Creation of a HoloLens Application To Support Instructors During Pilot Simulation Training

Jeanine Vlasblom

Jeanine Vlasblom (3900053)
j.i.d.vlasblom@uu.nl / jeanine.vlasblom@nlr.nl
master Toegepaste Cognitieve Psychologie
Thesis (27,5 ECTs)

Nederlands Lucht- en Ruimtevaart centrum (NLR)
Jelke van der Pal (mentor)
Jelke.van.der.pal@nlr.nl

Jelle van Dijk (first assessor)
j.a.vandijk@uu.nl

Jeroen Benjamins (second assessor)
j.s.benjamins@uu.nl

With special thanks to:

Bram Helleman, Jaap Groeneweg, Jeroen van Rooij and Tim van Breda.

Abstract

This study investigates the product development of a HoloLens-application which enables instructors to see pilots’ gaze behaviour in real-time. Gaze behaviour is an important aspect of flying and can be difficult to learn, especially for pilots in training for their first type rating when performing an approach and landing simulation. However, the current training instruments lack objective information about the pilots’ gaze behaviour. Instructors can base their feedback only on the pilots’ movements, the results in the simulation, and by asking about flight parameters to estimate their gaze behaviour.

Product development followed several iterations, in which instructors of a major Dutch airline were involved. The results of a short online questionnaire (N = 20) confirmed the claims in literature that instructors desire a tool to see the pilots’ gazes in real-time, with an optional playback during the debriefing. The pilot experiment (N = 5) yielded similar results and additional advice to enhance the tool. Answers on a common, validated questionnaire indicated a positive user experience, with ‘above average’ to ‘good’ evaluations of the application. Results from the newly constructed questionnaire regarding the experienced support for feedback, indicated that the instructors felt supported by the tool. All five instructors would implement a similar tool in their training and recommend it to others. The pilots did not have any problems with the application.

Many training situations may benefit from this application. The tool needs further development and should be tested on a larger scale to prove its value.

Table of contents

Introduction

Approach and landing

Gaze behaviour is an important aspect of flying and can be difficult to learn, especially for pilots in training for their first type rating like a Boeing 737 or Airbus A320 when performing an approach and landing simulation (van Rooij, J. M., personal communication, January 26, 2018). During approach it is important that the pilot gradually shifts their gaze from the Primary Flight Display (PFD, located on the cockpit display) to the outside windows. This instructor's remark (van Rooij, J. M., 2018) is supported by several studies on pilots' gaze behaviour during approach and landing. Pilots look more through the outside windows than at their primary displays during landing phase (Brown, Bautsch, & Wetzel, 2002; Goodman, Hooey, Foyle, & Wilson, 2003; Lefrancois, Matton, Gourinat, Peysakhovich, & Causse, 2016). Professional airline pilots make this shift gradually (Anders, 2001). However, most pilots in training have trouble with this gradual shift: their shift is often too direct or fast, or they keep looking at their PFD (Dubois, Blättler, Camachon, & Hurter, 2015).

Simulation training

Instructors know that this is a difficult part of the training but cannot see whether the pilots' gaze behaviour is correct or incorrect in the simulators. It is impossible to see where their students are looking at, as opposed to car driving instructors who can base their feedback on the head movements of their students. Both driving school and flight school instructors would like to see where their students are looking at specific moments during training. During the simulation training in flight school, the instructor will be seated behind the student; they are both looking at the displays of the cockpit and the outside windows, as shown in Figure 1. In this set-up, instructors can base their feedback only on the pilot's movements (e.g. pushing a button) and the results in the simulation (e.g. change in altitude). This means that they cannot entirely be sure if these results are based on 1) whether the pilot did or did not see the situation, 2) whether the pilot noticed the problem, and / or 3) whether the pilot knows the appropriate action (Carroll et al., 2013).

Figure 1. Typical set-up simulation training with instructor seated behind students.

Well-designed (corrective) feedback should inform the learner that there was an error and why there was an error, but without saying what the correct action is (Balzer & Doherty, 1989; Merriënboer & Kirschner, 2012). Instructors notice errors by checking the simulation: they can look at the values on the PFD to know if the (altitude) levels are correct according to the situation. However, if the error is caused by incorrect scan behaviour, the instructor is not able to obtain this information objectively. Asking students where they have looked is subjective and incomplete information, and therefore not a solid foundation for well-designed feedback. To put it in other words: if the cause of the error could be displayed objectively, the chance of well-designed feedback should increase.

Eye tracking

Modern technology could eliminate this lack of information. Eye tracking – a tool that measures the location of the fovea and uses an algorithm and screen information to show where the eyes are fixating (Holmqvist et al., 2011) – gives the opportunity to see where people are looking at a specific moment in time. In theory, eye tracking could be used to track the eyes of pilots, which could give the instructors more information as input for the debriefing session; the feedback given after the simulation flight. Instructors have consequently indicated that eye tracking could be a useful tool to improve training and/or flight safety (Henneman et al., 2014; O’Meara et al., 2015; van Oijen & Voskamp, 2016; Wetzel, Anderson, & Barelka, 1998).

Despite the desire to see the gaze behaviour of pilots in training, it is not yet common to implement this in training. Only few airlines use eye tracking to add value to their flight training and safer flight operations (Emirates, 2017; Lufthansa, 2016). However, in educational research, the use of eye tracking is starting to become increasingly popular.

Although most research has been done on the difference in eye movement behaviour between novices and experts (Ashraf et al., 2018), eye tracking has also been studied as a tool to support assessment feedback.

One of the first researchers to test this was the group of Wetzel, Anderson, and Barelka (1998). During a flight simulation, the eye-fixations of the pilots in training were tracked and could be displayed to the instructor operator station during the debriefing session. After seven months of usage, the instructors evaluated the system as an improvement of the training. The system was expected to be especially useful for the initial flight training program, so that bad habits can be detected and corrected in an early stage. A similar experiment showed similar results, even without the accuracy of an eye tracker. Waag, Wetzel, and Anderson (1993, as cited in Wetzel et al., 1998) used video recording to reveal the head and eye movements of students in an airplane simulation. Both interventions were quite intrusive and did – according to the participants and researchers - not have optimal functionality because they lacked real-time data, efficient event-related play-back options, and a clear visualisation of the eye movements.

O'Meara et al. (2015) used a less intrusive eye tracking method to support the debriefing session after a nursing simulation. Every participant performed an eight-minute simulation three times. The performance of the medical students was significantly better in the third simulation compared to the first. Since there was no control group, it remains unclear if the eye tracking had additional value to their learning curve compared to receiving feedback without the eye movement recordings. Despite this lack of information, nearly all students thought this intervention aided the feedback session (90%), improved their learning (87%) and benefitted from it (97%).

