

System porting to mobile devices at the example of the SEE project

Master Thesis

Roman Gressler

Matriculation number: 3217822

June 16, 2022



Faculty 3 — Mathematics and Computer Science
Computer Science

1. Supervisor: Prof. Dr. Rainer Koschke
2. Supervisor: Dr. Robert Porzel

ABSTRACT

TODO: Hier das Abstract der Arbeit. Kann deaktiviert werden.

ERKLÄRUNG

Ich versichere, diese Arbeit — sofern dies nicht explizit anders gekennzeichnet wurde — ohne fremde Hilfe angefertigt zu haben. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen sind, sind als solche kenntlich gemacht.

Bremen, den June 16, 2022

Roman Gressler

ACKNOWLEDGEMENT

I want to thank me and only me for putting all this hard work in.

CONTENTS

1	Introduction	1
1.1	Motivation	1
1.2	Research Question	1
2	Concept	3
2.1	Interface	3
2.2	Interaction	7
2.3	Requirements	8
3	Implementation	11
3.1	Mobile Player	11
3.1.1	PlayerMenu Script	12
3.1.2	Prefab interactions	12
3.2	Player Movement	12
3.3	Mobile Menu	12
4	Evaluation	13
4.1	SEE Desktop	13
4.2	Aim and hypothesis	14
4.3	Experiment set up	16
4.4	Realization	19
4.4.1	Survey tool	19
4.4.2	Questionnaires	20
4.4.3	Pilot study	22
4.4.4	Final experiment set-up	23
4.4.5	Execution	24
4.5	Results	25
4.5.1	Demographic data	26
4.5.2	Performance	28
4.5.3	Usability	32
4.5.4	Impact of experience	38
4.5.5	Comments of the subjects	41
4.6	Threads to validity	43
4.6.1	Internal Validity	43
4.6.2	External Validity	45
5	Conclusion	47
5.1	Outlook	47
A	Glossary	49
B	Acronyms	51
C	List of Figures	53
D	List of Tables	55
E	Related Files	57
	References	61

INTRODUCTION

1.1 MOTIVATION

1.2 RESEARCH QUESTION

The central research question of this thesis is: Are Android smartphones suitable of working with *Software Engineering Experience* (SEE)?

To answer this question this thesis will discuss a general concept to implement SEE on mobile devices in chapter 2. Afterwards the specific details of implementation will be introduced in chapter 3. The research question will then be answered in chapter 4 with an evaluation where the implemented mobile version of SEE will be compared with the desktop version. Last but not least this thesis will discuss further research and give a conclusion in chapter 5.

SEE: An interactive software visualization that uses the code-city metaphor and enables collaborative multiplayer interactions via multiple platforms like desktop, virtual reality and soon Android devices.

2

CONCEPT

In this section a concept of a mobile SEE version will be presented. Therefore, a prototype will be created to point out the features that a mobile version of SEE requires.

Prototypes are a common way to express the needs of a system. It is a low-cost way of planning an implementation, that can highlight challenges regarding constraints of a system early on.

Even though a prototype will never be able to show every aspect and need of a complex system, it should still help to answering questions like: How should the system feel? How should it be implemented, and what are the key features? [Houde and Hill \(1997\)](#)

SEE is meant to be used by multiple platforms such as desktop devices, mobile devices and virtual reality devices. Each device has different interaction constraints. While a desktop user will control the player with mouse and keyboard a mobile user will interact with virtual joysticks on a touchscreen. Selecting nodes of a [Code-City](#) will be done by clicking it with a mouse on desktop devices, while a mobile device will require a touch input.

[Code-City:](#) In the Code-City metaphor, software components are represented by buildings in a city, and the properties of these buildings can express different metrics of the software. For example the height of a "Node" could represent the lines of code in that class.

2.1 INTERFACE

In the following a paper prototype will be presented that marks out a concept for the mobile interface. Since the field of mobile development is quite young there few guidelines regarding the design of mobile device interfaces. A guideline that is widely accepted is problematic to find. [Renaud and Van Biljon \(2017\)](#), [Punchoojit and Hongwarittorn \(2017\)](#)

Major differences to desktop environments are the screen size, forms of input and input feedback. To assure as much space is used for the actual interaction of the app the menu should just take as much space as needed. As a study has found out, a size of at least 8*8 mm is needed to reduce error rates selecting the right button. [Conradi et al. \(2015\)](#) [Parhi et al. \(2006\)](#) TODO WEITER AUSFÜHREN SHORTCUTS WIE STRG Z NICHT MÖGLICH [Adipat and Zhang \(2005\)](#)

Moving the player will be handled with virtual joysticks as seen in figure 2.1. The left joystick will move the player through the virtual room and the right will move the camera angle or in other word the direction the player looks at. The joysticks are placed in the left and right corner and should just take as much space as needed to be handled

comfortably. This way the player is able to navigate through the virtual room with his/her thumbs while still having enough space to work on the Code-City.



Figure 2.1: Joysticks for moving in SEE

The menu on the top left side seen in figure 2.2 will be called “quickbar” further on. The quickbar can be minimized to save screen space when not needed. The quickbar is designed to offer more general functions that are needed in various situations. Because there are no shortcuts on mobile devices each function has to have a button to be activated.

The functions are redo and undo which will do an action undone again or revert an action. Then there is a camera lock that will lock the players perspective to a certain Code-City so that the player can only move around the selected city and move closer or further away from it. The next function is to rerotate a Code-City. That means the Code-City that was last rotated will be set back to its initial state of rotation. Last but not least there will be a button for recentering the city, which will work quite similar to the rerotate button and center the last moved Code-City. The button on the right can be used to collapse or expand the quickbar.

On the top right side another menu will be placed that contains different interaction modes. By clicking a button an interaction mode will be selected and moved to the top right corner. Also, the menu will be collapsed and only the buttons regarding the selected interaction mode shall be shown. By clicking the button on the top right again the menu shall expand and the other interaction modes shall be selectable. The other buttons shall be kept in the same order to reduce confusion of the user.

The first interaction mode, seen in figure 2.3, is for selecting nodes. Nodes can be selected by being touched and deselected by being touched again. There can be multiple nodes selected at once. The hole selection can be deselected by clicking the deselect button next to the

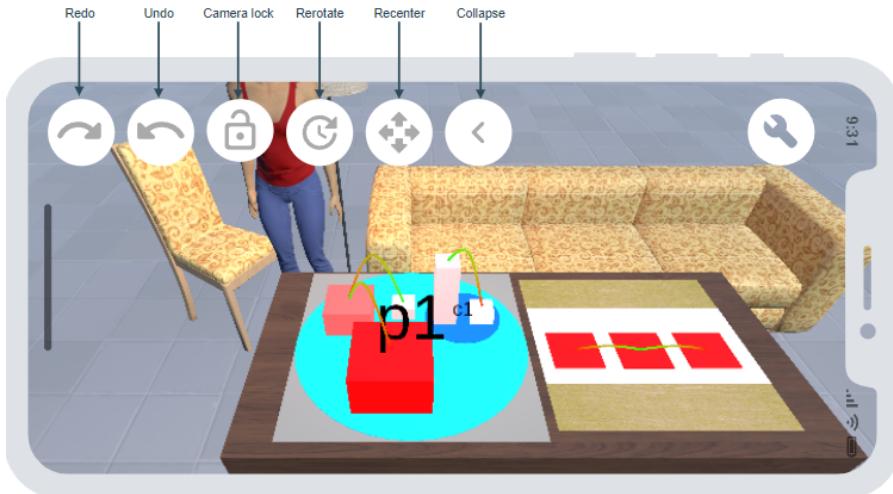


Figure 2.2: Quickbar for various interactions in SEE

select interaction mode button. Selected nodes shall be highlighted with a different node color and also display their name.

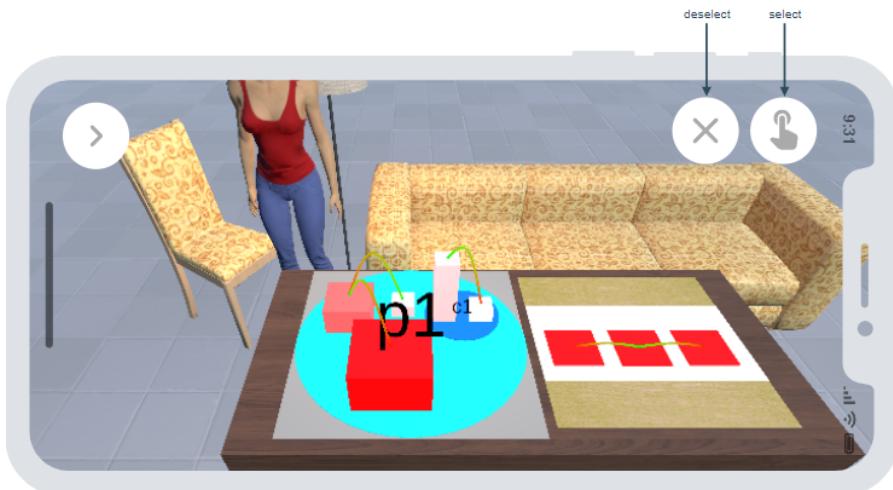


Figure 2.3: Selection mode in SEE

The second interaction mode, seen in figure 2.4, is for deleting node. It does not need additional buttons. Node will be deleted by being touched.- Unlike in the desktop version there will not be a group deletion interaction because it would require an additional menu panel. The added functionality would be minimal and selecting a group of nodes, confirming and finally deleting would require a handful more steps and would therefore most likely not be used.

The following interaction mode, seen in figure 2.5, is dedicated to the nodes and edges of a Code-City. Starting on with the “add node” button on the right. When activated the user can create new node by clicking on a certain spot on the Code-City plane. The following button on the left is for adding edges. By selecting two nodes a new edge will be created between them. Then, the button one further on the

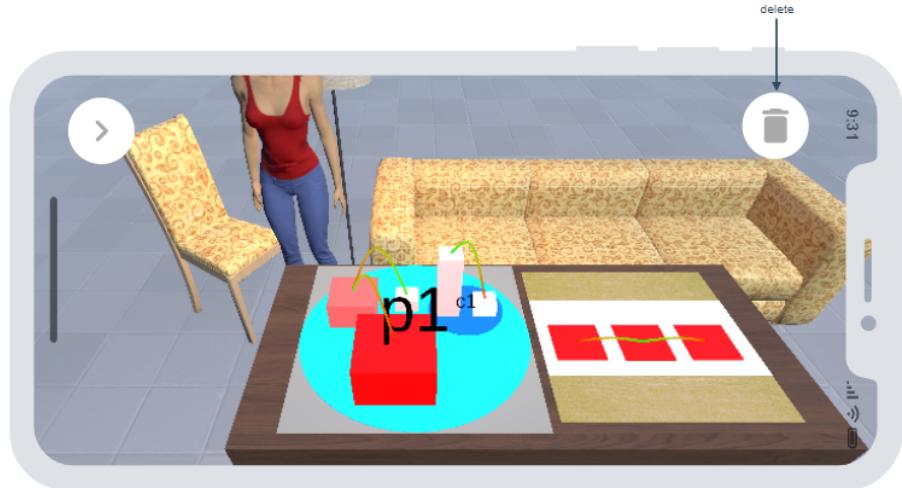


Figure 2.4: Delete mode in SEE

right is for editing nodes. By touching a node a window will pop up that allows the user to edit the node by changing its name and its type. Last but not least the button on the left-hand side will be used to scale nodes. That means the node height and width can be adjusted by first selecting it via touch and then hold a corner and slide it further away from the node center to increase the size or slide it towards the center to decrease the size of the node. Each button of the node interactions will be marked green after being pressed to indicate that it is active.

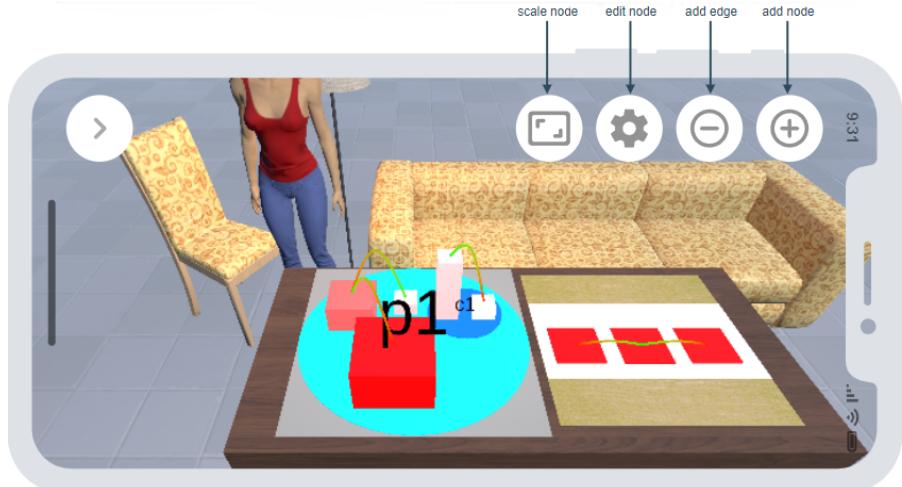


Figure 2.5: Node interactions in SEE

Then there will be a button for rotation interactions that can be seen in figure 2.6. Starting with the first activatable button that lets the user rotate the hole Code-City by touching any point on it and then sliding away from that point. Similar to that there will be a button that lets the user rotate just a single node on the Code-City. In addition to that there will be a button that activates the so-called “locked-rotation” mode. While in “locked-rotation” mode the rotation of a node or Code-City

will be done in eight predefined steps to a full rotation. Each step will have the same 45° range. The last button of this group will be for changing the center of the rotations. There are two options: the first option is a center of rotation in the middle of the Code-City and the second is in the middle of a node selection made with the interactions seen in figure 2.3. The second option can be activated by pressing the last button.

