



Department of Informatics, University of Zurich

# Kinect tracking for Affinity Diagram

Understanding and augmenting a paper arrangement-based process

**Master Thesis**

28<sup>th</sup> February 2014

Andrei Andriesi  
*Suceava, Romania*

Student-ID: 10-770-263  
[andrei.andriesi@uzh.ch](mailto:andrei.andriesi@uzh.ch)

---

People and Computing Lab

**Professor:** Elaine M. Huang, PhD, [huang@ifi.uzh.ch](mailto:huang@ifi.uzh.ch)  
**Supervisor:** Gunnar Harboe, [harboe@ifi.uzh.ch](mailto:harboe@ifi.uzh.ch)



## Abstract

The current master thesis is part of a larger PhD project, “*Understanding and augmenting a paper arrangement-based process*”, which aims to enhance the traditional paper based approaches in the HCI field with virtual elements. This thesis will study the use of Kinect in detecting and analyzing human position in relation with a wall containing an affinity diagram. With the collected users history data, the thesis will research different ways of visualizing and interacting with the visualization, focusing on displaying changes made in a specific selected region of the wall in a specific time frame. The visualization will be specially designed for being projected using a normal projector on a wall, overlapping with the existing affinity diagram. Interaction with the visualization will be done using the **Microsoft Kinect** sensor, the sensor being situated a few meters **behind** the person.

### **Keywords**

Kinect sensor, projector, user history data, visualization, augment



## Zusammenfassung

Diese Masterarbeit ist Teil eines grösseren PhD Projekts, „**Understanding and augmenting a paper arrangement-based process**“, welches darauf abzielt traditionel auf Papier basierende Ansätze im HCI Gebiet mit virtuellen Elementen zu ergänzen. Diese Arbeit studiert die Verwendung von Kinect zum Erfassen und Analysieren von menschlichen Positionen im Bezug zu einer Wand, welche ein Affinitätsdiagramm enthält. Mit den gesammelten Benutzerhistoriendaten, erforscht die Arbeit verschiedene Wege zur Visualisierung und der Interaktion mit der Visualisierung. Sie fokussiert dabei auf die Darstellung von Veränderungen in speziell ausgesuchten Bereichen auf der Wand innerhalb eines bestimmten Zeitraums. Die Visualisierung wird speziell gestaltet sein, mittels eines normalen Projektors auf die Wand zu projizieren, das bestehende Affinitätsdiagramm überdeckend. Die Interaktion mit der Visualisierung wird mittels Nutzung des **Microsoft Kinect Sensors** gemacht, welcher ein paar Meter **hinter** der Person positioniert ist.

### **Schlüsselwörter**

Kinect Sensor, Projektor, Benutzerhistoriendaten, Visualisierung



## Acknowledgments

With this occasion I would like to thank the following people for their help and patience throughout my master thesis.

I would like to first of all thank Professor Elaine Huang for the given opportunity, for the great advices and for the flexibility when choosing my master thesis subject. Special thanks to my supervisor, Gunnar Harboe, which did a great job in guiding me throughout the thesis, giving me great advices and most important showing me the importance of being structured and organized. The received hints made a big difference during the research and special in writing the final report. I would like to thank to all ZPAC members for the nice atmosphere, for the great moments and for integrating me in their group.

It was a pleasure to work with all the members from the ZPAC department and I look forward for future collaborations.





## Contents

Abstract .....	- 3 -
Zusammenfassung .....	- 5 -
Acknowledgments .....	- 7 -
Contents .....	- 9 -
Figures .....	- 12 -
1. Introduction .....	- 15 -
2. Related Work .....	- 17 -
2.1 Research Question 1 .....	- 17 -
2.2 Research question 2 .....	- 18 -
2.3 Research question 3 .....	- 19 -
3. Scientific Context .....	- 21 -
3.1 Research questions .....	- 21 -
3.2 Research boundaries .....	- 22 -
4. Microsoft Kinect .....	- 23 -
4.1 Interaction .....	- 24 -
4.2 Observations and limitations .....	- 25 -
4.3 Resources .....	- 25 -
5. Project Management .....	- 26 -
5.1 Project plan .....	- 26 -
5.2 Design process .....	- 29 -
Prototyping .....	- 29 -
Evaluation Methods .....	- 30 -
Methods and process .....	- 30 -
Observations .....	- 31 -
6. Implementation .....	- 32 -
6.1 Environment .....	- 32 -
Development environment .....	- 32 -
SQL environment .....	- 33 -
Tools .....	- 33 -
Devices .....	- 34 -
6.2 Architecture .....	- 34 -
6.3 Database .....	- 35 -

6.4 Development process .....	- 36 -
Research question 1 .....	- 36 -
Feasibility study and evaluation .....	- 36 -
Summary.....	- 39 -
Prototype.....	- 40 -
Research question 2 .....	- 41 -
Feasibility study .....	- 41 -
First prototyping session .....	- 42 -
Second prototype session .....	- 43 -
Final prototyping session.....	- 46 -
Research question 3 .....	- 49 -
Feasibility study .....	- 49 -
First prototyping session .....	- 51 -
Second prototyping session .....	- 53 -
Third prototyping session.....	- 55 -
Final prototyping session.....	- 56 -
6.5 Observations.....	- 60 -
7. Summary and Conclusions .....	- 61 -
7.1 Findings.....	- 61 -
7.2Future work.....	- 62 -
7.3 Conclusions.....	- 63 -
Bibliography.....	- 64 -
Abbreviations .....	- 66 -
Appendix.....	- 67 -
A. Database Schema .....	- 67 -
B. Software Engineering .....	- 68 -
C. Research Question 1 .....	- 70 -
Calibration .....	- 70 -
Important Images .....	- 70 -
D. Research Question 2 .....	- 72 -
Session 1 .....	- 72 -
Sketches.....	- 72 -
Findings for the Kinect prototype.....	- 74 -
Findings for the PC application prototype.....	- 76 -
Session 2 .....	- 78 -

Evaluation Sheet prototype 1.....	- 78 -
Evaluation Sheet prototype 2.....	- 85 -
Evaluation Sheet prototype 3.....	- 89 -
Findings.....	- 94 -
Session 3.....	- 98 -
E. Research Question 3.....	- 99 -
Session 1.....	- 99 -
Brainstorming session results document .....	- 99 -
Session 2.....	- 109 -
Consent Form .....	- 109 -
Sketches.....	- 111 -
Findings.....	- 113 -
Session 3.....	- 118 -
Evaluation Sheet.....	- 118 -
Images & Sketches.....	- 119 -
Findings.....	- 122 -
Session 4.....	- 126 -
Sketches.....	- 126 -
Findings.....	- 127 -
Session 5.....	- 128 -
Evaluation Sheet and interview protocol.....	- 128 -
Images .....	- 130 -
Findings.....	- 132 -
Statutory Declaration .....	- 133 -

## Figures

FIGURE 1 - COMPUTER BASED VS PAPER BASED AFFINITY DIAGRAM .....	- 15 -
FIGURE 2 - AFFINITY DIAGRAM LABORATORY SET-UP .....	- 16 -
FIGURE 3 - KINECT SENSOR .....	- 23 -
FIGURE 4 - PRESS GESTURE Z DATA .....	- 24 -
FIGURE 5 - VISUAL STUDIO 2010 .....	- 32 -
FIGURE 6 - POSTGRE SQL PGADMIN TOOL.....	- 33 -
FIGURE 7 – MICROSOFT KINECT STUDIO.....	- 33 -
FIGURE 8 - SYSTEM DIAGRAM.....	- 34 -
FIGURE 9 - NEW DB TABLES .....	- 35 -
FIGURE 10 – CALIBRATION .....	- 38 -
FIGURE 11 - PROJECTED PATTERN VIEWED BY KINECT.....	- 38 -
FIGURE 12 - CALIBRATION UI .....	- 40 -
FIGURE 13 – INTERNAL CALIBRATION PROCESS.....	- 40 -
FIGURE 14 - RQ2 / SESSION 1.....	- 42 -
FIGURE 15 - RQ2 / SESSION 2.....	- 43 -
FIGURE 16 - MAPPING PROCESS .....	- 46 -
FIGURE 17 - KINECT MAPPING UI.....	- 47 -
FIGURE 18 - PC MAPPING TOOL.....	- 47 -
FIGURE 19 - RQ3 / SESSION 1.....	- 51 -
FIGURE 20 - RQ3 / SESSION 2.....	- 53 -
FIGURE 21 - RQ3 / SESSION 3.....	- 55 -
FIGURE 22 - RQ3 / FINAL PROTOTYPE .....	- 57 -
FIGURE 23 - DATABASE SCHEMA .....	- 67 -
FIGURE 24 - SYSTEM DIAGRAM.....	- 68 -
FIGURE 25 - API CLASS DIAGRAM.....	- 68 -
FIGURE 26 - APPLICATION CLASS DIAGRAM (M-V-VM) .....	- 69 -
FIGURE 27 - ENVIRONMENT & DEVICES.....	- 70 -
FIGURE 28 - KINECT SKELETON .....	- 70 -
FIGURE 29 - JOINT ORIENTATION .....	- 71 -
FIGURE 30 - SKELETON TRACKING .....	- 71 -
FIGURE 31 - USER MAPPING (USING KINECT) .....	- 72 -
FIGURE 32 - MENU OPTIONS.....	- 72 -
FIGURE 33 - HEATMAP VISUALIZATION PROPOSAL .....	- 73 -
FIGURE 34 - USER MAPPING (PC APPLICATION) .....	- 73 -
FIGURE 35- Q2 / SESSION 2 / PROTOTYPE 1 .....	- 79 -
FIGURE 36 - Q2 / SESSION 2 / PROTOTYPE 2.....	- 85 -
FIGURE 37- Q2 / SESSION 2 / PROTOTYPE 3 .....	- 89 -
FIGURE 38- Q2 / SESSION 3 / PROTOTYPE 1 .....	- 98 -
FIGURE 39- Q2 / SESSION 3 / PROTOTYPE 2 .....	- 98 -
FIGURE 40 - Q3 / SESSION 2 / PROTOTYPE 1.....	- 111 -
FIGURE 41 - Q3 / SESSION 2 / PROTOTYPE 2.....	- 111 -
FIGURE 42 - Q3 / SESSION 2 / PROTOTYPE 3.....	- 112 -
FIGURE 43- Q3 / SESSION 3 / PROTOTYPE 1 .....	- 119 -
FIGURE 44- Q3 / SESSION 3 / PROTOTYPE 2 .....	- 119 -
FIGURE 45- Q3 / SESSION 3 / PROTOTYPE 3 .....	- 120 -
FIGURE 46- Q3 / SESSION 3 / PROTOTYPE 4 .....	- 120 -
FIGURE 47- Q3 / SESSION 3 / PROTOTYPE 5 .....	- 121 -

FIGURE 48- Q3 / SESSION 4 / PROTOTYPE 1.A .....	- 126 -
FIGURE 49 - Q3 / SESSION 4 / PROTOTYPE 1.B .....	- 126 -
FIGURE 50- Q3 / SESSION 5 / NOTES OVERLAPPING TEST .....	- 130 -
FIGURE 51- Q3 / SESSION 5 / PROTOTYPE 1.A .....	- 130 -
FIGURE 52 - Q3 / SESSION 5 / PROTOTYPE 1.B .....	- 131 -
FIGURE 53 - Q3 / SESSION 4 / PROTOTYPE 1.C .....	- 131 -



# 1. Introduction

“The **affinity diagram** is a business tool used to organize ideas and data. It is one of the **Seven Management and Planning Tools**.

The tool is commonly used within **project management** and allows large numbers of ideas stemming from brainstorming<sup>1</sup> to be sorted into groups, based on their natural relationships, for review and analysis.<sup>2</sup> It is also frequently used in **contextual inquiry** as a way to organize notes and insights from field interviews. It can also be used for organizing other freeform comments, such as open-ended survey responses, support call logs, or other qualitative data.”<sup>3</sup> from Wikipedia.com

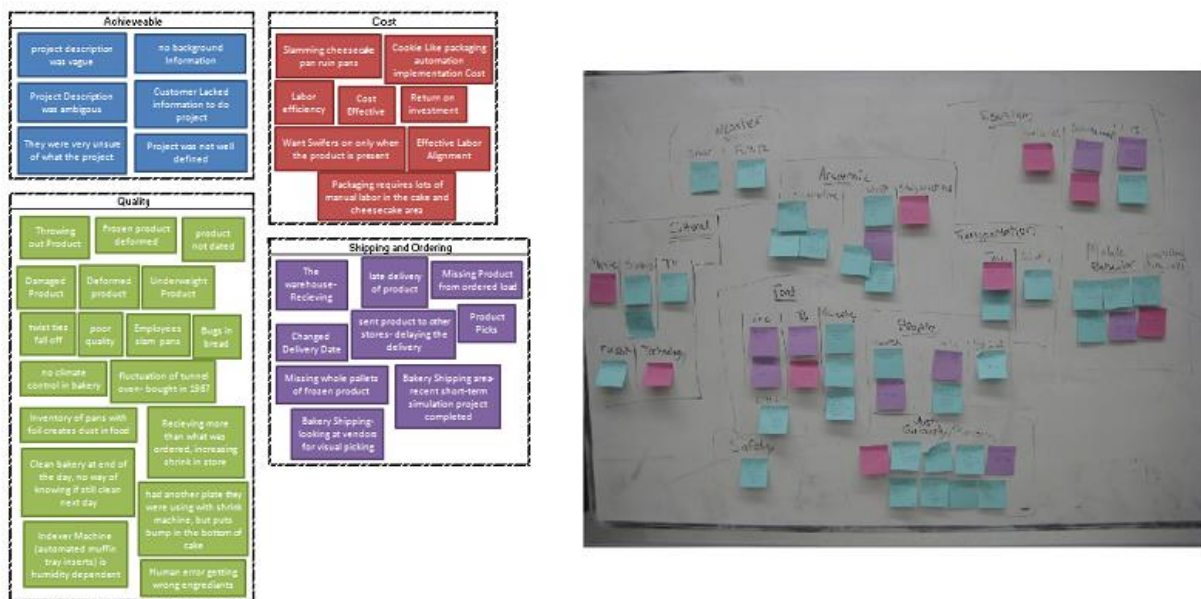


Figure 1 - Computer based vs paper based affinity diagram <sup>4 5</sup>

Affinity diagrams can be designed and developed either on a wall (for example using post-its) or either on a PC using different applications specially designed to support this process. Each method has its advantages and disadvantages. Using digital solutions offers advantages like fast and advanced search, sharing and global access but on the other side makes the development of the affinity diagram constrained to a normal computer screen. A wall based solution is cheaper and apparently preferred by most users (Harboe, Minke, & Huang, 2012) (Judge & P.S., 2008) and offers more flexibility and encourages group work because of the bigger working area.

Much research has been done in this area, in both **wall based** and **PC based** solutions. The current master thesis continues the work of Gunnar Harboe, more specifically the topic **Understanding and augmenting a paper arrangement-based process**. This thesis focusses on exploring and

<sup>1</sup> <http://www.pmhut.com/affinity-diagram-kawakita-jiro-or-kj-method>

<sup>2</sup> <http://www.leanyourcompany.com/methods/Using-Affinity-Diagrams.asp>

<sup>3</sup> [http://en.wikipedia.org/wiki/Affinity\\_diagram](http://en.wikipedia.org/wiki/Affinity_diagram)

<sup>4</sup> <http://edge.rit.edu/content/R12700/public/Affinity%20Diagrams>

<sup>5</sup> <http://pages.cs.wisc.edu/~ambrozia/breadcrumbs/ideation.html>

understanding how a Microsoft Kinect sensor can help in achieving in augmenting this paper based process.

If traditional electronic solutions focus on creating software and hardware devices for better creating and analyzing affinity diagrams, the approach proposed by (Doksam & Gunnar, 2012) proposes the use of available resources found in any laboratory in order to make this interaction better. In the context of the current master thesis, a projector is used to display any controls or visualizations on the wall and a Microsoft Kinect sensor serves as a mean to control the designed application and possible visualizations. The interaction between the person and the PC is done using natural body gestures, gestures which are common in Kinect or Wii games. Because of this, the user will be able to focus on the task in hand which is creating affinity diagrams, and just overlay on the diagram different information without needing to continuously switch his working environment between the wall and the PC. This approach tries to take advantage of the natural advantages of a wall based affinity diagram, exploiting the flexibility, team work and fast development but in the same time improving it by adding virtual elements which should integrate seamlessly with the laboratory itself.

The image below presents the envisioned setup.

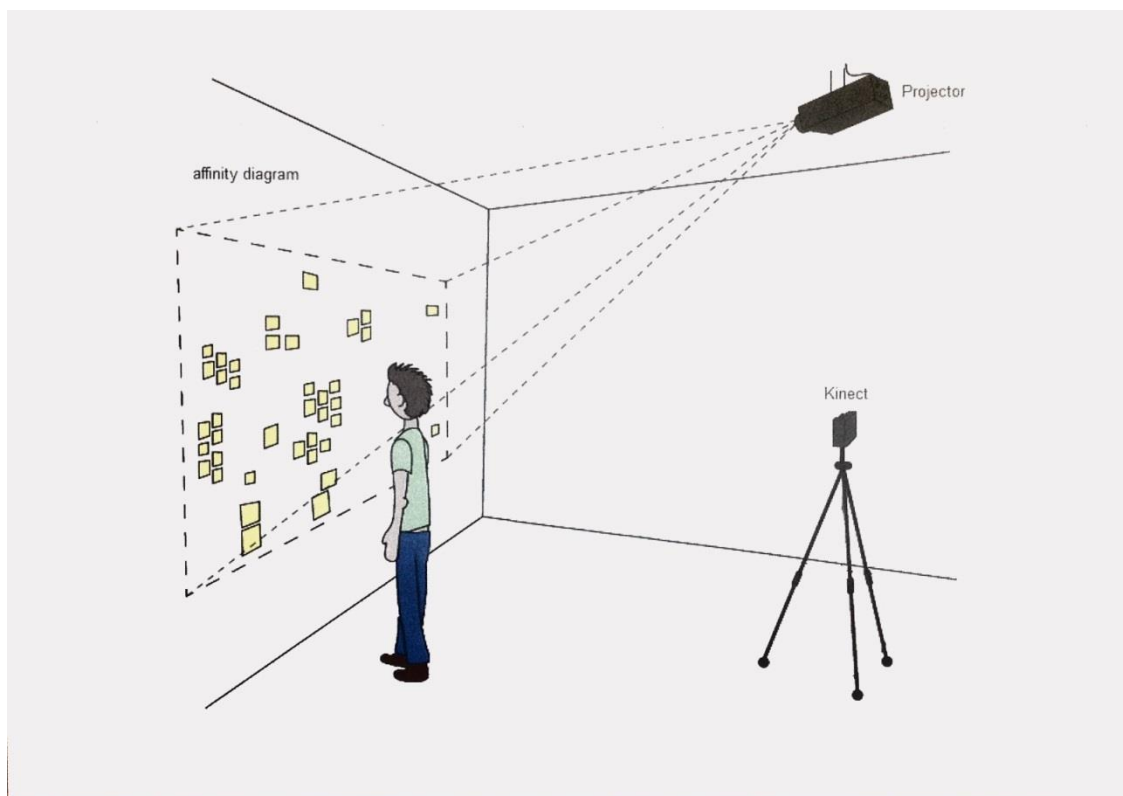


Figure 2 - Affinity diagram laboratory set-up

The objectives of the thesis are to research ways of calibrating existing devices and manually, automatically or semi-automatically identify the people working on the affinity diagram and the actions which they do. Having this history data, the last step is to research visualizations which can provide to the user information about what happen on a specific region of the wall in a specific time frame.



## 2. Related Work

Because of the nature of the thesis and because it consists from three main research questions, the related work is treated in details in the feasibility study section of each question. This chapter presents the highlights of what is presented in these sections.

### 2.1 Research Question 1

**QUESTION:** *How can we track multiple users' presence and activity using Kinect while affinity diagramming, and how would users set up this component as part of the envisioned system?*

#### **Technology**

Different libraries (OpenNI/NITE, Libfreenect, CL NUI SDK, Evolute SDK and Microsoft Kinect SDK) and research papers (Sinthanayothin, Wongwaen, & Bholsithi, 2012) were analyzed.

#### **Bounding box calculation**

From the point of view of calculating the bounding box (a rectangle surrounding the person) different ways were researched. Much research (Sinthanayothin, Wongwaen, & Bholsithi, 2012) (Nambiar, Correia, & Soares) (Xia, Chen, & Aggarwal) (Gulshan, Lempitsky, & Zisserman, Humanising GrabCut: Learning to segment humans using the Kinect) and commercial available solutions (Alahi) exist in this area but none of them have a similar setup like in the case of the affinity diagram, where the Kinect sensor is behind the person.

#### **Calibration**

The table below presents different ways of calibrating the Kinect sensor with the projector.

<b>Manually</b>	<ol style="list-style-type: none"> <li>1. Manual measurements of the distance offset between the sensor and the projected area.</li> <li>2. Manual alignment of the sensor at a pre-set position which should not change during the whole project duration.</li> </ol>
<b>Semi-automatically</b>	<ol style="list-style-type: none"> <li>1. Use the projector to project a pattern on the wall and then use the sensor depth data in order to calculate the distance from it. After calculating the distance, manually align the sensor so that the RGB frame from the sensor aligns with a pre-set position in software.</li> <li>2. Manually align the sensor to match with the top part of the projected area and use a person in order to touch the wall where the bottom left and right rectangles are projected. Using the position of the right hand, we could calculate the position offsets of the sensor.</li> </ol>
<b>Automatically</b>	<ol style="list-style-type: none"> <li>1. Automatically calculate the distance from the wall using the depth stream and use shape recognition in order to automatically detect the red rectangles from the projected shape.</li> </ol>

- |  |  |
|--|--|
|  | 2. Automatically detect a projected shape and its position and compare the detected size with the real size. Use the difference in size in order to detect the distance offsets. |
|--|--|

## 2.2 Research question 2

**QUESTION:** *How can we design and implement a user experience to support automatic or semi-automatic identification of users in the affinity diagramming context?*

Much research is done for automatic and semi-automatic recognition of people. Some of these approaches are presented below.

### Use Excel or write a tool in which we can manually assign to a skeleton-id<sup>6</sup> a person

In this way the information will be filled in the database in 2 steps. First we add him using the skeleton id and after we bundle together the actions of each skeleton and ask the administrator to map the skeleton with the actual user.

#### Problems with this approach

- It is time consuming in order to write the tool.
- A person will need to manually make the mapping.

### Using face recognition

Using face recognition we can semi-automatize the previous solution. If the credibility of the face recognition is big enough, then we assign the person, if not then the person should be mapped manually.

#### Problems with this approach

- Requires machine learning or images with people in order to recognize them
- Can be slow to recognize all the people in the picture
- Requires the people to turn towards the camera so they are recognized
- Problems in detecting dark skinned people

### Using speech recognition

When a person's skeleton is detected for the first time, the system will ask for the user id. At this moment the user will need to identify itself verbally by specifying his user id.

#### Problems with this approach

- Requires the person to turn towards the sensor and specify his user id
- Noise interferences can create problems
- English (language) accent

---

<sup>6</sup> A skeleton-id is a number which Kinect gives to a person when the person enters in the field of view.

- Distance from the sensor
- Specify the used id every time when he comes back in the field of view.

### **Using a predefined body gesture**

Every person can register himself with a predefined body gesture. When a person skeleton will be identified for the first time, he will have to stay in the predefined body gesture so the system can identify him.

### **Problems with this approach**

- The person has to stay in his predefined body gesture every time when he comes back in the field of view.

### **Make users wear different colors and detect the different colors**

Ask the people to associate themselves with a different color and the system will map them based on that.

### **Problems with this approach**

- The person has to wear the same color all the time, or someone needs to associate the person with that color
- No 2 people can wear the same color
- Sometimes colors are very similar from the point of view of the detection.
- The detected color it depends of the amount of light in the room

## **2.3 Research question 3**

**QUESTION:** *What user experiences can we enable with the user activity history and how can they be visualized?*

Because the visualization intends to display the changes in the affinity diagram over time and from the point of view of different people working on it, similar approaches can be enumerated. Among these approaches we can specify:

### **Video recording**

Video recording of the wall is a basic and simple technique which offers all the information needed in order to find out what happened with a specific area of the wall. One can simply use a video camera, record the affinity diagramming process and afterwards simply visualize the changes.

### **HD Photo camera**

Similar to the approach presented above, an HD photo camera can be used and programmed in order to take pictures of the wall at specific time intervals. The stored photos can be later compared automatically and a log or visualization showing the changes between photos could be created.

Examples of applications which can be used to compare two photos are ImageMagick<sup>7</sup> and Photoshop. In Photoshop, images can be loaded in different layers and afterwards a subtractive effect on the layers could be applied to see the difference. ImageMagick can also be used using the command line, therefore a Windows BAT files could be written in order to make the analysis more automatic.

### ***Affinity diagram log***

A designated person can keep a log of all important changes during the affinity session. These changes could be aggregated in a Microsoft Word document. Simple features as find, heading styles and tables can be used in order to organize and find the data. This is a very manual process and valuable intermediate changes could be ignored.

---

<sup>7</sup> <http://www.imagemagick.org>

### 3. Scientific Context

Many software applications exist for creating virtual walls with affinity diagrams, but even if it offers different advantages like search, filtering, smart tagging and so on, people still prefer creating affinity diagrams on a real wall using post-its (Judge, Pyla, McCrickard, & Harrison, 2008). Usually the reasons for this are very simple, post-its are more flexible and easier to use, do not require such a big extra cost and more important they offer the possibility to have an overview of the whole wall due to its bigger size. As stated before, the thesis is part of a larger project which aims to augment the physical world with virtual elements, which can prove to be a very good compromise from both approaches. Specific to this case, in the field of HCI, the project goal is to research ways of automatically or semi-automatically detect the people which are working on the affinity diagram and store their actions in order to be later processed and visualized in a meaningful way.

In order to detect the people working on the affinity diagram, a Microsoft Kinect sensor will be used. The sensor will provide skeleton and depth information which will be later used to derive the actions and position in a 3D space. A feasibility study on the Kinect sensor has already been done and the initial results show that it is actually possible, within a low cost, to detect the actions of a person.

#### 3.1 Research questions

The thesis is built around the following 3 research questions.

**QUESTION 1:** *How can we track multiple users' presence and activity using Kinect while affinity diagramming, and how would users set up this component as part of the envisioned system?*

A preliminary research revealed different ways of detecting and calculating the bounding box of a person. Part of the scope of the current thesis is to research which of the possible ways is best fitting for the project scope. Detecting a person interaction, movement and position in 3D space is a core component of the project.

**QUESTION 2:** *How can we design and implement a user experience to support automatic or semi-automatic identification of users in the affinity diagramming context?*

When working with the affinity diagram different people may be present in the field of view of the camera. Each person may be a spectator or an actual project member, and they may come in or go out from the field of view. Part of the scope of the current thesis is to research the best way to automatically or semi-automatically detect the project members in order to store their actions under the right user in the database.

**QUESTION 3:** *What user experiences can we enable with the user activity history and how can they be visualized?*

Only having information about a user does not help in any meaningful way if the results are not displayed properly and the interaction with the visualization is not adequate. The last part of the thesis aims to find the best way of rendering and interacting with these results.

## 3.2 Research boundaries

### ***Algorithms***

Because the thesis focusses on researching the best ways of visualizing and interacting with the designed prototypes, no special focus will be put on the different algorithms which will be used. We do recognize that the use of algorithms with better performance can change the way results are visualized and the way the user interacts with the final prototypes, but because the thesis is done from an HCI perspective we assume that better algorithms can only improve our findings. Beside this, no special attention from the point of view of software quality (application speed, testing ...) was made when designing the final software prototype.

### ***Kinect gestures***

If the first part of the thesis focusses on understanding how Kinect can be incorporated in the paper augmented process and how can the data provided by Kinect can be used in this process, no special attention was and will be put in researching, designing and implementing Kinect gestures. By default, Kinect does not provide gestures identification but it does provide the proper platform for such extensions to be implemented <sup>8</sup> (Ryan). Much research and many prototypes can be found as 3<sup>rd</sup> party extensions, therefore we will only assume that the desired functionality can be easily implemented and incorporated in the final prototype.

### ***Images and Icons***

As visualizations and interactions are the target of the thesis, any icons and images used in the final prototypes are not considered part of the research. The current research showed that it is very important to use high quality and intuitive images when describing natural hand gestures which should be understood by the Kinect sensor, but no special attention or effort was put in drawing these images. It is noted that using better images can substantially improve the user understanding of the prototype, but no research will be done in finding out how exactly these images should look like.

### ***Light Sources***

Because the visualizations are displayed using a projector on a wall, interferences from the natural light affects how the colors are interpreted (Ryan). In order to minimize the effect of the light on the colors and their intensity (especially when the light changes because of the clouds), the final prototypes were tested in a restricted environment in an HCI laboratory where all the blinders were down and only artificial light was used.

---

<sup>8</sup> [http://hackanui.com/resources/Human\\_Interface\\_Guidelines\\_v1.7.0.pdf](http://hackanui.com/resources/Human_Interface_Guidelines_v1.7.0.pdf)

## 4. Microsoft Kinect

Because of the nature of the project and because of user trends when working with affinity diagrams (Harboe, Minke, & Huang, 2012) a natural way of making the interaction between a computer and the natural environment (the laboratory and the papers on the wall and boards) was needed. In order to achieve this natural interaction, new devices were needed, but in the same time cost and already present devices in the laboratory should be considered. Different options were available but in the end Microsoft Kinect sensor proved to be the best solution due to its stable SDK provided by Microsoft, because of the vast types of data which were provided and in the same time at a lower cost.

Microsoft Kinect sensor for Windows was first released in spring 2011<sup>9</sup> and since then it is vastly used in research, especially in HCI and robotics departments. The sensor comes together with an SDK which gives the possibility to process raw data related to 3D depth and RGB but in the same time due to its 4 microphones located at perfect aligned distances on the sensor, gives the possibility not only to give audio commands but also to detect the region from where the sound is coming from.



Figure 3 - Kinect Sensor<sup>10</sup>

Some of the most important elements of the sensor and SDK are the ability to give the 3D depth information and the position of the detected skeletons. This data is out of the box and does not require much programming skills in order to use it. The depth data gives a feeling of how far or near objects are, while the skeleton information gives valuable information about how many people are in the field of view and where are the joints localized (in 3D space). The values returned by the sensor are in meters and are calculated with the sensor position at the 0 position.

Different other solutions are available which can provide the same functionality but most of them involve using normal cameras connected to a computer. The cameras have similar role as the sensor and are used in order to get the image which is later on processed by different algorithms, the end

<sup>9</sup> [http://en.wikipedia.org/wiki/Kinect#Kinect for Windows](http://en.wikipedia.org/wiki/Kinect#Kinect_for_Windows)

<sup>10</sup> <http://channel9.msdn.com/Series/KinectSDKQuickstarts/Understanding-Kinect-Hardware>

result being the location of different body parts (hands, legs, head ...). Since the appearance of Kinect sensor, this last method is less used mainly because of higher costs in buying an HD camera and because no commercial available SDK's are available which can provide the same functionality as Kinect does out of the box.

## 4.1 Interaction

As specified in the last paragraph, Kinect offers two main ways of interaction. The first one is using vocal commands. This method gives the possibility to instruct the system to react to specific keywords or phrases. Different examples which come together with the SDK are available and give a very good feeling of how to use this feature. Another way of interaction (and one of the most used) is using body gestures or body position. Comparing with the first type of interaction which was based on sound, this one is based on the depth and RGB streams. Combined, these two streams give the possibility to detect the position in space of nearly all human joints. Having the position of the joints one can instruct the system to react when one of the joints is in a specific position or when specific gestures are being executed. Kinect SDK does not come with gesture recognition included but the available resources make it easy to implement. Besides this, 3<sup>rd</sup> party libraries exist which can do gesture recognition fairly easy.

Some of the things which can be easily done using Kinect (from the interaction point of view):

- Detect if the person is pushing his hand in front (in this way simulating the push of a button)
- Detect the position of a person in relation with other objects
- Detect the position of the hands of the person (ex: for drag and drop purposes)
- Detect if a person is seated or not.

Simple data recordings (ex: position of the right hand) and comparing the values over time can easily indicate specific gestures without the need of specific libraries. Of course, better libraries can make the integration and the interaction more natural, in this way giving better results when controlling different UI elements.

The figure below demonstrates how recording the Z coordinate of the right hand over time can detect the "press gesture". Knowing such peaks, the system can automatically react to this specific gesture and automatically adjust and learn the behavior of the user. The X axis represents the time while the Y represents the distance from the sensor in meters.

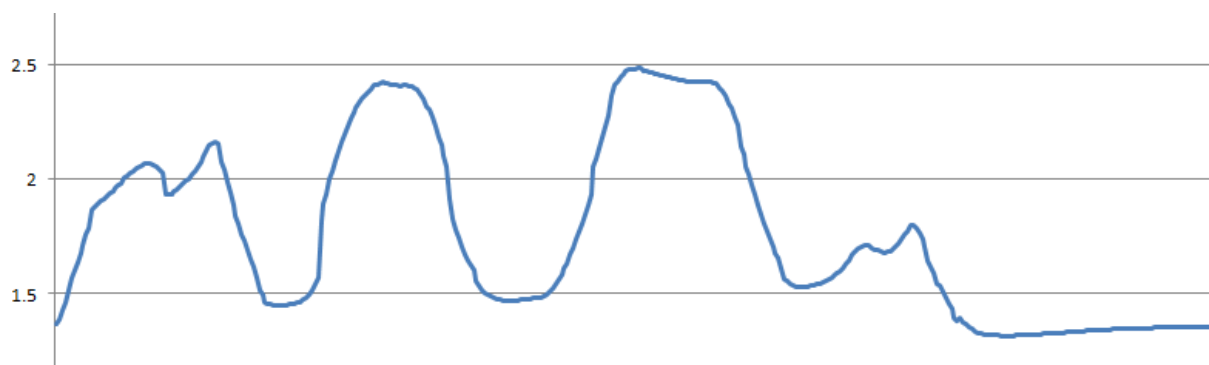


Figure 4 - Press gesture Z data



## 4.2 Observations and limitations

As specified in the last paragraph, Kinect has many advantages from the point of view of the data which it offers, and offers many future possibilities. Unfortunately even at version 1.4 it is still limited from the point of view of the out of the box functionality. As specified, Kinect does not offer features like gesture recognition, object detection, body shape and people counting (more than 6) but fortunately this can be achieved using the depth, RGB and sound streams (Xia, Chen, & Aggarwal) (Gulshan, Lempitsky, & Zisserman).

Based on the **Human Interface Guidelines v1.7.0** published by Microsoft, Kinect can recognize maximum 6 people but can track maximum 2, by tracking meaning that Kinect will offer the joint information only for maximum 2 people but it will be able to inform the system of the number and location of the other detected people. Also according to Microsoft, the sensor works best in the 0.8m to 2.5m interval but this can be extended to maximum 0.4m – 3m. Only one Kinect sensor should be used in a specific area. The sensor projects a pattern of infrared light to calculate the depth of the people, allowing the recognition of different body parts. If multiple Kinect sensors are used, the infrared light will interfere with the light from other sensors and one may notice a reduction in the accuracy and precision of the skeletal tracking.

Because of the nature of the two streams (RGB and depth) different objects or light sources can also interfere and as a result generate wrong data. For example, having reflective objects can reflect the infrared light emitted by the sensor; besides this, strong light from different artificial or natural light sources can decrease the performance of the RGB stream. Another limitation of the sensor consists in the detection of the skeleton which can be increased or decreased depending on the shape and color of what people are wearing.

The Kinect sensor can detect if a person is seated or not and can adjust the motor angle from +27 to -27 degrees.

## 4.3 Resources

In order to work with the Kinect sensor, besides the hardware elements, some drivers, APIs and SDKs need to be installed. No specific configuration is needed.

Kinect SDK	<a href="http://www.microsoft.com/en-us/kinectforwindows/Develop/developer-downloads.aspx">http://www.microsoft.com/en-us/kinectforwindows/Develop/developer-downloads.aspx</a>
XNA	<a href="http://www.microsoft.com/en-us/download/details.aspx?id=23714">http://www.microsoft.com/en-us/download/details.aspx?id=23714</a> <a href="http://www.microsoft.com/en-us/download/details.aspx?id=20914">http://www.microsoft.com/en-us/download/details.aspx?id=20914</a>
Speech platform	<a href="http://www.microsoft.com/en-us/download/details.aspx?id=27226">http://www.microsoft.com/en-us/download/details.aspx?id=27226</a>
Visual Studio	<a href="http://www.microsoft.com/visualstudio/eng/downloads">http://www.microsoft.com/visualstudio/eng/downloads</a>

## 5. Project Management

The project plan played an important role in the current thesis. From the beginning of the work, each research question and big milestones were planned. Extra time was added to each important phase in order to cope with any delays which may occur. Interviews and methods used were planned in advance in order to have a clear view of what type of findings, which answers will be available and when. Based on the timeline, the prototype was gradually designed and implemented in such a way to take advantage on one hand on the findings but on the other hand on the available time.

### 5.1 Project plan

The project plan presented below represents the original planned time. Each research question defined a more detail plan where each evaluation method and main questions are described. Slight delays appeared during some evaluation methods (especially during the last research question) and this was mainly because of the vacation time which made it difficult to find people for running the evaluation sessions.

Weeks	RQ	Title	Milestone
1 - 4		Tech investigation and initial design	
5,6	Q1	Bounding Box Research & Implementation	
7		People Recognition Research	
8		Run first paper prototypes with users	
9	Q2	First Implementation	
11		Run implemented prototype with users	
12		Gather feedback and implement them	
13		Research related to visualizing user actions	
14	Q3	Implement initial ideas in a prototype	
17		Test prototype with users and gather feedback	
18		Implement feedback	
19		Test entire system with users	
20 -24		Finish tests Finish thesis report and other documentation	

The tables presented below represent the individual research question project plan with individual/specific evaluation methods and sub-questions.

#### Research Question 1

Status	Due By	Task
	03.09.2013	Start with the thesis

10.09.2013	Prepare development environment and reading <ul style="list-style-type: none"> <li>• Visual Studio</li> <li>• Databases</li> <li>• Kinect</li> </ul>
17.09.2013	Create a <b>dummy Kinect application</b> . Take decisions about what tools will be used.
24.09.2013	Redesign the DB in order to add the bounding box data and integrate the database with the existing Kinect dummy application.
30.09.2013	Research different ways of calculating and creating the bounding box.
03.10.2013	Implement the findings in a <b>new software prototype</b> .
09.10.2013	Research different ways of doing the calibration between the projector and Kinect.
16.10.2013	Have the final calibration method implemented.
18.10.2013	Finish the software testing <ul style="list-style-type: none"> <li>• Calibration</li> <li>• Database I/O</li> <li>• Kinect connection and streams processing</li> <li>• Bonding box calculations</li> </ul>

### Research Question 2

Status	Due By	Task
	<b>22.10.2013</b>	Initial ideation sessions
	25.10.2013	Run first prototypes with experts. ( <b>think aloud</b> )
	01.11.2013	Gather feedback from the interviews and transcribe them to text.
	04.11.2013	Design new prototypes
	05.11.2013	Run a <b>cognitive walkthrough</b> with HCI experts.
	11.11.2013	Gather feedback
	22.11.2013	Use the feedback to : <ul style="list-style-type: none"> <li>• create new prototypes</li> <li>• test their potential by creating small software prototypes</li> <li>• have <b>open discussion with users</b> about how realistic and useful a prototype is</li> </ul>
	26.11.2013	Consolidate all the feedback and refine the paper prototype for implementation.

