ARSV

Augmented Reality Speakers View

Augsburg University of Applied Sciences
Interaction Engineering
WS 19/20
Prof. Dr. Michael Kipp

Simon Geier
Jan Niklas Schlichting

Abstract

A commonly used feature of PowerPoint presentations is the speaker view, which provides an overview for notes and upcoming slides. While usually a presentation clicker is used to switch the slides, standing in front of the presentation computer can not be avoided when it comes to reading the notes while presenting. Therefore, a concept for transferring the speaker view onto an Augmented Reality Smart Glass (ARSG) is developed and evaluated.

Using the Augmented Reality Speaker View (ARSV), free standing in front of the audience can be achieved, enabling a hands-free presentation and slide control via head-driven gestures. This new interaction method shall simplify holding presentations, as the speaker can talk to the audience freely without the need of changing focus for reading the notes. It therefore is expected to help with a more fluid speech and less loss of focus onto the audience.

Within user tests, the focus and the slide control are evaluated in a comparison to the classical speaker view of PowerPoint. Furthermore, the readability and the positioning of the holographic information elements are analyzed both for a head-up display (HUD) layout and a spatial layout within ARSV.

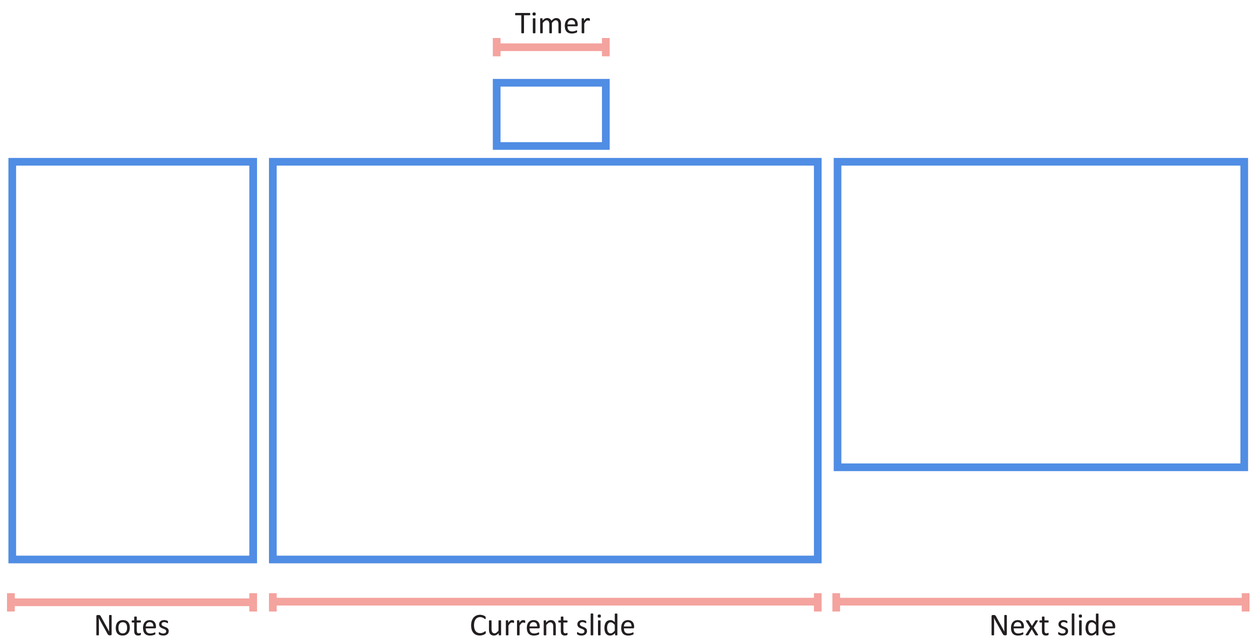
Table of Contents

Motivation

Holding presentations is usually required of many people, whether at school or at work. Often the speaker aims to perform a speech freely, without being bound to index cards containing notes or the speaker view of the presentation computer. To enable a free performance, the approach of this project is to use Augmented Reality in order to support speakers with their slides and notes.
To allow free gestural movement while presenting, the Augmented Reality functionality is provided by a head mounted device (HMD).
Technical assistance for speakers is a topic poorly explored up to now. The approach of this project using Augmented Reality HMDs is therefore a first advance into this field. Still, the interaction method itself has been researched. Hyrskykari, Istance and Vickers focused in their work “Gaze gestures or dwell-based interaction?” onto the limited accuracy of gaze interaction while working with an immersive device. Therefore, it has been evaluated how this accuracy can be outdone by combining it with dwell-based interaction and furthermore compared to gestural interactions using gaze.
For Head-Up displays (HUD) the automotive sector is currently the most popular field of application. In “Survey and Classification of Head-Up Display Presentation Principles” Tönnis and Plavšić focused on concepts for information presentation using HUDs and new capabilities of displaying contents three-dimensionally.
To compare interaction methods and information presentation exposed in those papers, the project Augmented Reality Speaker View (ARSV) is introduced, which uses the Microsoft HoloLens as current state of the art technology for Augmented Reality HMDs. The focus of this project is then laid onto the comparison between the usage of a HUD layout and a Spatial layout for presenting the required information to the speaker.
Furthermore the potential of the concept is evaluated. Firstly the question whether the concept can be pursued and improved and secondly how it compares with the most common form of presentation at the moment namely Power Point.