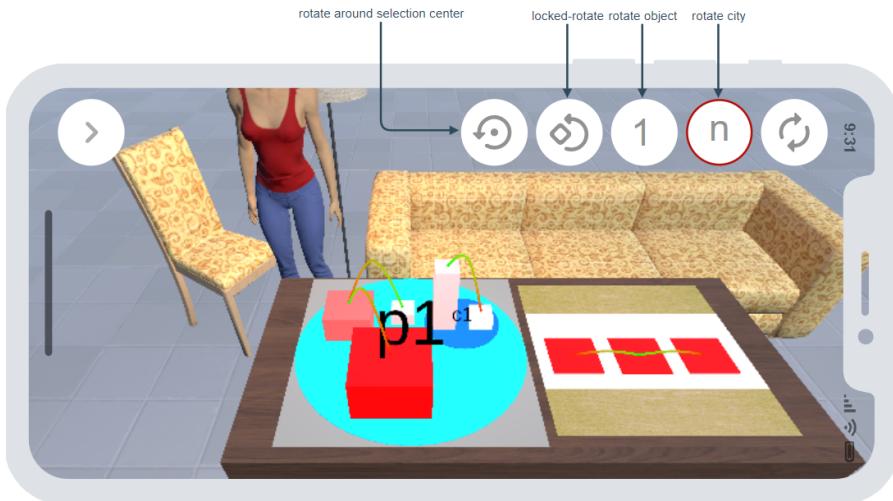


Figure 2.6: Rotation mode in SEE

The last interaction group, seen in figure 2.7, is for moving the Code-City or a single node. The move interactions are quite similar to the rotation interactions. There will be a button to move a hole Code-City as well as a button to move only single nodes. In addition to that there will be a button that restricts the movement of the Code-City or node to a predefined direction. The directions will be again in 45° angles and objects can be moved on a straight line on that angle. Moving a node or a Code-City can be achieved by touching and holding it and then moving it to the desired position.

2.2 INTERACTION

Smartphones are quite limited in space and there are few input possibilities. Unlike a desktop computer there is no mouse and there is no physical keyboard. Smartphones use virtual keyboards but due to the restriction of screen space the keyboard is hidden most of the time. Which would make keyboard shortcuts uncomfortable because the user has to open the keyboard first. Therefore, smartphones need different ways of interaction such as touch gestures.

Zooming in to a Code-City happens by scrolling on a desktop environment. There is no option to scroll on mobile devices, but there are at least two popular alternatives. The first option would be to double tap on the Code-City to zoom in. The double tap would zoom in, in



Figure 2.7: Movement mode in SEE

predefined steps and after reaching a certain level of closeness it would trigger to zoom out again. In SEE zooming in, in predefined steps might not be precise enough because there could be a quite large Code-City or a rather small one. Finding predefined steps that would fit every situation is rather hard. Therefore, a second option by zooming in with a two finger gesture might be better. In this option the user uses two fingers and slides them towards each other to zoom in or slides the two fingers away from each other to zoom out. This way there are no predefined steps necessary and zooming interactions can be done precisely.

2.3 REQUIREMENTS

In the following a list of requirements will be given, which will specify in detail what the implementation of a mobile version has to take care of. The list will be referred to multiple times in the upcoming realization part in chapter 3. Requirements are essential for the planning phase as they give a good fundamental structure for the developer to rely on. [Robertson and Robertson \(2012\)](#); [Stevens and Pooley \(2005\)](#)

Android: An operating system for portable device such as smartphones and tablets.

- [R1] The application shall run on *Android* devices
- [R2] The application shall be controlled via touchscreen
 - [R2.1] The player and camera shall be moved with virtual joysticks
 - [R2.2] Needed shortcuts of the desktop version shall be handled with buttons
 - [R2.3] Zooming shall be handled with a two finger gesture
- [R3] The user shall be able to select a node of a Code-City
 - [R3.1] After selecting the name of the node shall be shown

[R3.2] The user shall be able to deselect single nodes or a group of nodes

[R4] The user shall be able to delete nodes

[R5] The user shall be able to interact with nodes

[R5.1] The user shall be able to add nodes

[R5.2] The user shall be able to add edges

[R5.3] The user shall be able to edit nodes

[R5.4] The user shall be able to scale nodes

[R6] The user shall be able to rotate a Code-City

[R6.1] The user shall be able to rotate a Code-City in 45° steps

[R6.2] The user shall be able to rotate single objects

[R6.3] The user shall be able to rotate around a center of selected nodes

[R6.4] The user shall be able to undo the rotation

[R7] The user shall be able to move a Code-City

[R7.1] The user shall be able to move single object of a Code-City

[R7.2] The user shall be able to restore the Code-City initial position

[R7.3] The user shall be able to move a Code-City or single node in predefined directions

[R8] The user shall be able to undo and redo actions

[R9] The user shall be able to lock the camera to a selected Code-City

3

IMPLEMENTATION

3.1 MOBILE PLAYER

In this section the mobile player *Prefab* will be discussed. *SEE* is supported on multiple platforms and with each platform having different requirements each platform needs to be divided. Therefore, each platform gets its own player Prefab. The mobile player prefab for the Android version of *SEE* can be seen in figure 3.1.

Prefab: *Predefined GameObject that can be loaded into a scene.*

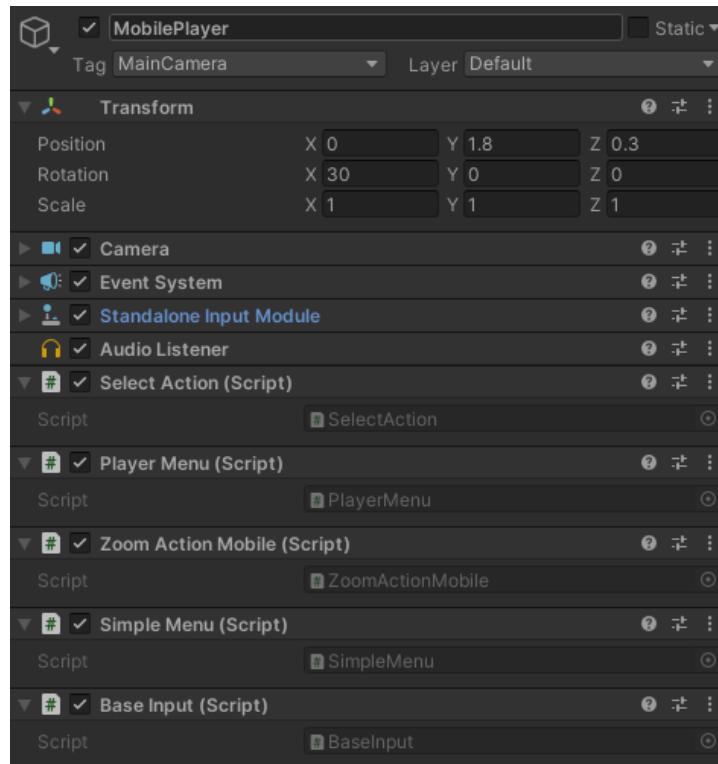


Figure 3.1: The mobile player Prefab

3.1.1 PlayerMenu Script

3.1.2 Prefab interactions

3.2 PLAYER MOVEMENT

3.3 MOBILE MENU



Figure 3.2: The *quickmenu* on the left top side of the mobile device. Pressing the button with an orange marked arrow will expand the menu.

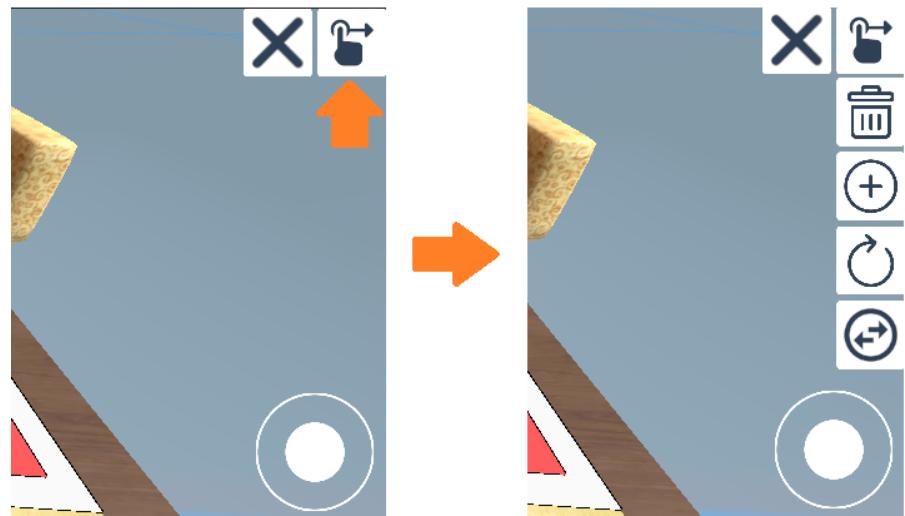


Figure 3.3: The player interaction menu on the right top side of the mobile device. The button on the top right side indicate the active interaction mode. Pressing the same button also expands the menu.

4

EVALUATION

In the following chapter the mobile implication of SEE will be evaluated in a user study. Therefore, the mobile application will be compared with the desktop version.

This chapter will start with a description of the desktop version and its main differences in section 4.1. Continuing with a defined aim and precise hypotheses for the user study in section 4.2. After sketching the first experiment set up in section 4.3 the actual experiment set up will be discussed in detail in section 4.4 including the used survey tool, questionnaires and the pilot study. This chapter will then be closed by discussing the results of the user study in section 4.5 and talking about the threads of validity in section 4.6.

4.1 SEE DESKTOP

In this section the desktop version of SEE will be explained. In this evaluation the mobile version of SEE will be compared with the desktop version. Therefore, it is necessary to take a deeper look at the differences between those two versions. Especially at how the interactions differ and what impact it could have on the user experience.

One outstanding difference from the desktop version to the mobile version is the selection of the interaction modes. While in the mobile version the menu for the interaction modes is always visible in the desktop version by pressing space a menu screen opens as seen in figure 4.1. Alternatively interaction modes can be changed by pressing one of the “1-9” keys, which, however requires the user to memorize which number belongs to which mode.

Another difference is the type of user input. The desktop version uses mouse hovering to display the name of a hovered *Node* or *Plane*. This is a faster method than touching the object first in the mobile version. In addition to that in the mobile version the object also has to be deselected otherwise there will be a lot of Node and Plane names displayed, and it will soon get quite messy. Also, the precision of object selection differs because touch input can never be as precise as selecting with a mouse cursor. This could force the mobile user to zoom further in because with a touch input it will not be possible to select small objects like it might be with a cursor. Which, of course, would require more time.

Node: A point in a diagram where lines intersect. In “SEE” it usually displays a software class

Plane: An area that bundles “nodes”. Could for example represent a namespace.

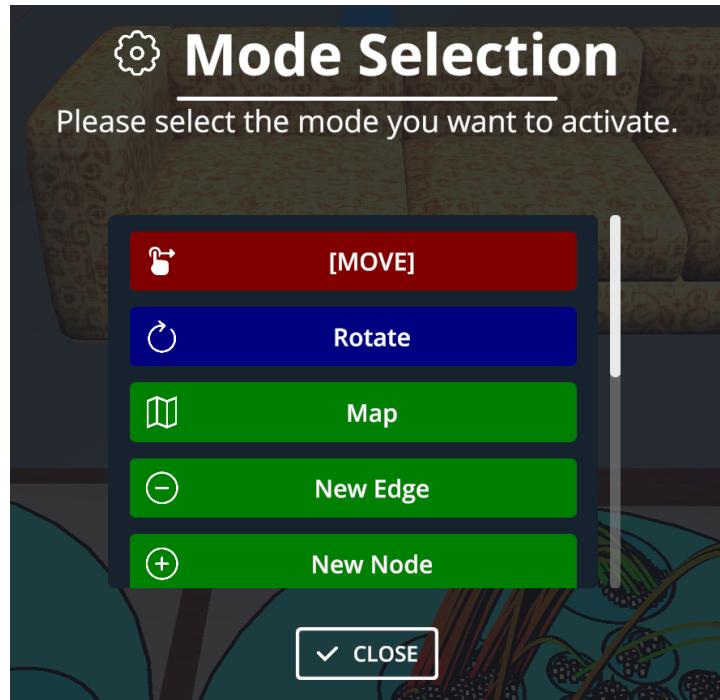


Figure 4.1: The desktop menu for selecting interaction modes.

One more key difference is the available keyboard for desktop users. It allows using *Shortcuts*, which makes some menu items unnecessary but also requires the user to memorize those Shortcuts. The desktop version for example uses the “R” key in the move and rotation mode to recenter or rerotate a Code-City. In the mobile version on the other side the user will find a button for both actions. With the right amount of training both actions should probably equal in the amount of time they need but the mobile version sacrifices screen space for those buttons. If however the user has to type more text like in renaming objects, the common desktop keyboard should come in handy as a study from [Kim et al. \(2014\)](#) shows that even at a same keyboard size, a virtual one will lack in productivity.

Shortcut: A combination of key that will call an action like for example “ctrl” + “c” for coping a text.

4.2 AIM AND HYPOTHESIS

The aim of this user study is to answer the research question discussed in section 1.2. In order of answering the research question the finished prototype of the mobile extension shall be evaluated. Therefore, the system shall be compared on Android smartphones as well as desktop computers. Comparing these two use cases shall give insight on how much impact the constraints of mobile devices have on the usability and overall user experience. To measure the difference between the desktop and the mobile version the following hypotheses will be used. The two aspects performance and usability will be measured in the

following study and each aspect will have a null hypothesis and an alternative hypothesis.

