VRIJE
UNIVERSITEIT
BRUSSEL

Graduation thesis submitted in partial fulfilment of the requirements for the degree of
Master of Science in Applied Sciences and Engineering: Computer Science

AN EXTENSIBLE PRESENTATION FRAMEWORK FOR REAL-TIME DATA ACQUISITION, INTERACTION AND AUGMENTED REALITY VISUALISATIONS

KWINTEN PARDON
Academic year 2017-2018

Promoter: Prof. Dr. Beat Signer
Advisor: Reinout Roels

Science and Bio-Engineering Sciences

VRIJE
UNIVERSITEIT
BRUSSEL

Proef ingediend met het oog op het behalen van de graad van
Master of Science in de Ingenieurswetenschappen: Computerwetenschappen

AN EXTENSIBLE PRESENTATION FRAMEWORK FOR REAL-TIME DATA ACQUISITION, INTERACTION AND AUGMENTED REALITY VISUALISATIONS

KWINTEN PARDON
Academiejaar 2017-2018

Promotor: Prof. Dr. Beat Signer
Begeleider: Reinout Roels
Wetenschappen en Bio-ingenieurswetenschappen

Abstract

Although it is one of the most used presentation tools, PowerPoint is often criticised for the shortcoming and limitations it imposes on the presenter. For instance, movement of the presenter is restricted because non-trivial interactions with the presentation content have to be initiated through the presenter's computer. Besides limiting movement, the opportunities for interaction are limited and as a result, PowerPoint enforces a linear presentation of relatively static content. Research has been conducted in an effort to resolve some of these shortcomings, but public speaking and presenting is a complex multifaceted field. Some related work focusses on improving one of the many aspects, but there are no practical solutions that address the problem as a whole.

We start from the fact that existing presenter views are too limited in terms of mobility and functionality. We present a new presentation framework with the goal of better supporting the presenter during presentations. The framework allows data from different applications to be gathered and visualised during the presentation, and also offers new methods for interacting with the presentation or other applications. This allows not only the presentation's content to be part of the presentation view, but also data generated by the audience, or for instance by a real-time presentation feedback system. The framework is based on requirements and findings derived from related work and was made to be generic and extensible to allow the integration of future interaction devices and new data sources. As a proof of concept we also present a new type of presenter view based on augmented reality, allowing the presenter to remain mobile while simultaneously offering flexible widget-based visualisations and interactions not visible to the audience.

Declaration of Originality

I hereby declare that this thesis was entirely my own work and that any additional sources of information have been duly cited. I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this thesis has not been submitted for a higher degree to any other University or Institution.

Acknowledgements

First and foremost, I must acknowledge my parents and their never ending support, guidance and advise throughout the entirety of my life. I'm grateful for all the opportunities they have given me and for the never-ending guidance which has lead me here. Thank you for your support, advice, guidance and for pushing me to explore my capabilities.

Furthermore, wish to extend my gratitude to the people closest to me, who had to bear with me through my frustrations, celebrated along with my accomplishments, and supported me through the entirety of my studies.

Gratitude are also in order for the student group InfoGroep and the friends I was honoured to meet there. It also gave me the opportunity to broaden my studies by exploring technologies not covered in courses and the possibility to use new found knowledge in real life applications.

Last but definitely not least, I wish to thank my promoter Prof. Dr. Beat Signer and my advisor Reinout Roels for their support and feedback for the entire duration of my thesis as well as for the received guidance and advise.

Kwinten Pardon
June, 2018

Contents

1 Introduction

1.1 Preface

PowerPoint is a commonly used presentation tool. PowerPoint has made an enormous impact on the way presentations are given nowadays and it has reduced the complexity and the time required to prepare a presentation by a significant amount. Coming from overhead projectors and photographic slide projectors which could take weeks to produce to a digital solution where the same presentation can be created in a matter of hours [2].

However, over time shortcomings have surfaced and many have critised Powerpoint and its use. PowerPoint did not become an extension of the presentation and has taken a more prominent role as being the presentation. PowerPoint has also made the presentation dull and predictable as it reduces the presentation to a sequence of bullet points lacking further explanation or motivation [55, 21]

The limited functionality of the presenter views in existing tools can be considered as one of the causes. By providing better tools, presenters could maintain a better grasp on their content and the relationship between presenters and audience members could be improved for more efficient knowledge transfer. For example, a better presentation view could help presenters navigate their content better, interact with their content more efficiently and also integrate data that is not produced by the presentation tool it self. It could help to increase audience participation and audience interaction by providing a way to visualise collected feedback as well improve the presentation in general by providing a way to be notified of bad practises and providing information on how to correct them.

1.2 Current State

PowerPoint[1] is a commonly used presenter tool which runs on the computer or laptop of the presenter. Figure 1.1 shows the presenter view. The main element on the presenter view is the current slide which is being displayed to the audience. To the right of the current slide, the next slide is shown. Underneath the next slide, notes related to the current slide and written by the presenter as aid is displayed.

Underneath the current slide being displayed, there are five buttons the presenter may use. From left to right these buttons are:

- Pen interaction: Writing on the slide as well as highlighting certain words. This feature works with any device that acts as computer mouse input such a mouse or a touch screen.
- Opens the slides overview as displayed in Figure 1.2. This allows for the selection of any random slide.
- Zoom button which allows to zoom in on a certain element on the current slide.
- A button to blacken the screen that displays the presentation.
- More options button which does not offer any mentionable functionalities.

The use of these functionalities requires the user to use a computer mouse or touch screen provided they have a touch screen on the device they are presenting with. This limits the ability of the presenter to move around. If the presenter wants to use any of these functionalities, they are forced to return to their laptop. The interaction with the presenter view also requires some focus on the screen which results in a loss of eye contact with the audience.

At the bottom of Figure 1.1, the navigational buttons are displayed. The presenter may navigate to either the next or the previous slide by pressing these buttons. The presenter may also use the right arrow and left arrow key

[1] https://products.office.com/en/powerpoint

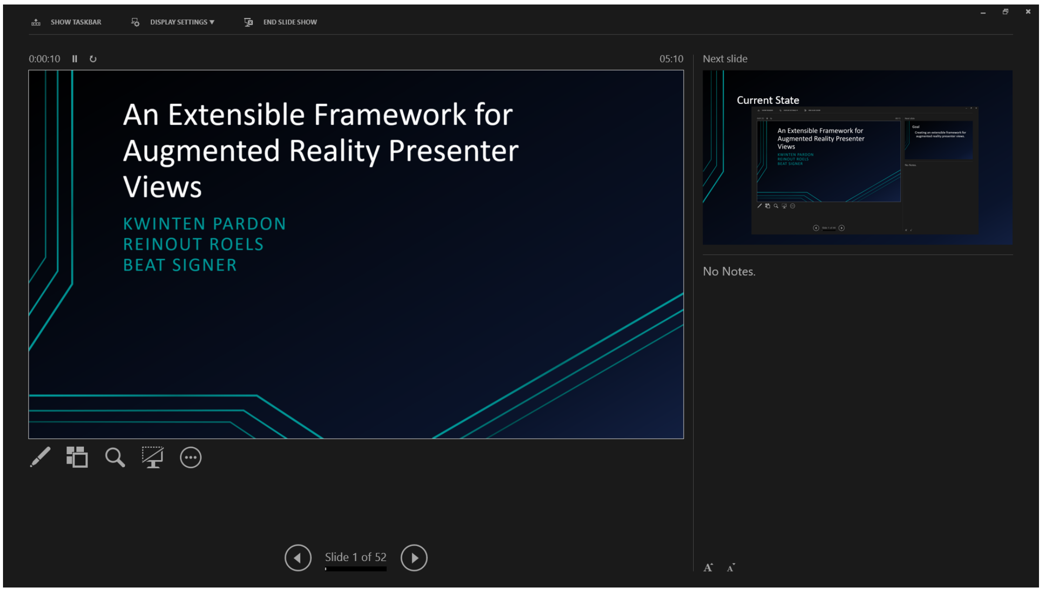

Figure 1.1: The PowerPoint presenter view

for these interactions as well as the page up and page down keys. Without the use of a wireless interaction device which mimics those key presses, the presenter is once again forced to remain within close proximity to their laptop.

An overview of all slides within the presentation can be shown within the presenter view as depicted in Figure 1.2. This allows for a random slide to be selected. However, this view requires the presenter to manually scroll through the overview and select the slide they wish to present. Due the nature of interacting with this overview and the effect it has on the presentation, one should frequent use of this overview in order to keep a smooth presentation. This implies that the feature can not be used to skip slides frequently causing the presenter to be stuck with the slide order and excess of slides for the entire duration of the presentation. This may be unfortunate when the available time has been reduced just prior to the start of the presentation.

Besides restricting movement, the use of the PowerPoint may have other negative aspects. Having the laptop reside on a lower surface may have a negative aspect on the overall posture of the presenter as they will have to bend over to be able to interact with their presentation. When the laptop resides on a surface at eye level, it may block the view of the audience breaking eye contact or negatively influence the use of gestures by blocking the audience view or restricting the space of the presenter to safely use gestures.

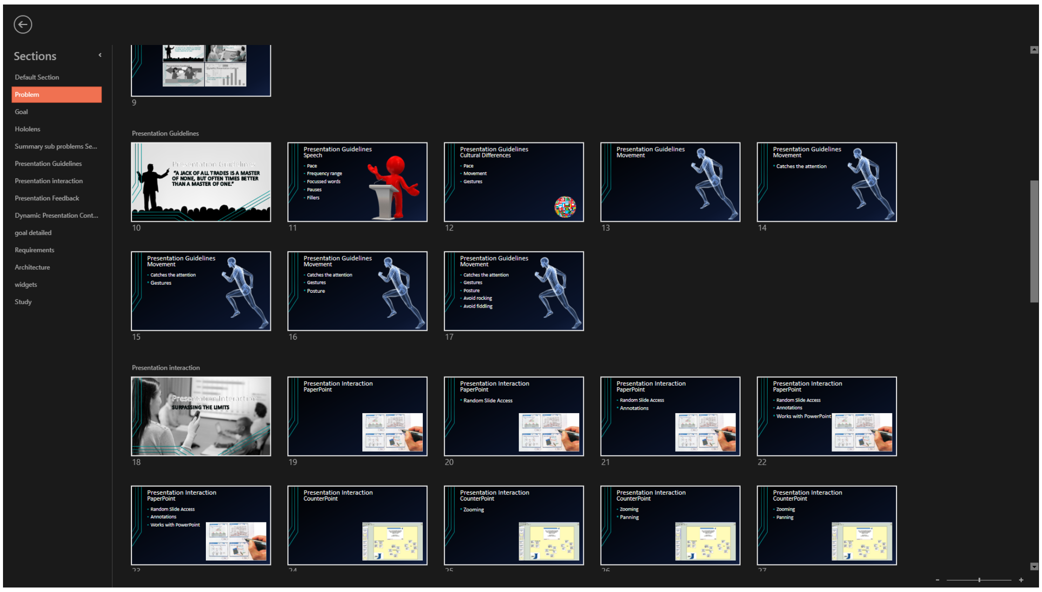

Figure 1.2: Slides overview view

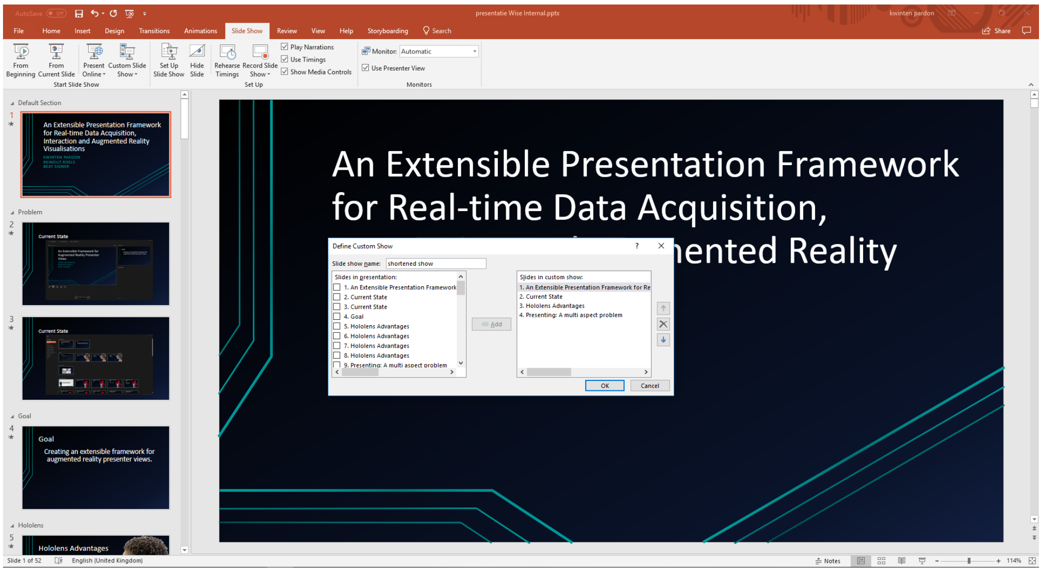

Figure 1.3: Interface for custom slide show within a PowerPoint presentation

PowerPoint allows custom slide shows to be created from the same set of slides within an existing PowerPoint presentation. As depicted in Figure 1.3, no real graphical user interface exists for the creation of a slide show. Slides have to selected through the name of the section they reside in. The overlay window used to create the custom slide show blocks any interaction with the main window making it impossible to watch the slides while creating a custom slide show.

Another issue with PowerPoint is that all content is static. It is not possible to integrate a chart in the slides or any raw data to be visualised. The presenter is required to create the visualisation or chart in a different tool and take a screenschot of it or export it as an image in order to use it on a slide in PowerPoint. The drawback of this is that the presenter is not able to change the chart during the presentation to answers questions about the influence of a certain parameter more visually. The presenter would be required to close his presentation and open a different tool to adjust the visualisation.

1.3 Problem Disquisition

Now that the current state has been discussed, we will continue by discussing presenting and public speaking in general and a more in depth introduction of the different aspects which are all important with regard to public speaking. In total, four main aspects have been identified which will be discussed in their relative sections. These four aspects are:

- presentation guidelines
- presentation interaction
- presentation feedback
- dynamic presentation content

1.3.1 Presentation Guidelines

Presentation guidelines are the rules, so to speak, that help presenters give a more successful presentation when applied. These guidelines range from the way the presenter should act to how they should speak. Knowing the psychology behind presentations and the effects of different behaviour helps to prepare the presentation and to ensure the correct message is conveyed to your audience [38].

The guidelines are being discussed in order to compare them with the offered possibilities and imposed limitations by the commonly used presentation tools. It is also important to have knowledge about the guidelines in order to design a solution which enables the presenter to follow these guidelines and deliver an overall better presentation.

It is important to note that the guidelines differ depending on your audience. The guidelines are culture dependent which means a different set of guidelines is to be used when giving an presentation in an Swedish environment as opposed to an Arabic or an American environment. A difference between public speaking in an Arabic culture and an American culture is that in the Arabic culture you may not come over as desperate while in the American culture, a bigger focus lies on not coming over as being ordinary. This has been shown in a study across these cultures in a comparison research of charismatic speech [10]. It is important to be perceived as a charismatic person towards your audience when giving a presentation as it enables the presenter to more easily influence them. Depending on the culture of your audience, a different approach should be used to reach this goal.

It is also important to remain calm for the duration of the presentation. Anxiety may have a negative effect on the presentation as it may affect your speech. Anxiety may cause the presenter to start talking at a faster pace which makes it harder for the audience to understand the presenter. It may also cause the presenter to lose focus causing unexpected pauses during the presentation which may divert the audience from the message one tries to convey. Another effect that it may have is stumbling over words and repeating oneself making it harder for the audience to remain focussed or to follow the story. When struck with anxiety, people also tend to play with their hands among other nervous ticks which may distract the audience [39].

Another important aspect of a presentation is the content of the slides. Slides do not always add value to the presentation and may even have a negative effect on the presentation when improperly made. Slides often contain too much text which distracts the audience from the presenter as they are focussed on reading the text writtin on the slides. The audience may miss important parts of the presentation said by the presenter as they are focussed on the text and not the speech of the presenter. Another problem that may arise is when the slides do not contain enough text to fully convey the message which keeps the audience distracted as they ponder about the meaning of the slide. Some presenters also neglect to focus on the content of the slide in an effort to impress the audience with the slides for example through the use of animations and colours. These slides are often harder to remember due to the absence of a clear structure within the slide show [15, 55].

Body language is another major aspect contributing to the effectiveness of a presentation and the way it is perceived by the audience. Body stance is a form of non-verbal communication and may unintendedly influence the audience in a wrong way. Giving the presentation while taking on a defensive posture, for example crossing your arms, is a way that may influence your audience. The crossing of the arms is a defensive posture showing insecurity towards your audience and shielding yourself off. As a presenter one should employ a body language which is inviting towards the audience. Besides body posture, hand gestures have an equal important role during presentations as they help the audience to better understand the message you are trying to convey. These hand gestures transfer feelings to the audience. An audience may become enthusiastic or passionate about a certain subject because the presenter who is talking about it is perceived to be enthusiastic or passionate [47]. However, one must remain cautious when using different hand gestures in different cultures as they may have different meanings depending on the culture. Some gestures may have a positive or encouraging meaning in some cultures while being perceived as insulting in other cultures. For example, thumb up in most western cultures is seen a sign off approval and agreement while the same gesture in the Middle East is considered offensive[2].

As mentioned before, speech is of the utmost importance when giving a presentation. It is the main way of communication while all other attributes are resources to enhance and improve the presentation. It is important to speak clearly at a pace fast enough not to bore the audience but yet slow enough for the audience to comprehend and follow what you are saying. Another important aspect is the volume at which the presenter speaks. The presenter must make sure to speak loud and clear enough for the entire audience to hear the presentation. It is also important not to speak too loud or to shout as it may be uncomfortable for the audience and may distract them from the actual message [47, 39, 30].

When looking back to the current state as discussed in Section 1.2, it becomes noticeable that these tools indeed are too limiting to take into account all the guidelines. The commonly used tools make it harder for the presenter to walk around the stage during the presentation as a close proximity to their device is required. Depending on the location of the device, it may also be hard to remain a good posture. For example, when the keyboard and mouse are located on a lower surface forcing the presenter to bend over.

Looking at the guidelines one might get overwhelmed by the amount of rules and details one must take into account. Mastering public speaking means mastering each element of the guidelines from design to delivery. It

[2]https://blog.busuu.com/what-hand-gestures-mean-in-different-countries/

becomes apparent why both practice and feedback about ones performance during presentation becomes important. Presentation feedback will be further introduced in Section 1.3.3.

1.3.2 Presentation Interaction

Under interaction we define two different kinds of interactions. On the one hand there is the interaction between the presenter and their presentation. On the other hand there is the interaction between the audience and the presenter. However, within the related work in Section 2, interaction with the audience will be discussed within Section 2.4.1 as part of dynamic presentation content because of the way how audience interaction is achieved by dynamically changing the content of the presentation based on their input.

With slide show interaction we mean interactions such as annotating slides or navigating through the slide. Navigation could be a linear operation of going to a predetermined next or previous slide without any viable alternative as is the case with PowerPoint.

Audience Interaction

With interaction between audience and presenter we mean feedback from the audience returned to the presenter. Feedback may be useful to let the presenter know whether or not the presentation was clear. The difficulties surrounding audience interaction are comparable with the difficulties surrounding public speaking and anxiety. Both are a result of social phobia and a fear to harm ones own reputation. An audience member may not come forward to ask a question out of fear that the majority of the audience may find the question to be unnecessary.

The most commonly used way of interaction between a presenter and their audience is through holding up hands. Audience member are asked to raise their hand to signal the presenter if they want to ask a question. Alternatively, a voting system where the presenter may ask a question to audience in an attempt to verify if the audience understood the given explanation, are most commonly done through a show of hands. One may answer this question dishonestly out of fear for consequences to their reputation by being the only who raised their hand, indicating that they did not fully comprehend the explanation [6, 42].

Audience interaction can also be achieved through *Classroom Communication Systems* which are further discussed in Section 2.4.1 as a part of dynamic presentation content. This is due to the influence dynamic content

Figure 1.4: A photographic slide projector
https://www.freeimageslive.co.uk/free_stock_image/slideprojectorjpg

has on audience participation in group discussion. *Classroom Communication Systems* allow the audience to vote without group pressure which may cause people to follow the majority vote. The visualisation of these answers may then contribute in reaching more audience interaction through a group discussion about the results.