02.12.2013	Finish implementing the prototype
06.12.2013	Finish testing the software: <ul style="list-style-type: none"> <li>• Kinect gestures</li> <li>• Database I/O</li> <li>• User mapping</li> <li>• Image import /export</li> </ul>

### Research Question 3

Status	Due By	Task
	<b>09.12.2013</b>	Start working on the visualization part
	13.12.2013	Create more prototypes
	17.12.2013	Define 2-3 main prototype lines which I want to focus on
	17.12.2013	Find some people for testing prototypes: <ul style="list-style-type: none"> <li>• Create a doodle form and send an invitation for helping in the questionnaire to as many people as I can find</li> <li>• Send the invitation to participate in the questionnaire to a few Facebook friends</li> </ul>
	19.12.2013	Run a <b>think aloud</b> prototype session in order to find out if: <ul style="list-style-type: none"> <li>• Do they understand what the visualization does?</li> <li>• Do they understand how to interact with the visualization?</li> <li>• Does this visualization offer the info they want?</li> </ul>
	23.12.2013	Gather feedback from the evaluation session and improve the visualization. Focus also on answering the question: "What form of visualization do you prefer?"
	07.01.2014	Have some <b>expert discussions</b> with the department colleagues
	13.01.2014	Gather feedback, improve visualization and start implementing it
	23.01.2014	Do a <b>functional prototype test</b> with some of the users from above and see <ul style="list-style-type: none"> <li>• How are they using the prototype?</li> <li>• What sections are they focusing on and what are they ignoring?</li> <li>• How is the interaction?</li> </ul>
	28.01.2014	Gather feedback; improve visualization and do a second <b>functional prototype test</b> .
	31.01.2014	Gather feedback, improve visualization and make the final prototype.

## 5.2 Design process

### Prototyping

The prototype sessions represent one of the main parts of the thesis. Different types of prototypes were made, all depending on the questions and the evaluation method which were intended to run on.

#### ***Brainstorming***

This method, together with paper prototyping, was used in the moments when was needed to gather fast and large amounts of different ideas related to a specific subject. Usually this method is suited for the beginning of a stage and that is why it was only planned and executed in the beginning of research question 1 and research question 3. The data provided by the brainstorming session was used in solving technical issues, in drawing basic prototypes and answering general questions either related to the research question itself or either related to a sub-question. Examples of results given by this method can be found in the Appendix.

#### ***Paper Prototyping***

This is the main type of prototype used throughout the thesis. This is a fast and easy to apply method which does not require extra cost and which can be executed in any environment and without any specific tools.

#### ***White Board prototyping***

White board prototyping was used when prototypes were hard to draw on paper. This particular method was used because of the nature of the project where mobile small white boards played the role of obstacles and people, when designing the prototypes, which would be projected by the projector on the wall. The same method was used to test other prototypes when working with the Kinect sensor.

#### ***PC drawing prototyping***

This type was used in more advanced prototypes where special graphics and UI elements were important for the evaluation method. When working with images and charts it was better to give a more realistic feeling about the prototype.

#### ***Mockups***

Similar to PC drawing prototypes, some evaluation methods required more advanced and detailed prototypes. This method is also very useful when defining prototypes in detail and where copy/paste and undo/redo options play an important role for the speed.

#### ***WPF Prototyping***

WPF prototyping is a special case of PC drawing where instead of using tools like Paint or Power Point, the Visual Studio environment was used to build the UI in WPF using the integrated designer and XAML code. This method was used in the final stages of the prototypes where UI and interaction

was very important. This method was also used in order to speed up the development process because the final prototype was anyway intended to be implemented in WPF.

### ***Simulations***

Simulations were used in combination with WPF prototyping in order to animate or simulate part of the prototype. The simulations were written in C# or XAML code.

## **Evaluation Methods**

Evaluation methods and their associated questions were planned in the beginning of each research question. Even with this planning, changes were made in terms of the method used and this was necessary because of previous findings and because of the change in sub-questions.

### **Methods and process**

Throughout the project different evaluation methods were used. The evaluation method defined how the prototype was designed; the method being selected based on the type of input which was needed but also based on how young (from the point of view of development) the prototype was. Consent forms and interview protocols were made when needed, especially in the case of working with people outside the HCI department.

### ***Analysis of existing research***

This was the method chosen especially for the first research question and it was used mainly because of the technical nature of the research. This method involved analyzing existing research and deriving advantages and disadvantages of existing methods; the final decision being taken based on the combined findings from different other research papers.

### ***Interviews***

Interviews were used in the first stages of the prototype in order to test and answer big questions related to interactivity and intuitiveness, without going in details related to the data which is presented. The scope of the interviews was mainly to identify which UI elements, shapes and interactions are preferred by users and where are the main design problems. Interviews were always conducted based on a script and at the end of the interview participants were encouraged to give any feedback desired. If allowed by participants, interviews were audio recorded and afterwards transcribed.

### ***Think aloud***

This method was also used in the early stages of the prototypes when it was desired to make a comparison between the envisioned design and the one interpreted by the user. Similar to the interviews, the evaluation session followed a specific script where the participant was asked to explain or to resolve specific tasks. In order to speed up the process, almost all sessions were audio recorded and afterwards transcribed.

### ***Expert discussions***

Expert discussion evaluation method (similar to think aloud) was done in the mid stages of the prototype and was executed with different people from the HCI laboratory or HCI experts. The scope of this method was mainly to get more valuable feedback of the prototype but from the point of view of an expert targeting elements which normally would not be observed by non-experts.

### ***Cognitive walkthrough***

In the case of the thesis this method made use also of the available HCI experts and was chosen in order to get written feedback about specific mockups in relation to a task and a specific set of heuristics.

### ***Heuristic evaluation combined with expert discussions***

Heuristic evaluation was used in order to identify problems in the user interface based on a given set of heuristics and on a pre-set grade scale. This type of written evaluation offered valuable feedback about existing problems and about how important/damaging they are for the final user. The session was always conducted in two parts, first in writing-individually and the second as an open discussion with the HCI experts. Before the evaluation itself, an introduction to the system was given, the same introduction which a normal end user would receive.

### ***Functional prototype test***

This evaluation method was only used in the end of the prototype where test users were asked to solve specific tasks or to interact with the final version of the implemented prototype.

## **Observations**

Evaluation methods were planned based on previous experience in UI design, but problems shortly appeared because of the nature of the prototype. In normal conditions people would easily relate to UI elements because of the similarity with other applications. In the case of the project, because of the use of Microsoft Kinect sensor as the mean of interaction and because of the use of a projector as a mean of visualization, it was hard for a test user to judge a prototype from these new points of view.

Designing prototypes for a new system in which new kinds of visualizations and interaction means are envisioned can be challenging because test users need extra training before doing the actual evaluation. The amount of information given during the training minutes can be challenging to determine and can easily influence the quality of the results. Even after training, many users tend to forget and still use the system as they would use it with the mouse and assume they have UI elements as scrollbars and popup menus, elements which are not available when designing Kinect UIs. Many elements, when working with Kinect sensor, are hard to prototype especially because of the requirement of having a more clear design and with bigger UI elements.

## 6. Implementation

### 6.1 Environment

From the point of view of the work being done, part of the thesis consists from the research and evaluation of different prototypes and part from the actual implementation of the final prototypes. Because of this, the environment can be described in two ways, from the point of view of implementation and from the point of view of evaluation.

From the point of view of evaluation the prototypes were designed and evaluated in an HCI laboratory using available materials and devices as pen and paper, whiteboards and post-its, projector, Kinect sensor and a Windows computer. The people which evaluated the prototypes were part of four different categories:

- HCI experts
- Students
- Researchers
- Business people (working as managers but with engineering knowledge)

From the point of view of implementation, the prototypes were developed in C# in Visual Studio using WPF technology.

#### Development environment

Because implementation was in second from the point of view of priority of the thesis, C# and WPF were chosen because of the student vast knowledge in the technology, knowledge which would increase the speed of development and would allow more focus on the prototype research and evaluation. Besides this, Windows platform and Visual Studio were chosen because of fast integration with the Kinect sensor SDK and all resources offered with it.

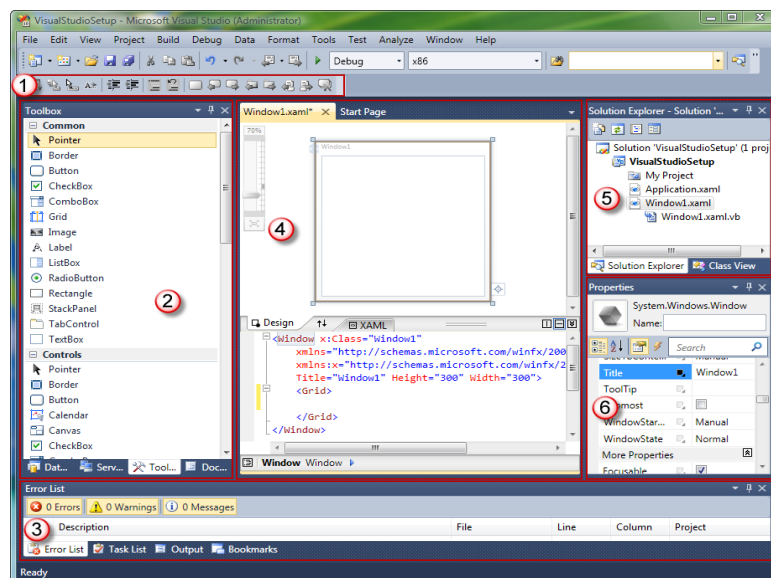


Figure 5 - Visual Studio 2010 <sup>11</sup>

<sup>11</sup> <http://blogs.msdn.com/b/wpfsdesigner/archive/2010/01/14/setting-up-visual-studio-for-wpf-and-silverlight-development.aspx>



## SQL environment

Because of the nature of the project recording, storing and visualizing the data collected from the Kinect sensor was a priority. From the point of view of storing, an SQL based environment was used and because of integration with past projects, Postgre SQL was used. pgAdmin Tool was used to configure and extend the existing database.

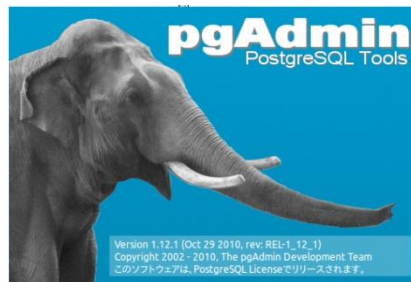


Figure 6 - Postgre SQL pgAdmin Tool

## Tools

Besides the pgAdmin Tool for database management, Kinect Studio was used for testing purposes. Kinect Studio is part of the Kinect SDK and is mainly used to create video recordings which can be later used for testing purposes. The recording saves the RGB, depth and audio streams and is the perfect alternative to live video feed. In order to use a saved recording one needs to open the file in Kinect Studio and connect the Kinect enabled application to it. This tool offers the possibility to run the streams in a loop and in specific intervals, making the debugging experience a lot better comparing with the live feed.

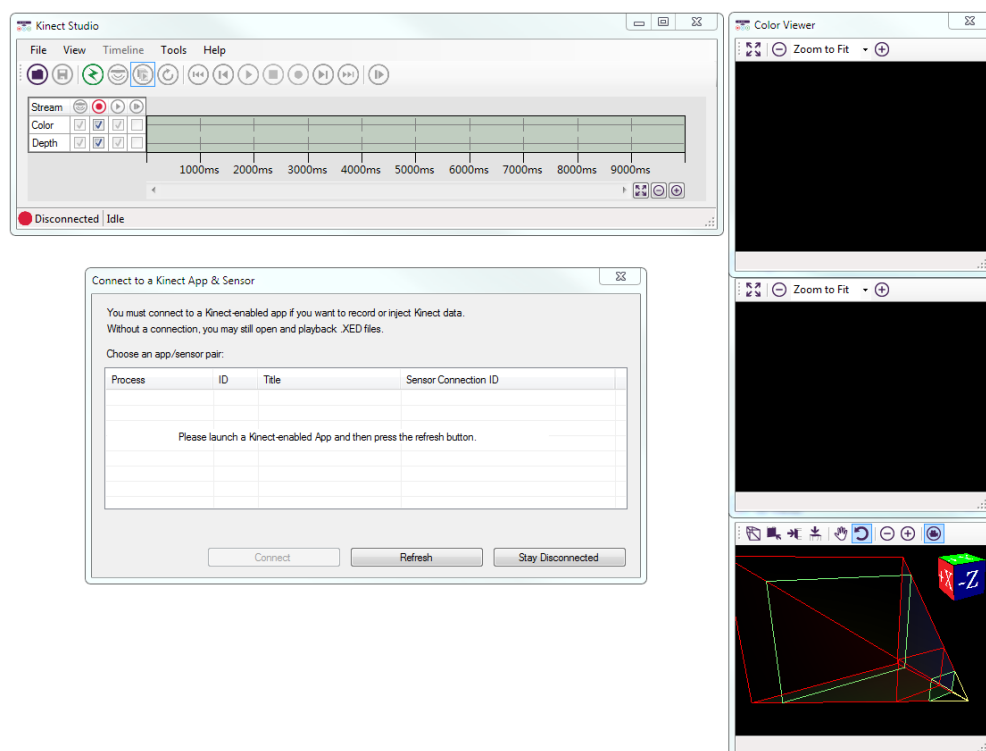


Figure 7 – Microsoft Kinect Studio

## Devices

No special devices besides a Kinect sensor and a projector were used for this project.

## 6.2 Architecture

In order to take advantage of the commonalities between prototypes, all three research questions were implemented in one application. Different menu items give access to different visualizations and prototypes associated with different research questions. The architecture of the application and its building blocks are presented in the image below.

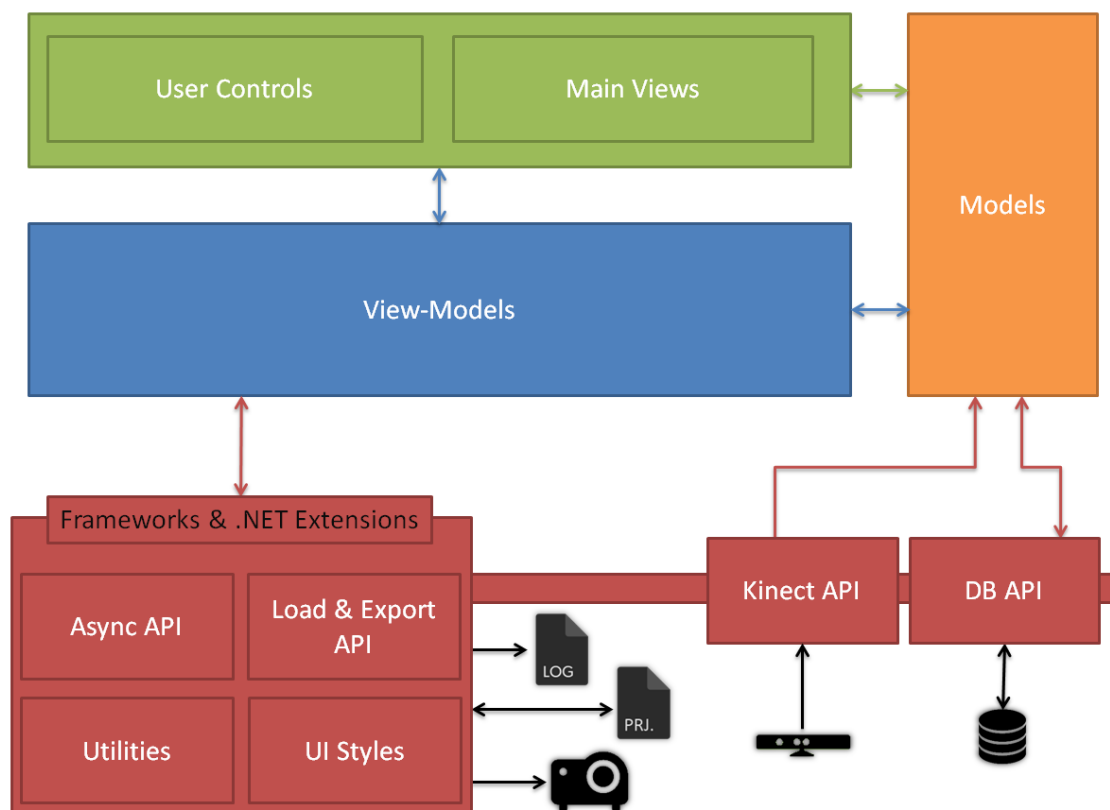


Figure 8 - System diagram

The application follows the MVVM (Model - View - View Model) architectural pattern which is the standard recommended by Microsoft when building WPF applications. Using this standard makes it easy to separate the logic of the application from the model part (the way the raw data is represented in memory) and from the UI part. The communication between view-models and models is done using normal events or direct relationship while the communication between view and view model is done using XAML bindings and commands. Bindings are used to automatically keep the UI in the same state reflected by the properties from the view model. Any update in a property in the view model will automatically update that property at the linked UI element. For example the update of a property "Name" from the view model which is linked with the property Text of a textbox from the view, will be done automatically in both directions. Because of this separation of layers, one can design the UI of the application only by knowing the interface to the view model, and other can

implement the logic of the view model only by knowing the properties which need to be displayed (without thinking about what kind of UI elements render the specific property and how).


Beside from models, views and view models, the application contains a DB API customized for our database which makes the read and write transparent to the type of database which is used. The Kinect API contains extensions on top of the Microsoft Kinect API which are used for specific tasks related to this project.

The direction of the arrows show the direction in which the data flows.

## 6.3 Database

In order to meet new requirements, the database was extended with the following tables:

- **affinity\_projects**: stores information about a project
- **user\_activities**: stores information about an activity. The activity (action row) can be MOVE, IN or OUT and defines the actions which a person does while working on the affinity diagram.

affinity_projects		
	<b>project_id</b>	INTEGER
	<b>main_user_id</b>	TEXT
	<b>informants_id</b>	TEXT[]
	<b>project_name</b>	TEXT
	description	TEXT
	resource_link	TEXT


user_activities		
	<b>activity_id</b>	INTEGER
	<b>user_id</b>	TEXT
	<b>time_code</b>	TIMESTAMP(6) WITHOUT TIME ZONE
	<b>project_id</b>	INTEGER
	bounding_box	TEXT
	position_offset	TEXT
	action	TEXT
	skeleton_id	TEXT

Figure 9 - New DB tables

## 6.4 Development process

### Research question 1

***How can we track multiple users' presence and activity using Kinect while affinity diagramming, and how would users set up this component as part of the envisioned system?***

A preliminary research revealed different ways of detecting and calculating the bounding box of a person. Part of the scope of the current thesis is to research which of the possible ways is best fitting for the project scope. Detecting a person interaction, movement and position in 3D space is a core component of the project.

Before calculating the bounding box of a person and storing the information in the newly added tables, the calibration between the projector and the Kinect sensor needs to be done. This is needed because the coordinates given by the sensor are all the time relative to the sensor itself. Positioning the sensor at different distances from the wall will give us different measurements over time which is not compatible with the overall system. In the end, the system will need to use the coordinates to calculate the distance in 3D space of the user in relation with the wall, therefore first knowing the position of the sensor itself gives us a mean of calculating the user absolute positions.

From the point of view of the research, the question was split in three main areas:

- Technology research and setup
- Bounding box calculation
- Sensor calibration

### Feasibility study and evaluation

From the point of view of technology research, different SDKs and frameworks were researched. This was mainly done consulting existing web resources and research papers and in the end deciding taking into consideration their advantages and disadvantages. According to (Sinthanayothin, Wongwaen, & Bholsithi, 2012) and taking into consideration the new Kinect SDK version and limitations of the project itself, Microsoft Kinect SDK was chosen as the best choice. Other libraries like OpenNI/NITE, Libfreenect or CL NUI were discarded because of complicated integration procedures or low support for skeleton tracking. Only one SDK (Evoluce SDK) came closer to the native Microsoft Kinect SDK, but taking into consideration the resources in terms of examples and documentation available on the Microsoft website, the Kinect SDK version was declared as the winner.

As specified earlier, Kinect SDK comes together with an impressive set of examples, with the native SDK, an offline/online examples browser (gallery) and with Kinect Studio which gives the possibility to record and store different streams (audio, RGB, depth) and later be used for debugging, making the development and testing of the application much faster.

From the point of view of calculating the bounding box (a rectangle surrounding the person) different ways were researched. Much research (Sinthanayothin, Wongwaen, & Bholsithi, 2012) (Nambiar, Correia, & Soares) (Xia, Chen, & Aggarwal) (Gulshan, Lempitsky, & Zisserman, Humanising GrabCut:

Learning to segment humans using the Kinect) and commercial available solutions (Alahi) exist in this area but none of them have a similar setup like in the case of the affinity diagram, where the Kinect sensor is behind the person.

The list below brings together some of the libraries and approaches which can potentially (at least partially) meet the project needs.

**Point Cloud Computing (PCL)**<sup>12</sup> library is written in C++ and uses a RGB stream in order to detect the people which are moving in the field of view of the camera. According to reviews, this method seems promising and source code is available to download for free, but it may be problematic to port it to C# and integrate it with Kinect. Because of the differences in programming language, this method was dropped.

**Using the research of PhD. Alexandre Alahi**<sup>13</sup> from Signal Processing Lab, EPFL which created an algorithm based on Kinect SDK in order to detect people, do noise reduction and calculate their bounding box. Unfortunately the algorithm and the research are patented and obtaining the source code and the research can prove problematic and time consuming.

**Using the Kinect depth stream** in order to calculate the body shape and based on that to calculate the bounding box. According to (Xia, Chen, & Aggarwal) (Gulshan, Lempitsky, & Zisserman, Humanising GrabCut: Learning to segment humans using the Kinect) it is possible to use the Kinect depth stream to not only identify the number of people but also their body shape which can give better information comparing with the bounding box. Unfortunately, calculating the body shape from a depth stream is outside the scope of the project and no open source algorithms were found which can do the required task. The papers do provide a step by step method in how to achieve the human detection but implementing and testing the algorithm is time consuming.

Taking into consideration already existing research and the possibilities provided by Kinect SDK, an estimation approach based on skeleton information was implemented. Natively, Kinect SDK provides based on the RGB and depth streams the estimated position of all joints from the tracked skeletons. The skeleton does provide the position in space but it does not provide a way of calculating the depth of the bounding box. Because of this, the depth value is estimated based on the height of the person. The height of the person is calculated by summing the values of the head, center of shoulders, spine, center of hip, right hip, right knee and right ankle. Because some of the joints can be obscured by obstacles (tables, chairs ...), depending on how many of the joints are not visible, the algorithm will try to estimate the rest of the height based on the available information. Taking into consideration that Kinect can only track two people at the same time, this means that only two skeletons will have joint information available. In order to solve this problem, the algorithm is assigning each skeleton an ID and rotates through the detected users in order to calculate and update the bounding box information. By decreasing the update interval, one can increase the performance of the application but risk of having older information regarding the detected skeleton. An asynchronous version of the update can be implemented which can solve the performance problem but increase the computation in the case the value is requested often and this is because no caching mechanisms can be safely implemented.

<sup>12</sup> [http://pointclouds.org/documentation/tutorials/ground\\_based\\_rgbd\\_people\\_detection.php](http://pointclouds.org/documentation/tutorials/ground_based_rgbd_people_detection.php)

<sup>13</sup> <http://actu.epfl.ch/news/connecting-kinects-for-group-surveillance/>

From the point of view of the sensor calibration different automatically and semi-automatically methods were researched.

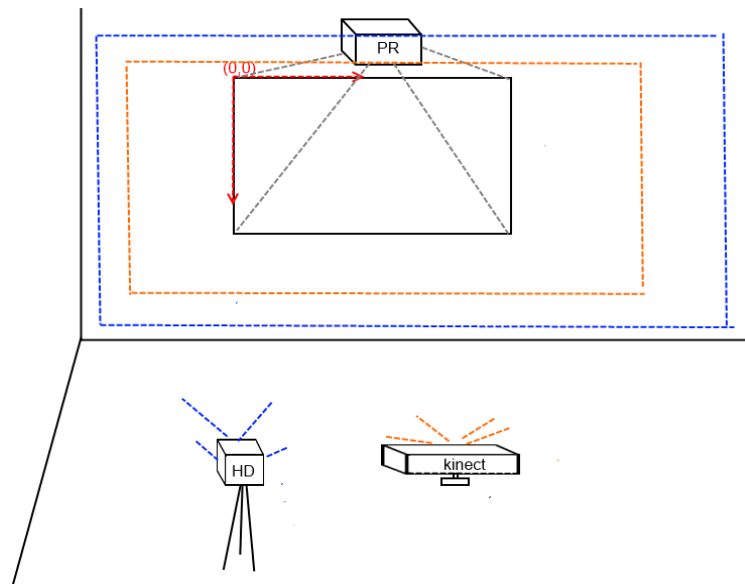


Figure 10 – Calibration

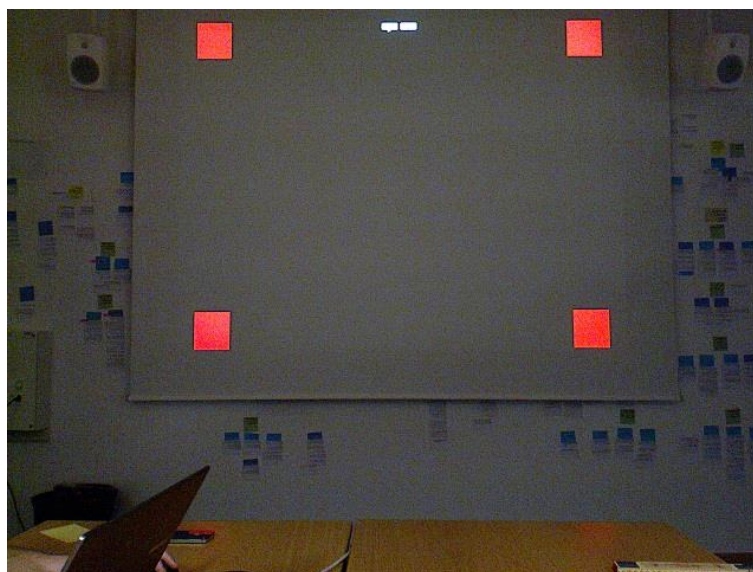


Figure 11 - Projected pattern viewed by Kinect

<b>Manually</b>	<ol style="list-style-type: none"> <li>1. Manual measurements of the distance offset between the sensor and the projected area.</li> <li>2. Manual alignment of the sensor at a pre-set position which should not change during the whole project duration.</li> </ol>
<b>Semi-automatically</b>	<ol style="list-style-type: none"> <li>1. Use the projector to project a pattern on the wall and then use the sensor depth data in order to calculate the distance from it. After calculating the distance, manually align the sensor so that the RGB frame from the sensor aligns with a pre-set position in software.</li> </ol>

	<ol style="list-style-type: none"> <li>2. Manually align the sensor to match with the top part of the projected area and use a person in order touch the wall where the bottom left and right rectangles are projected. Using the position of the right hand, we could calculate the position offsets of the sensor.</li> </ol>
<b>Automatically</b>	<ol style="list-style-type: none"> <li>1. Automatically calculate the distance from the wall using the depth stream and use shape recognition in order to automatically detect the red rectangles from the projected shape.</li> <li>2. Automatically detect a projected shape and its position and compare the detected size with the real size. Use the difference in size in order to detect the distance offsets.</li> </ol>

The evaluation of different envisioned methods was made starting with the automatic calibration methods. Because both methods rely on the automatic detection of a projected shape, different algorithms and libraries were tested (Emgu CV<sup>14</sup>, AForge .NET<sup>15</sup>). After testing both libraries, AForge.NET was chosen mainly because of the documentation, examples and because it was a native .NET library comparing to Emgu CV which is written in C++.

The shape identification algorithm was tested in different light conditions using natural and/or artificial light. The image to be processed consisted from the image rendered using the RGB stream from the Kinect sensor which was built from the projected image on the wall. Because of the unpredictable change in the brightness of the light sources and because of the presence of post-its on the wall, the identification of the red rectangles was unsuccessful in most of the cases. Because of this disadvantage, the automatic calibration methods were dropped.

Manual calibration was also dropped because of the discomfort in continuously manually measure the offsets between the projected area and the sensor.

Because it was desired to minimize the number of people involved in the calibration, but also to speed up the process, the second method of semi-automatic calibration was also dropped. In the end the first option of semi-automatic calibration was implemented. Tests were done for the calibration and the results are encouraging, but they depend on how well the user aligns the calibration rectangle with the projected shape.

## Summary

Because of the technical nature of the research question, the evaluation consisted mostly from analyzing existing research and existing libraries. In the end of the evaluation session the following results were identified:

## Technology

From a technological point of view, different SDKs were analyzed and in the end Kinect SDK was the chosen library.

<sup>14</sup> [http://www.emgu.com/wiki/index.php/Emgu\\_CV](http://www.emgu.com/wiki/index.php/Emgu_CV)

<sup>15</sup> <https://code.google.com/p/aforge/>



### Bounding box calculation

Different methods were identified, some using the bounding box concept, others using the body shape concept, in order to calculate the position and size of the people working on the affinity diagram. Even if we admit that the shape calculation concept can give better results when working with the affinity diagram because it specifies exactly which area the person occludes, the bounding box concept was implemented because no algorithm for the shape detection was found.

### Calibration

Different methods for manual, automatic or semi-automatic calibration between Kinect and the projected area were analyzed. A semi-automatic solution was chosen. The UI of this method is presented below.

### Prototype

The final prototype in the screenshot below presents just one part from the calibration process.

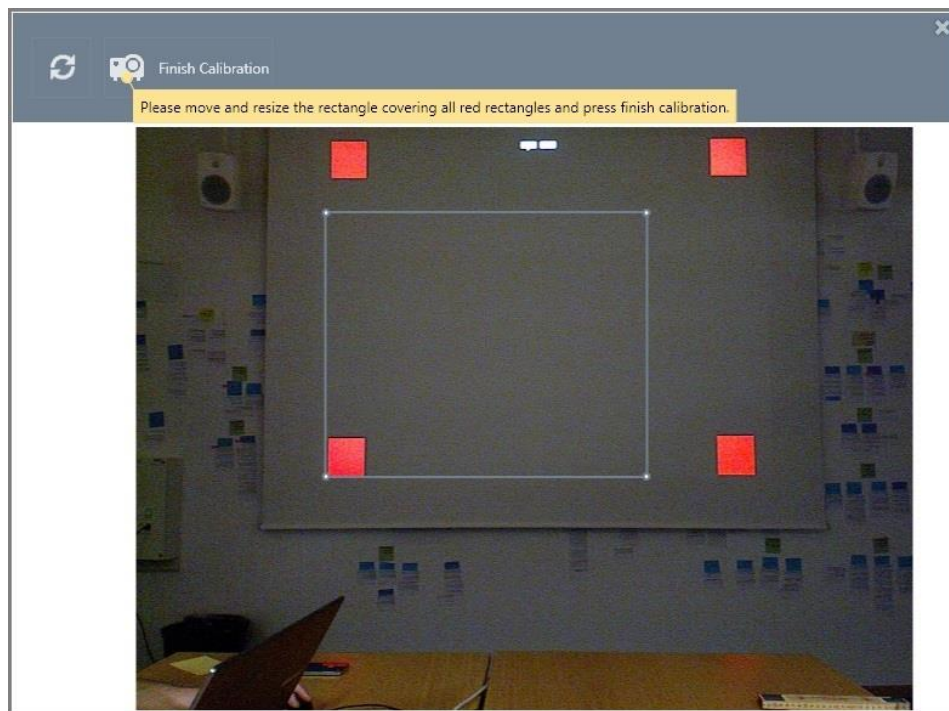


Figure 12 - Calibration UI

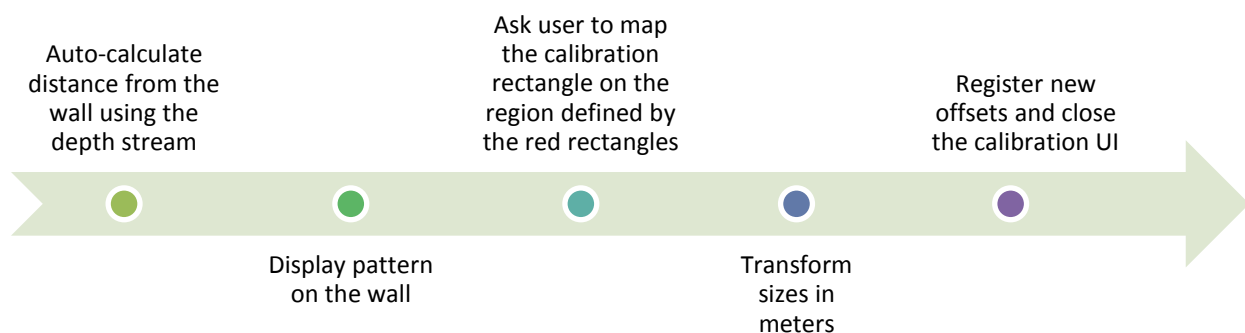


Figure 13 – Internal calibration process



## Research question 2

### ***How can we design and implement a user experience to support automatic or semi-automatic identification of users in the affinity diagramming context?***

When working with the affinity diagram different people may be present in the field of view of the camera. Each person may be a spectator or an actual project member, and they may come in or go out from the field of view. Part of the scope of the current thesis is to research the best way to automatically or semi-automatically detect the project members in order to store their actions under the right user in the database.

### Feasibility study

From the beginning of the research question, based on the envision setup, different boundaries were set for the future prototypes. These boundaries were related mainly to the type of interaction one can have with the system and only in the case of Kinect sensor. Because of the position of the sensor, which is behind the person, specific gestures can't be detected. Unfortunately, no research was found which resembles this particular arrangement of the sensor-person, therefore any design decision had to be tried out and discovered in prototypes and tests. In order to test which gestures can be used and implemented in prototypes a study was performed. The table below shows the results of the study.

<b>Grip</b>	Not working because the palm and fingers can't be recognized from behind
<b>Gesture libraries</b>	Different gesture libraries were tried and results are very poor because of the sensor position comparing to the human body
<b>Hand push</b>	Is working only if the hand is situated on the far right/left side of the body
<b>Hands</b>	Because the sensor is behind the person, the right and left hand are reversed
<b>Joints</b>	If the hand is not situated on the far right side of the body, elbow and wrist joints can be confused by the sensor
<b>Occlusion</b>	If gestures are made in a "normal hand position", the hand can be occluded by the human body
<b>Depth stream</b>	Because the body is reversed, the depth stream can't detect too well the body movements in front
<b>Movement</b>	Because of the sensor position and because of the projection on the wall, the movement on the right of the body will be detected as movement to the left; therefore the movement needs to be also reversed.

Taking into consideration the results of the study, two main prototype branches were made, one based on Kinect and one based on PC. The research consisted from three prototype sessions, each session containing a different set and number of prototypes.

## First prototyping session

This prototype session consisted of two prototypes, one based on Kinect and one based on a normal PC application concept.



Figure 14 - RQ2 / Session 1

The Kinect prototype was inspired from different Wii games and Sci-Fi movies where avatars are used to play different games. The prototype displays the members of the project as icons where each icon consists from the person head and his name. In the beginning, all members are aligned in a straight line on top of the projected area, the icons are decorated with a green border if the avatar has already a person associated with it, and a red border if not. An associated person is a person which is detected by the Kinect sensor and is mapped with a specific user (in the case of the prototype = avatar) from the database. When a person is mapped, the system will be able to associate all the actions done by that person on the affinity diagram with the specific user in the database. When a new person enters in the field of view of the sensor, all unassociated avatars (the ones with red border) will slowly float around the person's position. In that moment, the user needs to grab a head in order to complete the mapping process.

If in the first prototype the interaction was done using Kinect and the mapping process was done during affinity diagramming, the second prototype is a PC application. In this case the mapping is done after the affinity diagram session was completed and the interaction is similar to any PC application, by using the mouse and a keyboard. This prototype is composed from three regions; one hosting the video feed recorded using Kinect, one the main menu buttons including a list of project members and one containing a timeline with all identified people. This prototype was inspired from different movie making applications and follows the same interaction principles. In order to map a person with a database user, the administrator needs to go through the movie and drag and drop the right head over the bar which is represented by that person. Clicking on a bar from the timeline regions selects the corresponding person in the video feed. This representation gives a feeling on who worked in the project, between which time frames and how often.

## Evaluation

The evaluation of the presented prototypes was conducted as a think aloud method using HCI experts. Two people were used, each session being conducted independently. Each session was split in two parts of maximum 30 minutes discussion, each part focusing on one prototype. The table below displays the main important findings.

### Kinect prototype

- Too many terms related to the user can be ambiguous
  - Real user: the person making the affinity diagram
  - Database User: the virtual profile of each participant
  - User's skeleton: the virtual user identified by the Kinect sensor
- The skeleton drawn in the prototype was meant to describe the real user and not something which will be projected but it was drawn in order to infer the position in relation to the wall. This created much confusion for the participants.
- Floating head are considered a distraction
- A confirmation is needed in case someone would wrongly associate himself with another user
- A cancel button is needed
- The system should make proposals and ask "Who are you? Are you ...? "
- Menu items should be at relative position according to the person and not absolute in order to not disturb other people if one needs to press them.
- A visual feedback is needed.

### PC prototype

- Drag and drop is not intuitive for everyone, combo boxes are more intuitive
- "Tag" and "Approve Tag" directly on the video is more intuitive and requires less clicks
- If the video is too long then the administrator will not be able to watch it entirely. The timeline needs to be shortened.

Participants in general had only some problems using the Kinect prototype and this was mainly because of the new interaction principles. Besides this, problems in the designing of the prototype itself were noted down and planned to be mitigated in future prototypes.

### Second prototype session

Based on the prototypes from the first session four more prototypes were designed, two focusing on the PC concept and two focusing on the Kinect sensor.

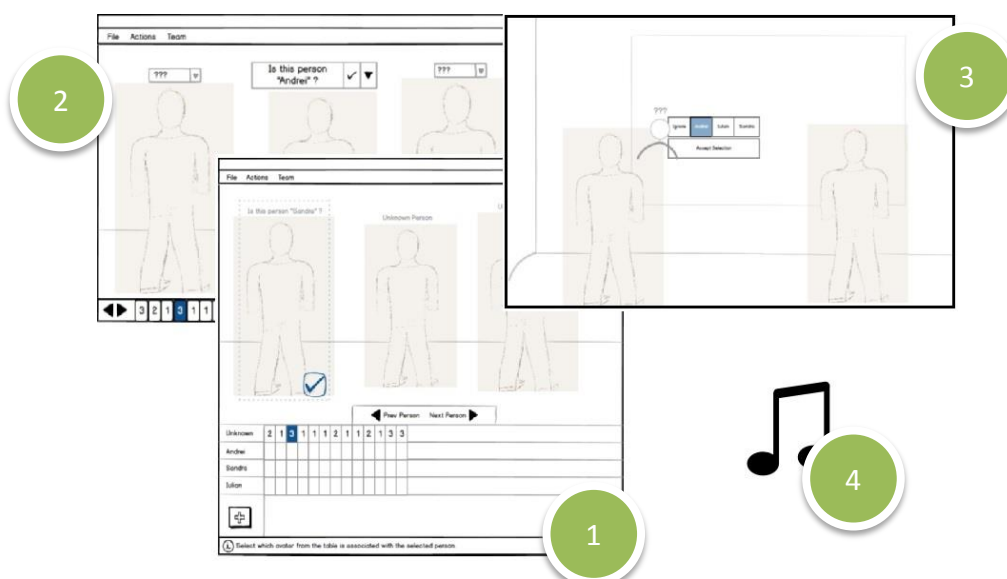


Figure 15 - RQ2 / Session 2

The first two PC prototypes (the ones from the left side of the RQ2/Session 2 Figure) are based on the previous PC prototype design, but adapted to the feedback received from the evaluation session. They both contain bigger image space, where the administrator can see in a better resolution the recording. Comparing to the previous prototype, instead of a video feed, the prototypes display only images from the moment when someone new comes in the field of view, in this way focusing more on the mapping process and not forcing the administrator to go through the whole video. When a new person is recognized by the sensor a frame is added to the bottom grid / list. The first line of the grid and the list itself display the number of people identified in that frame.

If from the point of view of visualization the two prototypes seem quite similar, the difference is in the way the mapping between the visualized person and the database user is made. In the first prototype (the one with the bottom grid) the administrator needs to use the arrows in order to cycle through the detected people and for each person select the cell associated to the right database user in the current column (the one with the blue background) in the grid.

The second PC prototype was inspired from the Facebook tagging feature and instead of using a grid; the administrator can directly chose the right user from the combo box located on the top of each person. In this case, the arrows are not used to navigate through people but instead to navigate through frames.