- a) **Performance:** The time required for a task in SEE desktop will be called t_D and for mobile t_M .
 - *Null Hypothesis H_{a0} :* The time required in SEE desktop is higher or the same as the time required in SEE mobile: $t_D \geq t_M$
 - *Alternative Hypothesis H_{a1} :* The time required in SEE desktop is lower as the time required in SEE mobile: $t_D < t_M$
- b) **Usability:** Two aspects are measured for *Usability*. First the *After-Scenario Questionnaire (ASQ)*-Score as a *Post-Task* result and second the *System Usability Scale (SUS)*-Score as a *Post-Study* result.
 - i) **ASQ:** Once again the aspect has to be split into three child aspects, because the three questions of the ASQ are independent:
 - 1) The ASQ-Score for *complexity* for SEE desktop is called A_{cD} and for SEE mobile is called A_{cM}
 - *Null Hypothesis H_{b0} :* The ASQ-Score for *complexity* is higher or even for SEE mobile than on SEE desktop: $A_{cM} \geq A_{cD}$
 - *Alternative Hypothesis H_{b1} :* The ASQ-Score for *complexity* is higher for SEE desktop than on SEE mobile: $A_{cD} > A_{cM}$
 - 2) The ASQ-Score for *effort* for SEE desktop is called A_{eD} and for SEE mobile is called A_{eM}
 - *Null Hypothesis H_{b0} :* The ASQ-Score for *effort* is higher or even for SEE mobile than on SEE desktop: $A_{eM} \geq A_{eD}$
 - *Alternative Hypothesis H_{b1} :* The ASQ-Score for *effort* is higher for SEE desktop than on SEE mobile: $A_{eD} > A_{eM}$
 - 3) The ASQ-Score for *information* for SEE desktop is called A_{iD} and for SEE mobile is called A_{iM}
 - *Null Hypothesis H_{b0} :* The ASQ-Score for *information* is higher or even for SEE mobile than on SEE desktop: $A_{iM} \geq A_{iD}$
 - *Alternative Hypothesis H_{b1} :* The ASQ-Score for *information* is higher for SEE desktop than on SEE mobile: $A_{iD} > A_{iM}$
 - ii) **SUS:** The SUS-Score is called S_D for SEE desktop and S_M for SEE mobile.

Usability: A term that describes how well a (software) system can be used.

ASQ: A post-task questionnaire consisting of three questions that is used to access how difficult a user perceived a task (See "Post-Task").

Post-Task: A questionnaire that is taken after every task of an experiment.

SUS: The System Usability Scale consists of ten questions that measure "Usability"

Post-Study: A questionnaire that is taken after every block of an experiment.

- *Null Hypothesis H_{e0}* : The SUS-Score is higher or even for SEE mobile than for SEE desktop: $S_M \geq S_D$
- *Alternative Hypothesis H_{e1}* : The SUS-Score is lower for SEE mobile than for SEE desktop: $S_M > S_D$

The experiment will be participated different groups:

1. **SEE-developer**: They are already experienced with SEE-desktop. They are also experienced with software development and with first person games because they tried at least SEE itself, which counts as first person game experience.
2. **Non-SEE-developer**: This group has to be divided into four subgroups as follows:
 - **Software development and third person game experience**: They are more likely to understand the Code-City metaphor and are also more likely to be comfortable with the controls in SEE.
 - **Software development experience**: They are more likely to understand the Code-City metaphor and therefore might be able to find nodes faster.
 - **First person game experience**: They are also more likely to be comfortable with the controls in SEE.
 - **No experience**: They do not benefit from experience and therefore have to learn the most to interact in SEE

After the experiment the two versions of SEE can be compared as well as all above listed groups. This should give a detailed answer to the research question of this thesis.

4.3 EXPERIMENT SET UP

The system shall be tested in two groups each starting with a different device. Each group does the test on both devices, but one group will start with the mobile application and the other one with the desktop application. The participants will be assigned random to the groups. The testers will have various tasks to test the usability of the two applications. Afterwards the users will get a survey in English to document their impressions.

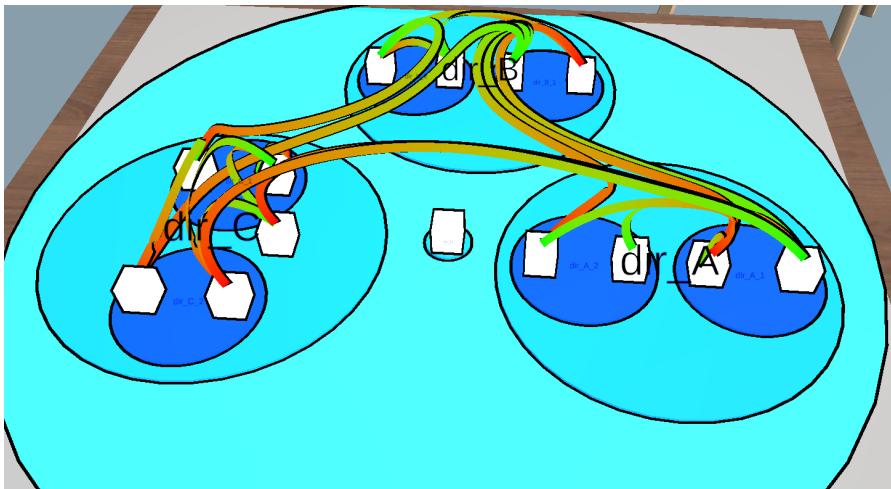


Figure 4.2: The first Code-City for the user study

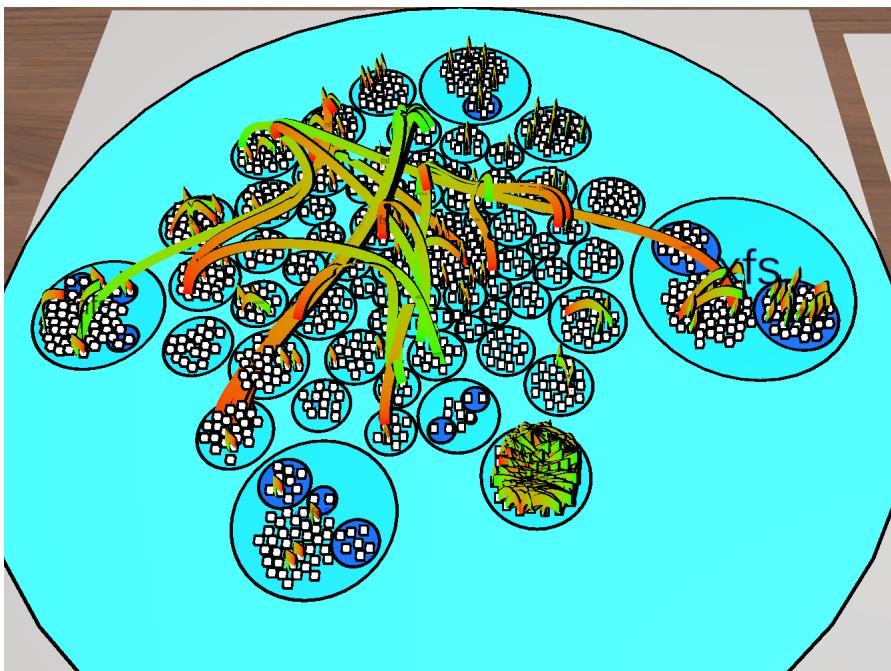


Figure 4.3: The second Code-City for the user study

In this survey the subjects will be asked various demographic questions as well as what Android device and version they will be using. In addition to that the subjects will be asked if they are experienced with SEE and if they are experienced with software development. Before the subjects will be asked to solve various tasks they will be asked to watch a short tutorial video on each application. After the video they will get a training task where every subject can get used to the system and ask questions if they have trouble solving the training task. The overseer will also make sure that every essential action will be practiced such as zooming and moving the Code-City. Figure 4.2 shows a small arranged

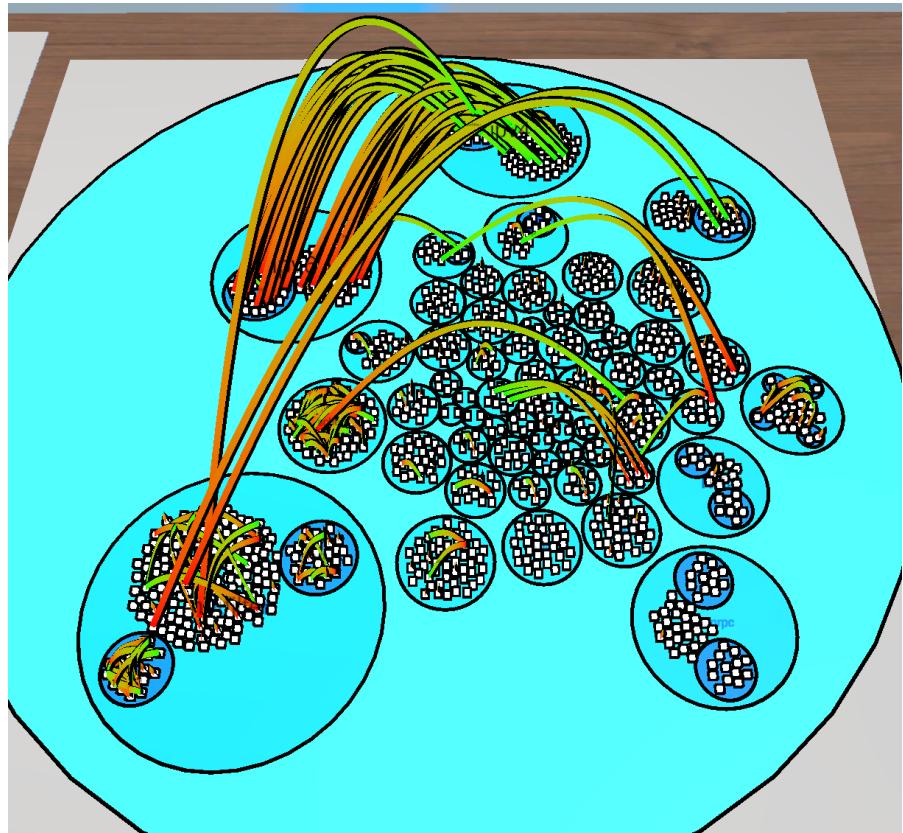


Figure 4.4: The third Code-City for the user study

Code-City that shall be used for the training tasks. The structure of the training Code-City is generic and follows a simple pattern. That shall ensure that the user can focus on the training and that the user does not get overwhelmed.

Following the first questions and the training, the subjects can start with the main tasks. For each application there will be two tasks and after each task the subjects will be handed a Post-Task questionnaire. Last but not least there will another questionnaire that aims to scale the Usability of the two applications. For the Post-Task questions the ASQ will be used and for the Usability questions SUS will be used. Both questionnaires will be discussed later on in section 4.4.2. For each main task the overseer will also take the completion time of every main task. The first and second task on the first device will be performed on the Code-City that can be seen in figure 4.3 and the third and forth task on device two will be performed on the Code-City that can be seen on figure 4.4. These examples are much larger than the training Code-City and represent real life code. The second Code-City shows the file system of Linux and the third one shows the network component of Linux. That way the tasks might reflect better on real world uses for SEE.

To not exhaust the testers too much the experiment shall not take longer than one hour. This also ensures that there is no to little variance

due to exhaustion. Each participant might have a different concentration span, but this shall not be the focus of this experiment.

4.4 REALIZATION

The following sections will cover the realization of the previously planned study. The choice of the used survey tool and questionnaires will be explained in section 4.4.1 and section 4.4.2. Afterwards a pilot study will be executed to test the study and possibly find missing aspects in section 4.4.3. Finally, the final experiment set up will be discussed in section 4.4.4.

4.4.1 Survey tool

As a survey tool Google Forms¹ will be used. The survey tool has to fulfill the following requirements:

- The study will be online because an overseer has to attend every experiment, and therefore it comes in handy to be flexible in terms of location. For this reason the survey shall be fully in a browser.
- The survey tool should be free to use.
- Subjects shall be anonymous.
- The results shall be exportable in a data format like *Comma-Separated Values (CSV)*
- Subjects should have the option to presave their answers. In addition to that answers should not be lost on reload.
- There should be an option to embed the introduction videos in the survey.

Google Forms fulfills all these requirements and will therefore be used. The final form can be seen in figure 4.5 which shows the intro of the survey and in figure 4.6 which shows the embedded intro video for SEE mobile.

CSV: A file format used for example to store table data. Each line represents a data record and each value is separated with a comma.

¹ <https://www.google.com/forms/about/> (last visit: 05.06.2022)

See Desktop - See Mobile

Thank you for participating in this study. The following questions will give insight on how well the different versions of SEE perform. The survey will take about an hour and the taken data will be kept anonymous.

SEE is project for code visualization that aims offer a collaborative space where developer can come together and have a look at various software projects. Therefore graphs are used that resemble cities. So called Code-Cities.

[In Google anmelden, um den Fortschritt zu speichern. Weitere Informationen](#)

[Weiter](#) [Alle Eingaben löschen](#)

Geben Sie niemals Passwörter über Google Formulare weiter.

Dieser Inhalt wurde nicht von Google erstellt und wird von Google auch nicht unterstützt. [Missbrauch melden](#) - [Nutzungsbedingungen](#) - [Datenschutzerklärung](#)

Google Formulare

Figure 4.5: The intro of the survey

See Mobile - See Desktop

[In Google anmelden, um den Fortschritt zu speichern. Weitere Informationen](#)

See Mobile

Please watch this short instruction of the app before you continue



[Zurück](#) [Weiter](#) [Alle Eingaben löschen](#)

Figure 4.6: The introduction video of the survey

4.4.2 Questionnaires

There will be three questionnaires used for the study that will be discussed in detail in the following. The study will start with a demographic questionnaire that covers general information about the

subject. After every task there will be ASQ and after every block of tasks for each of the two covered devices there will be a SUS questionnaire.

DEMOGRAPHIC QUESTIONNAIRE

The subjects shall start the survey with a demographic questionnaire. In that section they will be asked for their age, gender, the highest degree, experience with SEE, experience with first person video games, their Android device name, their Android version and their experience with software development. These specifications will be used to form different groups to see if there is any impact on the result of the following measurements. [Mclellan et al. \(2011\)](#) has shown that the user experience can have a significant impact on the SUS-score. It is therefore important to view the measured results in context of the paired demographic data.

The mobile version was only tested on a single Android device. It is likely that the performance of the application varies on different Android versions or devices.

POST-TASK QUESTIONNAIRE

The Post-Task questionnaire will supplement the Post-Study questionnaire on a micro level. The main focus here will be on single tasks, which allows to have a look at different aspects like effort, complexity and information provided by the system.

