Voice interaction as a control modality in a real-time mobile game

The limitation of latency and accuracy in a cloud-based speech recognition

Christian Lindberg & Joakim Linna

Department of Computer and Systems Sciences

Degree project 15 HE credits
Computer and Systems Sciences
Degree project at the Bachelor Programme in Computer Game Development
Spring term 2021
Supervisor: Thomas Westin

Stockholm University

Abstract

Cloud-based Automatic Speech Recognition (ASR) is a technology that has evolved a lot over the last decade. It is being utilized ever more in small commercial devices as the computational power is in the cloud rather than the user device. Even though this technology is readily available for developers, it is rarely implemented in games as a primary control modality. Voice control is an excellent tool for accessibility enhancement in a mobile context. Every smartphone has a microphone and possible access to the internet. With 5G currently being introduced to the world, the mobile networks are predicted to get a big boost in performance with a technical possibility of latencies lower than 1 ms. Slow System Response Time (SRT) and low Command Success Rate (CSR) is two of the relevant challenges for implementing speech recognition successfully as a control modality for real-time interactions. The problem is that it is unclear how viable a cloud-based ASR would be in these regards. The following research question was formed: *How viable is a cloud-based ASR as the single control modality, in a mobile game, in relation to system response time (SRT) and command success rate (CSR)?*

To answer the research question, a simple mobile game called “Voice Snake” was developed and the interactions between participants and the game was investigated in an experiment designed as playtesting. Playtesting was chosen as they are useful for investigating UX of games. The experiment design considered SRT the independent variable and UX and UP the dependent variables. There was one experimental group of 3 participants and one control group of 3 participants. Both groups partook in two consecutive playtesting session, where the control group had its SRT increased with 70 ms for the second session. The study regards UX as something complex that needs qualitative data to be described. Hence, following common methods of data collection during playtesting, that could collect qualitative data regarding UX. This was notetaking during observations and discussions, which was then analyzed with a thematic analysis. Thematic analysis was identified as a useful method for analyzing this qualitative data. Quantitative data regarding UP and SRT was also automatically collected by the developed system and screen-recordings were used to calculate the CSR. As this data was numerical, statistical analysis was identified to be a good method of analysis.

The results could not say anything about how the manipulated SRT affected the UX or UP. Furthermore, the results showed how the positive and negative UX manifested. That the UX was positive up until a certain point within the game’s rounds and that the ASR performed with a mean SRT of 370 to 450 ms and a mean CSR of 92 to 96%. A correlation test indicated that higher CSR correlate with higher UP ($r = 0.27$ $p < 0,05$).

Regarding the SRT, it was concluded that the cloud-based ASR was viable up to a point where the snake was moving too fast. Furthermore, it was concluded that the implemented techniques to alleviate slow SRT provided support for the users. It was also concluded that feedback, provided to the user, about the system status would have enhanced the UX. Lastly, it was concluded that a CSR closer to 100% is preferable for the kinds of fast paced interactions that can result in user failure.

Keywords: Automatic Speech Recognition, Voice Control, System Response Time, Latency, Mobile Game, Accessibility, 5G

Synopsis

Background	Cloud-based Automated Speech Recognizers (ASR) implements deep learning techniques such as neural networks. Modern smartphones have personal assistants that use cloud-based ASRs to perform tasks e.g., voice dictation and voice dialing. Voice recognition is, however, seldom utilized as a primary control modality in mainstream games. Rapid and precise input has previously not been a good fit for slow voice recognition. After 140ms the mind will no longer perceive one's action to cause a specific reaction. Modern cloud-based ASRs have a system response time that is not consistently below 140ms. The research of this paper relates to the domains of Game Development and Human-Computer Interaction within Computer- and System science.
Problem	The problem is that it is unclear how viable a cloud-based ASR could be as the single control modality in a real-time mobile game as of today. This study is important since voice control could be a suitable or more accessible control modality for some people and/or situations.
Research Question	The research question: How viable is a cloud-based ASR as the single control modality, in a mobile game, in relation to system response time (SRT) and command success rate (CSR)? The research question is interesting as cloud-based ASRs are good candidates to use more in mobile devices for voice interaction. By exploring its viability in this study, the findings may solve the problem as they give an indication as to how cloud-based ASR performs in the defined context.
Method	The approach of this study was empirical research based on empirical investigation through an experiment. Both quantitative and qualitative data was required to answer the research question. With the developed voice-controlled game "Voice Snake" data was to be collected during playtesting sessions. The quantitative data collection method was automatically collected system/game metrics, relating to UP and SRT. Screen-recordings were done to measure the CSR. These data were analyzed through statistical analysis. The qualitative data collection method was semi-structured notes through observation and structured discussions. This data was analyzed with a thematic analysis.
Result	The results showed that the user *does not* experience real-time control of the game, at all times, using cloud-based ASR as the single control modality in the game Voice Snake. It was concluded that the cloud-based ASR was viable up to a certain point, regarding the SRT. The queue-system implemented was concluded to be a way of alleviating the slow SRT and enhance the viability of the voice interactions. Furthermore, implementing feedback about the system's status would have

	enhanced the UX further. Regarding the CSR, the cloud-based ASR could have been viable if the implementing system were closer to, or at, 100% CSR.
Discussion	Two limitations identified, were that the observational notes did not mark the time of their occurrence and the SRT datapoints collected were extremely varied, without any proof to why. Exploring the viability of a cloud-based ASR as a control modality may lead to more inclusive design in applications and video games, in the society. However, by shifting the control modality from a local ASR to a cloud-based may lead to more users' data being collected by big companies. Some valuable findings were that the queue-system supported the participants and alleviated the slow SRT and that findings suggest how sensitive users may be regarding the SRT and CSR in a game such as Voice Snake. This study may be of interest to mobile application and game developers that are looking to include voice control for real-time interaction. By taking the findings of this study in regards, developers can improve their interaction design.

Acknowledgement

We wish to thank our supervisor Thomas Westin for his guidance. Alvin Jude Hari Haran and Gunilla Berndtsson, at Ericsson, for their help and valuable input. As well as our families and friends for helping and supporting us throughout this project.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ASR – Automatic Speech Recognition

CSR – Command Success Rate

HCI – Human-Computer Interaction

SRT – System Response Time

UI – User Interface

UP – User Performance

UX – User Experience

VUI – Voice User Interface

1 Introduction

1.1 Background

Automatic Speech Recognition (ASR) has been around for over half a century but has only during the last decade emerged as an ever more viable technology, as stated by Yu and Deng (2015). The big cloud-based ASRs from, e.g., IBM, Google, and Microsoft, implements deep learning techniques, such as neural networks, are powered by big data and continually keep improving their recognition capabilities, as explained by Yu and Deng. These powerful cloud-based ASRs are today readily available for application developers, offering the inclusion of Voice User Interfaces (VUIs). Locally implemented ASRs, on the other hand, may require more setup, e.g., pre-training to get a large enough database of reference for the recognition. A local ASR also puts extra performance requirements on the device as well, as the computation and storage are relied on it (Erić *et al.*, 2017). Intelligent personal assistants like Google Assistant, Amazon Alexa and Apple Siri make use of cloud-based ASRs and are implemented in a diverse array of small mobile devices used in people's everyday life (Pearl, 2016; Wiggers, 2019), showcasing the potential of the cloud-based ASR. Smartphones have services such as voice dictation and voice dialing which have largely become standard features in many applications. But voice as a *primary* control modality is rarely seen utilized in mainstream games (not counting genres like karaoke and similar music games), but is more often a secondary control modality or just a gimmick (Allison, Carter and Gibbs, 2017; Summa Linguae, 2017; Kiiski, 2020).

Harada, Wobbrock and Landay (2011) state that, for fast paced, real-time interactions, voice as a control modality, is inherently less efficient than traditional physical interaction since the act of thinking what to say and vocalize the words naturally takes more time than, e.g., moving a finger to touch a touchscreen, or a hand to move a computer mouse. However, accessibility is a strong motivator when it comes to incorporating voice as a control modality. Especially for people with motoric impairments in their upper limbs where many prefer or need to use voice as the primary control modality (Bierre *et al.*, 2005; Feng *et al.*, 2011; Naftali and Findlater, 2014). But also, for anyone who has limited hand availability during a certain activity leading to situations where voice interaction could be preferred. Unlike other assistive technologies like eye- or head trackers, a cloud-based ASR does not require elaborate or expensive hardware, only an internet connection and a microphone, making cloud-based ASR a perfect candidate as a control modality in mobile games. This brings about the need for understanding the possibilities and limitations ASRs have when used as a control modality for playing mobile games. Any control modality used for real-time interactions requires a low latency, or System Response Time (SRT), otherwise the User Performance (UP) and User eXperience (UX) get impaired (Kaaresoja, 2016; Attig *et al.*, 2017; Winkler *et al.*, 2020).

There are several studies exploring the use of voice as a control modality in different computer systems and games. Though it seems that many do not state what kind of ASR they utilize and those that do seem to make use of a local ASR or a 'Wizard of Oz' method, i.e., a faked ASR. One study (Scovell *et al.*, 2015) measured the impact SRT and Command Success Rate (CSR) had on the UX while users were giving commands to a tablet using natural language interaction, i.e., conversational speech, with the Wizard of Oz methodology. The study's results indicated that the experience was considered 'good' with SRT up to 4 seconds and with CSR up to 70%. CSR is a metric for measuring

the accuracy of ASR when used for commands and tells the percentage of successfully recognized commands by a system:

$$CSR\ (\%) = 100 - \frac{No.\ of\ Wrongfully\ Executed\ Commands}{No.of\ Reference\ Commands} * 100$$

Some studies have tried to combat this natural limitation of users requiring time to formulate their speech, with systems that use non-speech input in real-time games (Sporka *et al.*, 2006; Harada, Wobbrock and Landay, 2011). Another study explores how emotions in the voice can be used to control a real-time game (Hagerer *et al.*, 2017).

In the study by Harada, Wobbrock and Landay (2011), the authors describe two types of input signals, for a control modality, relevant in gaming: *discrete* and *continuous*. The discrete constitute of separate interactions, e.g., firing a single round from a weapon with one button press. The continuous constitute of one interaction that has a varied length or intensity, e.g., holding down a button to walk forward. Both inputs can range from being non-time-critical, to demanding fast, rapid, and precise executions. The demands put on a control modality by a game heavily depends on what type of game it is (Harada, Wobbrock and Landay, 2011). Turn-based strategy games are typically slow paced and do not require fast inputs of either discrete or continuous nature and are hence probably suitable for voice interactions. While a first-person shooter would require precise and rapid inputs of both variants and hence be hard to play via voice.

A conference paper by Attig *et al.* (2017) show that there are many guidelines on what acceptable thresholds of SRT there are, all with varying limits defined for similar use cases. The authors argue that even SRTs below 100 ms have been shown to be perceivable to users and impact their UP negatively. To identify possible acceptable limits of SRT when used as a control modality for real-time interactions in a game, this study also looks to what the human mind is capable of. According to Johnson (2013) the shortest time a visual stimulus can be shown and still affect the human mind (i.e., the subliminal perception) is 5ms (milliseconds). Johnson further state that after 140ms the mind will no longer perceive one's action as having caused the reaction (i.e., the perception of cause and effect). Hence, a SRT of 140ms could be a possible upper limit for real-time interactions.

A bachelor thesis compared two ASRs, the cloud-based Google Speech and locally-based Pocketsphinx, as the means of controlling a robot in real-time with short commands (Stenman, 2015). The thesis showed that Google Speech was performing with a SRT of 3203ms – 3398ms and Pocketsphinx with a SRT of 145ms – 166ms. A quick test, by the authors of this study, of a cloud-based ASR today in 2021, gives that it performs a lot better than 3300ms, but that it is not consistently below 140ms. However, depending on how fast the game is and by using some techniques, such as utilizing the cloud-based ASRs faster 'hypotheses' results, the cloud-based ASR could still be viable as a control modality in some cases. Furthermore, with the coming of 5G, the network latency is predicted to be less than 1ms at best (Rodriguez, 2015). This would mean that the SRT of cloud-based ASR would predominantly be its recognition speed when used over a mobile network, benefiting mobile games utilizing cloud-based ASR.

1.2 Problem

A suitable or more accessible control modality for some people or in certain situations, could be through voice control. Cloud-based ASRs are becoming ever more ubiquitous and mobile devices are good candidates for implementing cloud-based ASRs. However, the problem is that it is unclear how

viable a cloud-based ASR could be as the single control modality in a real-time mobile game as of today.

By designing and developing a mobile game that utilizes cloud-based ASR as its single control modality, the authors of this thesis set out to investigate the limits of these interactions. The motivation of the study is to provide answers regarding the viability of using a cloud-based ASR for real-time voice control and explore techniques of enhancing this interaction, to support researchers of Human-Computer Interactions regarding VUI and developers looking to implement voice interaction with cloud-based ASRs in applications and games.

1.3 Research question

This study regards the viability of a control modality to be described by two key factors. The user experience (UX) and the user performance (UP). The interactions that the study will investigate are defined as real-time discrete interactions, navigating in a 2D space, using absolute directions. With this in regard, the following research question was formed:

How viable is a cloud-based ASR as the single control modality, in a mobile game, in relation to system response time (SRT) and command success rate (CSR)?

1.4 Ericsson collaboration

For the sake of transparency, the extent of the collaboration with Ericsson will be described. Ericsson provided a general theme for the study, that to investigate how emergent technologies may benefit from 5G. The researchers were provided with information about 5G and networking. Guidelines for the technical setup of the experiment were also provided in the form of open-source tools for Linux. Nothing was extended to the researchers that may create a conflict of interest, such as funding or payment.

2 Extended background

2.1 Human-Computer Interaction and User Experience

Human-Computer-Interaction (HCI) as a topic is simply about people's interaction with computers or technology. Relatable to this study is research into UX when interacting with technology. The interaction with a computer is done via an interface referred to as the User Interface (UI). The goal of this interaction is to facilitate effective control from the user (human) end whilst simultaneously providing information, on the computer end, that aids the user in their decision making. When the user's input, or control modality, is enabled via voice the UI is referred to as Voice User Interface (VUI).

This study's relation to what good UX is will be Shneiderman's *et al.* (2017) description that a good user interface is eliciting positive feelings of success, competence, and mastery among the users. Furthermore, this study will use Heidegger's and Merleau-Ponty's approach to interactivity as described by Svanaes (2013). Svanaes describe the Heidegger approach as the idea that humans are not primarily thinking, planning, and reflecting beings, we are in the environment and act accordingly. Svanaes further describes the Merleau-Ponty approach as the idea of perception as an interaction. That we humans use our whole bodies to interact with the world. These are 'softer' approaches compared to the traditional cognitive science approach which, according to Svanaes, use more 'hard' values in describing an interaction, where the user side is modeled as if it was a computer. One more important prospect is Heidegger's analysis of tools use. In short, Svanaes (2013) describe this view as when a tool is designed well, it does not require the user to think about how to use it. The interactions should flow and if a breakdown occurs the tool should be enough transparent in use that the problem can easily be fixed.

2.2 Voice User Interface

In a conference paper, Aylett *et al.* (2014) remarked that speech technology is comparatively marginalized, calling researchers to do more research with the technology. The authors argue that voice interaction can play a significant role in eyes-free and hands-free interaction and that it is relevant on small mobile devices. The authors further predicted that ASR would become increasingly ubiquitous and identified the challenges of VUIs as: allowing for multi-tasking, user interruption and producing acceptable latencies. In another conference paper, Munteanu *et al.* (2017) also stated in their Abstract that "very little HCI attention has been dedicated to designing and developing spoken language, acoustic-based, or multimodal interaction techniques, especially for mobile and wearable devices." (p. 601) A journal article (Clark, Philip Doyle, *et al.*, 2019) doing a review of several research papers, regarding VUIs in HCI, further showed that a majority of the reviewed studies focused on desktops or laptops as a device context. Only 6 out of 68 were in a device context of mobile or intelligent personal assistant. Indicating that research on VUIs in HCI may have been trending towards desktop use.

A journal article (Allison, Carter and Gibbs, 2017) reviewed how voice interaction has been used in main stream games since 1970. The article argues that voice interaction in gaming has been coming and going but that its use is stabilizing and, in the future, might become more ubiquitous as the technology is becoming more viable and accessible to players and developers. When it comes to voice interaction in games another journal article (Allison *et al.*, 2018) constructed an initial language model describing different game design patterns concerning voice interaction. Using that model, this study can define the voice interaction game design pattern of interest as being the category 'Navigation' which consist of three different patterns 'Waypoints', 'Absolute directions', and 'Relative directions'. The pattern used in the study's developed game is 'Absolute directions'.