Concept

ARSV provides a novel way of presenting PowerPoint presentations, which allows the speaker to present freely without looking at a screen or using a keyboard or clicker to change slides. All information is displayed by the HoloLens in front of the user, allowing the speaker to focus onto the audience. Slide transitions are realized by gestural movements of the speaker's gaze and holographic elements, which are located at a fixed position of the room and can be controlled by dwell-based gaze interaction. The viewport includes the following elements within a fixed layout:

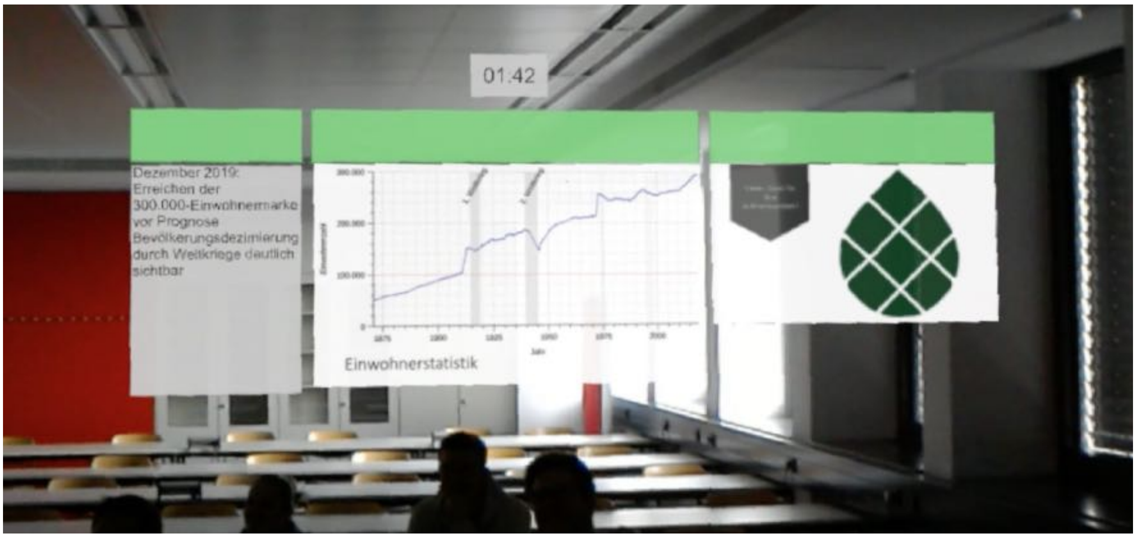


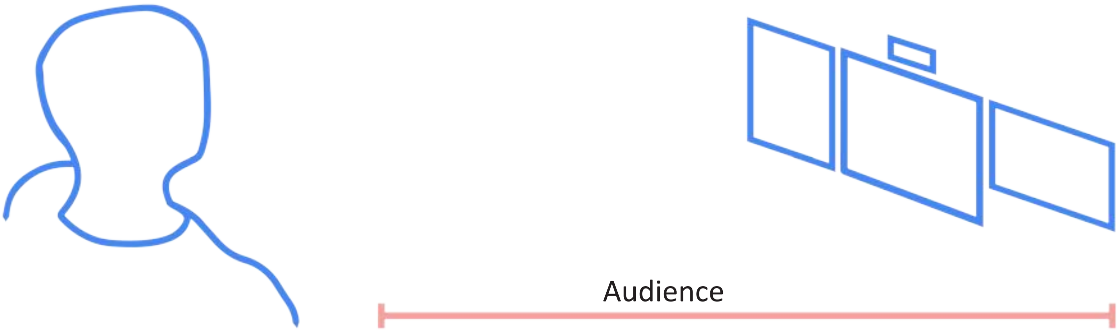
Static information elements:
- Presentation timer

Dynamic information elements:
- Mirror of the current slide
- Notes
- Preview of the upcoming slide

The speaker can control the system with head-driven gestures and dwell-based interaction. Therefore, it provides two different variants of layouts.