### ***Design decisions***

<u>Frames</u>	The movie bars were replaced by frames in order to shorten the space occupied in case of long videos, but also to not psychological stress the administrator with too much information
<u>Drop boxes</u>	These UI elements replace the drag and drop functionality which is not intuitive for everyone
<u>Image size</u>	Bigger image size in order to give better information about the recording
<u>Clear design</u>	Too many UI elements can make the mapping process unintuitive and confusing
<u>Visual feedback</u>	The prototype will display the mapping results in an obvious way (using names or symbols)

The last two prototypes (based on Kinect) are based partially on the last prototype session and partially on a new idea which was developed. The prototype resembling the prototype from the last session was slightly modified and the feedback from the last evaluation session was integrated. The main changes consist in the way the users are displayed and the way the person is selecting the right user. In this version of the prototype the users are displayed in the horizontal list, with the first item being "Ignore", for the case when the person does not want to map itself. The button underneath the user list is the apply button.

When a new person enters the field of view of the sensor, the list is displayed. If no action is taken in 3 seconds the list will disappear automatically. If the person selects a user and taps on "Apply" the selection UI will disappear and a virtual avatar containing the selected name will appear next to the person. The avatar will fade away automatically after 3 seconds.

## Design decisions

<u>Horizontal user list</u>	The list on top of the projected area and the floating heads were replaced with a normal and familiar horizontal list which is more intuitive
<u>Ignore and apply buttons</u>	Ignore and apply buttons were added in order to give more freedom to the person working on the affinity diagram
<u>Visual feedback</u>	The visual feedback was added in order to make the chosen option more obvious
<u>Localized interaction and visualization</u>	All interactions and visualizations are relative to the position of the person in order to not disturb the work of other people when affinity diagramming

The last prototype consisted from a sound based prototype. The envisioned system is an automatic mapping system and is composed from a mobile application and a PC application. When a new person enters the field of view of the sensor the PC application sends an instruction to all mobile phones of all unassociated users to emit their unique signal. If a signal is detected from the region in space where the detected person's skeleton is localized, than the system can map that user with the person. This is possible only because Kinect can detect the region in space from where a specific sound comes from, but it does not work well if multiple unassociated people are near to each other.

A feasibility study of this method was done in order to answer two important questions.

### How well can Kinect detect the region in space?

In order to find the answer for the first question a prototype was made in C# which can detect, using the Kinect SDK, the location of a sound sent by the phone. Unfortunately the SDK detects all sounds in the room and searching for a specific sound wave is not natively implemented and requires extra work.

### Do people have their phones in their pockets while working?

A small interview was done with 25 people. They were asked if they have their phones in their pockets at work. The interviews were done in 1 day and it consisted in asking people at school, work, at meetings with friends and over Facebook. The results are presented below.

**18 said NO** These people are not wearing the phones with them and this is mostly because the phone is too big, too inconvenient, gets warm or the cloths do not have pockets (in case of most of the girls/ladies). Usually the phone is located on the desk or in the bag.

**7 said YES** The people who said yes are mostly man.

Because of the low numbers from the interview and because of limitations in the Kinect SDK, this prototype was dropped.

## Evaluation

The evaluation of the first three prototypes was done with HCI experts as a cognitive walkthrough. The presented walkthrough can be found in appendix. 6 people participated in this evaluation. The table below presents the most important findings.

Kinect prototype
<ul style="list-style-type: none"> <li>• Delays and animations are hard to implement in a paper prototype and this created confusion.</li> <li>• The people must be facing the wall in order to observe the mapping UI. If they don't, they will lose the chance to do the mapping</li> <li>• The mapping should not be done by touching the wall because of the problems with shadows. Mapping from the distance is preferred.</li> <li>• Accept and ignore should be on the same row.</li> </ul>
PC prototype (with bottom grid)
<ul style="list-style-type: none"> <li>• The validate button should be next to the head and not on the bottom which is anyway not very observable</li> <li>• No way to "Disconfirm" if the system makes a proposal for a detected person</li> <li>• Click directly on the person and not use the arrows to navigate through them</li> <li>• No way to reassign users</li> <li>• The grid looks more like a summary.</li> <li>• No overall progress about the mapping process</li> </ul>
PC prototype (with bottom list)
<ul style="list-style-type: none"> <li>• No overall progress about the mapping process</li> <li>• Good familiarity because of similar tagging applications</li> </ul>

Overall, the Kinect prototype and the PC prototype with the bottom list proved to be very intuitive and to solve the given tasks.

## Final prototyping session

Based on the feedback from the previous session one final prototype was created incorporating both the Kinect and PC concepts. The figure below presents the envisioned mapping process.

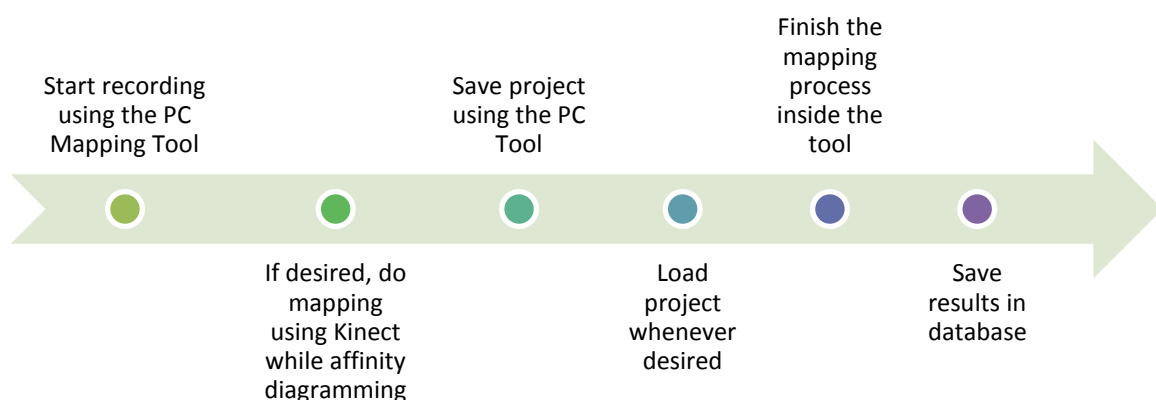


Figure 16 - Mapping process



Figure 17 - Kinect Mapping UI

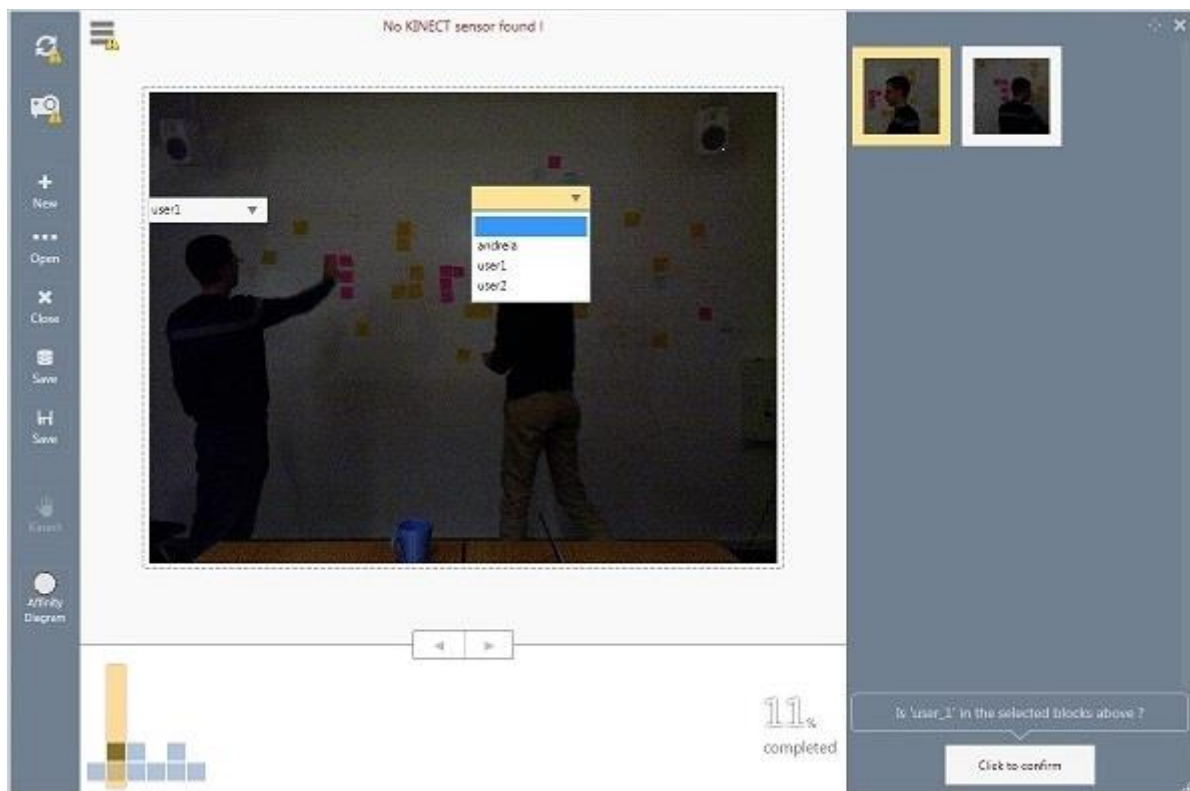


Figure 18 - PC Mapping Tool

Once a project is loaded, all users which were mapped using the Kinect approach are automatically loaded and configured. The combo boxes displays all database users registered for this project. The bottom bar represents as a chart the frames and the number of detected people. The administrator has the freedom to jump at any frame and map the person with the proper user. If the specific user is found in any or the other frames than the system proposes him in the right column. The administrator has the flexibility to continue the mapping process or to select the users and confirm the selection. This speeds up the mapping process, status which is presented at the bottom right side of the application. The administrator can at any time save the changes, close the application and resume when desired.

When doing the Kinect mapping, the person has the freedom to do the mapping from the distance or by tapping the wall. By moving the right hand, a virtual hand will be projected which can be used similar to a mouse.

### ***Design decisions***

**Combo boxes** Combo boxes are the chosen mapping UI element because it offers the list with available options, is familiar and offers the perfect way to reassign users.

<u><i>Propose users</i></u>	Speed is a big requirement because of the big length of the recording therefore proposing users is a must.
<u><i>Status</i></u>	Due to the large number of frames and users, having the status was necessary in order to give a feeling on how much was mapped and how much still need to be done.
<u><i>Flexibility</i></u>	The prototype was designed with the thought of giving the flexibility to the administrator and not forcing him to follow a specific process.
<u><i>Virtual hand</i></u>	In order to make things more natural, intuitive and flexible the virtual hand was added giving the possibility to do the mapping from the distance or from next to the wall.



### Research question 3

#### ***What user experiences can we enable with the user activity history and how can they be visualized?***

Only having information about a user does not help in any meaningful way if the results are not displayed properly and the interaction with the visualization is not adequate. The last part of the thesis aims to find the best way of rendering and interacting with these results.

#### **Feasibility study**

Taking into consideration the feasibility study related to Microsoft Kinect from the research question 2, the visualization and interaction were designed to take full advantage of the presented limitations.

Before starting with the prototype sessions a brainstorming session was conducted. The scope of this session was to generate as many ideas related to visualization and interaction as possible, but most important to answer some of the most important questions. These questions are presented below.

- *How to specify the time window and what does it consist from?*
- *How can I visualize the people who interacted with a region of the wall?*
- *How can I choose/select and scale a specific region of the wall?*
- *How far/near the wall should a person be in order to include him/her in the selection process?*
- *How can we visualize the amount of attention a specific region of wall have had?*
- *How can we indicate time when visualizing changes made on a specific region of wall?*
- *How can we represent user movement paths and their connection with the wall?*
- *How can we visualize the changes in notes?*
- *How can we select a note?*
- *How do we visualize the zoom out/in of the region of the wall in order to get a feeling of how big and where the note is in relation with the wall?*
- *How can we select the visualization type?*

The presented questions serve not only as a base for the feasibility study but are also meant to find out which features are most important to the people working on the affinity diagram. Gradually, different features related to different questions will be integrated in the prototype and presented in evaluation sessions to test users. The results from these sessions should indicate the desirability of specific functionality and guide the prototyping process to the final visualization. The results of the feasibility study are displayed in the appendix.

Because the visualization intends to display the changes in the diagram over time from the point of view of different people working on the affinity diagram, similar approaches can be enumerated. Among these approaches we can specify:

#### **Video recording**

Video recording of the wall is a basic and simple technique which offers all the information needed in order to find out what happened with a specific area of the wall. One can simply use a video camera,

record the affinity diagramming process and afterwards simply visualize the changes. Even if this approach seems sufficient for the presented requirements, some problems can be identified.

<u>Cost</u>	The video camera need to be placed at a relatively far distance from the wall in order to capture the entire area. Because of this, an expensive HD camera needs to be used.
<u>Quality of data</u>	A normal affinity diagramming process takes a few hours and is done over a time period of a few days. Because of this, the length and number of the videos can be quite big.
<u>Disk space</u>	Because of the big quantity of data the requirements of the disk space can be big and simple features as copy and share over the network can be problematic.
<u>Manual job</u>	Even if the above requirements are ignored, going through hours of recording is a very manual job.
<u>Speed</u>	Because of the manual nature in analyzing the data, the speed of finding the desired information is very low and the materials risk to never be used.

### **HD Photo camera**

Similar to the approach presented above, an HD photo camera can be used and programmed in order to take pictures of the wall at specific time intervals. The stored photos can be later compared automatically and a log or visualization showing the changes between photos could be created. Examples of applications which can be used to compare two photos are ImageMagick<sup>16</sup> and Photoshop. In Photoshop, images can be loaded in different layers and afterwards a subtractive effect on the layers could be applied to see the difference. ImageMagick can also be used using the command line, therefore Windows BAT files could be written in order to make the analysis more automatic. As the video recording approach, this approach is also limited and suffers from the same problems. Additional problems are described below.

<u>Loose important data</u>	Because the images are taken at specific time intervals, it is possible to miss small changes between time intervals, changes which can be crucial.
-----------------------------	---

### **Affinity diagram log**

A designated person can keep a log of all important changes during the affinity session. These changes could be aggregated in a Microsoft Word document. Simple features as find, heading styles and tables can be used in order to organize and find the data. This is a very manual process and valuable intermediate changes could be ignored.

The next prototype sessions uses, besides the previous research, all findings from the presented similar approaches.

---

<sup>16</sup> <http://www.imagemagick.org>

## First prototyping session

The first prototype session was designed in order to find out what types of visualizations were preferred by users. The prototypes tested also how easy users will understand the 3D and 2D sketches, what interaction approaches are intuitive and what prototype elements are preferred.

Three prototypes were designed, each focusing on different approaches. The prototypes are aggregated below and displayed in full form in appendix.

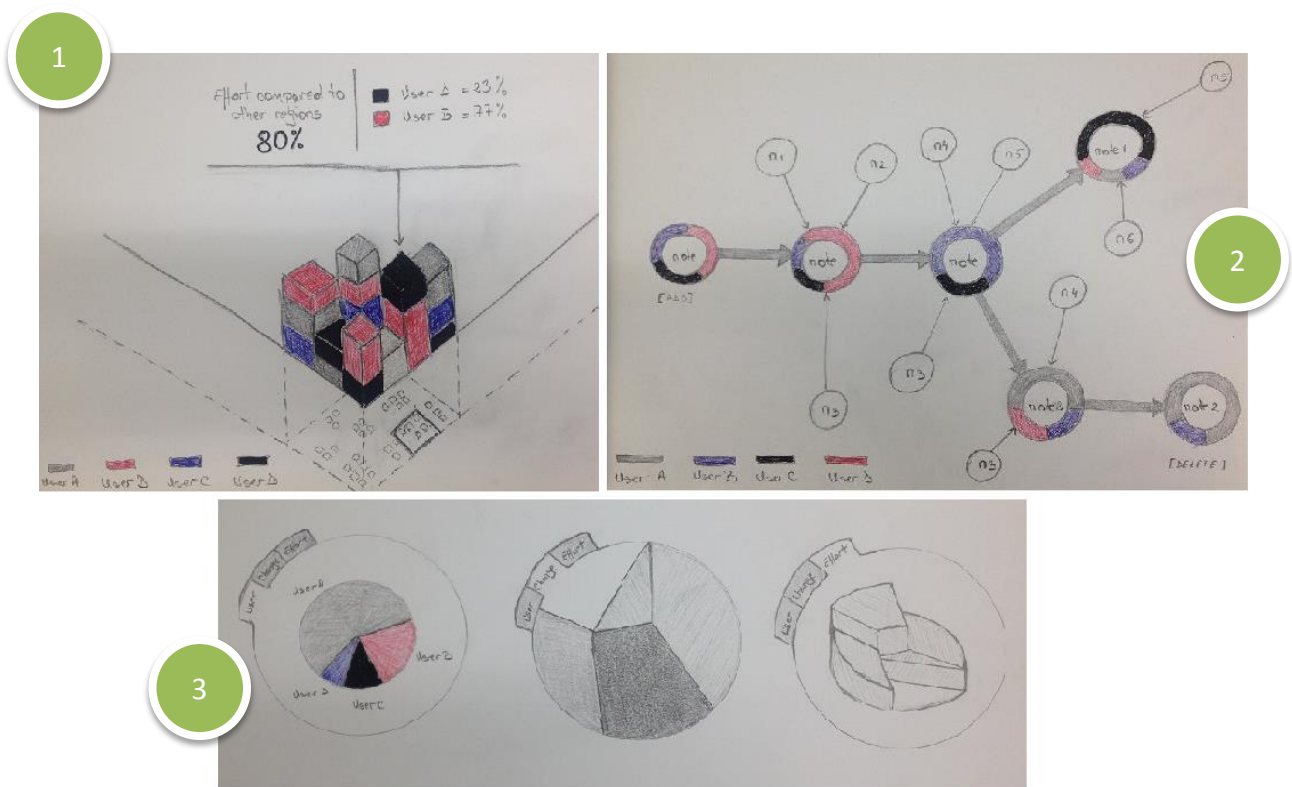


Figure 19 - RQ3 / Session 1

The first prototype uses a 3D approach to display the level of effort a person is putting on a specific region of the selected diagram. A snapshot of the diagram can be found on the bottom side of the prototype. The colors represent different users, users which are displayed on the left-bottom side of the prototype. Selecting a block will display extra information like the percentage of different colors and the percentage of the total block in comparison with the whole selected diagram. The interaction with the visualization is done using Kinect and normal gestures like zoom, rotate and tap are at the core of the prototype.

The second prototype displays the evolution of a note during the whole selected time frame. If the first visualization shows the effort on a selected part of the diagram, this prototype focusses on the individual notes and is displayed by simply selecting a note from the diagram. The main big circles represent the note itself and the changes it suffered. A ramification of a note consists in splitting a note in 2 different notes. The small circles are other notes which affected the evolution of the main note to that step while the colors represent the effort made by different users in taking the decision for the note to change to each stage.

The last prototype is represented by a magnifier glass with different layers. Normally, the person will only see one layer at the time but he will be able to switch between them. The magnifier glass can't be zoomed in or out but instead can be moved around to see details about specific regions.

- The first tab shows the amount of time different people put over time in that region
- The second tab displays the amount of change in a particular area (the bigger the change the darker the color)
- The last tab shows amount of effort people put in that region, displaying on the height the amount of change compared with the rest of the blocks.

### **Design decisions**

2D vs 3D It was desired to find out in the end of the prototype what type of visualization it is preferred and how easy it is understood. Besides this the type of interaction was also researched.

Full screen vs isolated display It was desired to find out if people prefer seeing the whole wall or prefer focusing on a specific region

Region vs Note It was desired to see if people prefer analyzing regions of the wall or individual elements.

### **Evaluation**

The evaluation was made using four people in a normal type of interview focusing on finding out the answer of the following questions.

- Do they understand what the visualization does?
- Do they understand how to interact with the visualization?
- Does this visualization offer the info they want?

Because the evaluation was conducted with external people (all software engineers having knowledge in diagramming) a consent form was created and signatures were gathered. Each interview took around 1 hour (even if the planned time was only 30 min). After the first interview was noted that more time was required (1 hour total) and people were asked if they are comfortable of having it for 1 hour. Interviews were audio recorded and transcribed afterwards. The results can be found in the appendix. The main results are presented below.

#### Things which were all the time clear

- The colors represent the users
- The pie chart is very easy to understand and it was understood by all
- Arrows are easy to understand and most often they represent evolution or history
- Normal gestures like zooming and rotating are easy for the ones which know about Kinect
- Most often when seeing charts people relate to Excel and Word.

#### Things which were not clear and are needed

- 3D shapes are hard to correlate with the wall
- Gradients are easy to understand if a legend is provided
- The interaction is not necessary obvious and it may be useful if hints or footnotes are provided
- The time frame needs to be displayed because it confuses people if is not shown
- Showing too many 3D layers confuse people and is hard to interact with them

## Second prototyping session

Taking into consideration the findings from the previous session other five prototypes were designed. After filtering out many of the inconveniences from the UI (like 3D shapes) these prototypes were designed in order to find out how the data is better represented from the point of view of the HCI experts. The prototypes used for this session are displayed below.

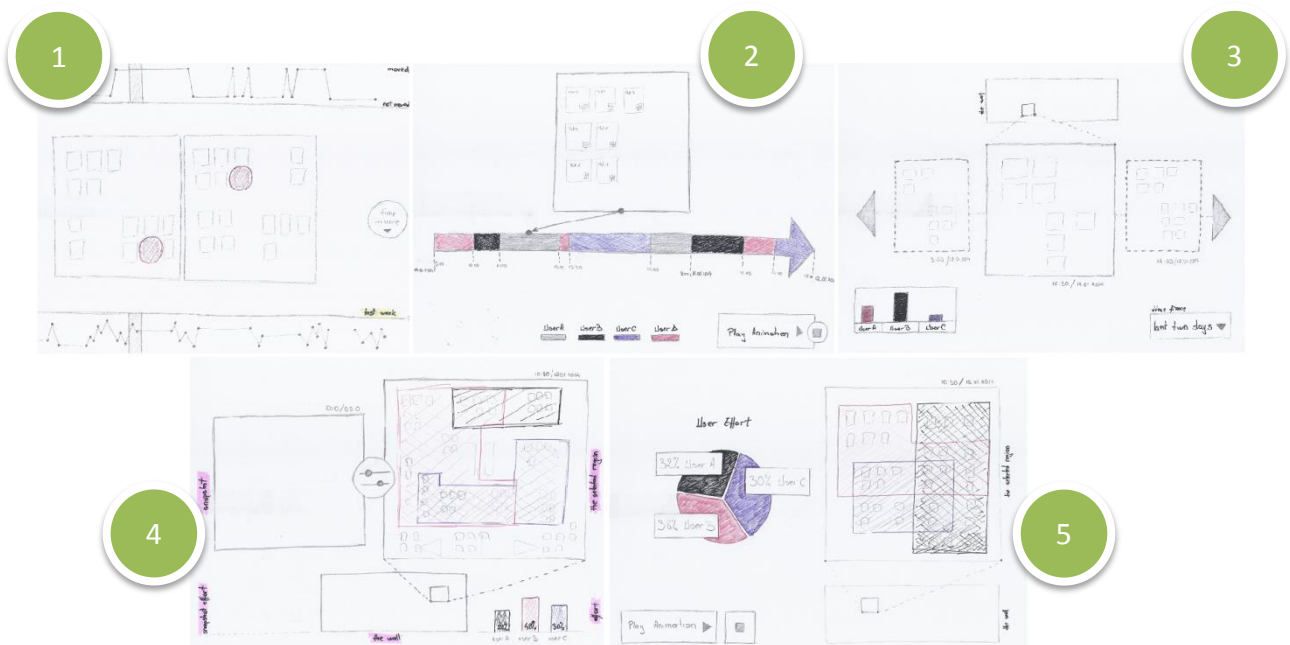


Figure 20 - RQ3 / Session 2

Prototypes 1, 3 and 4 focus on the idea of manually comparing two or more snapshots from the affinity diagram from different moments in time. The motivation of this approach is that a person can easily enough detect changes in notes if the selected area is small enough. Beside the normal screenshot regions, people are helped with different other visualizations which help them focus on particular snapshots.

Prototype 1 uses the bottom bar in order to display the effort (as a chart) put in the selected region over a specific period of time. The top part of the prototype displays the evolution of a selected note (the red circle) in relation with the selected time frame.

Prototypes 3 and 4 additionally displays charts indicating the level of effort different users put in the selected area and sometimes even highlighting the areas which they affected (prototype 4 and 5).

Prototypes 3, 4 and 5 use the panning concept in order to display the location of the selected area comparing with the entire wall. Beside the panning tool, some prototypes have a “Play Animation” button which trigger the prototype to automatically go through snapshots and highlight the user affected areas in a form of an animation.

Prototype 2 uses an arrow in order to display the main user which worked on the selected region of the wall in that time frame. Selecting a block will animate the change in the snapshot.

### ***Design decisions***

<u>2D</u>	As found out from the previous prototype stage 2D shapes are better understood and people relate faster to them because they seem more natural.
<u>Charts</u>	Charts are always easy to understand because of the familiarity with other applications.
<u>Snapshots</u>	Snapshots can be an easy way to compare data. According to the last session people would like to see the history of the diagram itself.
<u>Animations</u>	Using animations in order to show historical data was proposed by users in the last prototyping session.
<u>Heat-map</u>	Heat-map was a visualization form proposed by both brainstorming session and past prototypes.
<u>Legend</u>	Because of the heat-map colors which can be misinterpreted a legend is a must.

### ***Evaluation***

For these prototypes the evaluation was conducted with 6 people, all HCI experts. A heuristic evaluation session was conducted, focusing on the following four heuristics.

- Visibility of system status.
- Recognition rather than recall
- Minimalistic design
- Familiar Metaphors

These heuristics were particular chosen in order to test the familiarity, clean design and recognition of UI elements. The evaluation took place 45 minutes and was followed by an open discussion (around 30 min) where opinions were debated and more feedback was given. The evaluation form and a complete list of the findings can be found in the appendix. The main results are presented below.

- Concepts like “user effort”, “note” and “wall” are hard to understand only from a label.
- The link between graphs and other elements should be obvious.
- The UI should be obvious to the user and little instructions of how to use it should be given.
- The UI should be clean and clear without too many graphical elements which distract the user.
- The UI shouldn’t have too many overlaying elements.
- The UI should adapt to the current situation.
- Interactive elements should be position so that they are easily accessible to the user.
- Interactive elements should be obvious.
- Provide a way to see the overview of the whole wall.
- Don’t force the user to do things manually.

### Third prototyping session

This prototyping session contains only one prototype and it combines the findings of all other prototypes. The visualization offers two modes, one focused (by using a circle) and one as a whole wall which is triggered if the circle is resized to the maximum. The border of the circle represents a track bar, which can be resized in two directions, in this way setting the beginning and the end of the displayed time frame from the whole time frame. The inside of the circle displays the aggregated heat map of all users and by using opacity and transparency different areas where more changes occurred can be identified. The circle can be moved freely on any position of the wall but in case the selected position is outside the projected area, the circle will be resized to a default size, be placed on the right or left side of the projected area (depending where the selection was made) and an arrow will display the direction where the selection was made. If the selection is in the projected area no arrow will be displayed.

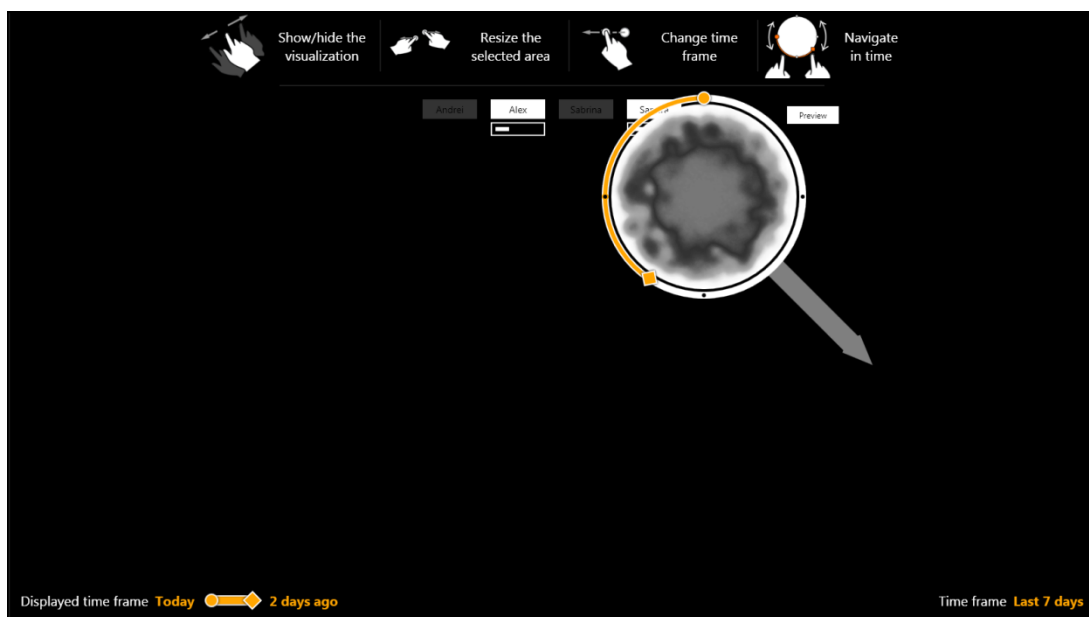


Figure 21 - RQ3 / Session 3

While navigating through time the progress bars underneath each user will display the amount of change each one did in comparison with the other in the displayed interval. If a person did not have any activity during the displayed interval they will be displayed more transparent. On the bottom of the visualization, different labels are strategically placed in order to display the information about the time frames. The top of the visualization gives a hint of available gestures, gestures which can be used to control the visualization. A list of available gestures is displayed below.

<u>Wave</u>	Hide the visualization
<u>Resize using two hands</u>	Changes the size of the circle. Shows the heat map of the whole wall in case the circle is resized to the maximum.
<u>Move hand horizontal from right to left</u>	Changes the total time frame (not the displayed one).
<u>Circular rotations with the right and left hand</u>	Changes the beginning (left hand) and the end (right hand) of the displayed time frame.

## ***Design decisions***

### **Clean design**

The design should be simple and not use too many visual elements. Each element should have a purpose. The visualization will be projected on top of the existing affinity diagram therefore the chosen colors should not interfere with the affinity diagram.

### **Familiar metaphors**

A circle was chosen as the default representation of the selected area. Circles are often used in practice when needed to see better or more details about a specific item or area (microscope, telescope, binoculars, glasses, magnifier glass ...).

### **Familiar interaction**

Specific gestures commonly used in Kinect games and Sci-Fi movies were used in order control the visualization.

### **Smart labels**

Dynamic labels were used in order to display information about selected time frame and displayed time frame.

## ***Evaluation***

Similar to the last evaluation, a heuristic evaluation was used, with 6 participants, in order to detect the same heuristics presented in the last prototype session. These heuristics are presented again below.

1. Visibility of system status.
2. Recognition and recall
3. Minimalistic design
4. Familiar Metaphors

The evaluation session lasted for 1 hour and 15 minutes and was composed from 45 minutes individual evaluation (written) followed by an open discussion.

The following findings can be mentioned.

- Gesture images are not well designed. More effort needs to be invested in creating better icons in order to make the interaction more intuitive.
- "Preview" label which appears next to the arrow is unclear and it should be removed.
- Link between the heat-map and the users is unclear.
- Heat-map should be colorful and not black/white.

## ***Final prototyping session***

The final prototype visualization is similar to the one from previous prototype but with small corrections, corrections made according to the findings from the last heuristic evaluation session. The UI is represented in the figure below. Comparing with the last prototype, the following features were added or changed.



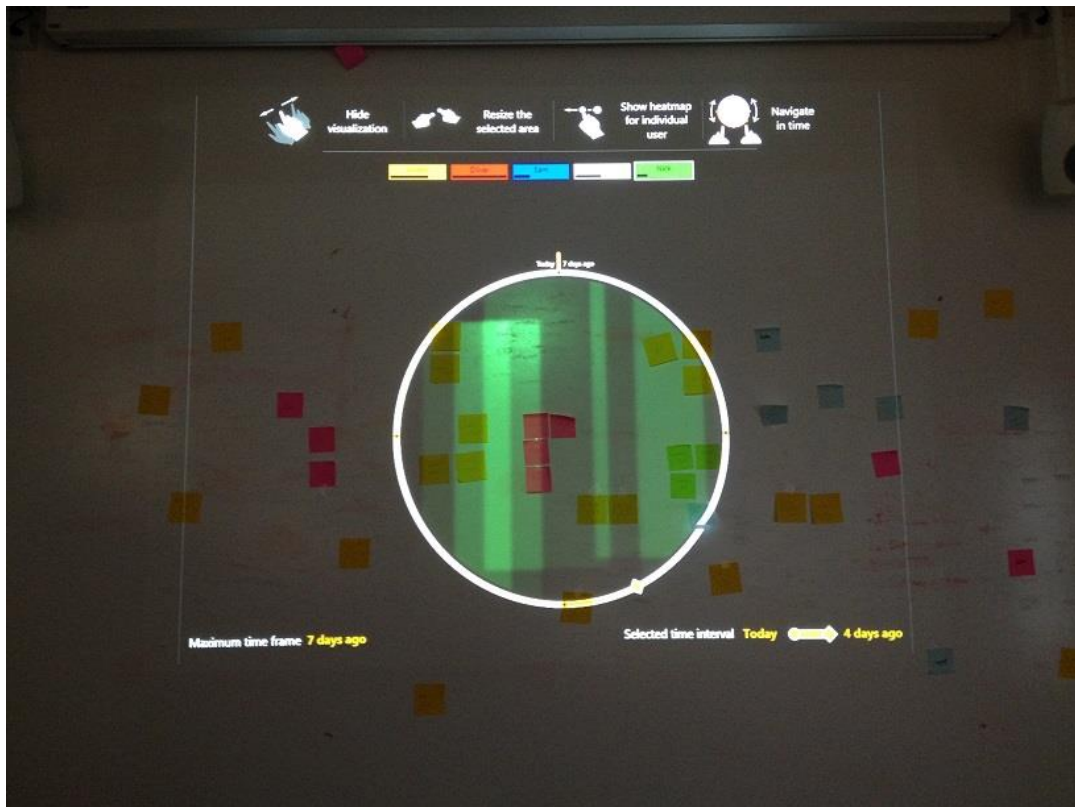


Figure 22 - RQ3 / Final prototype

### ***Design decisions***

#### ***Displayed time frame***

The track bar displaying the time frame was modified in order to change only the end point in time. According to the open discussion from the last heuristic evaluation it was commonly agreed by the HCI experts that changing the beginning of the point in time is not needed and it does not bring any value. The visualization should always present from the current point in time to the selected one.

#### ***Colors***

Colors were added to the heat-map and to the user boxes in order to easily connect the changes done by each person. These colors were specially selected in order to give the best results when creating the heat-map.

#### ***Heat-map***

The heat-map can now display either the aggregated results from all people or (as shown in the example above) display the heat-map of a selected user. In order to select a user the horizontal movement of the hand was redesigned in order to give the possibility to cycle through the available users.

#### ***Labels***

Labels were changed and new labels were added (top of the circle) in order to give better information about the status of the visualization.

## Evaluation

The evaluation method used for the final prototype is a functional prototype test. Seven people (other than the ones interviewed until this moment) were asked to answer specific questions and to solve specific tasks. Each interview lasted between 30 and 45 minutes and tested the real implementation of the prototype. People used for the test belonged to four different groups: HCI experts, students, researchers and business people (with engineering knowledge). Before the start of the interview an introduction (according to the protocol which can be found in the appendix) was made. Each person received a small introduction related to the affinity diagramming, used devices and a scenario was presented. This scenario is presented below.

*Suppose that you (user named “Sam”) are an HCI expert and you are currently working together with your other 4 colleagues on an affinity diagram for an internal project. The diagram is located on a wall and towards facing the wall are a projector and a Microsoft Kinect sensor. The affinity diagram (the notes on the wall) was started 7 days ago and you usually work a few hours per day on it. During the last 2 days you were in vacation and as your colleagues continued to work on the diagram, now it looks completely different compared with as you left it. Because you are unsure on one hand about the changes which were done during the last days and on the other hand about who did specific changes, you are using the UAV system in order to visualize all this information. Once the system is on, the visualization will be projected on the wall (overlapping partially the affinity diagram). The interaction with the system is done NOT using a computer but instead, using natural hand gestures which are interpreted by Kinect. Because Kinect is behind the person, any gesture needs to be clearly executed with the right hand, the hand being situated on the far right side of the body. In order to start the system one needs to only “wave the hand”.*

The main areas which were evaluated are:

- Can the person understand the gestures which he can use in this visualization?
- Does the user observe and understand all the displayed information (labels, shapes and heat-map)?
- For the given scenario, how does the user find out what happen in the selected area and what steps is he making towards achieving this goal?

The results and main observations are displayed below.

### Gestures

- Because of the wrong icons people do not understand the meaning of the gestures in the beginning.
- People tend to tap on the wall and not necessary use the gestures from the distance as a normal Kinect application would work (this needs to be learned).
- The list of gestures on the top of the projected area was not all the time being noticed.
- Even after the correct gestures were explained, people still continued to go at the wall and interact with the visualization as they would do with a normal computer and not using Kinect gestures.

## Information

- The labels on the bottom of the projected area were not all the time being noticed.
- Both focused and full-screen is needed.

*"I don't want to be distracted by regions which I don't care but in the beginning I may want to have an overview of the whole wall."*

*"My eyes would get tired because of too many colors and I would not use the visualization too much if I have to only use the full screen mode."*

- Cycling through people is not very intuitive; almost all participants wanted a way of selecting the desired user.
- Depending on the light conditions in the room, the heat-map can be partially hidden.
- The orange color of the track-bar situated on the border of the circle should be changed. One user believed that the orange color was connected with the orange color of a user.

## Solve the given task

- People used all given features in order to solve the task (move and resize the circle, change time frame, cycle through people and analyze differences, switch between the wall and full-screen visualization).

## Arrow

- Some problems in understand the purpose of the arrow. Some people understood it from the beginning but others needed more time to think. A label or some explanations may be required. In one case, it happened that the arrow was pointing to a specific note; this made the person believe that the visualization was somehow connected with that note.

## 6.5 Observations

Since the release of Kinect SDK to the developers and researchers in 2011 many Kinect related applications, research papers and projects appeared and this is mainly because of the flexibility, because of the natural interactions and new frontiers, and most important because of the freedom which Kinect offers it terms of imagination. For many, probably Kinect is the device which can make all those Sci-Fi movies partially come to life. In terms of what's under the hood, Kinect offers the perfect amount of data and at an affordable cost.

Even if Kinect offers plenty of freedom in terms of what you can do and how you can interact, slowly you find yourself restricted, not by the technology, but by what will, the users of your future system, be able to imagine. Unfortunate when working with new systems and new interactions, people tend to be resilient and tend to run back to what is proven to work. Because Kinect applications are not yet so popular outside games and because most of the people which use computers interact with it in a normal and old fashion way, it is hard to introduce a new interaction to someone, especially in the case when that someone is not a gamer or a Sci-Fi fan. Psychologically, people tend to stick to what they know best and is proven it works, and they would not change unless clear advantages are obvious (as found from different interviews).

*In one of the interviews during the final evaluation, a manager from an external company said that he will only introduce this kind of system in his company only if the advantages are so big that the shift in mentality for the people working for him is really worth it.*

At first, when one starts developing for Kinect he does not realize the shift in mentality which he needs to make. Being used with normal PC applications one may think that using Kinect means replacing the mouse with a virtual hand and maybe adding some hand gestures, but exactly as one needs a different mentality to create web or mobile applications comparing with PC applications because of differences in screen sizes and asynchronous nature, also Kinect needs a different mentality, where distance from the wall, calibration with the wall and projector, position of the sensor in comparison with the person (in front or behind), color of the wall, light intensity and multi-user need to be taken into consideration when designing the interaction and visualization of the application. Much of the research is currently focusing on controlling different hardware devices (like robots) using Kinect but this is only one use. Other parts of the research focus on analyzing the interaction but by using a normal screen as a visualization device. The current research and the current thesis took a different approach where Kinect is fully integrated, an approach envisioned already by the ubiquitous computing where the devices are part of our working environment. What makes this research different (and this is described in the current thesis) is that by simply using a projector to project the visualization instead of using a normal computer, the visualization itself is not adequate anymore and needs to be almost totally redesigned taking into consideration the environment in which is introduced. Because of this, visualizations become more than simple shapes and text; they become an extension of the environment itself.