PowerPoint does not offer any tools to enhance or improve audience interaction. Audience interaction is also gained through other tools which when used would break the presentation view. As discussed in the next section, PowerPoint does not offer a way to skip slides with ease. Therefore, it becomes impracticable to change the order of the slides depending on audience input. As the audience may not see any impact on the presentation as a result of their input, it may be hard to convince them to partake in polls or group discussion as the flow of the presentation is predetermined and static.

Slide show Interaction

Slide show interaction is any interaction the presenter wishes to perform on their presentation. A common interaction is navigating through the slides. A problem with common presentation software is how they are based on mechanical devices such as reversal film projectors and photographic slide projectors, depicted in Figure 1.4. Common presentation tools therefore only

Figure 1.5: An overhead transparency projector
https://www.bhphotovideo.com/images/images1000x1000/Buhl_Projectors
_9013EDC_9013EDC_Open_Head_Overhead_356205.jpg

have an easy way to navigate to either the next or the previous slide. Random slide access often requires the presenter to close the preventer view and search for the slide required. Random slide access is not uncommon as it is useful to answer questions raised by the audience. A natural way of random slide access also allows the presenter to reuse the presentation on occasions where time is more constrained. This is because the presenter would be able to skip slides or sections of slides without the audience knowledge. At the same time, the slides would still be available when required e.g. due to a question of an audience member to go more in depth on a certain topic [55, 49].

Another interaction is the annotation of slides. The presenter may wish to highlight certain elements of the slide. Annotations of slides is possible but often awkward as it requires the use of the mouse and keyboard. It also limits the freedom the presenter has since he can hardly step away from their computer during the presentation as he will no longer be able to interact with their presentation [49]. Annotation of slides was possible when overhead projector, seen in Figure 1.5, were being used in combination with transparencies. The presenter was able draw on the transparency during their presentation to enhance the slide or to put the focus of the audience on certain elements.

Figure 1.6: A pen tablet
https://imagesna.sslimagesamazon.com/images/I/61Ucag7VxTL._SL1200_.jpg

As discussed in Section 1.2, the current commonly used tools, such as PowerPoint, only have a limited array of interaction functionalities. PowerPoint does offer a pen functionality to write on the current slide. Writing however has to be performed using the computer mouse which may be hard and awkward to write properly. Also free hand drawing such as drawing a circle around a keyword may be hard to achieve through a computer mouse. PowerPoint also offers a highlighter but with the same limitations of using the computer mouse.

One may choose to use a pen tablet, depicted in Figure 1.6, as an alternative for the computer mouse. However, The Weight of the tablet may have an impact and wear down the presenter. It is not a lightweight solution and therefore may be hard to carry around for the entire duration of the presentation.

The random slide access by PowerPoint is, although available, not implemented in a way which allows slides to be skipped with ease without breaking the flow off the presentation. It does not close the presenter view for the audience but looking for the right slide may prove to be tedious for the presenter. It may also negatively affect the audience who have to wait without being able to see any changes or progress.

1.3.3 Presentation Feedback

As already discussed in Section 1.3.1, the way a presentation is given directly affects the effectiveness of the presentation and research has been performed to figure out the requirements of a good and effective presentation and guidelines have been written out to help the presenter prepare their presentations [38].

However, public speaking and presenting is an art, difficult to be mastered. To make it more complex, it is a combination of different aspects one must grasp in order to completely master the art of public speaking. It becomes clear that practice and training is required to improve the creation and delivery off better presentations. As discussed in Section 1.3.1, different aspects are important when giving a presentation such as speech, body language and posture and each of these aspects should be reviewed and trained.

Research has proven feedback to be important in a learning envirnoment and being a great influence in progressing. The way feedback is given remains equally important as it may also negativly impact ones skill. This thesis uses the definition of feedback given by Hatti, John "*information provided by an agent regarding aspects of one's performance or understanding*" [25]. Research has also shown three major questions with regards to feedback. These questions are:

- "*Where am I going?*" [25]
  What is the end goal of the trainee?

- "*How am I going?*" [25]
  What are the means of reaching the end goal?

- "*Where to next?*" [25]
  What is the next milestone or step towards the end goal?

In other words, feedback must take into account the end goal of the trainee. With regards to the end goal, feedback about the current position must be formulated and advise about how to proceed next should be given [25].

Another issue is tackling anxiety. Proper preparation and training prior to the actual presentation may help reduce anxiety during the presentation. It may also help the presenter if they were to receive feedback during the presentation to inform them of their mistakes and to help the presenter to calm down when anxiety is detected. This may cause the presenter to feel less alone during the presentation and therefore less vulnerable [14].

One of the ways presenters practice their presentation nowadays is by video taping their presentation and performing a post presentation analysis themselves. The presenter, however, may be biased towards their own presentation. Another method is presenting in front of a mirror but it is difficult to focus on both assessing and giving the presentation. Some people are fortunate enough to practise their presentation in front of a test audience but most people are not. This increased the need for tools to aid people in practising their presentation [38].

Creating those feedback tools is non-trivial. Multiple aspects must be taken into account and the presenter should be made aware of any mistakes they make. Tools that wish to cover every aspect needs to asses speech, both in speed as in volume, as well as posture and gestures. These tool must be able to distinguish different gestures and be able to give them meaning as well as identify certain postures and interpret them as a good posture or bad. The tools must also be able to track general body movement and warn the presenter if they have been stale for too long as well as being able to track their speech and warn them if there is a lack of pauses or if the pace is either to high or to low. Assessing the volume is also a non-trivial problem as the volume depends on the layout of the room. In a small room with the audience seated in close proximity to the presenter will require the presenter to speak at a lower volume. Speaking for a larger audience in a larger room without any volume enhancing tools, such as microphone and speakers, will require the presenter to speak up and increase their volume. A feedback tool should be able to help the presenter with their presentation disregarding the location where the presentation will be held.

Feedback can also be provided for the use of pauses and the use of fillers. Fillers are words which contain no meaning but are expressed when searching for the rights words to say. Fillers can be sounds such as “eum” [23]. Besides avoiding the use of fillers, it is also important for the presenter to included pauses within their speech. Pauses allow for the presenter to catch his breath as well as allows to let the content sink in with the audience and to let them process before continuing [47]. As discussed in Section 1.3.1, the guidelines and rules differ depending on the culture you are presenting in. A tool containing only the rules from the American culture will not properly prepare the presenter for a presentation in an Middle Eastern country [10].

Another aspect that must be considered is the way feedback is given to the presenter. Different possibilities may be considered. It must be decided whether or not the feedback should be disruptive forcing the presenter to stop their practice run or if a non-disruptive approach should be used where mistakes are collected and available for review after the practice run. One should weight the advantages and disadvantages of each possibility. The disruptive approach may provide the presenter with the opportunity to focus the remainder of the presentation on that specific issue and to practice that aspect of presenting. On the hand, disruptive feedback forces the presenter to stop his presentation breaking his flow. It may be hard to continue the presentation from the point where the disruption occurred.

We also have to distinguish two different kind of feedbacks. Feedback can be given either ad-hoc during live presentations as well as during training with the sole goal of practising. When feedback is given ad-hoc during live presentations, it becomes more important to display the feedback in a non-distracting way for the presenter yet in a way that that is fast to process and interpret. This is different from feedback during training where we may even consider to force the trainee and obligate them to adhere to the feedback if the mistakes are to grave to continue.

PowerPoint is a tool for slide show creation and delivery. It does allow training by rehearsing the timings of each slide. However, it is unable to provide feedback about the posture of the presenter, their speech or body movement. Beside the lack of feedback of the performance, PowerPoint does not provide any feedback about the created slides neither during the creation of the slides nor post creation. Feedback about the created slide show could help the presenter follow the presentation guidelines surrounding slides. Rehearsing the timings of the slides in the presentation therefore needs to be accomplished in front of a mirror or with a practice audience the presenters found themselves.

PowerPoint also displays the duration off the presentation which may help the presenter to adhere to imposed time limits. No feedback is given about the other aspects of public speaking as discussed in Section 1.3.1.

1.3.4 Dynamic Presentation Content

Dynamic presentation content embeds all aspects of slide reuse, slide show generation as well as changing the content of the slide at real-time during a presentation. Some argue that the creation of a slide show is too cumbersome and requires t0o much time to complete. A participant in the study performed by Edge et al. [18] stated that the effort put in the creation of the slides should not exceed the effort in the actual intellectual work you want to discuss.

The limited navigational interactions one may take during the presentation makes the presentation to be linear in kind. As a result the ability to reuse the presentation gets reduced as slides may not be skipped easily. This feat however, may prove useful when time is more limited. Another issue is that the slides are static and can not be changed in real-time during the presentation. This goes for text and bullet list in the slides but more importantly for the images of the charts where data is visualised.

Sometimes, an alteration to a chart is required to answer questions from the audience. E.g. the influence on the chart if one variable was no longer taken into consideration. The ability to remove the variable from the chart at runtime would answer that question and show the influence to the entire audience in a visual way which helps with retention. The lack of this features forces the presenter to either close the presenter view and generate the graph in another tool without the variable or verbally communicate the effects of the removal of the variable without providing visual aid. In a best case scenario, the presenter would have anticipated the question and can resort to a backup slide answering the question. Yet again, the limited navigational interactions may make it cumbersome to navigate to the backup slide and back to pick up the presentation.

Dynamic content can also be used to get the audience more engaged with the presentation. The feedback could be collected during the presentation and visualised in the presentation to show the answers the audience have given. The presenter can then use this information to take an alternative path in their presentation to accommodate the received feedback or can be seen as an indication that another explanation is due. The collection and visualisation of feedback can also be used by the presenter to incite a group discussion and have each other listen to their opinion and view which made them pick that answer. A presentation tool allowing for more dynamic content would make the presentation less predictable or dull as criticised by Tufte [55].

PowerPoint does not allow for dynamic content on the slides. Pictures or screenshots have to be taken of any chart, graph or other visualisation the presenter wishes to include in their presentation. The presenter is unable to these visualisations during the presentation since the actual visualisations are not part of the presentation, only a picture of them. One may wish to alter the visualisation to better answer a question or to better explain the effect of a certain parameter on the whole visualisation. The presenter should leave the presenter view and alter the visualisation through an other tool.

PowerPoint does allow the creation of custom slide shows with a different ordering off slides within one PowerPoint file. This may prove useful when an already existing presentations needs to be shortened or with a different focus or angle. However, storing all slides in one presentation and using this feature to create the actual slide shows may cause the tool to clutter and impracticable to work with. It also does not offer the possibility to alter the presentation in real-time during a live presentation. Besides the previous criticism, the custom slide show has to be created without a proper user interface before the presentation takes place.

1.4 Problem Statement and Goal

Commonly used presentation software is too limiting in the functionalities they provide while simultaneously imposing restrictions on the presenter. By using these presentation tools, one binds themself to their laptop or computing, forcing them to remain within close proximity of their device which restricts movement. However, as further discussed in Section 2.1, movement is one of the important categories within presentation guidelines.

These presentation tools are also too restricting on how one may interact with their presentation by limiting navigation to going either to the next or previous slide. There exists a slide overview which allows random slide access but frequent use of this functionality is impracticable. Besides interactions with their slides, these tools lack the functionality to enhance or improve audience interaction.

The tools do not offer any feedback to the presenter which may help him adjust certain elements to improve the overall audience experience. Feedback is only given in the form of a timer which shows how much time has passed since the start of the presentation. No other feedback is give even tough feedback can help the presenter in improving themselves in order to positively influence the presentation.

The last caveat is the lack of dynamic content. Dynamic content may be useful to change visualisations in real-time to show the influence of a certain parameter or to answer questions more clearly. It may also help to improve audience interaction and participation.

The goal of this thesis is the creation of a framework which allows:

- the presentation guidelines to be followed
- multiple ways of interacting with the presentation
- feedback to be given both during presentation as well as after the presentation
- the use of dynamic content within the presentation

In order for guidelines to be followed, the presenter must be able to walk around freely. The best way to accomplish this is by avoiding the use of physical devices that can not be carried around easily. A possible solution is the use augmented reality. Augmented reality is extending ones real world view with computer generated visualisations. Since the visualisations are virtual, one may easily walk around and while the visualisations move along or scale depending on the distance.

The HoloLens[3] has been determined to be the most viable option when considering a wearable augmented reality device. Other devices which can be used for augmented reality are smartphones and google glass. A well known example of an augmented reality game on the smartphone is PokemonGo[4]. However, the display of the smartphone is too small. Besides screen size, the smartphone is not a wearable device but must be carried in ones hand defying the purpose of augmented reality as a solution for this problem. Google glass also has a limited display making it a less attractive device to work with.

The HoloLens has the possibility to offer the presenter with more freedom by removing them from their computer. It also offers possibilities for interacting with the presentation in different ways as well as collect data and visualise feedback. The writer is aware that the HoloLens in its current phase may not be practical to be worn for longer durations at a time. The assumption has been made that with further research, devices with the same capabilities as the HoloLens will appear in a smaller, lighter and more sophisticated form.

This thesis will present an extensible framework for augmented reality presenters views using the Microsoft HoloLens. This framework enables the creation of both real-time and post presentation feedback applications. Due to the extensible nature, multiple interaction devices may be included to allow more intuitive interactions to be used by the presenter in order to manipulate the presentation. The framework also allows for dynamic content to be included and allows the guidelines to be followed more through the use of the HoloLens.

This thesis started off with the introduction in Chapter 1. We will continue with the related work in Chapter 2. Within the related work we will discuss different research performed within the different aspects of public speaking and presenting. The related work formed the baseline for our requirements which will be discussed in Chapter 3. This chapter will discuss both the technical and functional requirements off our framework. After the Requirements, we will discuss the implementation in Chapter 4 where we will

[3]https://www.microsoft.com/nl-be/hololens

[4]https://www.pokemongo.com/

talk about our design choices, the implemented architecture as well the implemented functionalities. After the implementation, an evaluation has been performed which will be discussed in Chapter 5. Within this chapter we will discuss the methodology of the study as well as the results and the ideas shared to us by the participants. This thesis will end with the conclusion in Chapter 6 where we will talk about what has been achieved and the future work.

2
Related Work

As mentioned in Section 1.4 the goal is to create a framework for better presenter views, allowing presenters to interact with presentation content, get real-time feedback and visualise external data streams. With this in mind we discuss some related work from each of the four main aspects we identified before: presentation guidelines, presentation interaction, presentation feedback and dynamic presentation guidelines. These solutions address some of the issues previously introduced in 1.3. We will discuss the solutions and approaches to draw conclusions and gain inspiration for the creation of a more generic framework.

2.1 Presentation Guidelines

Many guidelines apply to public speaking which all improve the impact and influence of the presentation. However, it is difficult for presenters to follow these guidelines. The guidelines will be discussed within this section so we can aid developers and presenters to follow these guidelines in our framework.

Assuming that the content of the presentation is on par with the expectations of the audience there are still other factors that may affect the delivery off the presentation. Besides the content of the slide, three other factors are going to be discussed. These factors are speech, body movement and cultural differences.

2.1.1 Speech

Discussing an interesting topic within your presentation does not guarantee a successful presentation. One must also deliver the information during the presentation in a way that gets the message across. The delivery of the information is as important as the content to be delivered.

One of the aspects of the delivery in public speaking is the element of speech. Strangert and Gustafson [52] researched the effects of acoustic-prosodic features of the speaker utilised during the presentation and the way the presenter is perceived as a public speaker and how well the message was delivered to the audience. The research also included the influence of fluency and speech rate. This study has been performed on public fragments from the Swedish parliament and presented to students who rated the speakers according to their own opinion on a series of given statements. Those statements measured how the participants perceived the speakers as to be objective, aggressive, insecure among others. The participants were asked to rate each speaker they heard as either a good or bad public speaker from their own point of view.

The samples were subjected to an acoustic analysis. The analysis included the measurement of speech rate, the pace at which words were being said, as well as the use of pauses, duration and frequency among other acoustic features. The results of the analysis was compared to how the speakers have been perceived by the participants in an effort to research the possible correlations between the acoustic features one may employ and how that affects the perception of them as a public speaker.

The samples of the speakers perceived to be disfluent were subjected to a disfluency calculation. The calculation took into account how often the speaker had to repeat himself as well as the frequency they experienced a slip of tongue among other factors. The study found that a correlation exists between the fluency of the speakers and how they are perceived as a public speaker.

Another correlation found through this study was the range of the pitch used within the speech and the frequency wherein focussed words were correctly emphasised. Focus words are words which are key to the understanding of the sentence. Speakers utilising a wider frequency range in their speech, speakers who were less monotone, were perceived to be better public speakers compared to more monotone speakers who made less use of focussed words.

Speech rate also influenced on how speakers were perceived as public speakers. A ranking or ordering in how important those influences are could be made. Speech rate has the lowest influence while a wide frequency range was determined to be the most important influence. In general speakers who were dynamic in their speech, had a wide frequency range, made use of pauses and changed pace frequently, among others, were perceived as the better public speakers. Using those results, the researchers altered the speech samples to improve them as public speakers. The altered versions were again subjected to the opinions of different students. The altered version was perceived as a better public speaker confirming the results from the first phase of the study [52].

Other research has been performed by Rosenberg and Hirschberg [45] on the correlations between charismatic speech and the acoustic-prosodic features of the speaker. The study has been conducted using the same methodolgy as the study by Strangert and Gustafson [52]. The participants were asked to give their opinion on given statements based on a speech fragment similar to the research of Swedish speech discussed earlier. The research concluded the same correlations to apply to American speech. Cultural differences and similarities are discussed further in Section 2.1.4.

2.1.2 Body Language and Movement

Body movement, posture and gestures are also an important aspect of public speaking. Non-verbal communication, even when done subconsciously, influences the audience and the effectiveness of the presentation. One aspect which may cause subconscious non-verbal expression is anxiety, but other emotions may also influence behaviour. There have been cases where the audience have been more conviced due to the non-verbal expression employed by the presenter than the content of the speech itself [38].

Arina in her book “Successful public speaking” [39], also states the importance of appropriate body language and facial expressions because body language portrays are real thoughts and feelings. The book states that the presenter should keep a relaxed an upright posture which portrays confidence. One should avoid abrupt movements, tense posture or holding onto your lectern as some sort of safety net. It also advises to avoid the constant shifting of ones weight from one foot to the other as the rocking may be distracting for the audience.

Body placement is also important. One might be persuaded to hide behind some sort off visual wall to create distance towards the audience. One example is a presenter who does not move away from their lectern. While it may be tempting to reduce anxiety by creating distance between ones self and the audience, it has a negative effect on the presentation. The presenter should move away from the lectern and move towards the audience. Moving towards the audience is inviting towards the presentation and allows a connection to be build up between the presenter and the audience. Body movement can also be used to attract the users attention again or to generate a spatial awareness between slides by physically moving a certain distance on stage [39].

The arms of the presenter are important to take note of as well as they easily show the anxiety one may have. An anxious person may carry his notes around with him all the time and check them too often. The presenter should appear to be calm and take on an inviting and open posture by keeping their arms in a relaxed state next to their body. One should avoid fiddling with clothes hiding hands in their pockets, clenching their hands together or crossing their arms [39, 47].

Facial expression also have an impact on the audience and the presentation. In order to gain the sympathy and interest of the audience, one must first shown interest in their audience. Arina states that the most effective way to gain sympathy from the audience is through a genuine smile [39].

The book also states that the presenter should provide visual aid but should be wary not to turn their back on the audience. One may turn subconsciously turn their back on the audience as a result of turning to the projection screen to better explain the visual aids being projected. However, this may result in a loss off connection between the presenter and the audience as the presenters closes themselves off from the audience.

2.1.3 Slides

Several studies have been performed about the layout and content of the slides. One such study has been performed by Mackiewicz [35] about the use of fonts in slides. They measured the differences of four variables between serif and sans-serif fonts. The research concluded no significant difference between either kind off fonts except for the profesional variable. Both types of fonts were equally attractive, comfortable to read and interesting. The presentations using a sans-serif font was perceived to be more profesional. The researches also concluded that the *Gill Sans* is the best font to use within presentations [35].