2.3 Accessibility in games

With the Entertainment Software Association reporting that approximately 21% of the total amount of gamers in the US have some form of disability (ESA, 2020) we can imagine there is a large portion of the global gaming population that have a disability as well. Making games accessible an important subject. The disability mostly targeted by this study is motoric or mobility disabilities. Data from the American Centers for Disease Control and Prevention, shows that out of every adult American with some form of disability, 12.8% where considered a mobility disability, which is one of the more prevalent disabilities (CDC, 2019).

The accessibility of games, mainstream games in particular, has long been overlooked as expressed by several authors when compared to the accessibility of other software technologies (Bierre *et al.*, 2005; Miesenberger *et al.*, 2008; Yuan, Folmer and Harris, 2011; Cairns *et al.*, 2019). One reason for that might be because of the help other software technologies get from governmental laws that regulates how workplaces and public services need to be accessible for people with disabilities. For example Section 508 in the US, associated with the requirement of Federal agencies to make their electronic and information technology accessible to people with disabilities (GSA, 2020). Among other things, this means that Federal websites has to be made accessible to people with disabilities and when the law came into rule in 1998 a response from W3C (World Wide Web Consortium) was to developed standards for accessible websites (W3C, 1999; Bierre *et al.*, 2005).

In recent years however, there has been an increased motivation for accessibility in games. As stated by Cairns *et al.* (2019), an explanation might be that more people with disabilities want to play games and that there are growing legal requirements for accessibility in products and services as well. With, for example, the European Accessibility Act being adopted in 2019 (EUR-Lex, 2019). Many of the recent big contributions to accessibility in mainstream games have mainly been for games on computers and consoles. With examples such as Microsoft releasing the "Xbox Adaptive Controller" in 2018, a low-cost, commercial product designed specifically for players with disabilities (Warren, 2018). Electronic Arts launching their accessibility portal in 2018 catered to guiding disabled gamers regarding accessibility options in their games (Key, 2018). Naughty Dog's release of "The Last of Us II" in 2020 (Gallant, 2020) which was regarded as one of the most accessible games in the mainstream market at its release, with more than 60 different accessibility settings (Carter and Molloy, 2020).

2.4 Accessibility in mobile games and applications

Digital games have grown into one of the biggest sources of entertainment globally and according to Newzoo (2020) the global games market will generate revenues of $159.3 billion in 2020, and is estimated to continue growing. According to Newzoo, mobile gaming (including both smartphone and tablet) accounts for $77.2 billion, constituting almost 50% of the revenues of the market. Newzoo further estimates that there will be a total of 2.7 billion players of digital games worldwide by the end of 2020.

Considering how big the market of mobile gaming is, there has so far been a surprisingly low motivation in making mobile applications accessible. An example is the study by Ballantyne *et al.* (2018) where the researchers compiled a list of guidelines for mobile application accessibility and scrutinized the top 25 popular applications on the android market "Google Play". They concluded that most applications could be accessible at the system level but are largely inaccessible at the usage level due to poor design and content. Another study, by Wilson and Crabb (2018), stated that, in regards to the World Wide Web Consortium (W3C) accessibility guidelines, very little attention has been given to mobile phone games. One of the aims of their study was to examine how well the W3C accessibility guidelines support the creation of accessible mobile games by interviewing six students of Computer Science or Digital Media. Wilson and Crabb concluded that the W3C WCAG (Web Content Accessibility Guidelines) can be used to assist in the development of accessible mobile games but that there is room for improvement for as to cover all accessibility issues players may experience in mobile games. However, researchers are attempting to tackle all these issues (Krainz, Miesenberger and Feiner, 2018; Westin *et al.*, 2018).

2.5 5G

The typically main 5G service types considered are (Marsch *et al.*, 2018):

- Enhanced mobile broadband (eMBB)
  Related to enhanced access to multiple media content, services, and data which could, for example, enable a more ubiquitous usage of virtual and augmented reality.
- Massive machine-type communications (mMTC)
  Related to services providing better communications between machines which could, for example, enhance the "Internet of Things".
- Ultra-reliable and low-latency communications (URLLC)
  Related to user-cases with demanding requirements for capabilities such as, latency, reliability, and availability.

When measured in the real world, the round-trip latencies (or SRT without computational time) of different 4G networks are around 40-70ms (Fogg, 2019; Kota, 2019; Statista, 2020). Technically, the fastest round-trip latency possible of the 4G system LTE-A is around 20ms and is expected to diminish to less than 1ms for 5G (Rodriguez, 2015). Such a fast round-trip latency opens a lot of doors, for example the "Tactile Internet" is expected to become viable with 5G (Gupta *et al.*, 2019).

To avoid the hop latency of underlying networks completely, 5G is also often discussed in conjunction with Edge Computing. The idea of moving cloud computing off central servers and onto servers closer

to the client (Mach and Becvar, 2017). This can be seen as a middle ground between keeping complex tasks on the cloud and on the user device. Cloud services reduces the load on client devices while at the same time letting consumers use complex and resource intensive applications, such as ASRs. The edge computing decreases the latency by being closer to the client while, perhaps, sacrificing some computational power.

2.6 Automatic Speech Recognition

The idea of ASR has its origins in dictation machines. These devices automated the task of recording notes and letters without requiring pen and paper. Automatic transcription was the idea that directly followed these machines. An ASR system introduced in 1952 used the theory of phonetics, the basic sounds of any specific language, to record phonetic resonance patterns and distinguish between different vowels. During the early 70's the U.S. Department of Defense funded a research program which resulted in one system that could recognize speech with a vocabulary of 1011 words (Juang and Rabiner, 2005).

A more modern approach is to implement machine learning techniques which were applied to ASRs as early as the late 1980s. One system developed in the early 1990s which used a machine learning technique called Deep Neural Network was able to perform with a 11-17% lower error rate when compared to conventional speech recognition methods (Yu and Deng, 2015).
The development of ASRs that implement machine learning techniques has continued to this day, and big tech companies such as Microsoft and Google now provide cloud-based ASRs which are available to both companies as well as individual developers.

Common metrics, other than CSR, of ASRs recognition accuracy are Word-Error-Rate (WER) and Single-Word-Error-Rate (SWER) (Karpagavalli and Chandra, 2016). Simply put, WER is a calculation of the percentage of errors in the recognition, based on the number of *insertions*, *substitutions* and *deletions* compared to the number of reference words:

$$WER\ (\%) = \frac{Insertions\ +\ Substitutions\ +\ Deletions}{No.\ of\ Reference\ Words} * 100$$

While SWER tells the percentage of incorrect recognitions for each different *word* in the system vocabulary:

$$SWER\ (\%) = \frac{No.\ of\ Incorrectly\ Recognized\ Words}{No.\ of\ Reference\ Words} * 100$$

3 Methodology

3.1 Research strategies

The approach for this study was that of an empirical research aimed at answering the research question based on empirical investigation conducted in an experiment. To accomplish this, a version of the game "Snake" was developed. Where the single control modality was voice commands recognized by a cloud-based ASR. This developed game is referred to as "Voice Snake" in this study. This was a mobile game that would require real-time interaction to successfully control the snake's movement in two dimensions. The score of collected fruits was to be considered the UP. It was determined that the research question could be answered by investigating players' interactions with the game in playtesting sessions. As explained by Pozzi and Zimmerman (2016), "playtesting is a methodology borrowed from game design where unfinished projects are tested on an audience" (p.1). This method is used since playtesting is considered useful when studying the interaction between a created experience and a user, as stated by Pozzi and Zimmerman.

The playtests were decided to be conducted with the research strategy of experiment. The motivation for choosing experiment was that experiment is considered an empirical investigation under controlled conditions where the properties and relationships between factors can be examined (Denscombe, 2014). Experiment is thus closely resembling how this study aims to answer the research question, with several specific variables to be examined in relation to each other. The exact experiment design is described in section 4.1.1 Experiment design. Following in this section are arguments for the experiment's suitability to this study.

Denscombe (2014) points out five conditions that need to be met for experiments to be suitable in research. (1) Experiments are to be used for explanatory research rather than exploratory. (2) Research should be drawn on well-established theory. (3) A hypothesis should be formed with the existing knowledge. (4) An experiment should produce quantitative data. (5) An experiment should also have the ability to manipulate variables and implement controls.

How were the experiment considered to fulfill these conditions? By (1) looking at the research question the study could be defined as explanatory as it is aimed to explain the viability of a cloud-based ASR, in a set context. (2) This study relies on previous research and theories which have explained the relevant variables as UX, UP, SRT and CSR. These factors play a role in describing the viability of ASR and were used to form the research question. (3) The forming of this study's work hypothesis was based on the research question and in turn the existing knowledge. (4) The data produced in this experiment was quantitative, but also qualitative. (5) The experiment was designed with the ability to manipulate the SRT and included a control group. Thus, all conditions are considered fulfilled.

Denscombe (2014) also describes what an experiment should include to be considered a "true" experiment: a pre-test and a post-test to get measurements before and after the independent variable has been altered; a control group for comparison; random allocation of people to the two groups and the control of a variable whose impact the experiment will investigate. The experiment was designed to include this and thus regarded as a true experiment.

3.2 Methods

3.2.1 Data collection

To acquire data on the UP, SRT and CSR quantitative data was to be *automatically collected metrics* related to these variables. Data regarding SRT and UP was to be saved by the developed game during play. Furthermore, a *screen recording*, including the microphone input, was to be saved with the data needed to calculate the CSR. Experiments are described to mainly produce quantitative data (Denscombe, 2014). However, this experiment design incorporated the ideas of playtesting and produced both quantitative and qualitative data.

According to Fullerton (2018), there are many different ways of playtesting, some of which are more informal and qualitative, and others more structured and quantitative. But that they all have the common goal of gaining feedback to improve the experience of the game. This feedback can be regarded as the UX. There are ways of collecting UX metrics that are quantifiable (Tullis and Albert, 2013). However, the types of collection methods described for "self-reported" metrics (i.e., when the participants themselves tell about their experience) are mostly via forms, questionnaires, rating scales or similar. The motivations for not opting these kinds of methods are discussed in section 3.2.3 Alternative methods.

The playtests were designed to follow the data collection methods described by Fullerton (2018), that in summary consists of observations and discussions. The qualitative data related to UX was collected with *semi-structured notes* through *observation*, and *structured discussions*. By taking notes the observer are less likely to miss crucial details of the playtesters' reactions and to control the impulse to talk too much one can use a test script (Fullerton, 2018). (The test script used is in Appendix A.) The notetaking was done by two researchers and therefore considered "group notetaking". As explained by Farrell (2017), each individual writes down observations during the session and then, after each session, the main findings are summarized as a group activity. Further following Farrell's and Fullerton's descriptions for notetaking, the technique applied was in part that of "chronological logs" and in part "free form", hence in the study called semi-structured. This be put in parallel to Denscombe's (2014) description of "systematic observation", that structured notes minimize variations in the data arising from factors influencing the individual researchers' perception of events.

Continuing with Fullerton's (2018) descriptions of discussions, there are two approaches; freeform or structured. This study opted for a structured approach. Where the participants were free to talk about what they want, but the researchers had questions meant to guide the discussions. Hence referred to as *structured discussions*. Parallels can be drawn to Denscombe's (2014) descriptions of semi-structured interviews, where the researcher has a clear list of issues/questions to be addressed while still being flexible on the topic. According to Denscombe, the advantage of this approach is that it can develop and change through the course of the study and rather than keeping every interview the same, new lines of enquiry can be followed up. The difference between interviews and discussions is that interviews are more of a verbal quiz (Fullerton, 2018).

3.2.2 Data analysis

A *thematic analysis* was chosen to analyze the qualitative data and investigate what the major UX for every playtest session was. Braun and Clarke (2006) regard thematic analysis as the foundational method of qualitative analysis and one of the benefits of thematic analysis is its flexibility. The motive for choosing this method is that Braun and Clarke describe it as a useful tool with the potential of

providing a rich and detailed account of complex data. A *deductive* (also called theoretical) and *latent level* (also called interpretative) approach was used as described by Braun and Clarke (2006). Deductive as there was a specific research question coded for and the themes of interest was already identified. Latent in order to find a deeper meaning in the data that could be related to the themes of interest.

The quantitative data was numerical and therefore lent itself to be analyzed through *statistical analysis* (Denscombe, 2014). Denscombe (2014) describe two types of approaches to statistical analysis, one using *inferential* statistics and one using *descriptive* statistics. Inferential covers more advanced procedures that, e.g., predict characteristics of the population as a whole (Willard, 2020). While descriptive statistics can be used to describe the data, explore connections between the data and summarize the findings (Denscombe, 2014). Following Denscombe's reasoning, by looking for patterns and relationships in this descriptive profile it was argued possible to answer the research question with descriptive statistics.

3.2.3 Alternative methods

If there already existed a suitable game that could be used to answer our research question, an alternative approach could be *case study*. As Denscombe (2014) points out "… one strength of the case study approach is that it allows the use of a variety of methods depending on the circumstances and the specific needs of the situation" (p.56). In our case we could examine players interactions of this already existing game that make use of a cloud-based ASR as a control modality for real-time interaction by the means of, for example, observational data collection. The case study approach was discarded due to the low availability of natural settings where data could be collected from. I.e., no appropriate game could be found, that could fit our context.

An alternative data collection method could be through a *form* or a *questionnaire*. After the playtest, the participants could be able to fill out a questionnaire about their UX. The main issue with these type of methods, is the lack of depth and detail in the qualitative responses (Denscombe, 2014). Drawing a parallel with interviews, discussion will have a higher likelihood of producing a deeper and more detailed data (Denscombe, 2014). This study regards UX as something complex that requires qualitative data to describe and therefore discarded the alternative of questionnaires.

3.2.4 Ethics

Denscombe (2014) describes four key principles regarding established research ethics for social researchers: protect the interests of the participants; participation must be voluntary and based on informed consent; researchers should operate in an honest way with scientific integrity; and comply with local laws and regulations.

The investigation must protect the interest of the participants, the biggest issue for this study's approach was its non-digital nature and the risk of transmissions of Covid19. To minimize this risk, which could cause harm to the participants, the study made sure to follow certain safety guidelines. (Details are described under the Method application, 4.5 Ethical consideration.) An alternative way that would remove this risk would be to use *online survey* as a research strategy, using online playtesting and discussions held online. However, with proper safety precautions, following safety guidelines and a convenience selection of participants the risk was deemed to be minimal.

A risk with discussions (as well as other interview methods) is that tactless interviewing can be seen as an invasion of privacy (Denscombe, 2014). I.e., the risk of data being used to identify a participant, or

that questions are upsetting for the participant. However, these risks were deemed exceptionally low as the theme of posed questions only regards the user's experience with the game. Furthermore, if any identifiable responses were to be found they were to be edited and anonymized.

Voice, a biometric data, is considered personal data under the protection of the General Data Protection Regulation in the EU (Mediartis, 2019). This study recorded the voice of the participants in screen-recordings during the experiment. How the data were to be handled, and used, had to be discussed with the participants and their consent collected with the consent form (Appendix B). Furthermore, voice data is sent to the cloud-based ASR. The study failed in finding information about how the selected cloud-based ASR collects and saves this data. However, the account connected to the cloud-based ASR was one of the researcher's and thus any data sent would not as easily be connected to the natural person of the participants. This ethical concern was discussed with the participants before the experiment was conducted and is also described in the consent form. To ensure that participation in the study was voluntary and based on informed consent, the participants would have to sign the consent form (Appendix B) before the experiment began. This also avoided deception and ensured that the study operated with scientific integrity.

4 Method application

Based on previous research, it was predicted that real-time interaction via voice will *not* be achieved, if the SRT is too slow in relation to the speed of the game. To investigate the research question, the following work hypothesis was specified:

H_0: The user experience real-time control of the game, at all times, using cloud-based ASR as the single control modality in the game "Voice Snake".

H_1: The user *does not* experience real-time control of the game, at all times, using cloud-based ASR as the single control modality in the game "Voice Snake".

