers, users had complaints about the communication aspect of the system, however, most users conveyed

satisfaction with the exhibit and gave a high rating to the installation in the survey questionnaire. Hence, we consider that the exhibit's different appeals such as the problem-solving challenge, the history facts and interactivity made users forgiving towards possible flaws in the communication side.

2. Consider preceding interactions: Interactive exhibits' attractiveness can sometimes cause users to line up or wait around for previous users to leave the exhibit to immediately start interacting. Preceding users can affect how the interaction unfolds for following users [Heath et al., 2005] and in many cases users are likely to start interaction when the system is not at its initial state [Block et al., 2015]. In the Vapriikki optical telegraph installation, preceding users affected the next users in two ways: 1) the switches initial positions depended on where the previous user had left them (ON or OFF), 2) the chatbot was designed to reset conversations only after a certain idle time, therefore, continuous users were perceived as the same user. Thus, it is of paramount importance to carefully evaluate the effects of preceding interactions during the design phase of a new interactive exhibit. Whenever possible, exhibits should aim to enable continuous interactions, where the actions of previous users do not affect newcomers.

3. Challenge users with a puzzle: This study revealed that the problem-solving challenge was the most attractive aspect of the exhibit. Users were observed celebrating victory when they successfully constructed a code and they expressed in the interviews how rewarding it was to understand the coding principle of the telegraph. We consider that designing exhibits that challenge users can contribute to the exhibit's appeal and enhance visitors' experience. Nevertheless, it is important to be aware that users' may require different times to complete the challenge provided (see guideline 6).

4. Support open-ended interactions: Sandifer [2003] identified that exhibits which offer users a certain liberty to perform tasks can hold visitors' attention for longer times. In the case of the Vapriikki optical telegraph, users had freedom to communicate with the chatbot sending few or several messages, in any given order. We argue that this open-ended quality contributed to user engagement with the exhibit in 54% of the sessions, as indicated by the log file analysis. Furthermore, the interview results show that users wished to increase their freedom to send messages by expanding the user codebook. Hence, we recognize the advantages that open-ended interactions can offer in terms of increasing dwell times and enhancing engagement. Additionally, designing open-ended interactions can allow short-term and long-term explorations of the exhibit, a need identified in previous studies [Hinrichs et al., 2008].

5. Evaluate tangibles' trade-offs: Tangible elements were considered a nice element in the telegraph exhibit. During the evaluation, observers noticed how users – particularly children – were attracted to move the switches. Furthermore, in the interviews adult users repeatedly mentioned that playing with the switches was a nice element in the interaction, a couple of users stated that they preferred switches because they offer an "old-fashioned" quality as opposed to touch screens. Therefore, we infer that the switches played an important role in attracting users and contributed to a pleasant user experience, even appealing to the nostalgia of some users. Nonetheless, it is noteworthy to say that the use of tangibles also hindered some aspects of the interaction. Given their physical quality, tangibles have some restrictions from the physical world that virtual objects do not have [Manches et al., 2009]. In the telegraph replica, the switches restricted the possibility to provide the same entry-point to all users, since it was impossible to control the initial state or position of the switches. Hence, we emphasize the importance of assessing the benefits and possible consequences of utilizing tangible elements in future interactive exhibits.

6. Aid novice users: Given the diverse backgrounds and interests museumgoers have, it is likely that they will require different amounts of time to understand how the exhibit works. Thus, interactive exhibits need to provide a challenge level that matches the user's skills [Allen, 2004]. In the evaluation of the Vapriikki telegraph, though the challenge level was rated as medium, it could be inferred that in 46% of the sessions users did not understand the coding principle. Hence, we suggest offering help to novice users to balance the challenge level of the exhibit. Help could be provided explicitly (with a Help option) or through hints and cues that appear after a certain time. Aiding novice users early in the interaction can avoid users leaving the exhibit because of boredom and frustration, possibly contributing to improve the exhibit's dwell time and its communication power.

These guidelines are intended to aid the design of interactive exhibits by providing a comprehensive set of recommendations that can help exhibit designers and developers make well-informed choices early in the design stage. We consider that these recommendations may improve several factors in the exhibit such as its attractiveness, dwell time, communication power and contribute to a positive User Experience.