As any software application, Kinect applications are better designed for a specific group of people and visualizations need to be tailored for those users. As any new system, a learning curve will always be necessary and as any new territory, things can also be discovered through trial and error.

## 7. Summary and Conclusions

### 7.1 Findings

Many software applications exist for creating virtual walls with affinity diagrams, but even if they offer different advantages like search, filtering, smart tagging and so on, people still prefer creating affinity diagrams on a real wall using post-its. Usually the reasons for this are very simple, post-its are more flexible and easier to use, do not require such a big extra cost and more important they offer the possibility to have an overview of the whole wall due to its bigger size. As stated before, the thesis is part of a larger project which aims to augment the physical world with virtual elements which can prove to be a very good compromise from both approaches. Specific to this case, in the field of HCI, the project goal is to research ways of automatically or semi-automatically detect the people which are working on the affinity diagram and store their actions in order to be later processed and visualized in a meaningful way.

In order to detect the people working on the affinity diagram, a Microsoft Kinect sensor was used. The sensor provides skeleton and depth information which can be later used to derive the actions and position in a 3D space.

More specifically, the thesis researched the following three research questions.

1. *How can we track multiple users' presence and activity using Kinect while affinity diagramming, and how would users set up this component as part of the envisioned system?*

Different SDKs were analyzed and Kinect SDK was the chosen library. Different methods for calculating the bounding-box were analyzed and taking into consideration already existing research and the possibilities provided by Kinect SDK, an estimation approach based on skeleton information was implemented. From the point of view of the sensor calibration, different automatic and semi-automatic methods were researched and in the end a semi-automatic method was chosen where the administrator needs to manually align a "mapping rectangle" over an image displaying a projected pattern on the wall.

2. *How can we design and implement a user experience to support semi-automatic identification of users in the affinity diagramming context?*

Different prototypes were tested. The most important findings are presented below:

- Special care needs to be taken when working with the Kinect because the sensor is located behind the person; some gestures will not work properly and some data (mostly related to joint information) will not be accurate.
- When doing a semi-automatic mapping speed is essential, therefore recommending people and just validating the proposals is an important requirement.
- From all researched methods, a simple combo-box combined with an image tagging approach is the best solution in mapping people.
- Because of the big number in frames which need to be considered, showing the status of the mapping process and having the flexibility in when to do the mapping is an important requirement.

- Kinect mapping should be kept optional and it should not force the people to get distracted from affinity diagramming.

### 3. *What user experiences can we enable with the user activity history and how can they be visualized?*

Different prototypes were tested. The most important findings are presented below:

- 2D shapes and visualizations are preferred over the 3D ones
- Displaying snapshots is not recommended because of the interferences with the other notes on the wall.
- Manual analysis of two or more snapshots is not perceived well by test users.
- Even if is obvious for users that Kinect is the way one interacts with the visualization, they tend to interact by tapping on the wall and not by making Kinect related hand gestures.
- Much focus needs to be put in designing icons in order to make the gestures understandable.
- Artificial light is preferred when working with visualizations projected on the wall because the natural light changes too much and interferes with the displayed colors (especially when heat-maps are used).

## 7.2 Future work

Throughout the paper, different limitations and potentials for future work were proposed. Some of the most important areas are presented below.

### **Algorithms**

Better algorithms for detecting the bounding box or the body shape can be implemented which can increase the performance of the prototype.

### **Better images and icons**

Images and icons play an important role in the visualization, especially when describing Kinect gestures.

### **Natural light**

Using only artificial light is a limitation of the prototype. Researching better colors and ways of limiting the effect of brightness changes can improve the performance of the visualization.

### **User recognition**

When proposing users in the mapping process, face recognition, voice recognition, body shape recognition or other means can be used to better propose and make the mapping faster and easier.

## 7.3 Conclusions

The tests done on the technical part and on the two final prototypes show that the proposed visualizations and interactions can help in mapping people with database users and visualize the collected data. The given use case and tasks from the final prototype test show that even with a small learning curve people can use the system in order to answer different questions related to affinity diagrams. Even if the visualizations do not provide the answer to some questions it does help in finding out **what change, when and who are the people involved in those changes.**

## Bibliography

- Alahi, P. A. (n.d.). *Connecting Kinects for Group Surveillance*. Retrieved November 2013, from <http://actu.epfl.ch/news/connecting-kinects-for-group-surveillance/>
- Baudisch, P., & Rosenholtz, R. (2013). *Halo: a Technique for Visualizing Off-Screen Locations*. Florida, USA: Interaction Techniques for Constrained Displays.
- Chattopadhyay, T., & Roy, S. (2013). Human Localization at Home Using Kinect. *HomeSys 2013*, 821-828.
- Danino, N. (2001, September 3). *Heuristic Evaluation – a Step By Step Guide Article*. Retrieved January 2014, from <http://www.sitepoint.com/heuristic-evaluation-guide/>
- Doksam, G., & Gunnar, H. (2012). *Where's the note? A multimodal system to augment affinity diagrams*. University of Zurich.
- Gulshan, V., Lempitsky, V., & Zisserman, A. (n.d.). *Humanising GrabCut: Learning to segment humans using the Kinect*. UK: University of Oxford.
- Gulshan, V., Lempitsky, V., & Zisserman, A. (n.d.). *Humanising GrabCut: Learning to segment humans using the Kinect*.
- Gustafson, S., Baudisch, P., Gutwin, C., & Irani, P. (2008). *Wedge: Clutter-Free Visualization of Off-Screen Locations*. Florence, Italy: CHI 2008.
- Harboe, G., Minke, J., & Huang, E. (2012). Computer support for collaborative data analysis: Augmenting paper affinity diagrams. *Computer Supported Cooperative Work* (pp. 1179-1182). New York, USA: ACM.
- Hsieh, C.-T., Wang, H.-C., Wu, Y.-K., Chang, L.-C., & Kuo, T.-K. (2012). *A Kinect-Based People-flow Counting System*. ISPACS 2012).
- Human Behaviour in Group Settings. (n.d.).
- Judge, T., & P.S., P. (2008). Using multiple display environments for Affinity Diagramming. *ACM, Computing Systems*, 9-12.
- Judge, T., Pyla, P., McCrickard, D., & Harrison, S. (2008). *Using Multiple Display Environments for Affinity Diagramming*. Retrieved from <http://workshops.fxpal.com/cscw2008/submissions/tmp6F.pdf>.
- Klemmer, S. (n.d.). *Human-Computer Interaction*. Retrieved from <https://class.coursera.org/hci/lecture>
- Li, B. Y., Mian, A. S., Liu, W., & Krishna, A. (2013). Using Kinect for Face Recognition Under Varying Poses, Expressions, Illumination and Disguise. *IEEE 2013*, 186-192.
- Microsoft. (2013). *Human Interface Guidelines v1.7*. Microsoft.



Nambiar, A. M., Correia, P., & Soares, L. D. (n.d.). Frontal Gait Recognition Combining 2D and 3D Data. 145-150.

Preis, J., Kessel, M., & Werner, M. (n.d.). Gait Recognition with Kinect.

Ryan, D. J. (n.d.). *Finger and gesture recognition with Microsoft Kinect*. Stavanger, Norway: University of Stavanger (Department of Electrical and Computer Engineering).

Sinthanayothin, C., Wongwaen, N., & Bholsithi, W. (2012). Skeleton Tracking using Kinect Sensor & Displaying in 3D Virtual Scene. *International Journal of Advancements in Computing Technology*, 213-223.

Xia, L., Chen, C.-C., & Aggarwal, J. K. (n.d.). *Human Detection Using Depth Information by Kinect*. Texas: The University of Texas at Austin.

## Abbreviations

<b>MVVM</b>	Model view view-model
<b>XAML</b>	XAML is a declarative markup language. As applied to the .NET Framework programming model, XAML simplifies creating a UI for a .NET Framework application. <sup>17</sup>
<b>API</b>	Application programming interface
<b>SDK</b>	Software development kit
<b>Windows BAT</b>	Scripting file used in Windows in order to execute different commands

---

<sup>17</sup> [http://msdn.microsoft.com/en-us/library/ms752059\(v=vs.110\).aspx#what\\_is\\_xaml](http://msdn.microsoft.com/en-us/library/ms752059(v=vs.110).aspx#what_is_xaml)

## Appendix

### A. Database Schema

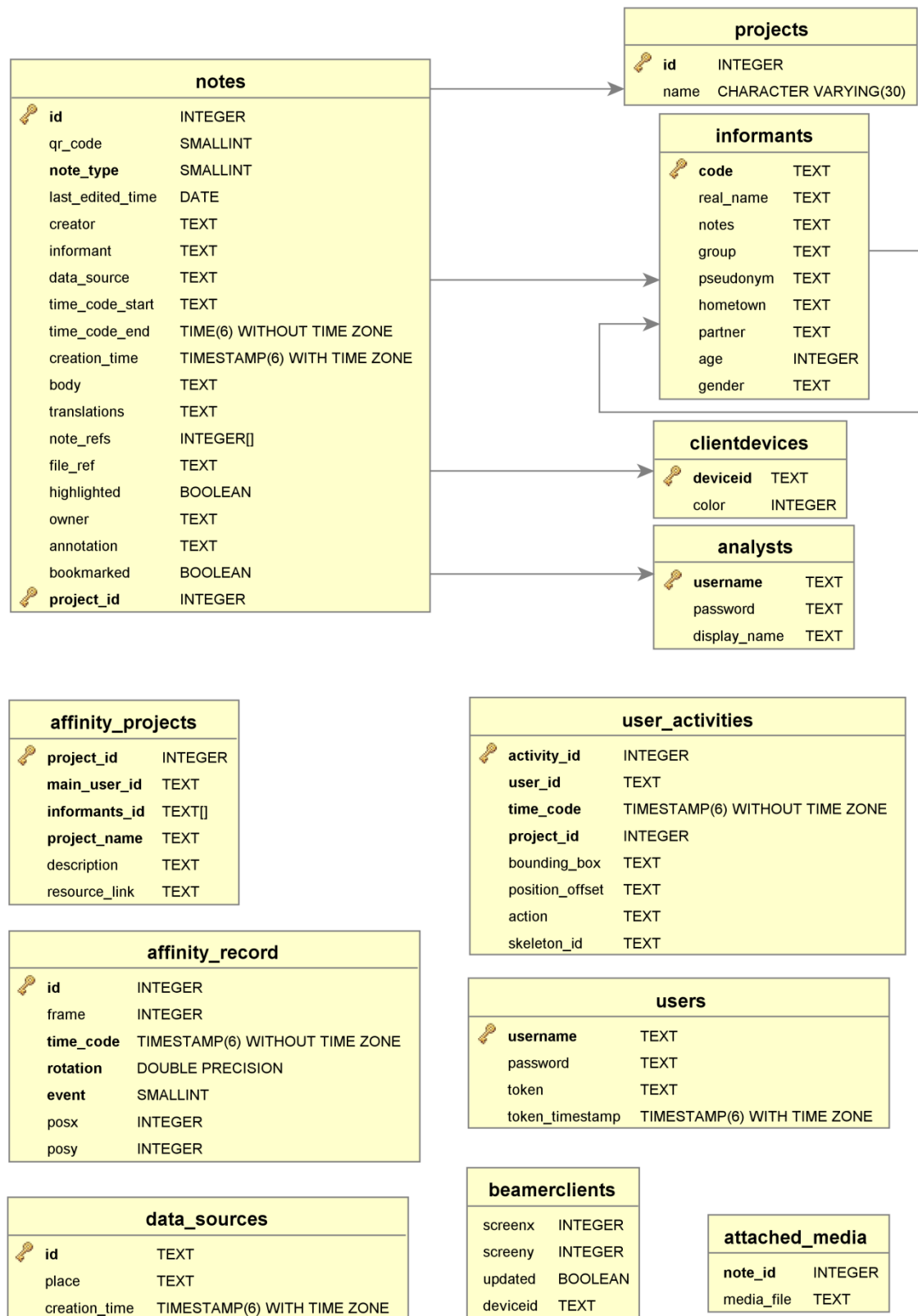


Figure 23 - Database Schema

## B. Software Engineering

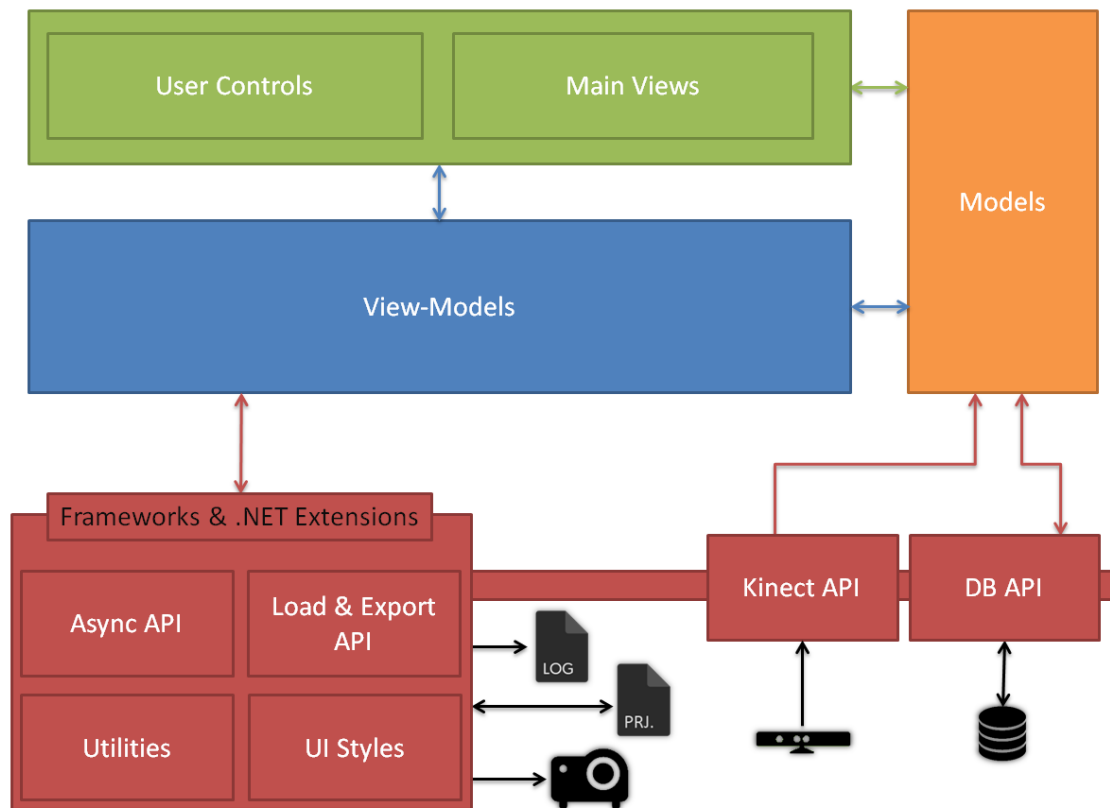


Figure 24 - System Diagram

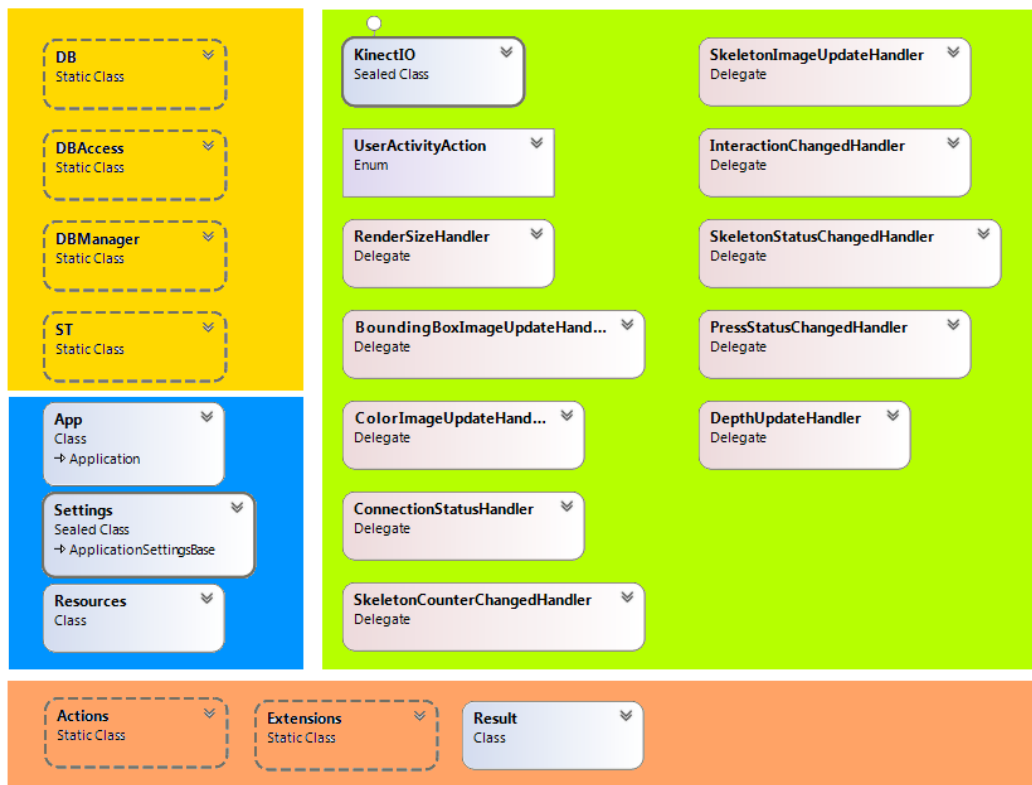


Figure 25 - API Class Diagram

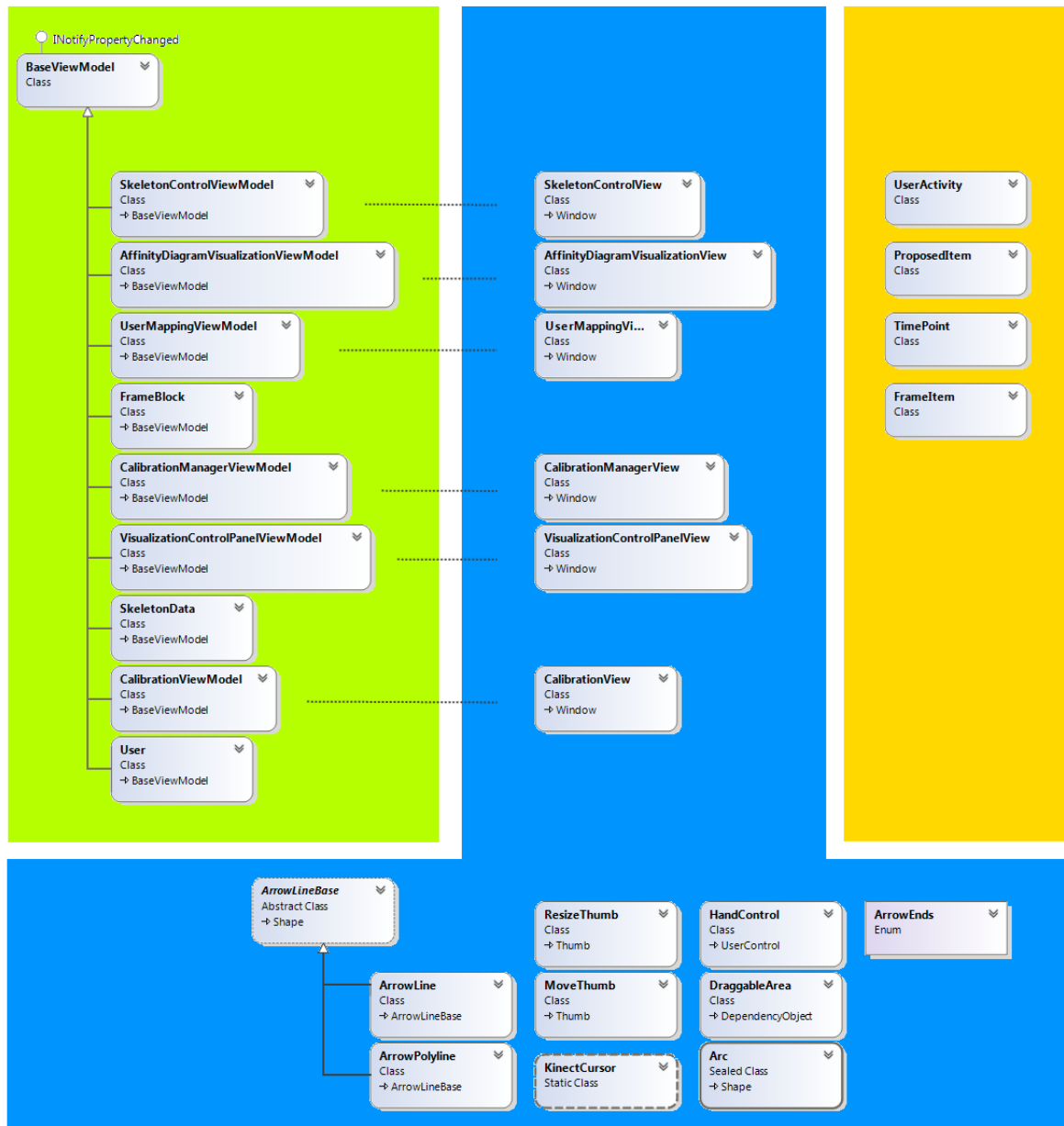


Figure 26 - Application Class Diagram (M-V-VM)

## C. Research Question 1

### Calibration

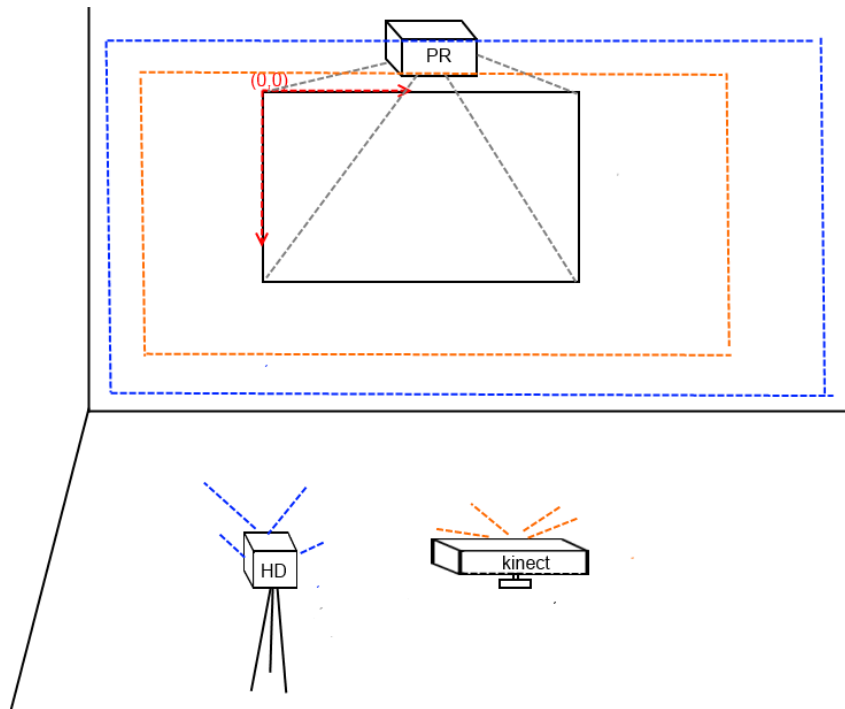


Figure 27 - Environment & devices

### Important Images

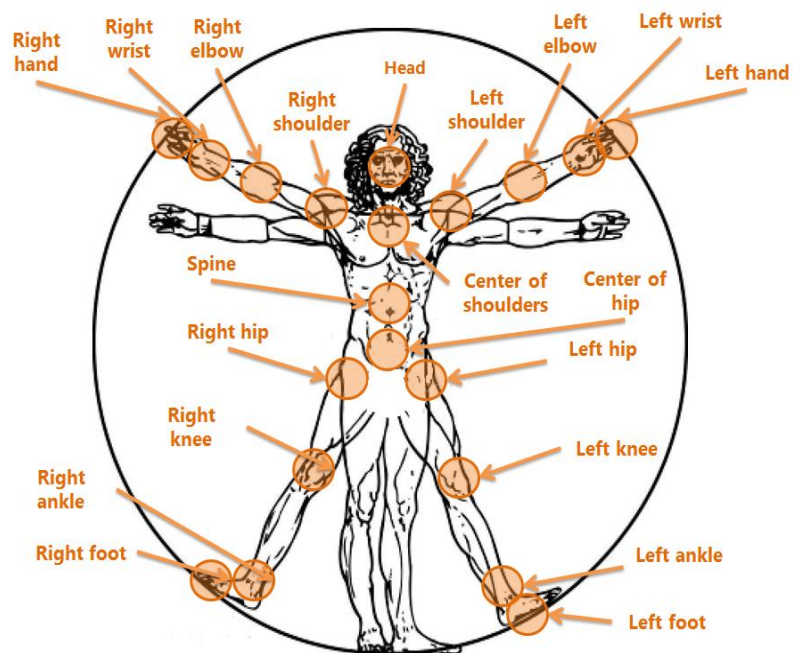


Figure 28 - Kinect Skeleton<sup>18</sup>

<sup>18</sup> Copyright <http://praveenitech.wordpress.com/2012/01/04/35/>

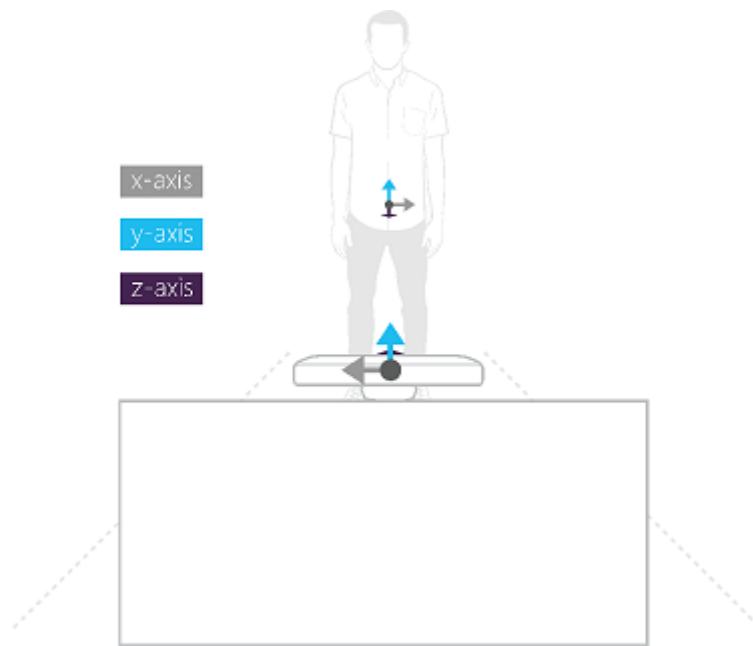


Figure 29 - Joint Orientation<sup>19</sup>

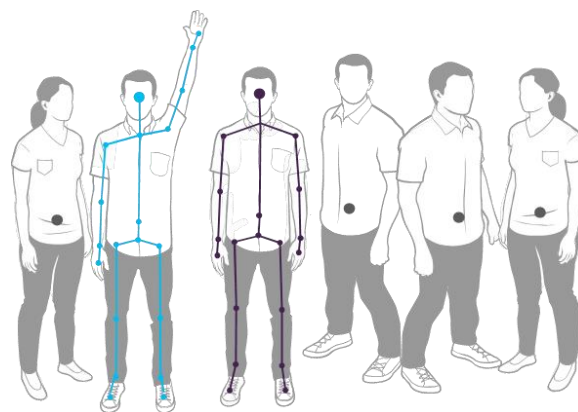


Figure 30 - Skeleton Tracking<sup>20</sup>

<sup>19</sup> Copyright <http://msdn.microsoft.com/en-us/library/hh973073.aspx>

<sup>20</sup> Copyright <http://msdn.microsoft.com/en-us/library/hh973074.aspx>

## D. Research Question 2

### Session 1

#### Sketches

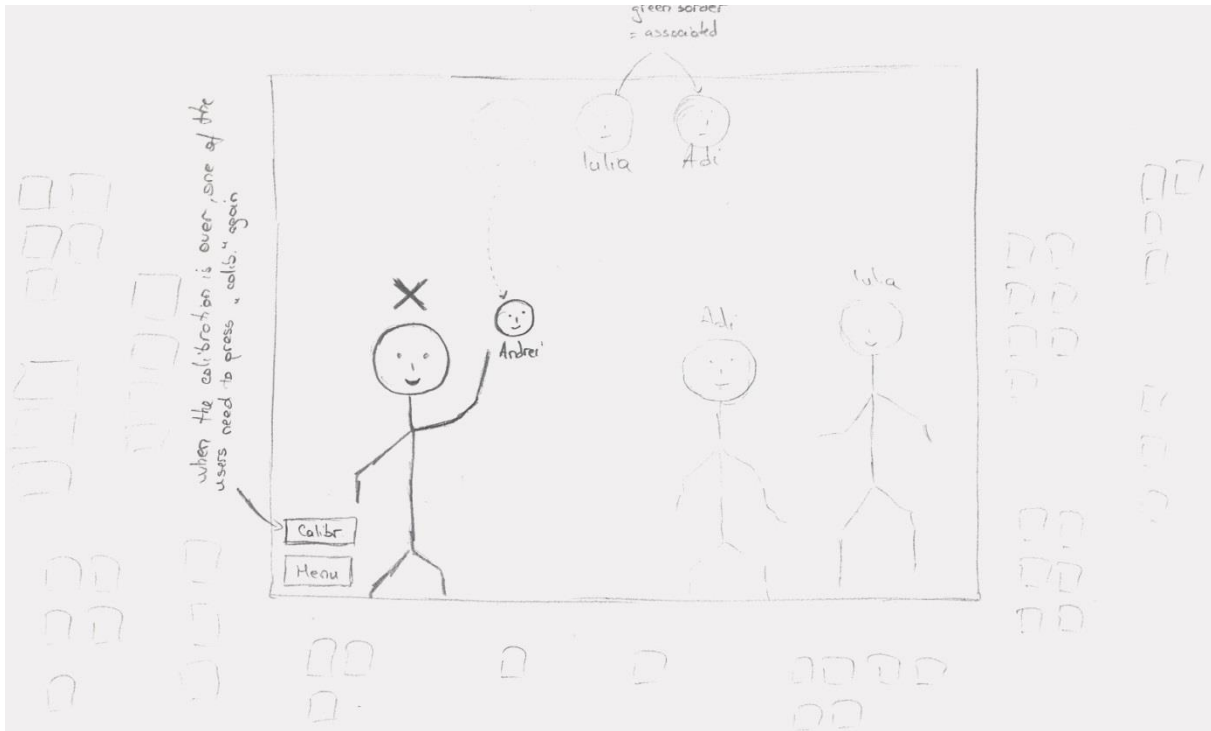


Figure 31 - User mapping (using Kinect)

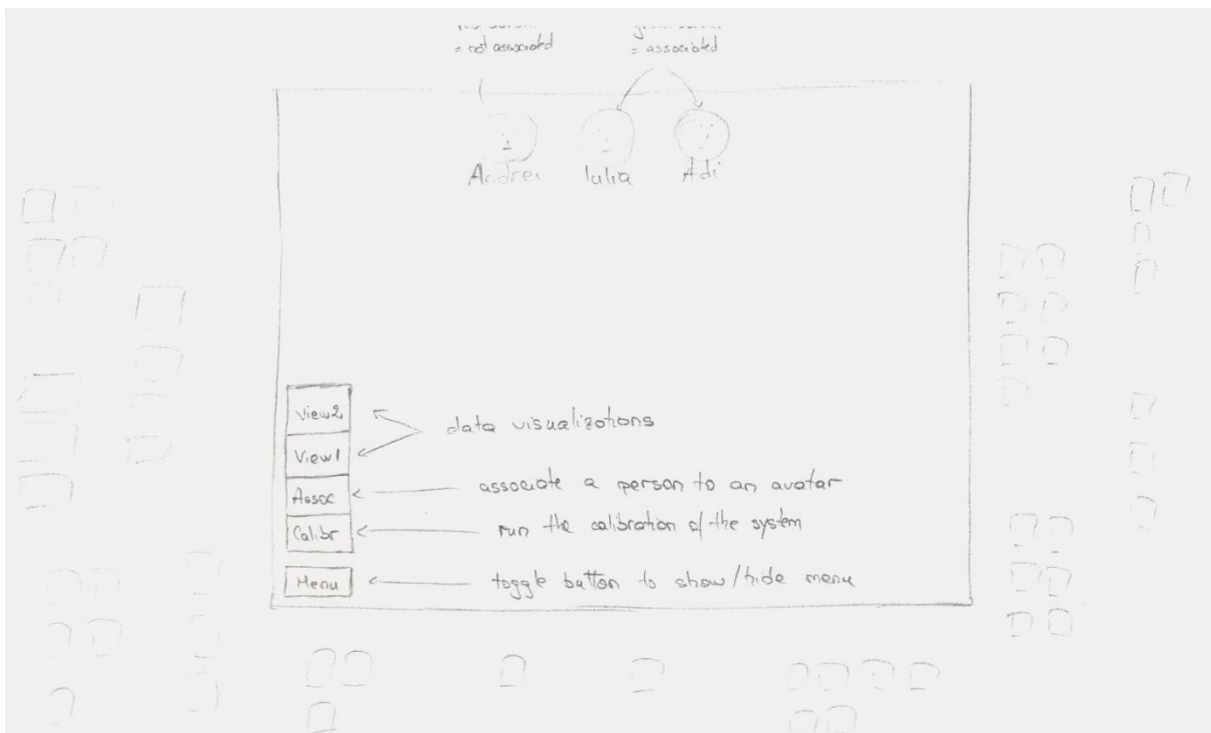


Figure 32 - Menu Options



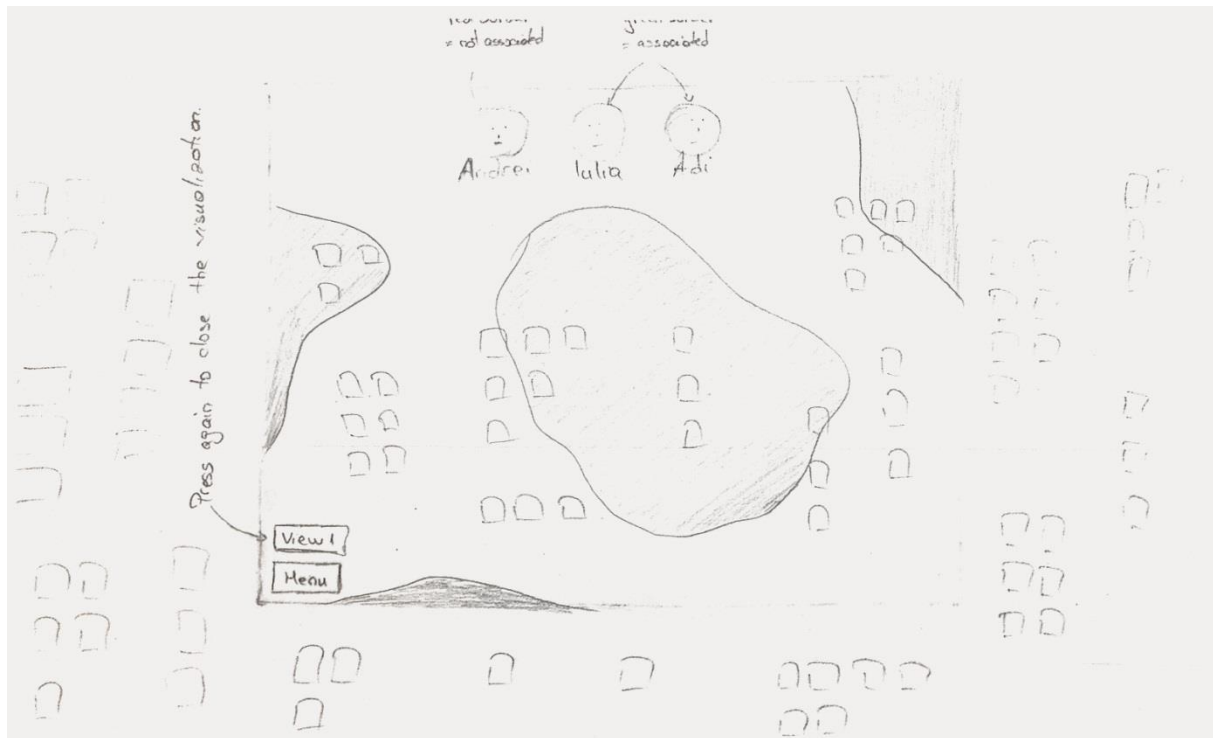


Figure 33 - Heatmap visualization proposal

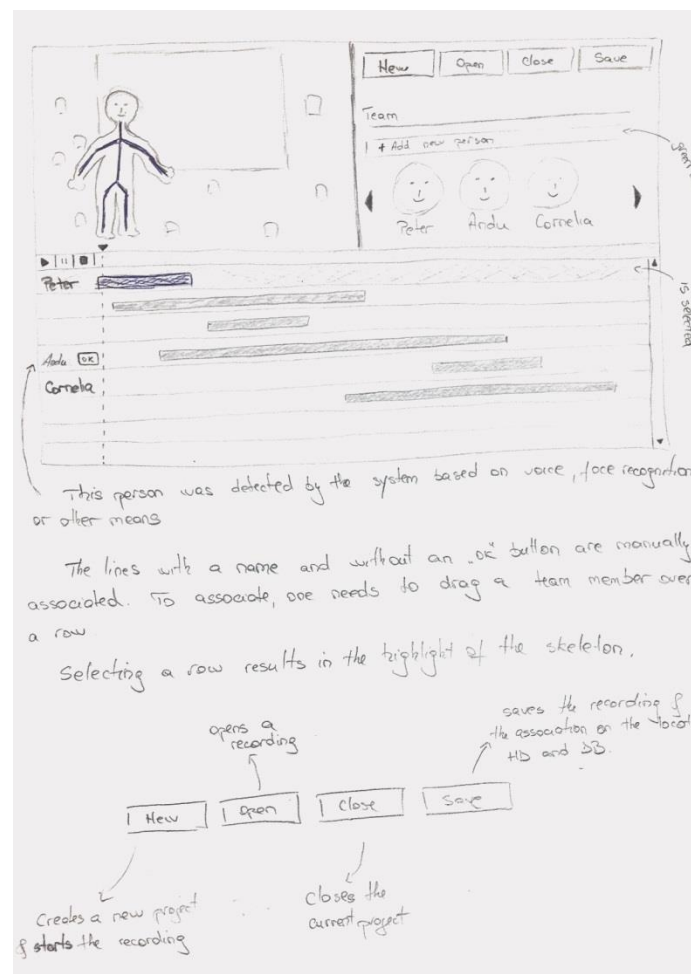


Figure 34 - User Mapping (PC application)

## Findings for the Kinect prototype

### Person 1

The Kinect prototype was the **second presented prototype** from the two available ones (Kinect and desktop app). In the beginning of the session, because of the nature of the other prototype it was hard to describe what and how the Kinect prototype works. The first prototype is an after work mapping prototype, where the user works with a recording while this prototype is an on the fly mapping system which requires more interaction from the user but less interaction in the end. Because of the after and on the fly nature of the prototypes, it was hard to explain to him what does the notion of a “real user”, “recorded user” and “skeleton” mean because he was all the time referring to the other prototype which needed a different mindset.

It is much better to make a clear separation and to give a clear definition about what each notion mean in order to not confuse people. Besides this, if it's possible the notions should be the same on all of the prototypes.

Even after thinking that the “user” notions are clear, for making the mapping between the real user and the avatar the person said at one moment that for doing the mapping he would drag the avatar head on the top of the projected person (which in fact there wasn't one but instead the prototype described the person as a real user and not as a projection). For the future, a better explanation of what is real and what is projected is needed.

Some ambiguous things which this person pointed out:

- He does not see the sense of having the heads on the top of the screen.
- When doing the association, he considers the floating heads as a distraction and he would prefer having a normal list of items.
- He is also concerned that someone would misuse or associate with a wrong avatar by mistake.
- For linking he would expect to touch the avatar for a few seconds in order for the link to be made.
- He does not like the idea to associate all the time with the avatar when he comes back to work, the system should detect him and just ask to validate.
- If a new person enters the field of view of the sensor, the system should ask something like “Who are you? Are you .... ? “

**Notes:** Person 1 did not have any prior knowledge of Kinect and/or how the technology works.

### Person 2

The Kinect prototype was **the first prototype** which was presented. When the interface was presented and he was asked how he would interact with the system, the simple said: *I would just click my name, that's it.* According to him, the association would be done by touching the wall on the

region where the avatar is located. The principles how the avatar and the “real person” interacts with it were understood as it was designed.