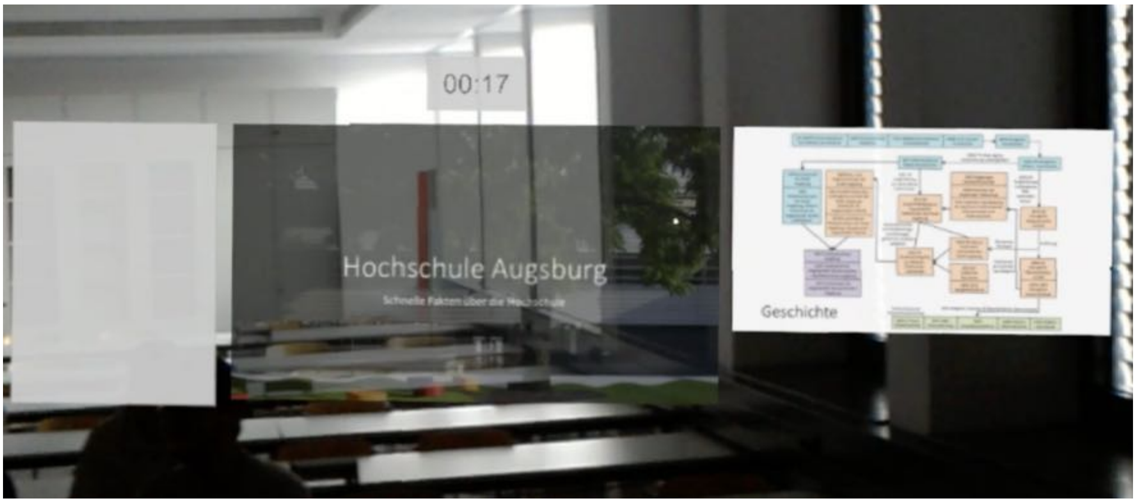
Variation 1: Spatial Layout

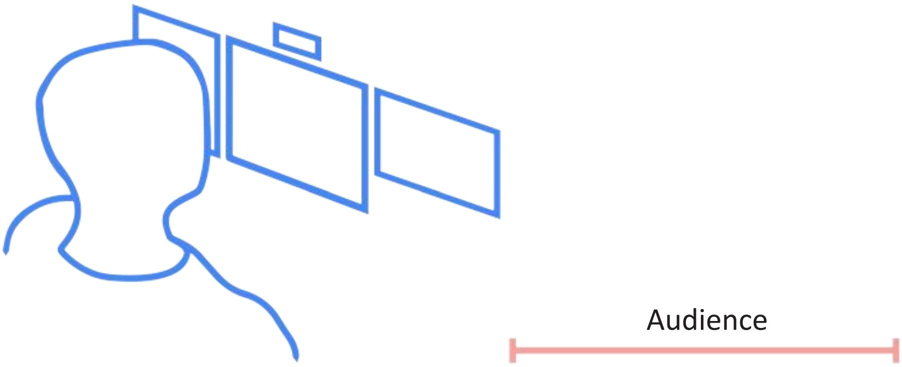




The elements are fixed in space and located in the depth of the room. Dynamic elements can be emphasized by the speaker, allowing the user to gain currently demanded information with enhanced readability.

Variation 2: HUD Layout





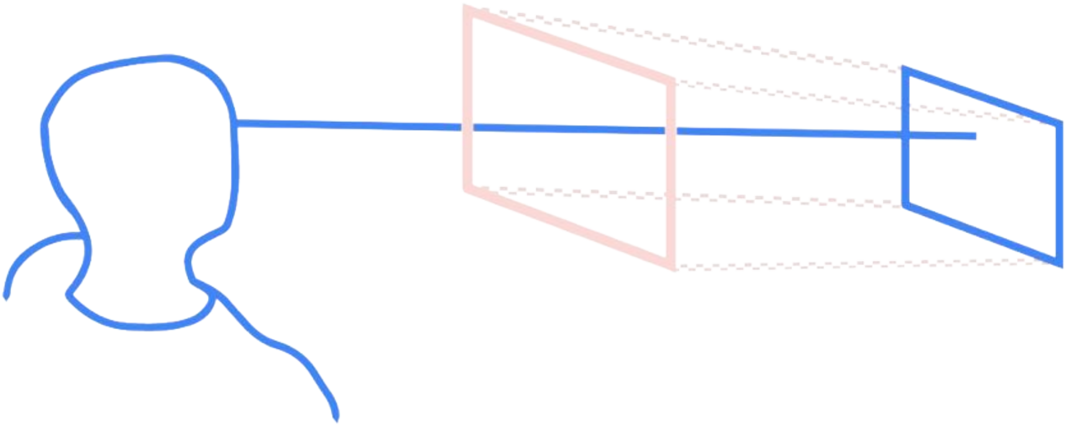
In comparison to the spatial interface, all elements are linked to the user. Therefore, the elements are moved alongside the sight of the speaker. Thus, elements cannot be emphasized.

Interaction techniques

Slide changing

The slide changing is triggered in the same way for both layouts by head-driven gestures. By tilting the head to the upper-right for the next or upper-left for the previous slide, the speaker can fire an event to the presenter's computer to perform slide changes on the computer, which then updates the slides of the HoloLens over the network.