8.10. Semi-automated log analysis

Besides the findings previously described, which mostly concern the design of interactive exhibits, one of the main contributions of this thesis work is that it showed that it is possible to programmatically separate visitor sessions by assuming an idle time between sessions. Though the same rationale has been applied in other studies before, the separation was performed manually using video recordings [Peltonen et al., 2008]. Additionally, this work used a semi-automated log analysis to extract information from every session as means to explore deeper aspects of the interaction patterns, such metrics helped assess the exhibit's dwell time, user engagement and communication power (Section 7.4.2).

Furthermore, this study showed the viability of performing longitudinal analyses using a semi-automated log file analysis, which would require extensive human resources otherwise. Finally, this work demonstrated how triangulating data from qualitative and quantitative methods can enhance and validate the results to provide a meaningful, holistic overview of the system being evaluated.

8.11. Limitations of the study

This work had some limitations. As mentioned in Section 7.4.1, the algorithm that separates sessions is not reliable when the idle time between two different sessions is marginal. In other words, in those special occasions where the museum is overcrowded and the time gap between interaction sessions is reduced to a couple of seconds, different sessions can be clustered into one, making it impossible to separate sessions reliably. Moreover, the device used to detect user presence was not accurate and therefore its data could not be used to complement the session-separation algorithm.

8.12. Future work

For future work, it is necessary to utilize a reliable device to track the presence of users around the installation. For instance, a Kinect-based tracking method could enhance the confidence of the semi-automated analysis, provide an automated form of collecting passerby information and bring light to other aspects of interaction. Additionally, exploring the use of machine-learning techniques for pattern recognition could improve the session-separation algorithm and potentially increase the confidence of its results, especially when used in combination with reliable user-tracking.

9. Summary

This thesis work presented the design and evaluation of an interactive replica of Abraham Edeldrantz's shutter telegraph, using tangible elements and mixed reality. This chapter answers the research question that guided this thesis and compiles the findings of this work:

How to utilize a semi-automated log data analysis to evaluate interactive exhibits over a long period of time? Using an algorithm that assumes a certain time gap between different users can be used to automatically separate sessions and extract relevant data from each session afterwards. The data extracted can facilitate an in-depth analysis of interaction patterns and allow longitudinal analyses that would otherwise require extensive human resources.

Interaction and collaboration in groups differs depending on the age difference among participants. In groups where members have similar ages (non-family groups), users are more likely to have all their participants interacting with the system and scaffold the interaction. In groups with large age difference between members (family groups), users are more likely to leave direct manipulation to the youngest members while adults act as guides.

Communication options should be broad to enable varied, meaningful human-agent conversations. Users desire a wide range of possibilities to have more "real" conversations with chat agents.

A set of design guidelines for interactive exhibits was presented on Section 8.9. The guidelines are a compilation of the lessons learned from the empirical work conducted in this thesis and considering previous related work. Briefly, these guidelines are:

1. *Design for multiple interests:* providing a variety of elements in the interaction, which satisfy the interests of different personas, can make an exhibit attractive to a broad spectrum of visitors.

2. *Consider preceding interactions:* previous users can affect how the next user will experience the interaction, exhibits should avoid these situations and aim for designs that favor continuous interactions.

3. *Challenge users with a puzzle:* giving users a challenge or puzzle to solve as part of the interaction can increase the exhibit's attractiveness and hold visitors' attentions for longer.

4. *Support open-ended interactions:* giving users freedom to perform the exhibit's tasks in different ways can help increase the exhibit's dwell time and favor engagement.

5. *Evaluate tangibles' trade-offs:* switches and buttons offer an attractiveness given their physicality and "old-fashioned" appeal. However, they negatively impact other aspects of the interaction, for example, they impede a consistent entry-point for interaction.

6. *Aid novice users:* users require different times to solve the challenges posed by the exhibit, assisting novice users with help and hints can prevent them from stopping interacting early.