4.1 Data collection

4.1.1 Experiment design

The experiment was designed as playtesting sessions held one-on-one with the participants. As in one participant and one group of two researchers. The independent variable was SRT, and the dependent variables were UP and UX. There was one experimental group and one control group. Both groups did two playtesting sessions, where the experimental group had the SRT increased by 70 ms in their second session. The SRT of the control group was not manipulated. The experiment was held in lecture room at Stockholm University DSV. The only people in the room were the two researchers and the participant. No disturbances, e.g., loud noises, occurred during the experiments. The researchers followed the playtest script in Appendix A, for each playtest. The process is described below.

The playtest sessions were conducted with one participant at a time with both researchers of this thesis present. After the participant was informed of what they were expected to do, they were allowed to try out the game for approximately five minutes to warm up and get acquainted with the game. After the warm-up, the playtesting began, and the data collection started. A screen- and microphone recording was started, and while the participant was playing the game, any notable observations were written down by the researchers. Concurrently, the system was saving metrics at each command and at game over. After about 12 minutes the participant was allowed to finish the current round and was then told to exit the game. The participant was then asked to discuss their experiences while the researchers wrote down their thoughts. The pre-defined questions were used to guide the discussion if the participants had not already answered them. When answers to the pre-defined questions were exhausted and the participant did not have anything more to say, the researchers' notes were verified by the participant.

Following was a second playtest session with the participant. The second playtest session was identical to the first. Unless the participant was in the experimental group. Then 70ms of latency was added to the local network, increasing the SRT of the second session. The exact same data collection methods were applied again and followed by a second discussion with the same structure as the first. When the participant had left the room, the researchers merged and summarized their notes as a group activity.

4.1.2 Voice Snake description

Voice Snake was designed as such:

- when a round starts, the snake continually moves forward by itself and the direction of the snake's forward movement is controlled by voice interaction, using absolute directions, via cloud-based ASR;
- there is always one apple somewhere on the playfield and when the apple is eaten a new is spawned randomly, on a vacant square, on the 9x16 playfield;
- at the start of the round the snake moves 1 grid space per second (i.e., a step-delay of 1000 ms);
- for every apple eaten the snake moves faster by 0.05 seconds (i.e., step-delay decreases 50 ms);
- the goal is to collect as many fruits as possible (the score);
- there are no "walls" (i.e., moving outside the playfield results in an entrance on the opposite side of the playfield);
- the snake dies if it bites into its own body (game over).

Hence, the difficulty *to control* the snake continuously increases with every collected fruit. The difficulty *to survive* spikes at 9 collected fruits as the snake becomes larger than the vertical space and furthermore at 16 as the snake becomes larger than the horizontal space.

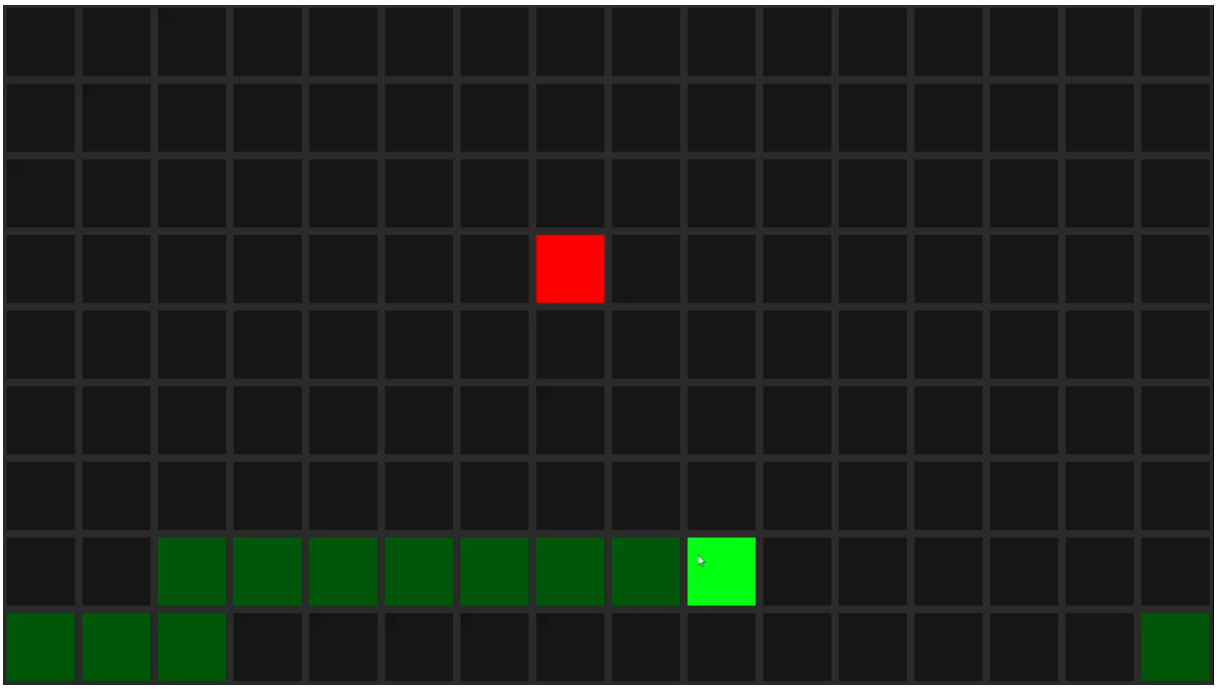

Figure 1 *Screenshot of the constructed game. Showing the 9x16 playfield, the green snake and the red fruit*

Table 1 *How the step delay decreased with every score of collected fruit*

Score (collected fruits)	Step delay (milliseconds)	Score (collected fruits)	Step delay (milliseconds)
0	1000	9	550
1	950	10	500
2	900	11	450
3	850	12	400
4	800	13	350
5	750	14	300
6	700	15	250
7	650	16	200
8	600	17	150

Note: post a score of 17 the step delay was always 100 ms.

4.1.3 Technical setup

The public GitHub repository containing the Unity project is available at the following address: https://github.com/kimcodekill/VoiceSnake

The chosen cloud-based ASR was the Microsoft Azure Cognitive Services Speech API. It was implemented using Microsoft's Cognitive Services package for Unity and set up per the quick start documentation. When a word was recognized, it was looked up in a string/Action dictionary. If the key existed, the action for that key was invoked which changed the direction of the snake. The possible commands were absolute directions in relation to the computer screen. The commands were Up, Down, Left and, Right. When the system was tested, it became clear that waiting for a result took too much time. To get past this issue it was decided that catching recognition hypotheses would be a simple way to speed up the response of the system. This was also combined with a list of similar keywords e.g., down/done, or right/write to make the system act on guesses that were close but not exact. This could be done since the base keywords (up/right/down/left) were clearly distinct from each other. If a distinction between down/done was needed, this technique would not work as effectively.

Furthermore, it became clear that if words were spoken within a brief time span of each other, the recognizer would bundle them together. As an example, if the words stated were "Up, Left, Down" the recognizer would first return "Up", then "Up, Left", and finally "Up, Left, Down". This led to only the first word in the phrase being handled. To deal with this, these phrases were split, and if there was more than one word, the final word was used. This let phrases of unlimited length be captured and each word in the phrase handled one after another in a queue. The common name for this became "queueing words" and is referred to as the "queue-system".

To simulate latency, a Raspberry PI was used as a network bridge between the testing computer and the internet. The Raspberry PI was set up with the tool called Traffic Control, commonly referred to as "TC". This tool made it possible to add any amount of latency between the client and the server. The

scripts for setting up the network bridge and adding/clearing latency were set up in advance to minimize the risk of issues during the experiment.

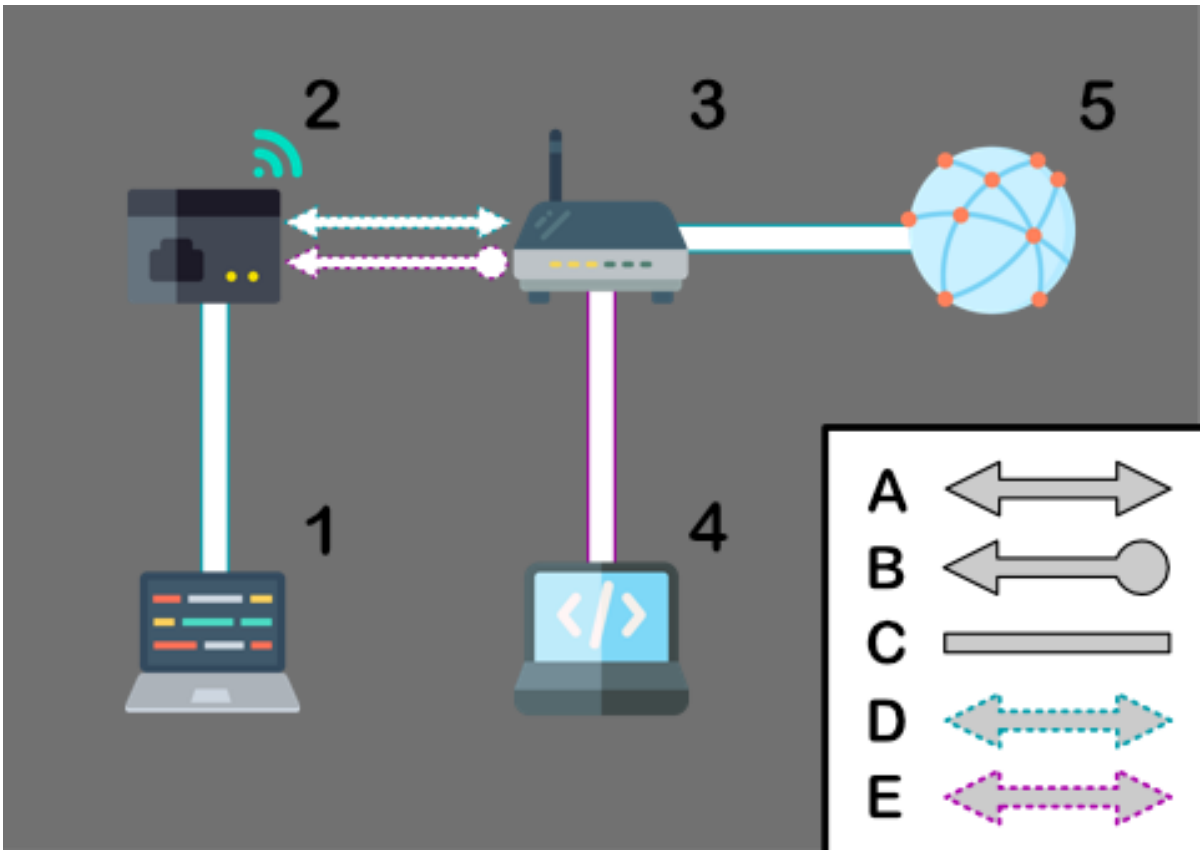

Figure 2 *Experiment Setup*

1 – Windows laptop for gameplay, see (D) for data color

2 – Raspberry Pi for latency spoofing

3 – Generic router as device hub, connected to the Internet (5)

4 – Windows laptop for controlling Raspberry Pi (2), see (E) for data color

5 – Connection to Azure Cloud Services

A – Back and forth communication

B – One-way communication

C – Wired communication (A)

D – Wireless communication (A) with color of game laptop data (1)

E – Wireless communication (A) with color of controller laptop (4), also type (B)

D & E – Color also present for type (C)

4.1.4 Collecting data

During each one-on-one playtest, qualitative data related to UX and quantitative data related to UP, SRTs and CSR was collected. The qualitative data were collected using observational semi-structured notes taken by the researchers during the playtest following the notes sheet in Appendix C. Post playtest, during structured discussions data were collected through notetaking. During each playtest

session, there were two researchers taking notes individually. Their notes were compared after the experiment and the findings were summarized as a group activity.

The quantitative data were collected in two ways. One part was metrics saved by the game into a text file that mainly consisted of the SRT, the command, and the snake's step-delay for every interaction. Furthermore, for each finished round, it saved the total play time and total number of fruits collected (the score). The other part of the quantitative data was in the form of screen-recordings of the playtests, including the input from the microphone, to measure the CSR. This recording was done with the program OBS Studio version 26-1-1.

4.2 Data analysis

4.2.1 Thematic analysis on observation and discussion notes

The raw data consisted of observational notes and discussion notes taken during the experiment. Both authors took notes individually during the experiment. With 6 participants playing 2 sessions each, this resulted in 24 handwritten documents. As a group activity, the notes were merged, rewritten in a digital document, cleaned up, and sorted by group and session, resulting in 4 documents: Experiment group Session A + B and Control group Session A + B. These 4 digital documents were imported to the online coding tool eMargin (emargin.bcu.ac.uk).

This data was analyzed using a thematic analysis following the six steps defined by Braun and Clarke (2006) to identify how three major themes of the UX manifested for every playtest session. These steps were iterated back and forth. A deductive and latent level approach was used, as defined by Braun and Clarke, during the coding process. Deductive in the way that there were three predetermined themes of interest to code for: *positive-*, *neutral-* and *negative UX*. Latent in the way that underlying concepts in the data was looked for. The two authors of this thesis did a first coding individually and compared the result to see how reliable the themes were, this showed that the coding for these themes was, to a big degree, similar. Both authors identified that the codes could be further categorized under their respective themes to better describe the data. It was also identified that the data was too similar between the control and experiment groups, as well as between individual sessions to single out any discrepancies. Hence the continued analysis was applied to the data as a whole.

With a combination of inductive and deductive approach, three basic categories were formed: Feelings, System Behaviors and Naturality of the Interactions which were split under the three themes, resulting in nine categories. Inductively as several categories were first identified in the data during the coding. Deductively as the categories were condensed into those three final categories based on their relevance to UX and this thesis. As described in section 2 Extended background, good UX can be seen as eliciting good feelings, hence feelings were coded and put in their relevant categories and themes. E.g., signs of frustration were identified as "Negative feelings" and "Negative UX". Furthermore, to identify the naturality of the interactions, the approach to interactivity as described in 2.1 Human-Computer Interaction and User Experience. Where, in short, a good interaction should be as natural as possible so that the user does not have to reflect on it. Hence, e.g., signs of a broken "flow" was identified as "Unnatural interactions" in the "Negative UX". The category of the system's behaviors identifies the computer side of the interactions. E.g., codes saying that the SRT was good were identified as "System behaviors supporting/enhancing the interactions" in the "Positive UX". A second coding was done, by one of the authors, with regards to the categories and then reviewed by

the other author. The theme ‘Neutral UX’ was the least prominent of the three and did not describe the data very well regarding UX and was thus discarded. Similar codes were also merged.

4.2.2 Statistical analysis on system metrics

The raw data was indexed by giving it a unique serial number connected with the participant, experiment/control group and session (P1A, P1B, P2A, etc.). The xml data files produced by the game and the saved screen-recordings were transferred to a stationary computer, via cable. The screen recordings were manually screened by counting the stated commands as well as determining which ones were successful, and which ones were not. The CSR could then be calculated along the formula presented in section 1.1 Background. When the CSR had been calculated, it was double checked for errors and then spliced into the xml data files. The xml files were then combined to one single xml file containing all participant data, imported to excel, and finally imported to the statistical analysis program SPSS.

The researchers familiarized themselves with the data and handled extremes or errors found as follows. Data for two rounds were removed completely as the screen recording had not started, hence did not provide a CSR. These rounds had been marked to be removed during the experiment. Some SRT datapoints had negative, and zero, values, hence regarded as invalid and replaced with null values. SPSS was configured to ignore null values so they would not interfere with statistical tests.

Statistical analysis was conducted on the automatically saved data, with descriptive statistics describing and comparing the data. Correlation analysis was also used to investigate the correlation between CSR and score.

4.3 Verifying the data

4.3.1 Qualitative

This section relates to Denscombe's (2014) explanations of credibility (validity) and dependability (reliability) in qualitative research. Qualitative data (in this case UX) are subjective and hard to verify scientifically as being the “truth” (Lincoln and Guba, 1985 cited in Denscombe, 2014). Concerning the credibility, the accuracy of the observational- and discussion notes taken by the researchers were verified by each participant at the end of each playtest. The findings have been grounded in the empirical data and drawn conclusions in the section 6 Discussion and conclusion are based on the data. In the section 6 Discussion and conclusion, the results of the thematic analysis is supported with the results of the statistical analysis. The confidence of the qualitative data is hence bolstered by indicating the data are ‘on the right lines’, referred to as “triangulation” by Denscombe (2014).