Another eye tracking study on medical students while performing a training simulation revealed similar results (Henneman et al., 2014). After the simulation, the students received either a debriefing from their teacher, a video with their eye movements, or both. All students performed better on a similar simulation after this feedback, but the group who received only their eye tracking records performed the best. These results should be interpreted with care because the debriefing was received immediately after the simulation, and the eye tracking data arrived four days after the simulation. A possible explanation for the conclusion that only eye tracking would be more beneficial than receiving both assessment feedback options could be that the 'both-condition' never checked the video.

Recently, new assessment tools to support flight training have been developed. Carroll et al. (2013) developed a tool to measure brain activity (EEG) and eye fixations during a flight simulation, to be played back by the instructor during debriefing. It was also possible to check the simulator control inputs of the student and the instructor’s notes made during the simulation. This is a promising tool that presumably allows instructors to accurately and efficiently diagnose a performance issue. However, no research has been performed yet to test its efficacy.

Similar research has been carried out including usability testing. The European A-PiMod project (van Oijen & Voskamp, 2016) developed many cockpit interventions, including a training tool that presents real-time information on the cognitive capability of flight crew to the instructors during training. When evaluating the tool, it appeared that the tool was appreciated for providing the instructors with objective data, which could support the debriefing sessions with the pilots. However, the intervention was a distraction from observing live interactions between pilots; the data visualisation was far from intuitive, and instructors’ notes could not be traced back because of a lack of event markers.

Most of these methods seemed promising to both the participants and the researchers, but were not yet functional because of the intrusive eye tracking methods (Wetzel et al., 1998), the unintuitive visualisation (van Oijen & Voskamp, 2016; Wetzel et al., 1998) or because the tool or experiment simply was not elaborate enough (Carroll et al., 2013; Henneman et al., 2014; O’Meara et al., 2015). None of these studies have used the eye tracking data to show the instructors in real-time where their students are looking.

Current study

Considering the positive expectations as well as the known drawbacks, this study will explore if pilot simulator training can be supported by showing the instructor where his students are looking during the simulation. The main research question is:

Do instructors feel supported in their feedback by the ability to see where their students are fixating during a simulation?

It is expected that the ability to see where the pilots in training are looking in real-time will give instructors a feeling of support for the debriefing session, since many instructors have expressed their positive expectations about this intervention.

Before an empirical experiment can be carried out, a tool must be made that enables the instructor to see the gaze behaviour of the pilots. The conceptualisation (design) and creation (development) of this new tool will be discussed in the following chapter. The first prototype will be tested in a pilot experiment to get insight into the possible advantages of the tool. The results of this pilot experiment can serve as input to optimize the tool.

Design & Development

Designing usable interactive products requires considering who is going to be using them, how they are going to be used and where they are going to be used (Rogers, Sharp, & Preece, 2009). Although the user might not know exactly what they need (Rogers et al., 2009), they do know a lot about how and where the product is going to be used. In attempt to involve the user in the design process, the process followed several iteration cycles. Each iteration cycle consisted of an analysis-, design- and test-phase, in which a subject matter expert (SME) was involved in every test-phase and in some analyses.

Iteration 1

To analyse the possibilities and constraints for the product, experts were heard and literature on product development, eye tracking and the HoloLens was used.

In a first brainstorm session, SME J. M. van Rooij (2018) highlighted the need for a feedback tool with which instructors can see where their students are fixating in the simulator. This would be particularly useful in training for people who know how to fly a small airplane but have not yet flown a specific civil airplane type. Cockpits differ per airplane and therefore require different procedures from the pilots. Pilots in training who start to fly a type have difficulty with the gradual gaze shift during approach and landing. Based on this information, a choice of simulation procedure, simulator type and target group has been made. Regarding the design of a usable interactive product (Rogers et al., 2009), the following consideration was made: instructors (who) will use a tool (how) to see where the pilots are looking during a simulation training (where).

To answer how this tool will be used, several functionalities have to be ensured. Previous studies indicate that the applications should not be too intrusive and that the visualisation should be intuitive and displayed in real-time (O’Meara et al., 2015; Rogers et al., 2009; van Merriënboer & Kirschner, 2012). During the brainstorm session, several ideas arose including the use of an eye tracker and the HoloLens. The analysis was extended with literature on these subjects and on visualisation techniques.

According to Holmqvist et al. (2011) the least invasive eye trackers are the remote eye trackers, which are completely detached from the body. In this way, pilots keep their freedom to move their heads while flying. Most eye tracking brands have remote eye trackers and most of those can forward the data in real-time. However, this depends on the software that comes with the eye tracker. To choose the best eye tracker for this situation, an eye tracking expert was included in the decision-making process.

Displaying the data intuitively to the instructor is a more extensive subject to explore and requires extra insight of experts and literature. A second brainstorm session with several experts was organized to construct requirements for the user experience and the conceptual design of the application. The following requirements came forward:

- The system should show where the pilots are looking.
- The system should present these data in real-time, and these should be available during the debriefing session.
- The data should be visualised in different ways, to test which visualisation is the best. The choice of colour has to be considered and the option to plot the raw data (real-time fixation location) or relative data (average fixation loation over time or relative to a set benchmark). Both options are manifested in typical eye tracking visualisations that come with the software, e.g. a heatmap or scanpath. The instructor should be able to see the preferred visualisation on demand to facilitate his sense of ownership; although this must be limited to some extent. Having infinite options diminishes the usability of a product (Rogers et al., 2009) and the likability of the option (Iyengar & Lepper, 2000).
- Both a 2D-screen and a HoloLens could be used to show the data to the instructor. Displaying the information on a screen will probably be easy to implement since the training set-up already uses an instructor screen. The HoloLens could be of extra value by showing the data in the instructor’s Field Of View (FOV) when looking at the pilot, while showing the data on a screen requires a change of gaze for the instructor. It is therefore expected that the HoloLens will result in a more intuitive visualisation of the data than a tablet.
- The eye tracker should have the accuracy to distinguish the localizer from the glide slope and the speed indicator, which are all located on the Primary Flight Display.

These requirements form the conceptual design of the product. Details like colours, icon design and the timeframe of dataflow are evaluated in the physical design, elaborated under “Visualisations”.