To sum up, this thesis presented a comprehensive literature review about interactive museum exhibits, the evaluation methodology and different metrics used in museum exhibit studies. Furthermore, it described the design and implementation of an interactive replica of the shutter telegraph as part of the Vapriikki MIRACLE project. In this study, the replica exhibit was evaluated utilizing a combination of a semi-automated log file analysis and ethnographically-oriented methods. Finally, this work presented the lessons learned on the design of interactive exhibit, offering a set of design guidelines to help the design of future exhibits, as well as the results of the methodology used to evaluate this interactive exhibit. Future work could involve utilizing a more accurate user-presence tracking and using machine learning to increase the confidence of the automated session separation method.

References

[Allen, 2004] Sue Allen, Designs for learning: Studying science museum exhibits that do more than entertain. *Science Education*, **88** (S1), S17–S33.

[Allen and Gutwill, 2004] Sue Allen and Joshua Gutwill, Designing with multiple interactives: five common pitfalls. *Curator: The Museum Journal*, **47** (2), 199–212.

[Block et al., 2015] Florian Block, James Hammerman, Michael Horn, Amy Spiegel, Jonathan Christiansen, Brenda Phillips, Judy Diamond, E. Margaret Evans and Chia Shen, Fluid grouping: quantifying group engagement around interactive tabletop exhibits in the wild. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, ACM Press, 867–876.

[Boehner et al., 2005] Kirsten Boehner, Jennifer Thom-Santelli, Angela Zoss, Geri Gay, Justin S. Hall and Tucker Barrett, Imprints of place: creative expressions of the museum experience. In: *CHI '05 extended abstracts on Human factors in computing systems*, ACM Press, 1220-1223.

[Caulton, 2006] Tim Caulton, *Hands-On Exhibitions: Managing Interactive Museums and Science Centres*. Routledge, 2006.

[Ciolfi et al., 2003] Luigina Ciolfi and Bannon Liam, Learning from museum visits: shaping design sensitivities. In *Proceedings of HCI International Vol 1. 2003*, 63-67.

[Cone and Kendall, 1978] Cynthia A. Cone and Keith Kendall, Space, time, and family interaction: visitor behavior at the science museum of Minnesota. *Curator: The Museum Journal*, **21** (3), 245–258.

[Crowley et al., 2001] Kevin Crowley, Maureen A. Callanan, Jennifer L. Jipson, Jodi Galco, Karen Topping, Jeff Shrager, Shared scientific thinking in everyday parent-child activity. *Science Education*, **85** (6), 712–732.

[Davies and Brember, 1994] Julie Davies and Ivy Brember, The reliability and validity of the “smiley” scale. *British Educational Research Journal,* **20**, 447–454.

[Dean, 1996] David Dean, *Museum Exhibition: Theory and Practice*, Routledge, 1996.

[Falk and Dierking, 2000] John H. Falk and Lynn D. Dierking, *Learning from museums : visitor experiences and the making of meaning*. AltaMira Press, 2000.

[Falk et al., 2004] John H. Falk, Carol Scott, Lynn Dierking, Leonie Rennie and Mika Cohen Jones, Interactives and visitor learning. *Curator: The Museum Journal*, **47** (2), 171–198.

[Falk et al., 1985]. John H. Falk, John J. Koran Jr., Lynn D. Dierking and Lewis Dreblow, Predicting visitor behavior. *Curator: The Museum Journal*, **28** (4), 249–258.

[Fernández and Benlloch, 2000] Guillermo Fernández and Montserrat Benlloch, Interactive exhibits: how visitors respond. *Museum International*, **52** (4), 53–59.

[Fitzmaurice, 1993] George W. Fitzmaurice, Situated information spaces and spatially aware palmtop computers. *Communications of the ACM*, **36** (7), 39–49.

[Gaver, 1991] William W. Gaver, Technology affordances. In: *Proceedings of the SIGCHI conference on Human factors in computing systems Reaching through technology - CHI '91*, ACM Press, 79–84.

[Gebhard and Karsten, 2009] Patrick Gebhard and Karsten Susanne, On-Site evaluation of the interactive COHIBIT museum exhibit. In: *Intelligent Virtual Agents*, Springer, 174–180.