Concerning dependability, the research process of this study has been presented in detail and should be possible to replicate for other researchers. With the selection of participants described in section 4.3 Selection and limitations readers can assess the transferability of the findings. Concerning confirmability (or objectivity), the researchers of this study acknowledge that our identities, values, and beliefs play a role in the analysis process. The researchers have been conscious about this, trying to distance themselves in order to be objective and have an open mind.

4.3.2 Quantitative

Concerning the quantitative data, the validity was enhanced by the following. The quantitative data were collected and saved automatically by the game and structured in such a way that it was easily imported to a statistical analytic software, which avoided manual data entry errors. The data was screened for extremes and errors which were dealt with, in a case-by-case basis (described in section 4.2.2 Statistical analysis on system metrics). The CSR count of the screen recording was done by two researchers to minimize errors in the measurement. The experiment used a simulation of the 4G and 5G networks latencies which enhances the external validity of the data. By using real instances of mobile networks there may be unknown variables that could affect the results, such as high/low internet traffic.

The reliability of the quantitative data was also enhanced by the automatic data collection system; the collection method will hence be consistent across multiple occasions. An unstable network connection of the area where the experiment was conducted could have negatively impacted the reliability of the SRT data. However, during the experiment, the wired connection was confirmed to be stable with repeated domain ping tests to “google.com”. The tests had an acceptable range of 1-3ms latency. The reliability of the UP variable was highly dependent on the skill level of the users. The participants of this study were required to confirm that they had experience with computers and gaming before they were selected to participate. This enhanced the chance that they would perform similarly to one another.

4.4 Selection and limitations

Based on Denscombe's (2014) principles of sampling, the study was identified to benefit more from an *exploratory* sample of the population. By selecting people for the sample based on their experience with technology and gaming the data collected was more likely to cover the aspects of interest and not, e.g., social, cultural, or educational aspects. People with no previous experience using technology or playing games could have corrupted the experiment results. Since the selection was influenced by the researchers the approach of the sampling cannot be considered completely random, hence the sampling approach is called *non-probability sampling* (Denscombe, 2014). The selected group of participants was adult students of computer sciences at the Stockholm University. Out of the sampling techniques described by Denscombe, regarding non-probability sampling, it was decided that the sampling technique used would be *convenience sampling*. The main reason for this was the ongoing pandemic of Covid-19 and it was deemed important to minimize the amount of travel people have to do to get to the location of the experiment. To minimize the risk of extraneous variables affecting the experiment, the experiment took place one on one, and all sessions occurred on the same day. Therefore, the selection of the participants was those that were available at the time of the planned experiment.

Regarding the size of the sample, Denscombe (2014) describes three different approaches: the statistical, pragmatic and cumulative approach. Because of the limited time and resources of the study and the ongoing Covid-19 pandemic, the sample size of the participants for the experiment had to be kept small out of convenience. The approach to the sample size of this study is therefore considered *pragmatic*. However, this does not necessarily reflect negatively on the results. As Denscombe (2014) states,

> … the size of exploratory samples generally reflects the fact that researchers want to probe deeper than they would do with representative samples. Second, the size of exploratory samples is not governed by matters of accuracy but by considerations of how informative the sample is. (p.46)

Since it was of interest to this study to gain deep understanding of UXs it can be considered an advantage to keep the sample size small. Tullis and Albert (2013) further state that they believe a sample size of 5-10 is acceptable when the scope of a usability test is not that big. Tullis and Albert point out that some research suggests that about 80% of usability issues will be observed with the first five participants, if the test is not too big and the system/interface not too complex. With this in regard, it was determined that the sample size would consist of six people. The six chosen participants were all Swedish residents within the age of 20-30. Five of the participants were male and one female.

Regarding the choice of the cloud-based ASR it was determined that Microsoft Azure would be the easiest to implement in the designed game. The study was limited to only testing one ASR and no other ASR was tested. The utilized mobile device during the experiment was a laptop/tablet hybrid of the model Microsoft Surface Book 2 and no other device was tested.

4.5 Ethical consideration

The ethics of social research as Denscombe (2014) describes them is structured along four principles of conduct, discussed in the section 3.2.4 Ethics. During the study, Denscombe's principles were followed in general by having 3rd party oversight (the study supervisor), by operating with full transparency, by conducting the experiment with informed consent of the participants, and by keeping their identities anonymous. Lastly by also complying with the local laws and recommendations regarding Covid-19.

Nothing in the planned experiment was identified as harmful to the participants. The biggest issue was the ongoing pandemic of Covid-19. In accordance with the principle of "following the laws of the country" the study complied with the regulations and guidelines of the Public Health Agency of Sweden (Folkhälsomyndigheten, 2021), as they were in March 2021. In practice, the physical playtesting provided adequate distance between participants and researchers, all individuals present in the room were provided with masks and required to always wear them, and anything that was needed to be handed between the researchers and participants was disinfected. Furthermore, the playtesting was held one-on-one with the participants, to minimize close contact between people.

Before the experiment begun, the agreement to participate in the study was documented with the consent form in Appendix B. Each participant was given a pseudonym in the raw data and used for reference in the study. The raw data collected was not kept together with information that could identify the participants (e.g., the consent forms). The raw data was digitalized, uploaded, and stored on the Stockholm University's internal network SciPro. The digitalized signed consent forms were also, separately, uploaded to SciPro and immediately deleted from the local computers. At the end of the study the digital raw data and signed consent forms were downloaded from SciPro and archived on a local hard drive and deleted from SciPro. The original signed consent forms were shredded and trashed, and any raw data stored on the researchers' computers were deleted.

5 Result and analysis

5.1 Thematic analysis

During the analysis it was identified that, the observational notes would have been more reliable if they marked how "far" into the round of the game an observation was made. The notes only had a chronological order from start to finish of the playtest sessions. However, with statements by the participants, it was still possible to show when/how the UX was changing during a round. Furthermore, the analysis could not identify any conclusive discrepancies between the groups, nor between the sessions. Hence, the analysis was applied to the dataset as a whole. I.e., the analysis could not say anything conclusive about the effect the manipulated SRT had on the participants. The dataset can be found in Appendix D.

Table 2 *Overview of themes and categories*

Theme	Category
Negative UX	Negative Feelings
	System behaviors hindering/worsening the interactions
	Unnatural interactions
Positive UX	Positive Feelings
	System behaviors supporting/enhancing the interactions
	Natural interactions

A full table of the themes, categories, and codes can be found in Appendix E. Three themes had been identified beforehand and nine categories that described the dataset were generated during the initial coding process. The theme 'Neutral UX' was discarded during the final steps for lacking relevance. 'Negative UX' was the most prominent followed closely by the 'Positive UX'. When comparing the observation and discussion notes, the discussion notes were showing more positive UX and the observational data suggested a lot of negative UX. This may be a sign of the 'social desirability bias (Nancarrow & Brace, 2000 cited in Tullis and Albert, 2013), i.e., that the participants are inclined to give positive feedback, which was taken in regard during the analysis process. However, self-reported data is valuable information when it comes to users' perception of an interaction and how they feel about the system (Tullis and Albert, 2013). Another reason for this could be that it is naturally easier to spot when something is wrong compared to when things are running smoothly while observing the playtest. Hence resulting in more observational notes about negative UX.

The excerpts presented in the analysis are not necessarily the full responses provided by the participants but have been selected to describe the themes and categories. Furthermore, the excerpts have been translated into English from Swedish, by the authors, and, in many cases, given a more natural language as they are notes and not transcripts. Any spelling and grammar errors have also been corrected.

5.1.1 Negative UX

This theme was predominant in the observational notes. Under the category ‘Negative Feelings’ there was a lot of frustration observed with notes such as “Frustrated”, “Sighing and scratching in frustration”, and “Shaking head when interaction goes wrong”. Frustration was also explicitly expressed in one discussion:

> “It was frustrating to have to think ahead, but you learned, after a while, approximately how far ahead you had to speak.” (P2)

The participant expressed frustration in having to deal with the slow SRT of the game, as it required the participant to say the voice commands much earlier than what was natural. The participant continued to state that this was something he learned to deal with, however.

Which brings us to the category of ‘Unnatural Interactions’. Two major parts of this category was discussion notes about the participants having to plan their interactions, together with observational notes about mental confusions while interacting. As seen above, P2 talked about having to plan his interactions. Another said something similar:

> “You had enough time to execute a command after you learned how many steps before you had to act.” (P1)

The participant stated that, after he learned how long the delay was, he felt he had enough time to execute the commands as he knew how many steps the snake would take before it executed the command. By having to think a lot about their interactions, it was observed that the participants showed a lot of confusion when saying their commands. Examples of observational notes are: “Said wrong direction”, “Meant down, said left”, “Said left, meant down”, and “A lot of commands when it is going fast leading to a muddle”. As the speed of the snake rose, the participants had less and less time to perform their interactions. While planning their interaction, what direction to go, how long beforehand to say the command, and formulate the word in their mind, the participants struggled to do it all in a brief time and confusion caused them to think or speak incorrectly.

Some of these negative UXs presented so far were the cause of category “System Behaviors Hindering/Worsening the Interactions”. This category described what the system did badly regarding the interactions and resulted in bad UX. The major points were that the SRT was too slow at times, that the CSR was too low at times and that the system did not provide enough feedback to the participant. After a command had been spoken the snake would take several steps before the command was executed. I.e., the SRT was slow. However, this only manifested as a negative UX at higher speeds, further into the rounds. At the beginning of a round, the snake moved slowly enough that a command was executed within its next step. For example, when the participants were asked: Did it feel like you had enough time to exercise the commands? Some responses were:

> “Yes, I thought so. But there is not enough time when you are bigger than one dimension of the play field.” (P3)
>
> “At the beginning [of the round] it felt like you had too much time and at 14 points too little.” (P4)
>
> “Yes, but at 15 points it was hard to think through and interact.” (P6)

This indicates that the participants were satisfied with the SRT, up to a point where the movement speed of the snake was too fast compared to the SRT. Concerning the CSR, an interaction with an unsuccessfully interpreted command could lead to total failure in the game and cause the player to die, thus having the possibility of making a significant impact on the UX. The observational notes showed that there were a lot of failed recognitions by the system and that some participants experienced a bit

lower CSR than others. When asked “How did you experience the game understood your commands?”, almost every participant stated that their experience was ‘good’.

> “The system had trouble with recognizing my ‘Down’ commands.” (P1)
>
> “It understood nine times out of ten and that one in ten it did not understand at all. It never misinterpreted and went in the wrong direction.” (P6)
>
> “It understood nine times out of ten I felt like.” (P3)
>
> “It was fairly good. But hard to know if it was my own fault or the system’s.” (P2)

The participants, in general, seemed satisfied with the CSR. However, no one experienced a perfect CSR of 100% across a whole session and system failure to recognize a command is a bad interaction in a real-time game. P1 felt that the system was having trouble recognizing him when he said ‘Down’, which was one of the more terrible experiences. Both P6 and P3 estimated the CSR to 90% and P6 continued to explain that the system never misinterpreted his commands. P2 stated the CSR was “fairly good” but that it was hard to know if it was because of himself, e.g., that he talked too quiet for the system to hear him, or if the system had received his spoken word and failed to recognize it. Which brings us to the last point, that the system did not provide enough feedback. Relatable to Heidegger’s idea of ‘transparent in use’ (Svanaes, 2013). When a system is not transparent in how it operates it can negatively affect the interaction. The UX during this experiment was inhibited as the participants did not know if the system had heard them. Not understanding whether they would have to repeat their command or if the system were just slow on executing the command. This was mainly expressed by the participants as self-critique, that they blamed themselves for commands not being executed.

> “Sometimes you had to speak [a command] several times but it did not feel like the system’s fault.” (P2)
>
> “I might have said ‘double commands’ too fast.” (P4)
>
> “… I think it might be me who spoke badly, so that the system did not understand me.” (P3)
>
> “It would be good if you could see that the system had heard you.” (P4)

P2 talked about how he had to repeat his commands sometimes and expressed that the fault laid on himself. P4 referred to the queue-system, that he might have spoken several, consecutive, words too fast for the system to understand him. P3 had a similar experience and blamed himself for not speaking properly when the system did not understand. P4 expressed his desire of feedback from the system.

5.1.2 Positive UX

This theme was predominant in the responses and statements of the participants. In the category ‘Positive Feeling’, fun and joy were the major points from the observations and discussion. The participants stated that they enjoyed the game and its control modality.

> “It gives a fun challenge, when having to plan and think fast.” (P3)
>
> “It is fun and cool. It is different.” (P5)
>
> “It was fun once you got into it.” (P6)

P3 stated that the game is fun and challenging when planning and fast thinking is required. When asked about his thoughts about the game, P5 answers that he thinks it is fun, cool, and different. Answering the same question, P6 answers that the game was fun once he got used to the voice interactions.

In the category ‘Natural Interactions’, the only identified point was when participants were in a ‘flow’. Being in a flow means that the interactions, while playing a game, are flowing without unintended hinderance or interruption. Flow and its unintended interruption is relatable to Heidegger’s descriptions of tool use (Svanaes, 2013). Where a tool should be natural in its use, the user should not have to think about using it. If a tool stops working as intended, it should be transparent enough so that the problem causing the breakdown can be fixed. Examples of observational notes referring to the user being in a flow is: “Very concentrated”, “Focused” and “Very focused, in a flow when things are running smoothly”. One participant also made a statement about being in a flow:

> “… if you got a flow, it went well.” (P5)

P5 stated that interactions went well when he was in a flow. Things that broke the flow were the system behaviors brought up in the category ‘system behaviors hindering/worsening the interactions’ under the ‘negative UX’ theme.

The last category in the ‘Positive UX’ theme was ‘System Behaviors Supporting/Enhancing the Interactions’. As already identified in the ‘negative UX’ theme, the SRT of the system seemed to be acceptable up to a point where the speed of the snake was too fast. Statements by the participants shows when they felt the SRT contributed to a positive experience. This is what some participants answered when asked “Did it feel like you had enough time to exercise the commands?”:

> “Yes, I panicked when it [the snake] went faster but if you got a flow, it went well.” (P5)
>
> “In the beginning it felt as if you had too much time [to interact] and at 14 points too little.” (P4)
>
> “It was tight towards the end [of the round], but really yes.” (P6)

P5 answered yes, he felt there was enough time to exercise the command and indicate that he struggled when the snake started to move faster but that it went well if the flow was not broken. P4 stated that there is enough time to exercise commands in the beginning of the rounds, when the snake is moving slower. P6 felt the same, that there was enough time but not towards the end of the rounds. The CSR was also brought up in the ‘negative UX’ theme, and as stated the experience with CSR was varied. Observational notes concerning good CSR was “The system understood anyway” and “Bad pronunciation works well” referring to the system performing the correct command despite the word spoken being terribly pronounced. When asked “How did you experience the game understood your commands?” these participants answered:

> “It worked well. It only failed once I think.” (P4)
>
> “It understood well.” (P6)
>
> “The system understood but there was a delay.” (P5)

All these three participants stated they were well satisfied with the CSR of the system. The last major system behavior providing positive UX was the queue-system. The queue-system was implemented by the authors to counteract the slow SRT by enabling the user to queue commands while the system was still processing a spoken word. I.e., the users were not forced to wait for the system to execute a command before another command was spoken. Examples of observational notes about the queue-system supporting the interactions are “Utilizing the queue-system well” and “Seems to go better when using the queue-system”. When the queue-system was utilized by the participant they seemed to perform better. Some statements about the queue-system were:

> “It was easier when queuing commands.” (P1)
>
> “It was practically impossible to control the snake if you did not queue commands.” (P6)

When asked if it felt like there was enough time to exercise commands, P1 stated in his answer that the queue-system made the interactions easier. When P6 talked about his general thoughts about the game, he stated that towards the end of a round it was hard to control the snake (since it was moving fast and the SRT was slow) but that the queue-system alleviated this.

5.2 Statistical analysis

5.2.1 SRT compared to score

To investigate how system response time affects the user, the group Session2Experiment had 70ms added latency. A higher SRT should have led to less control for the user which should have led to a lower score. To investigate how SRT affected the score of the separate groups the variables SRT and "score" (collected fruits) were analyzed descriptively.