Some research has also been conducted to define slide design guidelines. Alley and Neeley have defined a set of guidelines for slides [2] based on criticism from e.g. Tufte [55]. These guidelines state that a sentence should be used as a title instead of a phrase since it has more potential to convey the topic to be handled to the audience. It states that the title should be left aligned in the top left corner or the slide. The slide should contain visual evidence such as images, graphs or text arranged in a visual way. Similar to Mackiezicz [35], the guidelines from Ally and Neeley [2] state that a sans-serif font should be used. It is appropriate for the title to have a bigger point type then the body. It is also advised to avoid bullet lists as they to do not portray the relations between bullet points. Farkas [21] says that the presenter should reduce the number of slides and spent at least a minute per slide. The time per slide may increase depending on the total duration of the presentation. Also the amount of content on each slide should remain restriced. By this Farkas means content of which the audience is required to invest time in order to comprehend the content of the slide to the fullest [21].

Figure 2.1: An example of a badly designed slide
https://www.emaze.com/2014/11/25/6-worst-presentation-slides-ever/

The use of colour is also an important aspect we have to address. Colour is an important tool which can enrich the presentation when used in a proper manner. It can help to categorise element on a slide which belong together by colour coding them. Colours can also influence how the audience may perceive things. The green colour is generally perceived as something positive while the red colour is perceived to be negative [57].

When used improperly or without due care, it may render the presentation unreadable or cause distress with the audience. Colour is relative medium which means that it is perceived differently depending on the other colours neighbouring it. It may also cause the objects of equal size to be perceived differently due the colour the object may have [58].

Another aspect of slides is the visualisation of data. Some argue that the best way to visualise data is by maximising the data-ink ratio, or the amount of ink in the chart used to visualise the data. This corresponds with the notion of avoiding too much text on the slides. However, charts embedded within a drawing are remembered better by the audience. This corresponds with the guideline to use images and visualisations to aid the audience in better understandig the topic which is being discussed. [4].

Figure 2.1 shows us a badly designed slide. The amount of content invites the audience to read which diverts their attention from the presenter and his speech. Another problem with this slide is de colours which are being used. The background meant to enhance the topic of the presentation makes it hard to read the white coloured text. This may again distract the audience who will be more focussed on trying to read the written content instead off listening to the presenter.

2.1.4 Cultural Differences

Cultural differences also have to be taken into account. For speech, a comparison study has been performed between the Palestinian, Swedish and American culture by Biadsy et al. [10]. The study focussed on how charismatic speech is perceived within different cultures. With charismatic speech, the researchers meant the ability to influence others through the use of speech. This is similar to the goal of presenting and public speaking in general. Therefore we can use this research as an indication of the importance of the cultural differences. The comparison is made using the two studies already discussed in Section 2.1.1 [52, 45] and another comparison between American culture and Palestinian culture [9]. The Palestinian study has been conducted using the same methodology as the American and Swedish studies.

If we look at the comparison in acoustic-prosodic features between cultures, the study tells us that there are no differences. In all cultures where the speakers who utilised a wide frequency range in there speech perceived as the better public speaker. Same applies to speakers who spoke fluently as well as kept a decent pace and utilised phonetic pauses correctly.

2.2 Presentation Interaction

Presentation Interaction is any interaction that may occur while giving the presentation. This can be split up into audience interaction as well interaction with the slide show. Audience interaction, however, has been improved by integrating their answers into the presentation and gathering feedback through tools so other people their opinion does not influence their own answer. These applications are known as *classroom communication systems* and will be discussed in Section 2.4.1 As a result, this section will primarily focus on work surrounding interaction between the presenter and his presentation. Some of these solutions also positively influenced the interaction between the presenter and their audience.

As mentioned in Section 1.2, navigation off the slides is still hard or at the least limited within the current tools. Random slide access requires the presenter to search for the right slide in a sequence of all slides. Searching for the right slide may be made easier by dividing the slides into sections. However, PowerPoint does not support any sub sections which means that the additional functionality does not aid that much with improving random slide access. Due to the nature of interaction and the hardship of random slide access, skipping slides or the use of back up slides only to be used in case of great audience interest, are hard to achieve feats. In short, Current applications expect a presentation to be linear without only exceptional deviations. This however, is no always they case.

2.2.1 Camera-tracked Laser Pointer

Another mean of interacting with presentations is by using a laser pointer. The researchers found the use of a mouse and keyboard during the presentations to be uncomfortable. They limit the freedom of the speaker by binding them to the location of the devices of interaction. The use of the laser is to offer a solution to that problem and to provide more freedom to move around to the presenter. A laser pen, in contradiction to a mouse, is a wireless handheld device which is easy to carry around and does not enforce any constraints on the presenter their movements while using the system.

The laser mimics the actions normally provided by the computer mouse. Video cameras directed at the presentation screen track the laser pointer. Its location, relative to the content presented, will be calculated before the new information is interpreted. Using timers to track how long de pointer stood still on a certain spot or how long it was turned off, the system is able to simulate mouse button press and mouse button release events [28]. Figure 2.2 shows the workings of the system.

2.2.2 SketchStory

SketchStory [31] is an interactive whiteboard with touch and pen interactions. The pen can be used to draw and write annotations on the whiteboard. The touch can be used to move drawings and text around on the board as well as

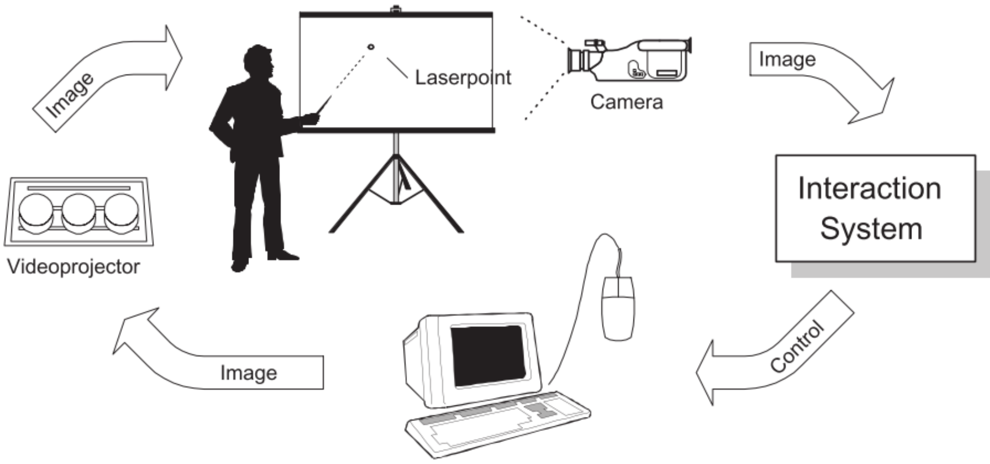

Figure 2.2: Chart of the camera-tracked laser pointer interaction solution [28]

zoom functionalities. SketchStory mainly focussed on the storytelling aspect of presentation giving and takes different approach than the other solutions discussed within this section by not integrating any slides.

SketchStory, depicted in Figure 2.3, is an application developed to be used for storytelling through data, with a focus on the information visualisation of the data and the benefits gained from freeform sketching and annotating. The researchers were inspired by other techniques such as whiteboard animation where a story is told in a video and enhanced through animations, drawings and graphs. Examples of such whiteboard animations can be found on the RSA-animates playlist[1].

To work with SketchStory, the presenter must first enter a sequence of charts in the system and synchronise the story they wish to tell with the sequence of charts in the system. Through the pen interaction, SketchStory is able to recognise sketch gestures. Those sketch gestures invoke the next chart of the sequence. The pen interactions allow the presenter to annotate the chart in real time with more information, such as written text and sketching. It also invites the audience to take part in discussions by allowing the presenter to change the chart in real time by excluding a certain element from the chart enabling the presenter to more precisely answer any question or provide more context. The zoom and panning functionality make it possible to keep the charts on the whiteboard while continuing with the presentation allowing the presenter to return to a previous chart in order to

[1]https://www.youtube.com/playlist?list=PL39BF9545D740ECFF

Figure 2.3: SketchStory interactive whiteboard in use [31]

answer questions which may be asked later on when more explanation has been giving through further storytelling, or to show the connections between charts.

While SketchStory has it advantages, it does not provide an answer to the linearity of presentations. One might even say it facilitates linear presentations. The presentation is required to enter a sequence of charts and to synchronise the story they wish to tell with this sequence. During the presentation, the presenter may not divert from this sequence. They are able to return to a previously drawn chart because it persists on the whiteboard. However, they are not able to skip charts or jump ahead in the presentation.

SketchStory is a solution for interaction between presenter and their presentation as it allows free-form annotations as well as panning and zooming interaction. In the meantime it also offers a solution for audience interaction as invites the audience to partake in discussions and ask questions due to the options provided to the presenter to easily and accurately answer those questions. The possibility to annotate the charts in real-time and to change the charts to enable the presenter in answering question more specifically also makes SketchStory a solution facilitating dynamic presentation content [31].

2.2.3 Digital Pen and Paper

Paper still excels in certain areas making it a more preferred medium for certain tasks. Some of these actions appear in presentations such as making annotations during a presentation to enhance your explanation. For this reason, research has been performed to link pen and paper back to digital presentations.

When presentations were performed with the use of overhead projectors and transparencies, one was able to perform on the fly annotations when the need arose. The ability to do this has somewhat withered with the introduction of new technology. It is still possible with the use of tablet computers and touch screen devices but they tend to be quite heavy and awkward to hold. Tablet computers are able to capture handwriting as well but, this is accomplished by writing with you finger or a touch-screen pen on an awkward, heavy device while the same can be accomplished on paper with a pen which is light weight, cheap and easy to handle [34].

PaperPoint

PaperPoint [49] is such a solution which links paper back to a digital presentation. It takes the user away from his computer by allowing interaction through a digital pen on lightweight paper which one can carry around during their presentation providing the presenter with the freedom to walk around during their presentation.

PaperPoint itself is not a new tool for presentation giving but offers a new interface to interact with PowerPoint presentations. This has the benefit for the presenter that the presentation can still be made using the presentation tool they are used to. It also allows the user to switch between interacting through PaperPoint and PowerPoint itself. Therefore no extra effort is required when creating the presentation. When printing the handouts of the presentation, the user has to make sure that the correct template is selected.

PaperPoint [49] makes use of an unique Anoto pattern. This unique pattern enables the linking of actions performed on the paper with actions in the digital copy. It also makes it possible to have buttons on the paper which translate to actions on the presentation software.

Buttons made available are a *next* and *previous* button which allows to go through the slides in the order as set in the presentation. Each slide also has a *show* button allow random slide access. The presenter therefore has the freedom to skip slides with ease without having to close the presenter view. It also offers the possibility for slide reuse since the presenter may create one elaborate presentation and skip slides or sections within this version when

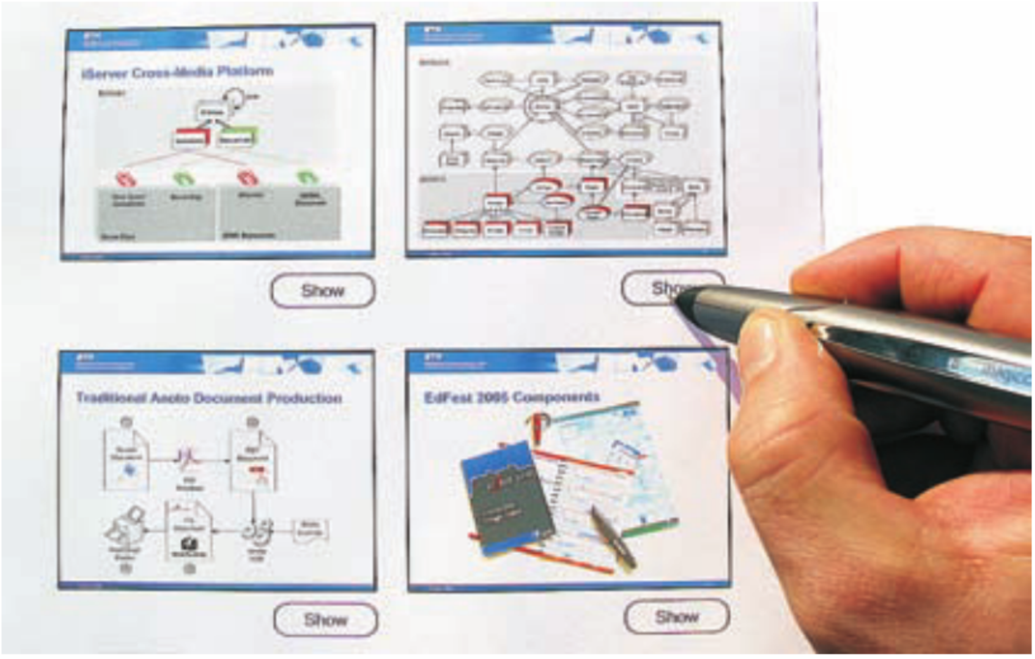

Figure 2.4: PaperPoint handouts with digital pen [49]

time is limited. This is also useful when wanting to return to a slide to answers questions from the audience. Figure 2.4 shows us a handout of a PaperPoint presentation with a show button underneath each slide.

PaperPoint also allows the annotations of slide through the paper handouts. This feature provides the ability to answers questions better by real-time augmenting the presentation by putting a larger focus on one of the aspects on the slide. The same could be achieved with mouse and keyboard interactions on the slide but enabling the presenter to perform the same actions on paper with a pen is more natural and offers a greater freedom and flexibility. The writing of the presenter is captured through the digital pen and sent over the digital presentation which augments the exact same slide the presenter was writing on with his writing. This might not necessarily be the slide which is currently being displayed to the audience [49].

Palette

Palette [37] is a paper interface using index cards as one may employ during a presentation. The cards have been adjusted to be readable by both humans and computers and can be used by the presenter to navigate through the slides. The researchers noticed how the audience had to wait from time to

Figure 2.5: The Palette system with presentation cards and scanner [37]

time due to the presenters searching through their file system searching for the right slides or having difficulties with the controls of the computer. The researches wanted to provide a new way of interacting with slides to reduce time loss and create an overall smoother experience when presenting.

The presenter first has to design their slides which can be done through the presentation software of their own choosing. When the presentation has been designed, index cards are generated and linked to the digital slide they represent. Each index card consist of a thumbnail of the slide it represents and a machine-readable code. This code is what links the physical index card to the digital slide. During the presentation, the presenter can freely jump to any slide by scanning the related index card.

This system allows slide reuse since the presenter is able to decide their own path at runtime, skip slides or entire sections as well as choose to present the content in a different order. It allows the presenter to more discretely look for the right slide without having to exit the presentation view and have the audience watch your search for the right slide. Palette allows the presenter to move around more freely. The presenter still has to pass by the static scanner to navigate to the next slide the presenter wants to discuss. Alternatively, the presenter is able to choose to use a portable scanner to allow more freedom in their movement. Palette allows the presenters to lay out their presentation in a physical form and can arrange them before his presentation in any way they prefer without the requirement of the computer [37]. Figure 2.5 shows us the scanner and presentation cards used within palette. Each presentation card has a barcode next to the slide content.

2.2.4 Zoomable User Interfaces

Zoomable User Interfaces are an attempt for a more natural and user friendly interface. In a zoomable interface, all data is displayed on a single two dimensional plane for the user to navigate through. The user is able to navigate through the plane by panning and zooming in or out. Panning and zooming allow the navigation through al large display off data. Panning and zooming are also more natural interactions to us which may therefore create a more inuitive interface [8].

The interactions are often animated to increase the resemblance of actions performed on paper. Sliding pieces of paper over a table is translated to panning on the plane. Bringing a piece of paper closer to your eyes for a more detailed field is translated to zooming in on the plane while the zoom out is increasing the distance off a piece of paper from the eyes.

By executing a zoom in interaction, the user is provided with a more detailed view of the data they zoomed into. Zoom out interactions are used to be able to see the bigger picture and to provide context the detailed data one was looking at [22].

Prezi[2] is a well known online presentation tool which makes of a zoomable interface. It allows the creation of both linear as freeflow presentations. Prezi makes use of a large canvas where the entire content of the presentation is displayed and uses the zoom and pan functionality of a zoomable interface to navigate through the slides of the presentation [41].

KidPad [16] is an example of a zoomable interface in the academic world. Kidpad is a childrens' story authoring board. The interface has been created through the input of children who prefered the zoomable interface and were visibly immerserd when a story was told using the zoomable interface. The experience offered by a zoomable interface was refered to the children as *a ride* and they would often augment the zoom with sound [16].

Another example was the creation of PhotoMesa [7] which is a zoomable image browser and an answer to the information retrieval query systems. The results of these systems still had to be browsed by the user in order to find

[2]https://prezi.com/

the correct picture. The authors mention that pictures are most often looked at for enjoyment. For this reason, they wanted to create a more intuitive image browser that allows easy retrievel of a specific picture as well as offer the enjoyment seeing pictures they were not necesseraly looking for, but fit well within context [7].

CounterPoint

CounterPoint [22] is a zoomable user interface based presentation tool. As discussed in Section 2.2.4, the interactions offered by a zoomable interface make it possible to move away from the otherwise usual linear presentations.

Since all content for the presentation reside in one large plane, seen in Figure 2.6, no order is enforced. This enables the presenter to switch more easily between slides which contradicts the imposed order. It also allows multiple versions to coexist in the same due to the presenter being able to freely choose in what order to present which slide. Those are the benefits of the single large plane and the panning interaction.

Besides the panning interaction, there also exists a zoom interaction. The zoom interaction has the purpose of zooming in to a more detailed view or to zoom out and get a more general context view. This has also been used in CounterPoint by providing different information based on the the current zoom level. Another benefit gained from the use of a zoomable interface in presentation is that the way to interact with them lies closer to what we are familiar with. They are a digital representation of everyday interactions in the physical space. Due to the given freedom and the more familiar interactions, the cognitive resources are put to use in a more efficient way .

CounterPoint is a tool which helps with the creation and presentation of a zoomable presentation. It helps the user in creating a hierarchical structure within the presentation which helps to determine what to display at each zoom level. Hierarchies are a natural way of ordering content. This thesis is an example of that since it is split into chapters, section, subsections and so on. The higher you are in the hierarchy, the more context you receive. The lower in the hierarchy, the more details are provided.

CounterPoint also addressed some of the issues zoomable interfaces have. One problem that may occur is map schock. Map schock refers to the feeling of being either overwhelmed, confused on unmotived by the vast amount of data or complexity portrayed on the plane. To address this, CounterPoints makes use off incremental revealing content. The zoomable interface helps with the semantic distance. Slides may sometime transition into a new subject with no real semantic relation between them. However the slides were

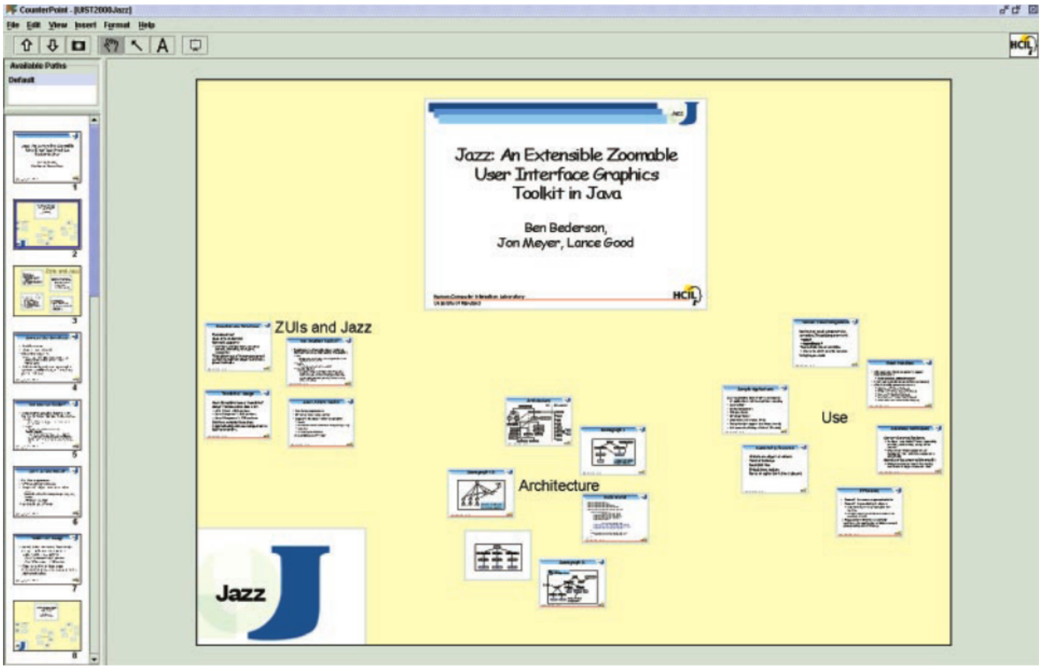

Figure 2.6: CounterPoint zoomable presentation interface [22]

in close proximity as the distance as the distance between slides is always the same. However, when using a zoomable interface, the shift in topic can be spatially augmented with the zoom out, pan and zoom-in function to enforce the semantic distance by an actual visual distance in the presentation canvas [22].