[Gibson, 1977] James J. Gibson, The Theory of Affordances. In: *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 67–82).

[Grinter et al., 2002] Rebecca E. Grinter, Paul M. Aoki, Amy Hurst, Margaret H. Szymanski, James D. Thornton and Allison Woodruff, Revisiting the visit: understanding how technology can shape the museum visit. In: *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, ACM, 146–155.

[Hall and Bannon, 2005] Tony Hall and Bannon Liam, Designing ubiquitous computing to enhance children's interaction in museums. In: *Proceeding of the 2005 conference on Interaction design and children - IDC '05*, ACM Press, 62–69.

[Hassenzahl, 2008] Marc Hassenzahl, User experience (UX). In: *Proceedings of the 20th International Conference of the Association Francophone d'Interaction Homme-Machine on - IHM '08* , ACM Press, 11-15.

[Hassenzahl, M., Burmester, M., Koller, 2015] F. Hassenzahl, M., Burmester, M., Koller, AttrakDiff, 2015. [Online]. Available: http://www.attrakdiff.de

[Heath et al., 2005] Christian Heath, Dirk Vom Lehn, and Jonathan Osborne, Interaction and interactives: collaboration and participation with computer-based exhibits. *Public Understanding of Science*, **14** (1) , 91–101.

[Heinecke, 1995]. Andreas M. Heinecke, Evaluation of hypermedia systems in museums POI systems in museums evaluation of POI systems. *ICHIM, Multimedia Computing and Museums*, 67–78.

[Hinrichs et al., 2008] Uta Hinrichs, Holly Schmidt and Sheelagh Carpendale, EMDialog: bringing information visualization into the museum. *IEEE Transactions on Visualization and Computer Graphics*, **14** (6), 1181–1188.

[Holzmann and Pehrson, 1994] Gerard J. Holzmann and Björn Pehrson, *The Early History of Data Networks* (1 edition.). Wiley-IEEE Computer Society Press, 1994.

[Horn et al., 2012] Michael S. Horn, Zeina Atrash Leong, Florian Block, Judy Diamond, E. Margaret Evans, Brenda Phillips and Chia Shen, Of BATs and APEs: an interactive tabletop game for natural history museums. In: *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, ACM Press, 2059-2068.

[Horn et al., 2012] Michael S. Horn, R. Jordan Crouser and Marina U. Bers, Tangible interaction and learning: the case for a hybrid approach. *Personal and Ubiquitous Computing*, **16** (4), ACM, 379–389.

[Horn et al., 2008] Michael S. Horn, Erin Treacy Solovey and Robert JK Jacob, Tangible programming and informal science learning: making TUIs work for museums. In: *IDC '08 Proceedings of the 7th international conference on Interaction design and children*, ACM, 194–201.

[Hornecker et al., 2006a] Eva Hornecker and Matthias Stifter, Digital backpacking in the museum with a SmartCard. In: *Proceedings of the 7th ACM SIGCHI New Zealand Chapter's International Conference on Computer-human Interaction: Design Centered HCI*, ACM, 99–107.

[Hornecker et al., 2006b] Eva Hornecker and Matthias Stifter, Learning from interactive museum installations about interaction design for public settings. In: *Proceedings of the 18th Australia conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments*, ACM, 135–142.

[Ishii and Ullmer, 1997] Hiroshi Ishii and Brygg Ullmer, Tangible bits: towards seamless interfaces between people, bits and atoms. In: *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '97*, ACM, 234–241.

[Korn and Jones, 2000] Randi Korn and Johanna Jones, Visitor behavior and experiences in the four permanent galleries at the tech museum of innovation. *Curator: The Museum Journal*, **43** (3), 261–281.

[Lazar et al., 2014] Jonathan Lazar, Jinjuan Heidi Feng and Harry Hochheiser, *Research Methods in Human-Computer Interaction*. Wiley Textbooks, 2014.

[Lee et al., 2015]. Seung Ah Lee, Engin Bumbacher, Alice M. Chung, Nate Cira, Byron Walker, Ji Young Park, Barry Starr, Paulo Blikstein and Ingmar H. Riedel-Kruse, Trap it!: a playful human-biology interaction for a museum installation. In: *Proceedings of the ACM CHI'15 Conference on Human Factors in Computing Systems*, **1**, ACM, 2593–2602.