HoloLens

The idea to use the HoloLens came from the expected non-invasiveness of the application. According to Rogers et al. (2009, p.281), the possibility of having instant information before one's very own eyes that is contextually relevant to an ongoing activity and that can be viewed without having to pull out a physical device is very appealing. By using the HoloLens, it is possible to display the locations of the eye fixations within the FOV of the instructor. While both the instructor and pilot in training are looking at the simulator cockpit, the pilot will see nothing but the simulator and the simulation. The instructor, however, will also see a holographic projection of where the pilot is looking at in real-time. This technique is called mixed reality (MR) or augmented reality (AR), in which the real world is expanded with extra, digital information. An AR system runs interactively in real-time and supplements our senses with virtual (computer-generated) objects that appear to coexist in the same space as the real world (Azuma et al., 2001). In this case, virtual representations are superimposed on physical devices and objects (Rogers et al., 2009).

Figure 2. The HoloLens and its additional features.

The HoloLens is one of the tools to create this augmented or mixed reality with and to interact with holograms in the world around you (Microsoft, 2018). Most demonstrations and advertisements show educational purposes. This is consistent with the rise of a new learning method: extending learners' interaction with and perception of their current environment using electronic devices to create a mixed reality (Sheehy, Ferguson, & Clough, 2014). AR is especially effective in activities where students can see things that could not otherwise be seen in the real world or without a specialized device (Bacca, Baldiris, Fabregat, & Graf, 2014). This finding could be externalised to the expectation that the additional eye tracking information will support the instructor in its feedback. Very few studies have reported a personalized process (meeting the needs of the student, e.g. by scaffolding) while using AR in education (Bacca et al., 2014). This study could be one of the first personalized AR applications in education, by using objective AR-data regarding the real-time behaviour of the student.

When using Mixed Reality, the conceptual design should – according to Rogers, et al. (2009) – be expanded to answer when the digital augmentation should take place (in real-time and afterwards), where it should appear in the physical environment (displayed over the cockpit) and what form it should take. The last one will be further explored and tested in this study, starting in the chapter “Visualisations”.

The intuitiveness of the HoloLens has not been investigated yet but can be expected based on: several statements in AR-literature that the HoloLens is an intuitive application, the 3D presentation of 3D objects, the freedom of the user to move naturally and the ways to interact with the application. The intuitiveness, overall user experience and experienced support of the HoloLens will be tested in this study.

Based on the fact that the information will be displayed in the instructor’s FOV when looking at the cockpit, the enthusiasm of the instructors (van Rooij, J. M., 2018), the positive experiences of using AR in education (Bacca et al., 2014; Sheehy et al., 2014) and the intuitiveness of the HoloLens, it is expected that the HoloLens application will be evaluated positively.

Visualisations

Intuitiveness

Working with an intuitive or easy-to-learn tool can enhance the usability of the product. An intuitive design should not conflict with the user’s cognitive processes involved in achieving the task (Rogers et al., 2009) to minimalize the cognitive load of the instructor and to reach optimal attention of the instructor (Proctor & Van Zandt, 2008). To present the eye tracking data in an intuitive manner to the instructor, the visualisation of the data must be thought through carefully. The information needs to stand out, but not distract the person from their ongoing activity in the physical world (Rogers et al., 2009). Information that is superimposed on the physical world (in this case digital information overlaying the real world) needs to be simple and easy to align with the real-world objects (Rogers et al., 2009). The ease of alignment depends on the accuracy of the eye tracker and HoloLens and is part of the development process. The simplicity of the information display is a subjective experience of the user and will therefore be measured with a questionnaire after using the tool.

Raw versus relative

There are several options to visualize the data. In this situation, raw data represents the fixation locations of the pilot in coordinates on the attended screen. Relative data can relate to an average fixation location over a chosen period of time, an average amount of time spent looking to a specific Area Of Interest (AOI) or the difference between the raw data and a benchmark. During this first iteration cycle, the preference of seeing the raw fixation locations came forward. In training situations, it is less important to see the scan behaviour compared to a standard, because the instructors have no difficulty assessing the pure behaviour: they know which behaviour is best for which situation. Since instructors know where the students should be looking at these specific moments during the simulation, there is no need to reconstruct the eye movements to norm-based values. In addition, raw data would cost the instructor less cognitive load than interpreting relative data or graphs. Analysing the eye tracking data and plotting it in graphs or comparing it to a standard will probably give the instructors more cognitive tasks to solve while looking at this information (Proctor & Van Zandt, 2008). Therefore, raw data (dots) will be displayed on top of the simulation screen, displaying the exact fixation locations of the students.

However, since the gradual shift from the PFD on the inside displays to the outside windows is important during approach and landing, it would be interesting to test whether relative information about this factor supports the instructors in their feedback. Therefore, each couple of seconds, the percentage gazed towards the windows versus gazed towards the PFD will be plotted (Figure 3). This will not be in the complex form of a graph, but contains only the relevant data, including normative data.

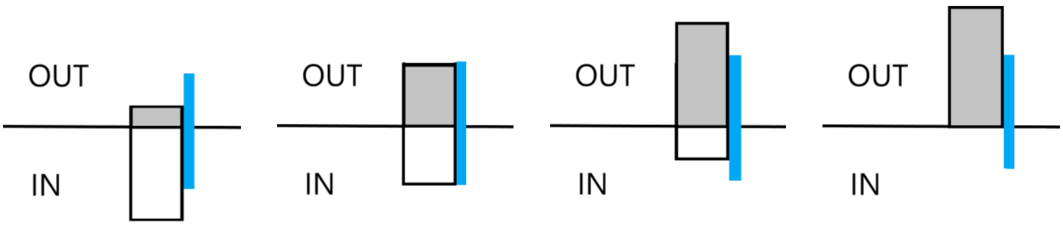


Figure 3. Dynamic graph showing the percentages of time looking at the outside windows or inside displays, compared to a benchmark in blue.

Real-time data

Showing the data in real-time has benefits for both the debriefing session and the feedback given during the simulation. It gives the instructors more information to take with them before they start the debriefing session. When eye movements need to be played back during the debriefing, these sessions will take up too much time to complete with every student (Wetzel et al., 1998). The real-time data can also be used to give feedback during the simulation flight, since feedback is most effective when it is given directly after the specific situation and behaviour of matter (O'Meara et al., 2015; van Merriënboer & Kirschner, 2012). However, while unexperienced pilots could benefit from receiving feedback during the simulation training, more experienced pilots prefer to receive the feedback during the debriefing session (SME, 2018).

Dynamic versus static

When raw data is shown in real-time, this will automatically be dynamic. Relative data can be shown statically or (semi-)dynamically. To establish optimal interpretation in the instructor, dynamic representation is especially important in this situation, since the fixation locations in the cockpit are close to each other (van Wermeskerken, Litchfield, & van Gog, 2018).