As a Post-Task questionnaire the ASQ will be used. The ASQ was first introduced in 1991 by [Lewis \(1991\)](#). It is designed for task based surveys and contains three questions. The ASQ will be used because it brings the following advantages:

- The questionnaire has been used many times over the years and has proven its validly ([Hajesmael-Gohari et al. \(2022\)](#); [Lewis \(1991\)](#); [Lewis \(1995\)](#)).
- With its three questions it is short and does not exhaust the subjects. This is especially important because it will be required to finish this questionnaire a total of four times.
- It fits well for this study because it is a questionnaire designed for task based evaluations.

The ASQ consists of three questions that scale from one to seven where one means “strongly disagree” and seven means “strongly agree”.

POST-STUDY QUESTIONNAIRE

The Post-Study questionnaire is mainly to obtain as much information about the Usability of the two systems as possible. The questionnaire can be longer than the Post-Task but still should not be to long to keep the processing time of the survey at around an hour.

The SUS questionnaire was first published in 1986 by John Brooke ([Brooke \(1996\)](#)) and is therefore widely used and proven as citations in

more than 1200 publications up until 2013 show (Brooke (2013)). The SUS is used in this study for the following advantages:

- It consists of ten questions and has to be done two times. Twenty questions in total fit well into the planned one hour total time span of this experiment.
- It has been made publicly available and is free to use (Brooke (1996)).
- It is widely used and therefore already proven to give useful results (Brooke (2013); Lewis (2018); Grier et al. (2013)).
- According to Peres et al. (2013) and Bangor et al. (2008) it is suited to compare two systems.

The SUS questionnaire will contain as already mentioned ten questions. Each question will give a five-level rating from “strongly disagree” to “strongly agree”. The results will then be combined into a Usability-score from zero to 100 where a high value represents a good result and a low value a bad one.

4.4.3 Pilot study

In a first test the pilot study was executed with one subject. Afterwards the study was discussed and checked for errors. It stood out that the example Code-City of task one was too different to the one in the second task. Therefore, the Code-City of the first task was exchanged with a larger and better comparable one. Further on a Code-City with 1288 nodes (see figure 4.3) as well as one with 1464 nodes (see figure 4.4) will be used.

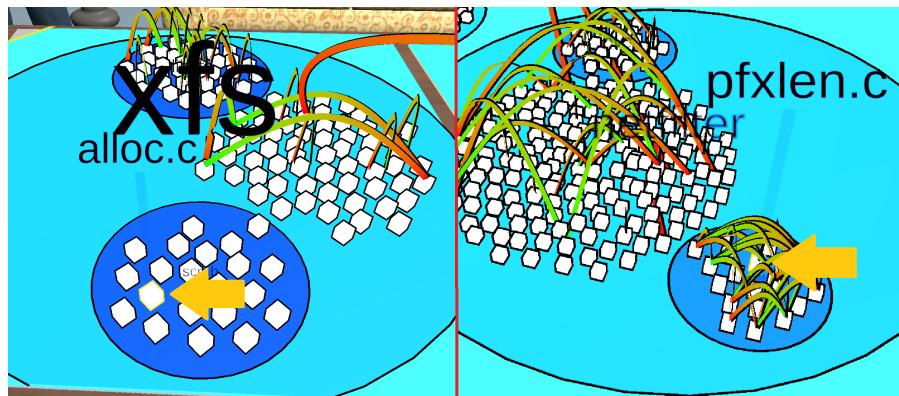


Figure 4.7: The two key nodes are marked with a yellow arrow

Also, the tasks were not comparable because they differed in the types of interactions they used. In one task the user was asked to rename a node and in the other one the user shall add four nodes. For

renaming a node the user has to use a keyboard which does not make it comparable to just click and add nodes in the second task.

To also ensure that task one and task three (first task of each version) were comparable both task are similar in their set-up as seen in figure 4.7. As one part of the task is to find the right node both are on a plane with a similar amount of nodes.

The same goes for task three and four (second task of each device) as the target planes that have to be found have a similar size. Both tasks also include a hint where to find the desired planes.

4.4.4 Final experiment set-up

After the pilot study the final set-up can be discussed. The demographic questionnaires will be designed as described in section 4.3. The final questions will be as listed below.

DEMOGRAPHIC QUESTIONS:

- Age
 - 0-15 years old
 - 16-30 years old
 - 31-45 years old
 - 46+ years old
- What gender do you identify as?
 - Male
 - Female
 - Other ...
 - Prefer not to say
- What is the highest degree or level of education you have completed?
 - Some High School (Hauptschule/Realschule...)
 - High School (Abitur)
 - Bachelor's Degree
 - Master's Degree
 - Ph.D. or higher
 - Prefer not to say
 - Other ...
- Questions regarding used hardware and experience
 - Are you already experienced with See?
 - Do or did you play first person video games?

- Do or did you develop software?
- On which Android device will you attend?
- Which Android version are you using?*

After filling out the demographic questionnaires the subjects will be asked to watch a short tutorial video. Which video they will watch first will depend on with group they are in. If they start with the mobile version of SEE they will be asked to watch [SeeMobile.mp4](#) otherwise they will be asked to watch [SeeDesktop.mp4](#). After completing their first training and their first two tasks, as described in table 4.1, they will be asked to watch the video they have not started with.

Nr.	Task	Expected time
Training	Navigate through the planes “dir_root” >“dir_B” >“dir_B_2”. On that plane select “b2_b.cpp” and rename it “b42”.	1 - 5 mins
	Detect the largest plane “xfs”. On that plane find plane “scrub”. Then find and delete node “alloc.c”.	0.5 - 5 mins
1	Find the plane with one blue child plane (“btrfs”). On the blue child plane “tests” add four new nodes.	1 - 5 mins
Training	Navigate through the planes “dir_root” >“dir_C” >“dir_C_2”. On that plane select “c2_b.cpp” and rename it “c42”.	1 - 5 mins
	Detect the largest plane “netfilter”. On that plane find plane “ipset”. Then find and delete node “pfxlen.c”.	0.5 - 5 mins
3	On the plane with the most edges (“ipv6”) find the smallest plane “ila” and connect all four nodes on it.	1 - 5 mins
4		

Table 4.1: The tasks used for the experiment. The device will be switched after task 2.

After each task the subjects will be asked to fill out a ASQ and after each block of task, as described in table 4.2, they will be asked to fill out a SUS questionnaire. Therefore, each participant has to fill out four ASQs and two SUS questionnaires. The full study should not take longer than an hour.

4.4.5 Execution

Before the study the subjects got an installation instruction as well as the two applications. For the execution of the user study every subject got one of two links to a Google form. Each form starts with a

Phase	Description		
	City	Group 1	Group 2
Pre-Experiment			Demographic questionnaire
Training	Figure 4.2		
Task 1	Figure 4.3		
ASQ			
Task 2	Figure 4.3	Desktop	Mobile
ASQ			
SUS			
Training	Figure 4.2		
Task 3	Figure 4.4		
ASQ			
Task 4	Figure 4.4	Mobile	Desktop
ASQ			
SUS			

Table 4.2: Experimental procedure per subject. The procedure is swapped per group.

different device. The forms were handed out in alternating order and the subjects should therefore be assigned random to the two groups.

The study was executed within a week and a total of 20 subjects participated. From the 20 subjects, two could not finish the tasks on their mobile device, which leaves $n = 18$ participants that finished the survey.

The overseer hosted an instance of SEE from his home network and watched the subjects doing their task. He also gave instructions for the training tasks and made sure that all essential interactions were trained such as zooming and moving the Code-City. Every subject got the same training.

4.5 RESULTS

The `calc_data.ipynb` script can be used to reproduce the result data for the following section. Also, all shown diagrams in this section can be reproduced with the script.

In the following the group that started the study with the desktop version of SEE will be called *Group 1* and the group starting with the mobile version will be called *Group 2*. The task order of the two groups remains the same. Both groups start with the same task. That way not only both groups but also the desktop and the mobile SEE version will be compared.

In the coming up section the *Mann-Whitney-U-Test* will be used multiple times to see if there are significant differences between to data

Mann-Whitney-U-Test: A test that compares two independent samples with ordinal scales to find out whether there is a difference between the two groups. The Mann-Whitney-U-Test does not require a normal distribution.

sets. The Mann-Whitney-U-Test brings the advantage that the tested data sets do not need to fulfill the requirement of a normal distribution ([Gibbons and Chakraborti \(1991\)](#)). This is important because the sample size of $n = 18$ is too small to assume a normal distribution. The used level of significance will be $\alpha = 0.05$.

4.5.1 Demographic data

In this subsection it will be verified if the two groups differ significantly in their demographic data. Therefore, a two-tailed Mann-Whitney-U-Test that checks whether a set of demographic values from *Group 1* ≠ a set of demographic values from *Group 2* will be used. Both groups have a sample size of $n = m = 9$ with $\alpha = 0.05$ which leads to a critical value of $U = 17$ for a two-tailed Mann-Whitney-U-Test ([Zar \(2010\)](#)).

GENDER

Of 18 participants, 17 reported being male. One participant preferred not to say his or her gender and can therefore not be grouped. Due to this fact a grouping does not make sense here in general.

AGE

The subjects were asked to choose their age in grouped options. Most of the participants were aged between 16 and 30 years. Only two participants were aged between 31 and 45 years. Therefore, the age of the participants should not fall into account of the results of the study.

HIGHEST DEGREE

Nine of the participants answered with “Bachelor’s Degree” for their highest achieved degree. It is therefore the most common option as pictured in figure [4.8](#).

To measure the Mann-Whitney-U-Test the options were put into an ordinal scale which follows the same order as in the survey. The test shows there is no significant difference between the two groups ($U = 33.0; p \approx 0.51$).

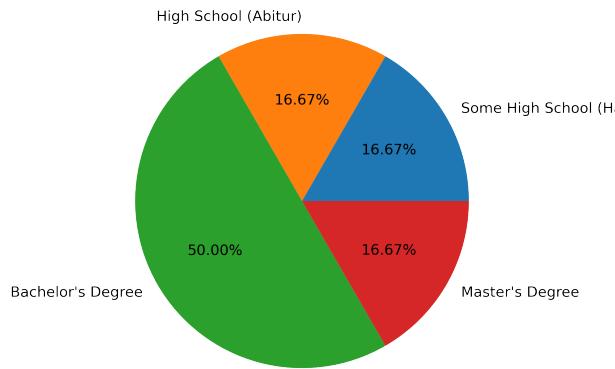


Figure 4.8: The distribution of the highest completed degrees of all 18 subjects

EXPERIENCE

The subjects were asked if they had experience regarding SEE, software development and first person video games. All but one subject had experience with first person video games and the single subject without experience had at least experience with SEE which could arguably count as a first person video game. Therefore, experience with first person video games should not have an impact on this study.

Five of 19 subjects had experience with SEE. The Mann-Whitney-U-Test shows there is no significant difference between the two groups ($U = 27.0; p \approx 0.23$). Even though there is no significant difference, four of the subjects with SEE experience are in *Group 2* and only one in *Group 1*.

The majority of ten subjects had experience with software development. The other eight subject had no experience. The Mann-Whitney-U-Test shows there is no significant difference the two groups ($U = 49.5; p \approx 0.38$).

ANDROID DEVICE AND VERSION

A large variety of 17 different Android devices were used. Therefore, the distribution of Android devices should not have impact on the study results. Nonetheless, it has to be considered that the devices “Huawei P10 lite”, “Huawei P20 Pro” “Samsung A52S”, “Samsung S9” and the “Poco X3 Pro” had performance issues. These performance issues showed in bumpy movement and player interactions. Unfortunately, it could not be measured how strong the impact was on the performance on the single phones. The threads of validity this could cause will be further discussed in section 4.6. Three of the devices with bad performance belong to *Group 1* and the other two to *Group 2*.

The Android version was not distributed as much as the Android devices used. The most common used one was Android 12, which is also the most current one to the time of this writing. Android 12 was used by seven subjects. The second most used version was Android 11. One subject used Android 7 on the “Huawei P10 lite”. Which is

remarkable because Android 7 was released in 2016² and the “Huawei P10 lite” was released a little later in March 2017³ but SEE still managed to work, if with lower performance. The distribution of the used Android devices can be seen in 4.9. The Mann-Whitney-U-Test shows there is no significant difference the two groups ($U = 39.0; p \approx 0.93$).

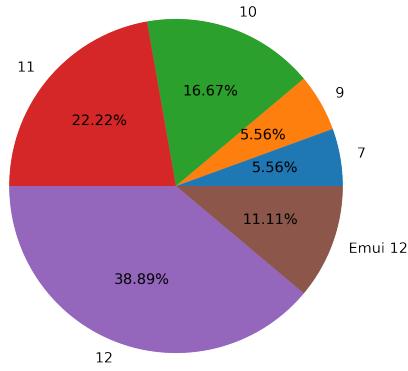


Figure 4.9: The distribution of the Android versions the subjects were using

Surprisingly there has been no significant difference with the demographic data between the two groups. In the following section the more exciting data will be analyzed.

4.5.2 Performance

For the performance the time needed for each task was measured. The time for the trainings task, however, was not measure because it would not fit the purpose of a training to be measured by time. Therefore, the overseer asked the subjects to tell him when they begin the task. The time was taken after the subject has read and understood the task, and it was stopped when the subject successfully finished. This will help to prove or discard hypothesis H_a .

Let's have a look at the two groups in general. Is there any significant difference? The total time needed for all subjects in the two groups can be seen in figure 4.10. The average time $Group 1$ needed to complete all tasks was $\bar{t}_1 = 487s$ and for $Group 2$ it was $\bar{t}_2 = \approx 388s$. The median time for $Group 1$ was $\tilde{t}_1 = 315s$ and for $Group 2$ it was $\tilde{t}_2 = 384s$. As figure 4.10 and the values mentioned above suggest the Mann-Whitney-U-Test shows there is no significant difference between the two groups with a $U = 43.0$ and $p \approx 0.86$.