Table 3 *SRT per round split by session and Collected fruits per round split by session*

Descriptive Statistics

Session & Group	Variable	N	Minimum	Maximum	Mean	Std. Deviation
Session1Control	SRT (ms)	665	11	2241	430	164
Session1Experiment	SRT (ms)	433	61	1591	372	137
Session2Control	SRT (ms)	438	141	1441	432	149
Session2Experiment	SRT (ms)	546	91	1121	446	131
Session1Control	score	16	9	16	12,63	2,19
Session1Experiment	score	13	9	16	11,85	2,27
Session2Control	score	12	9	16	13,00	2,34
Session2Experiment	score	15	9	17	13,00	2,78

Comparing the data visually it was identified that the variance in SRT was remarkably high, in the magnitude of seconds rather than milliseconds, ranging from 11 ms to 2241 ms. The reason for this was not clear. The internet latency could have suddenly increased for some interactions, or the ASR took longer to compute for some interactions.

Also, it does not seem to be any correlation between high/low mean SRT and high/low mean score. There was no decrease in the mean collected fruits for the group Session2Experiment which had 70ms of added latency, compared to the group Session2Control which had no added latency. Since both of these groups had a mean score of 13, i.e., no difference, a T-test to see statistical significance cannot be done. However, both groups showed an increase in mean score during Session 2 compared to Session 1.

5.2.2 System state at user failure

The baseline grade of the system is not something calculated. Instead, the outer bounds and means of the variables CSR and "score" can tell us at what points the user failed and what the state of the system was at that point. To investigate this, CSR and score were analyzed descriptively and sorted by session.

Table 4 *Collected fruits and CSR per round split by session*

Descriptive Statistics

Session & Group	Variable	N	Minimum	Maximum	Mean	Std. Deviation
Session1Control	CSR (%)	16	77	100	92,1	8,3
	score	16	9	16	12,6	2,2
Session1Experiment	CSR (%)	13	80	100	92,0	7,1
	score	13	9	16	11,8	2,3
Session2Control	CSR (%)	12	82	100	93,4	5,7
	score	12	9	16	13,0	2,3
Session2Experiment	CSR (%)	15	82	100	95,2	6,0
	score	15	9	17	13,0	2,8

The maximum score was 17 which in turn represented a step delay, or time to react, of approximately 150 milliseconds. This time to react between grid movements was not survived by any user. The minimum score was 9 which was the point where the snake surpassed the vertical size of the playfield. Both groups, during session 2, showed a slightly higher minimum and mean CSR and also showed a slightly higher mean score. Overall, the lowest CSR was 77%, and all groups had at least one round where the CSR was a perfect 100%. All groups had a mean CSR of more than 90%.

5.2.3 CSR as predictor of score

As CSR is the percentage of user commands that were interpreted correctly by the system, it is probable that a higher CSR could be a predictor of higher score as indicated above. To investigate this, the correlation between CSR and score was analyzed through a Pearson correlation test. Both variables had a normal distribution and a linear correlation in a scatter plot.

Table 5 *Pearson correlation on CSR and Collected fruits*

Correlations

		score	CSR
score	Pearson Correlation	1	.270*
	Sig. (2-tailed)		0,044
	N	56	56
CSR	Pearson Correlation	.270*	1
	Sig. (2-tailed)	0,044	
	N	56	56

*. Correlation is significant at the 0.05 level (2-tailed).

The test determined that there was a positive correlation between CSR and score, $r = 0.27$ ($p < 0,05$). This indicates that the game rounds that had fewer errors tended to lead to a higher score.

6 Discussion and conclusion

6.1 Discussion

This study set out to answer the research question:
How viable is a cloud-based ASR as the single control modality, in a mobile game, in relation to system response time (SRT) and command success rate (CSR)?

The key factors in explaining the viability were regarded as User eXperience (UX) and User Performance (UP). To answer the question, the game 'Voice Snake' was developed. To keep the scope within limits, the game was designed to be navigated in just two dimensions using only discrete inputs. To test the limits of the SRT for real-time voice-controlled interactions the speed of the snake was designed to gradually increased with every score. Techniques to combat inaccurate and slow speech recognition was implemented to investigate if these would help make the cloud-based ASR more viable. An experiment designed as a playtest was conducted to investigate the interactions in the game. Based on previous knowledge researched and the fact that the speed of the snake increased ever more, the working hypothesis during the study was: H_1: "The user *does not* experience real-time control of the game, at all times, using cloud-based ASR as the single control modality in the game 'Voice Snake'".

UX was considered as requiring qualitative data to investigate. This was collected in the form of semi-structured notes from observations and discussions and analyzed with a thematic analysis. UP was considered the score of collected fruits in Voice Snake. Along with other system/game metrics, this quantitative data was automatically collected by the system. Screen recordings were made to manually count the CSR which was imported to the quantitative dataset. The quantitative data were analyzed with statistical analysis.

In summary, the thematic analysis showed that both positive and negative UX were prevalent during the experiment, indicating that the experience went from positive to negative at some point during the rounds. Furthermore, no major discrepancies between the groups or sessions were found to be conclusive enough. Hence, unable to identify if the manipulated SRT had any effect on the participants.

In summary, the statistical analysis also showed that the manipulated SRT seemed to have no effect on the UP. Both groups had an increase of the UP from the first to second session, indicating a skill increase. However, both groups also had a slight increase of the CSR from the first to second session. A Pearson correlation test showed a slight positive correlation between the CSR and UP, indicating that higher CSR may lead to better UP.

Continuing the discussion in detail, it is important to remember that the speed of the snake rose with every collected fruit. Refer to Table 1 for how the step delay decreased (i.e., movement speed increased) with every score.

Furthermore, the reason for why no one surpassed a score of 17 was probably because at a score of 16 the length of the snake exceeded the length of the playfield (which was a 9x16 grid). Requiring the user to continuously act to survive. Before the score of 16, the snake can move endlessly on the horizontal plane without dying.

6.1.1 SRT

Concerning SRT, the results indicate that the participants had enough time to interact up to a certain point into the rounds, where the UX went from positive to negative. Based on the participants statements, this point seemed to be at around a score of 8 to 14 collected fruits (equivalent to the step delay of one step per 600 ms to 300 ms). This point is supported by the results that showed the mean scores of collected fruits ranging between 11.85 to 13 across all sessions. The results also showed that the maximum score of any round was 17. Indicating that the participants struggled with the interactions post a score of 13 and that no one could manage the interactions post 17. Furthermore, the results showed that the mean SRT ranged between 370 ms to 450 ms across all sessions. Which means that, at a score of 9 to 11 the step delay would begin to be faster than the mean SRT.

Hence, it can be argued that a cloud-based ASR, with the mean SRT of 370 ms to 450 ms, could be a viable control modality for real-time discrete interactions, navigating in 2D space, in a mobile game, when the thing controlled has a minimum step delay of around 300 ms to 600 ms. It is hard to say any definitive values since the SRT varied extremely during the experiment and the participants had varying levels of acceptance for the SRT. That a step delay *below* 300 ms makes the interactions *non*viable, with a mean SRT of 370 ms to 450 ms, is in line with a brief test by Sporka *et al.* (2006). They briefly concluded that a SRT of around 500 ms is not viable for precise discrete/continuous hybrid interactions using voice commands in a game of Tetris. However, Tetris has two different dimensions of interactions to consider, the rotation of figures and horizontal movement. Hence more sensitive to higher SRT than Voice Snake.

We consider the game Voice Snake to be real-time, just that the movement speed of the controlled snake is slow in the beginning. Since the step delay of 300 ms to 600 ms is relatively high, the user does not expect a fast response. Therefore, the required SRT is not in line with the SRT limit of 140 ms reasoned in the Background. This is because 140 ms presumes the user expects an “immediate” response by the system. After further research into literature, we believe that it is more relevant to talk about the responsiveness of a system and that a user’s perception of what is responsive dictates viable interactions. Shneiderman's *et al.* (2017) seventh rule of interface design state that users wish a technical system to respond as intended so that they feel in charge of this system. Johnson (2013) describe responsiveness as being important to users and that it is not simply about fast performance, but rather that real-time deadlines are dictated by the users’ perceptions of what is responsive. With the measured mean SRT being viable up to a point is partly in line with the study by Winkler *et al.* (2020). They concluded that, when using voice commands to interact with a virtual button, a SRT above 450 ms should be avoided for the users to have a stronger sense of being in control.

Based on what has been discussed, this study argues that the findings indicate that the participants experienced the interactions as real-time, and were in control, up to a point in Voice Snake. This point is where the step delay was around 300 ms to 600 ms, while the SRT was around 370 ms to 450 ms.

6.1.2 Queue-system

The results further indicated that the implemented queue-system provided big support to some participants in managing the slow SRT. By using the hypotheses results of the ASR and queueing of

commands, the system did not have to wait for the participants to finish their utterances before initiating the recognition. Hence, the participants did not have to worry about speaking commands too fast consecutively. This combats one of the problems with using speech recognizers, as a control modality, as described by Sporka *et al.* (2006) that "most speech recognizers [have to] wait for the user to finish their utterance before initiating the recognition" (p. 214). That there are techniques for achieving more responsiveness, in VUIs, is also supported by Johnson (2008). In his book he gives examples of these techniques, where queuing is one of the described.

6.1.3 Feedback

The results also showed that feedback from the system could be improved upon in order to improve the UX. With feedback implemented, e.g., showing that the system had heard a command and is processing it, it could be argued that the viability of voice control could be enhanced even further. This is in line with several design principles, such as Shneiderman's *et al.* (2017) third golden rule of interface design and Pearl's (2016) guidelines on designing VUIs. That for every user action the system should provide feedback on what happens. As stated by Shneiderman *et al.* on the importance of feedback, "the availability of a display can greatly speed up interaction by presenting the proposed action in detail…" (p. 322). Winkler *et al.* (2020) also showed that, apart from low SRT, feedback is important for the users to feel in control of a system.

6.1.4 CSR

Concerning the CSR, the thematic analysis showed that the experiences were varied. This is supported by the statistical analysis where the CSR ranged from 77% - 100% for each game round, across all sessions.

Results showed that the participants were generally positive regarding the CSR but that a CSR below 100% for critical interactions could have a big negative effect on the UX. The results also showed that when the CSR was higher the UX was more positive. Many participants expressed that they felt the CSR was around 90%. Which is also supported by the statistical analysis that showed the mean of the CSR ranging between 92% - 95% across all sessions. The results also indicated that there was a weak positive correlation between high CSR and a high score (i.e., high UP). Supporting the statement of positive UX when the CSR is high. If the participant performs better the UX should naturally be more positive. The CSR in this game could be even further improved upon, by teaching the system how the current user speaks or to react on even more similar words as the speech commands.

Scovell *et al.* (2015) reported users having low tolerance to CSR below 70%, this was however for non-critical interactions with a tablet. This study argues that a CSR close to 100% is needed for a system with critical interactions, as a CSR below 100% can result in complete user failure. Just as Shneiderman *et al.* (2017) states, "errors remain a significant challenge, and not all situations benefit enough from speech input to balance the cost of errors and the frustration of error correction" (p. 312). This 'balance' is easily disturbed by errors if the interactions are critical.

Hence, it can be argued that the cloud-based ASR is not a viable control modality, regarding its CSR, in Voice Snake, with the current implementation. However, if the implementing system can implement techniques that improve the CSR to 100%, or very close to it, the ASR would probably be viable. Negative UX, when the CSR is below 100% in games with vital interactions is in line with reviews of previous games that utilize voice control as the primary control modality (Whitehead, 2015; Ip, 2017). The reviewers state that the gaming experience crumbles when the interactions fail as they are often left to die.

6.2 Conclusion

As there was a point in the game where the interactions started to falter and the UX became negative, it can be concluded that the working hypothesis H_1 is supported: The user *does not* experience real-time control of the game, at all times, using cloud-based ASR as the single control modality in the game Voice Snake.

In summary the study concludes that, regarding the SRT, the cloud-based ASR was viable up to a point where the snake was moving too fast. Furthermore, the queue-system was an effective way of enhancing the viability of the voice control by alleviating the slow SRT. It is also concluded that implementing feedback, about the system status, would have enhanced the UX. Lastly, regarding the CSR, the interactions were not viable, but the cloud-based ASR itself could still be viable if the implementing system would have got the CSR to 100%, or very close to it.

6.3 Implications

These findings could be of interest to any mobile application developer or mobile game designer that are looking to include speech recognition for real-time interaction in an application. The findings give an indication as to how the cloud-based ASR performs. E.g., game developers, developing a game with similar interactions, may adapt the speed of their game with the help of this study. Furthermore, the techniques implemented to alleviate slow SRT, i.e., hypothesis result and queue-system, is a guide for other developers on how to improve the viability of a cloud-based ASR. The findings may also be of interest to researchers of HCI within VUI as they support at least one general guideline of VUI and voice interaction, i.e., that a system should provide feedback for every user action.

With the above described support this study can provide, the societal implications could be that more inclusive design in applications and video games are implemented, in the form of voice interaction. Increasing the accessibility of using these digital products. However, by shifting the control modality from local ASRs to cloud-based ASRs, the users' integrity may be negatively affected as some of their data would be shared with the owner of the cloud-based ASR. Voice is a biometric data and regarded as a personal data under the General Data Protection Regulation (Mediartis, 2019) and transferring and storing users' voice is a great privacy concern (Siegert *et al.*, 2020). Considering that big companies can collect and save this data indefinitely (Osborne, 2019), and that this data can end up in the wrong hands (Murnane, 2018), there is valid concerns about increased use of cloud-based ASRs in the society.

6.4 Limitations

The observational notes did not record at what time in the round something occurred; they were only chronologically ordered over each whole session. This fact could be used to question the validity of this study since the observations have no reference of time as to where in the game round it occurred. Furthermore, the statistical analysis showed that the participants had increased their UP from session 1 to session 2. We show that this could be because of higher CSR during the session 2. But it could also be that the participant's skill level had increased and that the 5 minutes of "Warm up" was not enough to mitigate this.

The statistical analysis also showed that the SRT collected was extremely varied, without any indication of cause. This heavily impacts the validity of the SRT. Lastly, the researchers of this study

have little experience and training in conducting research. Hence, there is a risk of bad *inter-observer reliability*. I.e., that other researchers would interpret similar observations differently to what this study have. These facts could hence also be used to question the reliability of this study.

Furthermore, the ASR's ability to recognize the participants speech may have been negatively affected. During the experiment it was noted how the participants were speaking ever more rapid, affecting their pronunciation, as the speed of the snake rose. This was not incorporated in the results but is something to note as it may have negatively affected the ASR's ability to recognize their speech further into the game rounds. Possibly leading to lower CSR towards the end of the rounds and ultimately affecting our conclusion about the viability related to CSR.

Also, when the voice control system was implemented in Voice Snake, the selection of protocol to use to connect with Azure was completely disregarded. In general, TCP is more reliable but slower, and UDP is less reliable but faster (Kumar and Rai, 2012). Which protocol was used could have an impact on the reliability and performance of the system. Unfortunately, it is unclear whether the automatically configured connection used TCP or UDP to communicate with Microsoft Azure during the experiment.

Finally, the nature of how Voice Snake runs may have occasionally resulted in added SRT. The implementation of the "snake" in Voice Snake had two separate modules for control. The speech recognition which was asynchronous, and the movement timer which ran on Unity's "Fixed Update". Fixed Update is an event that is triggered 50 times per second, approximately every 20ms. (Unity Technologies, n.d.) The snake steps on a Fixed Update call if it has waited for enough real-time milliseconds, the limit of which is defined by the Step Delay. When player commands are registered, they are added to the queue of commands and the oldest command will be executed when the next Fixed Update is triggered. As commands are registered asynchronously it is possible that a command is registered just after the snake was allowed to move. This causes the SRT of that command to be increased by the current Step Delay. If this happened during the experiment and, if so, how often, is unknown and could have an impact on the reliability of this study.

6.5 Future research

Future research could investigate what more techniques or technologies could be used to make a cloud-based ASR a more viable control modality in real-time games. For example, a study showed that edge devices with ASRs can outperform cloud-based ASRs for low-complexity tasks (Sridhar and Tolentino, 2017), such as the case of Voice Snake that utilize short and few commands. With 5G, edge computing is predicted to become more ubiquitous and a simpler, but faster, ASR on the edge could be the middle ground between local ASR and cloud-based ASR. Furthermore, research could be conducted to see how other cloud-based ASRs would perform in a similar setup of this study. It could be the case that other cloud-based ASRs are more viable than Microsoft Azure's.

References

Allison, F. *et al.* (2018) 'Design Patterns for Voice Interaction in Games', *Computer-Human Interaction in Play*, pp. 5–17. doi: 10.1145/3242671.3242712.