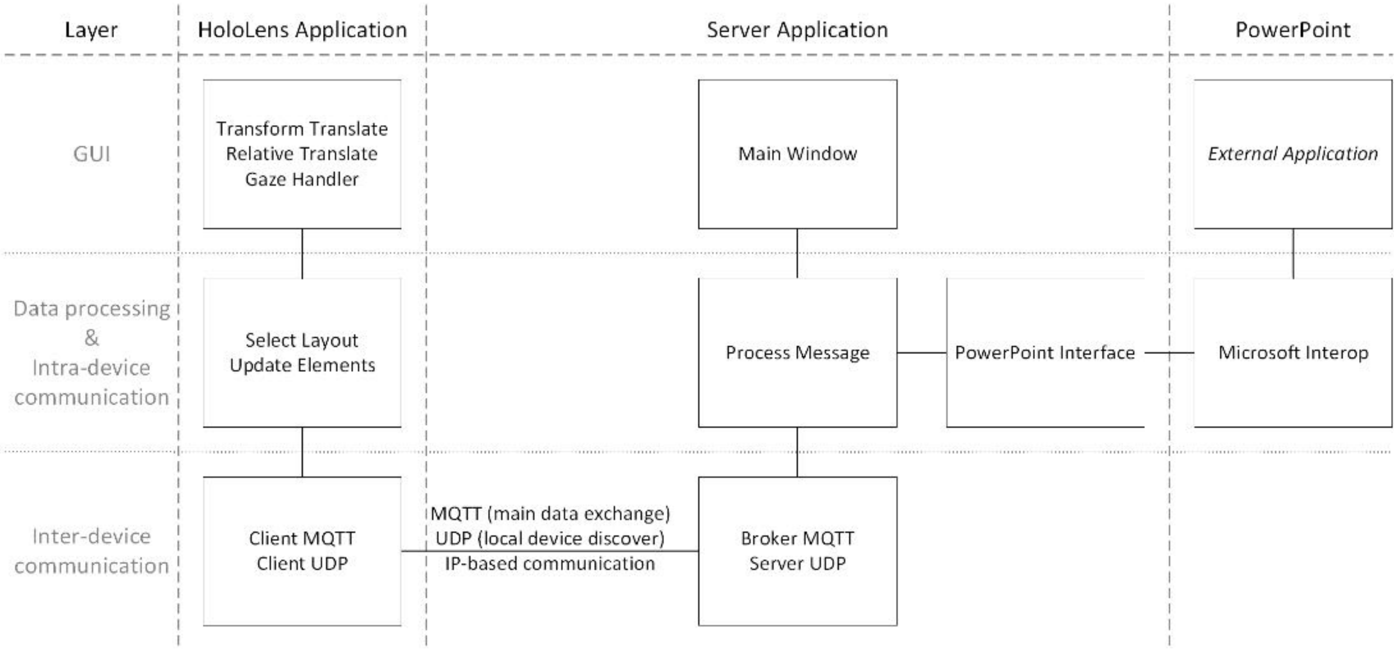
Emphasizing Elements



Using the spatial layout, the speaker can interact by dwell-based focus onto a dynamic element for a specific time interval. Interacting with the dynamic information elements will temporarily move them towards the speaker. If the speaker's focus is moved outside the emphasized element, it will reset its position to the default.

Implementation

Within this project, two different applications have been implemented. The application for the Microsoft HoloLens acts as main GUI and is simply called ARSV, as it provides the main entrance point for the user. The other application runs on a Windows 10 device as a server and is therefore called ServerARSV. It contains an interface to Microsoft PowerPoint for slide control and offers a simple GUI for log messages.
Both independently developed applications communicate with each other IP-based via UDP and MQTT. UDP broadcasts allow the Microsoft HoloLens to discover the server within a local network so that no hard IP declaration is required when using both applications. This allows an independent usage within any network. MQTT then serves as main data exchange protocol between Microsoft HoloLens and the server. An overview of implemented classes and scripts is shown in the following figure.



The development of the Microsoft HoloLens application has been implemented in Unity 2018.4 LTS with the Mixed Reality Toolkit (MRTK) 2.1.0. As programming language C# has been used with .NET Framework 4.6. As communication interface compatible with the Microsoft HoloLens M2Mqtt was integrated into the application.
For the server application the .NET Framework 4.8 has been used as well with C# as programming language. As interface between server application and PowerPoint the Microsoft Interop API has been used. As MQTT library for the broker MQTTnet was integrated into the server.
The following table shows the function of all scripts displayed in the previous figure.

Name	Device	Function
Main Window	Server	Handle control elements of server GUI
Process Message	Server	React onto queries from HoloLens and events from PowerPoint
PowerPoint Interface	Server	Perform slide changes, process information of presentation, export slides to images
Microsoft Interop	Server	Provide an API to start an instance of PowerPoint and control this instance externally
Broker MQTT	Server	Publish data to subscribed clients on various topics
Server UDP	Server	React onto discovering broadcasts with directional answer messages
Transform Translate	HoloLens	Handle transform translation of emphasize function
Relative Translate	HoloLens	Lock user relative slide changing rotation when looking above the horizontal line to trigger changes
Gaze Handler	HoloLens	Provide dwell-based triggering of interactive elements via gaze direction of head mounted device
Select Layout	HoloLens	Create Sprites from received image data, assign to current layout and provide layout selection methods
Update Elements	HoloLens	Update notes, current and next slide of layout used

For transmission between server and HoloLens a protocol using MQTT has been developed. In the following table single chars sent are displayed commonly, strings sent are wrapped into quotes, byte packages into angle brackets with data details in square brackets.

Message	Topic	Action
+	HoloLens	Query for next slide
-	HoloLens	Query for previous slide
=	HoloLens	Query for transmission of slide data
?	HoloLens	Query for slide index and notes synchronization
“00:00”	Server/Time	Active time transmission of PowerPoint presentation
“Log”	Server/Log	Erroneous logging message transmission
“Notes”	Server/Notes	Transmission of current slide’s notes
<[Index] [PNG]>	Server/Slides	Transmission of all slides, where the first byte marks the index of the slide and all following bytes the actual image data encoded in PNG
<[Index]>	Server/Current Server/Next	Transmission of the index of the current and the next slide of the active PowerPoint presentation