Fly

Fly [27] is presentation tool which focusses on the spatial mapping of the slides. It uses spatial relations between slides as well as colour coding. The researchers based their presentation tool on mind maps. This was chosen due to the way mind maps make use of the Gestalt principles [56]. All information resides within one large information plane. Navigation through the plane is achieved through panning and zooming classifying it as a zoomable interface.

In Fly, slides are not placed in a hierarchical structure but rather grouped together by association and linked using weighted edges as seen in Figure 2.7. It makes use of colour coding and labels off the subgroups. When presenting, the applications zooms in on the first slide and visit each slide in a clockwise order. Random slide access is available by clicking on a slide in the graph [27].

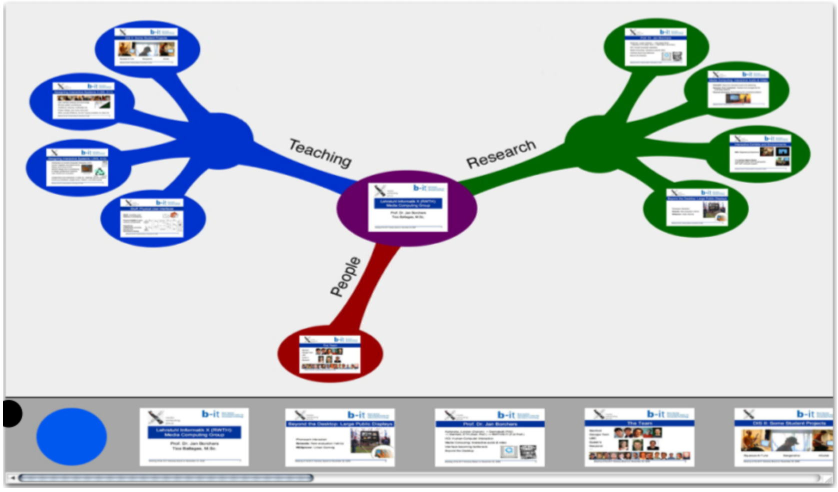

Figure 2.7: Fly, mind map based presentation tool [27]

2.3 Presentation Feedback

This section is going to cover the research and work which has been conducted resolving presentation feedback. Within this section we distinguish between feedback during training and feedback during an actual presentation. This distinction has been made due to the nature and form the feedback may be provided. When in training, the feedback may be interruptive and force the presenter to stop and take note of the provided feedback in order to improve. While feedback provided during a live presentation must be subtle and should not confuse or distract the presenter.

2.3.1 Training

Feedback systems used in training focus more on pinpointing mistakes to be corrected. These systems are focussed on improving the presentational skills of the presenter and providing instructions on how to improve their presentation practises.

Presentation Trainer

Presentation Trainer [47] is an application designed to prepare and train the skills going along with public speaking and presenting. This tool assesses multiple aspects of public speaking such body movement and posture as well as voice to assess the overall performance of the presentation. The importance of the presentation guidelines have previously been introduced in Section 1.3.1, as well in Section 1.3.3 where the problems surrounding current training practices have been written out.

The Presentation Trainer makes use of a Microsoft Kinect sensor V2[3] to track the body of the trainee. The Kinect is able to track the coordinates of different joints of ones body enabling the detection of certain gestures and postures. In order to reduce the amount of false positives, a threshold of 0.3 seconds has been implemented. The trainee must stay still, within reasonable bounds, for the threshold amount before being recognised as a posture. The application also holds a predefined set of postures which one may not assume during a presentation.

The Presentation Trainer acknowledges the importance of gestures and body movement. However, it does not try to identify which gestures are being used but only assesses whether or not gestures are being used. Gesture detection is accomplished through the use of the Microsof Kinect sensor v2 by keeping track off the different joint through their coordinates. It also keeps track of the angles between body parts. Gestures are being identified through the changes in angle between body parts. If the change in angles exceed a predefined threshold, it is perceived as a gesture. It takes notion of the time between gestures. This way it is able to provide the trainee with feedback if the use of gestures seems to be lacking.

The presentation tool watches the shifting of their weight between their legs. This causes the presenter to make a rocking movement. The application detects this practice through the coordinates of the trainees hip, provided through the Microsoft Kinect sensor V2. It counts the weight shifting within a four second period. This count may not exceed a predefined threshold.

Also the speech of the trainee is assessed through the use of a microphone. The application has three treshold to identify normal, soft and loud speaking volume. It takes into account the different possible stages one may present at enabling the trainee to set these tresholds manually to accommadate their specific needs.

Phonetic pauses is also an aspect of public speaking which is assessed by this application. It detects pauses through the absence of sound captured by the microphone. The silence must last for at least a quarter of a second. The application expects a pause to occur every fifteen seconds.

Presenters may use fillers as phonetic pauses. The use of fillers, however, is a bad practice. The application makes use of the speech recognition ability of the Microsoft Kinect V2 to make the trainee aware of its use. The accuracy in detecting the filler sounds is rather low but one may assume a rather high frequency of fillers with presenters who should be made aware of their habit.

[3]https://msdn.microsoft.com/en-us/library/hh438998.aspx

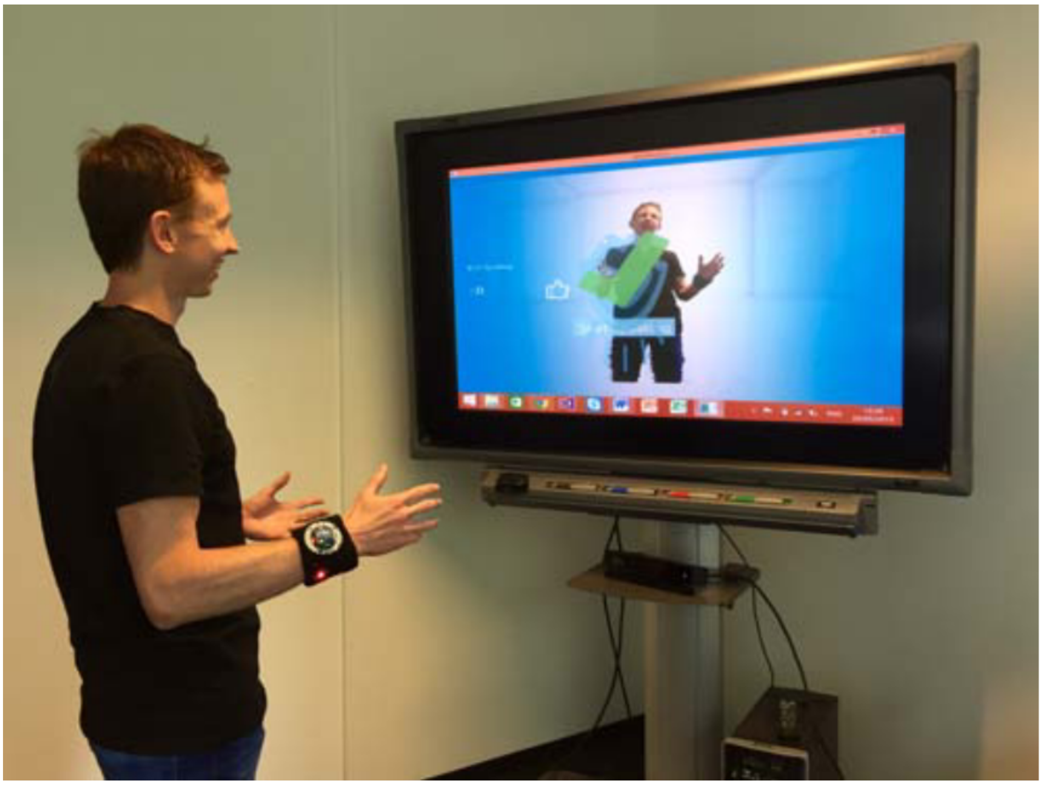

Figure 2.8: Presentation trainer in use [47]

It is important that the feedback is useful for the trainee when presented. The researches learned that a constant stream of feedback is confusing for the trainee, even when the feedback was all of the same type. To mitigate this problem, the application only shows one feedback item at a time and waits at least six seconds after it stopped displaying the feedback, before presenting the trainee with the next feedback item. It stores all feedback generated by the assessments and selects the most relevant feedback to be presented when the moment for feedback is appropriate.

The feedback consist of two parts. Haptic feedback, through means of a vibrating wrist band is used to notify the trainee whenever feedback is presented. A graphical interface is used to provide visual feedback. The visual feedback consist of a mirrored image of the trainee, as seen in Figure 2.8, to reduce cognitive load and increase self-awareness and a short written statement describing how to correct the mistake.

It makes a difference between corrective and interruptive feedback. The application stops assessing the presentation on interruptive feedback and waits until the trainee acknowledges and corrects the mistake. Interruptive feedback is therefore reserved for severe mistakes. When interruptive feedback is presented to the trainee, the possibility is offered to focus the remainder of the training on that aspect so no other feedback will be displayed for the remainder of the presentation [47].

Presentation Sensei

Presentation Sensei [30] is a presentation feedback tool which focusses on five aspects of presentation giving. The first aspect is the pace of speech of the presenter, which may not be to slow. Another aspect is that the speech may not be too monotone. The third aspect being measured is the use of fillers which should be avoided. Fillers are noises which have no semantic meaning the presenter may produce while searching for the right words. The fourth aspect is the gaze of the presenter. The presenter should pay attention to its audience and not present while reading slides or notes out loud off of his screen. The last aspect is that the presenter should adhere to the given time limit if one is given.

Presentation Sensei provides feedback during the presentation in real time through visual alerts notifying the presenter of bad practices when detected. The visual alerts are similar to traffic signs. there is also a constant display of the short term statistics for the presenter to follow their own progress. Even though the Presentation Sensei is able to deliver real-time feedback, it is not meant to be used during an actual presentation as the alerts may be distracting for the presenter.

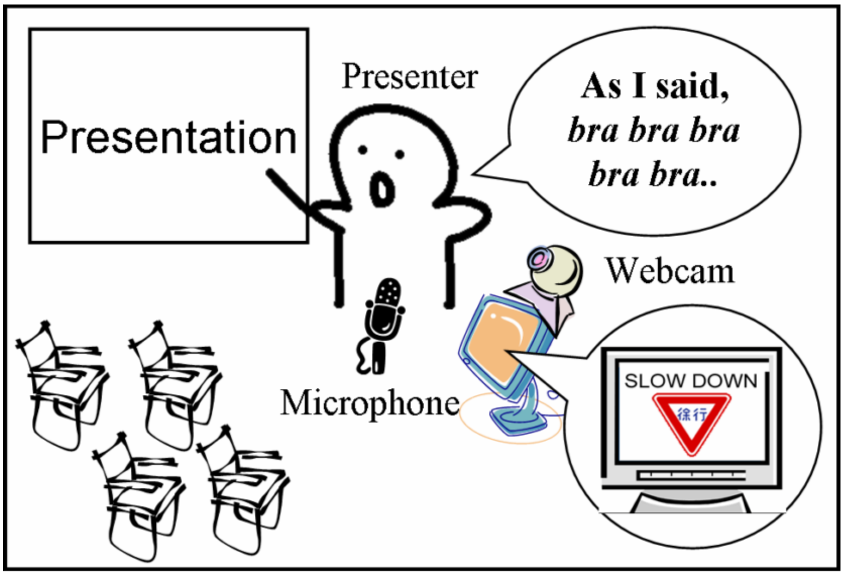

Figure 2.9: Cartoon describing the working of Presentation Sensei [30]

Presentation Sensei provides a visual summary at the end of the presentation to provide the presenter with the means to reflect on his performance. The summary consists of a series of charts displaying the overall performance of the presenter. Charts are generated from both the entire set of raw data as well as the averages off each slide. The charts of each training session are placed vertically respecting the order in which the sessions occurred. This allows for the presenter to see their performance improve over time.

Detection is achieved through the use of a microphone and a webcam as seen in Figure 2.9. the microphone is used to input the speech of the presenter for analysis of the speech related aspect: pace, fillers and tone. Since the solution is Japanese, the pace is determined by tracking the number moras per second which is an intuitive phonological unit for the Japanese language. The pitch is analysed through the change in volume of the presenter. When too few changes in pitch occur, the systems will conclude the presenter to be monotone. For the detection of fillers, the researcher used the system created by Goto et al. [23] who created a system to detect fillers in pauses during speech.

The webcam is used to detect the position and the orientation of the presenters head which is then analysed to determine what the presenter is looking at. The presenter may be looking to much at his slides or the notes displayed on his computer. The systems therefore measures if the presenter spends enough time looking at his audience [30].

2.3.2 Real-Time Feedback

Real-time feedback systems are systems which provide feedback in real-time during the presentation. They may not distract the presenter as it may abrupt the flow of the presentation. The information must be displayed for the presenter to be able to look at in order to improve the presentation. However, it may not stall the presenter or the presentation. The information must therefore be displaying in a way that can be interpreted at ease. This section will discuss some solutions which provide feedback to the presenter in real-time

Rhema

Rhema [53] is an application on the Google Glass for providing real-time feedback to the presenter during the presentation. Rhema assesses the speech of the presenter in terms of volume and pace. It acknowledged the difficulties of keeping the presenter informed without distracting the presenter. Distraction may induce weird behaviour such as sudden pauses or stuttering which is not desirable during a live presentation.

Google glass is a small display in the corner of the presenters eye, mounted on the their head. This imposed the possibility of attention shift or for the display to be completely ignored or forgotten making the application pointless. The application may however not capture the attention of the presenter for the entire duration off the presentation as it would reduce eye contact between the audience and the presenter which will have a negative effect on the presentation. In order to prevent a cognitive overload, information is displayed in a way that only requires a short glance of it to be perceived and interpreted [53]. Figure 2.10 depicts the architecture of Rhema. Rhema makes use of a server for the computations and analysis and only uses Google Glass as a screen for the presenter in order to visualise feedback.

2.3.3 Virtual Audience

Virtual reality is being used in helping people getting rid off their fears or phobias. This branch in the medical health care is refered to as Virtual Reality Treatment or Virtual Reality Therapy. The Virtual Reality Therapy can act as a replacement for behavioural therapy where the anxiety inducing stimuli may be expensive or hard to come by. One such therapy is being used to help people get over their acrophobia, the fear of heights [46]. Research showed that Virtual Reality Therapy for acrophobia was equally effective compared to exposure in vivo [19].

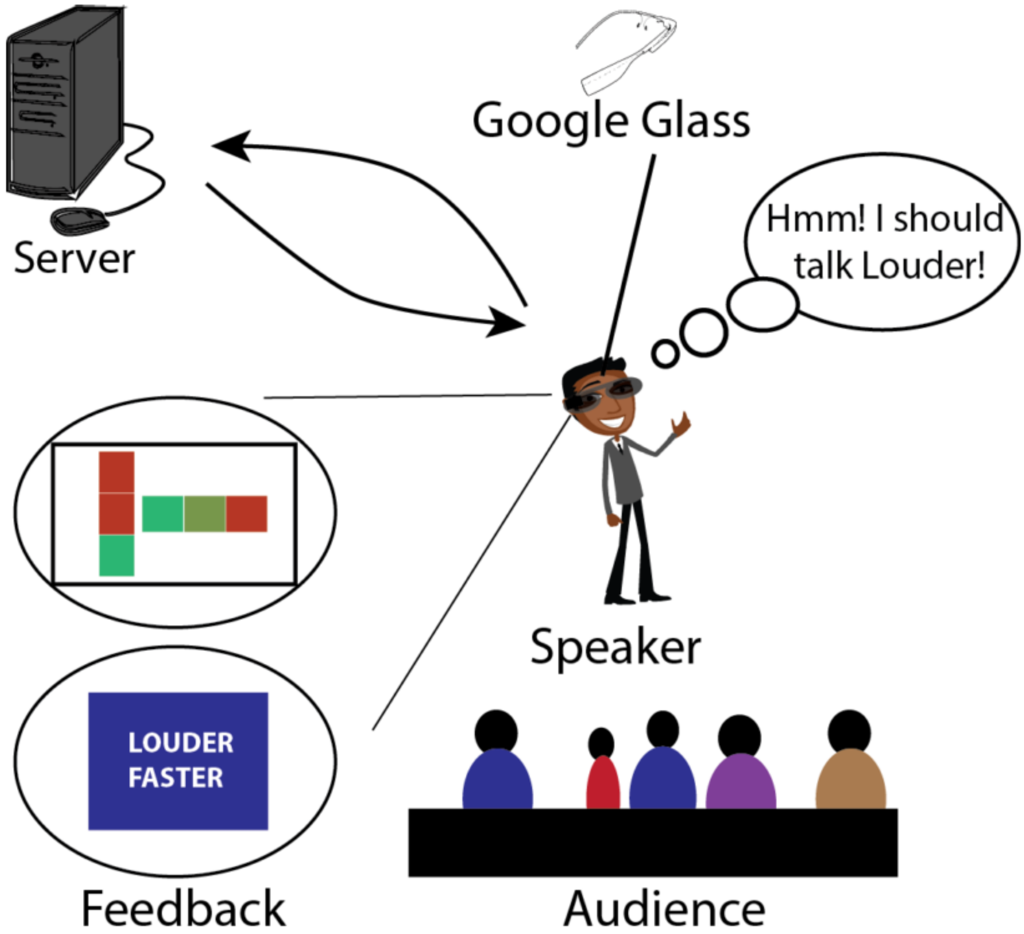

Figure 2.10: Rhema, real-time feedback through google glass [53]

Figure 2.11: Cicero virtual audience [5]

Anxiety in public speaking is a social phobia where one fears to humiliate or embarrass themselves or generally, show behaviour that may cause others to negatively think about them. Anxiety may impact the voice of a persons as well as their behaviour. Research has concluded that practising public speaking in front of virtual audience positively influences the trainee. The experiment proved that anxiety can be invoked by the attitude of an audience even if the audience is known to be virtual [42].

Cicero

Cicero [5] is a virtual audience training tool, as depicted in Figure 2.11, that has been established though using expert ratings and correlating it with automatic behaviour descriptors. Those descriptors are now used within Cicero to determine the behaviour of the virtual audience. The changing behaviour of the virtual audience is one of the ways in which the system is able to provide the presenter with feedback alongside visual gauges which change according to the presenters performance.

The system is split into three different components. The first component is the performance handler which asses the performance of the presenter during his presentation. The performance handler uses the detected behaviour from the behaviour recogniser. The behaviour recogniser relies on sensors watching the presenter in order to analyse the performance. The performance handler uses the information and sends the most relevant information to the feedback producer. The relevant information is determined by the given training condition. Alongside the virtual audience, gauges are also displayed which show the presenter the reports of the behavioural descriptors. The systems is capable of providing feedback through these gauges as well as provide feedback through changing the behaviour and facial expressions of the virtual audience [5, 13].

2.4 Dynamic Presentation Content

This section will discuss solutions to create a more dynamic presentation setting. As already mentioned in the introduction, Section 1.2 and in Section 1.3.4, presentation nowadays consist of static content. E.g images of charts which can not be altered. This is a consequence of the presentation software used to enhance the presentation with slides. The more commonly used slideware solutions impose a certain sequence to be followed throughout the presentation as they do not allow the presenter to easily switch to a slide that breaks the expected order. These programs also do not have the functionality to create or change the slides on the fly during a presentation to accommodate audience input.

2.4.1 Classroom Communication Systems

Classroom communication systems are applications which are meant to increase communication in the classroom and extend presentations through the use of generated content based on audience input. A classroom communication systems is a system through which the presenter can collect answers from the audience after presenting a problem statement or a question. The system collects those answers and creates visualisations for those answers.

Those systems are meant to increase audience participation by allowing lectures to be changed to a sequence of small discussion sessions. The cycle, as described in Figure 2.12, used within classroom communication systems attempts to change the audience to active participants in the learning process. Such systems are seen as the complete combination of both hardware and software [6, 11].