Allison, F., Carter, M. and Gibbs, M. (2017) 'Word Play: A History of Voice Interaction in Digital Games', *Games and Culture*, 15(2), pp. 91–113. doi: 10.1177/1555412017746305.

Attig, C. *et al.* (2017) 'System Latency Guidelines Then and Now – Is Zero Latency Really Considered Necessary?', in Harris, D. (ed.) *Engineering Psychology and Cognitive Ergonomics: Cognition and Design*. Cham: Springer International Publishing, pp. 3–14.

Aylett, M. P. *et al.* (2014) 'None of a CHInd: relationship counselling for HCI and speech technology', in *CHI '14 Extended Abstracts on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery (CHI EA '14), pp. 749–760. doi: 10.1145/2559206.2578868.

Ballantyne, M. *et al.* (2018) 'Study of Accessibility Guidelines of Mobile Applications', in *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*. New York, NY, USA: Association for Computing Machinery (MUM 2018), pp. 305–315. doi: 10.1145/3282894.3282921.

Bierre, K. *et al.* (2005) 'Game not over: Accessibility issues in video games', in *Proc. of the 3rd International Conference on Universal Access in Human-Computer Interaction*, pp. 22–27. Available at: https://www.researchgate.net/publication/267403944_Game_Not_Over_Accessibility_Issues_in_Video_Games (Accessed: 21 January 2021).

Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology', *Qualitative Research in Psychology*, 3(2), pp. 77–101. doi: 10.1191/1478088706qp063oa.

Cairns, P. *et al.* (2019) 'Future design of accessibility in games: A design vocabulary', *International Journal of Human-Computer Studies*, 131, pp. 64–71. doi: 10.1016/j.ijhcs.2019.06.010.

Carter, P. and Molloy, D. (2020) 'Last of Us Part II: Is this the most accessible game ever?', *BBC News*, 20 June. Available at: https://www.bbc.com/news/technology-53093613 (Accessed: 26 January 2021).

CDC (2019) *Disability and Health Data System (DHDS), Centers for Disease Control and Prevention*. Available at: https://dhds.cdc.gov (Accessed: 9 June 2021).

Clark, L., Doyle, Philip, *et al.* (2019) 'The State of Speech in HCI: Trends, Themes and Challenges', *Interacting with Computers*, 31(4), pp. 349–371. doi: 10.1093/iwc/iwz016.

Denscombe, M. (2014) *The Good Research Guide: For Small-Scale Social Research Projects*. 5th edn. McGraw-Hill Education (UK). Available at: https://books-google-se.ezp.sub.su.se/books?hl=sv&lr=&id=C5BFBgAAQBAJ&oi=fnd&pg=PR3&dq=the+good+research+guide+for+small-scale&ots=gvNh_q6ycC&sig=ybiS5WhFOeBnMmX1E1YELeQltDE&redir_esc=y#v=onepage&q=the%20good%20research%20guide%20for%20small-scale&f=false.

Erić, T. *et al.* (2017) 'Voice control for smart home automation: Evaluation of approaches and possible architectures', in *2017 IEEE 7th International Conference on Consumer Electronics - Berlin (ICCE-Berlin)*, pp. 140–142. doi: 10.1109/ICCE-Berlin.2017.8210613.

ESA (2020) *2020 Essential Facts About the Video Game Industry*. Available at: https://www.theesa.com/esa-research/2020-essential-facts-about-the-video-game-industry/ (Accessed: 21 January 2021).

EUR-Lex (2019) *Directive (EU) 2019/882 of the European Parliament and of the Council of 17 April 2019 on the accessibility requirements for products and services*, *151*. Available at: http://data.europa.eu/eli/dir/2019/882/oj (Accessed: 24 January 2021).

Farrell, S. (2017) *Group Notetaking for User Research*, *Nielsen Norman Group*. Available at: https://www.nngroup.com/articles/group-notetaking/ (Accessed: 11 March 2021).

Feng, J. *et al.* (2011) 'Speech-based navigation and error correction: a comprehensive comparison of two solutions', *Universal Access in the Information Society*, 10(1), pp. 17–31. doi: 10.1007/s10209-010-0185-9.

Fogg, I. (2019) *Latency is the new 5G speed battleground which will enable XR and AR*, *Opensignal*. Available at: https://www.opensignal.com/2019/03/08/latency-is-the-new-5g-speed-battleground-which-will-enable-xr-and-ar (Accessed: 7 March 2021).

Folkhälsomyndigheten (2021) *Regulations and general guidelines - The Public Health Agency of Sweden*. Available at: https://www.folkhalsomyndigheten.se/the-public-health-agency-of-sweden/communicable-disease-control/covid-19/regulations-and-general-guidelines/ (Accessed: 25 February 2021).

Fullerton, T. (2018) *Game Design Workshop: A Playcentric Approach to Creating Innovative Games, Fourth Edition*. 4th edn. Boca Raton, USA: CRC Press.

Gallant, M. (2020) *The Last of Us Part II: Accessibility Features Detailed*, *Naughty Dog*. Available at: https://www.naughtydog.com/blog/the_last_of_us_part_ii_accessibility_features_detailed (Accessed: 26 January 2021).

GSA (2020) *IT Accessibility Laws and Policies | Section508.gov*. Available at: https://www.section508.gov/manage/laws-and-policies (Accessed: 22 January 2021).

Gupta, R. *et al.* (2019) 'Tactile internet and its applications in 5G era: A comprehensive review', *International Journal of Communication Systems*, 32(14), p. e3981. doi: https://doi.org/10.1002/dac.3981.

Hagerer, G. *et al.* (2017) 'VoicePlay — An affective sports game operated by speech emotion recognition based on the component process model', in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, pp. 74–76. doi: 10.1109/ACIIW.2017.8272589.

Harada, S., Wobbrock, J. O. and Landay, J. A. (2011) 'Voice Games: Investigation Into the Use of Non-speech Voice Input for Making Computer Games More Accessible', in Campos, P. et al. (eds) *Human-Computer Interaction – INTERACT 2011*. Berlin, Heidelberg: Springer (Lecture Notes in Computer Science), pp. 11–29. doi: 10.1007/978-3-642-23774-4_4.

Ip, C. (2017) 'The next video-game controller is your voice', *Engadget*, 23 June. Available at: https://www.engadget.com/2017-06-23-voice-based-gaming.html (Accessed: 17 May 2021).

Johnson, J. (2008) *GUI Bloopers 2.0*. 2nd edn. Elsevier. doi: 10.1016/B978-0-12-370643-0.X5001-X.

Johnson, J. (2013) *Designing with the Mind in Mind: Simple Guide to Understanding User Interface Design Guidelines*. 2nd edn. Waltham, USA: Morgan Kaufmann/Elsevier.

Juang, B. H. and Rabiner, L. R. (2005) 'Automatic Speech Recognition – A Brief History of the Technology Development', p. 24.

Kaaresoja, T. J. (2016) *Latency guidelines for touchscreen virtual button feedback*. PhD. University of Glasgow. Available at: http://eleanor.lib.gla.ac.uk/record=b3144451 (Accessed: 25 May 2021).

Karpagavalli, S. and Chandra, E. (2016) 'A Review on Automatic Speech Recognition Architecture and Approaches', *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 9(4), pp. 393–404. doi: 10.14257/IJSIP.2016.9.4.34.

Key, R. (2018) *EA Creates Accessibility Portal For Disabled Players, Game Informer*. Available at: https://www.gameinformer.com/b/news/archive/2018/03/15/ea-creates-accessibility-portal-for-disabled-players.aspx (Accessed: 26 January 2021).

Kiiski, T. (2020) *Voice Games: The History of Voice Interaction in Digital Games*. Bachelor. KAMK University of Applied Sciences. Available at: http://www.theseus.fi/handle/10024/336352 (Accessed: 17 May 2021).

Kota, S. (2019) 'Online Gaming, Speed, and Network Latency on 4G LTE - EvdodepotUSA', *Evdo*, 16 January. Available at: https://www.evdodepotusa.com/speed-network-latency-required-online-gaming-4g-lte/ (Accessed: 7 March 2021).

Krainz, E., Miesenberger, K. and Feiner, J. (2018) 'Can We Improve App Accessibility with Advanced Development Methods?', in Miesenberger, K. and Kouroupetroglou, G. (eds) *Computers Helping People with Special Needs*. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 64–70. doi: 10.1007/978-3-319-94277-3_12.

Kumar, S. and Rai, S. (2012) *Survey on Transport Layer Protocols: TCP & UDP*.

Mach, P. and Becvar, Z. (2017) 'Mobile Edge Computing: A Survey on Architecture and Computation Offloading', *IEEE Communications Surveys & Tutorials*, 19(3), pp. 1628–1656. doi: 10.1109/COMST.2017.2682318.

Marsch, P. *et al.* (2018) *5G System Design: Architectural and Functional Considerations and Long Term Research*. John Wiley & Sons. doi: 10.1002/9781119425144.

Mediartis (2019) *GDPR : Why is voice considered a personal data ?, Mediartis*. Available at: https://mediartis.com/blog/gdpr-voice-is-personal-data/ (Accessed: 14 June 2021).

Miesenberger, K. *et al.* (2008) 'More Than Just a Game: Accessibility in Computer Games', in Holzinger, A. (ed.) *HCI and Usability for Education and Work*. Berlin, Heidelberg: Springer (Lecture Notes in Computer Science), pp. 247–260. doi: 10.1007/978-3-540-89350-9_18.

Munteanu, C. *et al.* (2017) 'Designing Speech, Acoustic and Multimodal Interactions', in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery (CHI EA '17), pp. 601–608. doi: 10.1145/3027063.3027086.

Murnane, K. (2018) *Amazon Does The Unthinkable And Sends Alexa Recordings To The Wrong Person, Forbes*. Available at: https://www.forbes.com/sites/kevinmurnane/2018/12/20/amazon-does-the-unthinkable-and-sends-alexa-recordings-to-the-wrong-person/ (Accessed: 14 June 2021).

Naftali, M. and Findlater, L. (2014) ‘Accessibility in context: understanding the truly mobile experience of smartphone users with motor impairments’, in *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility*. New York, NY, USA: Association for Computing Machinery (ASSETS ’14), pp. 209–216. doi: 10.1145/2661334.2661372.

Newzoo (2020) *Newzoo Global Games Market Report 2020*. Available at: https://newzoo.com/insights/trend-reports/newzoo-global-games-market-report-2020-light-version/ (Accessed: 21 January 2021).

Osborne, C. (2019) *Amazon confirms Alexa customer voice recordings are kept forever*, *ZDNet*. Available at: https://www.zdnet.com/article/amazon-confirms-alexa-customer-voice-recordings-are-kept-forever/ (Accessed: 14 June 2021).

Pearl, C. (2016) *Designing Voice User Interfaces: Principles of Conversational Experiences*. 1st edn. Sebastopol, USA: O’Reilly Media, Inc.

Pozzi, N. and Zimmerman, E. (2016) ‘Don’t follow these rules! A Primer for Playtesting’. Available at: www.ericzimmerman.com/s/A-Primer-for-Playtesting.pdf (Accessed: 17 February 2021).

Rodriguez, J. (2015) *Fundamentals of 5G Mobile Networks*. 1st edn. Chichester, United Kingdom: John Wiley & Sons. Available at: https://onlinelibrary-wiley-com.ezp.sub.su.se/doi/pdf/10.1002/9781118867464 (Accessed: 25 May 2021).

Scovell, J. *et al.* (2015) ‘Impact of Accuracy and Latency on Mean Opinion Scores for Speech Recognition Solutions’, *Procedia Manufacturing*, 3, pp. 4377–4383. doi: 10.1016/j.promfg.2015.07.434.

Shneiderman, B. *et al.* (2017) *Designing the user interface: strategies for effective human-computer interaction*. 6th edn. Boston: Pearson.

Siegert, I. *et al.* (2020) ‘Personal data protection and academia: GDPR issues and multi-modal data-collections’, *Online Journal of Applied Knowledge Management (OJAKM)*, 8(1), pp. 16–31. doi: 10.36965/OJAKM.2020.8(1)16-31.

Sporka, A. J. *et al.* (2006) ‘Non-speech input and speech recognition for real-time control of computer games’, in *Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility - Assets ’06*. Portland, Oregon, USA: ACM Press, p. 213. doi: 10.1145/1168987.1169023.

Sridhar, S. and Tolentino, M. E. (2017) ‘Evaluating Voice Interaction Pipelines at the Edge’, in *2017 IEEE International Conference on Edge Computing (EDGE)*, pp. 248–251. doi: 10.1109/IEEE.EDGE.2017.46.

Statista (2020) *Mobile provider latency in the US 2019*, *Statista*. Available at: https://www.statista.com/statistics/818205/4g-and-3g-network-latency-in-the-united-states-2017-by-provider/ (Accessed: 7 March 2021).

Stenman, M. (2015) *Automatic speech recognition An evaluation of Google Speech*. Bachelor. Umeå University. Available at: http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-108383 (Accessed: 30 January 2021).

Summa Linguae (2017) ‘Voice Controlled Games: The Rise of Speech Technology in Gaming’, *Summa Linguae*, 22 June. Available at: https://summalinguae.com/language-technology/voice-controlled-games/ (Accessed: 6 May 2021).

Svanaes, D. (2013) *Philosophy of Interaction*. Available at: https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/philosophy-of-interaction (Accessed: 18 April 2021).

Tullis, T. and Albert, B. (2013) *Measuring the user experience: collecting, analyzing, and presenting usability metrics*. 2nd edn. Amsterdam ; Boston: Elsevier/Morgan Kaufmann.

Unity Technologies (no date) *Unity - Scripting API: MonoBehaviour.FixedUpdate()*. Available at: https://docs.unity3d.com/ScriptReference/MonoBehaviour.FixedUpdate.html (Accessed: 16 June 2021).

W3C (1999) *Web Content Accessibility Guidelines 1.0*. Available at: https://www.w3.org/TR/WCAG10/ (Accessed: 25 January 2021).

Warren, T. (2018) *Microsoft's new Xbox Adaptive Controller launches in September, The Verge*. Available at: https://www.theverge.com/2018/6/11/17449206/microsoft-xbox-adaptive-controller-august-3rd-release-date-pricing (Accessed: 26 January 2021).

Westin, T. *et al.* (2018) 'Game Accessibility Guidelines and WCAG 2.0 – A Gap Analysis', in Miesenberger, K. and Kouroupetroglou, G. (eds) *Computers Helping People with Special Needs*. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 270–279. doi: 10.1007/978-3-319-94277-3_43.

Whitehead, D. (2015) 'There Came an Echo review', *Eurogamer*, 2 March. Available at: https://www.eurogamer.net/articles/2015-03-02-there-came-an-echo-review (Accessed: 17 May 2021).

Wiggers, K. (2019) 'Google Assistant will soon be on 1 billion devices, but still can't speak like John Legend', *VentureBeat*, 7 January. Available at: https://venturebeat.com/2019/01/07/google-assistant-will-soon-be-on-1-billion-devices/ (Accessed: 7 February 2021).

Willard, C. A. (2020) *Statistical methods: an introduction to basic statistical concepts and analysis*. Routledge.

Wilson, A. and Crabb, M. (2018) 'W3C Accessibility Guidelines for Mobile Games', *The Computer Games Journal*, 7(2), pp. 49–61. doi: 10.1007/s40869-018-0058-7.

Winkler, P. *et al.* (2020) 'How latency, action modality and display modality influence the sense of agency: a virtual reality study', *Virtual Reality*, 24(3), pp. 411–422. doi: 10.1007/s10055-019-00403-y.

Yu, D. and Deng, L. (2015) *Automatic Speech Recognition: A Deep Learning Approach*. London: Springer-Verlag (Signals and Communication Technology). doi: 10.1007/978-1-4471-5779-3.

Yuan, B., Folmer, E. and Harris, F. C. (2011) 'Game accessibility: a survey', *Universal Access in the Information Society*, 10(1), pp. 81–100. doi: 10.1007/s10209-010-0189-5.

Appendix A – Playtest script

1. **Introduction (5 Minutes)**

First, welcome the playtesters and thank them for participating.

Introduce ourselves—our names, that we are students and currently doing our thesis. Then give a brief explanation of the playtesting process and explain that the purpose of the session is to get their feedback on the experience. Let them know that after the whole playtesting we can share details on what the research is about if they wish to know. But not beforehand as it might affect what we intend to observe (aka the observer effect).