[Lee and Heller, 1997] Sungkyu Lee and Rachelle S. Heller, Use of a keystroke log file to evaluate an interactive computer system in a museum setting. *Computers & Education*, **29** (2–3), 89–101.

[Manches et al., 2009] Andrew Manches, Claire O'Malley and Steve Benford, Physical manipulation: evaluating the potential for tangible designs. In: *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction - TEI '09*, ACM, 77-84.

[Marshall et al., 2011]. Paul Marshall, Richard Morris, Yvonne Rogers, Stefan Kreitmayer, Matt Davies, Rethinking “multi-user”: an in-the-wild study of how groups approach a walk-up-and-use tabletop interface. In: *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI ’11*, ACM, 3033-3042.

[Marshall et al., 2016] Using tangible smart replicas as controls for an interactive museum exhibition. In: *Proceedings of the TEI ’16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction - TEI ’16*, ACM 159–167.

[McManus, 1994] Paulette McManus, Families in Museums. In: Roger Miles and Lauro Zavala, *Towards the Museum of the Future: New European Perspectives*, Routledge, 1994, 81–97.

[Nielsen, 1994] Jakob Nielsen, Enhancing the explanatory power of usability heuristics. In: *Proceedings of the SIGCHI conference on Human factors in computing systems celebrating interdependence - CHI ’94*, ACM, 152–158.

[Nielsen, 2012] Jakob Nielsen, Usability 101: Introduction to Usability. [Online]. Available: https://www.nngroup.com/articles/usability-101-introduction-to-usability. [Accessed: 09-Apr-2017]

[Norman, 1999] Donald A. Norman, Affordance, conventions, and design. *Interactions*, **6** (3), 38–43.

[Norman, 2013] Donald A. Norman, *The design of everyday things Revised and expanded edition.* Basic Books, 2013.

[Panagiotis et al., 2013] Panagiotis Zaharias, Michael Despina and Yiorgos Chrysanthou, Learning through multi-touch interfaces in museum exhibits: an empirical investigation. *Journal of Educational Technology & Society*, **16** (3), 374–384.

[Peltonen et al., 2008] Peter Peltonen, Esko Kurvinen, Antti Salovaara, Giulio Jacucci, Tommi Ilmonen, John Evans, Antti Oulasvirta and Petri Saarikko, It’s Mine, Don’t Touch!: interactions at a large multi-touch display in a city centre. In: *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI ’08*, ACM, 1285-1294.

[Pousette et al., 2014] Sandra Pousette, Martin Löfgren, Birgitta Nilsson and Anders Gustafsson, An extended method to measure overall consumer satisfaction with packaging. *Packaging Technology and Science*, **27**(9), 727-738.

[Rekimoto and Nagao, 1995] Jun Rekimoto and Katashi Nagao, The world through the computer: computer augmented interaction with real world environments. In: *Proceedings of the 8th annual ACM symposium on User interface and software technology - UIST ’95*, ACM, 29–36.

[Rizzo and Garzotto, 2007] Francesca Rizzo and Franca Garzotto, "The Fire and The Mountain": tangible and social interaction in a museum exhibition for children. In: *Proceedings of the 6th international conference on Interaction* *design and children - IDC '07*, ACM, 105-108.

[Sandifer, 1997] Cody Sandifer, Time-based behaviors at an interactive science museum: exploring the differences between weekday/weekend and family/nonfamily visitors. *Science Education,* **81**(6), 689-701.

[Sandifer, 2003] Cody Sandifer, Technological novelty and open-endedness: two characteristics of interactive exhibits that contribute to the holding of visitor attention in a science museum. *Journal of Research in Science Teaching*, **40** (2), 121–137.

[Schmitt et al., 2010] Bénédicte Schmitt, Cedric Bach, Emmanuel Dubois and Francis Duranthon, Designing and evaluating advanced interactive experiences to increase visitor's stimulation in a museum. In: *Proceedings of the 1st Augmented Human International Conference on - AH '10*, ACM, no. 15.