Furthermore, let's look at the single tasks. Has there maybe been a learning effect or difference in general between the groups? To find out

² <https://android-developers.googleblog.com/2016/08/taking-final-wrapper-off-of-nougat.html> (10.06.2022, 12:37)

³ https://www.gsmarena.com/huawei_p10_lite-8598.php (10.06.2022, 12:37)

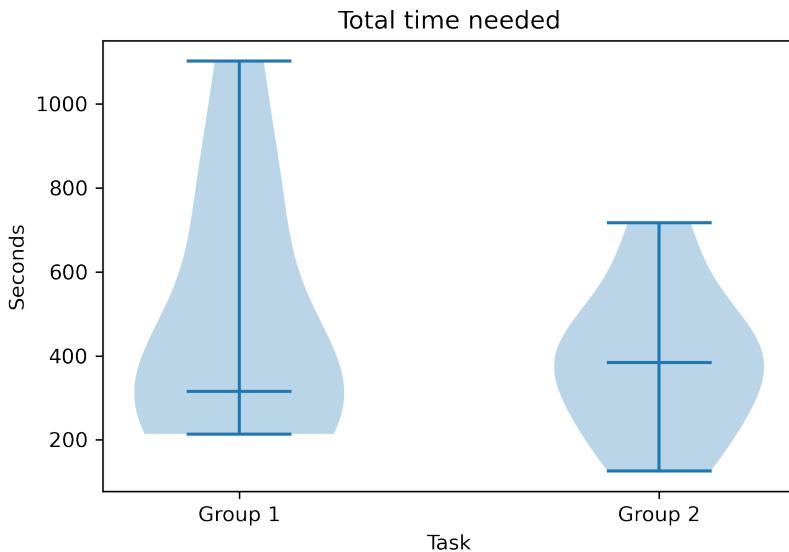


Figure 4.10: Total time each group needed as violin plot

task 1 from *Group 1* will be compared with task 3 from *Group 2* and so on. Every couple of tasks is comparable because they have been done on the same device and use the same set of user interactions as described in section 4.4.4. The Mann-Whitney-U-Test shows that there is no significant difference in any combination as listed below:

- *Group 1*, task 1 & *Group 2*, task 3: $U = 26.5; p \approx 0.23$
- *Group 1*, task 2 & *Group 2*, task 4: $U = 35.5; p \approx 0.69$
- *Group 1*, task 3 & *Group 2*, task 1: $U = 49.0; p \approx 0.48$
- *Group 1*, task 4 & *Group 2*, task 2: $U = 57.0; p \approx 0.16$

Finally, let's see if there is a significant difference between the two used versions of SEE in the single tasks. The time needed for every single task on both versions can be seen in figure 4.11. There are some results that stand out and took way longer than the medians would suggest. That could be caused by various reasons as for example:

- The mobile version of SEE did not run smooth on the subjects device and therefore the tasks became harder.
- The subject did not understand the task correctly and has to reread it.
- The subject overlooked a key object and continued the search at wrong areas.

To prove the hypothesis, H_a for every task will be checked if the subjects needed significantly less time with the desktop version of SEE than with the mobile version. Therefore, a one-sided Mann-Whitney-U-Test will be used. The critical value here is $U = 21$.

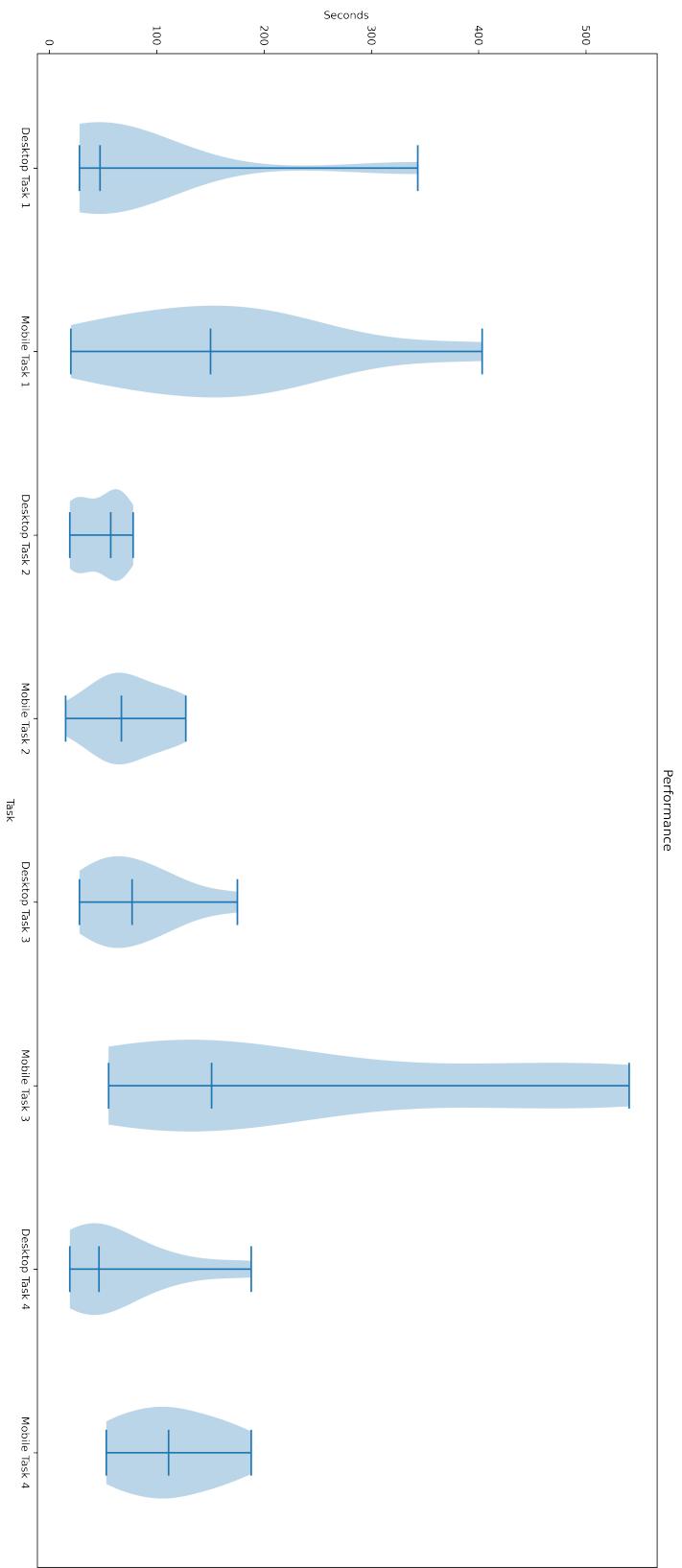


Figure 4.11: Violin plots of all tasks by device

TASK I The subjects needed an average time of $\approx 80\text{s}$ with the desktop version and $\approx 166\text{s}$ with the mobile version. The median here was 47s with the desktop version and 150s with the mobile version. The medians show that there are some outliers on both sides that took longer than the usual subject needed. The one-sided Mann-Whitney-U-Test shows, with $U = 19; p \approx 0.03$, that the subjects needed significantly less time than with the mobile version.

TASK II This task shows the lowest notable difference in figure 4.11. The Mann-Whitney-U-Test shall confirm this assumption with $U = 24; p \approx 0.08$. This results that the subjects did not need significantly less time to solve task 2 with the desktop version than with the mobile version. The average time of the desktop version here is $\approx 47\text{s}$ and of the mobile version ≈ 74 , while the desktop median lies at 57s and the mobile median at 67s .

TASK III In contrast to task 2 this task shows the biggest difference in figure 4.11 and once again the one-sided Mann-Whitney-U-Test approves this clearly with a result of $U = 12.0; p \approx 0.007$. This shows that the desktop version needed significantly less time than the mobile version once more. Also, the average time needed spread quite far with an average of $\approx 82\text{s}$ for the desktop version and $\approx 247\text{s}$ for the mobile version. The median values are 77s for the desktop version and 151s for the mobile version. The desktop median is almost the same as the desktop average while the mobile median is again quite far from the average, which shows that there are again some outliers that took much longer than the usual subject. This could, among other things, be caused by the reported bad performance of the devices “Huawei P20 Pro” and “Huawei P10 Lite”, which were used for task 3.

TASK IV The average time needed to complete the task was $\approx 67\text{s}$ for the desktop version and $\approx 113\text{s}$ for the mobile version. The medians here were 46s for the desktop version and 111s for the mobile version. The values differ quite much and the Mann-Whitney-U-Test validates one last time that the desktop version needed significantly less time than the mobile version to complete the tasks ($U = 17.5; p \approx 0.023$).

To close this section figure 4.12 shows the total time needed for all tasks by used device. The one-sided Mann-Whitney-U-Test for the null hypothesis has a clear result of $U = 11.0$ and $p \approx 0.005$. Therefore, the null hypothesis H_0 : “The time required in SEE desktop is higher or the same as the time required in see mobile” can be rejected, and the alternative hypothesis can be assumed true.

Summarized it can be said that the subjects were faster on the desktop version of SEE. All data sets regarding the required time to solve the four tasks are comparable even crossed over like for example task 1 and

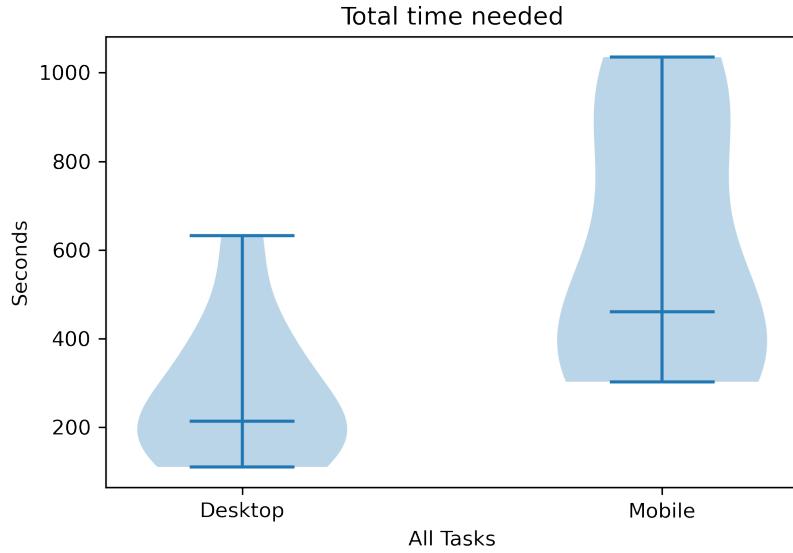


Figure 4.12: Total time the subjects needed on the different devices as violin plot

task 3. Only for one task the subjects did not need significantly less time on the desktop version, but the border value to a significant result was quite close. All in all the desktop version did perform better than the mobile version regarding the required time.

4.5.3 Usability

This section will discuss the results of the usability questionnaires presented in section 4.4.2. First the ASQ results will be analyzed and second the SUS results to test the hypotheses presented in section 4.2.

ASQ

As a start the Post-Task questionnaire will be looked at. Figure 4.13 shows the ASQ results by group independent of the used device. Unfortunately the result differ significantly as the Mann-Whitney-U-Test shows with $U = 392.5$ and $p \approx 0.0037$. Therefore, a deeper look is needed. The results of *Group 1* do not differ significantly from *Group 2* looking only at the tasks completed with the desktop device as the Mann-Whitney-U-Test shows with $U = 122.5$ and $p \approx 0.21$. On the other hand side the results only caused by completing tasks with a mobile device differ significantly as the Mann-Whitney-U-Test shows ($U = 73.0; p \approx 0.0048$). Having another look at figure 4.13 the result on *Group 1* only on the mobile stands out. Comparing the two groups without the described results of *Group 1* results in a Mann-Whitney-U-Test of $U = 266.5; p \approx 0.29$, which means without the subset of *Group 1* the two groups show no significant difference.

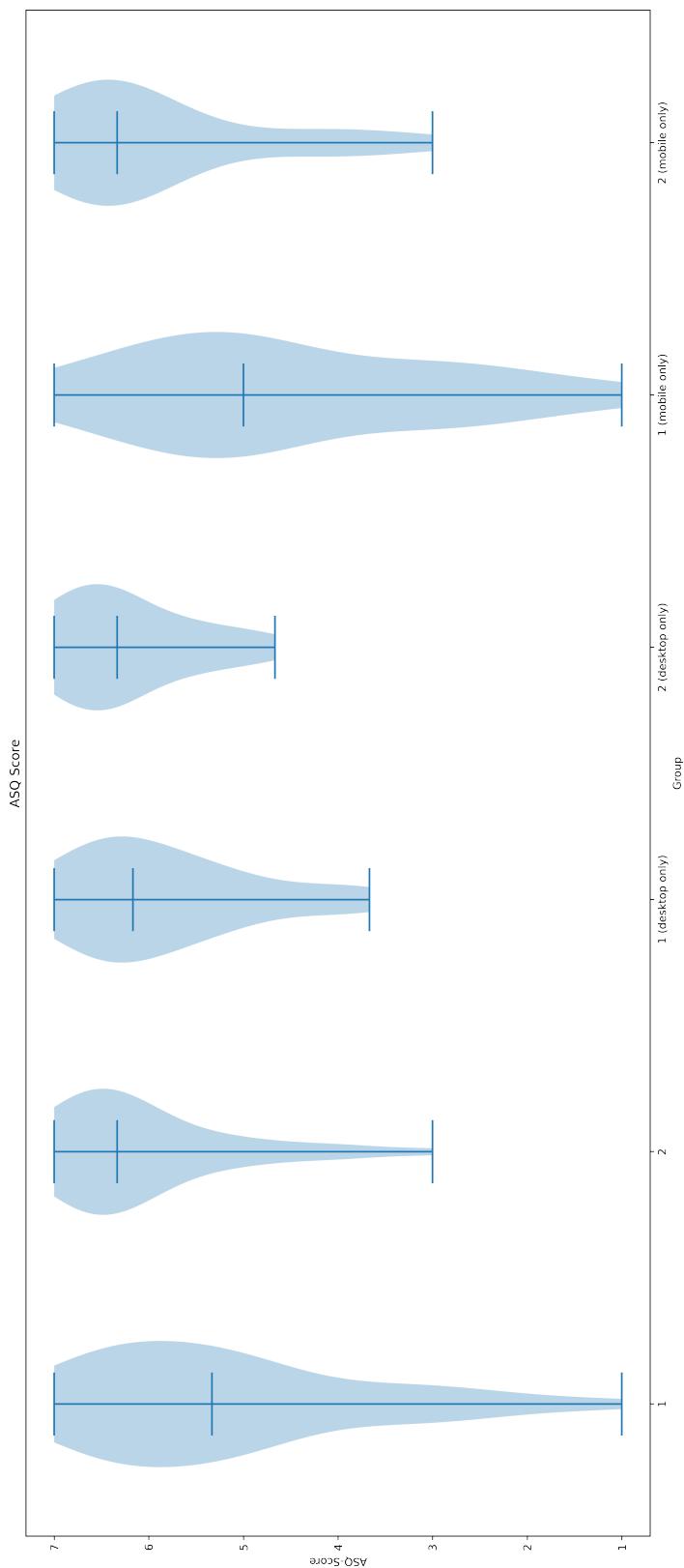


Figure 4.13: Violin plots of all tasks by group

To ensure a good comparability further on the all task will be looked at separately. The ASQ consists of three questions. Each representing a different aspect. In the following all tasks will be compared as shown in table 4.3.