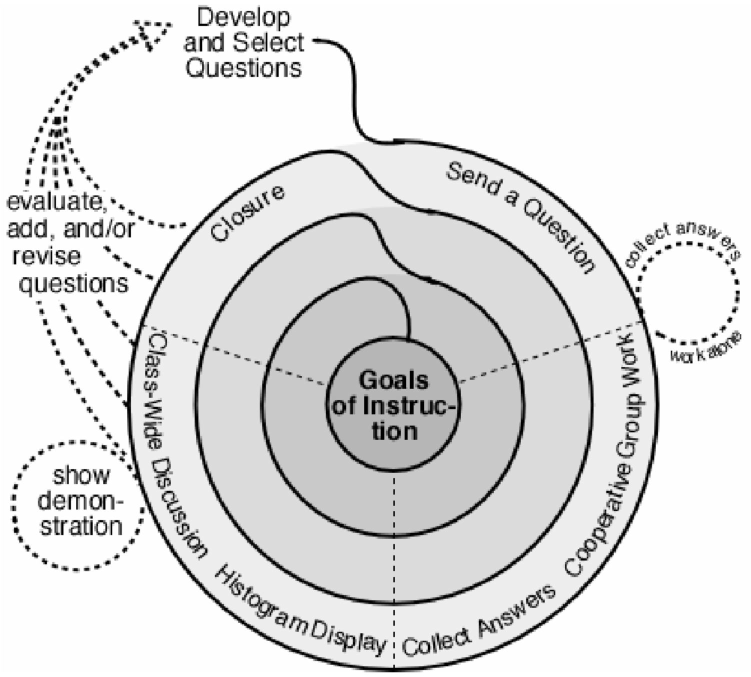

Figure 2.12: The classroom communication system cycle [17]

ClassTalk [17] was one of the first popular commercial classroom communication tools. With the growth in technology, classroom communication systems evolved to personal response systems such as Educue [48, 6]. As technology increased further, new applications have emerged which work on commonly installed network infrastructures using standard web protocols and able to work on nowaday commonly used hand held devices [6]. An example of such an application is the crowd feedback tool, which is a mobile friendly website, created by Teevan et al. [54].

Classtalk

Classtalk [17] is a classroom communication system which enables the presentation of questions and problem statement in a small group work settings. It follows the cycle described in Figure 2.12 where the teacher starts the cycle by presenting a problem statement or question to the classroom as depicted in Figure 2.13. In a second phase, the classroom is divided in small groups and discusses the problem and possible answers before submitting their answers to the system. When all groups have answered the question, a histogram with the collected answers is shown allowing for a class wide discussion. The

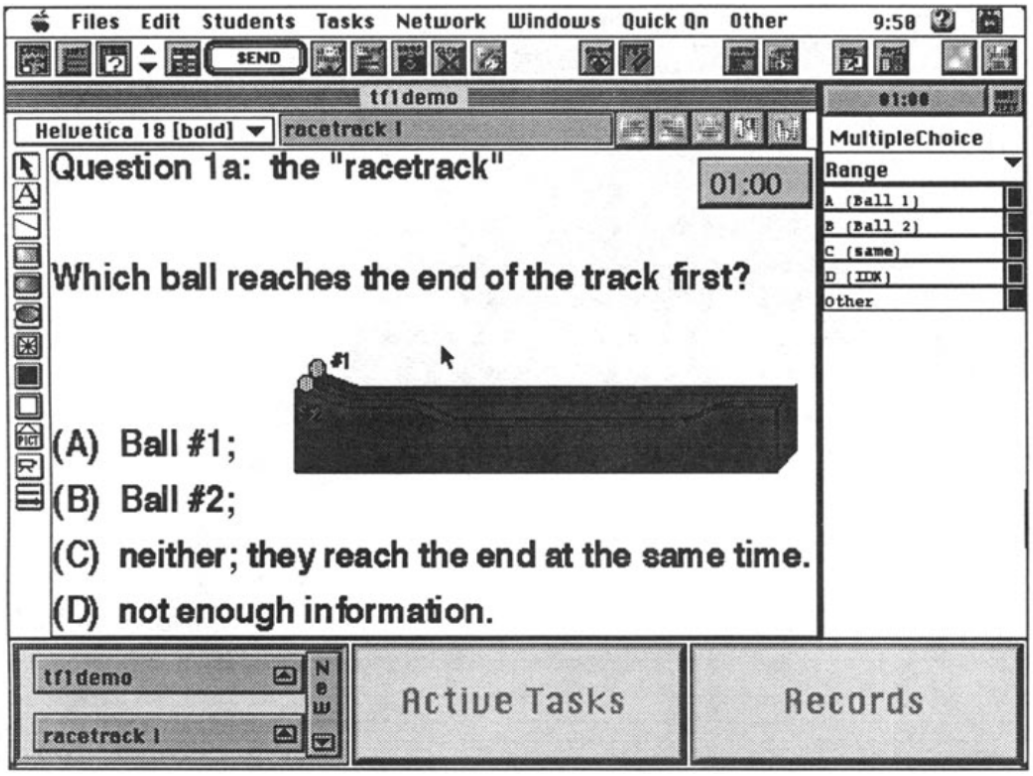

Figure 2.13: Classtalk [17]

end of the cycle comes from the lecturer summarising the discussion and optionally a more in-depth explanation of the solution. The lecturer can then again start the cycle by introducing a new problem statement [17].

Crowd Feedback

Crowd Feedback [54] is a tool which gathers explicit feedback from the audience. The feedback is helpful for the presenter to make adjustments to the presentation of the given explanation without distracting the audience. Feedback is gathered through either liking or disliking the content currently presented. This feedback is then combined and visualised next to the slide as depicted in Figure 2.14. The feedback is displayed constantly next to the slide and is updated in real-time as audience feedback is entered into the system.

The presenter is made aware of events containing interesting feedback, delivered by the audience, by means of a notification on the presenters' smartphone and the display of a badge on the current slide. Interesting feedback

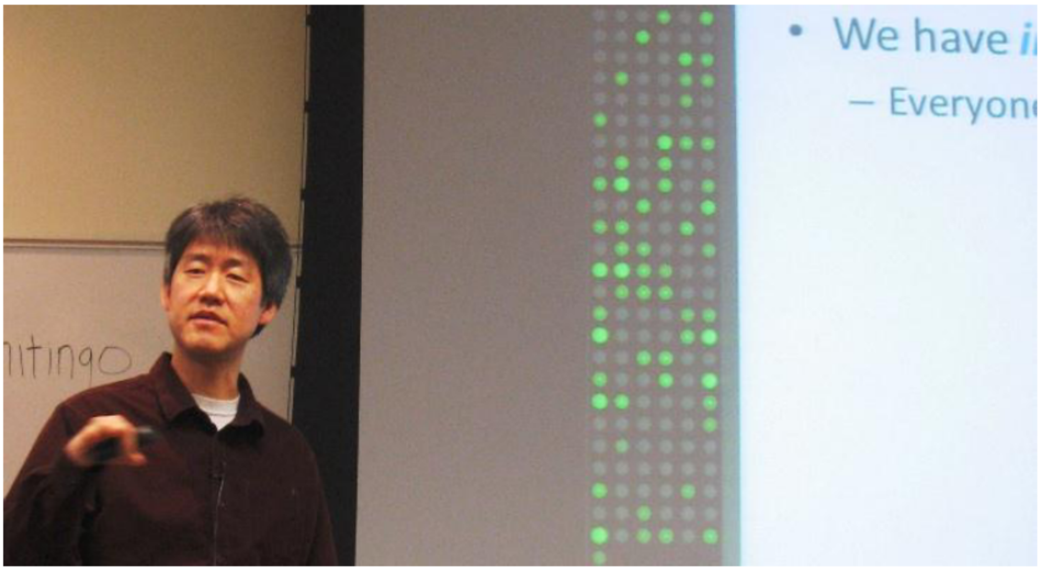

Figure 2.14: CrowdFeedback[54]

is determined through the participation grade. Another trigger is when a significant majority votes in the same manner, either positive or negative. The event may also be fired when the votes are split almost equally between the options.

A study was performed to measure the effects of Crowd Feedback. The Feedback Presentation client is a mobile friendly website. However, since it is a website the user may also navigate to the client through other devices such as laptops or desktop computers when the presentation can be followed over the internet. Most audience members declared that they would have used the device anyway during the presentation. The researchers therefore concluded these devices to be a good medium for audience participation and feedback since the use of the device is already a part of the presentation for audience members. Participants of the study did not pay better attention to the presentation as they would have without the feedback systems. Some even argued that the constant display of the feedback results, as seen in Figure 2.14, were a distraction from the actual presentation. However, participants did feel more engaged [54].

2.4.2 HyperSlide

HyperSlide [18] is a presentation generation tool which generates the presentation based on user input placed in a text file using a simple markup language. The generated presentation is a PowerPoint presentation which can then be opened in, and presented using, PowerPoint. The tool lets the user define scenes and bullet points for that scene. It does not except large portion of text but does also not take action in preventing it. The way the markup language works however, pushes the presenter into using a bullet point approach. Other options for the user are including media into the slides, including hyperlinks and to create verbal linkages between scenes.

The user is also allowed to enter styling rules. the available rules allows to set the font type, colour of the text as well as the background colour and aspect ratio of the background image. The user can also provide a background image with each scene. These images are used as background images and should represent the scene subject. They are also used as thumbnails in navigational slides.

The slides are generated based on three levels and the user is able to navigate vertically, between levels, or horizontally, between slides on the same level. The first level is the story level which shows the thumbnails of the different scenes. Horizontal navigation shows the verbal links between the scenes, vertical navigation is achieved by clicking on the thumbnail of the scene the presenters wishes to discuss next. Beneath the story level, the scene level resides. The scene level shows the image of the selected scene. Horizontal navigation on this level means switching between scenes. The presenter can also go back to the story level or go down to the detail level. The detail level is the lowest level where all details of a certain scene are depicted. Horizontal navigation is the cycling through all the detail slides of the current scene. The presenter can no switch to the detail slides of another scene without returning to the scene level first.

The story level where the slides with the verbal linkages between scenes are shown are meant to be used during rehearsal runs. It is not supposed to be used during the actual presentation. During the presentation, only the first slide of the story level is meant to be used as a way to more easily navigate to a specific scene [18].

2.4.3 NextSlidePlease

NextSlidePlease [51] is a tool which enables the user to create and deliver agile presentations. It focusses on the creation of presentation where the complex relationships between different sections is apparent. It has been critiqued that forcing the presenter to push everything in a linear structure blurs the complex relationships and makes it harder to understand for the audience. It also focussing on the reuse of previously created content as well as time management for the presentation giving. The researchers were aware that time estimations and management would become harder when the presentation was no longer linear in structure. The researchers used a graph based approach which allows complex relationships to be visualised between slides or sections of slides. The graph makes use of weighted edges indicating the time required to explain the target slide and the priority of this slide. This information can be used by the presenter to determine the path to be taken during the presentation.

NextSlidePlease helps with the content reuse by automatically detecting related section or subsection of slides and link them to the current presentation with meaningful weighted edges. In order to facilitate time management and take time constraints into account, the tool suggests paths in real-time during the presentation. It takes the given information about the slides, such as required time and priority, into account to suggest a path. Alongside the path suggestions it also provides time indications for the presenters to help them make decisions and make him aware of the time constraint [51].

2.4.4 MindXpres

MindXpres [44] is a tool for both generating and giving presentations. It separates the content of the slides from the actual presentation. As a result, MindXpres allows for different types of visualisations while the presenter only needs to focus on the content. The visualisation of the content is generated through MindXpres.

The content has to be written in an XML-based language. Within this language, the presenter is able to define a title, bullets points and image among other things without having to worry about how to visualise it for the presentation. The presenter can also integrate videos within the presentation and enhance these using captions and highlighting. MindXpres is extensible through a plug-in system allowing new visualisations to be added.

MindXpres allows the visualisation of source code [43] which will automatically indented and made to fit on the slide through the use of scrollbars. Besides the visualisation of the source code, MindXpres allows for the execution off the source code in an iterative way. This means enabling the presenter to change the input of the program. The presenter is also able to step through the program to more easily explain the flow of the program. During execution, the program state is also visualised such as memory usage [44].

2.5 Critical Analysis

An overview of the related work and the features they employ is depicted in Table 2.1. We limited the features in order not to overwhelm the reader but at least one important feature of every aspect has been given. Before showing the overview of the systems and their features presented in Table 2.1, we first discuss the chosen criteria.

- **Presentation Guidelines**
  - **Movement:** Does the solution allow the presenter to move around? We concluded that when the presenter is tied to the hardware which can not be moved with ease that the presenters movement is restricted.
- **Presentation Interaction**
  - **Freeflow:** Is the presenter able to change the order of the slides or skip slides at any time during the presentation without breaking the flow of the presentation?
  - **Annotate:** Is the presenter able to annotate slides at runtime?
- **Presentation Feedback**
  - **RTF:** Short for real-time feedabck. Can feedback be displayed during the presentation?
  - **PPF:** Short for post presentation feedback. Is feedback generated and displayed to the presenter at the end of their presentation?
- **Dynamic Presentation Content**
  - **Audience:** Does it improve audience participation?
  - **RCM:** Short for runtime chart manipulation. Is the presenter able to change the content of a chart during a live presentation by, for example, including or omitting a parameter?
  - **SR:** short for slide reuse. Can the slides or content of the slides, be reused easily for other slide shows? With easily, we mean without the need to manually copy and paste it in the new slide show.

Name	Movement	Freeflow	Annotate	RTF	PPF	Audience	RCM	SR
Camera-tracked Laser Pointer	✓							
SketchStory			✓				✓	
PaperPoint	✓	✓	✓					
Palette	✓	✓						
CounterPoint		✓						
Fly		✓						
Presentation Trainer				✓				
Presentation Sensei				✓	✓			
Rhema	✓			✓				
Cicero				✓				
Classtalk						✓	✓	
Crowd Feedback						✓		
HyperSlide								✓
NextSlide-Please		✓						✓
MindXpres		✓				✓	✓	✓
PowerPoint								

Table 2.1: An overview of the related work

Looking at Tabel 2.1, one can easily conclude that every application focusses on a specific issue or a specific topic. No application has been found which provides an answer to all these issues. When reading about these applications, it was also easy to notice how closed off all the applications were. Out of all applications which have been discussed within this chapter, only MindXpres [44], with its plug-in system, is extensible and may possibly be integrated within other applications through the creation of a plug-in.
The other applications were all standalone solutions which can not be incorporated within other applications. Also some applications have been discussed which are not meant to be used during a live presentation but were created for training purposes such as Cicero. Besides the applications meant for training, the majority of the applications were meant to be used during live presentations. Each of these applications however, were created in such a way that combining applications in order to gain more of the functionalities, is hard to achieve.
For this reason, we believe that a framework is needed which would allow the creation of applications aimed to solving these problems in such a way that they could be used at the same times as other applications solving another problem. This would allow the developer to still focus on a single issues while the presenter would have opportunity to combine applications and come to a complete solution.

3 Requirements

Within this chapter, we will discuss the requirements gathered from the already existing solutions. Using the solutions that have been tried before, it was possible to extrapolate the requirements for our generic framework for better presenter views, allowing presenters to interact with their presentation content, get real-time feedback as well as visualise external data streams. Looking at the features of PowerPoint and the research which has been conducted surrounding public speaking and slide shows, we were able to determine the issues of the commonly used presentation tools as well as determine the requirements for our framework by looking at the related work and analysing the approach as well as the technology these researchers used we will discuss the requirements and how they apply to the already discussed solutions. Besides using the already existing solutions discussed in Chapter 2, the limitations of the HoloLens also had to be considered.

3.1 Technical Requirements

3.1.1 The HoloLens

As mentioned in Section 1.4, a possible approach is to make use of virtual reality. Movement of the presenter is restricted when their presenter view is located on a device which can not be carried along. When resorting to more lightweight devices such as tablets, we would enable to walk around more. However, eye contact could be lost by the presenter frequently looking down at the presenter view on their tablet. Augmented reality takes away the restriction of physical devices while improving eye contact between the presenter and the audience when the augmentations are located near members of the audience. As a result, the HoloLens is a good choice as a wearable augmented reality platform.

The HoloLens[1] is a wearable head device. For it to remain comfortable to wear, certain decision had to be made concerning hardware[2]. The HoloLens RAM is limited to 2GB. This memory is also used to run the HoloLens operating systems leaving an estimated 900 MB of RAM [3]. The HoloLens only has a passive cooling making it prone to overheating. This reduces the intensity on which an application may run as too heavy computations may cause the processing unit to heat up.

With the memory limitation and taking into account the lack of an active cooling system, the number of operations to be performed on the HoloLens has to be reduced to a minimum. In order to achieve this, all data gathered on the HoloLens is captured and sent to a remote server where the actual operations will be executed.

Ha et al. [24] also lays out some of the issues one has to take into account when using wearable devices for cognitive assistance [24]. It also speaks about the limited resources these devices have and the need to use a separate server for the computations. The paper also discusses the interactivity required within these wearable devices. The human brain is very skilled at cognitive tasks[33]. It is also very susceptible to delays. Delays of only half a second may already cause the user of the device to be distracted.

When looking at Table 2.1. We see that in the first row movement, four applications are marked. These applications are the *Camera-tracked Laser Pointer* [28], *PaperPoint* [49], *Palette* [37] and *Rhema* [53]. Rhema stands out in this list due to the combination of allowing movement while delivering feedback to the presenter. To accomplish this, Rhema has been developed to work on Google Glass which is another wearable augmented reality device. In comparison to the Google glass, the HoloLens has a greater field of vision and offers more possibilities. For this reason, we decided to work with the HoloLens to be able to visualise information to the presenter without limiting their movement.

[1] https://www.microsoft.com/en-us/HoloLens

[2] https://docs.microsoft.com/en-gb/windows/mixed-reality/HoloLens-hardware-details

[3] https://forums.HoloLens.com/discussion/9009/HoloLens-limitations

3.1.2 Modularity and Extensibility

A key requirement is to make the framework modular and extensible. The ability to be able to write a driver for an input device becomes apparent when reading some of the solutions concerning presentation interaction depicted in Section 2.2. These solutions divert from the commonly used input devices, such as the mouse and keyboard, and it is important not to think of the HoloLens as the only possible input device for this framework. It is not possible to predict what the future input devices will be or what researches will want to try. For this reason, the framework requires the ability for a developer to add their input device to the framework.
It must be possible for developers to integrate visualisations of data provided by external applications into the framework which can then be used by the presenter. These visualisations are also referred to as widgets within the framework The applications must also be able to use the different input devices currently plugged into the framework. A mapper must be foreseen to translate the actions off the input devices to actions performed on the widgets.
As mentioned in Section 2.5. Most already existing applications are closed off meaning they can not be incorporated within other applications. Most of these applications even can not be used alongside each other due the nature in which they were developed. This further motivates our decision to have the framework be modular and extensible.

3.1.3 Communication Protocol and Medium

To allow all modules and widgets to interact with each other, a communication protocol and medium has to be established. The protocol must be a standard followed by all modules as well as all clients. The use of such a communication protocol over a shared communication medium allows developers of input modules to communicate with the widgets and vice versa and to easily integrated new modules. This requirement should therefore already offer a solution for the Modularity and Extensibility requirement depicted in Section 3.1.2.
The use of this shared communication medium and communication protocol must allow interaction between the presenter and their audience to take place. It would be opportune if the solution for this requirement would offer a solution for the extensibility requirement discussed in Section 3.1.2 as well. The Communication channels must also enable the use of the framework as a classroom communication system such as Classtalk [17] and Crowd Feedback [54].

The communication protocol is also important due to the limitations on the HoloLens discussed in Section 3.1.1 which imposed the requirement of having to use the HoloLens as an input device and a screen but to send data over to a server for actual computations.
On the other hand, the use of a shared communication medium makes it easier for developers to create real-time presentation feedback applications such as Rhema [53]. As all event taking place in the framework are transmitted over a shared communication medium, widget developers are able to utilise the transmitted information and generate visualisations to the presenter to inform them about their performance.

3.2 Functional Requirements

3.2.1 Graphical User Interface

A user interface is required for the front end applications meant to be used by the presenter and to be visualised on the HoloLlens. It might be cumbersome for the developer having to revert to unity when they are trying to create a simple overlay application. The framework must offer the possibility for developers to create widgets. It would be advantageous if the developer would be able to use known technologies and if the developer were able to test his widget on his own local machine.
It is also important that we support a technology which is mature and offers multiple possibilities for data visualisation since the way data is visualised to the presenter is crucial. This mostly applies to applications meant to be used in real time. One must also be aware of change blindness [50] where the change in the interface may go unnoticed due to how subtle the change is. The presenter may however not be distracted by a sudden change in one of the widgets as this may confuse her, stall the presentation or make the presenter lose focus. The data must be perceived pre-attentively [26], change blindness should be avoided and widgets may not attract the attention of the presenter unless desired by the presenter. An already existing technology stack must be either found or created to enable the creation of widgets with those requirements in mind.
The widgets must also remain separated. Mashing all widgets in one general interface may cause unexpected behaviour or may be confusing for the presenter when looking for specific information of a specific widget. We have to be cautious in avoiding an information overload for the presenter. The presenter should therefore be in control of each separate widget. Whether it should be displayed or not and where it should be located in the room.