Relative data creates the opportunity for several options. The choice of time-frame, frequency and latency affects the dynamic presentation. A cumulative heatmap will finally result in a screen that is coloured red completely, whereas a real-time heatmap shows the mainly focused fixations over a fixed couple of seconds. With the current equipment, real-time information can be as fast as 50 data-points per second. This frequency will make it impossible for the instructor to see all the saccades and other scan behaviour.

Therefore, the raw data will be presented dynamically, in real-time, but with a slight delay. The relative data as seen in Figure 3 will be shown semi-dynamically: in time-frames of a few seconds. Examples of different dynamic visualisations can be found via the web links in Appendix A.

Multi-coloured versus mono-coloured

The use of colours is an important aspect of product design. None of the colours used should blend in with the background, which in this case is the simulator. The colours from the simulator used in this study (APERO, more information can be found in the method section of the pilot experiment) are mainly black, and some obtrusive colours on the screen are blue, light grey and brown. All colours should be intuitive to interpret and, depending on the situation, should therefore (not) interfere with basic instincts such as "red is bad" and "green is good". Expressing valence with colour-coding can be useful when direct action is preferred, e.g. after a warning, but can be confusing when the data is more neutral.

To plot the raw data, it is best to avoid the instinctively classified colours. The raw data is not compared to a benchmark, and its interpretation is supposed to be left completely to the instructor. Therefore, a neutral colour should be used to display the raw data.

The colours to represent the relative data are allowed to evoke an instinctive reaction because these data need to be categorized. As the pilots in training are no professionals, they will probably make a lot of mistakes. Therefore, it will probably be too chaotic to work with red and green for when students are performing perfectly compared to imperfectly. One should rather work with the current values displayed in a neutral colour, and the norm-reference displayed in grey (see Figure 3).

Typical display of eye tracking data

Eye tracking data is often displayed as dots or crosshairs on the same location on the screen or picture as the participant was looking at. Depending on the frequency of displaying the data, the dots will move around very quickly or stay at their place for about half a second. The dot could increase in size in line with the amount of time people have fixated on that location. When lower frequencies of display are used, fixations can be distinguished more easily from saccades, of which the latter could be displayed as lines between the fixations (Holmqvist et al., 2011). This visualisation is called the scanpath (Figure 4).

Figure 4. Scanpath visualisation.

Often, the data are displayed as a heatmap: colours indicate the locations where people have been watching (Figure 5). Usually the colour red indicates the most-fixated regions, with a colour-spectrum down to blue for the least-fixated regions, and no colour for regions that have not been fixated on (Holmqvist et al., 2011).

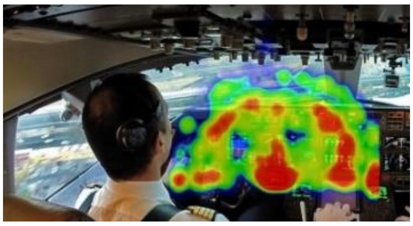


Figure 5. Heatmap visualisation.

A reversed heatmap is a completely dark image with a flashlight-like view over the locations where people have looked. A similar view could be created with a clear, sharp focus on the regions where people have fixated, with a blurry view for the unfixated areas (Holmqvist et al., 2011). This is called foveation (Figure 6).

Figure 6. Foveation.

It is possible to divide the complete FOV of the eye tracker into Areas Of Interest (AOIs). This is being used often in eye tracking research, because it makes the analysis a lot easier. AOIs can also be used to show relative information in graphs or other figures about the scan behaviour (Holmqvist et al., 2011).

Figure 7. Areas Of Interest (AOIs).

An atypical presentation of the data is one that arose from the first brainstorm session and is only possible to display in the HoloLens: the instructor sees the pilot’s gaze virtually displayed as a coloured line from the eyes to the point of fixation (Figure 8).

Figure 8. Eye gaze as lines through HoloLens.

Design and test

Based on the information from literature and the expectations of the (subject matter) experts, several options were conceptualized (Figure 3-8, see Appendix A for a complete overview). These were shown on paper to the subject matter expert involved in this study. A low-fidelity prototype like this is a good option to use in the first iterations (Rogers et al., 2009).

The options for the HoloLens and the tablet were explained. Next to these examples, dynamic videos (web links in Appendix A) demonstrating a few options were shown to the SME. This was mainly to familiarise him with the concept of presenting gaze behaviour, but also to show the differences when using certain delays and visualisation frequencies. Seeing the elaborated options and dynamic videos made it clear to the instructor that some would be dysfunctional, and others would be worth developing further.

Conclusion iteration 1

Based on product design literature and the assumptions and preferences of the user, it is concluded that the data should be presented to the instructor in real-time, raw (and possibly relative to a benchmark), dynamic, and delayed. All typical eye tracking presentations will be taken to iteration 2. Black, light grey and PFD-blue colours should not be used to avoid visually disappearing of data. A neutral colour should be used to display the raw and relative data. The reference norm can be displayed in grey. The HoloLens is expected to be appropriate for these requirements.

Iteration 2

The conclusions from the first iteration serve as an input for the analysis of the second design cycle. The concepts including the AOIs and the scan path were expanded to include in new 2D concepts (Figure 9-10).

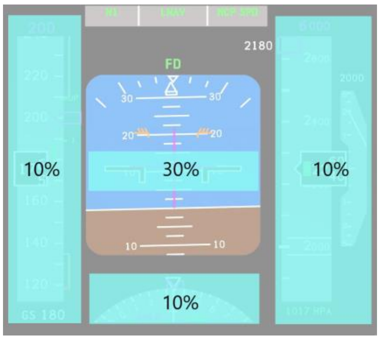


Figure 9. AOIs in PFD.

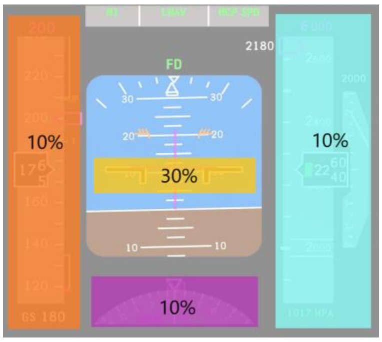


Figure 10. AOIs in PFD in colour.

Next to that, another possible concept emerged from the first iteration: Displaying the gaze behaviour data in a relative manner could be shown in a more intuitive way than displaying percentages on AOIs, by showing the percentage of time looking at the AOI as visual volume instead of in numbers. This resulted in graph-like concept with two bars representing the time spent looking through the outside windows and at the inside displays. A reference point indicates the optimal ratio. The concept in Figure 3 was created during the second iteration but was displayed in this paper earlier to improve understanding of the concepts for the reader. Figure 11 shows the second concept of this visualisation.