Briefly go through the Consent Form:
Let the participant know what data will be collected (I.e., notes through observation and discussion, audio- and screen-recording of the play session and game/system metrics).

Let them know that the data will be anonymized. Assure them that the data is for our reference only and will not be shown outside the research team and that details can be read in the Consent Form.

Hand over the Consent Form. Let the participant read it. If they agree and sign it continue with the playtest. Otherwise thank them for coming.

2. **Warm-Up Discussion (5 Minutes)**

Let the participant try out the game so they get familiar with the mechanics.

3. **Play Session (12 Minutes)**

Make sure the participant understand that we are testing the game, not their skill. Ask the participant to put their mobile in silent mode.

Remind the participant that they should focus on playing the game but that they are free to express any thoughts they have between their rounds. I.e., they do not have to "think out loud" when they are playing, since they need to use their voice for interacting with the game.

Stay in the room and watch quietly, from behind the participant, taking observational notes.

Make sure the environment is quiet and that other people can't interrupt the playtest. Turn computer notifications and phone ringers off.

Start the microphone + screen-recording and the game.

If the testers have a tremendous amount of difficulty with something, we can help them to move the session forward, but be sure to put in the notes where and why the problem occurred.

When you think of a question, you would like to ask the participant, please write it down for the discussion.

At the end of the play session, wrap it up, pause the recordings, make sure the data have been saved successfully and start the discussion with the participant.

4. **Discussion (10 Minutes)**

Make sure the participant understand that we are testing the game, not their skill. There are no wrong answers or experiences, and any difficulties they have in playing the game will help you improve your design.

Start by telling the participant that we are mainly interested in the interactions with the game but that they are free to talk about anything that comes to mind.

Use the notes sheet to fill out answers and let the following questions guide the discussion:

• Overall, what were your thoughts about the game?

• How did you experience the game understood your commands?

• How responsive did you feel the interaction was?

• Did it feel like you had enough time to exercise the commands?

Before ending the discussion, read back the notes step by step and ask the participant to verify that everything in the notes has been correctly interpreted.

<u>Repeat step 3 and 4 once more</u> for session B, manipulate the SRT if the participant belongs to the experimental group.

5. **Wrap-Up (6 Minutes)** ***Total 60 minutes***

Thank the playtester for coming in. Ask if they want details about exactly what research we are doing. Make sure we have the Consent Form and remind them that they can contact us if they wish to exit their participation. If we have a token gift, we can give it to them now.

Appendix B – Consent form

1. Background and purpose	This study is carried out within a Bachelor thesis at the Department of Computer and System Sciences (DSV), Stockholm University (SU). The research is investigating User Experience regarding voice control in games.
2. Description of study	The approach of the study is that of an experiment. In order to not affect the outcome of the experiment, detailed information about the study will be given after the experiment has been conducted. The experiment is designed as a playtesting session, consisting of two sessions. The participant will be playing a game of "Snake" controlled via speech commands. The experiment is estimated to take 60 minutes. Data will be collected via observation and discussions*. The data collected will be analyzed and used as a basis for the study.
3. Risks	The experiment, per se, poses no risk to the participant. There is however a risk of transferring the Covid-19 virus among the present people, with the ongoing pandemic. The responses given by the participant will, in the study, be referenced under a pseudonym. There is a risk that the participant can be identified through their answers, this risk is judged to be very small. Efforts will be made to obscure identifying responses in the data collected.
4. Data Management	The participant will be referenced with a pseudonym after the experiment is done. No written list of names and related pseudonyms will be held. Only the responsible researchers will have access to the data collected. The data will not be managed over email. Responses that are deemed to have a risk of identifying the participant will be edited. Data collected will be kept on the researchers' computers locally and on SciPro, the internal network of SU. The data and this consent form will not be kept together. The data will only be used within the scope of this study. At the end of the study this consent form will be destroyed and only a digital copy will be archived on a hard drive at SU. The cloud-based speech recognition service used is provided by Microsoft (Windows Speech**). The audio-data sent to this service may be stored and used for other purposes than of this game. However, the risk of this data being connected to the participants person is highly unlikely. The account used for this service is that of one of the researcher's and any data will most

	likely only be connected to that account. Further information on Microsoft's terms and other agreement-related documents can be found here: https://www.microsoftvolumelicensing.com/DocumentSearch.aspx?Mode=3&DocumentTypeId=46
5. Participation	Participation in the study is voluntary and the participant can at any time, during the course of the study, choose to withdraw without further explanation by contacting any of the researchers.
6. Responsible Researchers	The study is carried out by Game Development students at DSV Stockholm University Spring 2021 Christian Lindberg [crillefrille@gmail.com] Joakim Linna [joakimlinna1998@gmail.com]

Informed Consent

- I confirm that I have received, read, and understood this written information along with receiving verbal detailed information about the study, after the experiment.

- I agree to take part in the above study and acknowledge that my participation is voluntary.

- I am aware that I can at any point, during the course of the study, withdraw my participation from the study without further explanation.

- I allow the information provided by me and the collected data regarding my participation to be archived and handled electronically by the responsible researchers of this study.

……...	…….....	……
Date and Location	Participant Name	Participant Signature

……

Researcher [Christian Lindberg], Signature, Date and Location

…….

Researcher [Joakim Linna], Signature, Date and Location

* Error noticed after the held experiment. Data was also collected automatically regarding game/system metrics, and with a screen-recording including microphone input. However, this was communicated verbally to the participants before signing.
** Error noticed after the held experiment. The service was called Microsoft's Cognitive Service.

Appendix C – Notes sheet

Participant GUID: ….....

Test group: Exp / Ctrl Session: A / B Previous GUID: ….....

Observation

Event	Note

Discussion

A. Overall, what were your thoughts about the game? / SWE: Vad är dina tankar om spelet generellt?:

_____
_____
_____
_____
_____

B. How did you experience the game understood your commands? / SWE: Hur upplevde du att systemet förstod dina kommandon?:

_____
_____
_____
_____
_____

C. How responsive did you feel the system was? / SWE: Hur responsivt upplevde du att systemet var?:

_____
_____
_____
_____
_____

D. Did it feel like you had enough time to exercise the commands? / SWE: Kändes det som att du hade tillräcklig tid för att utöva kommandon?:

_____
_____
_____
_____
_____

Follow-up questions / Other (A/B/C/D + Number, Follow-up question, answer):

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

Appendix D – Collected qualitative data

Participant ID: P1

Test group: Ctrl Session: A

Observation

- Han artikulerar och time:ar in sina kommandon
- Blir förvånad att snabba kommandon köas
- Blir glad när han tar äpple
- Har svårt att säga "rätt" kommando ibland (t.ex. ska vänster men råkar säga höger)
- Suckade när han dog, fick 10p
- Blir för mycket att säga när det börjar gå snabbt
- Blir irriterad när kommando utförs för sent
- Pratar långsammare och lugnare i början när ormen rör sig långsammare
- Sade fel kommando och ändrade sig mitt i ett ord

Discussion

A. Overall, what were your thoughts about the game? / SWE: Vad är dina tankar om spelet generellt?:

Man måste förbereda sig för mycket. Man måste time:a in kommandon.

B. How did you experience the game understood your commands? / SWE: Hur upplevde du att systemet förstod dina kommandon?:

Enklare när man köar upp kommandon. Tycker att ordet "down" inte registreras

C. How responsive did you feel the system was? / SWE: Hur responsivt upplevde du att systemet var?:

Man behöver time:a in kommandon.

D. Did it feel like you had enough time to exercise the commands? / SWE: Kändes det som att du hade tillräcklig tid för att utöva kommandon?:

I början behövde man säga till 1 steg innan det skulle utföras och senare I rundan fler steg innan (då det gick snabbare). Fick mer kontroll när man köade kommandon.

Follow-up questions / Other (A/B/C/D + Number, Follow-up question, answer):

NOTE: Ta bort sista rundan, det var ingen seriös runda. Bar mask.

Participant ID: P1

Test group: Ctrl Session: B

Observation

- Frustrerad av kommandos responstid.
- Dog för att responsen var långsam.
- Blandar ihop kommandon när han köar flera kommandon. (tex. Down left VS Left down)
- Blir torr om halsen och hostar.
- Drar ut orden i början, som “hån” (mot systemet)
- Imponerad att lång kö med kommandon kunde göras.
- Klappar med händerna för att time:a kommandon
- När det går snabbt pekar han med armarna åt det håll han skall åt (för att säga rätt)

Discussion

A. Overall, what were your thoughts about the game? / SWE: Vad är dina tankar om spelet generellt?:

Märker att man måste säga kommandon långt innan de kommer utföras, med *flera* “steg” innan när det börjar gå snabbare.

B. How did you experience the game understood your commands? / SWE: Hur upplevde du att systemet förstod dina kommandon?:

Speciellt kommandot “down” var svår för systemet att förstå.

C. How responsive did you feel the system was? / SWE: Hur responsivt upplevde du att systemet var?:

Tycker det är mer responsivt med “kö-läggning” av kommandon

D. Did it feel like you had enough time to exercise the commands? / SWE: Kändes det som att du hade tillräcklig tid för att utöva kommandon?:

Samma som innan.

Follow-up questions / Other (A/B/C/D + Number, Follow-up question, answer):

Participant ID: P2

Test group: Exp Session: A

Observation

- Registrerar inte hens “down”
    - Fnissar
- “Panik” när spelet inte registrerar ord
- Har svårt att lyckas ta äpple och dör vid 11p
- Många kommandon som inte registreras
- Skrattar när spelet inte gör som förväntat
- “Flow” bryts märkbart
- Skakar på huvudet när kommando blir fel
- Repeterar kommandon flera gånger
- Lyckas inte med kommandots timing
    - Frustration
- Säger fel kommando
    - Vänster istället för ner

Discussion

A. Overall, what were your thoughts about the game? / SWE: Vad är dina tankar om spelet generellt?:

Imponerande att det funkar. Frustrerande att behöva tänka i förväg. Man lär sig ungefär hur långt man har innan man måste prata.

B. How did you experience the game understood your commands? / SWE: Hur upplevde du att systemet förstod dina kommandon?:

Hyfsat bra. Svårt att vet om det var jag eller systemet.

C. How responsive did you feel the system was? / SWE: Hur responsivt upplevde du att systemet var?:

Ganska responsivt. Sade man I tid så gick det bra. Fördröjningen stör.
“jag kanske uttalade saker för dåligt/snabbt och fick sämre respons”

D. Did it feel like you had enough time to exercise the commands? / SWE: Kändes det som att du hade tillräcklig tid för att utöva kommandon?:

Kändes bra bara man lärde sig hur tidigt man måste prata. Det blev svårare när det blev snabbare.

Follow-up questions / Other (A/B/C/D + Number, Follow-up question, answer):

Ibland behövde man säga flera gånger men det kändes inte som systemets fel.

Participant ID: P2

Test group: Exp Session: B

Observation

- Upprepar kommandon
- Lyckas inte med timing för kommando
- Besviken när hen missar äpple
- Missar ofta äpple
- Spel inte tillräckligt responsivt
    - Dör 2 gånger
- Skrattar när hen missar äpple
- Skrattar frustrerat när hen missar äpple
- Dog på grund av långsam respons
- Skrattar efter flera missade kommandon
- Suckar efter flera missade kommandon och dör

Discussion

A. Overall, what were your thoughts about the game? / SWE: Vad är dina tankar om spelet generellt?:

Hinner inte planera vad man ska göra då spelet reagerar för sent. Delvis lättare efter erfarenhet.

B. How did you experience the game understood your commands? / SWE: Hur upplevde du att systemet förstod dina kommandon?:

Samma som innan

C. How responsive did you feel the system was? / SWE: Hur responsivt upplevde du att systemet var?:

Samma som innan.

D. Did it feel like you had enough time to exercise the commands? / SWE: Kändes det som att du hade tillräcklig tid för att utöva kommandon?:

Samma som innan. För lång tid innan kommandon utförs.

Follow-up questions / Other (A/B/C/D + Number, Follow-up question, answer):

E. Varför ändrade du ett kommando du hade upprepat?:
För att ormen plötsligt blivit närmare från andra hållet. Kan även ha varit för att mina “up” var dåliga och ville ändra kommando.

Participant ID: P3

Test group: Ctrl Session: A

Observation

- Systemet svarade inte på kommando
  - Frustration, ledde till död
- Hann precis svänga
  - Lättnad
- Kommandot utfördes, men inte i tid
  - Frustration
- Frustrerad då han dog p.g.a. långsam respons
- Sade “meeen” när han dog, uppenbart besviken.
- Missar ofta äpplet, halvt frustrerad
- Dör p.g.a. långsam respons
- Väldigt koncentrerad
- Dubbelkommando misslyckades och han dog, besviket säger “fuck”.
- Suckar och kliar sig lite frustrerad.
- Dog p.g.a. långsam respons

Discussion

A. Overall, what were your thoughts about the game? / SWE: Vad är dina tankar om spelet generellt?:

Man spänner hela kroppen och fokuserar. Det är inga konstigheter med ett snake spel. Det svåra, p.g.a., fördröjningen är att man måste planera i förväg.

B. How did you experience the game understood your commands? / SWE: Hur upplevde du att systemet förstod dina kommandon?:

Den uppfattar 9 av 10 gånger tycker jag. Men tänker att det kan vara “jag” som pratar dåligt så att systemet inte förstår mig.

C. How responsive did you feel the system was? / SWE: Hur responsivt upplevde du att systemet var?:

Inte lika bra som man önskar. Att det är lite segt get karaktär till spelet. Ger också en kul utmaning då man måste planera och tänka snabbt.

D. Did it feel like you had enough time to exercise the commands? / SWE: Kändes det som att du hade tillräcklig tid för att utöva kommandon?:

Ja det tycker jag. Men man har inte tillräckligt med tid när man är större än en dimension av spelplanen.

Follow-up questions / Other (A/B/C/D + Number, Follow-up question, answer):

Participant ID: P3

Test group: Ctrl Session: B

Observation

1.1 Sa ett annat ord än vad som menades
    a. Vänster istället för ner
1.2 Upprepade kommando
    a. Skedde ett flertal gånger
1.3 Dog vertikalt
1.4 Kände sig taggad sade han
1.5 Sade fel kommando
1.6 Systemet hörde inte flera kommandon
1.7 Sätter fingrar vid tinningen för att koncentrera (frustrerat fokusera?)
1.8 Suckar när han dog
1.9 Frustrerad av att behöva upprepa sig när systemet inte hör/reagerar
1.10 Frustrerad av att dö
1.11 Väldigt fokuserad i ett "flow" när saker flyter på, vilket bryts av frustration
1.12 Utnyttjar bra "kö"-systemet

Discussion

A. Overall, what were your thoughts about the game? / SWE: Vad är dina tankar om spelet generellt?:

Ingen större skillnad från förra gången. Kände att jag hade bättre kontroll men det kan vara för att man tränat upp sig nu

B. How did you experience the game understood your commands? / SWE: Hur upplevde du att systemet förstod dina kommandon?:

Samma som förra gången. Lite fler kommandon som inte hördes på *förra* sessionen

C. How responsive did you feel the system was? / SWE: Hur responsivt upplevde du att systemet var?:

Samma som förra gången. Jag räknade liksom rutorna för hur långt innan man behövde prata, Upplevde responsen som första gången

D. Did it feel like you had enough time to exercise the commands? / SWE: Kändes det som att du hade tillräcklig tid för att utöva kommandon?:

Ja, samma som förra gången.

Follow-up questions / Other (A/B/C/D + Number, Follow-up question, answer):

Participant ID: P4

Test group: Exp Session: A

Observation

1. Sa fel riktning
    a. Menade down, sa left
2. I början av rundan lite trött att det går segt.
3. Fnissar åt när många kommandon blir fel och inte lyckas ta äpplet
4. Säger lätt “fel” kommando
5. Skrattar till när han dör, aningen frustrerad
6. Aningen obekväm med att styra med röst
7. Verkar gå bättre när han använder “kö”-systemet
8. Frustrerad när kommando inte hörs av systemet

Discussion

A. Overall, what were your thoughts about the game? / SWE: Vad är dina tankar om spelet generellt?:

Klankigt med delay. Egentligen bara svårt innan man vant sig. Det känns som snake, man behöver tänka på samma sätt. Bara en gång den inte reagerade tyckte jag. Bra om man skulle kunna se att systemet hörde en dock.