ASQ Question	Complexity			
Group 1, Task:	1	2	3	4
Group 2, Task:	3	4	1	2
Mann-Whitney-U	$U = 47; p \approx 0.58$	$U = 37.5; p \approx 0.81$	$U = 7.0; p \approx 0.003$	$U = 24.5; p \approx 0.16$
ASQ Question	Effort			
Group 1, Task:	1	2	3	4
Group 2, Task:	3	4	1	2
Mann-Whitney-U	$U = 44; p \approx 0.78$	$U = 37; p \approx 0.78$	$U = 13.5; p \approx 0.02$	$U = 23; p \approx 0.11$
ASQ Question	Information			
Group 1, Task:	1	2	3	4
Group 2, Task:	3	4	1	2
Mann-Whitney-U	$U = 39.5; p \approx 0.96$	$U = 36; p \approx 0.70$	$U = 20.5; p \approx 0.07$	$U = 20.5; p \approx 0.07$

Table 4.3: Two-sided Mann-Whitney-U-Test for each comparable question on the same device

The two marked Mann-Whitney-U-Tests in table 4.3 as they show that the pair *Group 1, Task 3 - Group 2, Task 1 - ASQ Question 1* as well as *Group 1, Task 3 - Group 2, Task 1 - ASQ Question 2* have significant differences. As those two pairs are not comparable, they will not be observed further on. All remaining results can be seen in figure 4.14 for the ASQ aspect effort, in figure 4.15 for ASQ aspect complexity as well as in figure 4.16 which covers the ASQ aspect information.

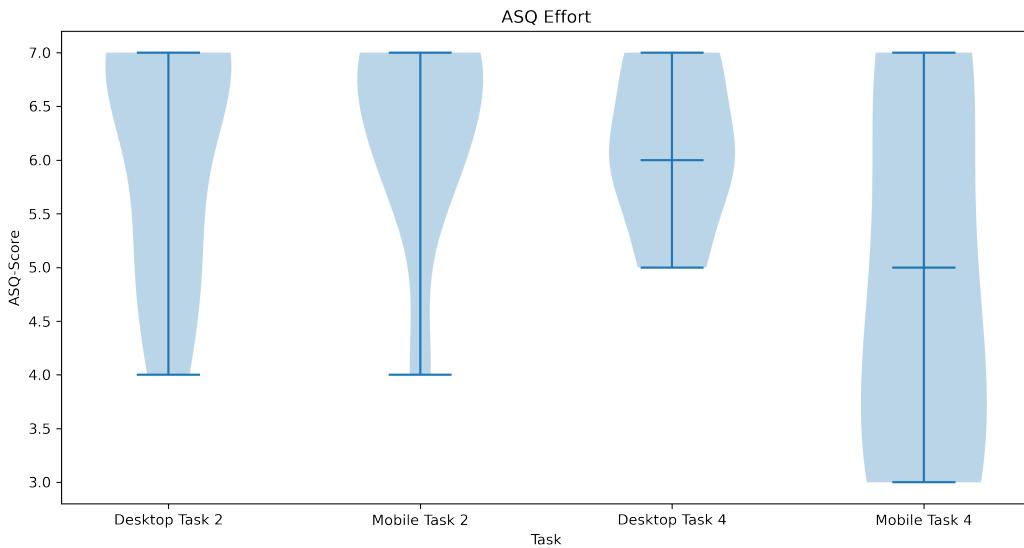


Figure 4.14: ASQ effort results violin plots of all tasks by device

The Mann-Whitney-U-Test only shows significant findings for two tasks:

TASK III The desktop version of SEE has significant better results than the mobile version in the aspect of information. The subjects

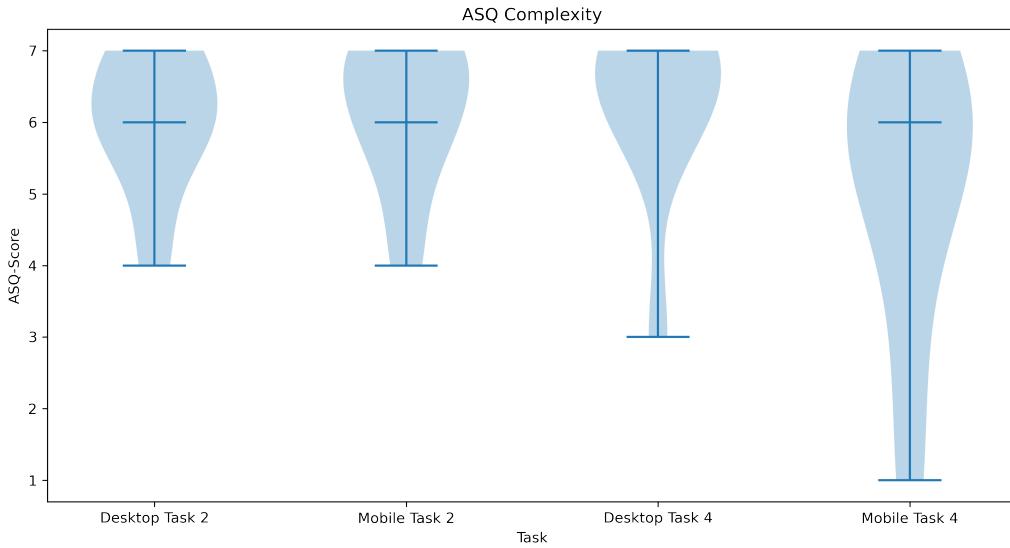


Figure 4.15: ASQ complexity results violin plots of all tasks by device

rated the desktop version with an average score of 6.22 and a median score of 7, while the mobile version was rated an average of 6.44 and a median of 5.22. The Mann-Whitney-U-Test results here are $U = 20.5$ and $p \approx 0.037$.

TASK IV The desktop version also showed better results for the aspect of information in this task. The subjects rated the desktop version with an average score of 6.44 and a median of 7, while the mobile version was rated with an average of 5.22 and a median of 6. The Mann-Whitney-U-Test resulted in $U = 20.5$ and $p \approx 0.035$ in a one-sided test, which means the desktop version got a significantly higher rating than the mobile version.

All other tasks show significant higher or lower results. It is also to mention that only the desktop version only scored a significant higher value at the aspect of information. That was to be expected because the mobile versions offer less information because of its constraints regarding the screen size. The mobile version uses for example only icon buttons and not text buttons like the desktop version, which save screen space but sacrifices information value since the user has to remember the meaning of each icon.

Finally, the hypotheses H_b , H_c , H_d can be rated:

H_b The null hypothesis H_{b0} can not be rejected since the Mann-Whitney-U-Test results in $U = 362.0$; $p \approx 0.23$. Therefore, the alternative hypothesis can be assumed true, which means that the desktop score for the ASQ aspect *complexity* is not significantly higher.

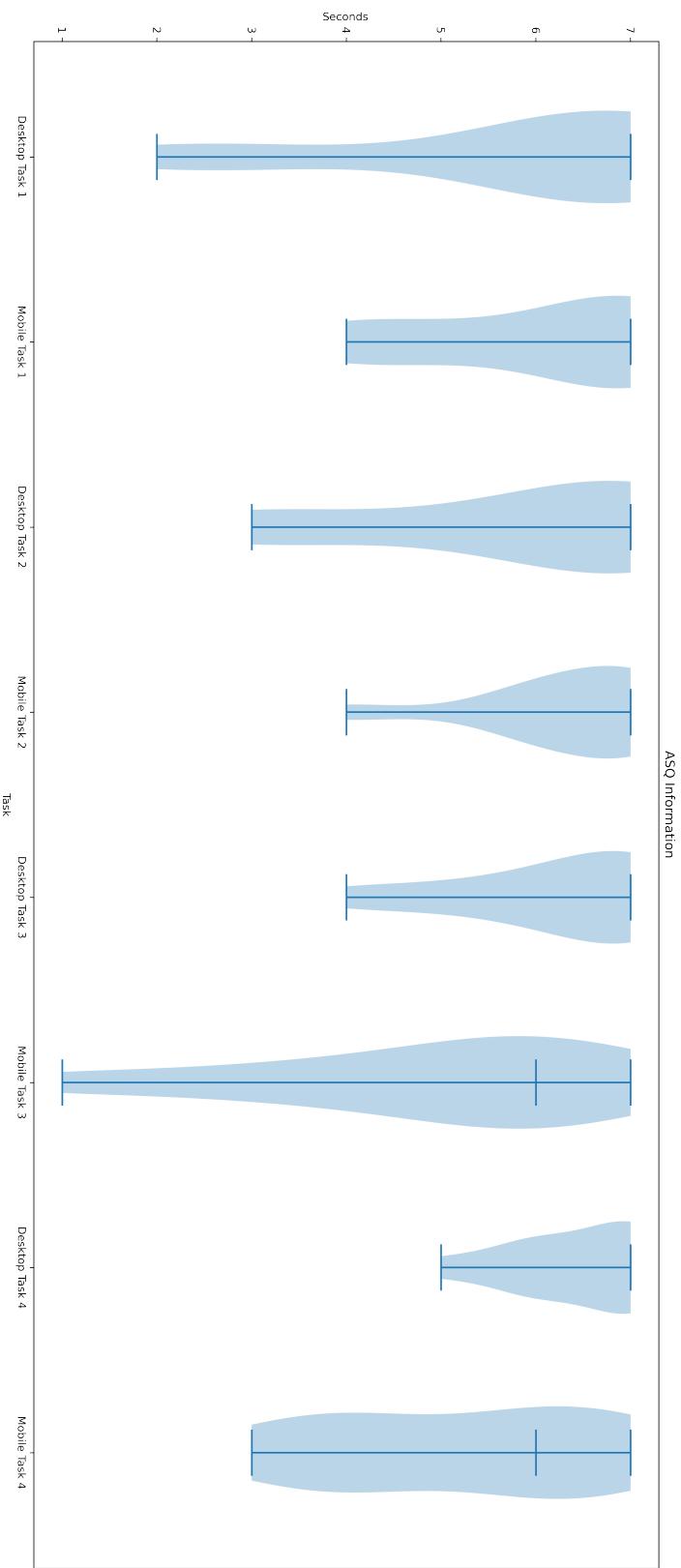


Figure 4.16: ASQ information results violin plots of all tasks by device

H_c The same goes for null hypothesis H_{c0} . It can not be rejected since the Mann-Whitney-U-Test results in $U = 348.5; p \approx 0.32$. That means that the desktop ASQ score for the aspect *effort* is also not significantly higher than the mobile ASQ score.

H_d Two of the four tasks showed a significantly higher score for the desktop version in the ASQ aspect *information* but for all four tasks combined the null hypothesis H_{d0} can not be rejected since the Mann-Whitney-U-Test resulted $U = 776.5; p \approx 0.06$. Although the results are much closer to the significance level of $\alpha = 0.05$ in this example.

The final ASQ scores, which are an average of all three questions, can be found in table 4.4. However, the final ASQ-Scores still differ significantly as the Mann-Whitney-U-Test of $U = 92.0; p \approx 0.03$ shows. Since, even after not taking into account the previously discussed pairs of tasks, the ASQ scores of the two groups differ significantly, which means the final scores cannot be compared.

D	5.8	4.7	6.3	5.2	6.8	5.8	6.7	6.3	4.8	6.7	6.7	5.8	5.3	6.8	5.0	6.5	6.8	6.3
M	6.8	6.8	6.0	6.8	6.3	4.0	6.0	7.0	6.3	6.3	5.0	6.0	5.3	5.3	2.3	7.0	5.3	3.0

Table 4.4: The ASQ-scores from all 18 subjects. The first row contains the ASQ-scores for the desktop application and the second row for the mobile application. The figures have been rounded to whole numbers.

Summarizing it can be said that the ASQ results show that the two versions of SEE show little significant difference. In total only task 3 and task 4 showed a significant difference in favor of the desktop version for the ASQ aspect *information*.

SUS

Last but not least the results of the SUS questionnaire will be discussed. The SUS questionnaire results in a SUS score which was calculated by the formula of Lewis (2018):

$$SUS = 2.5 \cdot (20 + \sum_{i=1}^1 (-1)^{i+1} \cdot S_i) \quad (4.1)$$

In this formula S_i is the score of one of the ten questions in the SUS. The question scores are alternating positive and negative for the SUS score. Therefore, the factor $(-1)^{i+1}$ is needed to either add or subtract from the final SUS score.