Results

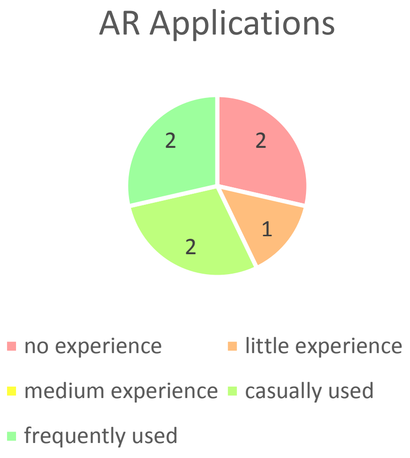
The developed prototype has been tested within a user study. Therefore, two different presentations have been presented with both layouts of the ARSV application as a within-subject user test. The order of the presentations did not change for each participant, but the starting layout has.
After testing, the presenter had to fill out a questionnaire regarding personal references, usage of the HoloLens itself, comparison between both layouts and comparison to a classical presentation with usage of the built-in PowerPoint speakers view.
Finally, open questions for favors and opportunities of ARSV have been inquired. As for the audience, for each presenter they also had to assume which layout was used for which presentation and rate the perceived focus of the presenter onto the crowd.
The both presentations were designed in a simple way, consisted of five slides each and contained information about the city of Augsburg and the University of Applied Sciences Augsburg, respectively.
The user test took place at room E4.02 of the University of Applied Sciences Augsburg on January 27, 2020 with seven participants aged between 20 to 30 years old. The order of the starting layouts was decided randomly as shown in the following table.

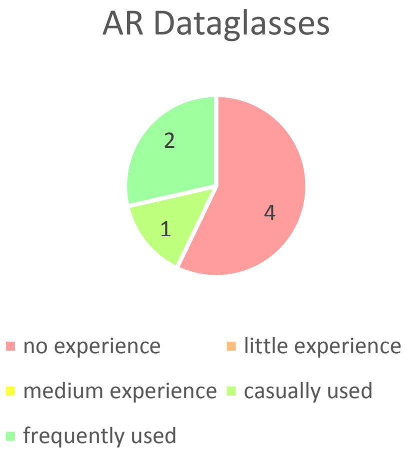
Speaker	Layout for presentation “Augsburg”	Layout for presentation “University”
1	Spatial	HUD
2	Spatial	HUD
3	HUD	Spatial
4	Spatial	HUD
5	HUD	Spatial
6	HUD	Spatial
7	Spatial	HUD

To create equal conditions, each presentation has been held by the test leader using each layout of ARSV once. For guidance purposes, the view of the test leader has been displayed alongside the current presentation.

The following questions have been asked to the speakers. Question I.1.1 is omitted as it only referred to the age of the user.

Personal preferences
Question I.1.2: Did you use Augmented Reality applications before?
Question I.1.3: Did you work with Augmented Reality data glasses (e.g. HoloLens) before?
The answer possibilities scaled from no experience to frequently used.

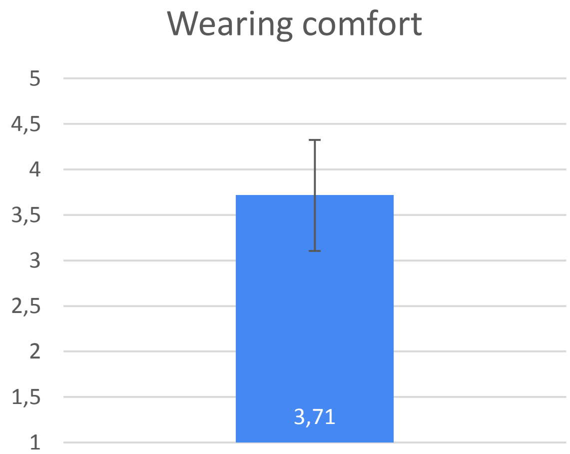


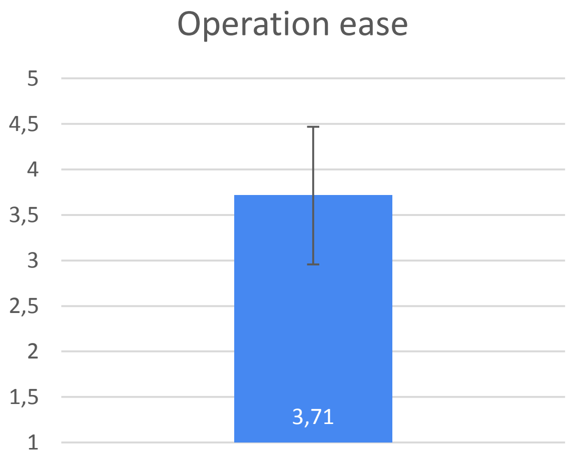


HoloLens usage

Question I.2.1: How was the wearing comfort?
Scaled from 1 – very uncomfortable to 5 – very comfortable

Question I.2.2: How easy was it for you to operate with the HoloLens?
Scaled from 1 – very hard to 5 – very easy