B. How did you experience the game understood your commands? / SWE: Hur upplevde du att systemet förstod dina kommandon?:

Det funkade bra. Den missade bara en gång tror jag.

C. How responsive did you feel the system was? / SWE: Hur responsivt upplevde du att systemet var?:

Inte jätteresponsivt. Blev jobbigt när det gick snabbare.

D. Did it feel like you had enough time to exercise the commands? / SWE: Kändes det som att du hade tillräcklig tid för att utöva kommandon?:

I början kändes det som att man hade för mycket tid på sig och vid 14 poäng för lite.

Follow-up questions / Other (A/B/C/D + Number, Follow-up question, answer):

NOTE: Stryk första rundan, vi avbröt för att sätta på inspelning.

Fråga: “Du sade att du tänkte ett kommando men sade ett annat, kan du förklara?”

Jag vet inte varför det blev så, det var en brainfart.

“Hade du koll på vad du skulle göra I stunden?”

Nä

Participant ID: P4

Test group: Exp Session: B

Observation

- Sa fel riktning
    - Menade down, sa left
- Kommandot utfördes tidigare än förväntat
- Vickar på huvudet som att det går långsamt I början.
- Skrattar när kommandona blir fel. Aningen frustrerad samtidigt.
- Glad när det gick bra
- Aningen långsamma uttalade kommandon
- Utnyttjar inte “kö”-systemet helt ut
- Suckar när han dör vid 14-15 poäng
- Har större problem än andra med att säga rätt kommandon
- Dog pga långsam respons, suckar.
- Ser ut att ha långtråkigt i början av rundor
- Frustrerad av långsam respons
- Skrattar när många försök/kommandon att ta äpplet misslyckas
- Problem med att säga rätt kommando

Discussion

A. Overall, what were your thoughts about the game? / SWE: Vad är dina tankar om spelet generellt?:

Tänker Det är snake.

B. How did you experience the game understood your commands? / SWE: Hur upplevde du att systemet förstod dina kommandon?:

Den förstod bra. Bara en gång nu också den inte förstod. Jag kan ha sagt dubbelkommandon för snabbt.

C. How responsive did you feel the system was? / SWE: Hur responsivt upplevde du att systemet var?:

Inte super-responsivt. Borde vara bättre, det känns omöjligt att klara spelet

D. Did it feel like you had enough time to exercise the commands? / SWE: Kändes det som att du hade tillräcklig tid för att utöva kommandon?:

Ja, I början av rundorna. Senare ej.

Follow-up questions / Other (A/B/C/D + Number, Follow-up question, answer):

Participant ID: P5

Test group: Ctrl Session: A

Observation

1. Tog flera försök att ta ett äpple I ett hörn
2. Systemet svarade inte på kommando
3. Upprepade kommando
4. Smått frustrerande att kommandon inte hörs (av systemet) och seg respons
5. Fnissar när det blir fel
6. Förstod att systemet var så segt att han skulle dö redan innan hen dog
7. Vissa ord eller dubbelord registreras inte
8. Pratar hårdare efter att systemet inte hört
9. Dog på grund av långsam respons
10. Fnissar när hen nästan dog
11. Snabb på att börja ny runda
12. Fundersamt uttryck I början
13. Svårt att få rätt timing på dubbelkommandon
14. Besviket ansiktsuttryck när hen missade äpplet på hög hastighet.
15. Frustrerad när hen dog

Discussion

A. Overall, what were your thoughts about the game? / SWE: Vad är dina tankar om spelet generellt?:

Kul och häftigt. Annorlunda.

B. How did you experience the game understood your commands? / SWE: Hur upplevde du att systemet förstod dina kommandon?:

Systemet förstod men det var delay. Vid 15p var det 3 steg innan utförande.

C. How responsive did you feel the system was? / SWE: Hur responsivt upplevde du att systemet var?:

Olika delay från början till slut.

D. Did it feel like you had enough time to exercise the commands? / SWE: Kändes det som att du hade tillräcklig tid för att utöva kommandon?:

Ja, fick panik när det gick fortare men fick man ett flow så gick det bra.

Follow-up questions / Other (A/B/C/D + Number, Follow-up question, answer):

Participant ID: P5

Test group: Ctrl Session: B

Observation

- Systemet svarar inte på kommando
- Fokuserad
- Dog på grund av långsam respons
- Långsammare tal i början
- Nickar med huvudet för att hitta timing i stegen
- I slutet när det går snabbt förstår hen att hen skulle dö redan innan hen dör
    - På grund av långsam respons 2 gånger
- Många kommandon när det går snabbt vilket blir strul

Discussion

A. Overall, what were your thoughts about the game? / SWE: Vad är dina tankar om spelet generellt?:

Kul fortfarande

B. How did you experience the game understood your commands? / SWE: Hur upplevde du att systemet förstod dina kommandon?:

Kändes likadant. Pratar om antalet steg vilket hen förstår bättre nu.

C. How responsive did you feel the system was? / SWE: Hur responsivt upplevde du att systemet var?:

Kändes likadant.

D. Did it feel like you had enough time to exercise the commands? / SWE: Kändes det som att du hade tillräcklig tid för att utöva kommandon?:

“Nu när jag förstod “antal steg” hos delayen var det bättre.”

Follow-up questions / Other (A/B/C/D + Number, Follow-up question, answer):

Participant ID: P6

Test group: Exp Session: A

Observation

- Systemet reagerade trots att hen inte utförde ett kommando
    - Systemet reagerade som att hen sa “down”
- Systemet svarade inte på kommando
- Dåliga uttal fungerar bra
- Väldigt besviken när hen dog
    - “NEJ!”
- Skrattar när det blir fel
- Blir pirrig när hen nästan dör
- Suck när hen dör
- Lite trött och hinner inte prata
- Fokuserad
- Vid 16p dog på grund av långsam respons
    - Skrattade och sa att det typ inte går att få mer
- Upptäckte att man kan kedja kommandon länge

Discussion

A. Overall, what were your thoughts about the game? / SWE: Vad är dina tankar om spelet generellt?:

Det var kul när man började komma in i det. Hög delay vilket är lite av utmaningen.

B. How did you experience the game understood your commands? / SWE: Hur upplevde du att systemet förstod dina kommandon?:

9/10 bra, 1/10 inte alls. Gick aldrig åt fel håll.

C. How responsive did you feel the system was? / SWE: Hur responsivt upplevde du att systemet var?:

Väldigt responsiv men hög delay. (?)

D. Did it feel like you had enough time to exercise the commands? / SWE: Kändes det som att du hade tillräcklig tid för att utöva kommandon?:

Ja men svårt vid 15p att tänka ut och interagera.

Follow-up questions / Other (A/B/C/D + Number, Follow-up question, answer):

Participant ID: P6

Test group: Exp Session: B

Observation

- Stakade sig
    - Systemet gjorde rätt ändå
- Dör vid 15 poäng och skrattar för att responstiden var så långsam att han förstod att han skulle dö innan han faktiskt dog
- Slår upp huvudet I liten frustration över att han dog.
- Sade fel kommando vilket ledde till en kedja av random kommandon, för att rätta till felet.
- Hittade knep att stega kedjor så att ormen rör sig mest horisontellt (större sidan av planen) och fick 17 poäng.
- Glad att ha slagit rekord.
- Fortsatte med nya tekniken nästa rundan.

Discussion

A. Overall, what were your thoughts about the game? / SWE: Vad är dina tankar om spelet generellt?:

Det kändes som att det var mycket längre delay denna gång. I princip omöjligt att kontrollera ormen om man inte köade kommandon.

B. How did you experience the game understood your commands? / SWE: Hur upplevde du att systemet förstod dina kommandon?:

Den förstod bra.

C. How responsive did you feel the system was? / SWE: Hur responsivt upplevde du att systemet var?:

Långsam på att svara

D. Did it feel like you had enough time to exercise the commands? / SWE: Kändes det som att du hade tillräcklig tid för att utöva kommandon?:

I slutet var det tajt, men egentligen ja.

Follow-up questions / Other (A/B/C/D + Number, Follow-up question, answer):

Appendix E – Thematic codes

Theme	Category	Code	Occurance
Negative UX	Negative feelings	Frustration	30
		Dissapointment	9
		Irritation	2
		Panic	2
	System behaviors hindering/worsening the interactions	Slow SRT	45
		Bad CSR	25
		Bad feedback from system	8
		Difficult to control with precision	3
		Hard to predict SRT	1
	Unnatural interactions	Planning interaction	21
		Mental confusion	20
		Physical or mental fatigue	2
		Broken "flow"	2
		Uncomfortable	1
Positive UX	Positive Feelings	Fun and joy	8
		Motivated	1
		Calm	1
	System behaviors supporting/enhancing the interactions	Queue-system	9
		Good CSR	8
		Good SRT	7
	Natural interactions	Player in "flow"	6
Neutral UX	Ambiguos feelings	Bored in the beginning	4
		Exciting	1
	System behaviors neither supporting/enhancing nor hindering/worsening the interactions	Adequate SRT	3
		Adequate CSR	3
		Experienced different SRT from start to finish	2
		Impossible at one point	2
		Surprised by queue-system	1
		Impressed by queue-system	1
		Impressed by voice interaction	1
		Does not utilize queue-system	1
	Interactions neither identifiable as natural nor unnatural	Involving the whole body	3
		Practice gives more control	3
		Focusing	2

Note: The ’Neutral UX’ was discarded in the final stage of the analysis.

Appendix E – Reflection Document of Christian Lindberg

"How does your study correspond to the goals of the thesis course? Why? Focus on the goals that were achieved especially well and those that were not well achieved."

I feel that this thesis, and the work performed, fulfills all the goals of the course to some degree. One of the goals I fulfilled the best is that the work was done independently. We received suggestions, feedback, and tips from our supervisor of course, but not more than any other thesis project would. In the end it was we ourselves who chose how to improve the thesis based on the feedback received. There was a lot of work put into the whole project and we managed the time we had well. We communicated our progress well with our supervisor and held the deadlines.

The goal I feel that we have been struggling the most to fulfill is that of searching and finding relevant research and literature. We struggled to find research that was concerned with the same context as our own. That is voice as the primary control modality in a real-time application. This is probably because voice does not seem to be considered viable in that regard. However, I think we failed in expanding our search pattern. Voice control might have been researched more in other areas, such as robotics, that we could have used as a base for the thesis.

"How did the planning of the thesis go? What could have been done better?"

Overall, the planning of the project was good and followed. It was hard to predict how much time the various steps and phases would take, and we underestimated the time required to do things. This led us to taking on more than we maybe should have. By having several data collection methods, with both quantitative and qualitative data to analyze, the workload increased more than we anticipated. Having several methods leads to a lot more to read up on, write about and motivate in the paper. It is hard to know before having done it "for real", and not just in a preparatory course such as METOD. Another area I would improve on, if we did the project all over, is the initial research. I would have liked to begin the work much earlier than we did as I knew that the research would take a lot of time. However, I felt that we could not. Since the project idea was our supervisor's, and it did not have that much of a description when we selected it, we felt that we had to have the first meeting before deciding anything. Furthermore, the project struggled with finding its identity in the beginning. Which, I identify as the consequence of not doing the initial research properly. The project's approach went from a design science to empirical research and trying to keep a relevance to 5G felt artificial for a long time.

"How does the thesis relate to Your studies? What courses and areas have been the most relevant doing your thesis?"

This thesis relates very well to my studies at DSV. I studied the bachelor's program Computer Game Development which also consisted of studies within Computer and Systems Sciences. The topics this thesis relates to the most are Human-Computer interaction, Game Development, and Game Accessibility. To investigate the interactions in the experiment, we had to know how research and literature defines interactivity and user experience. In the course MAPP I learned specifically about this and with that knowledge we define what this thesis regards as good or successful interactions. Furthermore, with the knowledge from several different courses of project work, using the program

Unity, and the programming courses in general we were able to develop our own game prototype that utilize a cloud-based ASR for voice interaction. Knowledge gained from the course PROD inspired us to adapt an accessibility approach in the initial exploration of what new emergent technologies could be enhanced by 5G.

"How valuable is this thesis for Your future work and/or studies?"

The work and studies I have conducted during this project will be valuable in my coming studies for a master's degree. I will be studying interactive media technology the coming years and the work done on this thesis has given me valuable knowledge and practice I will be able to draw upon. Knowledge regarding scientific methodology, research in the areas of accessibility, human-computer interaction, user experience and voice interaction. I have had valuable practical training in all the scientific methods we conducted. E.g., conducting an experiment, note-taking during observation, data analysis through thematic analysis and descriptive statistics. I feel very confident beginning the master's program with the new knowledge I have gained.

"How satisfied are You with the execution and result of the thesis? Why?"

The execution of this thesis was good considering this was our first ever thesis. But it is certainly not perfect. It has felt as if we have constantly learned and realized new things, always editing, and changing the thesis and never feeling as if it is finished. I am especially satisfied with how the data collection went, even though we identified areas that could have been improved. A lot more could have gone wrong but did not, and I attribute that to our hard effort. Thanks to the well-developed game and well-made data collection, we could write a good discussion and argument for our conclusions, and therefore am I very satisfied with the result of the thesis.

Thank you for a great education!

Christian Lindberg June 2021

Appendix F – Reflection Document of Joakim Linna

"How does your study correspond to the goals of the thesis course? Why? Focus on the goals that were achieved especially well and those that were not well achieved."

While I think we were able to reach all of the goals to some degree, the implementation of scientific methods is what I see us doing especially well. Even though we did not have any experience at all with conducting experiments, we designed both the experiment and developed the systems with very little input from external sources. I have to admit that our contacts at Ericsson were helpful with guidance at first, but one meeting was enough. I am very proud that we were able to handle it professionally and I am especially proud of us doing it mostly independently.

What I am a bit disappointed with was my own inability to find and summarize scientific literature. To be clear I did do it, but it was much more difficult than I expected. I have realized that while I'm interested in reading about a wide range of subjects, I need to practice summarizing it. Often times I also searched for relevant information but did not feel like the literature I found was appropriate. I intend to practice reading and retelling information.

"How did the planning of the thesis go? What could have been done better?"

The planning was not really an issue through any part of the project. We made a plan in the beginning that gave us room to breathe if we ran into any issues. When we were approaching a deadline, we were mostly satisfied with the work and did not have to worry too much. We did however run into some issues the day before the experiment. We had not planned much testing of the tools which resulted in us finding experiment breaking bugs less than 24 hours before the experiment was supposed to start. I had to stay up so late I overslept and ended up being late to the experiment. If I were to do something similar again, a playtest for a game for example, I would plan more time for testing the system before the playtest.

"How does the thesis relate to Your studies? What courses and areas have been the most relevant doing your thesis?"

We chose to discuss video games mainly because we are game developers. After we had finished developing the game, I immediately started working on a game for my internship course. Since I had just finished a game project, I was acutely aware of the difficulties I had and what I needed to change for this new project. That is a huge thing for me, finishing a project. I often end up switching project before the last one is done, so each finished project is a huge success. I think the most relevant course for this project was SPM as the workflow was very similar. We had requirements we needed to fulfill, and we had little time to do it. We wrote down what the game requirements were, started looking through documentation, and went through multiple prototypes before we were satisfied.

"How valuable is this thesis for Your future work and/or studies?"

Since I am looking to become a developer at a game studio, it is vital that I can show my passion, knowledge, and experience through produced works. Most of the time this would be game projects, such as "Voice Snake", but I think the subject of this study might be interesting to game development companies. I also found the function of voice control as a primary control modality very interesting,

and I want to explore this function further. Hopefully, the experience and knowledge I have gained through this project will serve me well if I work with voice user interfaces again.

"How satisfied are You with the execution and result of the thesis? Why?"

I am in general satisfied with the execution. We did what we intended to and managed the scope decently. I wish though, that our experiment would have had a wider scope. I do not feel satisfied with the reliability concern of only having 6 participants who are all W.E.I.R.D. The results also disappointed me slightly. I would have preferred if we could have proved that 5G makes a significant impact on voice user interface, and furthermore that VUIs would be more viable. It is unfortunate that the technology is still not available to have a smooth experience. Like I stated earlier however, I am proud that we were able to independently plan and manage a study of this scope as we have such little experience in the field. I am very happy that I worked with Christian as his knowledge and skills proved mighty useful and the teamwork was solid through the whole project.

Thank you for these past three years.

Joakim Linna June 2021