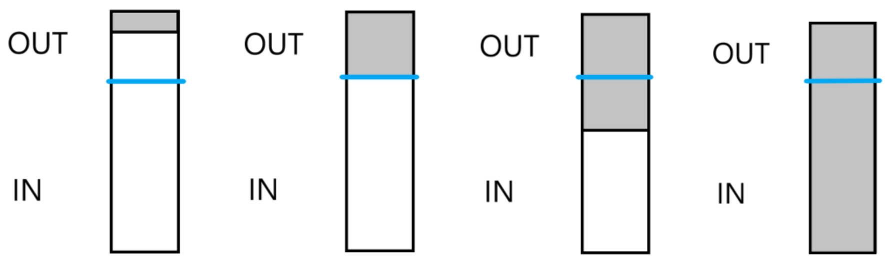


Figure 11. Dynamic graph showing the percentages of time looking at the outside windows or inside displays, compared to a benchmark in blue.

Showing the designs to the SME clarified which concepts were worth developing further and which were not. The concept displaying the AOIs (Figure 9-10) was evaluated positively, with a preference for displaying the information in one neutral colour. According to the SME, it is impossible to set a benchmark for correct gaze behaviour, since this is highly dependent on the situation. Even when each pilot flies the same simulation, their actions will result in such a varied range of situations that it is impossible to calculate the correct behaviour in advance. On top of that, instructors know the desired behaviour in every situation, which makes comparing raw data to a benchmark irrelevant.

In conclusion, the relative information should be displayed in one neutral colour. Showing the percentages in numbers or in volume are both promising according to the SME and are therefore taken to the third iteration. Since it is impossible to set a fixed benchmark for the perfect scan behaviour, this option has been dismissed.

Iteration 3

With the input from the first two iterations and the literature study, different concepts have been chosen to create for the HoloLens. Several experts were included to construct the application. During development, the two biggest subjects to discuss were data-flow and showing the data in the HoloLens. For optimal data-flow it is important to have the most efficient data-route and synchronization between the data acquisition and the data display. A connection between an eye tracker and the HoloLens was not yet available and had to be constructed. During the creation, several problems arose, of which some could not be fixed. These issues and their solutions are elaborated on in Appendix B.

The created concepts were demonstrated to the same SME as in iteration 1 and 2. The first demonstration showed the eye fixations of the pilot flying on the specific simulator screen that includes the PFD and Navigation Display (ND). Both pilot and instructor could see the eye fixations of the pilot in real-time. According to the SME, this concept would be good enough to test whether instructors benefit from seeing where their student is looking. However, this visualisation comes with an important disadvantage: if the pilot receives instrumental feedback about his gaze behaviour, he will be aware of this and therefore could react (in this case fly) differently than he would do without being aware of his eye movements. Since this feedback is not provided during real flight situations, it is undesirable to implement this option in training situations.

The second demonstration was not interactive but based on a HoloLens recording and viewed on a 2D screen. In the movie one can see the complete situation: a pilot flying a simulation, his gaze virtually displayed over the simulator screen including the PFD and the ND, and a graph-like visual showing the percentages of looking at this simulator screen versus looking through the outside windows. A screenshot of the movie is presented in Figure 12. This relative visualisation was an elaboration of the 2D concept, with the reference norm being left out. The SME was convinced about the functionality and decided to invite other instructors to participate in the pilot experiment.

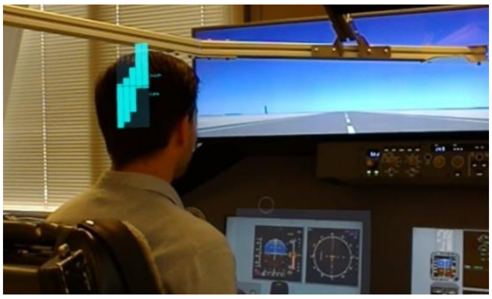


Figure 12. FOV of the HoloLens, including the blue dynamic graph (upper left corner) and the raw fixation location visualized as a white circle (bottom-centre).

Pilot experiment

In this phase, the tool will be tested to indicate the possible (dis)advantages, to obtain information about the visualisation and to decide on a prototype for further development and testing on a large scale. The main research question is:

Do instructors feel supported in their feedback by the ability to see where their students' eyes are fixating during a simulation?

Additional research questions to help answer the main research question are:
- *Do participating instructors wish to see students' gaze behaviour during simulation training?* This question tries to determine if instructors who are currently working at a major Dutch airline desire the same as literature studies between 1993 and 2016 propose. If this is the case, this application will more likely be of value for this airline's training.
- *Do instructors evaluate the user experience of the HoloLens application positively?*
A tool needs to have a positive user experience to be maximally supportive for the user. Therefore, it is important to look at both the user experience and the experienced support.

Before carrying out the pilot experiment, the application has been fine-tuned as much as possible based on the results from the third iteration. The results of the pilot experiment should be used as input to improve the application again. Implications and advice for future development will be discussed in the Conclusion of this paper.

Methods

Participants

There are two target groups in this experiment: flight instructors and (non-official) pilots with little flight experience. The intervention is designed to support (1) flight instructors in teaching (2) pilots who have some experience, but do not have a type-license yet.

Since it is hard to gather enough pilots in training, non-professional pilots - preferably with a Private Pilot License - were selected. This convenience sample existed of three men, of which two have never received any official pilot training, and one has his Glide Pilot License. The flying level of these men is close to the level of the target group considering the specific intervention: learning to approach and land in a new aircraft type.

A convenience sample of six instructors was selected to participate in this study. When testing the first participant, the application was not completely ready for testing. This distorted the answers and therefore he was left out of the analysis. The instructors were all male, had an average age of 46.2 years and decent experience ($M_{\text{years of experience flying}} = 26.2$, $M_{\text{years of experience teaching}} = 12.8$). Four instructors give training at a major Dutch airline and train pilots for Airbus A330 or the Boeing airplane types 737, 747, 777 and / or 787. The fifth participant is a former F16 instructor pilot.

The instructors group was extended with 14 instructors, gathered by a convenience sample of J. M. van Rooij's colleagues. A questionnaire examining the needs of the instructors for their simulation trainings was sent to these instructors and all participants in the pilot experiment. A total of 20 instructors ($M_{age} = 44.1$, 1 female), with an average experience of 13.6 years in training pilots completed the questionnaire.

Design

This pilot-study investigates if a new intervention could be useful in pilot training. For future testing, it is advised to check whether the application on a HoloLens is evaluated better than the application on a tablet, and which of at least 4 visualisations would be the best option for this simulation. This would conclude in a 2 (application: HoloLens vs. tablet) x 2 (visualisation: scanpath vs. heatmap) x 2 (colour: neutral vs. valent) design. The pilot study will focus on one visualisation on the HoloLens, namely a scanpath in neutral colours.