For him the interface is quite intuitive but there are some things which he pointed out:

- What is it happening if the person is trying to touch the wall but in the same time the hands come in-between the projector and the wall, this is a bit annoying and may make the person not react in a natural way
- He does not agree with objects which are set by their absolute position. As an example the menu button is not placed comfortable because he may be in another area of the room and he does not want to disturb other people in order to click on that button.
- According to him, everything should be kept “local” (next to the user)
- He does not see the sense of having to press the avatar and the associate button. In fact the associate button in that screen was meant to mean “finish association and return to first screen” but it was understood as a second step in the association process.
- He does not see the sense of having the button and the association should finish when the avatar is selected.
- The system was thought to be used for multi-association but according to him the association should be single user because too many problems can arise if people have to wait for each other. For this to work a leader should be appointed in order to manage the association process and this complicates things too much.
- Another point which we brought up was: *How would the system react if I want to do the association from the distance? Would it be able to detect the association or am I forced to actually go next to the wall?*
- Most of the time, especially when doing affinity diagram, people tend to stand back in order to have a good overview on the whole diagram. Depending how big the wall is, some people can go back quite much. Would in that case the person go out from the field of view and does he need to associate again or can the system detect him somehow?
- A visual feedback would be nice to have when the person was associated. In fact, the “X” above a person’s head was intended to have this meaning (to change from X to the selected name), but this was not intuitive for him because of the confusion what is presented text and what are labels/comments on the sketch.
- He does not necessary like the idea of having floating heads around him.

**Notes:** Person 2 did have prior knowledge of Kinect and/or how the technology works.

## Findings for the PC application prototype

### Person 1

The desktop application prototype was the **first presented prototype** from the two available ones (Kinect and desktop app). Unfortunately too much information was provided in order to gather a full and totally subjective feedback. Even so, some really interesting ideas were provided:

- It would be nice the person to be automatically/ semi-automatically detected and the just to just validate them
- He likes the idea of having the timeline because this gives a very good indication of
  - who worked in different time frames
  - what is the percentage of work
  - who was more active
  - what areas were changed by which person
  - what is the coverage which resulted from the work done by different people
- He likes the idea of drag and drop but he admits it is not very intuitive when you first get in contact with the application (it can be part of the learning curve). Instead he would prefer if he could “Tag” and “Approve Tags” for people directly in the video like on Facebook.
- Regarding the OK button located next to a person, it is not intuitive that the person was identified automatically and it needs to be validated. He recommends to be replaced by a “?” sign.
- Information regarding the time frame of a person:
  - He would associate a person.
  - If a person was before associated he would try to drag and drop a line on another line and aggregate them.
  - He would try to double click on a bar in order to play the selected section of the video, having in the same time the person’s skeleton selected in the video.
  - He does not like the idea having a person appearing multiple times in the time line.
  - If possible the system should already aggregate the lines and ask for validations instead of having many “?” buttons.
  - He is concerned about how many lines will be if the video is around 10 hours long.
- He is also concerned about the fact that the user will need to do too many steps in order to do the mapping. He is saying that maybe the user will do this for one video but if too much time is required to do the mapping, he will not use it again.
- It was not obvious for him that selecting a bar would highlight a skeleton in the video (this needs to be redesigned in the prototype).
- The way he would tag people is that he would jump to a section with many people and he will associate them, the system should make automatically other tags and ask for validations.

**Notes:** Person 1 did not have any prior knowledge of movie making software (which represents the interface from where the prototype was inspired from).

## Person 2

In case of this person, the Kinect prototype was **the second prototype** which was presented. He already had prior knowledge about video editing tools. The current prototype was inspired from such tools. In the beginning of the discussion, he was asked to present the interface and to describe what does each part of the interface. Because of his experience with other tools, he was able to relate to the basic functionality of the application. The video and the time line part were intuitive, but he did not see the sense of having the team (plus the smile faces). When he was asked what he thinks about having the faces of the users he answered: *"I would expect that each face has the face of the actual team –member and that the system will do face recognition on the video based on the given faces"*.

If the purpose of the timeline was understood, the way of interacting with it was not the same as we have envisioned. For example he did not understand what is with the OK button and what his purpose is. Because the first line had the name Peter, the next two were blank and the next one had the name Andu, he thought that the two empty slots between Andu and Peter are actually associated with Peter. Even with this logic, the representation was strange for him because the time frames did not make any sense. When explained the way it was envisioned, he said that this is not obvious.

Other things which were not obvious:

- He did not realize that people are associated using drag and drop of the faces. He expects that when he clicks on the empty region from the beginning of the bar a combo box with all the names will give him the possibility to choose a person.
- He did not understand the purpose of the application and he was thinking more that we can use the software for getting information what are people doing and not necessary that he has to tag people.
- "Do I need to associate all the time? This is time consuming."

Things which he suspects:

- Different time frames have different colors and when playing the video the people will have different colors (based on the time frame bar).
- Clicking on a line will play that section of the movie.
- He suspects that he is able to click on the name but he is not sure
- He would like the possibility to combine bars and have just one bar for each team member

Things which are important to him:

- Is it important to know which person did what, on what it worked
- To know what are the roles of a person in the affinity diagram

**Notes:** Person 2 did have prior knowledge of movie making software (which represents the interface from where the prototype was inspired from).

## Session 2

### Evaluation Sheet prototype 1

#### *Cognitive Walkthrough evaluation of user identification process*

#### **How can we design and implement a user experience to support semi-automatic identification of users in the affinity diagramming context?**

When working with the affinity diagram different people may be present in the field of view of the sensor. Each person may be a spectator or an actual project member, and they may come in or go out from the field of view. Part of the scope of the current thesis is to research the best way to semi-automatically detect the project members and to design and implement a user experience which will allow a person to finalize the identification process.

#### **Intended Users, Technologies and Environment**

The target users of this system are HCI experts. The system offers to the user the possibility to load a Kinect recording with the actual affinity meeting and after, to assign different avatars (virtual representation of a participant) to different participants. The participants in the recording are also HCI experts. The user handling the system could be one of the participants. It can be assumed that the total length of a recording is not more than 4 hours. The presented prototype is an application located and operated on a normal computer. It can also be assumed that the number of people participating in the affinity process is 4 and that they generated a total number of 30 frames. Each modification in the application is saved automatically; therefore the user can close and resume the process at any moment.

If necessary, participants can be added in the system at any point in time.

**TASK: Review the recording of an affinity session, identify users and associate them with their specific avatar (its virtual representation in the system).**

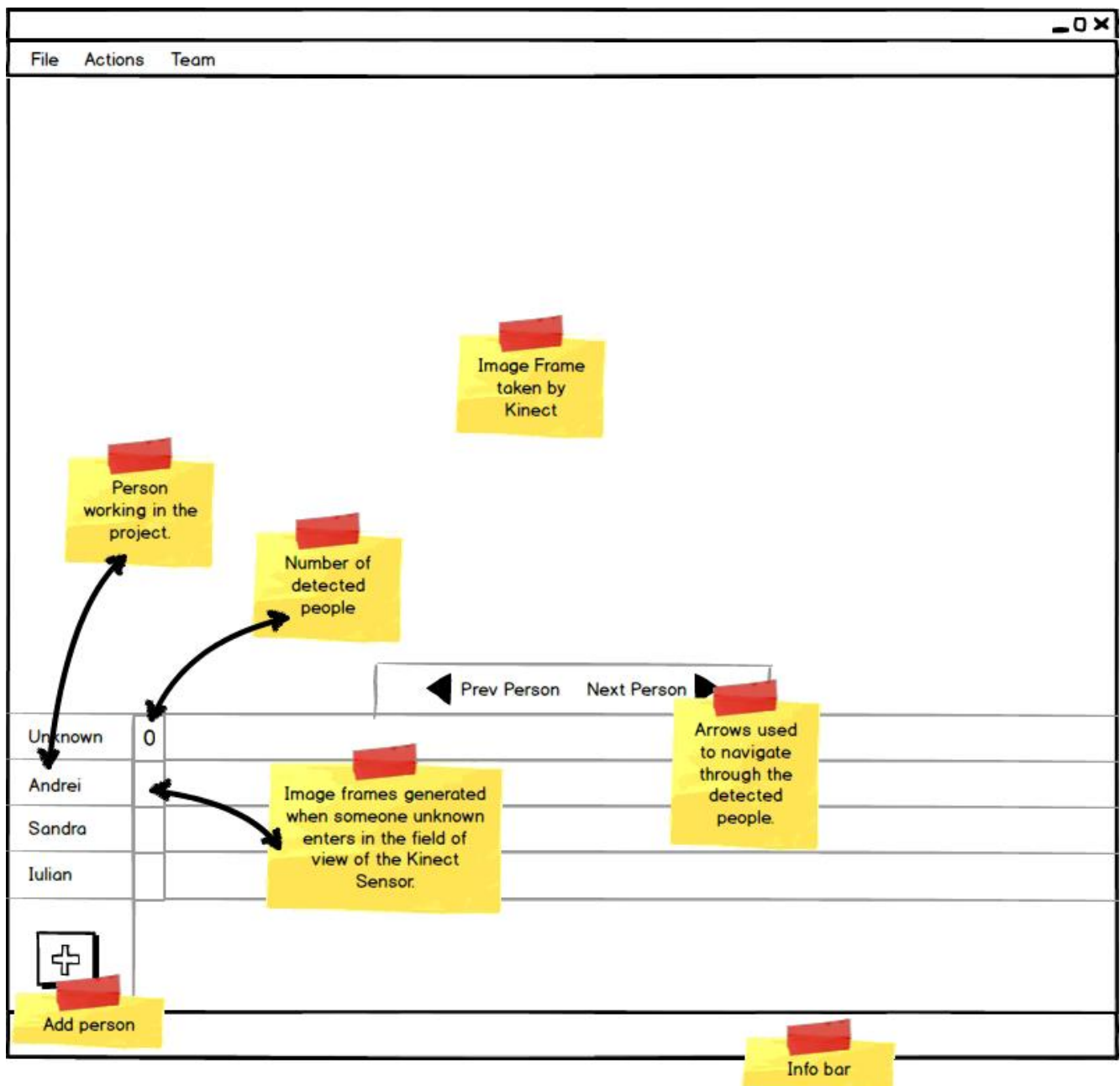


Figure 35- Q2 / Session 2 / Prototype 1

STEP1: Click on the validate button from the bottom right corner of the dotted rectangle in order to confirm the suggestion.

Unknown	2	1	3	1	1	1	2	1	1	2	1	3	3
Andrei													
Sandra													
Iulian													

☐ Select which avatar from the table is associated with the selected person.

1. Will the user be trying to produce whatever effect the action has?

2. Will the user be able to notice that the correct action is available?

3. Once the user finds the correct action at the interface, will she know that it is the right one for the effect she is trying to produce?



4. After the action is taken, will the user understand the feedback given?

STEP2: Press the “Next Person” button in order to highlight the next person in the frame.

	2	1	3	1	1	1	2	1	1	2	1	3	3
Unknown													
Andrei													
Sandra		✓											
Iulian													

Select which avatar from the table is associated with the selected person.

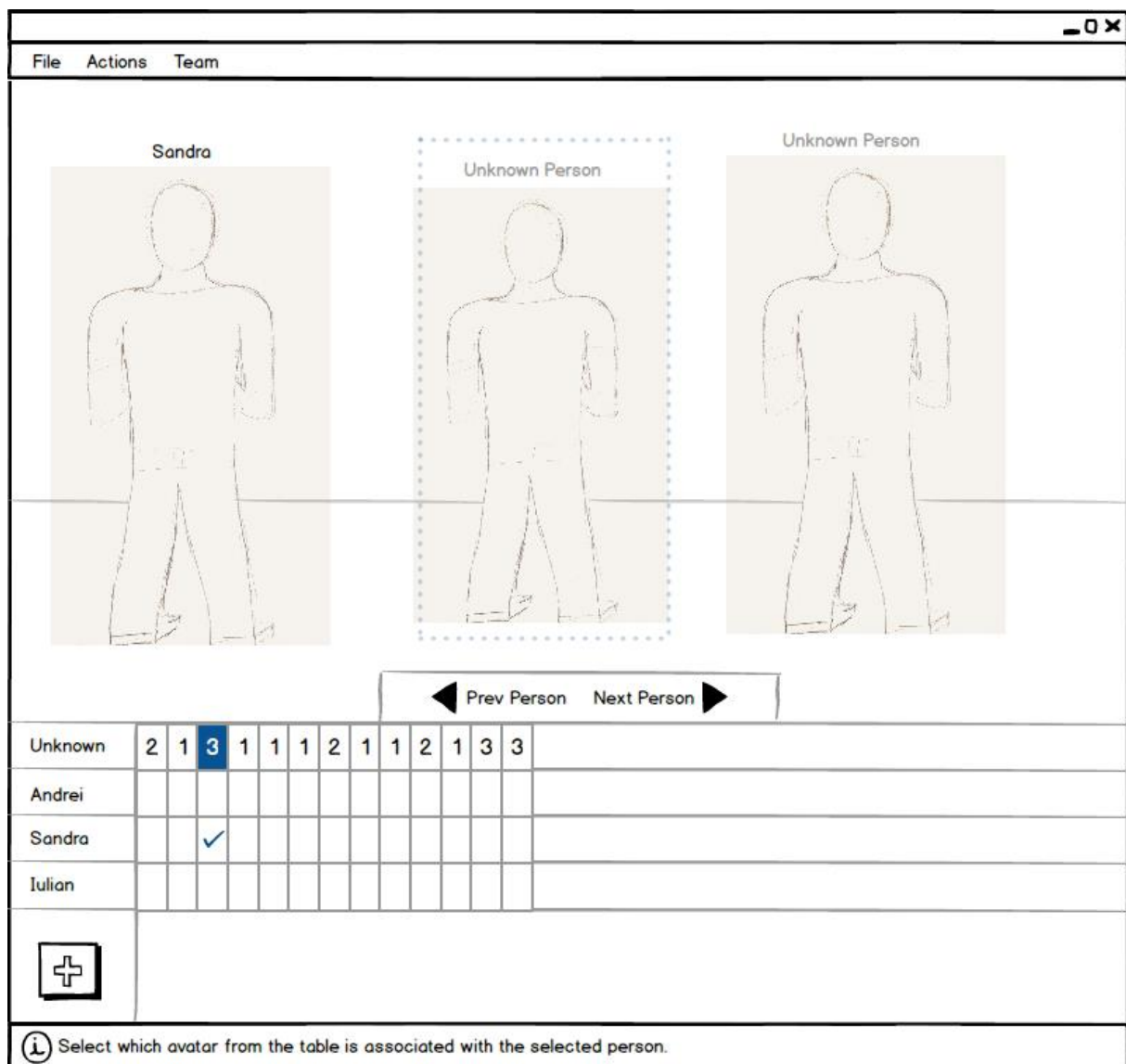
1. Will the user be trying to produce whatever effect the action has?

2. Will the user be able to notice that the correct action is available?

3. Once the user finds the correct action at the interface, will she know that it is the right one for the effect she is trying to produce?

4. After the action is taken, will the user understand the feedback given?

STEP3: Select the right avatar from the lower grid in order to assign it to the highlighted person.



File Actions Team

Sandra Unknown Person Unknown Person

Prev Person Next Person

Unknown	2	1	3	1	1	1	2	1	1	2	1	3	3	
Andrei														
Sandra		✓												
Iulian														

+

i Select which avatar from the table is associated with the selected person.

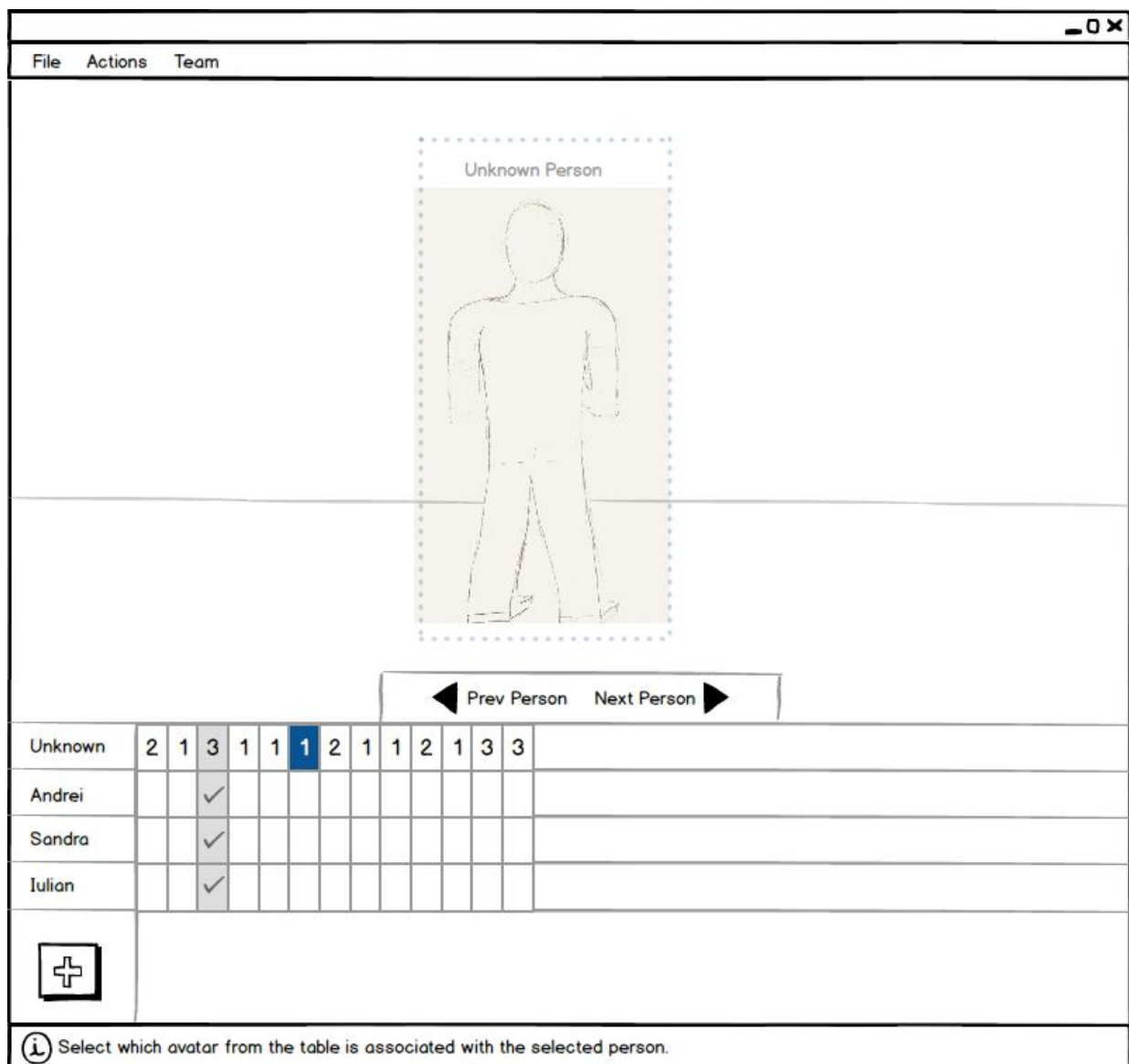
1. Will the user be trying to produce whatever effect the action has?

2. Will the user be able to notice that the correct action is available?

3. Once the user finds the correct action at the interface, will she know that it is the right one for the effect she is trying to produce?

4. After the action is taken, will the user understand the feedback given?

STEP 4: Click on another frame in order to assign/validate other people. Observe that after assigning all people in a frame, the frame changes the background color to gray.



File Actions Team

Unknown Person

Prev Person Next Person

Unknown	2	1	3	1	1	1	2	1	1	2	1	3	3
Andrei			✓										
Sandra			✓										
Iulian			✓										

+

Select which avatar from the table is associated with the selected person.

*1. Will the user be trying to produce whatever effect the action has?*

*2. Will the user be able to notice that the correct action is available?*

*3. Once the user finds the correct action at the interface, will she know that it is the right one for the effect she is trying to produce?*

*4. After the action is taken, will the user understand the feedback given?*

## Evaluation Sheet prototype 2

### *Cognitive Walkthrough evaluation of user identification process*

**How can we design and implement a user experience to support semi-automatic identification of users in the affinity diagramming context?**

When working with the affinity diagram different people may be present in the field of view of the sensor. Each person may be a spectator or an actual project member, and they may come in or go out from the field of view. Part of the scope of the current thesis is to research the best way to semi-automatically detect the project members and to design and implement a user experience which will allow a person to finalize the identification process.

#### **Intended Users, Technologies and Environment**

The target users of this system are HCI experts. The system offers to the user the possibility to load a Kinect recording with the actual affinity meeting and after, to assign different avatars (virtual representation of a participant) to different participants. The participants in the recording are also HCI experts. The user handling the system could be one of the participants. It can be assumed that the total length of a recording is not more than 4 hours. The presented prototype is an application located and operated on a normal computer. It can also be assumed that the number of people participating in the affinity process is 4 and that they generated a total number of 30 frames. Each modification in the application is saved automatically; therefore the user can close and resume the process at any moment.

**TASK: Review the recording of an affinity session, identify users and associate them with their specific avatar (its virtual representation in the system).**

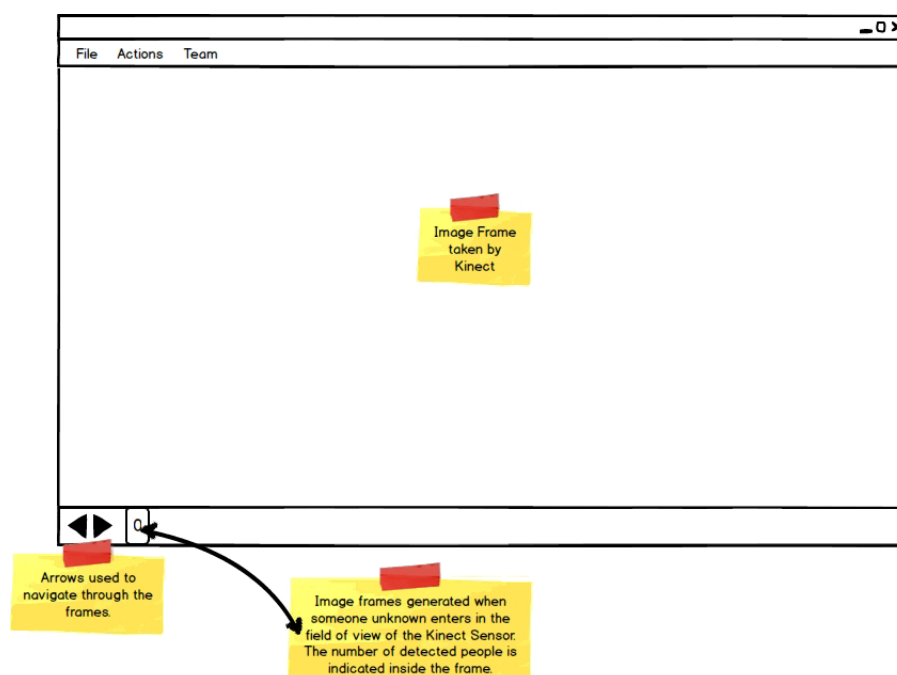
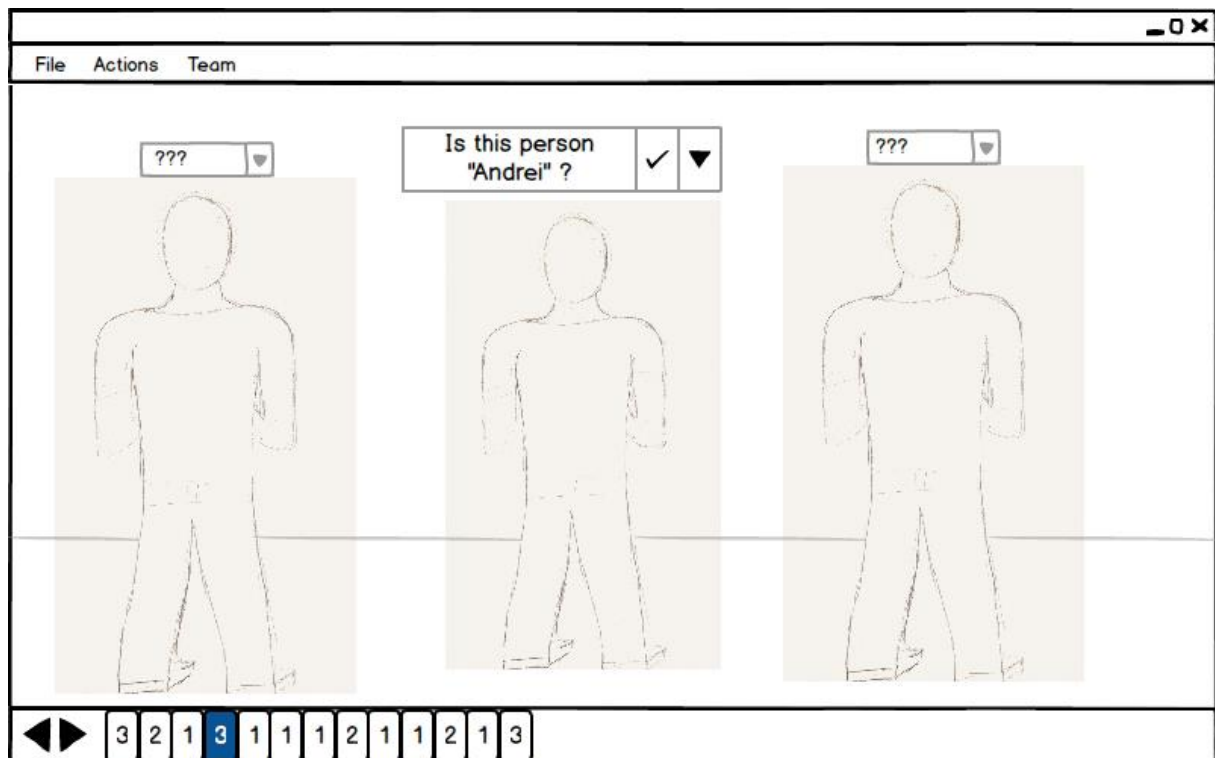


Figure 36 - Q2 / Session 2 / Prototype 2

STEP1: Click on the validate button next to the identified person's name in order to confirm the suggestion.



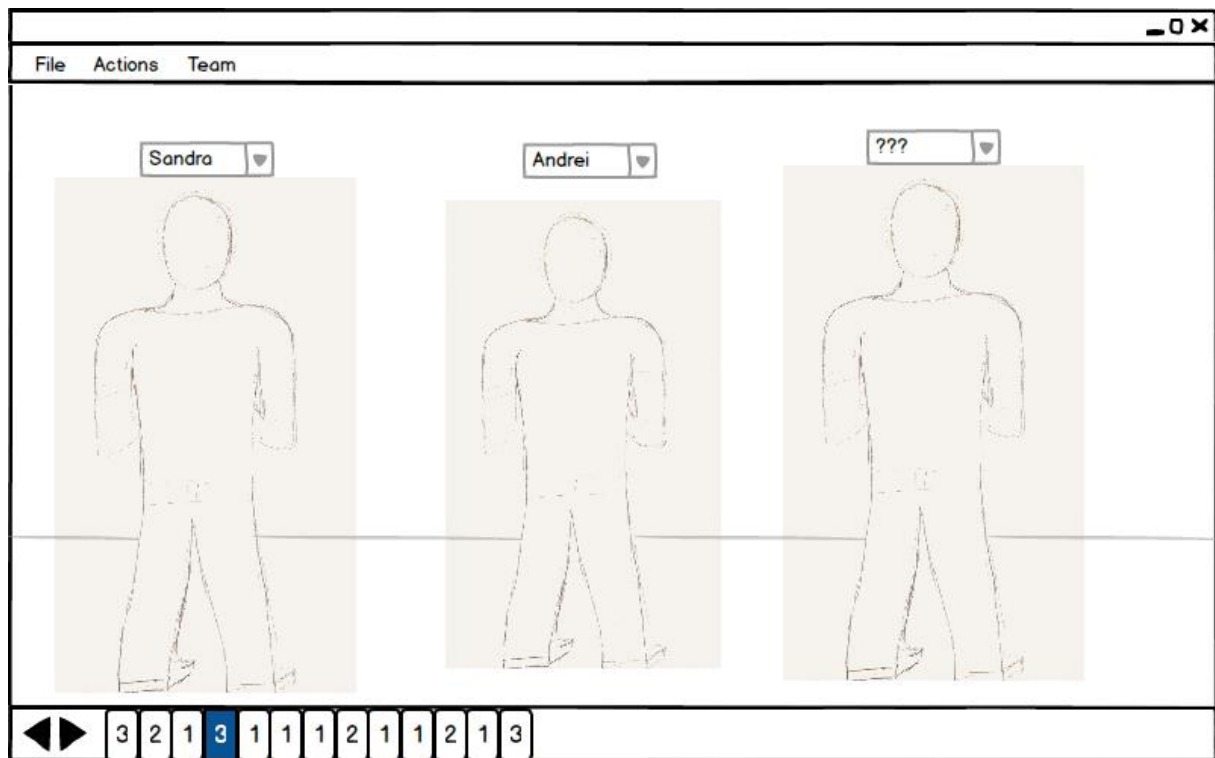
1. Will the user be trying to produce whatever effect the action has?

2. Will the user be able to notice that the correct action is available?

3. Once the user finds the correct action at the interface, will she know that it is the right one for the effect she is trying to produce?

4. After the action is taken, will the user understand the feedback given?

STEP2: Click on the arrow next to “???” in order to choose the right avatar for that specific person.



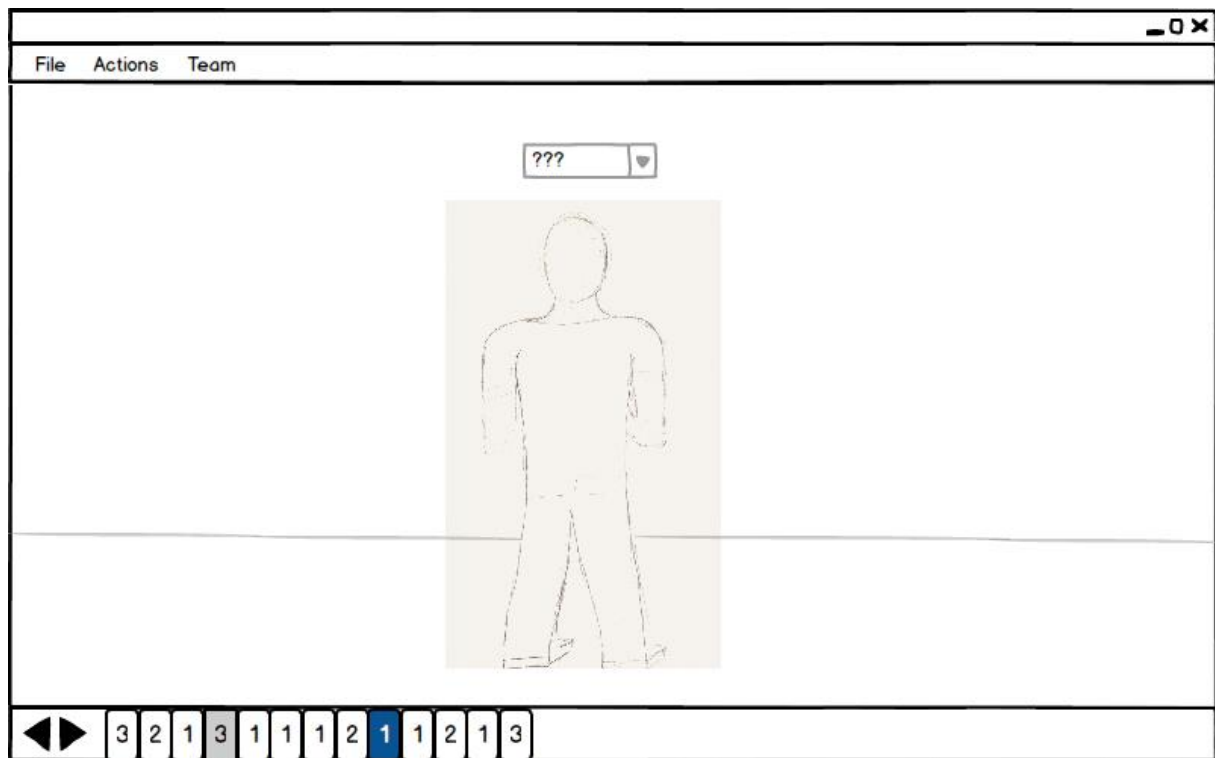
1. Will the user be trying to produce whatever effect the action has?

2. Will the user be able to notice that the correct action is available?

3. Once the user finds the correct action at the interface, will she know that it is the right one for the effect she is trying to produce?

4. After the action is taken, will the user understand the feedback given?

STEP 3: Click on another frame in order to assign/validate other people. Observe that after assigning all people in a frame, the frame changes the background color to gray.



1. Will the user be trying to produce whatever effect the action has?

2. Will the user be able to notice that the correct action is available?

3. Once the user finds the correct action at the interface, will she know that it is the right one for the effect she is trying to produce?

4. After the action is taken, will the user understand the feedback given?



## Evaluation Sheet prototype 3

### *Cognitive Walkthrough evaluation of user identification process*

#### **How can we design and implement a user experience to support semi-automatic identification of users in the affinity diagramming context?**

When working with the affinity diagram different people may be present in the field of view of the sensor. Each person may be a spectator or an actual project member, and they may come in or go out from the field of view. Part of the scope of the current thesis is to research the best way to semi-automatically detect the project members and to design and implement a user experience which will allow a person to finalize the identification process.

#### **Intended Users, Technologies and Environment**

The target users of this system are HCI experts. The system is designed to be used while working on the affinity diagram. The system records and detects new users asking them to self-identify. A person is considered as a new user if it spends at least a few minutes in front of the affinity diagram. A person which is just “passing through” will not be considered as a new participant. It can be assumed that the people involved in the affinity process have knowledge of the Kinect sensor and how to interact with it. When needing to interact, the sensor will intercept the user’s natural body moves and create a virtual skeleton from it which will mimic what the user does. The skeleton is projected on the wall by a projector. Besides the skeleton, different actions will be also projected. To interact with the projected buttons, imagine that your right hand is a substitute for your mouse. Clicking the mouse involves pushing your hand in front.

It can be assumed that the total number of people participating in the affinity process is 4 and that the wall where the affinity diagram is located has a length of 6 meters.

**TASK: Identify users and associate them with their specific avatar (its virtual representation in the system).**

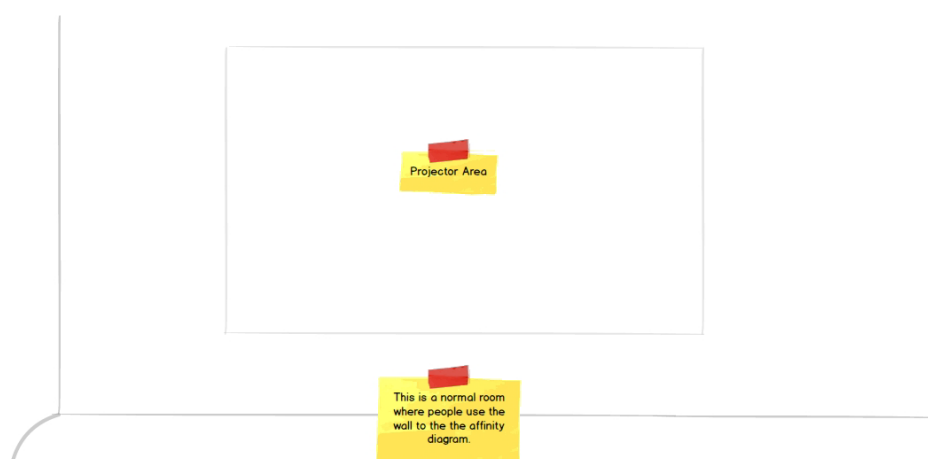
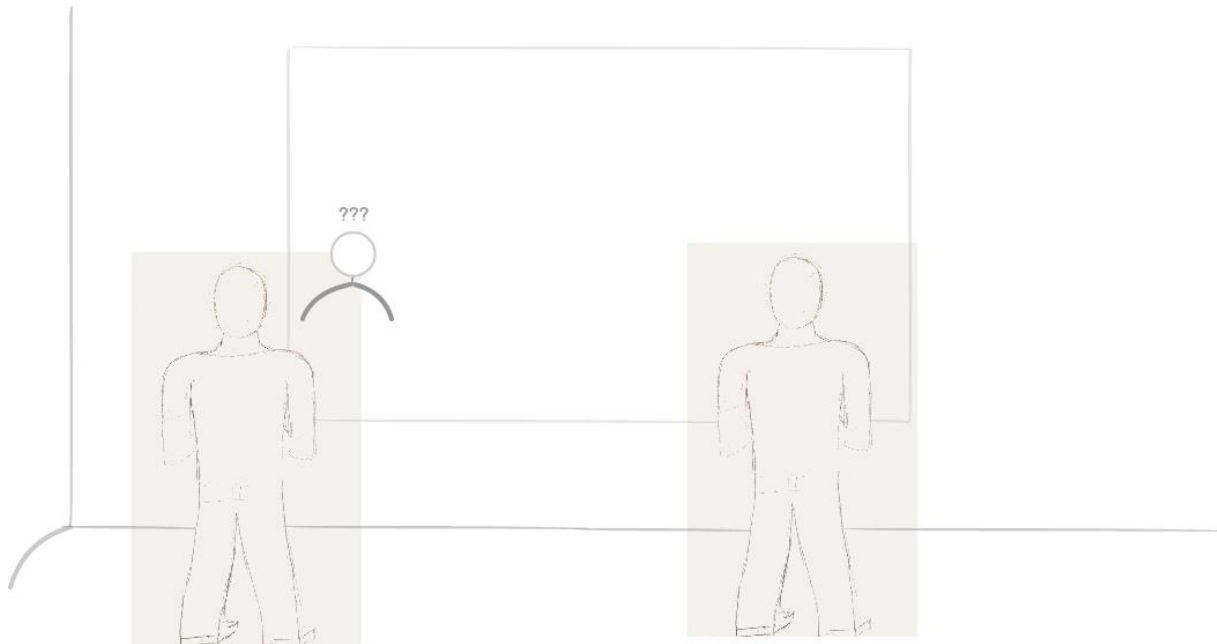


Figure 37- Q2 / Session 2 / Prototype 3

STEP1: Observe the projected skeleton of a new and unidentified person. Move a bit in order to observe the update in the projected skeleton.



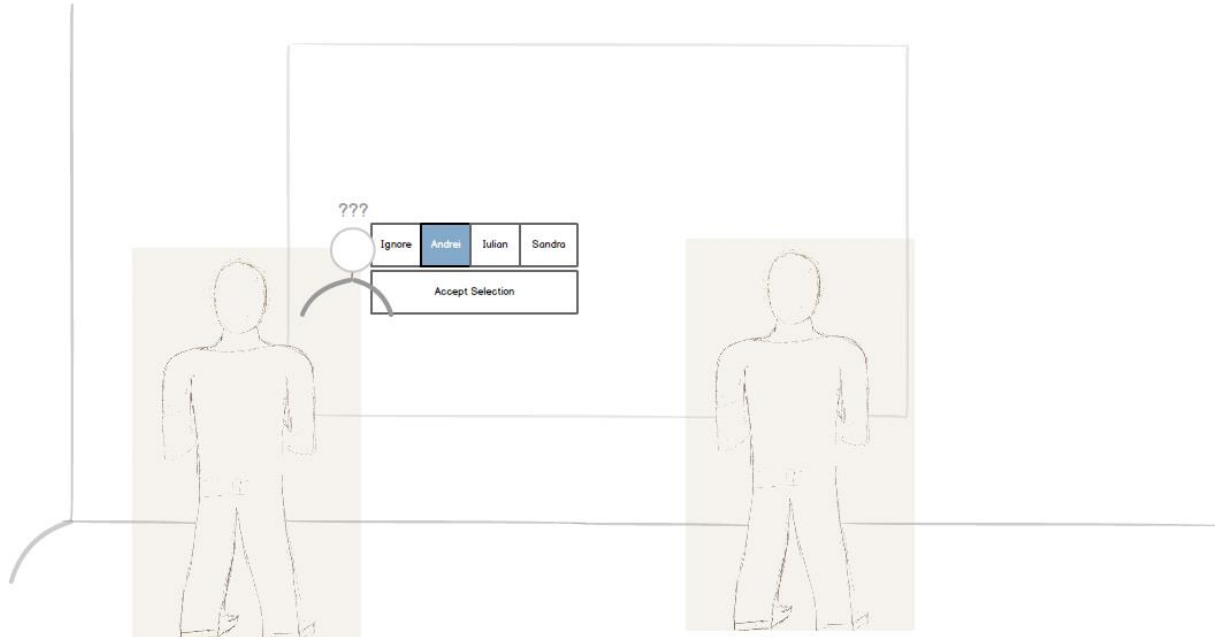
1. Will the user be trying to produce whatever effect the action has?