The SUS scores from both groups can be seen in table 4.5. Both groups can be compared as the Mann-Whitney-U-Test shows with a result of $U = 106.0; p \approx 0.08$. The group results can also be seen visualized as violin plots in figure 4.17. The average score of *Group 1* was ≈ 60.83 and the score of *Group 2* was ≈ 72.77 . The median of *Group 1* was 58.75 and of *Group 2* it was 82.5. The median of *Group 2* differs quite much from the average which indicates an uneven distribution

and outliers that are lower than normal. This can also be seen in figure 4.17.

D	60	58	80	60	78	70	85	53	80	88	80	50	63	85	50	93	100	85
M	38	53	83	53	53	30	88	35	43	86	48	48	55	90	50	88	98	58

Table 4.5: The SUS-scores from all 18 subjects. The first row contains the SUS-scores for the desktop application and the second row for the mobile application. The figures have been rounded to whole numbers.

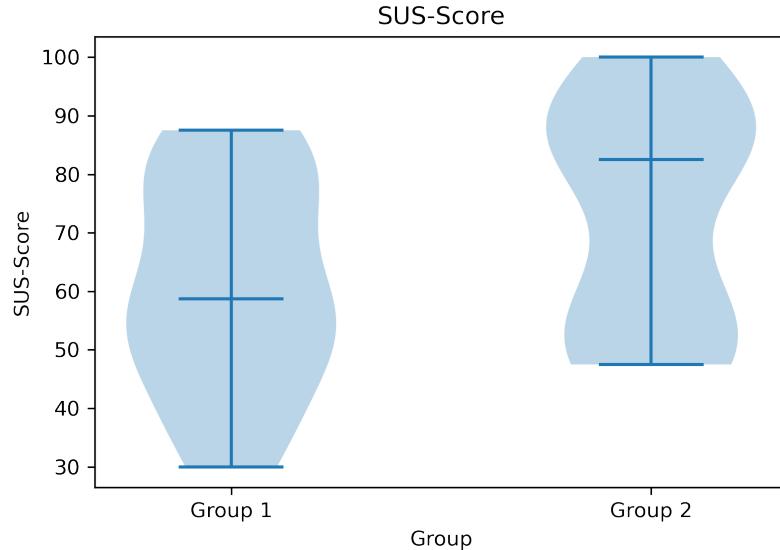


Figure 4.17: The SUS-Scores of the two SEE versions by group

Finally, comparing the SUS-Scores (figure 4.18) of the two different versions shows that the null hypothesis H_{e0} can be rejected since the Mann-Whitney-U-Test resulted in $U = 219.5; p \approx 0.035$. The desktop version had an average SUS-Score of ≈ 73.06 and a median of 78.75, while the mobile version had an average SUS-Score of ≈ 60.56 and a median of 52.5. After seeing the results it can be assumed that the Usability of the desktop version is significantly higher than the Usability of the mobile version.

4.5.4 Impact of experience

As already described in section 4.2 the subjects will be grouped by their level of experience and their results analyzed. Since only one subjects had no experience with first person video games but is still experienced with SEE the subjects can be divided into three groups:

- SEE-developer
- Non-SEE-developer with software development experience
- Non-SEE-developer without software development experience

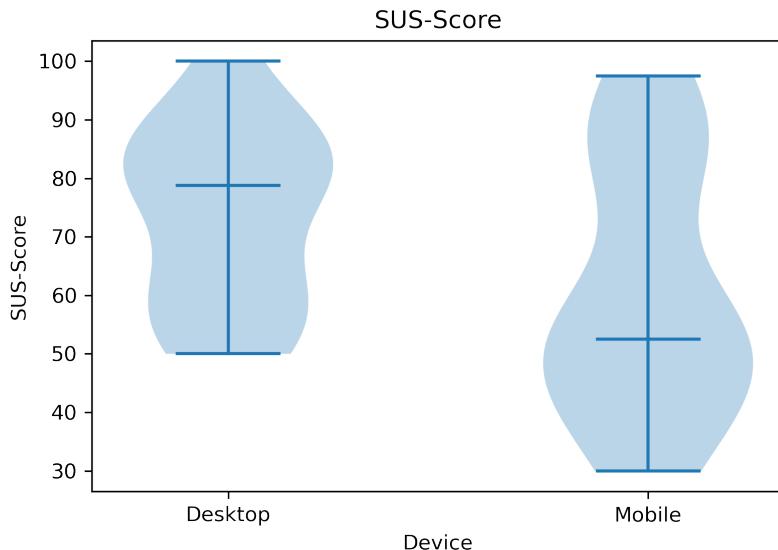


Figure 4.18: The SUS-Scores of the two SEE versions by device

In addition to that there will also be a look at the impact the highest achieved degree of the subjects has. In order to measure the impact the *Kendall rank correlation coefficient* will be used. Another option to calculate the correlation would have been Pearson's correlation coefficient, which has been more popular, but its distribution has slightly worse statistical properties (Hauke and Kossowski (2011)). In addition to that Khamis (2008) also suggest the use of the Kendall rank correlation coefficient, if there is a small ordinal scale of less than five or six levels, which is the case for our results.

Kendall rank correlation coefficient: The rank based correlation coefficient is a statistic measured with two ordinal or interval-scaled variables. The result is a range between [-1; 1], where a positive value indicate a positive correlation and a negative value a negative correlation. The value 0 indicates no correlation at all.

EXPERIENCE WITH SEE

Table 4.6 contains the correlation between the experience with SEE of the subjects and the dependent variable of the certain version of SEE. The significance level will be again $\alpha = 0.05$. The only significant correlation was found for the SUS-Score with $\tau \approx 0.47$ and $p \approx 0.02$. That means that there is a weak positive correlation between the experience with SEE and the SUS-Score. This could be because subjects that worked with SEE might be happy to see their work on a different device.

EXPERIENCE WITH SOFTWARE DEVELOPMENT

The correlations between the experience with software development and the corresponding dependent variable from one of the two versions are listed in table 4.7. For this comparison only the subjects without experience with SEE were looked at. Not a single significant correlation has been found in that context, which means that there is no impact the dependent variables by having experience with software development.

	SEE-Desktop	SEE-Mobile
Performance	$\tau \approx -0.32; p \approx 0.12$	$\tau \approx -0.37; p \approx 0.07$
ASQ - Complexity	$\tau \approx 0.29; p \approx 0.18$	$\tau \approx 0.30; p \approx 0.15$
ASQ - Effort	$\tau \approx 0.19; p \approx 0.37$	$\tau \approx 0.16; p \approx 0.46$
ASQ - Information	$\tau \approx 0.04; p \approx 0.87$	$\tau \approx -0.03; p \approx 0.88$
SUS	$\tau \approx 0.29; p \approx 0.17$	$\tau \approx 0.47; p \approx 0.02$

Table 4.6: Correlation between experience with SEE and the dependent variables calculated with the Kendall rank correlation coefficient

	SEE-Desktop	SEE-Mobile
Performance	$\tau \approx -0.14; p \approx 0.56$	$\tau \approx -0.07; p \approx 0.77$
ASQ - Complexity	$\tau \approx -0.08; p \approx 0.76$	$\tau \approx 0.06; p \approx 0.82$
ASQ - Effort	$\tau \approx -0.21; p \approx 0.41$	$\tau \approx -0.05; p \approx 0.83$
ASQ - Information	$\tau \approx -0.39; p \approx 0.15$	$\tau \approx 0.21; p \approx 0.41$
SUS	$\tau \approx -0.07; p \approx 0.77$	$\tau \approx 0.04; p \approx 0.88$

Table 4.7: Correlation between experience with software development and the dependent variables calculated with the Kendall rank correlation coefficient without the subjects that had experience with SEE

HIGHEST ACHIEVED DEGREE

The correlations between the highest achieved degree and the corresponding dependent variable from one of the two versions can be found in table 4.8. Also in this context there were no correlation even close to significance. This concludes that there is also no correlation between the highest achieved degree and the dependent variable.

	SEE-Desktop	SEE-Mobile
Performance	$\tau \approx 0.06; p \approx 0.77$	$\tau \approx -0.36; p \approx 0.06$
ASQ - Complexity	$\tau \approx 0.06; p \approx 0.76$	$\tau \approx 0.10; p \approx 0.62$
ASQ - Effort	$\tau \approx -0.11; p \approx 0.59$	$\tau \approx 0.19; p \approx 0.32$
ASQ - Information	$\tau \approx 0.16; p \approx 0.44$	$\tau \approx 0.28; p \approx 0.17$
SUS	$\tau \approx 0.02; p \approx 0.90$	$\tau \approx 0.14; p \approx 0.46$

Table 4.8: Correlation between highest achieved degree and the dependent variables calculated with the Kendall rank correlation coefficient

ANDROID VERSION AND DEVICE

The correlations between the Android version or device and the corresponding dependent variable from one of the two versions can be found in table 4.9. The devices were divided into the ones reported with a bad performance and the devices with and a mention of a bad performance. Unfortunately it is not safe to say if the division is correct because it is

possible that subjects did not mention the bad performance on their device. There is still a significant correlation between the ASQ effort aspect and the Android device. A higher performing device means a weakly higher ASQ rating in the effort aspect, which is a positive effect since a high ASQ score represents a low effort. It is also to mention that the correlation between the Android device and the ASQ complexity aspect is right on the edge the significance level of $\alpha = 0.05$. It is quite possible that on another day it would fall below, which would signify that there is also a positive correlation between a better performing device and a higher ASQ complexity score.

	Android Version	Android device
Performance	$\tau \approx -0.23; p \approx 0.22$	$\tau \approx -0.35; p \approx 0.09$
ASQ - Complexity	$\tau \approx 0.13; p > 0.05$	$\tau \approx 0.40; p > 0.05$
ASQ - Effort	$\tau \approx -0.02; p \approx 0.91$	$\tau \approx 0.44; p \approx 0.04$
ASQ - Information	$\tau \approx 0.13; p \approx 0.50$	$\tau \approx 0.33; p \approx 0.11$
SUS	$\tau \approx 0.07; p \approx 0.72$	$\tau \approx 0.31; p \approx 0.14$

Table 4.9: Correlation between the Android version or device and the dependent variables calculated with the Kendall rank correlation coefficient

4.5.5 *Comments of the subjects*

Last but not least the subjects were asked if they had suggestions for improvements or additional features. The original answers can be found in [See Desktop - See Mobile.csv](#) and [See Mobile - See Desktop.csv](#). All the result have been grouped, commented and listed below:

DESKTOP VERSION

- Information on key assignments x3
 - Since this was mentioned multiple times it might be considered to add some kind of information on the Shortcuts. An option would be to implement a tutorial mode, which offers this kind of information but can later be turned off after the user memorized the Shortcuts.
- Better contrast or font for nodes and planes x5
 - This was mentioned the most for the desktop version. This problem is already known and solutions are in progress.
- Easier switch between interaction modes
 - This is already implemented with the key 1-9 but was not mentioned in the tutorial video since it is not beginner-friendly and the subjects should not be confronted with too much information to memorize.

- Hard to use without a mouse
 - This is not optimized yet. Maybe it is worth considering an option to work only with a touchpad or just to clarify that a mouse is needed for an optimal use of SEE desktop.
- Visible UI elements, key combinations are hard to remember
 - Eventually pressing Shortcuts like “alt” in the rotate mode could be indicated as well.

MOBILE VERSION

- Different colors for menu buttons and sub menu buttons
 - The UI design seemed to be confusing for some subjects. Maybe a division of the main button that also reopens the menu again and the additional buttons should be considered. Different colors could be an option or different icons for the main buttons that also indicate the menu can be opened.
- Less battery usage
 - This is probably due to the high need of computing power of SEE. SEE is in ongoing development and on optimizing the efficiency is always a topic.
- Better icons x2
 - One subject mentioned that the minus icon for adding an edge is confusing since a minus rather indicates to delete something. Most of the icons are the same in both versions, but the desktop version also contains a button title. Designing a set of homogeneous and innovative icons might be worth considering.
- Deselect button should always be available
 - This can easily be fixed by moving the button into the “quickmenu”, which is available from every interaction mode.
- The camera notches of some devices overlap some buttons x2
 - A general approach for this would be to not allow screen space with a notch on it. This would sacrifice screen space but provide consistency over different devices. Having a design for every possible device seems to be too much effort.
- Joysticks to far in the corner
 - Since this did not apply for the testing device, there might need to be found a more dynamic approach which considers the actual screen size and adapts the UI size to it.

- Joystick handling to sensitive
 - This seemed to be an issue for more than one subjects even if only one mentioned it. The sensitivity should probably be reduced. That way the user can work more precisely.
- Zooming inverted x9
 - Since this was explicitly mentioned by half the subjects, this should definitely be changed.
- Better performance on older devices
 - This is probably hard implement. Maybe optimizing features of SEE that require heavy computing will help in the future.
- App layout should be turned by 180° x2
 - This can be adjusted in the Unity editor and is an easy fix.
- Add a search function
 - This could be considered for the future since the desktop version already has a search option.
- Sometimes planes were deleted after I changed the mode from delete
 - The menu and joysticks need to be safe to touch. Right now objects below buttons or joysticks get selected.

4.6 THREADS TO VALIDITY

This chapter will be closed with a discussion on the threads of validity to the executed study. For this purpose, there will be a distinction into *internal validity* and *external validity* as by [Campbell and Stanley \(1963\)](#)

4.6.1 Internal Validity

The internal validity is about the correlation between the independent variable, which is the version of SEE in this case, and the dependent variable, which is *performance*, *ASQ-Score* and *SUS-Score*. Some of [Campbell and Stanley \(1963\)](#) factors will be discussed in the following.

HISTORY

The subjects did the study asynchronously and there is no significant difference between the time it took place for the two groups as seen in figure [4.19](#). The study took place within five days. It is very unlikely that one of the subjects could have gotten an advantage by for example trying out SEE because every subject got the applications just shortly before their appointment.