Procedure

All participants are informed about the complete procedure and aim of this study prior to participating. This includes information about the tool and the flight simulation. Before the pilot experiment starts, the instructor fills in an online questionnaire. The pilot will take a seat in the simulator first to start eye calibration. The instructor takes a seat behind the pilot. Both participants are watching the same simulator (Figure 1). The instructor can see the eye movements of the pilot displayed in the HoloLens.

When all participants are ready, the landing simulation will start. If the pilot does not make a turn-around, the simulation should be completed after approximately five minutes. The instructor wearing the HoloLens will give feedback to the pilot. This can be done during and / or after the simulation.

If the pilot or instructor does not feel comfortable enough to give or receive feedback during the first flight, the participants will repeat the same simulation. This feedback will be summarized by the researcher.

After completion of the simulation and debriefing, both instructor and pilot fill in a questionnaire regarding their experiences.

Materials

Simulation and scenario

The approach and landing phase of an Airbus A320 simulation training will be investigated. NLR has its own simulator (APERO) which is suitable to execute the described simulation. This simulator is interactive: all the input of the pilot will have impact on the simulation. The whole simulation has to be handled manually, is without emergencies or dysfunctionalities, and has a clear view. The simulation starts at ten miles from the landing strip and 2000 feet altitude, while the landing phase should start at one mile out and 1000 feet high. Decision height is set at 200 feet, which should be a turning point in the pilot’s gaze behaviour.

Eye tracker: Tobii 4C

The eye tracking data will be collected with the screen-based remote eye tracker Tobii 4C, 90Hz. This is a black bar of 30 cm wide that is placed within the simulator, connected to the screen displaying the PFD and ND. The pilot can freely move his eyes, head and body, which makes the intervention non-intrusive for the pilot. The 2D-gaze fixations will be collected and forwarded to the HoloLens in real-time. The Tobii bar is excellent in calibration: it is sufficient to calibrate the eyes once and afterwards track the eyes of different persons of similar height when seated in the same chair as the person who was calibrated.

Since the intervention is used to show the instructors in real-time where the pilots are looking, no AOIs have to be defined. However, based on the AOIs found by Dubois et al. (2015), Lefrancois et al. (2016) and AOIs as indicated by our subject matter expert, the focus will be on the outside windows and the Primary Flight Display. Other regions in the simulator will not be analysed in this study, but the data will be saved to extract other (more specific) AOIs in the future.

HoloLens and set-up

For this study, the HoloLens 3.0 is being used. The device connects with a computer via Wi-Fi, is completely free from any cables and is based on Unity-applications. This makes the complete set-up involving: 1) the Tobii 4C tracks the eyes of the pilot flying, 2) the eye-gaze regarding the simulator screen including the PFD and ND is translated to this screen's coordinates, 3) the coordinates are sent to a Unity application on another computer, 4) this Unity application is connected with the HoloLens which makes 5) the real-time data visible in the HoloLens. In this setup, the frequency of the eye tracking data will be reduced from 90 Hz to 50 Hz because of bandwidth limitations.

Questionnaires

The following three questionnaires were used to answer the (sub-)research questions (for details, see Appendix C).

Expectations / needs of instructors

This online questionnaire consisted of demographics, questions about the instructors' wishes for extra information during training, and some of the 2D concepts from iteration 1 to show gaze data. The concepts were evaluated by asking whether these would be useful or not, and in which order the concepts were considered best to worst.

Instructors' experiences

This questionnaire regarded 1) the user experience of the HoloLens application, 2) the experienced support for the debriefing, 3) the possible situations to apply this application and 4) advice to enhance the application.

For the first part, the Dutch version of the User Experience Questionnaire (UEQ) has been used (Schrepp, 2017a). This is a quick analysis to measure the attractiveness, perspicuity, efficiency, dependability, stimulation and novelty of the intervention. The items are represented by semantic differentials (two adjectives with opposite meanings) and can be valued on a 7-pt scale, with scores ranging from -3 to 3. One question example is: "Wat vond u van de applicatie in de HoloLens? Ga af op uw eerste instinct. Onplezierig - - - - Plezierig".

Since there are no existing instruments to measure the quality of feedback or feedback support, a new questionnaire has been constructed. This questionnaire focuses on the experienced support for feedback. Price et al. (2010) state that it is difficult and perhaps impossible to accurately measure feedback effectiveness. However, they also find it likely that reviewing the allocation of resources will support an effective feedback process. Based on literature, the following topics were included: can the errors be distinguished, can the causes of the errors be distinguished, is the tool intuitive to use and what is the experienced support for the debriefing? An example question of the experienced support scale is “De HoloLens ondersteunde mij in het feedback geven (tijdens de debriefing)” with answer options ranging from 1 (completely disagree) to 7 (completely agree), reflecting a 7pt Likert scale.

Pilots’ experiences

After the flight simulation, each pilot answered a few questions regarding their experience with an instructor who could see where they are looking.

Results

Needs questionnaire

On average, the instructors who filled out the needs questionnaire said that the available information during a simulation training is usually sufficient to give good feedback. However, from the open answers they could give on this question, it appeared that all but one instructors think that the currently available technological support systems are lacking something. Examples of answers demonstrating their opinion about the currently available technological tools were: “minimally sufficient to give feedback”, “sufficiently realistic” and “room for improvement”.

All except one instructor believed that seeing the gaze behaviour could support them during the debriefing session, both because it gives the instructors more information to base their feedback on, and to show the pilots where they have been looking. The odd-one out explained that it is important to (also) ask the student where they have looked in order to gain and grow their awareness about it. All but one instructor believed that receiving information about the cognitive load of the pilot would support the training.

The most popular 2D concept of how to present the gaze behaviour was the scan path (Table 1). Most other concepts were evaluated neutral, except for a negative evaluation of the lines in the HoloLens. The results varied widely between the instructors. This could be a confirmation of the claim that users are not always sure what they need (Rogers et al., 2009).

Table 1

Evaluation of 2D concepts by 20 instructors

	Scanpath	Heatmap	AOIs intern	Lines	AOIs extern	Graph
Like	16	8	8	4	8	7
Neutral	4	6	7	14	8	8
Dislike	0	6	5	2	4	5
Score	16	2	3	-2	4	2

Note. Scores are calculated as follows: like = 1, neutral = 0, dislike = -1.

An open question revealed that most instructors agree that the desired gaze behaviour during approach and landing is a visual shift from the inside displays to the outside windows. Other answers included a combination of behaviours, or that the desired behaviour depends on the situation.