2. Will the user be able to notice that the correct action is available?

3. Once the user finds the correct action at the interface, will she know that it is the right one for the effect she is trying to produce?

4. After the action is taken, will the user understand the feedback given?

STEP2: The system identified you as being a person called “Andrei”. Accept the selected person by pressing “Accept Selection”.



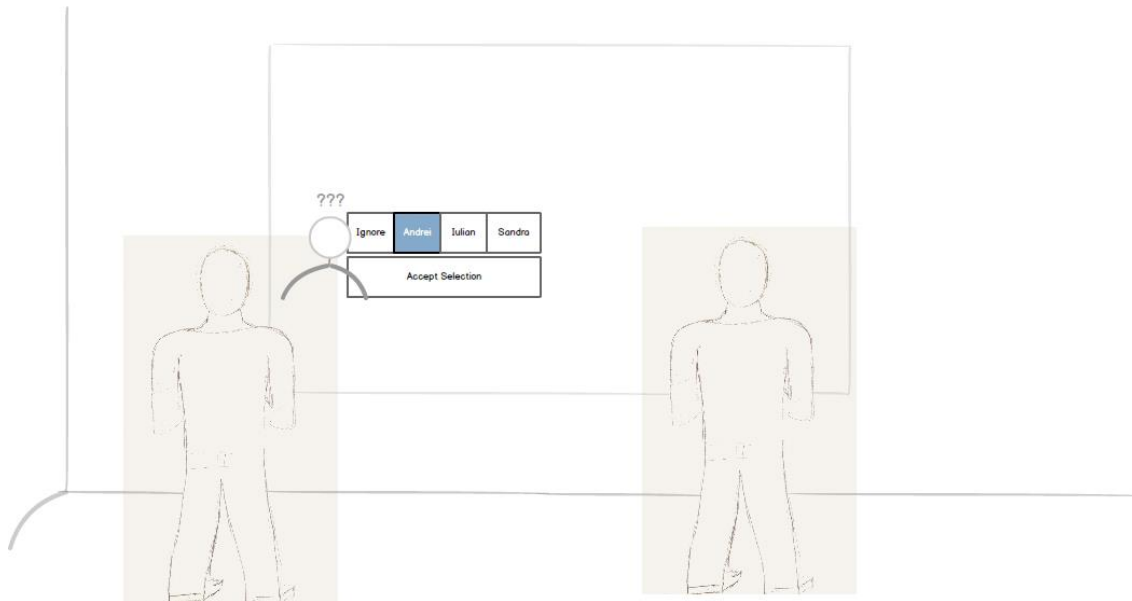
1. Will the user be trying to produce whatever effect the action has?

2. Will the user be able to notice that the correct action is available?

3. Once the user finds the correct action at the interface, will she know that it is the right one for the effect she is trying to produce?

4. After the action is taken, will the user understand the feedback given?

STEP3: The system identified you as being the wrong person. Select the right person (or Ignore) and press "Accept Selection".



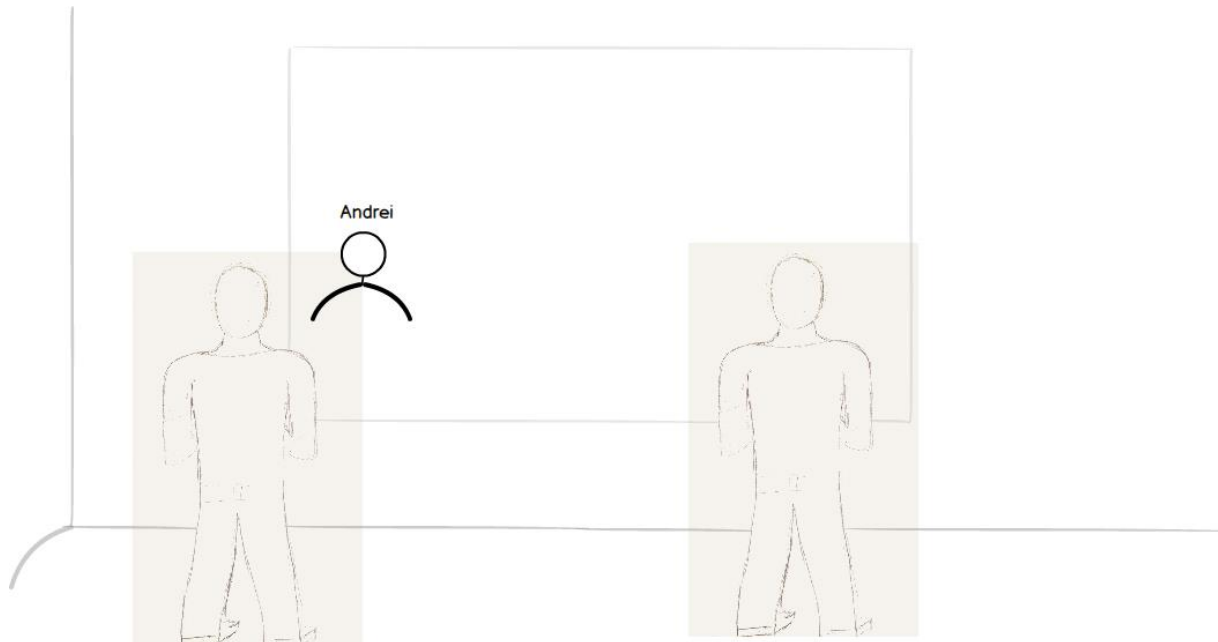
1. Will the user be trying to produce whatever effect the action has?

2. Will the user be able to notice that the correct action is available?

3. Once the user finds the correct action at the interface, will she know that it is the right one for the effect she is trying to produce?

4. After the action is taken, will the user understand the feedback given?

STEP4: Observe the change in the text above the skeleton's head from "???" to "Andrei" (the skeleton will disappear automatically after a few seconds).



1. Will the user be trying to produce whatever effect the action has?

2. Will the user be able to notice that the correct action is available?

3. Once the user finds the correct action at the interface, will she know that it is the right one for the effect she is trying to produce?

4. After the action is taken, will the user understand the feedback given?

## Findings

The findings presented below represent the aggregation of the answers given by **6 participants** in a cognitive walkthrough evaluation session.

### General questions from participants

- “The order in which present the prototypes affects my experience”
- When I have drop boxes is good to also show at least in one place the content
- How will the users know the information which I added on the yellow boxes on the first page of the cognitive walkthrough?
- Make the notions more clear. Some people did not understand the difference between the “new users” and people “passing through”.
- Set the number of the prototype when dealing with more prototypes.
- Set the number of the page on each prototype.

### *Prototype 1: Desktop Application + Bottom Grid*

#### Step 1: Confirm the suggested person

- The bottom info is too small
- The validate button should be next to the upper question and should have a text like “Please Confirm ☒”
- The checkmark is a bit confusing, it may say that the person is already validated (maybe use some text or write a question).
- Does the entire body need to be visible?
- There should be a “Disconfirm” button also.
- If the user need to disagree then it is confusing what actions one should take.
- The number 3 (total number of detected people) should better represent the remaining number of people to tag.
- There are many elements in the UI and the user may overlook the checkmark.
- The checkmark from the grid is very small and the user may not recognize the feedback.
- The question from the top of the avatar’s head is changing from a question to a name and the grid will have a checkmark next to it, therefore the feedback is rather clear.

#### Step 2: Press “Next Person” in order to highlight the next person

- It would be nice if the person can click on the avatar directly and not on the arrows
- The bottom text (info) is confusing
- “Pre” and “Next” person buttons are a bit confusing and not intuitive enough
- Maybe using a text something like “Next Unknown Person” is better

- There is confusion between the “Unknown” text from the top of the grid and the “Unknown Person” from the top of each person in the video.
- How do I reassign people?
- The new dotted frame appears so the feedback is okay.

### Step 3: Select the correct avatar from the bottom grid

- Does the user need to go through all the frames one by one?
- The grid looks more like a summary and people won't be motivated to click on the avatar and then in the grid.
- Some visual changes in the grid would be nice
- Not sure how to assign people and how to use the grid.
- The grid looks as it would be non-intractable
- What happens if the last person is not in the list (the “+” button is not obvious).
- It is confusing what does the row and what does the column represent.
- The user may expect a popup on the image or something.
- What if by mistake I press and reorder the grid, how is that affecting the content?
- The connection between the avatar and the grid is not clear.
- The new checkmark appears so the feedback is okay.

### Step 4: Click on another frame and assign other people

- It should move by itself to the next frame when I finish with one set.
- It is still a bit confusing the functionality which the grid provides.
- The user will learn from previous steps, therefore he might know what to do.
- The info text from the bottom of the screen is very distracting.
- A button “Next frame” would be nice.
- No overall progress about the validation and mapping.
- Is there a way to tell to the user that the session is over or how much time and validations he still has.
- “Avatar” is not a really good term. Maybe “account” or “identity” is better.
- The grayed out column may give the impression that they are disabled.
- Do I have to do this for every frame?
- The meaning of the “3” is still not clear and is very confusing (it should mean the number of people remained to validate).
- Why is a “Next Person” button but no “Next Frame” button.
- Can the user just jump from a frame to another?
- It seems to be a lot of work involved if we have to validate each person or the same person multiple times.
- The grid is not intractable.

### Prototype 2: Desktop Application + Tagging

#### Step 1: Confirm the suggested person

- It is clear what the user need to do in order to validate the person in the picture.
- The number in the frame is not entirely clear and some people assume the right behavior but they are not sure.
- The arrow and the check mark button are not entirely intuitive. What does the user need to do in order to select another person or to disagree with the question?
- Maybe use a Facebook approach by adding the “yes” / “no” / “select” buttons.

#### Step 2: Click on the arrow next to “???” to choose a person

- The “???” text is a bit confusing. Maybe use a more clear text like “Identify me”.
- It is not intuitive enough that the combo boxes contain names
- What if the person is staying in a posture in which I can’t recognize him? I would not know what to pick.
- A combo box is a very familiar element and is intuitive how to operate it.

#### Step 3: Click on other frames in order to assign other people

- A gray background on the frames doesn’t necessary mean that the people are assigned. I would have to learn this meaning / behavior. It is not very intuitive.
- A gray background gives more the impression that the frame is disabled and not that all persons were validated.
- People didn’t get the idea with the checkmark. They think it should be there to confirm the selection in the combo box.

### Prototype 3: Kinect

#### Setup

- The delay which the application uses to see if a person is working on the diagram or he is simply walking in front of the sensor is a bit confusing and annoying. The person should be detected when he comes in the field of view and if he is only passing through then the confirmation dialog should move with him and it should eventually disappear.
- Why is the person considered new each time when it enters in the field of view?

#### Step 1: Observe the update in the projected skeleton

- Coming / stepping in front of the sensor and seeing the skeleton makes it obvious that the skeleton responds to the persons’ moves.
- The person should also get labeled as “Unknown Person”.



- The users must be facing the wall in order to observe the skeleton.
- Will the skeleton disturb other people?

#### Step 2: Confirm the suggested person

- It is obvious what to do because of the provided set of actions.
- What if the user doesn't know how to use Kinect?
- Your hand shadow may block the projector.
- The user may not know how to press the buttons.
- Why not clicking directly on the button "Andrei"? Why do I have to also accept the selection?
- Maybe is better if the system asks "Are you Andrei?"?
- Why is it "Accept Selection" and not "Confirm Selection"?

#### Step 3: Select the right person or ignore the association

- It is happening automatically and it does not require interaction
- The "ignore" option should be separated from the names in order to be more obvious.
- The "ignore" and "accept" button should be together on the same line.

#### Step 4: Observe the update in the text above the skeleton

- No real feedback (besides the labeling which may not be visible enough).
- The users will only observe if they pay attention on the screen but they may want to go to work and they may take their eyes away.

## Session 3

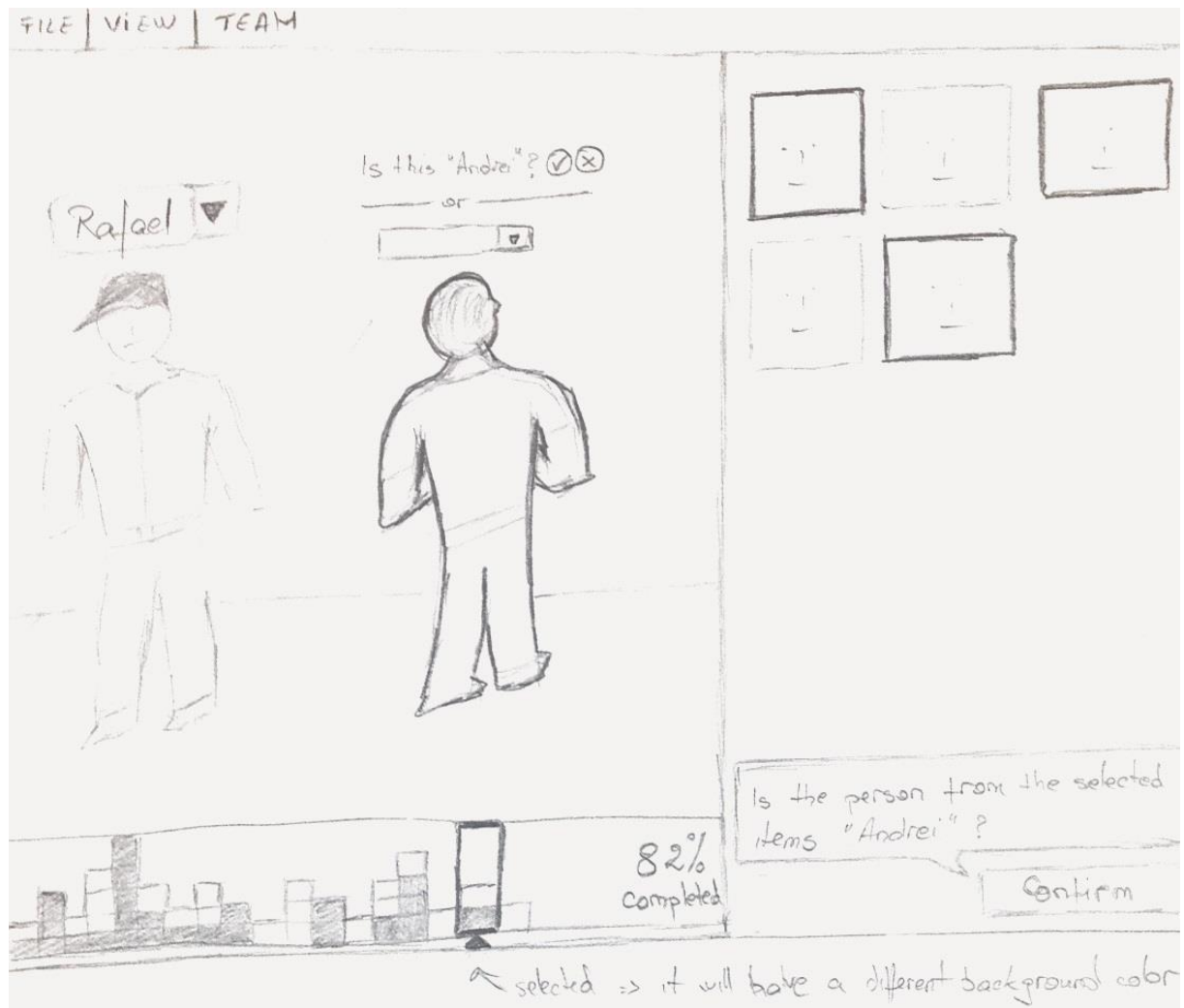


Figure 38- Q2 / Session 3 / Prototype 1

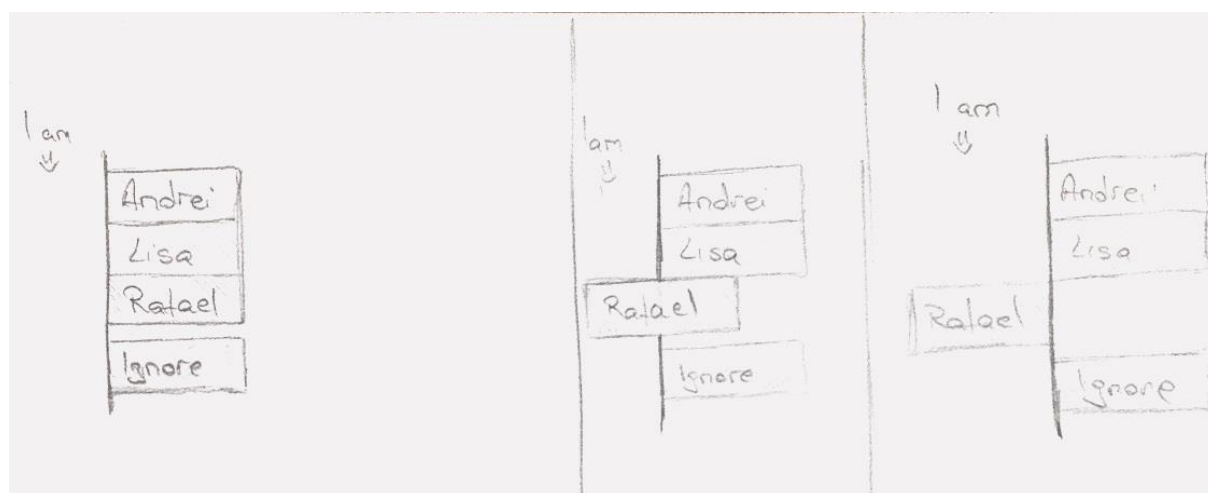


Figure 39- Q2 / Session 3 / Prototype 2

## E. Research Question 3

### Session 1

#### Brainstorming session results document

How to specify the time window and what does it consist from?

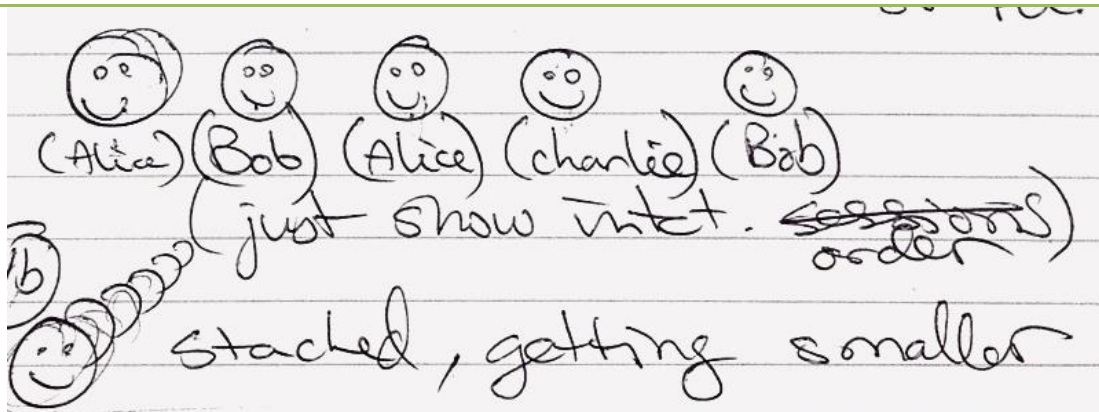
- Last day when I was in the lab
- Last week
- Last time or last X days when this region of wall has changed
- A specific time frame ...

How can I visualize the people who interacted with a region of the wall?

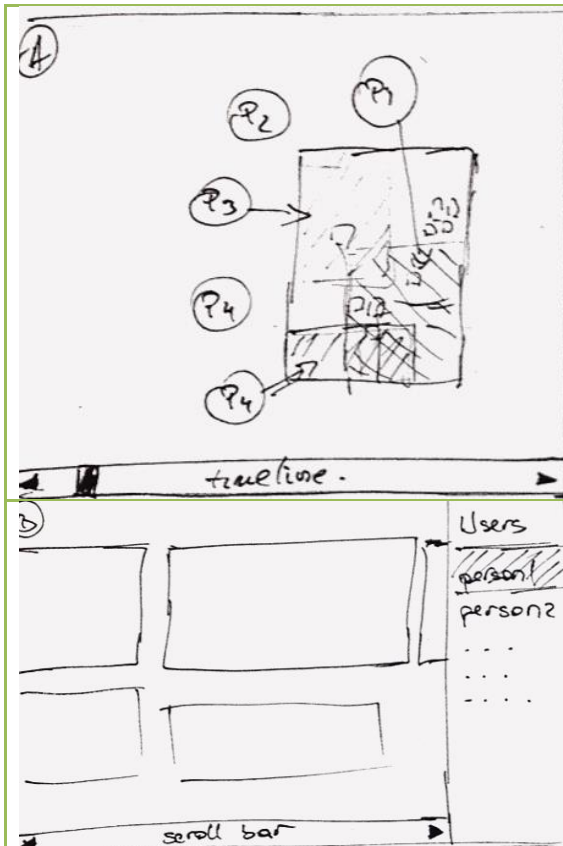
Most time	Most recent
😊 Bob 5:30m	😊 Alice - 6:38
😊 Alice 2:01	😊 Charlie - 16:13
😊 Charlie 04:8	😊 Bob - 34:51

↑ absolute, or rel. time?

People will be displayed in a normal table. The categories can be customized based on the general interest.



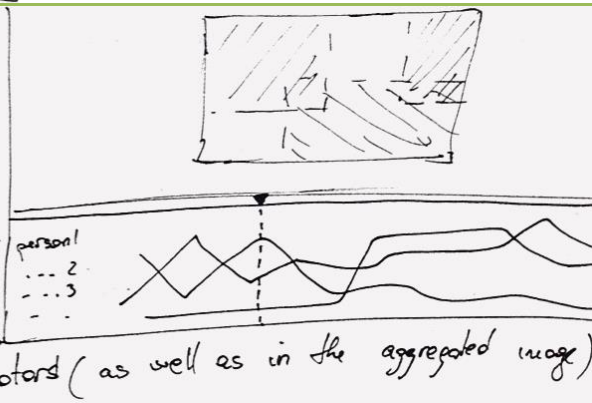
A 2D/3D list of people who worked on that section of the wall. The people are arranged based on the time line, the first in the queue being the last person who interacted with the wall.



The circles (represented by users) have different colors and they can be selected or not. The rectangle region will display different inner regions with semi-transparent color layer for all the selected circles.

Will present the changes done by the user as a photo album or a slideshow.

⑥ - the x lines represent small intervals of time and the y axes gives a feeling of how active a user (person) was in that smaller time frame.  
 - move the dotted line in order to see the aggregated frame from that smaller time frame.  
 - different lines with different colors (as well as in the aggregated image)



⑦ - show the percentage and the name of the person  
 - the % represents how much the person changed the notes in that time frame.  
 - next to each person show also the changes which were done in different time frames.  
 - each small rectangle represents a smaller frame from the original area.

50% Person1	
28% Person2	
12% Person3	
10% Person4	

⑥ - for a person show in an animated way a rectangle (with the size and position depending on the affected area). The animation consists in moving & resizing the rectangle (automatically) while the time from the selected time frame is slowly increasing.  
 - the same can be done with multiple rectangles (of different colors) for different people.

⑦ - compose all the regions changed by a person in one big region with a specific color. The more changes were done by that person in that area the more opaque that area will be. Do the same for multiple people in order to create a heat map, or a 3D map (if instead of transparency we are using cubes which we put on top of each other) ... the more cubes  $\Rightarrow$  the more changes in that area.

⑧ - have a mountain's / valleys map on the selected area. The mountains will represent the amount of changes. On the base of the mountains (the bigger area) will be the users with more changes and towards the peaks we will display the users with the least changes.

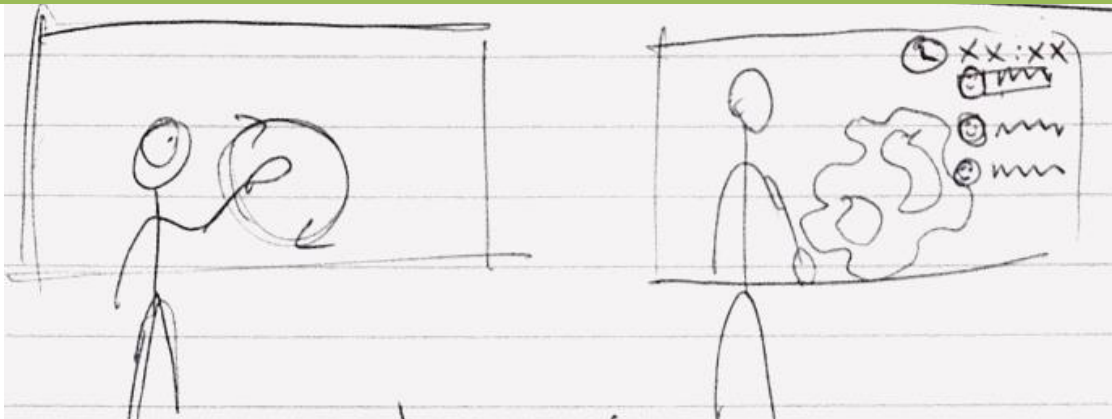


Use the wall as a big map and use a standard rectangle which will play the role of a magnifier glass. The data presented by the magnified glass represents the data about the selected region of the wall which is normally not visible. The user can move the magnifier glass using the mouse or using a normal Kinect interaction.

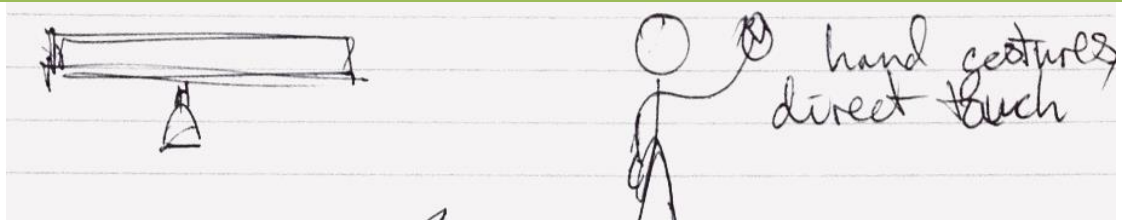


Use a directed graph (See below) where the border of the note changes the color in order to represent the color associated with different users. The users and the associated color should be specified below the graph.

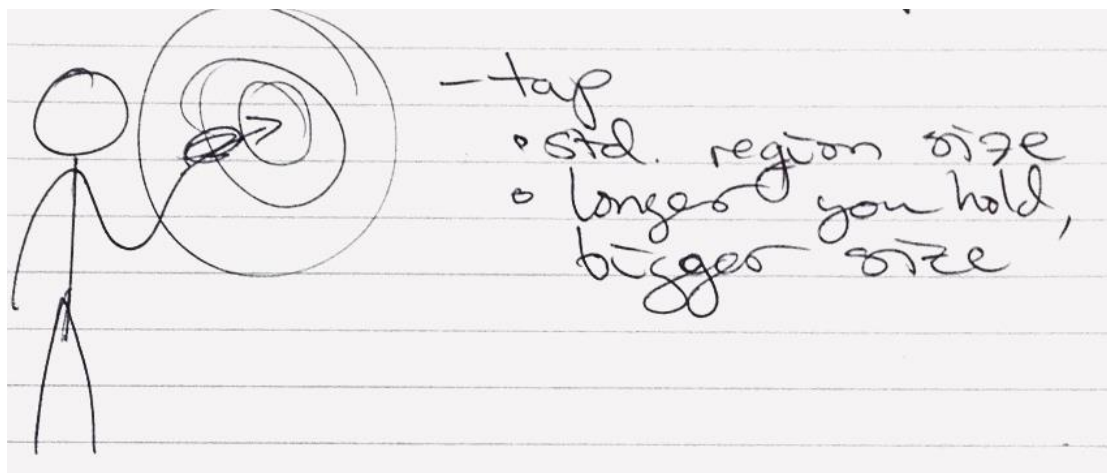
How can I choose/select and scale a specific region of the wall?



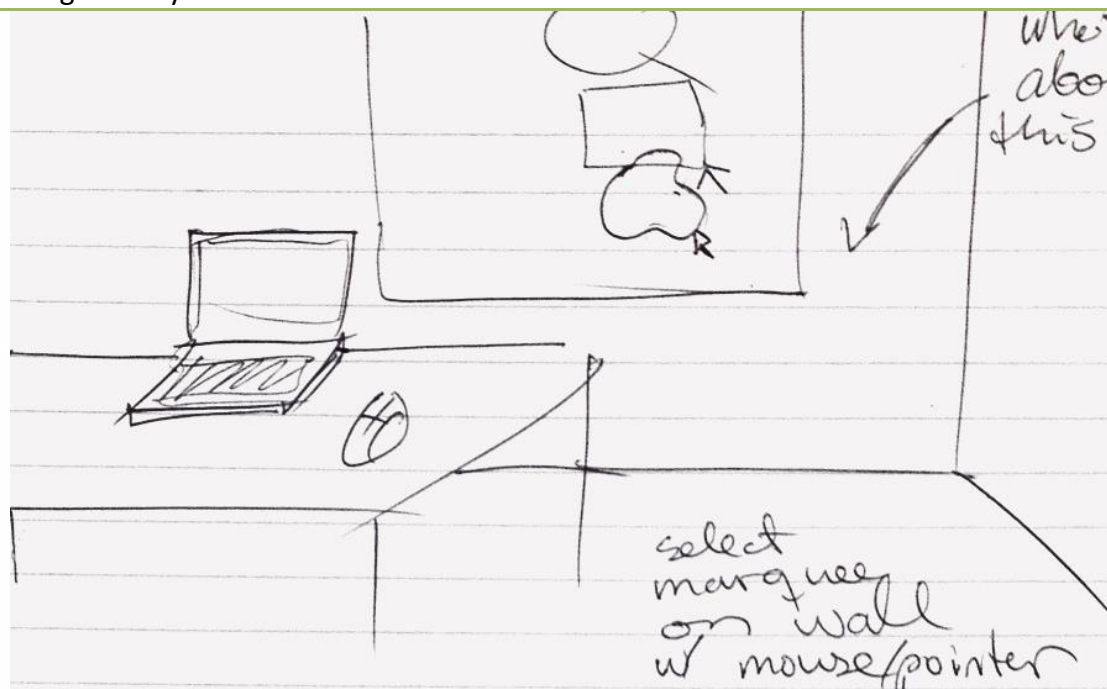




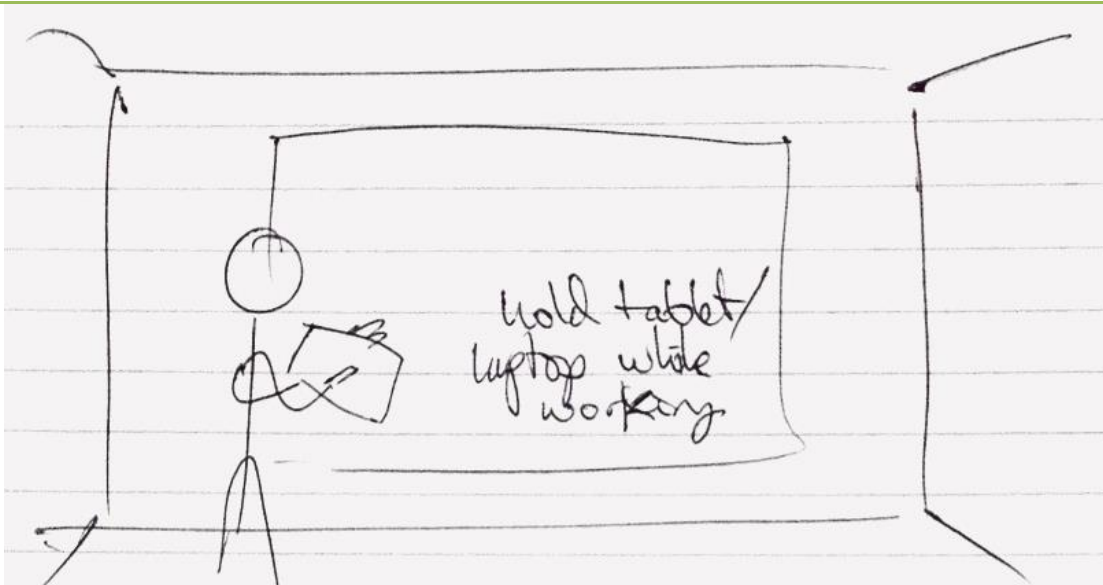
Use Kinect in order to mark an area of the wall.



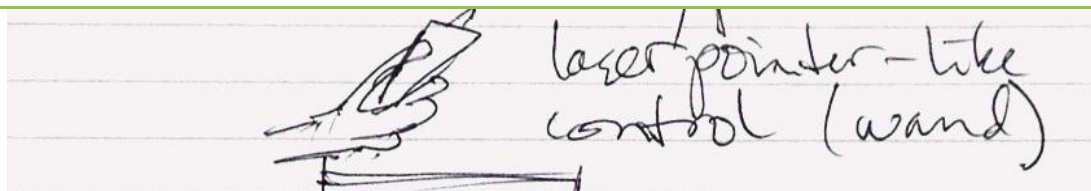
Tap with the hand on the wall and hold it until the area is growing or shrinking. The tap will be recognized by Kinect.



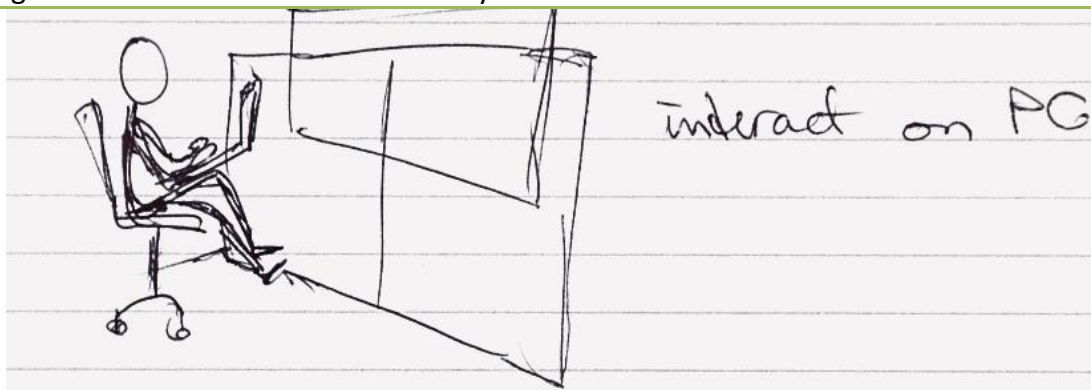
Use a normal computer which is tied to the projector in order to define using the mouse a specific area of the wall.



Use a tablet or a phone to mark the area of interest. The specific area will be also marked on the wall.



Using a remote which is a substitute for your real mouse.



Use a normal computer to select the area of the wall, but instead of having it tied up to the projector, use it also as a viewing display.

How far/near the wall should a person be in order to include him/her in the selection process?

According to the distance to the wall, there can be two areas (if working together):

- The people working physically with the wall; they are the one who actually move the notes around.
- The people which are a bit far away from the wall but which give support and indications to the people moving the notes. They have more the role of bringing arguments why a note should be in a specific place.

As the reason of why a note was moved from one place or another does not depend entirely on the position of the people to the wall, we can't filter out a specific category of people. Nevertheless, because the people working physically with the wall have a 100%

effect on the wall (moving the notes) we can safely say that the second category (support people) has a smaller impact in the final decision.

**Google search (images):** group interaction affinity diagram - [link](#)

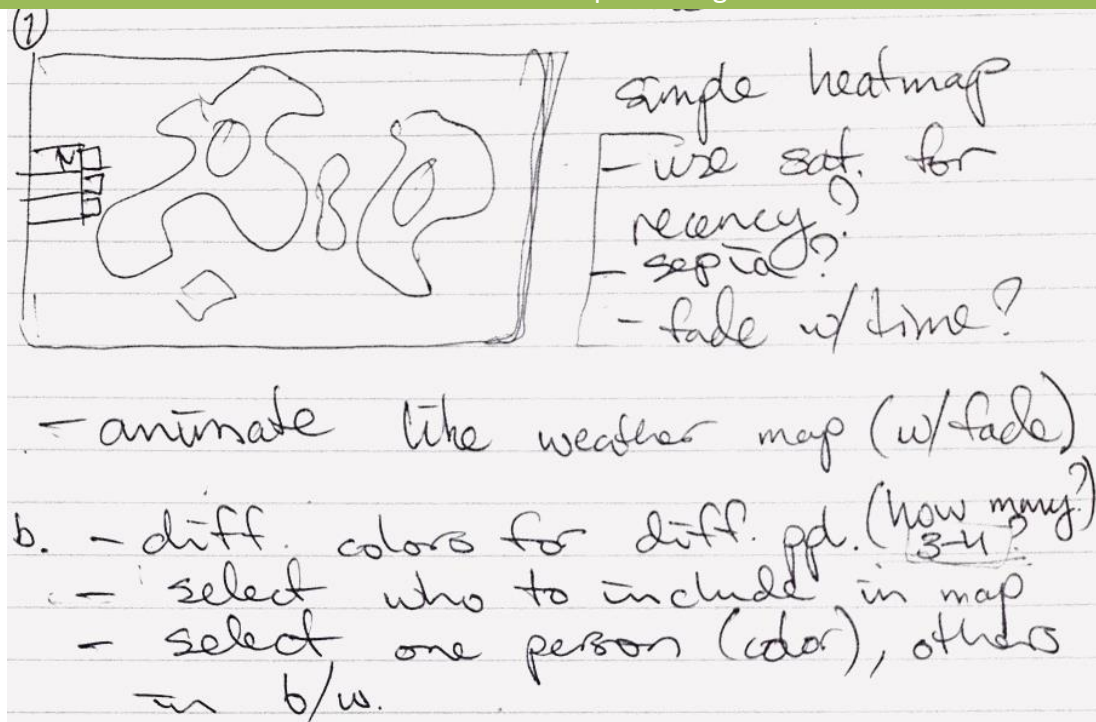
**YouTube search (movies):** affinity diagram - [link](#)

**Social Loafing** <http://old.nios.ac.in/secpsycur/unit-16.pdf> (Psychology - Human Behavior in Group Settings)

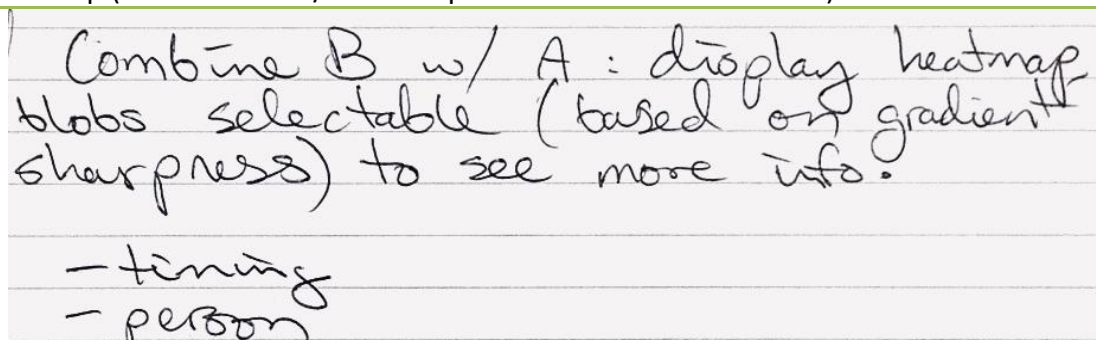
## Results

From the movies and images from above and also according to the social loafing theory, it can be safely assumed that the people staying closer to the wall are more engaged, while the others can give valuable ideas and reasons but they can be also easily distracted. A good measure would be to assign to people a percentage of engagement which depends on how far the person is from the wall.