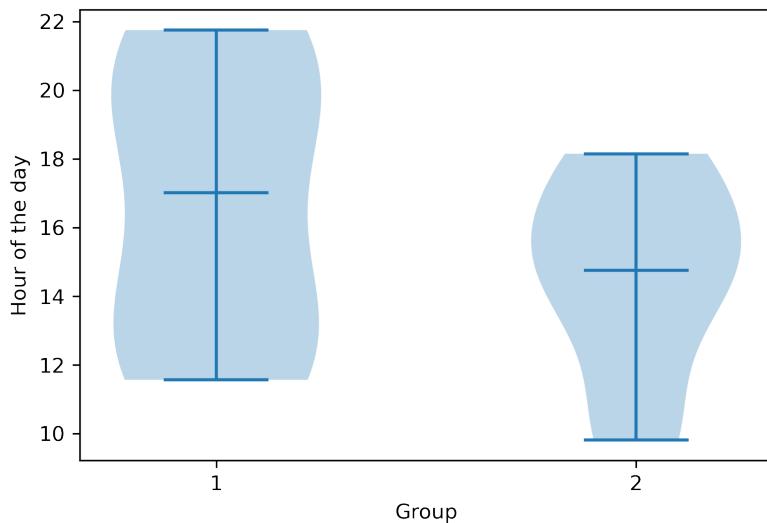


Figure 4.19: The time the subjects started the experiment

TESTING

There could be a learning effect since the tasks were comparable to each other. To prevent a learning effect the subjects got a training before the time was measured in the actual tasks. The training contained all necessary interactions to solve the tasks and since the interactions are executed in different ways in the two versions the learning effect should not be too high if even exciting.

INSTRUMENTATION

Unfortunately the performance on the subjects Android devices fluctuated quite much and there was also a significant correlation between the performance of the Android device and the ASQ aspect of *effort* as described in section 4.5.4. Furthermore, it can also not be ruled out that the performance of the desktop device had an impact on the results even if no subjects mentioned any troubles.

BIASES

Many of the subjects know me personally and most of them also knew that I did the mobile implementation of SEE. It is therefore possible that they gave a higher rating. There was also found a correlation between subjects experienced with SEE and the SUS score in section 4.5.4, which could be caused because SEE developers are most likely emotional involved in the project. Since there is no significant difference of the amount of SEE experienced subjects this should not have impact on the study results but could still be a threat to validity.

SELECTION

The selection of the subjects was random, however it is still possible

that the groups would have been formed uneven. In section 4.5.1 no significant difference between the two groups was found. Still in section 4.5.3 significant differences between the ASQ ratings the two groups gave were found. Therefore, this thread to validity can not be excluded.

OTHER THREADS

The Code-Cities of the two task blocks were not the same. This is to reduce the learning effect, but it is possible that they differed in complexity and had impact on the difficulty of the tasks. However, the Code-Cities are at least similar as discussed in 4.4.3

4.6.2 *External Validity*

External validity looks at the generalizability of a study. Are the findings applicable in general or just for this specific case? However, the question of generalizability can never be fully answered and there can only be approximations. Threads to the external validity could be:

NONREPRESENTATIVE SAMPLE

Unfortunately the subjects do not fully represent the population of the software development industry as all the subjects are male and 16 of 18 subjects are in the age of 12-30 years. In addition to that most of the subjects are not even software developers.

REACTIVE EFFECTS OF EXPERIMENTAL ARRANGEMENT

Since the study was supervised the subjects might feel observed, which could have an impact on their performance. The environmental conditions of the real use of SEE cannot be reproduced in an experimental arrangement. The external validity is therefore in doubt.

5

CONCLUSION

...

5.1 OUTLOOK

AR - [Santos et al. \(2016\)](#)

A

GLOSSARY

Code-City In the Code-City metaphor, software components are represented by buildings in a city, and the properties of these buildings can express different metrics of the software. For example the height of a “Node” could represent the lines of code in that class. [3–9](#), [14](#), [16–18](#), [22](#), [25](#)

Kendall rank correlation coefficient The rank based correlation coefficient is a statistic measured with two ordinal or interval-scaled variables. The result is a range between $[-1; 1]$, where a positive value indicate a positive correlation and a negative value a negative correlation. The value 0 indicates no correlation at all. [38–40](#)

Mann-Whitney-U-Test A test that compares two independent samples with ordinal scales to find out whether there is a difference between the two groups. The Mann-Whitney-U-Test does not require a normal distribution. [25–29](#), [31](#), [32](#), [34](#), [35](#), [37](#)

Node A point in a diagram where lines intersect. In “SEE” it usually displays a software class [3](#), [13](#), [16](#)

Plane An area that bundles “nodes”. Could for example represent a namespace. [13](#)

Post-Study A questionnaire that is taken after every block of an experiment. [15](#), [21](#)

Post-Task A questionnaire that is taken after every task of an experiment. [15](#), [18](#), [21](#), [31](#), [47](#)

Shortcut A combination of key that will call an action like for example “ctrl” + “c” for coping a text. [13](#), [40](#), [41](#)

Usability A term that describes how well a (software) system can be used. [15](#), [18](#), [21](#), [22](#), [37](#), [47](#)

B

ACRONYMS

ASQ A post-task questionnaire consisting of three questions that is used to access how difficult a user perceived a task (See “Post-Task”).
[15](#), [18](#), [21](#), [31](#), [32](#), [34](#), [35](#)

CSV A file format used for example to store table data. Each line represents a data record and each value is separated with a comma. [19](#)

SEE An interactive software visualization that uses the *code-city* metaphor and enables collaborative multiplayer interactions via multiple platforms like desktop, virtual reality and soon Android devices. [1](#), [3–8](#), [13](#), [15–19](#), [21](#), [25–29](#), [31](#), [32](#), [35](#), [37–39](#), [41](#), [42](#)

SUS The System Usability Scale consists of ten questions that measure “Usability” [15](#), [16](#), [18](#), [21](#), [22](#), [31](#), [35](#), [37–39](#)

C

LIST OF FIGURES

Figure 2.1	Joysticks for moving in SEE	4
Figure 2.2	Quickbar for various interactions in SEE	5
Figure 2.3	Selection mode in SEE	5
Figure 2.4	Delete mode in SEE	6
Figure 2.5	Node interactions in SEE	6
Figure 2.6	Rotation mode in SEE	7
Figure 2.7	Movement mode in SEE	8
Figure 3.1	The mobile player Prefab	11
Figure 3.2	The <i>quickmenu</i> on the left top side of the mobile device. Pressing the button with an orange marked arrow will expand the menu.	12
Figure 3.3	The player interaction menu on the right top side of the mobile device. The button on the top right side indicate the active interaction mode. Pressing the same button also expands the menu.	12
Figure 4.1	The desktop menu for selecting interaction modes.	14
Figure 4.2	The first Code-City for the user study	17
Figure 4.3	The second Code-City for the user study	17
Figure 4.4	The third Code-City for the user study	18
Figure 4.5	The intro of the survey	20
Figure 4.6	The introduction video of the survey	20
Figure 4.7	The two key nodes are marked with a yellow arrow	22
Figure 4.8	The distribution of the highest completed degrees of all 18 subjects	27
Figure 4.9	The distribution of the Android versions the subjects were using	28
Figure 4.10	Total time each group needed as violin plot . .	29
Figure 4.11	Violin plots of all tasks by device	30
Figure 4.12	Total time the subjects needed on the different devices as violin plot	32
Figure 4.13	Violin plots of all tasks by group	33
Figure 4.14	ASQ effort results violin plots of all tasks by device	34
Figure 4.15	ASQ complexity results violin plots of all tasks by device	35
Figure 4.16	ASQ information results violin plots of all tasks by device	36
Figure 4.17	The SUS-Scores of the two SEE versions by group	38
Figure 4.18	The SUS-Scores of the two SEE versions by device	39

Figure 4.19 The time the subjects started the experiment . . . 44

D

LIST OF TABLES

Table 4.1	The tasks used for the experiment. The device will be switched after task 2	24
Table 4.2	Experimental procedure per subject. The procedure is swapped per group.	25
Table 4.3	Two-sided Mann-Whitney-U-Test for each comparable question on the same device	34
Table 4.4	The ASQ-scores from all 18 subjects. The first row contains the ASQ-scores for the desktop application and the second row for the mobile application. The figures have been rounded to whole numbers.	37
Table 4.5	The SUS-scores from all 18 subjects. The first row contains the SUS-scores for the desktop application and the second row for the mobile application. The figures have been rounded to whole numbers.	38
Table 4.6	Correlation between experience with SEE and the dependent variables calculated with the Kendall rank correlation coefficient	40
Table 4.7	Correlation between experience with software development and the dependent variables calculated with the Kendall rank correlation coefficient without the subjects that had experience with SEE	40
Table 4.8	Correlation between highest achieved degree and the dependent variables calculated with the Kendall rank correlation coefficient	40
Table 4.9	Correlation between the Android version or device and the dependent variables calculated with the Kendall rank correlation coefficient	41

Regie: Kontrolliere am Ende, ob alle bibliographischen Angaben vollständig sind. Wird also die Zeitschrift oder Konferenz aufgeführt, in der ein Artikel veröffentlicht wurde? Sind überall die Seitenangabe aufgeführt? Bei Verweisen auf Web-Seiten, ist überall angegeben, wann der letzte Zugriff darauf erfolgte? Sind Umlaute und andere Sonderzeichen korrekt in LaTeX beschrieben worden?



RELATED FILES

Name	Description
SEE_Desktop.zip	
SEE_Mobile.apk	
calc_data.ipynb	A Python script that calculates all the results of the study.
See Desktop - See Mobile.csv	The survey answers from <i>Group 1</i> the results of the study.
See Mobile - See Desktop.csv	The survey answers from <i>Group 2</i> the results of the study.

BIBLIOGRAPHY

- Stephanie Houde and Charles Hill. What do prototypes prototype? In *Handbook of human-computer interaction*, pages 367–381. Elsevier, 1997.
- Karen Renaud and Judy Van Biljon. Demarcating mobile phone interface design guidelines to expedite selection. *South African Computer Journal*, 29(3):127–144, 2017.
- Lumpapun Punchoojit and Nuttanont Hongwarittorrn. Usability studies on mobile user interface design patterns: a systematic literature review. *Advances in Human-Computer Interaction*, 2017, 2017.
- Jessica Conradi, Olivia Busch, and Thomas Alexander. Optimal touch button size for the use of mobile devices while walking. *Procedia Manufacturing*, 3:387–394, 2015.
- Pekka Parhi, Amy K Karlson, and Benjamin B Bederson. Target size study for one-handed thumb use on small touchscreen devices. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services*, pages 203–210, 2006.
- Boonlit Adipat and Dongsong Zhang. Interface design for mobile applications. *AMCIS 2005 proceedings*, page 494, 2005.
- Suzanne Robertson and James Robertson. *Mastering the Requirements Process: Getting Requirements Right*. Addison-Wesley Professional, 2012. ISBN 978-0-13-294285-0.
- P. Stevens and R. Pooley. *Software Engineering with Objects and Components*. Springer, 2005.
- Jeong Ho Kim, Lovenoor Aulck, Michael C Bartha, Christy A Harper, and Peter W Johnson. Differences in typing forces, muscle activity, comfort, and typing performance among virtual, notebook, and desktop keyboards. *Applied ergonomics*, 45(6):1406–1413, 2014.
- Sam Mclellan, Andrew Muddimer, and S. Peres. The effect of experience on system usability scale ratings. *Journal of Usability Studies*, 7, 11 2011.
- James R Lewis. Psychometric evaluation of an after-scenario questionnaire for computer usability studies: the asq. *ACM Sigchi Bulletin*, 23 (1):78–81, 1991.
- Sadrieh Hajesmael-Gohari, Firoozeh Khordastan, Farhad Fatehi, Hamidreza Samzadeh, and Kambiz Bahaadinbeigy. The most used

- questionnaires for evaluating satisfaction, usability, acceptance, and quality outcomes of mobile health. *BMC Medical Informatics and Decision Making*, 22(1):1–9, 2022.
- James R Lewis. Ibm computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, 7(1):57–78, 1995.
- John Brooke. Sus: a “quick and dirty’usability. *Usability evaluation in industry*, 189(3), 1996.
- John Brooke. Sus: a retrospective. *Journal of Usability Studies*, 8:29–40, 01 2013.
- James R Lewis. The system usability scale: past, present, and future. *International Journal of Human–Computer Interaction*, 34(7):577–590, 2018.
- Rebecca A Grier, Aaron Bangor, Philip Kortum, and S Camille Peres. The system usability scale: Beyond standard usability testing. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 57, pages 187–191. SAGE Publications Sage CA: Los Angeles, CA, 2013.
- S. Camille Peres, Tri Pham, and Ronald Phillips. Validation of the system usability scale (sus): Sus in the wild. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 57(1):192–196, 2013. doi: 10.1177/1541931213571043. URL <https://doi.org/10.1177/1541931213571043>.
- Aaron Bangor, Philip T. Kortum, and James T. Miller. An empirical evaluation of the system usability scale. *International Journal of Human–Computer Interaction*, 24(6):574–594, 2008. doi: 10.1080/10447310802205776. URL <https://doi.org/10.1080/10447310802205776>.
- Jean D Gibbons and S Chakraborti. Comparisons of the mann-whitney, student’s t, and alternate t tests for means of normal distributions. *The Journal of Experimental Education*, 59(3):258–267, 1991.
- J.H. Zar. *Biostatistical Analysis*. Prentice Hall, 2010. ISBN 9780131008465.
- Jan Hauke and Tomasz Kossowski. Comparison of values of pearson’s and spearman’s correlation coefficient on the same sets of data. *QUAESTIONES GEOGRAPHICAE* 30(2), 2011.
- Harry Khamis. Measures of association: how to choose? *Journal of Diagnostic Medical Sonography*, 24(3):155–162, 2008.
- Donald T Campbell and Julian C Stanley. *Experimental and quasi-experimental designs for research*. Ravenio books, 1963.

Carlos Santos, Brunelli Miranda, Tiago Araujo, Nikolas Carneiro, Anderson Marques, Marcelle Mota, Jefferson Morais, and Bianchi Meiguins. Guidelines for graphical user interface design in mobile augmented reality applications. In *International Conference on Virtual, Augmented and Mixed Reality*, pages 71–80. Springer, 2016.