Evaluation application

During the experiment, instructors tended to give their advice and share their ideas and experiences immediately. These experiences were mainly focused on the enhancement of the tool and the situations suitable for using it. All ideas and advices were added to the results stated below.

Instructors: User Experience HoloLens

The User Experience Questionnaire gives six different output-scores that cannot be combined (Schrepp, 2017b). According to the UEQ results ($N = 5$), the HoloLens application was attractive (1.80), perspicuous (1.65), quite efficient (1.20), slightly dependability (1.00), stimulating (1.80) and definitely novel (2.0).

All scales score higher than neutral, which ranges from -0.8 to 0.8. The HoloLens application scores high on novelty and almost neutral on dependability. The items regarding dependability concerned predictability, supportiveness, trusted and conform expectations, of which predictability and supportiveness scored higher (1.2 and 2.0 respectively) than the other two items (0.2 and 0.6).

Instructors: feeling of support for feedback

This corresponds to the newly created supportiveness scale, which has an average score of 5.62, signifying a positive experience of support. Specific questions revealed that the instructors experienced support (6.0), think it is an improvement of the training (6.0), thought the tool was intuitive (6.4), and could give more specific feedback than usual (6.4). The application did not have impact on the workload of the instructors (4.0) nor their confidence about their feedback (4.8). Instructors were capable of noticing the error (5.3) and the cause of the error (5.8). After using the application, their expectation about the value of seeing the gaze behaviour was extremely high (6.6).

One of the five instructors did not answer the questions about the intuitiveness of the application and the possibility to recognize the errors. The internal consistency of the answers of the four remaining instructors was analysed with SPSS. Despite the very small sample, the internal consistency was high. The support scale including 11 items had a Cronbach's alpha of 0.88. With exclusion of question six regarding the workload, the internal consistency rises to a Cronbach's alpha of 0.90. Questions one, six and seven had an item-total correlation below 0.47, while all other items had an item-total correlation of 0.74 or higher. This signifies that questions one, six and seven might not represent the main subject (which is ought to be "support") as much as the other items do.

Instructors: most preferred situations to apply this application

According to the five instructors, this application could be useful for pilots with every experience level, although the useful situations will probably differ per experience level. For more experienced pilots, this tool could be used to eliminate their struggles with one specific part of the simulation. Using the application for training approach and landing seemed useful for all instructors. On top of that they named emergency situations and collaboration between pilot flying and pilot monitoring. Most instructors agreed that the tool could especially be useful for short parts of a training; they did not like the prospect of wearing the HoloLens for a regular three-hour session. All instructors would use this application and would recommend it to their colleagues.

Instructors: remarks for product development

The main advice for further development of the tool was to give instructors the option to choose what will be shown in the application. A larger FOV in the HoloLens would be a big improvement. Three instructors think that placing the graph-like inside-outside ratio-visualisation on a fixed location in the FOV of the instructor (instead of on a fixed spot in the environment) would improve the application.

Pilots' experiences

The pilots did not have any problems with the HoloLens application. They did not notice that their eyes were tracked during the simulation. All pilots had the experience that the instructor could base their feedback on the source of their behaviour in the cockpit. The pilot with a Glide Pilot License is used to receiving feedback on his flights. He noticed that the feedback he received with this tool was more specific, more accurate and more connected to his own learning experiences than the feedback he usually receives without this tool.

Conclusion

Based on literature, it was expected that pilot instructors would benefit from a tool that enables them to see where the pilots in training are looking at during a simulation training. The following research question has been studied:

Do instructors feel supported in their feedback by the ability to see where their students' eyes are fixating during a simulation?

Additional research questions to help answer the main research question were:
- *Do participating instructors wish to see students' gaze behaviour during simulation training?*
- *Do instructors evaluate the user experience and experienced support of the HoloLens application positively?*

The first additional research question has been answered by analysing the results of an online questionnaire filled in by 20 pilot instructors. The claims arising from literature that instructors of pilots expect to be supported in their feedback by seeing the pilot's gaze behaviour has been confirmed. Hence, the first sub-research question has been answered in line with existing literature: they believe that seeing the gaze behaviour of the pilots in training could support the debriefing.

To answer the main research question, an application was built. The process of product development followed three iterations over the span of this study. Each iteration included a an analysis, a designing phase and a test of the created concepts. The results were used to redesign and improve the application. The first functioning concept has been tested on experienced pilot instructors.

Even though every instructor still had a few remarks, they were enthusiastic about the application. All instructors who tested the application would want to implement (an advanced version) in their own training and would recommend it to other instructors.

Further results from the pilot experiment were mixed. The first participant was more negative about the application than the others. This can be explained by two factors: he is a glide instructor - which implicates that he does not have experiences with simulation training

- and the application was not functioning optimally at that moment. Eliminating his answers resulted in a more homogeneous and positive evaluation of the application.

Overall, the application had a positive user experience, especially regarding the novelty and experienced support for the training. According to this small sample of instructors, the tool was intuitive to use, was an improvement of the training and did not impact the instructors' workload. The second sub-research question can therefore be answered: instructors evaluate the HoloLens application as a positive user experience.

The newly created support scale showed a high internal consistency, signifying a high reliability. This is a remarkable finding considering the extremely small sample, or could be a coincidence. Based on these results, the scale could be improved by eliminating three questions since these do not have a high correlation with the main subject being investigated. However, to optimize the support scale, it is important to test it on a larger group of flight instructors. An important result is that the instructors said that they can notice the error and the cause of the error, which are both important to give good feedback (Balzer & Doherty, 1989; Merriënboer & Kirschner, 2012). This result is especially interesting since the instructors lacked most of the information they usually receive during simulation trainings. Apparently, only seeing the pilots' movements and gaze behaviour was enough to notice the error and cause of the error.

The results from the experienced support scale reflect that the instructors experienced support by the HoloLens application. This result, combined with the positive answers on the sub-research questions, can answer the main research question: With this new application, instructors do feel supported in their feedback by the ability to see where their students' eyes are fixating during a simulation.

In conclusion, adding the HoloLens application to the current training situation may be a promising improvement of the instructors' feedback in pilot training.

Future development

Despite the positive results from the pilot experiment, the application is yet far from perfect. The instructors' remarks from the pilot experiment and their ideas about the different 2D concepts should be used as input for a new iteration. More visualisations should be created, the location of the graph should be reconsidered, the instructors should get the opportunity to choose these visualisations during the training, and the option to play back the gaze behaviour during the debriefing should be created. The application should be created in the HoloLens and on a tablet. All options need to be tested to choose the best prototype. This test could have a similar design as proposed in the method section of this paper. If the resulting prototype leaves many or substantial remarks, the iteration process should be continued to refine the application. When the application is (almost) free of remarks, the experienced support, user experience and effect on the training can be tested on a bigger scale.