3.2.2 Screen Cloning

It may be advantageous to have the widget open on multiple screens but in synchronising states. When an action is initiated on one of the screens, it also gets executed on the other screens. This allows for every screen to serve as an input for the whole widget. Interaction on the HoloLens itself may be hard due to the lack of actual gaze tracking or a moveable cursor. The only way for the user to move the cursor is by moving their entire head until the element they wish to select is in the middle of the presenters screen. The HoloLens also does not have a keyboard. When writing is required, a digital keyboard will pop up on the HoloLens. The user can then interact with the digital keyboard through gaze and the select gesture.
Section 3.1.3 discussed the need for a communication medium and communication protocol to allow different kinds of input devices while Section 3.1.2 discussed the importance of the extensibility for developers to include their own input devices. We can consider the screen cloning as another input device which should publish the events on the shared communication medium to allow the action to be performed on each client. This would allow the use of a touch screen device, such as a tablet, to be used as an interaction device. The use of a tablet would be a solution to have a zoomable interface such as Fly [27] and CounterPoint [22]. Zoomables interfaces resemble interactions with paper which may be more natural to do using a tablet instead of a computer mouse and keyboard. With the use of screen cloning, interactions on a zoomable interface could be performed on a tablet while the audience can still see the results on the presentation screen.

3.2.3 Logging

Every message that has been posted on the communication channel must also be logged and kept in a file as a time line of events. The logging of said events makes it possible to analyse ones performance after the facts. Using timestamps in the log files would allow a more in depth analysis of the performance or would allow a sort of replay experience where the widgets would act as if he's just now receiving those events. The logged data can be used for post presentation feedback systems as discussed in Section 2.3. The logs could also allow a new screen in an active session to catch up to the current state for the screen cloning as discussed in Section 3.2.2.
Logging is important for the creation of post presentation feedback tools such as Presentation Sensei [30]. Using the logs, developers can create visualisation to be visualised on either the HoloLens or any other screen. Feedback can be generated based on how fast the presenter was speaking. The feedback

could show at what point in the presentation, the presenter was speaking either to fast or to slow. Besides the paces of the speech, feedback could be provided about the use of fillers and pauses within the speech. Feedback could be given on the movement of the presenter as well as the gestures used during the presentation. Another idea is the generation of statistics about time spend per slide as well as how often the presenter returned to a slide. Returning to the same slide for a number of times within the same presentation, may an indication that the slide is not clear enough.

3.2.4 Speech

Since one of the goals of this framework on the HoloLens is to enable presenters to walk around and give them more freedom on stage, it removes them from their keyboard and computer mouse. These input devices therefore have to be replaced by another means of input. Speech would be a possible replacement. It would require the widgets to register certain keywords within the framework they want to be notified about. Even further, complex keywords such as sentences containing parameters should be able to be processed. Complex sentences containing parameter would allow for generic command such as a command to go to a certain slide where the number of the slide is the parameter, or to navigate to a certain section with the name of the section as a parameter,

Speech is also important to extract different kinds of information from such as pace, volume and the uses of pauses during the presentation. Gathering this information would enable the developer to foresee the presenter with feedback about his performance.

The HoloLens has a built-in microphone. This microphone is relatively close to the presenter and can be used as speech input. However, due to processing being performed on a separated server, it is not mandatory for the HoloLens microphone to be the speech input. The speech processing unit has been developed with that idea in mind.

3.2.5 Body Tracking

Another aspect to be taken into account is tracking the body of the presenter. This is important for multiple reasons. For one, some of the gestures used during presentations may be considered offensive depending on the culture one is presenting to. It is also a matter of the culture whether or not it is expected, accepted, or wrong behaviour to walk around a lot on stage. Every culture has a different idea of how one should act given different circumstances and this also translates to different expectations from the audience

towards the presenter. The HoloLens could visualise feedback about the presenter's performance on stage if so desired by the presenter. It would be advantageous towards the presenter in both real-time applications as well as in training environments if they would be notified about the anxious fiddling with their hands or the defensive posture they are standing in. To enable the development of applications which rely on the presenter's movement the framework must first be able to capture the movement of the presenter. One such application could be a feedback widget correcting mistakes depending on the selected culture.
Multiple ways to track body movement already exist. There are 2 categories to be discerned. One category is a vision based approach whereas the other category makes use of sensor rich body suits. Microsoft Kinect has been used to develop a vision based approach which includes skeleton tracking of people standing in front of the device [59]. An example of a sensor rich body suit is PrioVR[4] which is intended to map the movement of the wearer onto a virtual avatar in games.

Hand Gesture Recognition

Body tracking can also be extended for hand gesture recognition. Hand gestures are a natural way for humans to interact with each other or to enhance their speech. Gesture recognition is a human computer interaction where the computer measures movements of body parts and tries to interpret them. hand movement tracking is one of the more important research topic within its field.
Using gestures and body movement tracking, the user is once again more free to move around the stage while giving his presentation. One concern however is that the presenter must remain within the detection field of the tracking device when opted for a vision based tracking solution. This concern can be circumvented by using glove-based hand movement tracking [12, 40].

3.2.6 Face Recognition

The HoloLens offers a wide arrange of possibilities augmenting the wearers world view in real time. One of these augmentations may be the scanning of the audience, recognising their faces and add some information about this person floating next to their head. This could be used to locate people who have questions or experience difficulties in understanding the matter. It may also be used to display the answers of the audience when the presenter asked

[4] https://www.vrs.org.uk/virtual-reality-gear/motion-tracking/priovr.html

a question. This would work best with closed multiple choice questions. It would have the same effect towards the presenter as counting a raise of hands but the answers would not be influenced by the other audience members.
HoloFace [29] is an application on the HoloLens which uses facial recognition to augment these faces. Depending on the face expression, different visualisations appear on the HoloLens. This shows that the HoloLens is clearly capable of handling face recognition and can be used to augment information alongside these faces.

Gaze Detection

Gaze of the audience may also be used as an indicator whether or not your audience is focussed. Using gaze detection, it could be determined whether or not a certain audience member is paying attention. When an audience member were to ask a question about a subject in the presentation and a application were to notify that the audience member in question was gazing out the window at the time this subject was handled, may be crucial information for the presenter. In this case the presenter would know that the audience member was not paying attention and is therefore no indication whether or not the original explanation was unclear and needs to be audited. Gaze detection of the presenter could also be advantageous to provide feedback on the amount of time they look at their audience.

Emotion Scanning

Face recognition can also be used for emotion scanning which can serve the same goal as gaze detection towards the audience. Through emotion scanning one may find audience members facing difficulties in understanding the matter, find audience members who are distracted or bored. This information can be useful for the presenter who can adjust his explanation or the pace of the presentation depending on the emotions portrayed by the audience.
Openface [3] is an open source framework for facial recognition. It is able to track the position of a head as well as gaze tracking and emotion scanning. The face is a key medium for non-verbal communication and a lot can be learned from analysing ones face.

3.2.7 Object Recognition

Object recognition as part of the framework would open an new array of interaction possibilities. One may switch to a certain slide of the presentation by looking, grabbing or selecting a real life object. Other possibilities are the augmentation of these object with holograms. An example could be a clipboard with augmented buttons on them enabling interaction with the framework. Another possibility would be to fill a number of detected chairs in a room with a virtual audience to increase the immersion of virtual audience solutions.

4
Implementation

Based on the requirements presented in Section 3, we present the implementation of our framework as well as the implementation of the functionalities.

4.1 Design Choices

The choice has been made early on to make use of common web protocols and web standards. This choice was made since we required a solution that would allow visualisations to be made as well as a communication to happen between different parts of the framework. Choosing the common web protocols and web standards allows the creation of a RESTFull interface, the use of WebSockets, and an easy way of integrating and serving widgets. The use of common web protocols also enables developers to create classroom communication systems within the framework without the burden of learning new integration protocols.
It was decided that the use of HTML, JavaScript and CSS would enable for easier creation of widgets. It would lower the threshold for developers to create widgets using the framework. Multiple JavaScript modules exist for data visualisation. These modules allow for dynamic changes within the visualisation. One such JavaScript module for visualisations is D3.js[1].

[1]https://d3js.org/

It can be further motived due to already existing solutions such as Crowd Feedback [54]. Besides the visualisations, this choice enables the creation of zoomable interfaces.
It was therefore decided that a web based solution would be a good approach for an extensible framework which allows easy creation of widgets. Since we are developing a framework to work with the HoloLens, which is a wearable device augmenting the reality seen by the wearer, precautions had to be put in place to prevent obstruction of the view. It was therefore decided that the background should be transparent and only visualisations of data should be shown.
One of the early complications was the absence of a component for visualising web pages on the HoloLens within our own developed application. The HoloLens does have a browser and one can open the browser to a certain link from within their application. However, this closes the application. One possible solution was to use a third party unity asset such as Awesomium[2] which provides a webview for the HoloLens. These assets however are not free and there was no certainty if we would be able to remove the background to have a transparent overview. Another solution was to run a XAML based webview in the background, take screenshots of the webview and display it in the 3D HoloLens application[3]. As a workaround we execute and render the widgets on a separate server and periodically send screenshots of the widgets over the the HoloLens to be displayed. This solution will be further discussed in Section 4.2.1.

4.2 Architecture

As mentioned in Section 3.1.1, the resources of the HoloLens are limited. It is unable to handle heavy computations as it will kill the application to prevent overheating due to the lack of active cooling. As well as the lack of active cooling, the HoloLens has RAM of which an estimated 900MB can still be used. We therefore decided to offload as much as possible from the HoloLens. The Architecture, depicted in Figure 4.1, consist of four parts. These parts are:

- a server which hosts the widgets
- the real core of the framework which renders the widgets and runs the WebSocket server

[2]https://github.com/bkeiren/AwesomiumUnity
[3]https://forums.HoloLens.com/discussion/comment/13011/#Comment_13011

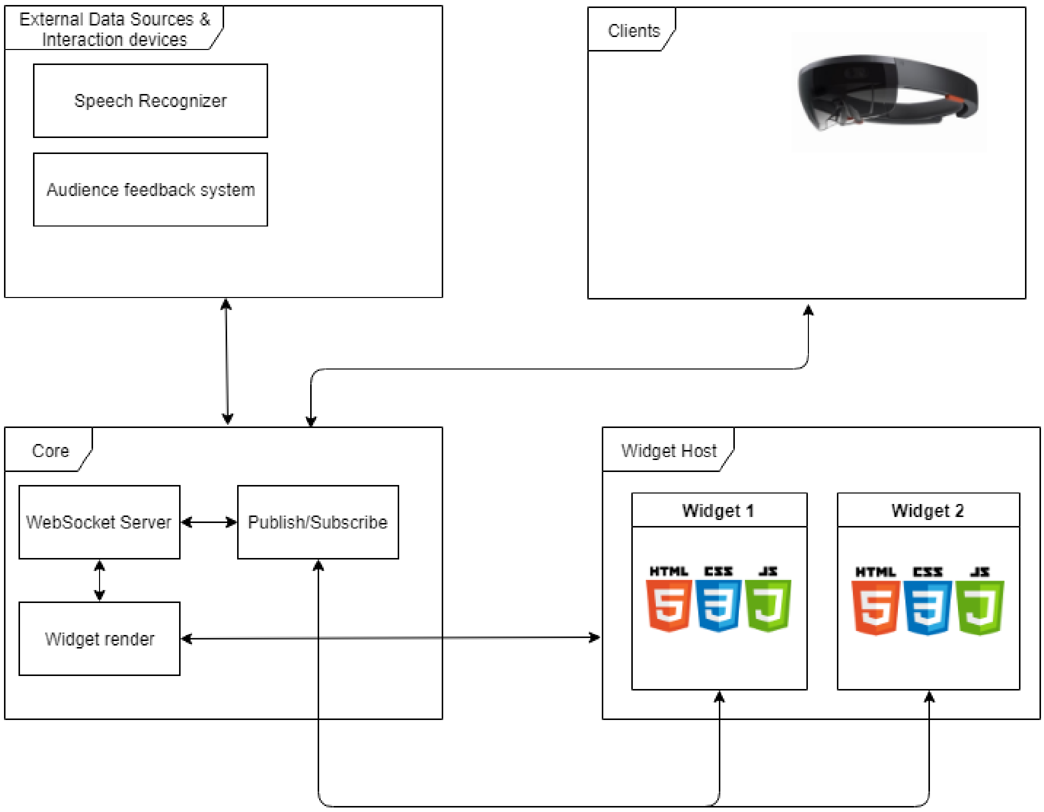

Figure 4.1: The frameworks architecture

- the clients which are the screens on which the widgets will be visualised
- external data sources and interaction devices.

Interaction is also possible through the clients via the widgets themselves. These interactions may be clicking on a button of the widget using a touch screen or a computer mouse, zooming in and out using a touch screen, among others.
Widgets are hosted through a webserver. The reason behind is that widgets are frontend web applications developed in HTML, JavaScript and CSS. A JavaScript module has been developed to help the developers in creation their widgets. The JavaScript module is responsible for connecting the widget to the rest of the framework to enable the widget to receive data from external data sources as well as interaction devices.
The HoloLens is a special type of client due to the absence of a webview that can be integrated into the applications. A webview is responsible for rendering webpages and processing the interactions, such as mouse clicks, on the webpage. The HoloLens does have a web browser. However, the web browser can not be made transparent which would obstruct the presenters' sight. Another aspect to be considered is that no two web browsers can be opened at the same time but it does allow the use of multiple tabs within the browser. This meant that the presenter would have to switch between the chose widgets during the presentation whenever required. We decided to create a graphical user interface on the HoloLens to allow the visualisation of multiple widgets at a time. To achieve this, the widget render has been implemented as part of the core depicted in Figure 4.1. The graphical user interface on the HoloLens has been developed in Unity[4] game engine using the provided *Mixed Reality Toolkit*[5].
To be enable communication between the clients, widgets, interaction devices and external data sources, middleware has been implemented. The middleware is responsible for managing the interaction devices and external data sources, which will be referred to as modules, as well as enabling the communication between all components. The middleware also enforces the communication protocol which is required for the publish/subscribe pattern to work properly. The middleware has been implemented in an extensible way to allow future integrations of new modules and widgets. The widget render is part of the middleware and enables the widgets to be visualised on devices without a webview. The middleware has been developed in Python[6] as it allowed fast iterative development of prototypes.

[4]https://unity3d.com/unity
[5]https://github.com/Microsoft/MixedRealityToolkit-Unity
[6]https://www.python.org/

4.2.1 Middleware

One of the main components is a web server which hosts a WebSocket server. The WebSocket server serves as the communication medium throughout the framework. This component will also act as the browser for the HoloLens as a solution for the lack of a webview in the HoloLens as mentioned in Section 4.2. This component is also is responsible for starting the other components.

The WebSocket is the main communication channel for all modules and widgets. The architecture is based on the publish subscribe pattern where modules can register on the WebSocket and fire events using the communication protocol while the widgets can listen to the WebSocket and bind actions to certain events on the WebSocket. This pattern was chosen since it automatically makes the architecture to be extensible.

The architecture can also been seen as a single bus architecture since all communication relies on a single WebSocket echo server. This is responsible for receiving events and transmitting it to all other connected modules or widgets.

The other main component is a web server hosting the widgets. Widgets are small websites that are connected to the rest of the framework over the WebSocket. The widgets are to be written using web technologies such as HTML, JavaScript and CSS. Each widget must be contained within their own directory. The directory must also contain a JSON file with basic information about the widget. To allow more freedom for the developer, the JSON file may contain an `origin` key. This `origin` key can point to another web server for widgets requiring a sophisticated back end.

HoloLens Webview

As discussed in within the introduction of this chapter, the HoloLens does not have webview available. The suggestion of running an application in the background was used as an inspiration to tackle this problem. In order not to burden the HoloLens with computations that can be executed elsewhere on a different server, and to not constrain the already limited data, a web server has been implemented to be used as the browser for the HoloLens.

The server opens the website using Selenium[7] which automates browser interactions. The driver used in Selenium is a Chromium webdriver[8] which can be run headless. The headless browser works as a regular browser with the exception of a graphical user interface. Interaction on the browser is

[7]https://www.seleniumhq.org/

[8]https://sites.google.com/a/chromium.org/chromedriver/downloads

accomplished through function calls in code. This allows the framework to run without a graphical interface which also speeds ups execution time. To allow widgets to be dynamic with the use of JavaScript, we keep the website open in Selenium. With the use of Chrome devtools[9], The default background can be set. In order to meet our design choice to have a transparent background, the default background was change to be transparent.
The web server acts as a middleware between the widget host server and the HoloLens. This server takes a screenshot of the website from Selenium and responds with this image to the HoloLens. The HoloLens uses this image to overlay it on a cube object to visualise it to the user.
As alternative, an attempt was made to use phantomjs[10] instead of Selenium. PhantomJS is a tool to test websites with. Due to the focus of website testing, PhangromJS lacks the features to easily use it as the webview for the HoloLens. PhantomJS expects a script to execute in order to produce a result stating if the test succeeded or not. It does allow for screenshots to be taken from the website but it does not allow the website to be kept open for an undetermined amount of time and screenshots to be taken on request. Within Selenium, these interactions are possible hence the decision to use Selenium.

Communication Channel and Protocol

The communication channel can been seen as a single bus architecture implemented using a WebSocket server. The WebSocket server keeps track of all of its clients. Every client is allowed to send messages to the WebSocket server which saves the message in a log file before echoing it to all clients. Every module or widget in the framework is therefore connected with each other. This offers future developers freedom to choose the programming language they wish to develop in as the the only requirement is the integration of the WebSocket connection and the communication of events over this connection. Every message over the connection should follow the communication protocol which states that the message must be in a JSON format with four mandatory keys. These keys are:

- **source**: the origin of the event. The name of the module or widget.
- **actionType**: The developers are free to chose any action type names they find appropriate.
- **action**: The name of the action.

[9] https://tatut.github.io/clj-chrome-devtools/api/
[10] http://phantomjs.org/

- **actionData**: supplementary data for the action which listeners may require when executing the corresponding function.

The WebSocket server has been developed in python using WebSocket-server[11] module. Clients are implemented using the WebSocket-client[12] module.
An issue occurred when implementing the WebSocket client for the HoloLens stating that WebSockets were not implemented. Since the HoloLens in its current state does not require the ability to subscribe to events, an HTTP endpoint has been made for the HoloLens to send its events too. These events were then transmitted over the WebSocket. The event has to follow the same protocol as any other module or widget.

Extensibility and Modularity

The framework allows for the implementation of new modules and new widgets. For the widgets, a widget host has been implemented as will be discussed in Section 4.2.2. Within this Section we will discuss the implementation of new modules for the framework. The developer has complete freedom over how to implement the module and in which language. The only constraint is that the module must be able to register to the WebSocket server to publish their events as well as take action upon events from other modules or widgets. The extensibility of the framework is achieved through basing the solution on the publish/subscribe pattern.
The main component of the system is responsible for starting all other modules of the framework. A folder called `drivers` exists where all modules should be placed. Every module must be contained within its own folder. The folder must contain a file `module.json` which contains information of the module. One of the keys within the json file must be `startCommand`. The value of this key should be the string entered in the terminal when one would like to start the module.
The publish/subscribe pattern is a paradigm used in distributed systems. We created the framework to be a distributed system with each module part of the system. This allows for the integration and removal off modules without touching the core of the framework. As is common with the publish/subscribe pattern, there exists a communication medium through which the system can be notified about the occurrence of events. In this case the communication medium is the WebSocket server. The systems does not distinguish between publishers and subscribers. Each connection made with the

[11]https://github.com/Pithikos/python-WebSocket-server
[12]https://github.com/WebSocket-client/WebSocket-client

WebSocket is interpreted as a subscription to be notified about the events even when the module is only interested in notifying the system about the events this module produces. The systems does not keep track which connection wants to subscribe to which events. The WebSocket channel is an echo server which notifies every client of every event it receives itself. The modules are responsible for how they wish to respond to these events [20].