[Shettel, 2001] Harris Shettel, Do we know how to define exhibit effectiveness? *Curator: The Museum Journal*, **44** (4), 327–334.

[Tahvanainen, 1994] Karl V. Tahvanainen, *Ord i sikte - Den optsika telegrafen i Sverige 1794 - 1881.* Stockholm: Nordstedts tryckeri AB, 1994.

[Turunen et al., 2009] M. Turunen, J. Hakulinen, A. Melto, T. Heimonen, T. Laivo, and H. Juho, SUXES - user experience evaluation method for spoken and multimodal interaction., INTERSPEECH 2009, 10th Annu. Conf. Int. Speech Commun. Assoc. Bright. United Kingdom, 2009, 2567–2570

[Vom Lehn et al., 2001]. Exhibiting interaction: conduct and collaboration in museums and galleries. *Symbolic Interaction*, **24** (2), 189–216.

[Vom Lehn and Heath, 2005]. Dirk Vom Lehn and Christian Heath, Accounting for new technology in museum exhibitions. *International Journal of Arts Management*, **7** (3), 11–21.

[Wellner et al., 1993]. Pierre Wellner, Wendy Mackay, Rich Gold, Back to the real world. *Communications of the ACM*, **36** (7), 24–27.

[Wojciechowski et al., 2004] Rafal Wojciechowski, Krzysztof Walczak, Martin White and Wojciech Cellary, Building virtual and augmented reality museum exhibitions. In: *Proceedings of the ninth international conference on 3D Web technology - Web3D '04*, ACM, 135-144.

[Wolf et al., 2007] Marius Wolf, Eric Lee and Jan Borchers, Education, entertainment and authenticity: lessons learned from designing an interactive exhibit about medieval music. In: *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07*, ACM, 1887–1892.

[Woods et al., 2004] Eric Woods, Mark Billinghurst, Julian Looser, Graham Aldridge, Deidre Brown, Barbara Garrie and Claudia Nelles, Augmenting the science centre and museum experience. In: *Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Austalasia and Southe East Asia - GRAPHITE '04,* ACM, 200-236.

Appendix 1: Full Edelcrantz codebook

CODE	MESSAGE	CODE	MESSAGE
000	0	**401**	MEN
001	1	**402**	MI
002	2	**403**	MIN
003	3	**404**	MÄ
004	4	**441**	N
005	5	**442**	NE
006	6	**443**	NG
007	7	**444**	NEN
010	SPACE	**445**	NN
011	8	**446**	NS
012	9	**447**	NU
013	A	**472**	O
014	AA	**473**	ON
015	AI	**474**	OLEN
016	AL	**475**	OI
017	AN	**476**	OS
020	AAN	**477**	OT
021	AR	**503**	P
022	AS	**504**	PA
023	ASS	**505**	PE
024	AT	**506**	PI
025	AAT	**507**	PO
031	B	**510**	PU
047	C	**515**	Q
063	D	**520**	R
064	DE	**521**	RA
065	DEN	**522**	RANSKA
066	DEL	**523**	RI
067	DO	**524**	RU
115	E	**525**	RUOTSI
116	EE	**526**	RO
117	EK	**544**	S
120	EM	**545**	SA
121	EN	**546**	SAKSA
122	ENSIM	**547**	SE

123	ES	**550**	SEN
124	ET	**551**	SILL
130	F	**552**	STA
173	G	**553**	STI
224	H	**554**	SIITÄ
225	HA	**654**	T
226	HE	**655**	TA
227	HI	**656**	TE
230	HO	**657**	TI
277	INN	**660**	TÄ
300	IM	**661**	TÖ
301	IS	**670**	TTI
302	IT	**671**	TÄSTÄ
231	HJ	**672**	TUOSTA
232	HT	**706**	U
273	I	**707**	UA
274	IA	**710**	UO
275	IE	**711**	UT
276	IN	**712**	UU
303	J	**713**	US
304	JA	**714**	V
305	JO	**715**	W
306	K	**755**	X
307	KA	**756**	Y
310	KE	**757**	YI
311	KI	**760**	YK
334	L	**761**	YN
335	LA	**762**	Z
336	LLA	**763**	Å
337	LAINEN	**765**	Ä
340	LE	**772**	-
341	LLE	**773**	‘
342	LLÄ	**774**	Ö
376	M	**775**	,
377	MA	**776**	?
400	ME	**777**	.