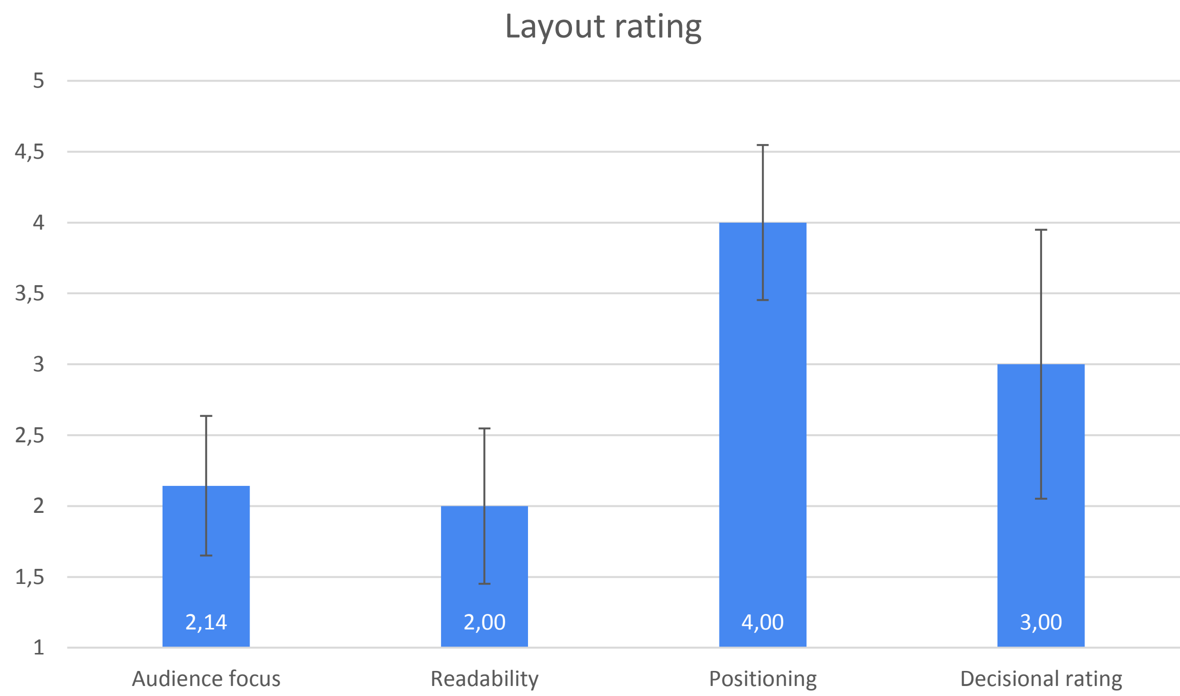
Rating for both layouts

Question I.3.1: How good was your overall focus onto the audience while presenting?
Scaled from 1 – no focus to 5 – very focused

Question I.3.2: How would you rate the overall readability of the elements?
Scaled from 1 – not readable to 5 – very legibly

Question I.3.3: How would you rate the overall positioning of the elements?
Scaled from 1 – very distracting to 5 – very organized

Question I.3.4: Which layout do you prefer more?
Scaled from 1 – HUD layout to 5 – Spatial layout

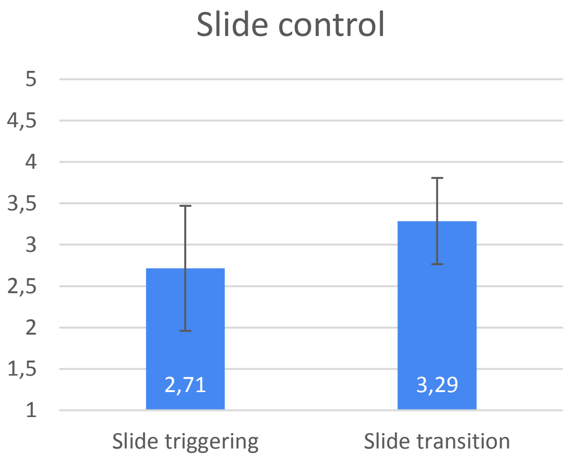


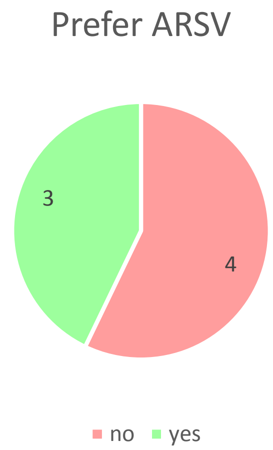
Comparison to classical PowerPoint

Question I.4.1: How useful would you rate the triggering of the slide changes on the HoloLens?
Scaled from 1 – not useful to 5 – very useful

Question I.4.2: How comfortable would you rate the slide transition on the HoloLens?
Scaled from 1 – not comfortable to 5 – very comfortable

Question I.4.3: Would you prefer the ARSV method itself over the classical PowerPoint?
Decide for yes or no





Open questions

Question I.5.1: What did you like of ARSV?
Open question

Summarization of similar answers:

- General:
  No need to turn away from audience (L1)
  Fast access to additional information (L2)
  Intuitive interaction (L3)
- Spatial:
  Slides are directed in direction of the audience (L4)
  Reading slides seems like focussing audience (L5)

Question I.5.2: What would you improve of ARSV?
Open question

Summarization of similar answers:

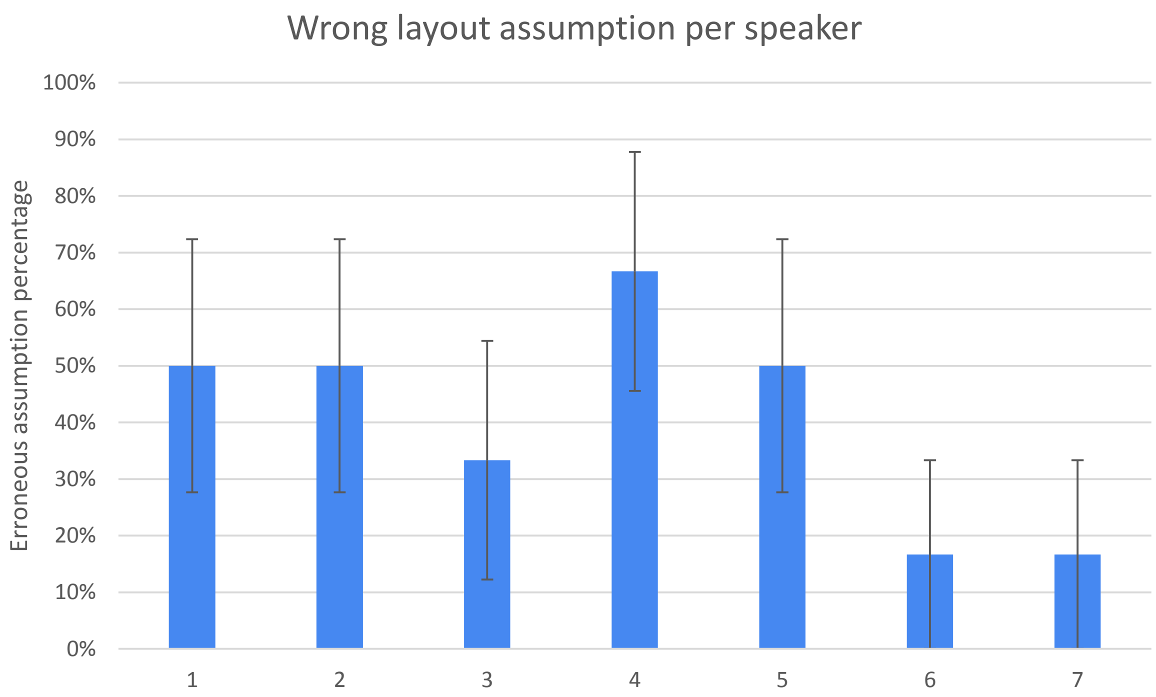
- General:
  Improve and customize gesture mechanism to change slides (I1)
  Reduce brightness for better focus onto audience (I2)
  Reduce actualization delays and duration of gaze timer (I3)
- Spatial:
  Elements are not visible while changing slides (I4)
- HUD:
  Improve readability of elements (I5)

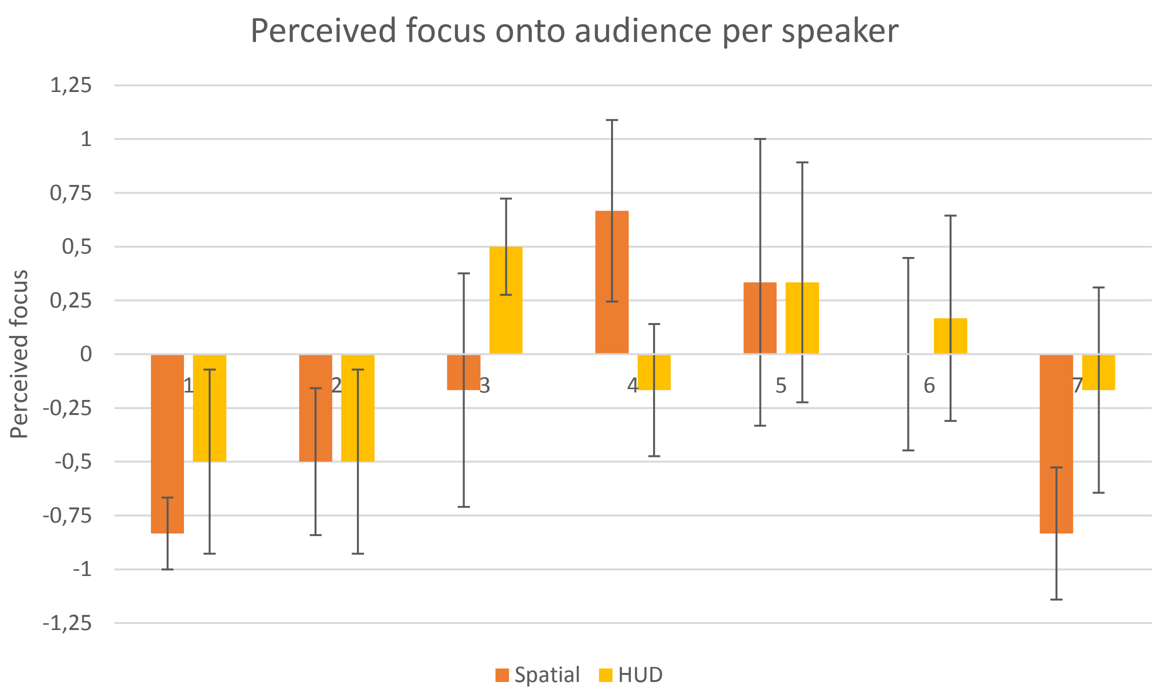
Audience questionnaire

Question II.1.x: What layout did the speaker use for which presentation?
Assign ARSV layout to presentation

Question II.2.x: Rate the focus of the speaker onto the audience
Scale from -2 – very unfocused over 0 – medium focused to 2 – very focused

The questions above have been asked independently for each speaker. As it was only known to the test personnel and the speaker what layout was currently used, thus an unbiased focus rating regarding the ARSV layout used could be retrieved from the audience. According to this information the error rate of wrong layout assignments for question II.1.x and the focus per layout for question II.2.x could be measured.