4.2.2 Widget Host

The widget host is a component which has been implemented as a module for the framework and will be started by the main component. The widget host is the web server which hosts the widgets. The widgets must be frontend web applications written using HTML, JavaScript and CSS. Each widget must be contained within their own directory. The main file that should be displayed when navigating to the widget through the browser should be called `index.html`. In order to find the appropriate files belonging to this widget, the system makes use of the referer header in the request. Using this header, the system is able to determine which widget is making the request and search within the directory of this widget. The system is therefore able to separate the stylesheets and JavaScript files from the different widgets.
The widget is also required to have a `widget.json` file in his directory containing information about themselves. In case that the developer wishes to their own back end for the development, one may enter an `origin` key within this json file. When this key is given, all other files within the directory will be ignored and the widget host will perform a redirect to the given origin. The developer is free to do so but must be aware that this may come with a cost in lost functionality.

4.3 Functionality

4.3.1 Widgets

A JavaScript module has been implemented to help with the development of frontend widgets. This framework will handle the WebSocket connection and any communication that passes through there. It offers the developer the ability to register to actions from certain action types. The developer must first register to which action types he wishes to listen through the `fwi.addActionType` function. This function requires the action type to be passed along as a string.
After we added the action type to our widget, we can subscribe to actions

of that action type through the `fwi.on` function. This function expects three parameters.

- **actionType**: The action type. This must be an action type you first registered using the `fwi.addActionType` function.
- **keyWord**: The actual action or keyword within the given action type we want to subscribe to.
- **callback**: The callback function to be called when the event is passed through the WebSocket. The callback is called with the entire event as parameter. The event is an object as defined in Section 4.2.1

The module also automatically sends the subscriptions over the WebSocket. This gives modules, such as the speech module, the appropriate actions to accommodate the subscriptions. When the connection on the WebSocket is already open, the subscription is sent immediately to allow for dynamic adding of subscriptions. All subscriptions which are predefined get published on the WebSocket once the connection has been opened.

Screen Cloning

To allow for the screen cloning functionality, clicks are captured and sent over the WebSocket but not executed. The click gets executed when the click event returns to the widget over the WebSocket from the echo server. To perform a click on the widget, a click event must be sent over the WebSocket. The WebSocket must contain following information:

- **source**: The source where the click originated from, may be the HoloLens, the widget itself or any other interaction device.
- **actionType**: In this case the action type is `click`.
- **action**: The name of the widget the click should be executed in.
- **actionData**: A JSON object containing the coordinates where in the browser, the click should be executed.

4.3.2 Speech

A speech module has been created to allow speech interactions through the framework. The speech module has been implemented as an UDP server that allows multiple client connections. Client connections are supposed to be microphones sending over audio data.

The speech module makes use of a speech recognition module[13] to transform speech to text. In this prototype, we transform it to a string to be able to extract information form it through the use of regular expressions.
Speech module accepts the registration of keywords and tries to retrieve them within the spoken text. This functionality will be further explained in Section 4.3.2. When a keyword has been detected, it notifies the event over the WebSocket.
Besides the speech server, a microphone client has also been developed in Python to use the microphone on the users laptop or desktop. The solution for the division of the speech analysis from the microphone input has been inspired by an example found on github from the user "fopina"[14]
Each module or widget can register keywords they want to be notified off. The registration off such keyword must be a message over the WebSocket using the protocol.

- **source**: The name off the widget or module registering a speech keyword.
- **actionType**: The action type must be `register` in order to make clear that this widget or module is registering to an action.
- **action**: Must also be `register`.
- **actionData**: Must be a dictionary containing following information
  - **type**: The action type that is being registered too.
  - **value**: The keyword that is being registered.

Complex keywords are a collection of words containing parameters. In order to cope with the parameters in keywords, the speech module transforms them into regular expressions. This process is achieved through regular expressions as well. Variables in a keyword must be indicated by the use of curly brackets. An example of a complex keyword could be "go to slide {slideNumber}". In this keyword we have a parameter `slideNumber`. A regular expression is used to find all parameters, indicated through the use of curly brackets. When the keyword is changed into a regular expression, the places where the keywords were located will be changed to `[a-z,A-Z,0-9]+`. The regular expression will match any sequence of characters with the following limitations:

- the sequence must be at least one character long

[13]https://github.com/Uberi/speech_recognition
[14]https://gist.github.com/fopina/3cefaed1b2d2d79984ad7894aef39a68

- the sequence may only contain following characters
  - a-z
  - A-Z
  - 0-9

When a match is found, the values are extracted from the keyword and matched with their parameters. The value of the parameters are passed along using the *actionData* field.

Example

The example keyword "go to slide slideNumber", used in the explanation, will now be discussed more elaborately as a keyword registration from a widget. In phase one, the keyword is registered through the system by sending the following JSON message over the WebSocket.

```
{
  "source": "Widget: revealjs",
  "actionType" : "register",
  "action" : "register",
  "actionData" : {
    "type" : "speech",
    "value" : "go to slide {slideNumber}"
    }
}
```

The string gets converted to a regular expression, The resulting regular expression is:

```
go to slide [a-z,A-Z,0-9]+
```

When the sentence is detected, the parameter gets linked to the value which has been said. The system is then notified of the event through the following JSON.

```
{
  "source": "SpeechAnalysis",
  "actionType" : "speech",
  "action" : "go to slide {slideNumber}",
  "actionData" : {
    "slideNumber" : 51
    }
}
```

The speech module makes use of the root mean square (RMS) of the audio to determine when the user is talking and when not. As long as the user is talking, we add the received audio packets to a buffer. When the user stopped speaking, we send over the audio data to Google Speech Recognition which returns a string. We can use this string to count the number of words that have been said for pace calculations. The RMS can also be used to keep track of the tone and to measure pauses.

Even when no registered keywords were matched within the speech, an action is sent over to the clients to inform them that something has been said. This event contains a timestamp indicating when it happened and a word count indicating the length of the sentence being said.
Besides the events around what has been said, it also measures pauses in the speech using the RMS. When the user has been too quiet, a timer starts measuring the duration of the pause. The timer stops on the next successful speech to text transformation. The module.
An example of both the speech action and pause action will be shown

Listing 4.1: A speech event

```
{
  "source": "SpeechAnalysis",
  "actionType": "info",
  "action": "speech",
  "actionData": {
    "wordCount": 8,
    "timestamp": 1526834838.108087
  }
}
```

Listing 4.2: A pause event

```
{
  "source": "SpeechAnalysis",
  "actionType": "info",
  "action": "pause",
  "actionData": {
    "duration": 7.882960557937622
  }
}
```

Listing 4.1 shows speech action of a sentence with a length of eight words. Listing 4.2 shows a pause action indicating a pause just ended which has lasted just short of eight seconds.

4.3.3 Logging

Every message which is sent over to the WebSocket server gets logged before being echoed to all other clients in the system. The logging of the event is important for two reasons. The first reason is to allow widgets being opened later on during the presentation to catch up with all previous events and to reach the same state on all screens. However, one must take into account that this is no automatic action. The use of this function is at the developers discretion.
The other reason is to allow feedback to be given after the presentation. One could use all the events that have been passed through the systems and

create a dashboard out of it. This dashboard could then be used as an aid in analysing the performance of the presenter.
A replay of events is when useful when a widget lost connection or gets connected later on during the presentation. When requested, the WebSocket server will transmit all events which have occurred thus far to the client. To request the replay events from the current session, one has to sent the following JSON.

```
{
  "actionType":"request",
  "action":"replay",
  "source":"Widget: Words per Minute",
  "actionData":{}
}
```

Within `actionData` an optional key `sessionName` could be provided to request the replay of a specific session in the past. this is mostly useful for post-presentation feedback widgets.

4.3.4 HoloLens

The HoloLens interacts with the web server, which also serves as its webview. When a screenshot is returned, it is overlayed over a cube. The HoloLens constantly updates the widget by requesting a new screenshot every time the previous screenshot has been received. The user is able to click on the widgets. The gaze of the user is use to determine where the click should be executed. The resolution is limited which causes the coordinates to be an estimation.

Figure 4.2: The first view on application start up

The HoloLens is to be worn during the presentation by the presenter and therefore, a graphical user interface had to be made for the presenter to be able to interact with the framework. Figure 4.2 depicts the interface of the framework on start up. The picture shows the screen on which is being presented with to the right, virtual buttons in augmented reality which can only be seen by the presenter. These buttons are toggle buttons and can be used to toggle the visualisation of the corresponding widget on or off. A button is created for every widget in the framework.
When the presenter selects a button of which the corresponding widget is not visualised, the HoloLens will create a virtual `cube`. The button can be selected with the use of an air tap[15] gesture. The virtual cube will be used to display the widget on. The presenter is able to drag the visualisation around after using the HoloLens tap and hold[16] gesture.

[15] https://docs.microsoft.com/en-us/windows/mixed-reality/gestures#air-tap
[16] https://docs.microsoft.com/en-us/windows/mixed-reality/gestures#composite-gestures

Figure 4.3: The virtual RevealJS widget alongside the presentation screen

Figure 4.3 depicts the HoloLens application in working with in the middle the physical presentation screen. To the left of the screen, a virtual presenter view is being displayed. The presenter view is implemented using RevealJS[17]. The presenter view displays the current slide to the left, the next slide in the top right corner and presenter notes in the bottom right corner. RevealJS has been modified to use the WebSocket to communicate between the presentation screen and the presenter view. This was required to keep the presenter view and presentation screen in sync across the different devices. This makes it possible for the presenter view to be also opened on a tablet, smartphone as well as a computer or laptop.
The HoloLens can visualise different widgets at the same time. Due to the background being made transparent, widgets can be placed over each other or on any location without blocking the view of the presenter.

[17] https://revealjs.com/#/

5 Evaluation

The usability of the framework has been evaluated by using the *The Computer System Usability Questionnaire* [32]. Since it was the framework that was being evaluated, the study required the participants to have some development experience.

The study consisted of five phases. The first phase consisted of an introduction to the problem and an explanation of the context. An explanation was given on how PowerPoint limits the presenter. In order to convince the participant, a small explanation was given about the different aspects such as the presentation guidelines discussed in Section 2.1 and how tools as PowerPoint make it hard to follow these guidelines by offering limited interaction possibilities and no dynamic content. The lack of tools to help with bettering one self as a public speaker has also been explained. The explanation focussed on the limitations but no other research to solve these issues had been given in order not to limit the creative thinking of phase four.

In the second phase, participants were asked to fill out a personality form to gauge their knowledge of programming in general and their experience in web development more specifically. The personality form also required the participants to fill out their gender. The personality form and the results from the *The Computer System Usability Questionnaire* have not been linked together to ensure the anonymity of the results.

The third phase was the explanation phase. The participant was given an explanation of the workings of the framework, the patterns behind it and how both the *module manager* and the *widget hoster* module work.
In the fourth phase, the participant was asked to think of a module or widget they would like to integrate in the framework and were given the opportunity to touch the code to explore the feasibility and to get an idea of how the implementation would be achieved. During this phase, the participants were free to ask questions or discuss their ideas with the framework developer to gain further insights on how to achieve the implementation of their idea.
The fifth phase consisted of the participants sharing their conclusion. The fifth phase started when the participant concluded that enough information had been acquired to determine the complexity for integrating their idea into the framework. The participant was given the *The Computer System Usability Questionnaire* [32] and asked to answer the nineteen seven-point Likert scale questions according to their opinion.

5.1 Participants

All participants were computer science students. Therefore they all had at least some development experience. This was a requirement since the study resolved around the evaluation of the framework. The participation pool consisted of both Master and Bachelor students. In total, 6 students participated in the study, of which two female and 4 male students, with each between a year and a half to twelve years of development experience with a median of three and a half years. Out of all participants, 50% stated their core activity in development to be *back-end development*, 33% had no chosen main activity and stated that working on *projects for classes* to be their core development activity. The last 17% stated their core activity to be *software development*
The personality form also gauged the familiarity of the participants with several technologies relevant for the framework. Only 33% of the participants had ever heard of the publish/subscribe pattern on which the architecture of the framework is based. All participants had experience with both JavaScript and JSON and all but one had experience with HTML and CSS as well. 67% had worked with WebSockets before. The participants were also asked to grade themselves on a seven-point Likert scale how familiar they were with distributed systems on the one hand and with web development on the other hand.
The participants gave themselves, with regards to distributed systems, an average score of 3,16 with a standard deviation of 1,34. The median was 3,5.

With regards to web development, the average participant gave a score of 4,16 with a standard deviation of 1,07. The median was 4. The participation pool is a mixture of both experienced and less experienced developers with differences in their familiarity with the technologies used in this framework.

5.2 Results

Based on the nineteen question, four scores have be calculated. The first score is determines how effective the framework is. The second score is how good the framework is documented and how well it provides information when errors occur. The third score shows how easy it is to use the framework. The last score is an overall score of the framework. The scores are calculated based on the answers given to us by the participants. Each question had to be answered using a seven-point Likert scale with one being the lowest score and seven the most positive score.

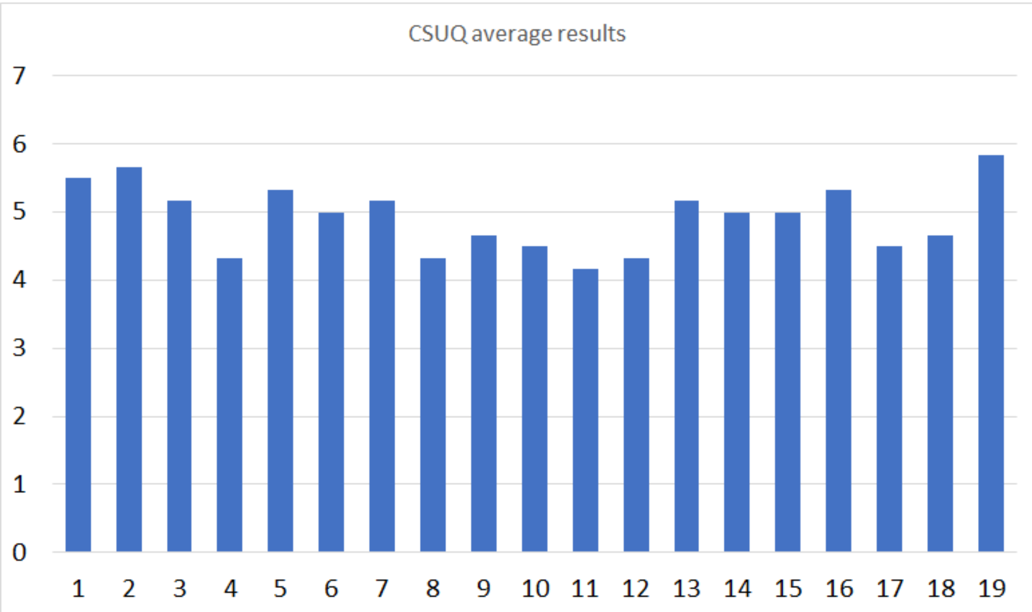

Figure 5.1: Averages scores of the CSUQ questionnaire

Figure 5.1 shows the scores given by the participants. As we can see, the framework received an overall positive evaluation with no score lower then four. The set of nineteen questions can be grouped in four different sections. Questions one to eight are used to calculate the Effectiveness. Questions nine to fifteen are grouped together to calculate the documentation of the framework as well as the available visualisation of available information and error handling. Question sixteen to eighteen are used to calculate the ease of use of the framework. The last grouping are all nineteen questions which are used to calculate the overall score of the framework.

Effectiveness	5,06
Information and documentation	4,69
Ease of use	4,83
Overall	4,93

Table 5.1: The scores of the CSUQ

Table 5.1 displays the four different scores which could be calculated from the *The Computer System Usability Questionnaire*. This shows that the framework received an overall positive evaluation but shows room for improvements to be made. The effectiveness received a score of 5,06 on the 7 point Likert scale, however some participants noted that some functionalities which they perceived to be core functionalities, were missing from the framework. Functionalities that had been named and are not yet implemented were face recognition and gesture recognition.
No real documentation of the framework exists yet, this may be the cause of a lower *information and documentation* score. Some Participants stated that however, the errors are handled in a informative way which helps them correct mistakes, could still be further improved by informing the system of other useful events. The ease of use score may have been influenced by the difference in experience of the participants and the difference in knowledge about the *publish/subscribe pattern*. One participant found that the framework relied too much on the technical knowledge of the user. They argued that non technical people are also required to give presentations and they would like to create their own visualisations as well.
At the end of the *The Computer System Usability Questionnaire*, participants were also asked to list up to three positive and up to three negative aspects of the framework. The most frequently mentioned positive aspect noted by the participants is the modularity of the framework and how easy it is to add new widgets and modules within the framework. Other positive aspects mentioned were:

- the freedom the framework offers to the developer. E.g. modules can be developed in any programming language which supports the use of WebSockets.
- the ease of learning and using the framework.

The most frequently mentioned negative aspect is the lack of documentation and a lack of informational messages through the systems. Informational messages that were deemed missing included information when a new widget or module established or lost their connection to the framework. This was a fair complaint as the framework is a distributed system which allows modules to be run on different physical servers if desired. One also argued that without this information, one is unable to check requirements. This makes the user responsible for handling all the requirements as the systems is not able to. Other negative aspects mentioned were:

- the user must be technical capable disregarding their goal.
- possibility of a name collisions between modules created by independent developers.

5.3 Participant Applications

In phase four of the study, participations were asked to conceptualise an application they wanted to include within the framework. When the participants were unable to come up with an idea of their own, they were informed about some of the related work discussed in Chapter 2. Most participants were able to come up with an idea themselves. Still most ideas mentioned resembled research discussed in the related work in Chapter 2.
One idea was to use the camera of the Hololens to capture written text on a paper version of the presentation. The text would then have to be processed though *optical character recognition* to digital text. The digital text should then be saved on the slide. The application would have to recognise the the slide and calculate the location on the digital version in order to place the annotation on the correct place. The slide being annotated did not have to be the slide currently displayed to the audience. Another feature the application would have was recognition of the slide the presenter is gazing at in order to navigate and display the slide to the audience. This idea resembles PaperPoint [49] where annotations can be made through the use of a digital pen. It also resembles Palette [37] which allows navigation through the scanning of the slide one wants to display.
Another participant was inspired by courses where a number of proofs are given and have to be studied. The participant opted that during a number of courses, these proofs are being written out on the blackboard and the students are required to follow along in their notes. However, the lecturer is able to clear sections of a certain statement on the blackboard and adjust the original statement with the correct information of the next step in the

proof. This makes it harder for the student to follow along and to break it down again in all the different steps. The participant opted for an application where each step of the proof was captured and logged within the framework and to allow the students follow along on their own screen and to individually go through each off the steps off the proof. It would also help the lecturer who often has to go back and explain a previous step. The system should allow the lecturer to copy the current state of his proof and have the current state of the proof visible side by side with the copy where the lecturer can then revert steps to give a more in depth explanation linking back to how that step helped reach the current state. This idea can be seen as a more advanced version of SketchStory [31].
A third idea was to allow interactions through gesture recognition to allow both navigation and real-time manipulations within the presentation. These manipulations include the alteration of graphs, charts or any other visualisation as well as changing the order of the slides to be displayed. Navigation would be linear by default, going to either next or previous slides, but the presenter would also be able to skip slides more easily by using two different gestures. One gesture to go to the next slide and one to display the currently selected slide to the audience. The participant stressed that they wanted the gestures to be minimal and really subtle. The presenter should not be required to perform large gestures to change the content of a slide that is not currently being displayed to the audience. Tap[1] might be a good device for such interactions. However, due to its novelty some research has to be performed before such claims can be made.
The next participant found inspiration in the feedback received when presenting. The participant had been made aware that they jump around in place quite often and fiddle a lot which distracts the audience. This corresponds with the guidelines discussed in Section 2.1, which state that a good posture is important and shifting weight from one foot to another is seen as distracting. The participant noted that it is hard to focus on this flaw. They argued that a choice had to made to either focus on the behaviour and movement during the presentation or the content which should be said. The participant therefore wanted an application which would analyse posture and movement of the presenter and notify him in real-time to correct the mistake. This would allow the presenter to focus on which content must be delivered through their speech while still being able to improve their behaviour and how it is delivered. This is what both presentation sensei [30] and presentation trainer [47] achieve. The difference here is that the participant does not wish for a separate training application but an real-time feedback sys-

[1] https://www.tapwithus.com/

tem during the actual presentation as an help to remind the presenter of his mistakes or to make the presenter aware when a mistake occurs in order to make adjustments and give an overall better presentation.
The following participant wanted to clone a system they had already seen being used during a presentation. The participant however found that the system was a single purpose closed application and wanted the benefits of this application in an open framework where it could be used in collaboration with other applications. The application is a voting system in which the presenter can ask a question and the audience is asked to vote through the framework using a separate view. The participants wanted to add a feature to allow the presenter to see the votes progress in real-time. The real-time visualisation would allow the presenter to already prepare the explanation to be given at the end of the voting process. The presenter should be able to close the poll and display the results in either a predefined visualisation, or choose a visualisation from a selection menu. The participant had also argued that this should be further expanded to allow open questions and answers. In their experience, students are often afraid to answer the question given to them by the lecturer. A fear the participant had experienced themself. The participant noted that there were always the same few students giving the right answer to each question and feared to become embarrassed by giving the wrong answer in an effort to actively participate. This is known as social phobia which has been discussed in Section 1.3.2 . This also has negative consequences for the lecturer who does not know the general understanding of the subject by the entire audience. Therefore, the application should be expanded by allowing the students to anonymously asking questions on which other audience members can then vote.
The last idea included facial recognition and emotion scanning of the audience to provide real-time feedback to the presenter. The participant argued that often people refuse to ask questions or notify the lecturer when they are experience trouble in understanding the subject. This due the social phobia as discussed in Section 1.3.2. The participant wanted to create an application to allow the presenter to see if people were having difficulties in understanding the matter by analysing the audience. He argued that body posture of the audience members can also be analysed giving valuable information to the presenter. It is known that the crossing your arms is a defensive posture and to be avoided when presenting since your shielding yourself from your audience. On the other hand, when the audience start shielding themselves from you during your presentation you may have to reconsider how your delivering your message. He therefore opted the use of the camera on the HoloLens to scan the audience member their posture as well as faces and analyse it for

emotions, defensive posture and other valuable information which could be extracted. This should then be visualised towards the presenter in an easy to understand way to help their adapt his presentation to gain an a better connection with the audience.