Appendix 2: User codebook (Finnish)

Code	Message
A001	VIESTI VASTAANOTETTU
A002	SAATEKO VIESTINI?
A004	YMMÄRRETTY
A005	TOISTAAKAA
A007	LOPPU
A010	KENEN KANSSA VIESTITÄN?
A020	KERTOKAA LISÄÄ
A077	KYLLÄ
A700	EI
A777	EHKÄ
A301	KUKA?
A302	MILLOIN?
A304	MISSÄ?
A311	MIKSI?
A312	MITEN?
A313	MITÄ?
A401	ENSIMMÄINEN
A402	TOINEN
A403	KOLMAS
A701	AIEMPI
A702	JALKIMMÄINEN

Appendix 3: Experience questionnaire

Experiences about *the optical telegraph – the Internet of its time*

Date: ___________ 2016 Time: ______:______

Age: ______ years Gender: ☐ Boy/male ☐ Girl/female ☐ Other

For the following statements, select the option that best describes your own experience.

	Totally disagree	**Neither agree nor disagree**			**Totally agree**
The first impression about the telegraph was interesting.	☐	☐	☐	☐	☐
The idea of the telegraph was clear.	☐	☐	☐	☐	☐
The telegraph was irritating.	☐	☐	☐	☐	☐
Communicating with the telegraph was fun.	☐	☐	☐	☐	☐
I would like to communicate with the telegraph again.	☐	☐	☐	☐	☐
Sending the messages was easy.	☐	☐	☐	☐	☐
Figuring out the coding principle was difficult.	☐	☐	☐	☐	☐
The application helped me to understand the principle of the optical telegraph.	☐	☐	☐	☐	☐
The application taught be about the telegraph's history	☐	☐	☐	☐	☐
These kinds of applications would increase my interest towards museum visits.	☐	☐	☐	☐	☐
I would like to tell my friend about the telegraph.	☐	☐	☐	☐	☐
I would like to share my coding experience to my friends in social media (e.g., Facebook, Twitter or Instagram).	☐	☐	☐	☐	☐
I am interested in technical things.	☐	☐	☐	☐	☐
I am interested in playing.	☐	☐	☐	☐	☐
I am interested in problem-solving tasks.	☐	☐	☐	☐	☐
I am interested in history.	☐	☐	☐	☐	☐

How much did you like the telegraph installation as a whole?	☹	☹	🙂	☺	😀
	☐	☐	☐	☐	☐

What was your main interest while using the application?

☐ Solving the code ☐ Chatting with Edelcrantz ☐ Both ☐ None ☐ Other _______________

Appendix 4: Observation form

Observation

User code: ____ Usage started (approx.) ___: ___, Usage ended(approx.) ____:____

Duration of the usage, (approx.) ________

Number of users: ☐ Single user ☐ Small group _____ ☐Family ______
Active participants ________ Passive participants __________
Gender: ☐ Male ☐ Female ☐ Other
Age group: ☐ < 12 ☐ 12-18 ☐ 19-29 ☐ 30-59 ☐ 60+

1. User's spontaneous comments when using the application:

   ____
   ____
   ____
   ____

2. The user seemed to be:

   ☐ Relaxed, interested ☐ Uncomfortable
   ☐ Surprised ☐ Confused, unaware, uninformed
   ☐ Performing to others ☐ Other _______________________________

3. Based on the user's actions, s/he seemed to feel:

   ☐ Positive ☐ Happy
   ☐ Negative ☐ Bored, disappointed
   ☐ Inquiring, curious ☐ Other ____________________
   ☐ Impressed

4. The user seemed to understand the coding principle:

   ☐ Immediately ☐ < 30 sec☐< 60 sec ☐> 60 sec ☐Did not understand

5. Other observations (e.g. estimate of time, if > 60 sec, unexpected function of application etc.)

____
____
____
____