Evaluation

As obtained from the personal references, the previous experience of the participants with Augmented Reality applications and belonging data glasses was very mixed. Still, the overall wearing comfort of the HoloLens and its ease of operation was rated positively on average. This was quite unexpected, as we would classify the HoloLens ourselves as an unwieldy device.

For the rating of both layouts, the focus of the speaker onto the audience and the readability of provided elements was rated negatively. This was also against our expectations, as we assumed that the focus onto the audience would be improved by being able to move freely and look at the audience while presenting. Same goes for the readability, as the content of the slides is directly provided into the view of the speaker. We justify those ratings by the information obtained within the open questions. Therefore, in question I.5.2 it was summarized that the focus onto the audience was generally distracted by the brightness of the holographic elements (I2), the duration for the activation of the dwell-based gaze timer was too long (I3) and the readability of the HUD interface needs to be improved in principle (I5). As for the latter part, this was not possible for the very moment due to technical limitations of the field of view of the HoloLens.

On the other side, the positioning of the elements for both variants were rated overall positively. As for the HUD interface, the elements have been positioned to fill the maximal width of the screen. The Spatial layout has explicitly been mentioned positively in the open questions, as the direction of the slides into the direction of the audience was liked (L4). Still, there was no clear preference of the participants onto one of both provided layouts. The average of the decisional rating was exactly between both layouts, but it is worth to mention that there has been a relatively high standard error of the mean. This shows that the rating of the participants has been split onto their preferred layout, while it is then leveled out on average. We therefore tend to declare the Spatial layout as coequal to the HUD layout, whereas the latter one might be interpreted as the more common variant of holographic applications.

The comparative questions between ARSV and the classical variant of PowerPoint showed that the triggering of the slide changes was rated negatively. This was confirmed by the open questions, as one of the general statements referred to the improvement and customization of the related gesture (I1). However, as there has been a relatively high standard error of the mean onto this question, it is shown that the opinions of the participants diverged. The transition of the slide, meaning the actualization of the interface, was on the other hand rated overall positively. As the images are transferred bundled at the start of the presentation and the elements are only updated per selection of the slide index transmitted, this meets our expectations. Still, as the result has been only slightly positive, not everyone felt comfortable with this transition. This was also stated within the open questions, as it was criticized that the slides are not visible in the Spatial layout while performing the gesture for slide changing (I4). The analyzed rating might therefore be deteriorated by influence of the triggering of the slight changes.

Using a questionnaire also for the audience, it was aimed to find out if the spectator can distinguish the layout used by the speaker's behavior and if there has been a noticeable difference between the perceived focus of the speaker onto the spectator for each layout of ARSV. It was overall shown that the error rate of classifying the layout used by the speaker leveled out to above 40 percent, meaning that the layout used could not be classified by the spectator. The perceived focus was overall slightly negatively rated for both layouts, but the difference within each speaker was clearly slighter than the difference between the speakers. This was also proven by performing a t-test, which declared with a paired probability level of 48.9 percent the difference between both layouts to be statistically insignificant. The speaker's focus onto the audience therefore has not been influenced by the layout used.

Finally, the idea behind ARSV has slightly been rejected, as 4 out of 7 participants still prefer the classical variant of PowerPoint over ARSV. Nevertheless, this quote might be improved with future work on the prototype of ARSV with the help of the critical feedback by the participants.

Conclusion

In this project, an Augmented Reality Speaker View (ARSV) has been implemented as a prototype. This prototype contained two different layouts, a HUD layout, where elements were located linked to the user, and a Spatial layout, where elements were located at a fixed position of the room.

The conduction of a within-subject user test comparing those both layouts lead to the finding, that the idea behind ARSV was at least accepted by little less than half of the participants. Overall positively rated has been the positioning of the elements containing the slides and the notes, so that the speaker has always been able to look towards the audience while presenting.

Due to the technical state of the art, the readability and the personal focus onto the audience of the speaker has been criticized. This may be improved in future work on ARSV, as the new generation of the HoloLens, which is going to be accessible by public within the year of 2020, offers a wider field of view and accelerated computing power. Those properties could help to improve the size of the elements used within the HUD layout and the reliability of the dwell-based triggering to emphasize the Spatial layout.

Also dependent on the technical state of the art are the results of the slide control. As criticized by many participants, the gestural movement to change slides should be reconsidered and improved in general. It was furthermore claimed that changes of the slides could not be recognized within the Spatial layout. To provide remedy, the interaction method of head-driven gestures could also be improved with help of HoloLens 2, as the new generation in fact offers gaze tracking. Slide control therefore could be realized by looking to the upper left or right by eye movement without being obliged to move the head. This would make it easier to perform the slide control, but it is yet unclear if it comes along with a reduced number of false positives.

Furthermore, we assume that with upcoming generations the required obscurity of Augmented Reality data glasses will be reduced. This might increase the perceived focus of the speaker onto the audience, as the audience can determine in which direction the speaker is currently looking at.

Finally, it was mentioned in open dialogue with participants to merge both layouts together. The notes would remain as part of the spatial layout, but current and next slide would be adopted from the HUD layout and adjusted in position, so that the center point of the application remains empty to look at the audience. This would fix the downside of constantly overlaying the main focal point with information and combine the ability to perceive slide transitions while performing the head-turning control gesture with the emphasizing of the hard-to-read notes element.