How can we visualize the amount of attention a specific region of wall have had?

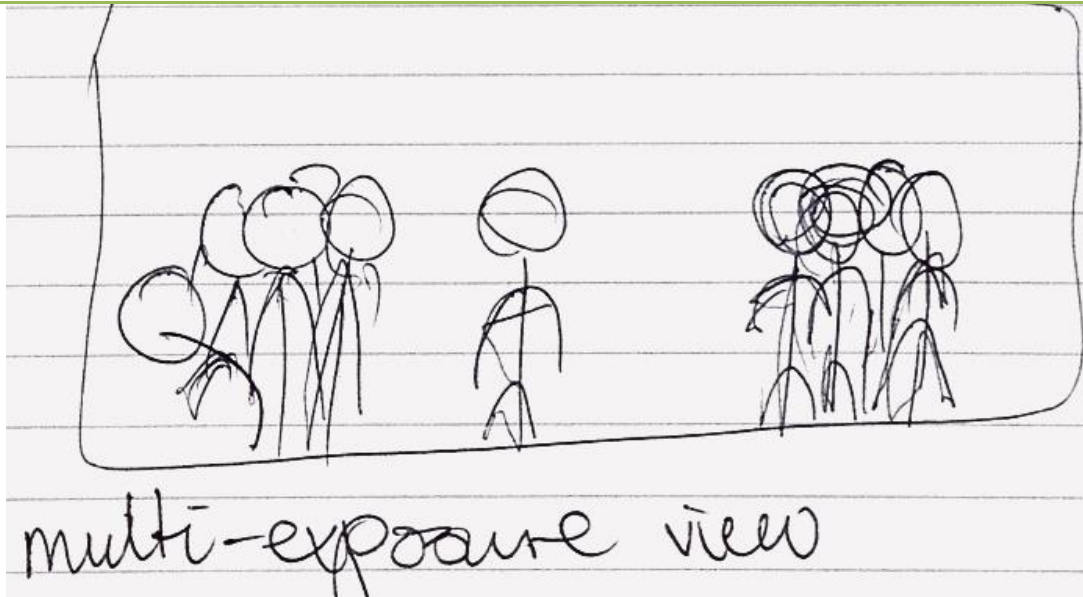


A heat map (the more color/less transparent => the most attention).





A heat map where each section (color) is selectable.



The users (people) who worked on the certain parts of the wall are grouped in that section. A person can be in different parts of the wall.

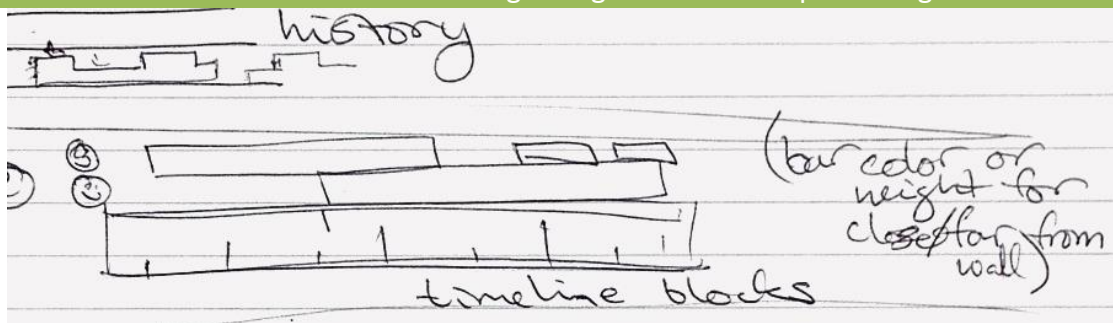
Look at question 2 in order to see other possibilities.

Use a directed graph (See below) where the border of the note changes the thickness.



Show as a graph the people, effort and amount of time invested in the change.

How can we indicate time when visualizing changes made on a specific region of wall?



It is a time line similar to the first prototype but the lines is not rectangle, is taller or shorter, where the size represents how far/near the wall the person was.

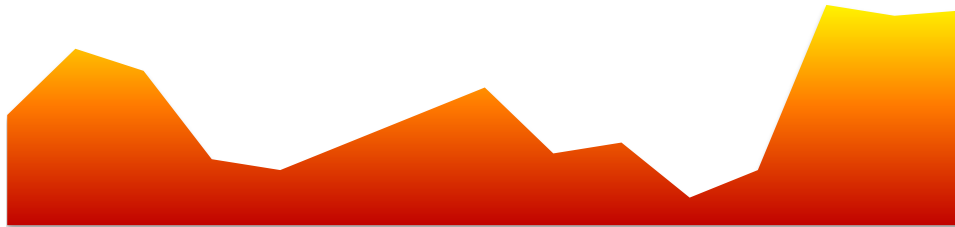
Look at question 2 in order to see other possibilities.

How can we represent user movement paths and their connection with the wall?

Using the near/far distance and the amount of time which they spend working on the wall per frame interval we can extract how much effort a person is putting in working on the affinity diagram. Taking into consideration the movement on the length of the wall, we can detect which areas and notes a specific user affecting.

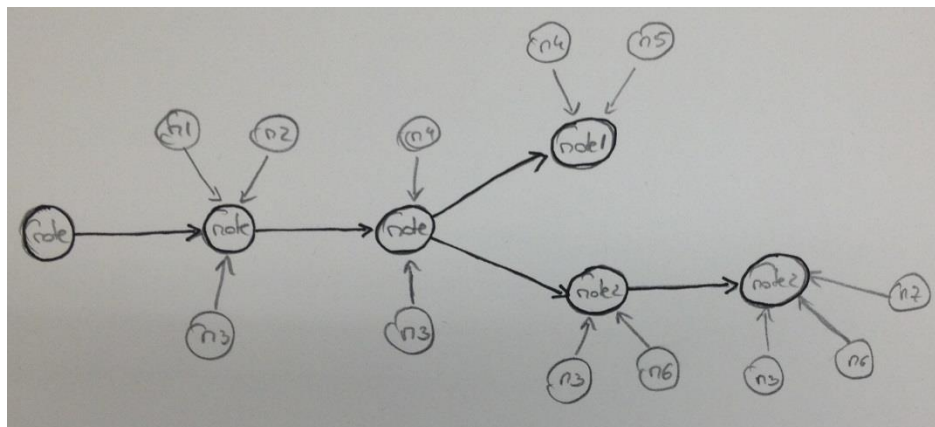
Displaying user movement paths:

- Using a direct graph as shown below and combining it with user/notes data
- Use a graph in order to show the path done by the user over time and the distance of it from the wall. The higher values represent closer to the wall.



How can we visualize the changes in notes?

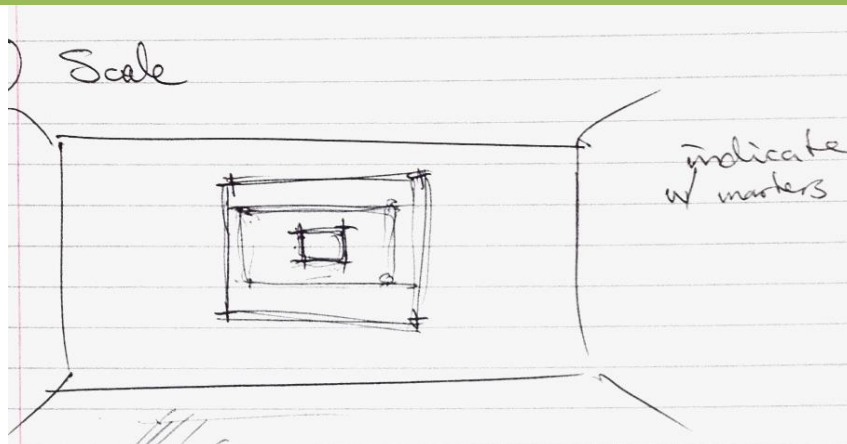
- Show the entire wall with a semitransparent effect except the note which is desired to be observed. A slideshow (time sorted) will display the movement and evolution of that note.
- Use a log in order to display as a list all the relevant places where the note was (which categories).
- Build a timeline directed graph which displays as a node the note itself and as inputs in the node other notes which influenced the change on the selected note.



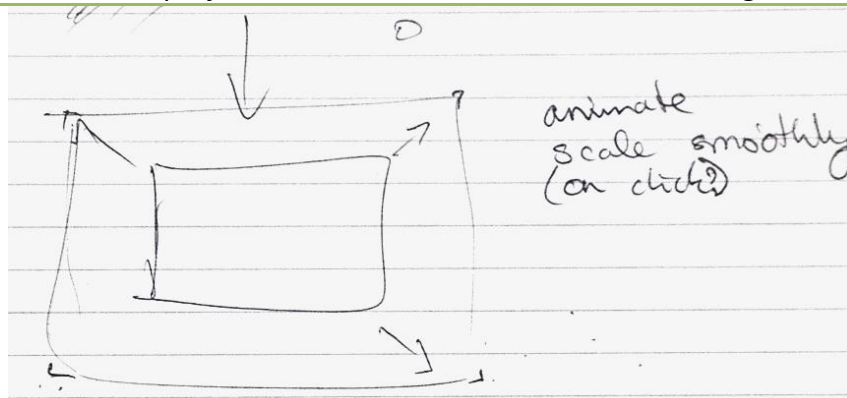
How can we select a note?

- Using Kinect in order to detect a tap on the wall in the region where the note is.
- Using a laser with a red color and detect the color.
- Using a mouse and select region of the image where the note is (on a laptop or desktop PC).
- Using a tablet or a smartphone and take a picture of the specific note.
- Use a tablet which displays the whole wall and tap on the region where the note is.

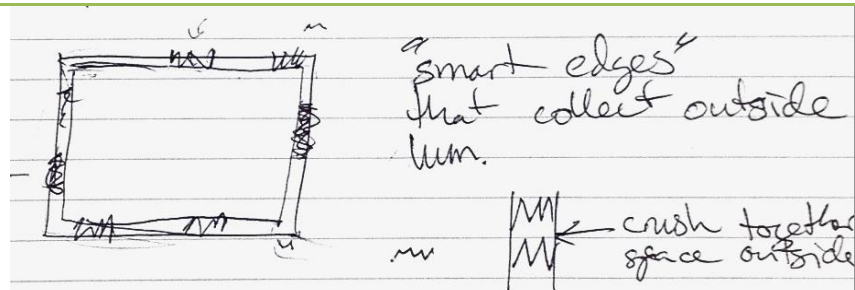
How do we visualize the zoom out/in of the region of the wall in order to get a feeling of how big and where the note is in relation with the wall?



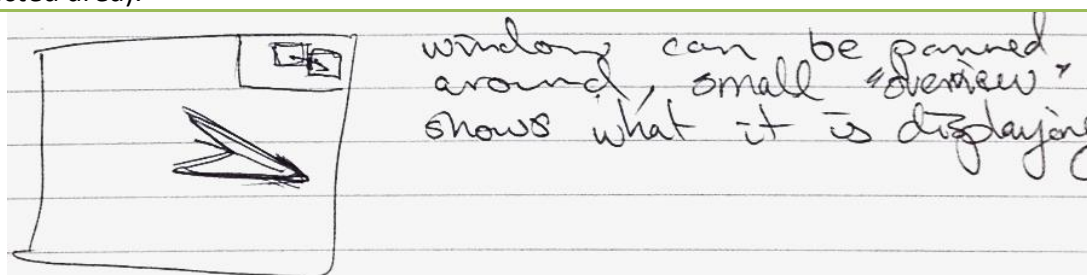
Show the full wall on the projected area and mark the selected rectangle.



Show in the beginning the full wall and after zoom in slowly. In this way the user should have a feeling on where the selected area is in comparison with the wall.



Show only the selected area on the projector, but on the side show the distance on the area which is not visible. (for ex: 2m on the right, 1m on top ... => represents the bounds of the selected area).

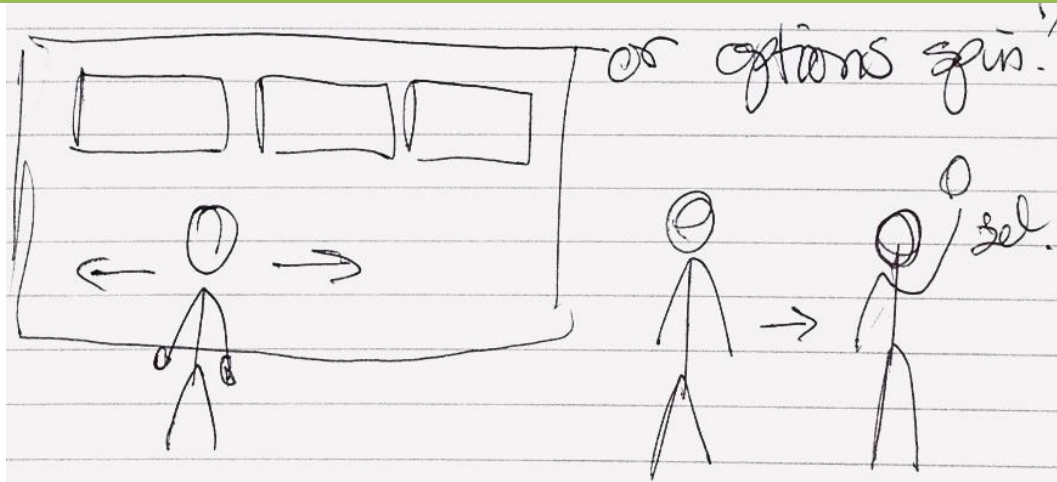


Show a small arrow which will point the direction where we made the zoom in (the position of the selected area in comparison with the whole wall).

Google Maps panning rectangle.

November 2013 Wed 20			November 2013 Sat 23	Use something similar to doodle (how the unimportant area gets folded).
7:00 AM - 9:00 AM	9:00 AM - 12:00 PM	12:00 PM - 3:00 PM	12:00 PM - 3:00 PM	
✓	✓		✓	
			✓	
✓	✓		✓	
			✓	

How can we select the visualization type?



The system will display the available visualizations as rectangles on the top of the projected area. The user should go and align itself on the X axes with one of the rectangles and stay there or a very short period of time.

Use the phone or the computer in order to quickly select the visualization type.

## Session 2

## Consent Form

<b>Purpose of Research</b>	<p>The research that we are currently conducting has to do with answering the following question:</p> <p><b>“What user experiences can we enable with the user activity history and how can they be visualized?”</b></p> <p>The user activity history described in the previous sentence refers to the movement of one or more people in front of a wall while working on an affinity diagram, the data being recorded by a Kinect sensor.</p> <p>We are interested in finding out what visualizations describe best the people which were involved in the affinity diagram; the changed areas and the correlations between an area and a user; all in a specified time frame.</p> <p>The information which we will gather from you will be used to optimize existing visualizations and to create new ones.</p> <p>Important keywords</p> <p><b>Affinity diagram, user movement data, Kinect, history.</b></p>
<b>Procedures</b>	<p>If you agree to participate in this study, you will be asked to do the following:</p> <ul style="list-style-type: none"> <li>• Participate in a preliminary interview (around 30 min) during which you will be asked about <ul style="list-style-type: none"> <li>▪ Do you understand what the visualization does?</li> <li>▪ Do you understand how to interact with the visualization?</li> <li>▪ Does this visualization offer the info you want?</li> </ul> </li> <li>• Grant researchers permission to audiotape the interviews, so that we may have access to your answers later.</li> </ul>
<b>Risks</b>	<p>There are no anticipated risks to you resulting from your participation in this research.</p>
<b>Confidentiality</b>	<p>The information that we collect in the study will be handled confidentially. Your information will be assigned a code number.</p>

**Voluntary Participation**

Your participation in the study is completely voluntary. You have the right to withdraw from the study at any time without penalty, and you may refuse to answer any questions you don't want to answer and still remain in the study. If you decide that you want to withdraw from the study, please tell the researchers.

**Audio Release**

Yes \_\_\_\_ No \_\_\_\_

I give consent for audio from my participation to be studied by the research team for use in the research project.

Yes \_\_\_\_ No \_\_\_\_

I give consent for audio from my participation to be used in reports and presentations about this research project.

**Questions**

If you have any questions about the research, please feel free to contact any of the researchers listed below.

Andrei Andriesi  
[andrei.andriesi@uzh.ch](mailto:andrei.andriesi@uzh.ch)

If you have any concerns about or objections to the research, you may report them to:

Andrei Andriesi  
[andrei.andriesi@uzh.ch](mailto:andrei.andriesi@uzh.ch)

**Agreement**

I have read and understand the above information on the purpose of the study, the procedures that will be followed, the use and confidentiality of the data, the risks and benefits, and the freedom to withdraw from the study.

I agree to participate in the research study described in these three pages.

**Signature**

Sign: \_\_\_\_\_

Date: \_\_\_\_\_



# Sketches

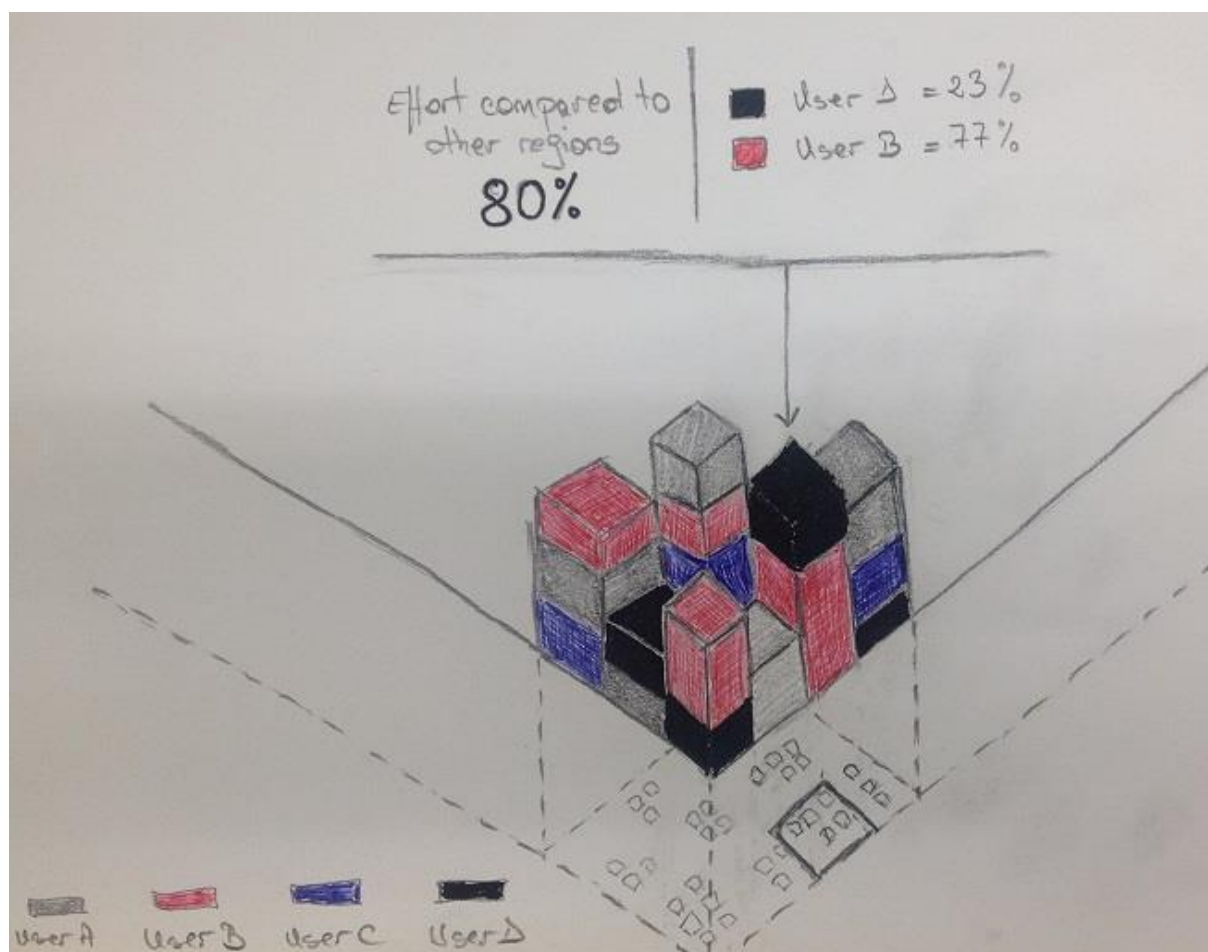


Figure 40 - Q3 / Session 2 / Prototype 1

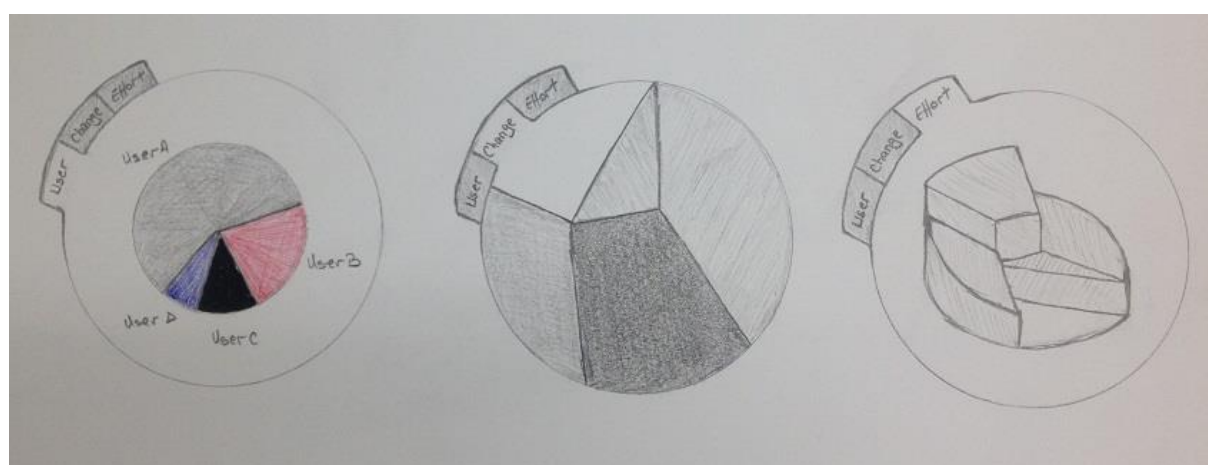


Figure 41 - Q3 / Session 2 / Prototype 2

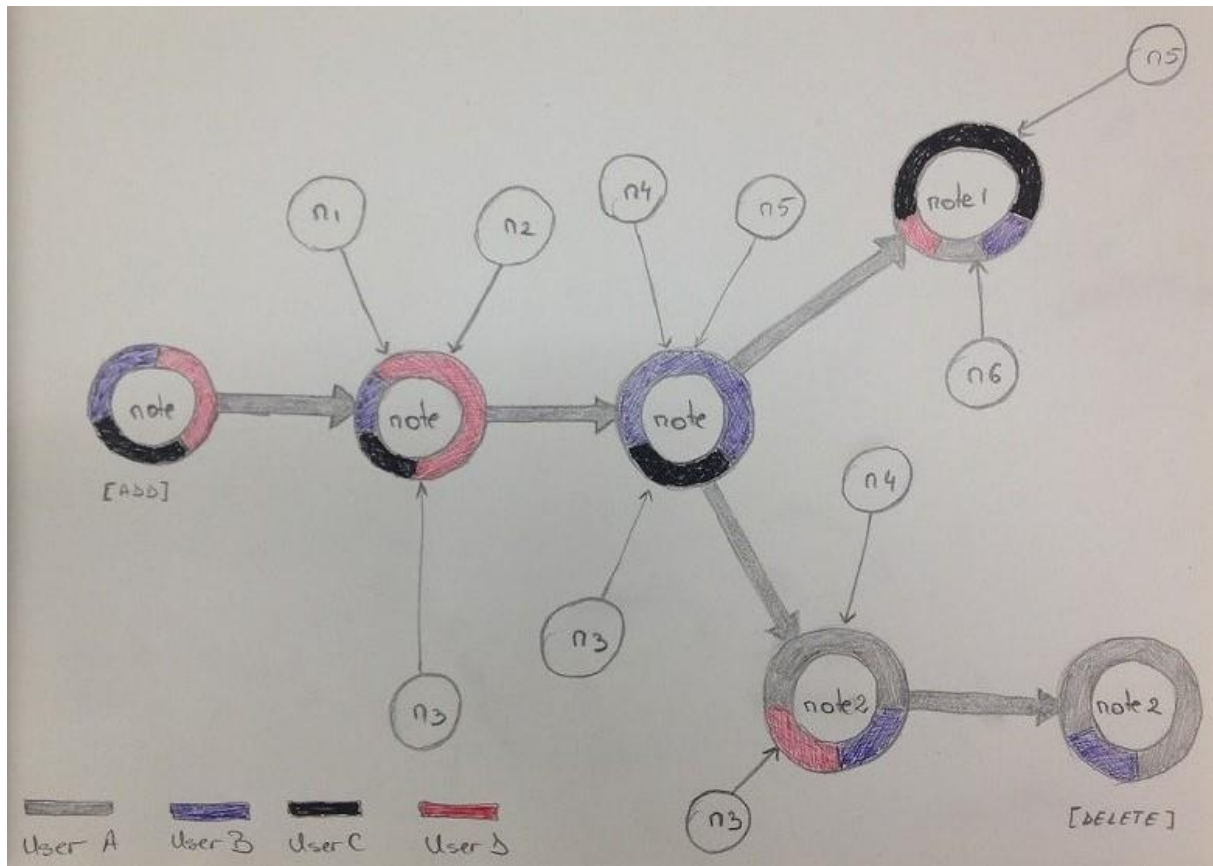


Figure 42 - Q3 / Session 2 / Prototype 3



## Findings

### Person 1

#### Prototype 1

- There are 4 users working in this diagram
- The layer underneath is the selected wall
- The column represents what the users changed on the diagram
- The size of the block represents how much the users work on that part of the diagram comparing with the rest of the diagram
- The diagram represents the change in the specified time frame
- P1 thinks there is a person outside which manages the diagram
- P1 didn't really know if there is a need of an interaction
- If P1 wants to see hidden areas then he will try to export the diagram in a format to better visualize the data, or to grab and drag a block in order to better visualize it. If not possible to move he will want to at least see the info about them floating above the diagrams.
- P1 wants to be informed about the hidden parts which are not so easily visible in the 3D diagram by just putting a sign on the top. If more info is required then P1 will just want to click on it.
- P1 expect the diagram to be arranged in a way that the important things are more in the front and the not important ones more in the back.

#### Prototype 2

- **D1**: shows the amount of time a person (user) worked in comparison with other users.
- **D2**: represents the change in the diagram.
- Why is D2 not colored?
- It is not clear what does each region represents.
- P1 does not understand how it correlates to D1.
- P1 thinks (not immediately) that the darker means more changes and while less.
- **D3**: is very unclear and it is not covering the whole diagram
- The size of the blocks in D3 is represented by the cumulated amount of work by all people.
- D3 does not show the relation with the affinity diagram ... that is very very unclear.

#### Prototype 3

- The first impression is that the small n1, n2, n3 are comments
- The diagram represents the correlation between notes
- The big arrow represents the evolution of a note.
- The split represents OR and not AND
- The color represents the users and the amount of color is the amount of work
- I don't understand why do I have to show delete

- There is a clear relation between the n1, n2, n3 ... and the colors
- Different users add different comments and there are displayed by the arrows from the "comments"
- The changes are based on a discussions
- I do not need to interact too much.
- If I need some more details I will try to click on a note.

### Feedback

- I would use revisions like in MS Word but only if there are not many changes
- I would show only the last version in a revision
- I would try to annotate the diagram like in MS Word
- I would display the changes in the diagram like in the worship game.

## Person 2

### Prototype 1

- It represents notes grouped together
- The size of the block is connected with the number of notes and the colors from the block are related to the users which added those notes to the group.
- The size of a color block is a number and not a percentage
- If I have 100 notes then the column (block) should be 100 times bigger
- P2 has no idea what is the region underneath the 3D shape.
- After more thinking P2 thinks the size of the block represents the percentage
- The blocks represent 100% all the time
- At the first guess P2 would not need any interaction.
- After asking about the blocks which are not visible, P2 said: I would try to put my hands on the wall and try to move the hands up and down and from left to right in order to rotate

### Prototype 2

- **D1:** amount of notes added by each user or the amount of changes done by each user
- **D2:** the amount of changes related to each region
- **D2:** the colors represent the group of note (I think). It is hard to understand the colors and the shape. Are the colors from D2 related somehow to users?
- Colors from D2 can mean also a type of change: white (added), black (remove) gray1 (moved from one place to another), gray2 (moved from another place to another).
- **D3:** somehow D3 is similar to D2 but only because of the missing part from the chart
- The colors represent again the: add, remove, or move from different places to others.
- D3 represents the effort base on the change types

### Prototype 3

- Is like a workflow or maybe a node is a child of another node.

- P2 does not realize that nodes are repeating
- N1, n2, n3 ... are notes but they are confusing why there are two types of notes.
- Is also confusing why we have note and note1, note2 ... why the first big nodes do not have numbers?
- Is like a history.
- The small arrows from the n1, n2, n3 ... are related to the colors from the big notes.
- The big arrow means that at one moment the note goes either in one direction or another.
- I do really need interaction.

**Interaction:** In general I would start doing moves and hope that I discover how to interact, or call a friend which has Kinect. I would try to simulate the moves from IronMan.

### Person 3

#### Prototype 1

- The blocks display how long did a person spent in front of a section of an affinity diagram
- P3 does not understand why is the diagram 3D when it can easily be 2D.
- Why isn't all the users display in each block?
- The height of the column represents maybe the cumulated effort in that certain area.
- When I see the blocks I think about excel charts.
- P3 has no idea how does a block relates to another block.
- The size of the blocks is not 100% even if it should be.
- Is hard to say how this is measured from the point of view of time.
- It needs to be clear how long the time frame is.
- P3 thought is a hologram and not projected.
- As interaction, P3 will try to touch the wall and move the hand from left to right hoping that the 3D shape will rotate

#### Prototype 2

- **D1:** looks familiar with excel and the size of the users is related to the effort that users give to that part of the diagram.
- **D2:** is very much not familiar and is represents the area where most of the changes are but I do not see the connection with the users
- The colors from D2 represent the amount of changes (single or cumulated ... is not clear) in that area.
- It should have a more clearly measuring unit.
- **D3:** why is the 3D shape smaller than the circle and how does it correlates with the affinity diagram? ... this is like a bird eye but I do not imagine how is connected with the affinity diagram.
- D3: how do I correlate the user and the changes?
- The only interaction is the tabs and I would want to use my hand to touch the area where the tabs are in order to change it.

### Prototype 3

- This is a tree of interactions with the selected node.
- N1, n2, n3 are other nodes which relate with the main nodes.
- The names do not mean much
- The arrows are maybe relationships or children-parent relationship
- The color represents the contribution of different users but it does not make sense why there are necessary more users.
- There's no relationship between the nodes and the colors. If I really think maybe I can find a relationship but it was not obvious in the beginning.
- The interaction: drag if is longer and tap on a node in order to have a fish view and have more details about the node (zooming effects).

### Person 4

#### Prototype 1

- P4 understands as the first bottom layer as being the selected part of the wall
- The colors represent which user was active in different parts of the diagram
- Users are arranged in the blocks based on the order how they worked (the bottom user works first, then the second and so on) : this is the first thought
- The height of the shape represents how active a person or a group of people are
- **Interaction:** I would try to rotate it to see the back side of the diagram (but this is only if I really think about this)

#### Prototype 2

- **D1:** shows the users which worked on the part of the diagram and the size of the pie shows how active a user was. P4 didn't consider the time.
- **D2:** is not very colorful but it does not matter
- The colors does not necessary mean something, but maybe it represents that the darker side is more often changed compared with the white one.
- Also looking at the size and at the locations of these blocks we can say that the black and white do not have any relationship between them and that the flow of papers between those regions is not that big.
- P4 thinks that everything has a meaning because is a diagram
- **D3:** is like a chart but it has different vertical layers (in this case 2). P4 does not understand what does the pieces mean, or at least is not very intuitive.
- In the end P4 thinks that each section represents a user because there are 4 pieces and 4 users. Why is a part missing? (This is a bit irritating).
- Why is a part missing? : Maybe because 4 users work on the diagram but they did not work 100%.
- **Interaction:** I would want to take the parts apart somehow and analyze them.

### Prototype 3

- The colors represent the users which moved the note
- The size of the color represents that the user X moved more the note comparing with others
- The thick arrows represent the movement of a note from position A to position B on the wall.
- The n1, n2, n3...do not say anything, is hard to understand.
- The split of the "note" in "note 1" and "note 2" represents a split from the text of the note which resulted in deleting the note and creating 2 different notes.
- **Interaction:** not so much interaction is needed

### All cumulated (most important information)

#### Prototype 1

- Block size: time, number of notes, effort, how active a group is
- Colors: the users
- Interaction: put hand on the wall and rotate

#### Prototype 2

- **D1:** easy to understand and it represents time, amount of notes added by an user, effort, how active a user is
- **D2:** is not very logical at first and no connection with users:
  - dark mean more changes
  - dark can mean a group, a category or a type of movement
- **D3:** very confusing. Why is it 3D? What's the connection with the users? How's related with the affinity diagram?
  - Cumulated amount of work
  - The colors and sizes can represent different category: add, remove, move
  - Maybe it represents users and how much they worked; but why as 3D?
- Interaction: press on the tabs (1 user)

#### Prototype 3

- Small notes are: comments added by users, note
- Big notes are: evolution of the same note
- Big arrows: evolution, history, child-parent relationship, move from posA to posB
- Split: OR, AND
- There is a relation between the small notes and the colors because of the arrows
- Size of color based on: discussion
- Interaction: click on a note, drag, zoom

[illegible]