6 Conclusion

Within this chapter we will discuss the contribution made within this thesis, the complications we encountered and discuss future work.

6.1 Contributions

We developed an extensible framework which allows the integration of external data sources and the development of widgets. These widgets are visualisations which be made available to the presenter in augmented reality through the use of the HoloLens. We have been inspired by the related work discussed in Chapter 2. The related work and their approach has been studied, evaluated and used as a baseline to determine the requirements of our framework explained in Chapter 3. Following these requirement we implemented an extensible distributed framework which, through the use of augmented reality on the HoloLens allows:

- presentation guidelines to be followed
- multiple ways to interact with the presentation content
- real-time feedback to be given as well as post presentation analysis to be made
- the visualisation of external data sources

Within Section 2.5, criteria were determined to compare different systems with each other. We will now use these criteria again in an analysis of our framework.

- ✓ **Movement:** As a result of augmented reality and the use of the HoloLens, the presenter is no longer constrained in their movement by immobile hardware
- ✓ **Freeflow:** Due to the extensible nature of the framework, different interaction devices, such as PaperPoint [49], can be integrated allowing freeflow presentations to be given.
- ✓ **Annotate:** As mentioned in the previous criteria, PaperPoint could be integrated within the system which would enable the annotation of slides
- ✓ **Real-time feedback:** Data can be send over the communication medium allowing widgets to visualise feedback as the actions take place in real-time
- ✓ **Post presentation feedback:** All actions and event which happen during the presentation are logged. This allows for post presentation analysis and feedback to be displayed from any of the previous presentations
- ✓ **Audience:** Widgets can be created for visualising external data streams. An external data stream may be classroom communication systems through audience feedback can be gathered
- ✓ **Runtime chart manipulations:** As widgets are created using HTML, JavaScript and CSS. This makes it possible to create charts through visualisation modules such as D3.js which allow runtime changes to take place
- ✓ **Slide reuse:** Presentations can be delivered as an external data stream which allows the separation of presentation content and its visualisation through the means of a widget.

6.2 Challenges

Working with the HoloLens itself was a challenge. It is not yet commercialised and only in hands of certain companies and research laboratories for early development or research purposes. The Hololens is a work in progress and the current version can be seen as an early alpha. One of the complications encountered was the lack of support when issues were encountered. It is safe to assume that most developers revert to a search engine when a complication in their code arises and that platforms as stackoverflow[1] often provided a useful answer. Due to the limited amount of people working with the HoloLens, finding solutions to certain issues proved to be more challenging.

6.3 Future Work

The system is developed to be extensible. An effect of this decision is that each functional requirements as described in Sections 3.2 can be developed as separated modules within the system.

6.3.1 Body Movement and Gesture Tracking

A module should be implement to track the presenters body movement and to provide gesture interactions. The HoloLens itself does not provide enough functionalities to properly track the wearers movement or to have proper gesture input.

The HoloLens is only able to process hand gestures which have been executed right in front of the HoloLens. This means that executing a gestures would obstruct the view of the presenter as well as break eye contact with the audience. The author of this thesis also found it unnatural to perform gestures right in front of their face. Also in order not to strain the limited resources of the HoloLens, the decision has been made that gesture recognition was best to performed on a different device. The camera of the HoloLens is also only able what the presenters sees meaning that if the decision was made to use the camera of the HoloLens, gestures still would have to be executed in front of their face. The author therefore feels that the gesture tracking should be implemented using a different input device.

The same applies to Body movement. Body movement tracking using the HoloLens can be achieved by using the gaze of the presenter and trying to calculate the distance between certain objects within the field of vision of the camera. Using these distances, the HoloLens could then calculate the position of the presenter. Doing this continuously would allow for movement to be tracked. The camera feed could also be send over to a remote server where the computations could then be executed. However, other input devices for the body tracking already exists and it would be advisable to make use of

[1] https://stackoverflow.com/

these systems. Real-time hand gestures recognition can already be achieved through the use of a singular colour camera [12] while skeleton tracking can be achieved through the Microsoft Kinect [59].

6.3.2 Face Recognition

Face recognition could be a great addition to the framework as it would allow the scanning of emotions, detect the gaze of the audience and provide feedback to the presenter which can be used to adapt and improve the presentation.
Some research already has been performed in the field of face recognition though the HoloLens such as HoloFace [29]. HoloFace is able to recognise faces and augment them in real time by providing them with a digital mask or to have a fire animation when the mouths opens. The system is also able to detect when the subject smiles. It relies on the CANDIDE-3 face mode [1] which is a face mask created to allow coding around face models. Cardiolens is able to use to power of the HoloLens and face recognition to calculate the heart rate of a patient. It calculates the heartbeat through colour changes on the face of the subject [36].
Face recognition should be performed on the HoloLens itself to allow the augmentation of the faces in mixed reality. It is believed that such a face recognition module could be implemented on the HoloLens itself and interact with the framework. Applications such as Cardiolens [36] and HoloFace [29] prove that the HoloLens is able to handle such computations. Due to the overall nature of the framework and the offloading off the computations where possible, it is assumed that such a module would fit within the framework and run on the HoloLens itself without issues.
Face recognition has also been mentioned by the participants during the study. We can therefore conclude that people are interested in using face recognition within the framework to further improve their presenter view.

6.3.3 Emotion Scanning

Emotion scanning can be achieved through through analysis of the speech [14] as well as through scanning the subjects face [3]. Emotion scanning has been mentioned by a participant of the study. This participant wanted to include a feedback widget to inform the presenter of the audience perception of the presentation. Emotion scanning can determine whether a person is bored, confused, anxious among others. Using this information, the presenter can change their approach, change the order of slides to be presented or revert to a detailed backup slide containing more information.

6.3.4 Object Recognition

Object recognition could be used to detect different objects such as chairs. If chairs could be detected, an application could be implemented to show a virtual audience on some of these chairs. A virtual audience can now already be implemented as widget, however, through object recognition a more immersive virtual audience can be implemented.
Object recognition could also be used to allow different ways of interacting with the presentation content. It would allow actions with the presentation to be linked to the position or movement of a certain object. For example, an automated timer which starts when the ball starts rolling and stops when the ball stops which could be use in physics experiments.

6.3.5 Screen Cloning Restrictions

At the current stage of development, no restrictions are enforced upon the screen cloning functionality. As a result, the presentation view can not be shared safely with the audience as they will be able to interact with these widgets which will influence the presentation view for the presenter. This would allow the current presentation to be shared among audience member and enable to see the same slide as the presenter. This could be useful for taking notes and automated tracking on which slide the notes were taken at what time during the presentation.

6.3.6 PaperPoint

As mentioned by one of the participants in the study. An application such as PaperPoint [49] can be integrated within the framework to allow annotation of slides and a more intuitive way of navigating through the slide show. An application as PaperPoint could be integrated by creating an external data source which transmits actions over the WebSocket server to allow widgets, such as the presentation screen, to react. These actions can be use to annotate the slide as well as navigating to the selected slide.

Bibliography

[1] Jörgen Ahlberg. Candide-3. An Updated Parameterised Face. *Technical Report LiTHISY-R-2326, Linköping University*, 1:1–16, 2001.

[2] Michael Alley and Kathryn A. Neeley. Rethinking the Design of Presentation Slides: A Case for Sentence Headlines and Visual Evidence. *Technical Communication*, 52(4):417–426, 2005.

[3] Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. OpenFace: An Open Source Facial Behavior Analysis Toolkit. *2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016*, 2016.

[4] Scott Bateman, Regan L Mandryk, Carl Gutwin, Aaron Genest, David Mcdine, and Christopher Brooks. Useful Junk ? The Effects of Visual Embellishment on Comprehension and Memorability of Charts. *Human Factors*, pages 2573–2582, 2010.

[5] Ligia Batrinca, Giota Stratou, Ari Shapiro, Louis-Philippe Morency, and Stefan Scherer. Cicero - Towards a Multimodal Virtual Audience Platform for Public Speaking Training. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8108 LNAI:116–128, 2013.

[6] Ian D Beatty. Transforming Student Learning with Classroom Communication Systems. *EDUCAUSE Center for Applied Research*, 2004(3):1–13, 2005.

[7] Benjamin B Bederson. PhotoMesa: A Zoomable Image Browser Using Quantum Treemaps and Bubblemaps. *The Craft of Information Visualization*, pages 66–75, 2003.

[8] Benjamin B. Bederson and James D. Hollan. Pad++: A Zooming Graphical Interface for Exploring Alternate Interface Physics. *Proceedings of the 7th annual ACM symposium on User interface software and technology*, pages 17–26, 1994.

[9] Fadi Biadsy, Julia Hirschberg, Andrew Rosenberg, and Wisam Dakka. Comparing American and Palestinian Perceptions of Charisma Using Acoustic-Prosodic and Lexical Analysis. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, volume 2, pages 925–928, 2007.

[10] Fadi Biadsy, Andrew Rosenberg, Rolf Carlson, Julia Hirschberg, and Eva Strangert. A Cross-Cultural Comparison of American , Palestinian , and Swedish Perception of Charismatic Speech. *Proceedings of the Fifth Meeting of the Speech Prosody Special Interest Group of the International Speech Communication Association (Speech Prosody 2008).*, (May 2014), 2008.

[11] Charles C Bonwell and James A Eison. *Active Learning Creating Excitement in the Classroom.* 1991.

[12] Feng Sheng Chen, Chih Ming Fu, and Chung Lin Huang. Hand Gesture Recognition Using a Real-time Tracking Method and Hidden Markov Models. *Image and Vision Computing*, 21(8):745–758, 2003.

[13] Mathieu Chollet, Giota Sratou, Ari Shapiro, Louis-Philippe Morency, and Stefan Scherer. An Interactive Virtual Audience Platform for Public Speaking Training. *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems, AAMAS 2014*, pages 1657–1658, 2014.

[14] Laurence Devillers, Lori Lamel, and Ioana Vasilescu. Emotion Detection in Task-Oriented Spoken Dialogs. 2000.

[15] Jean-luc Doumont. The Cognitive Style of PowerPoint: Slides Are Not All Evil. *Technical Communication*, 52(1):64–70, 2005.

[16] Allison Druin, Jason Stewart, David Proft, Ben Bederson, and James D. Hollan. KidPad. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '97*, (January):463–470, 1997.

[17] Robert J Dufresne, William J Gerace, William J Leonard, Jose P Mestre, and Laura Wenk. Classtalk: A Classroom Communication System for Active Learning. *Journal of Computing in Higher Education*, 7(2):3–47, 1996.

[18] Darren Edge, Joan M. Savage, and Koji Yatani. HyperSlides: Dynamic Presentation Prototyping. *Proceedings of the SIGCHI Conference*

on Human Factors in Computing Systems (CHI 2013), pages 671–680, 2013.

[19] Paul M. G. Emmelkamp, Mary Bruynzeel, Leonie Drost, and Charles A. P. G van der Mast. Virtual Reality Treatment in Acrophobia: A Comparison with Exposure in Vivo. *CyberPsychology & Behavior*, 4(3):335–339, 2001.

[20] Patrick Th. Eugster, Pascal A. Felber, Rachid Guerraoui, and Anne-Marie Kermarrec. The Many Faces of Publish/Subscribe. *ACM Computing Surveys*, 35(2):114–131, 2003.

[21] David K. Farkas. Toward a Better Understanding of PowerPoint Deck Design. *Information Design Journal*, 14(2):162–171, 2006.

[22] Lance Good and Benjamin B. Bederson. Zoomable User Interfaces as a Medium for Slide Show Presentations. *Information Visualization*, 1(1):35–49, 2002.

[23] Masataka Goto, Katunobu Itou, and Satoru Hayamizu. A Real-Time Filled Pause Detection System. *Eurospeech*, pages 227–230, 1999.

[24] Kiryong Ha, Zhuo Chen, Wenlu Hu, Wolfgang Richter, Padmanabhan Pillai, and Mahadev Satyanarayanan. Towards Wearable Cognitive Assistance. *Proceedings of the 12th annual international conference on Mobile systems, applications, and services - MobiSys 2014*, (December):68–81, 2014.

[25] John Hattie and Helen Timperley. The Power of Feedback. *Medical Education*, 44(1):16–17, 2010.

[26] Christopher G. Healey, Kellogg S. Booth, and James T. Enns. Visualizing Real-time Multivariate Data Using Preattentive Processing. *ACM Transactions on Modeling and Computer Simulation*, 5(3):190–221, 1995.

[27] David Holman, Predrag Stojadinovic, Thorsten Karrer, and Jan Borchers. Fly: An Organic Presentation Tool. *Chi 2006*, 2:863–868, 2006.

[28] Carsten Kirstein and Heinrich Müller. Interaction with a Projection Screen Using a Camera-Tracked Laser Pointer. *Proceedings 1998 MultiMedia Modeling. MMM 1998 (Cat. No.98EX200)*, pages 191–192, 1998.

[29] Marek Kowalski, Zbigniew Nasarzewski, Grzegorz Galinski, and Piotr Garbat. HoloFace: Augmenting Human-to-Human Interactions on HoloLens. 2018.

[30] Kazutaka Kurihara, Masataka Goto, Jun Ogata, Yosuke Matsusaka, and Takeo Igarashi. Presentation Sensei: A Presentation Training System using Speech and Image Processing. *ICMI 2007: Proceedings of the 9th international conference on Multimodal interfaces*, pages 358–365, 2007.

[31] Bongshin Lee, Rubaiat Habib Kazi, and Greg Smith. SketchStory: Telling More Engaging Stories with Data through Freeform Sketching. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2416–2425, 2013.

[32] James R. Lewis. IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. *International Journal of Human-Computer Interaction*, 7(1):57–78, 1995.

[33] Michael B. Lewis and Andrew J. Edmonds. Face Detection: Mapping Human Performance. *Perception*, 32(8):903–920, 2003.

[34] Paul Luff, Christian Heath, Moira Norrie, Beat Signer, and Peter Herdman. Only Touching the Surface. *Proceedings of the 2004 ACM conference on Computer supported cooperative work - CSCW '04*, 6(3):523, 2004.

[35] Jo Mackiewicz. Audience Perceptions of Fonts in Projected PowerPoint Text Slides. *IEEE International Professional Communication Conference*, pages 68–76, 2006.

[36] Daniel McDuff, Christophe Hurter, and Mar Gonzalez-Franco. Pulse and Vital Sign Measurement in Mixed Reality Using a HoloLens. *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology - VRST 2017*, pages 1–9, 2017.

[37] Les Nelson, Satoshi Ichimura, Elin Ronby Pedersen, and Lia Adams. Palette: A paper Interface for Giving Presentations. *Conference on Human Factors in Computing Systems - Proceedings*, (May):354–361, 1999.

[38] Anh-Tuan Nguyen, Wei Chen, and Matthias Rauterberg. Feedback System for Presenters Detects Nonverbal Expressions. *SPIE Newsroom*, 2013.

[39] Arina Nikitina. *Successful Public Speaking.* Number 105. 2012.

[40] Vladimir I Pavlovic, Rajeev Sharma, and Thomas S. Huang. Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):677–695, 1997.

[41] Brian E. Perron and Alyson G. Stearns. A Review of a Presentation Technology: Prezi. *Research on Social Work Practice*, 21(3):376–377, 2011.

[42] David-Paul Pertaub, Mel Slater, and Chris Barker. An Experiment on Public Speaking Anxiety in Response to Three Different Types of Virtual Audience. *Presence: Teleoperators and Virtual Environments*, 11(1):68–78, 2002.

[43] Reinout Roels, Paul MeÅ§tereagÄČ, and Beat Signer. An Interactive Source Code Visualisation Plug-in for the Mindxpres Presentation Platform. *Communications in Computer and Information Science*, 583:169–188, 2016.

[44] Reinout Roels and Beat Signer. MindXpres: An Extensible Content-Driven Cross-Media Presentation Platform. *Proceedings of the 27th BCS Conference on Human Computer Interaction (HCI 2014)*, pages 215–230, 2014.

[45] Andrew Rosenberg and Julia Hirschberg. Acoustic/Prosodic and Lexical Correlates of Charismatic Speech. *Proceedings of Interspeech*, pages 513–516, 2005.

[46] Barbara O. Rothbaum, Larry F. Hodges, Rob Kooper, Dan Opdyke, James S. Williford, and Max North. Effectiveness of Computer Generated (Virtual Reality) Graded Exposure in the Treatment of Acrophobia. *the American Journal of Psychiatry*, 152(4):626–628, 1995.

[47] Jan Schneider, Dirk Börner, Peter van Rosmalen, and Marcus Specht. Presentation Trainer, your Public Speaking Multimodal Coach. *ICMI 2015: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pages 539–546, 2015.

[48] Joann Segovia. Personal Response System and Its Effect on Student Learning. 2008.

[49] Beat Signer and Moira C Norrie. PaperPoint: A Paper-Based Presentation and Interactive Paper Prototyping Tool. pages 15–17, 2007.

[50] Daniel J. Simons and Ronald A. Rensink. Change Blindness: Past, Present, and Future. *Trends in Cognitive Sciences*, 9(1):16–20, 2005.

[51] Ryan Spicer, Yu-Ru Lin, Aisling Kelliher, and Hari Sundaram. NextSlidePlease. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 8(4):1–20, 2012.

[52] Eva Strangert and Joakim Gustafson. What Makes a Good Speaker? Subject Ratings, Acoustic Measurements and Perceptual Evaluations. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, pages 1688–1691, 2008.

[53] M. Iftekhar Tanveer, Emy Lin, and Mohammed Ehsan Hoque. Rhema : A Real-Time In-Situ Intelligent Interface to Help People with Public Speaking. *IUI 2015: Proceedings of the 20th International Conference on Intelligent User Interfaces*, pages 286–295, 2015.

[54] Jaime Teevan, Daniel J. Liebling, Ann Paradiso, Carlos Garcia Jurado Suarez, Curtis von Veh, and Darren Gehring. Displaying Mobile Feedback During a Presentation. *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services - MobileHCI 2012*, page 379, 2012.

[55] Edward R Tufte. The Cognitive Style of PowerPoint. *Graphics Press, Cheshire, Connecticut.*, page 23, 2003.

[56] Max Wertheimer, Kurt Koffka, and Wolfgang Köhler. A Brief Introduction to Gestalt, Identifying Key Theories and Principles. *Psychol Forsch*, 4:301 − 350, 1923.

[57] Bang Wong. Points of View: Color Coding. *Nature Methods*, 7(8):573, 2010.

[58] Bang Wong. Points of View: Avoiding Color. *Nature Methods*, 8(7):525, 2011.

[59] Zhengyou Zhang. Microsoft Kinect Sensor and Its Effect. *IEEE Multimedia*, 19(2):4–10, 2012.