## Images &amp; Sketches

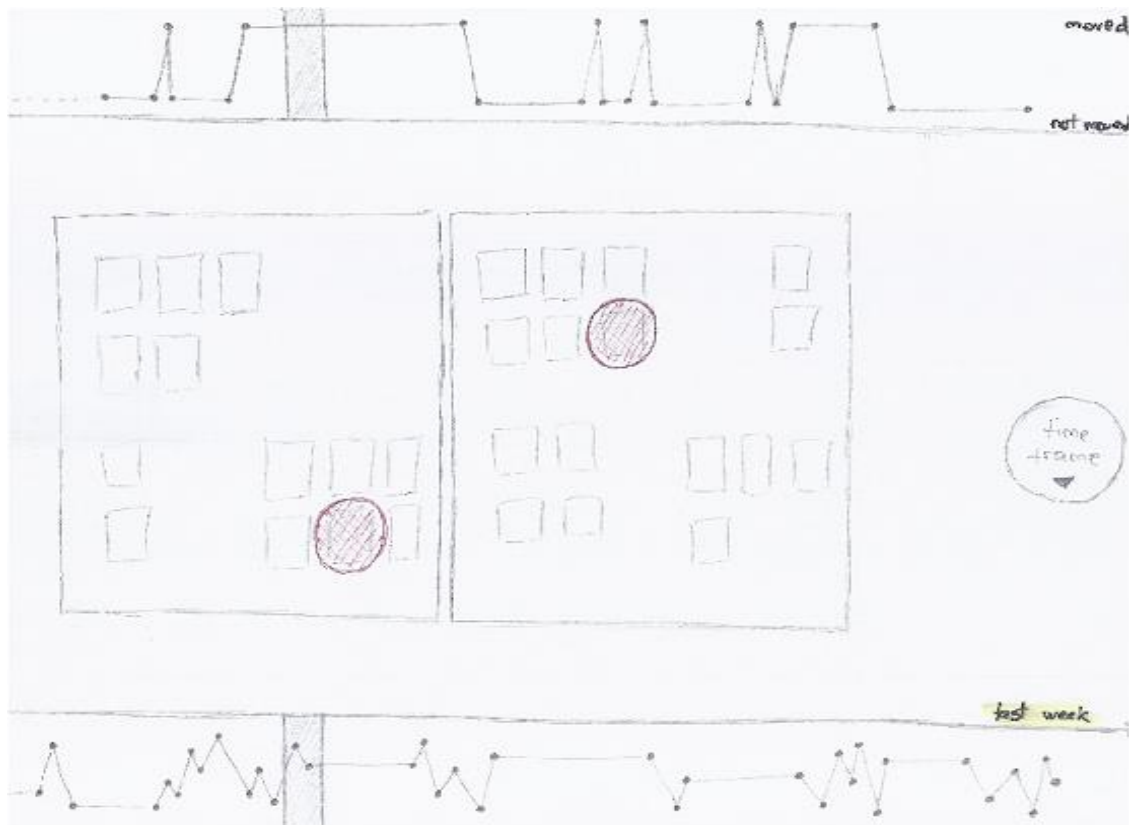


Figure 43- Q3 / Session 3 / Prototype 1

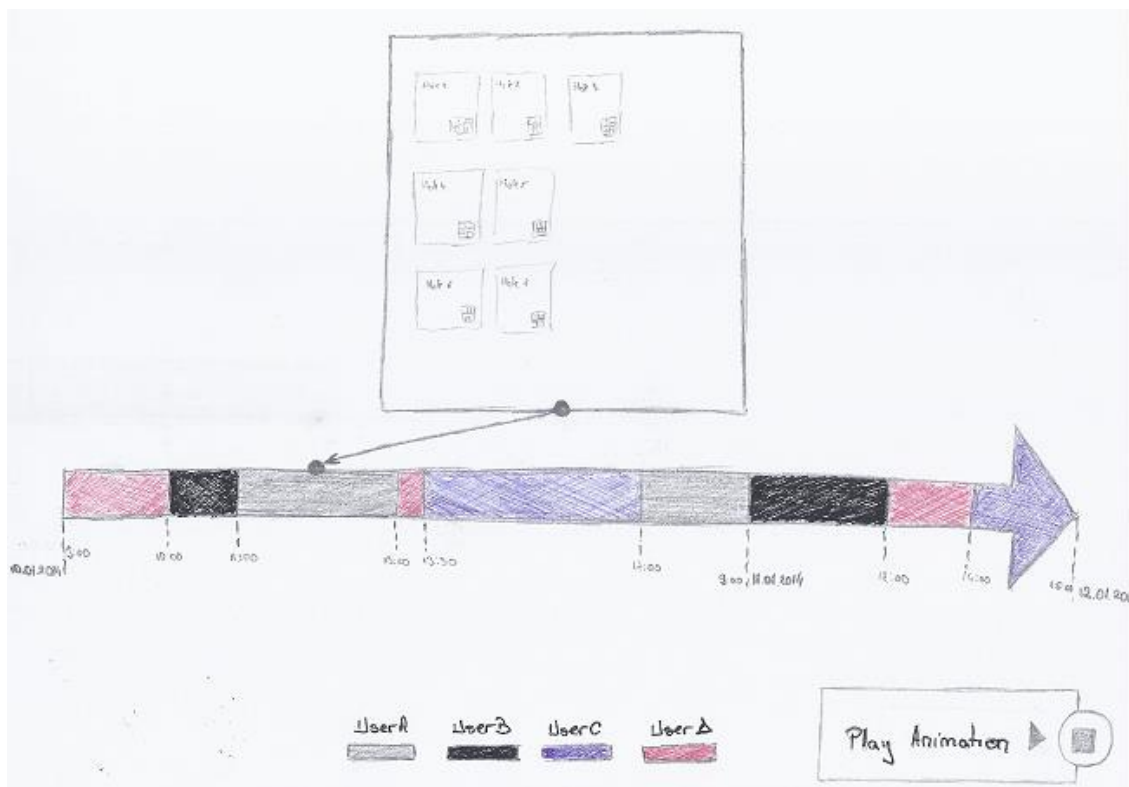


Figure 44- Q3 / Session 3 / Prototype 2

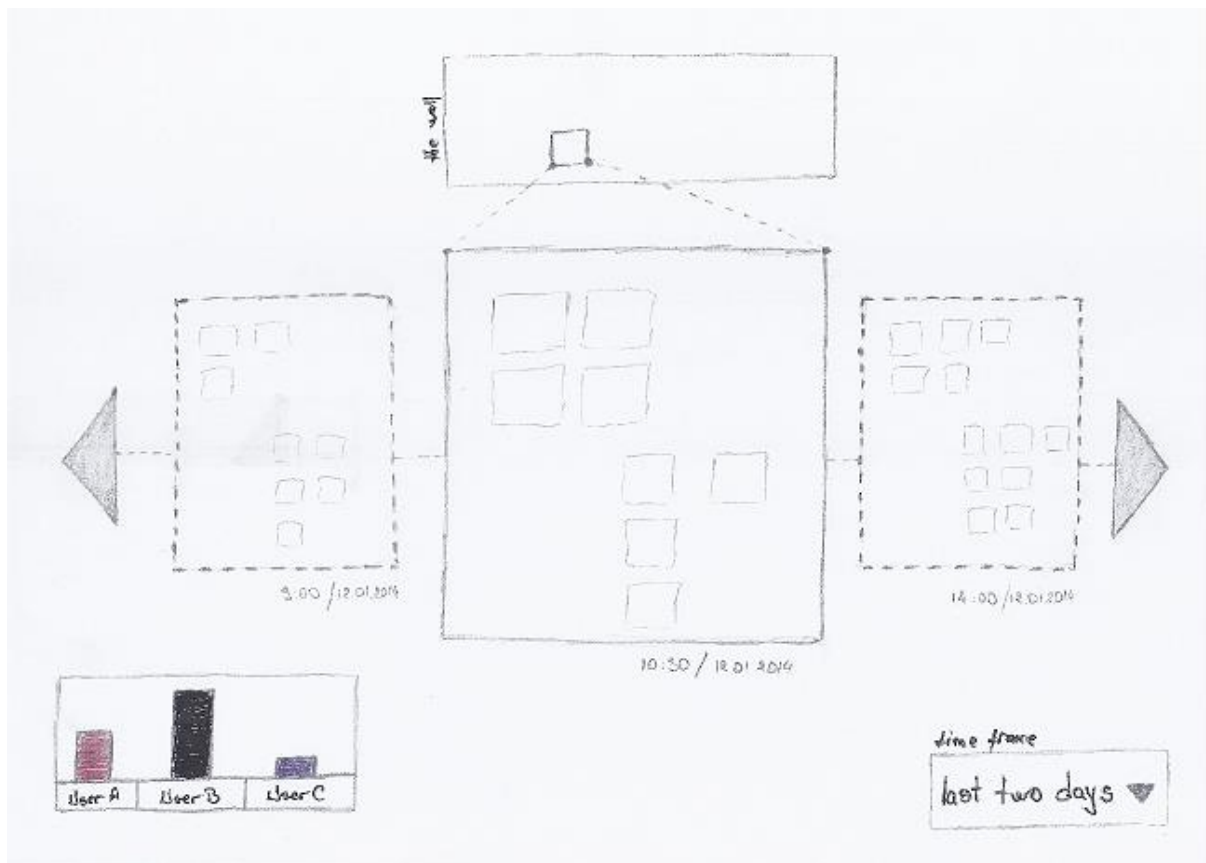


Figure 45- Q3 / Session 3 / Prototype 3

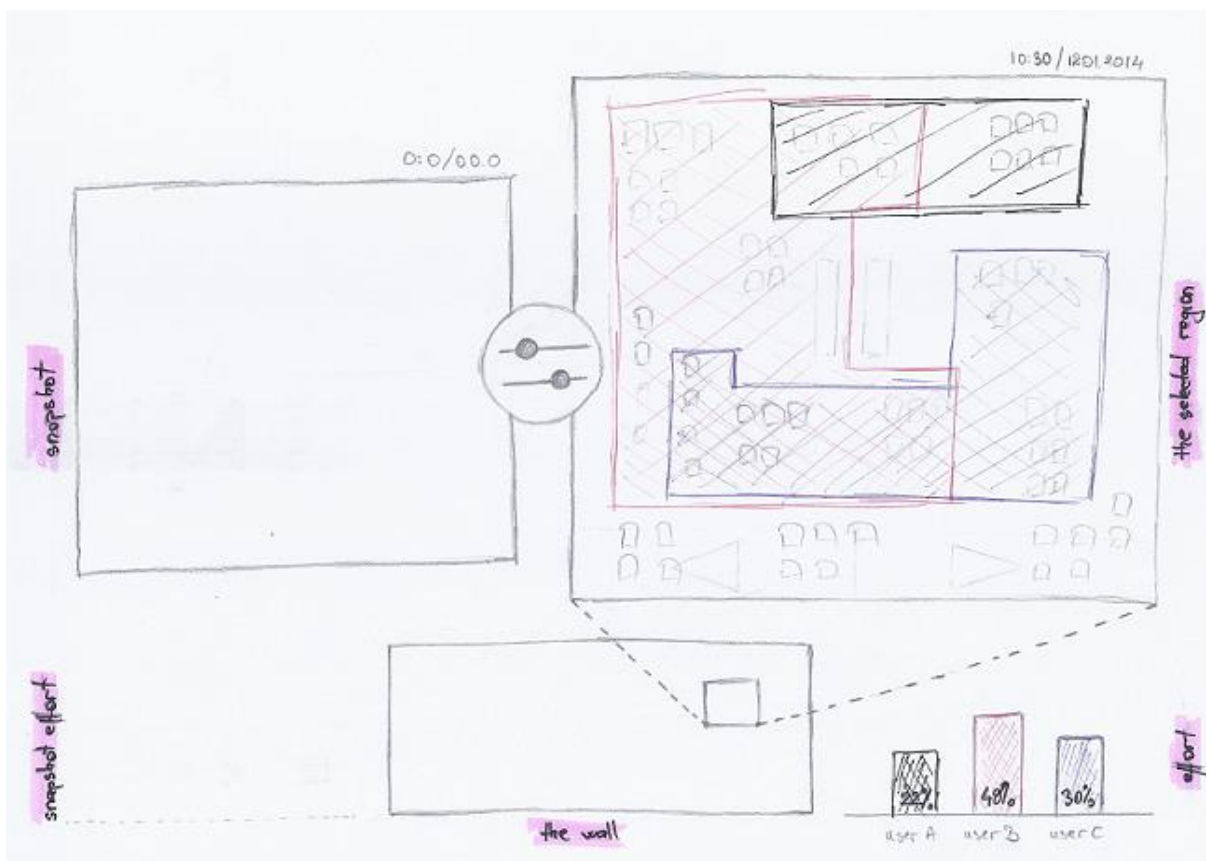


Figure 46- Q3 / Session 3 / Prototype 4



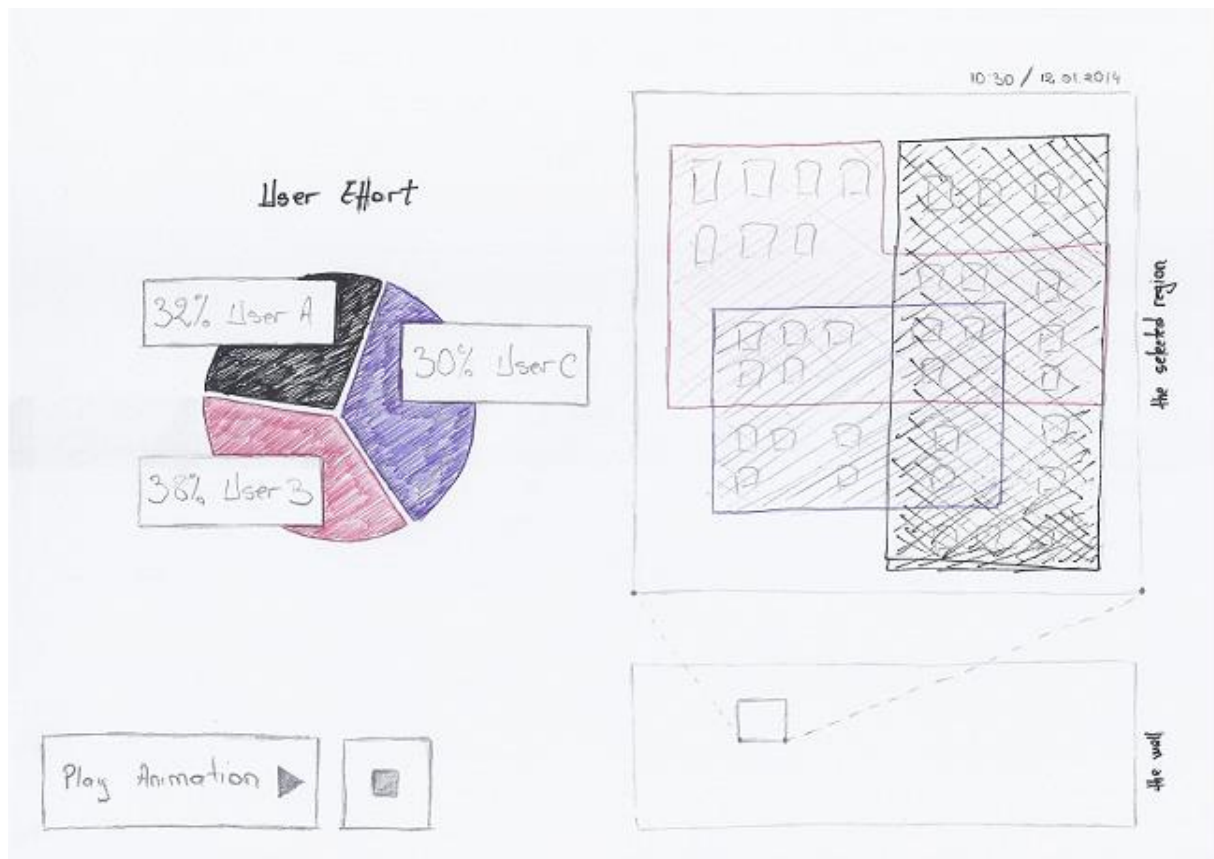


Figure 47- Q3 / Session 3 / Prototype 5

## Findings

The results presented below are based on **5 prototypes** and are cumulated from **6 users**.

No. Users	Sketch No.	Problem Description	Heuristic	Severity
6	1	Not sure what the timeline effort means.	5	4
4		Not sure what does the timeline "note" mean.	5	4
4		There is no label on the axis and in this way people don't know what they mean.	1	2
4		Unclear how "effort" scale works.	3	2
3		<b>Unclear what does "user effort" mean.</b>	4	3
2		The two visible points in the effort bar and the two screens don't have any visible connection.	1	4
2		No way to get an overview of the whole wall?	5	4
2		It's not clear what the position of the dots in the effort diagram means.	4	3
2		<b>User A/B/C. What is it for? Who are they?</b>	3	2
1		Why is the timeline note a binary but displayed as a line chart?	5	4
1		It is not clear that User A changed the data more comparing with User B.	2	4
1		The user bar looks more like a legend and is not intuitive.	4	4
1		I don't know why there are 2 snapshots.	2	4
1		I don't know why there are 2 red circles and I don't know what they mean.	2	4
1		I don't know what does time frame mean and I don't know how it relates to the other charts.	2	4
1		I don't know how and when to use the time frame bullet.	2	4
1		Unclear what part of the wall we're looking at.	1	4
1		Doesn't show what time we're looking at.	2	4
1		Can only see one step at a time.	5	4
1		Interactive parts not evident.	1	3
1		Needing to manually compare two frames for differences.	2	3
1		Labels are not self-explanatory. What do they mean?	4	3
1		<b>Effort and Note texts on the vertical alignment are harder to read comparing on writing them on the horizontal way.</b>	5	2
1		What if no note is selected or if the user wants to select multiple notes?	1	2
1		How can you tell who moved the note on the screen?	1	2
1		What does a long period of moved mean on the "note" chart?	5	2
1		So many different elements and so many views in one screen	3	2
1		The label "moved" and "not moved" might be more helpful if they were placed next to the axis label (note).	5	1
4	2	It looks like only one user worked on the affinity diagram at a moment	3	3

		in time.		
1		Not sure what I'm looking at? Is this a colored timeline?	4	4
1		This doesn't look like it tells me very useful info.	2	4
1		The arrow connecting the picture with the wall is unfamiliar.	4	3
1		Is there any other interaction besides pressing play/stop?	1	3
1		How do you know who the users A/B/C are?	2	3
1		<b>What can I do here, what is interactive?</b>	2	3
1		I'd say that the "play animation" button is inconveniently placed for the physical interaction.	5	3
1		I'm not sure what should I do there? Only press "start animation"?	1	3
1		This looks like a YouTube video of an affinity, not very interactive.	2	1
2		What are the user bars? No legend or axis labels.	2	2
1		Not sure what "the wall" means in relation with the other squares.	4	4
1		Doesn't visualize changes?	2	4
1		Some people did not understand the concept behind picking a region of the wall and visualizing the data about that particular region.	2	3
1		The user legend shows the information of which frame?	5	3
1		The dotted lines from the corners of the middle screen are confusing in this case.	3	2
1	3	It is not very clear that the visualization shows the same snapshot of the diagram but in different moments in time.	2	2
1		You can't tell the ownership of notes and what did the users do because no colors are present.	1	2
1		Not sure what timeframe does.	4	2
1		Confusion related to the use of the "time frame" combo box.	5	2
1		Some people tend to think that "the wall" rectangle is a progress bar.	1	2
1		<b>The label "the wall" should be rotated because is earlier to read.</b>	5	1
1		What can I do here, what is interactive?	2	1
3		Buttons blend into screens and are not intuitive. (the ones on the selected region)	2	3
3		Unclear what does "snapshot" and "snapshot effect" represent.	1	3
3		The central big circle and its icon are confusing. Can I interact with those lines? What does it do?	4	2
1		You are assuming that the users move notes only in one small area, but they can move notes from and to very far locations.	5	4
1	4	What is "snapshot effort"?	5	3
1		Manually comparing the snapshots.	5	3
1		Not sure what color does the overlay shows.	4	3
1		People don't know if they can or can't interact with "the wall" rectangle.	4	2
1		The "effort" bars are showing the effort for 10 min, 1 week or how much?	2	2
1		The arrows on the selected region and the pause button are unclear.	2	2

		What do they do?		
1		Rotate the labels on the horizontal view.	5	1
1		Pause control may be unnecessary.	3	1
2		The colors/regions will get much cluttered if users don't always work in continuous regions.	5	2
1		Why do I need to know the effort in %?	4	2
1		Play animation button is not good placed for a right handed person.	5	2
1	5	The colors are very important because of the color of the real notes on which the visualizations will be displayed.	3	2
1		"Effort" gives a good or bad judgment. Maybe is better to use a different word.	4	1
1		What is the timeframe for the pie chart?	5	1
1		<b>You don't need two separate buttons for play/stop; play could become stop and the other way around.</b>	3	1
General Questions / Notes			Heuristic	Severity
		Is this something a user needs to see?	5	4
		<b>The projection on the wall will interfere with the actual diagram.</b>	3	4
		System status is never visible.	1	4
		<b>The interaction is hard to understand.</b>	4	4
		I would like to see an animation regarding who touched what but how do I visualize it if no coloring is used?	4	4
		How will the transition between the notes be visualized?	1	4
		Not clear how to access this screen and how to get out.	1	3
		Selecting an area is unclear.	1	3
		<b>No clear connection between the wall and the visualization.</b>	5	3
		Why is it important to know the user effort?	4	3
		<b>What's the reason to project on the wall if we don't interact with the physical wall?</b>	5	3
		<b>Many screens are confusing.</b>	3	3
		Do you interact with the overlay or with the physical notes?	4	2
		If someone spends a lot of time in front of the wall but doesn't do anything, is that a high effort?	5	1
		The heat map should also show how much a person worked in a specific area and not only the area itself. Use transparency.	4	1
		The overall design is hard to understand without a context.	ALL	
		If people are really near to the wall then the image can be hidden because of shadows.	ALL	

## Main areas where issues were found

- User needs vs provided information
- Environment (light, room size, projected area, things on the projected area)
- Show current status
- User interaction
- Real vs virtual elements
- Displaying / hiding the visualization

## Main (general) challenges found in sketches

- Concepts like “user effort”, “note” and “wall” are hard to understand only from a label.
- Time stamps and other texts should be more obvious to the user.
- If graphs are used then labels and legends should be present.
- The link between graphs and other elements should be obvious.
- The UI should be obvious to the user and little instructions of how to use it should be given.
- The UI should be clean and clear without too many graphical elements which distract the user.
- The UI shouldn't have too many overlaying elements.
- The UI should adapt to the current situation.
- Interactive elements should be positioned so that they are easily accessible to the user.
- Interactive elements should be obvious.
- Provide a way to see the overview of the whole wall.
- Don't force the user to do things manually.

## Session 4

### Sketches

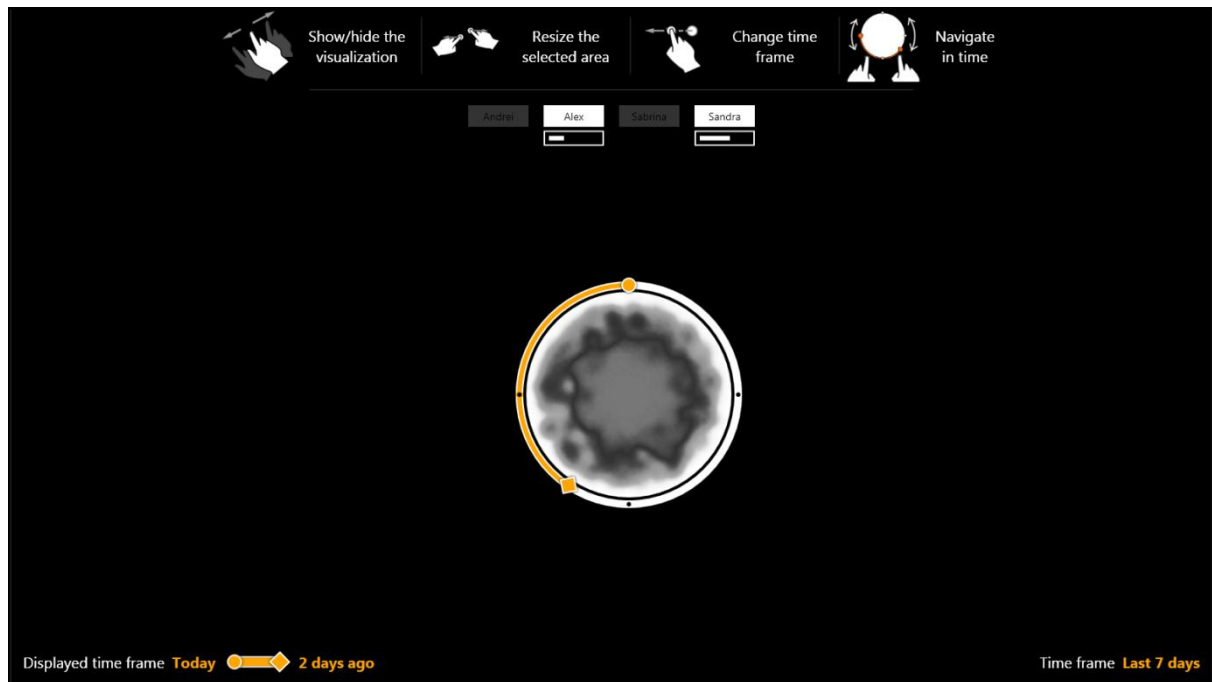


Figure 48- Q3 / Session 4 / Prototype 1.a

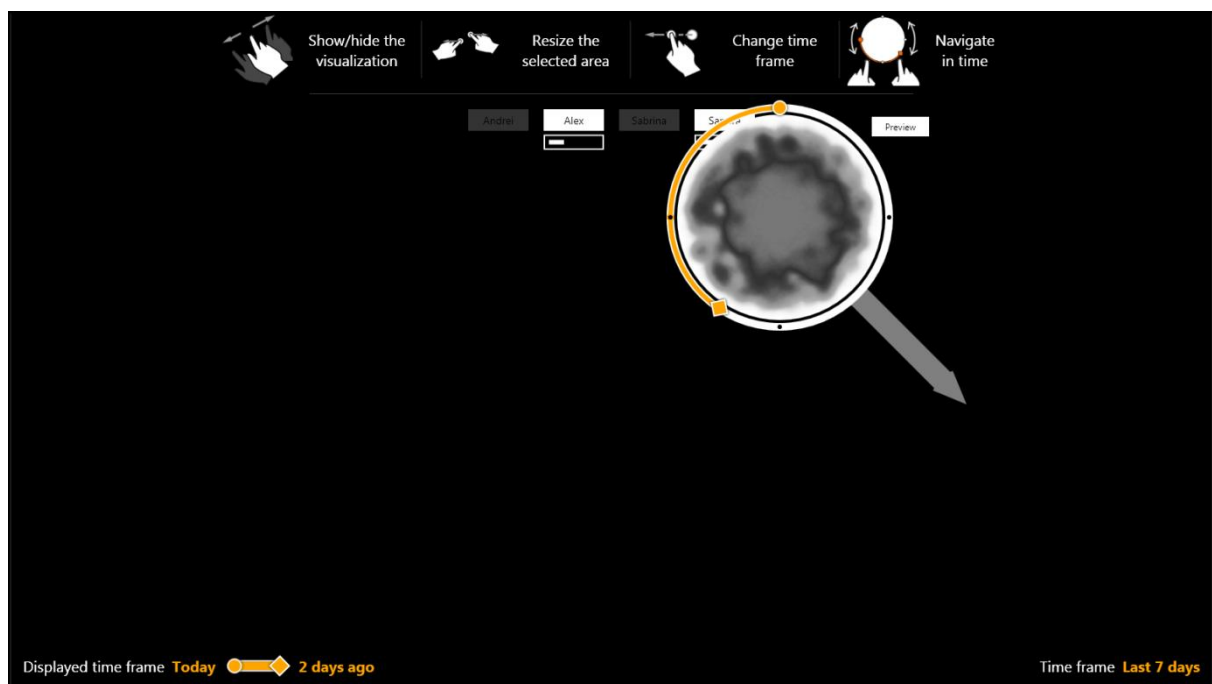


Figure 49 - Q3 / Session 4 / Prototype 1.b

## Findings

The presented results are gathered from **6 people**.

### Gestures

- I don't understand the "resize selected area" gesture.
- "Wave" gesture unclear.
- Gestures for navigation in time and resize area seem easily confused.
- No gesture for "select region".
- No "move visualization" gesture.

### Labels

- I don't understand why there are two "time frame information". What's the difference between "time frame today" and "time frame"?
- "Preview" label is unclear.
- Label "show / hide visualization" is confusing.

### Shape

- If you try to expand the visualization when the selection is outside the projected area, you will end up in occlusion problems.

### Visualization

- More detailed info about type of changes not available.
- Link between users and heat map is not clear.
- Scale for timeline unclear.
- I can't easily link a specific change with a specific user.
- Heat map doesn't distinguish users, can't filter users.
- Heat map seems to imply activity and not presence.
- Maximize visualization to show full screen rect.

## Session 5

## Evaluation Sheet and interview protocol

## User activity visualization (UAV)

*Interview protocol*

## Questions

1. Which (top gesture) icons do people relate to and which not?
2. Do people know how to use (simulate) the provided gestures?
3. Do people observe and interpret correctly all labels?
4. Do people recognize on the screen the selected time frame and the total available time frame?
5. Do people understand the selected time frame and the total available time frame?
6. Do people observe the visual feedback (time label and rectangle rotation) when changing the selected time frame?
7. Do people observe and understand the progress bar underneath the user name?
8. Are people able to correlate the user progress bars with the visualization (colored “heat map”)?
9. Will people focus on a specific region or prefer seeing the entire wall?
10. Are people able to map the off-screen selection point with the circle visualization and the pointing arrow?
11. Do people understand the meaning of opacity and transparency in the visualization (colored “heat map”)?

## Observations

1. Do people use the provided gestures consistently correct?
2. What is the order of the gestures people are doing when faced the first time with the system?
3. Do people get confused when working with the system?

## Bring to the interview

- Two copies of the consent form
- Copy of interview protocol

## Interview Protocol

- Briefly describe study and thank participant for participating
- Briefly go over consent form and collect signature
- Ensure participant that you want to learn from them
- Present use case
- Continue with the questions and write down the observations

## Purpose of research

The research that we are currently conducting has to do with answering the following question: **“What user experiences can we enable with the user activity history and how can they be visualized?”** The user activity history described in the previous sentence refers to the movement of



one or more people in front of a wall while working on an affinity diagram, the data being recorded by a Kinect sensor.

We are interested in finding out what visualizations describe best the people which were involved in the affinity diagram; the changed areas and the correlations between an area and a user; all in a specified time frame.

### Describe the use case

Suppose that you (user named “Sam”) are an HCI expert and you are currently working together with your other 4 colleagues on an affinity diagram for an internal project. The diagram is located on a wall and towards facing the wall are a projector and a Microsoft Kinect sensor. The affinity diagram (the notes on the wall) was started 7 days ago and you usually work a few hours per day on it. During the last 2 days you were in vacation and as your colleagues continued to work on the diagram, now it looks completely different compared with as you left it. Because you are unsure on one hand about the changes which were done during the last days and on the other hand about who did specific changes, you are using the UAV system in order to visualize all this information. Once the system is on, the visualization will be projected on the wall (overlapping partially the affinity diagram). The interaction with the system is done NOT using a computer but instead, using natural hand gestures which are interpreted by Kinect. Because Kinect is behind the person, any gesture needs to be clearly executed with the right hand, the hand being situated on the far right side of the body. In order to start the system one needs to only “wave the hand”.

### Gestures

Can you start by describing me what gestures you think the system will recognize?

- Enumerate the gestures
- Simulate the gestures
- Observe if the user does every gesture correctly and write down how they do it normally.

### Displayed information and interaction

- Please tell me what information you see on the projected area.  
*Probe for: labels and their meanings and for the recognition of familiar metaphors (circles, taskbars, rectangles, watch and heat map).*
- Think back about the presented use case. You just opened the visualization and now you would like to see what happened in the last days.
- What would you do?
- How would you do that?
- What information do you see?  
*Probe for: user and progress bars, the opacity and transparency of heat map*
- Consider observing other areas of the wall inside the projected area. How would you do that?  
*Probe for: touching the wall, resize the circle, maximize the circle*
- Consider observing other areas of the wall outside the projected area.
- Is the visualization intuitive?
- Is the arrow helpful?
- Is the interaction with the visualization intuitive?

## Images

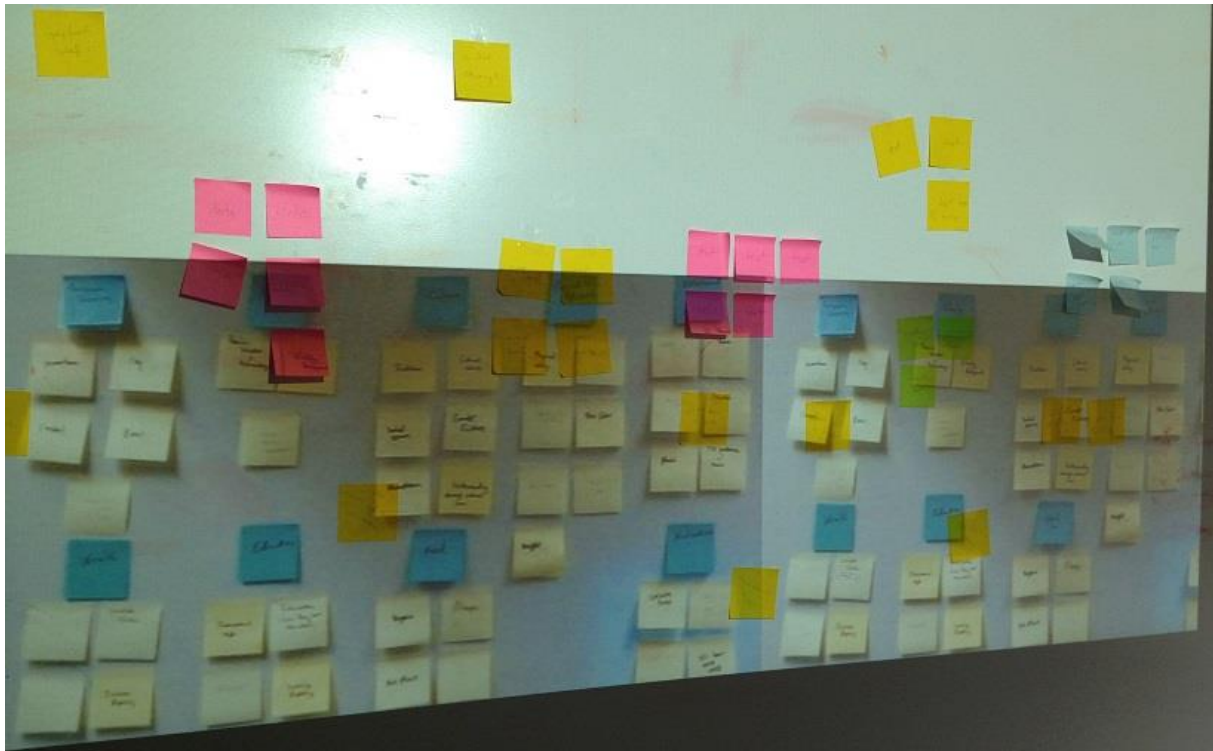


Figure 50- Q3 / Session 5 / Notes overlapping test

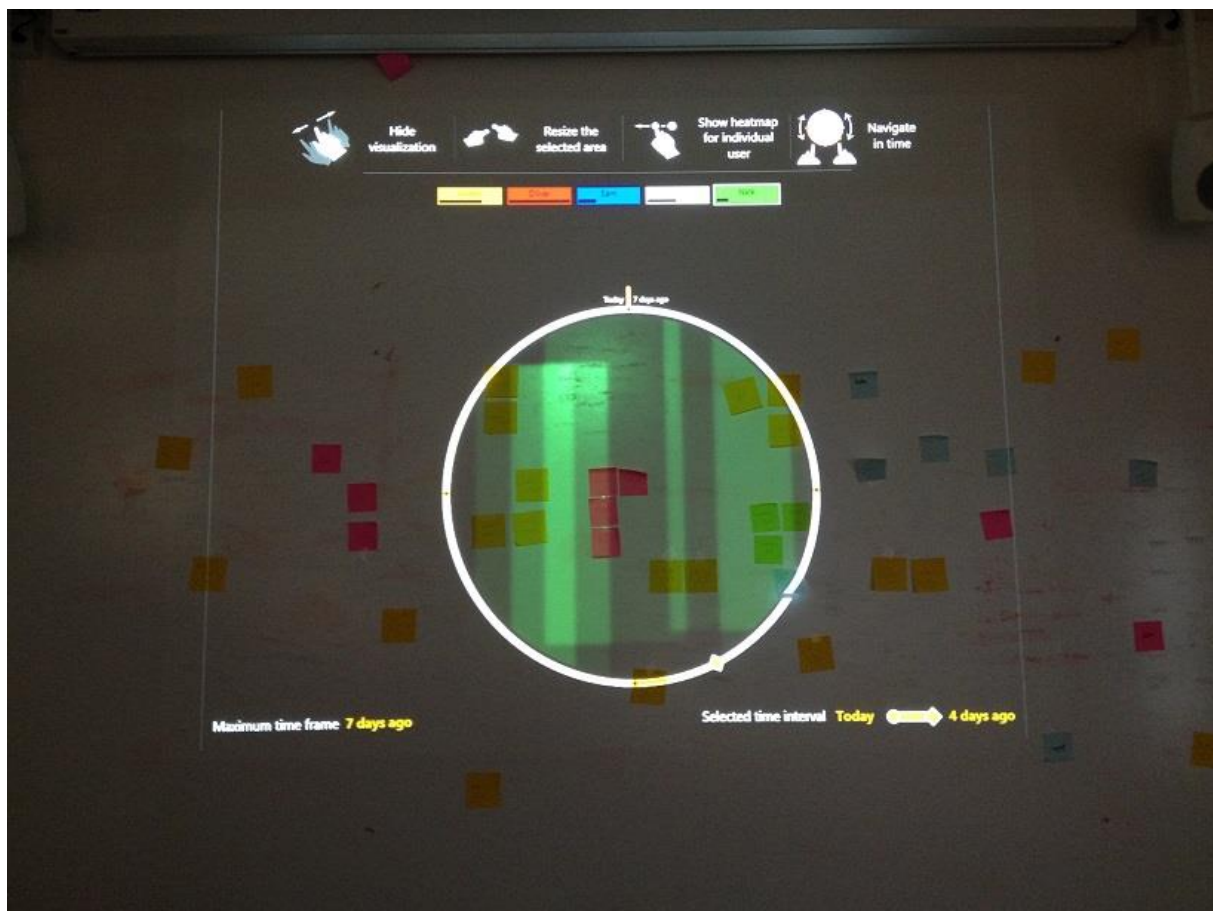


Figure 51- Q3 / Session 5 / Prototype 1.a

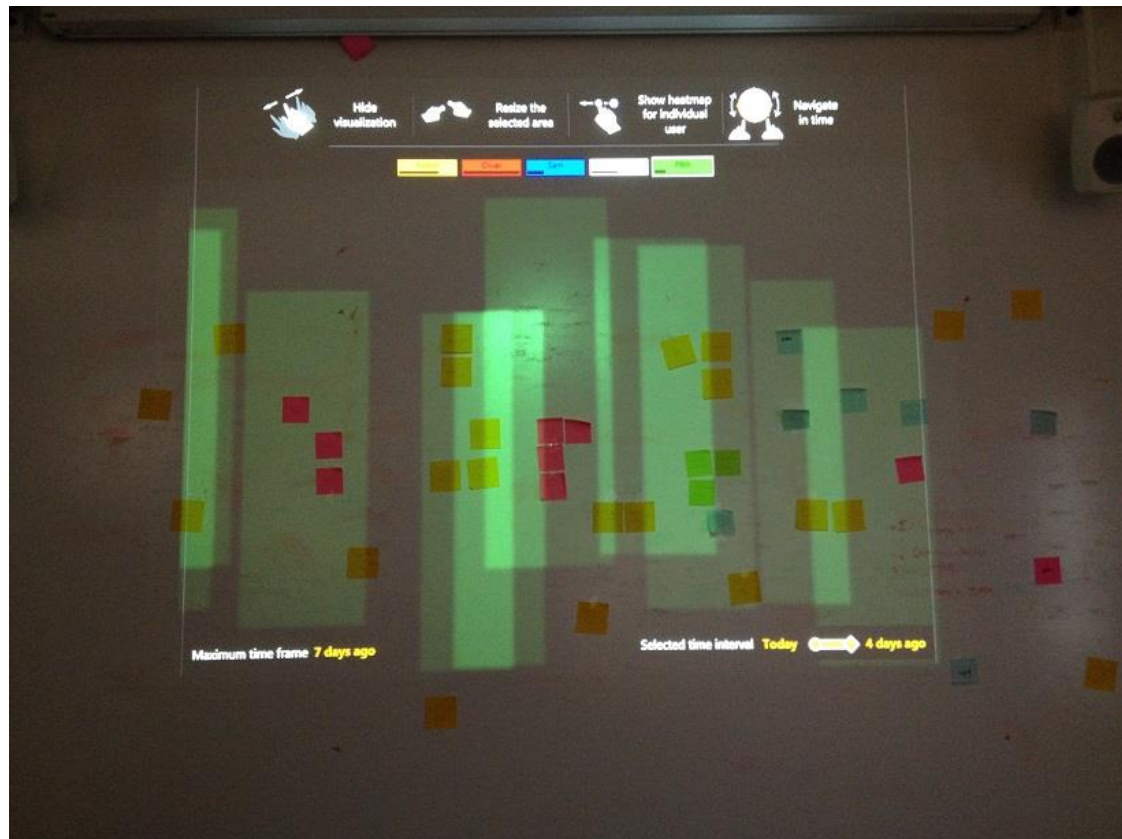


Figure 52 - Q3 / Session 5 / Prototype 1.b

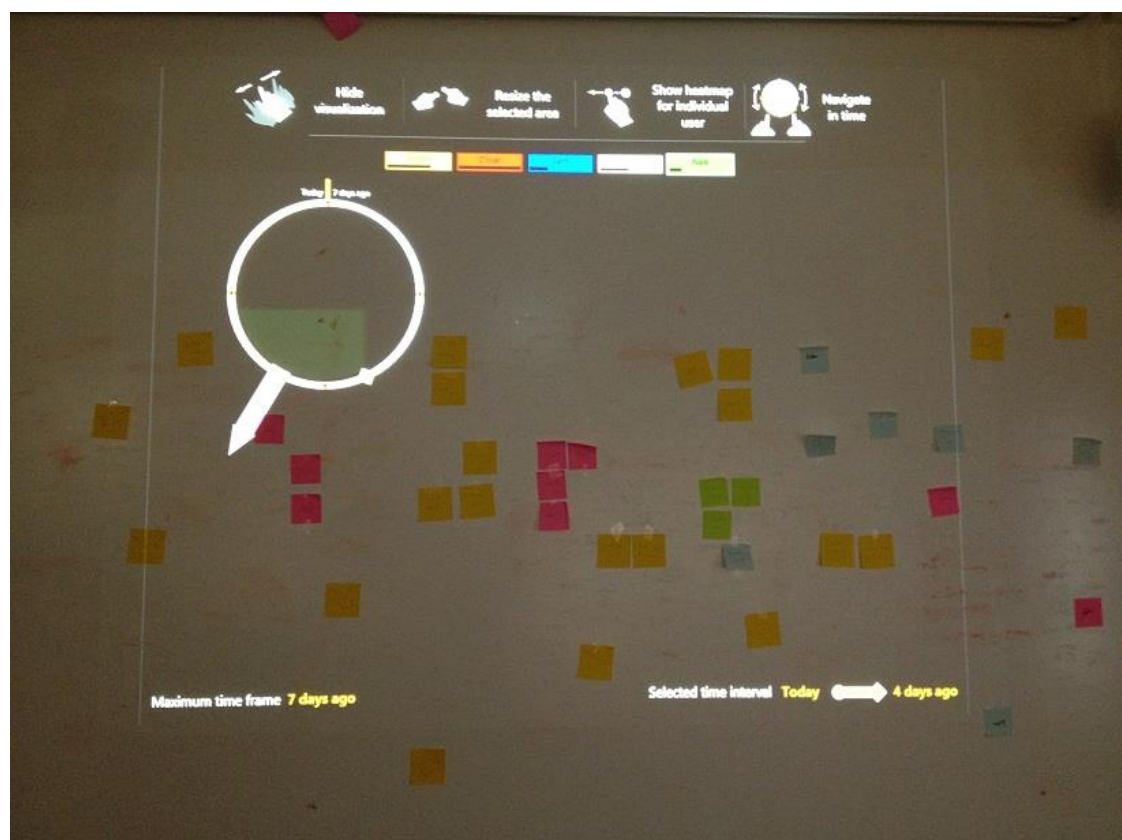


Figure 53 - Q3 / Session 4 / Prototype 1.c

## Findings

### Gestures

- Because of the wrong icons people do not understand the meaning of the gestures in the beginning.
- People tend to tap on the wall and not necessary use the gestures from the distance as a normal Kinect application would work (this needs to be learned).
- The list of gestures on the top of the projected area was not all the time being noticed.
- Even after the correct gestures were explained, people still continued to go at the wall and interact with the visualization as they would do with a normal computer and not using Kinect gestures.

### Information

- The labels on the bottom of the projected area were not all the time being noticed.
- Both focused and full-screen is needed.
  - “I don’t want to be distracted by regions which I don’t care but in the beginning I may want to have an overview of the whole wall.”
  - “My eyes would get tired because of too many colors and I would not use the visualization too much if I have to only use the full screen mode.”
- Cycling through people is not very intuitive; almost all participants wanted a way of selecting the desired user.
- Depending on the light conditions in the room, the heat-map can be partially hidden.
- The orange color of the track-bar situated on the border of the circle should be changed. One user believed that the orange color was connected with the orange color of a user.

### Solve the given task

- People used all given features in order to solve the task (move and resize the circle, change time frame, cycle through people and analyze differences, switch between the wall and full-screen visualization).

### Arrow

- Some problems in understand the purpose of the arrow. Some people understood it from the beginning but others needed more time to think. A label or some explanations may be required. In one case, it happened that the arrow was pointing to a specific note, this made the person believe that the visualization was somehow connected with that note.

## Statutory Declaration

I declare that I am the only author of this thesis and that I did not use any other sources, resources or materials other than the ones specifically declared and marked.

Date

-----

Name / Signature

-----