To evaluate the success of the HoloLens application, it is important to take several aspects into account. The experienced support should be measured again. This can be done by question 11 of the UEQ and by the newly constructed questionnaire for experienced support. The internal consistency of the support questionnaire needs to be tested on a larger target group to improve and validate this questionnaire. If the questionnaire has a good internal consistency (with a Cronbach's alpha higher than 0.8), it should be used to measure the experienced support of both the HoloLens (and similar tablet) application and the current training equipment, to determine the differences between these tools.

The user experience should be tested to gain more insight in the UX components. Regarding the feasibility of including this intervention in regular pilot training, the focus should be on the pragmatic quality of the application, measured by the degree of perspicuity, dependability and efficiency. The relative low scores on dependability resulting from the current study will probably rise when the instructors are used to the product. It is expected that the hedonic components will have a high score because of the nature of the application – a HoloLens is still novel, attractive and stimulating. However, these are probably less important for companies to start an implementation process compared to the efficiency and support of the tool. The user experience can be tested throughout different situations, for example training sessions of three hours or emergency situations.

A more interesting approach to evaluate the success of this application is to assess the quality of the instructors' feedback and the learning curves of their students. This will result in a more objective evaluation of the application than examining the experiences of the instructors. Based on the pilot experiment of this study, hopes are high that this tool will indeed have a positive effect on the learning curve of the pilots. The pilot who received consecutive feedback on his gaze and operating behaviour mentioned that he was improving on his (gaze) behaviour because the feedback was so specific and targeted on his main struggles. Measuring the learning curve of the students requires several students receiving the usual training from their instructor, and a similar number of students receiving the usual training from the same instructor while using the HoloLens application. Ideally, several instructors participate in this experiment, all training students with and without the HoloLens application. The performance of the students being trained with and without the new application should be compared. Possibly, the instructor will affect the performance of the students, in which case the difference between the students should be analysed per instructor instead of as a whole group.

When the experiment yields positive effects, the application could be used in actual training situations. Before implementing it on a large scale, it would be a good idea to test the possible training situations on a select group of instructors or training situations.

Future use

The range of possible situations to use this application is vast. From the small number of instructors participating in this study, many options came forward. The application could be used for pilots with every experience level, in simulated emergency situations, during approach and landing simulations, to assess collaboration between pilots, and to eliminate struggles with one specific part of the simulation. Most instructors agreed that the tool could especially be useful for short training sections; they did not like the prospect of wearing the HoloLens for longer periods of time.

Next to the immediate possibilities to use this application, there are many future options if the tool will be developed and expanded to other functionalities. Equipped with more data, it can become a smart application by creating algorithms to notice the difference between good and bad gaze-behaviour. If the application contains this knowledge, it might even be able to replace the current instructors by giving automated feedback, such as Adams and Strickland (2011) did with technology education.

Considering all these options, the application seems to be a tool that should be implemented as fast as possible. However, there are still some constraints. As explained, the tool needs further development and fine-tuning. Besides, the choice of using a new technology is linked to financial possibilities, outside pressures, and classroom management (Sheehy, Ferguson, & Clough, 2014). Aviation is a working field considered to have little to no financial constraints. Outside pressure to implement this intervention in the training situations could be coming from instructors who are familiar with this intervention and see the huge benefits for their training. The outside pressure could also come from the continuous development in aviation, with other airlines being ahead (Emirates, 2017; Lufthansa, 2016). A small-scale implementation could give insight in the most relevant training situations to use the application.

References

Adams, R. H., & Strickland, J. (2011). The effects of computer-assisted feedback strategies in technology education: A comparison of learning outcomes. *Journal of educational technology systems, 40(2),* 211-223.

Anders, G. (2001). Eye-Tracking Research in an A330 Full Flight Simulator. *Berlin University of Technology Institute of Aeronautics and Astronautics.* Retrieved from: https://www.bioing.units.it/GIMO/pdf/eye_tracking.pdf Retrieved on: 09-03-2018.

Ashraf, H., Sodergren, M. H., Merali, N., Mylonas, G., Singh, H., & Darzi, A. (2018). Eye-tracking technology in medical education: A systematic review. *Medical teacher, 40(1),* 62-69.

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. *IEEE computer graphics and applications, 21(6),* 34-47.

Bacca, J., Baldiris, S., Fabregat, R., & Graf, S. (2014). Augmented reality trends in education: a systematic review of research and applications. *Journal of Educational Technology & Society, 17(4),* 133.

Balzer, W. K., & Doherty, M. E. (1989). Effects of cognitive feedback on performance. *Psychological bulletin, 106(3),* 410.

Boeing. (2016). Statistical Summary of Commercial Jet Airplane Accidents – Worldwide Operations.

Brown, D. L., Bautsch, H. S., Wetzel, P. A., & Anderson, G. M. (2002). Instrument scan strategies of F-117A pilots. Logicon Technical Services Inc. Dayton OH.

Carroll, M., Surpris, G., Strally, S., Archer, M., Hannigan, F., Hale, K., & Bennett, W. (2013). Enhancing HMD-based F-35 training through integration of eye tracking and electroencephalography technology. *International Conference on Augmented Cognition (pp. 21-30).* Springer, Berlin, Heidelberg.

D'Mello, S., Olney, A., Williams, C., & Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of human-computer studies, 70(5),* 377-398.

Dubois, E., Blättler, C., Camachon, C., & Hurter, C. (2015). Eye movements data processing for Ab initio military pilot training. *In Intelligent Decision Technologies (pp. 125-135).*Springer, Cham.

Emirates. (2017). Emirates and Seeing Machines pave the way for enhanced safety and training optimisation across global aviation industry. Retrieved from: https://www.emirates.com/media-centre/emirates-and-seeing-machines-pave-the-way-for-enhanced-safety-and-training-optimisation-across-global-aviation-industry# Retrieved on: 09-03-2018.

Goodman, A., Hooey, B. L., Foyle, D. C., & Wilson, J. R. (2003). Characterizing visual performance during approach and landing with and without a synthetic vision display: A part task study. *Proceedings of the 2003 NASA Aviation Safety Program Conference on Human Performance Modeling of Approach and Landing with Augmented Displays.* NASA Conference Proceedings NASA/CP-2003-212267.

Henneman, E. A., Cunningham, H., Fisher, D. L., Plotkin, K., Nathanson, B. H., Roche, J. P., ... & Henneman, P. L. (2014). Eye tracking as a debriefing mechanism in the simulated setting improves patient safety practices. *Dimensions of Critical Care Nursing, 33(3),* 129-135.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). Eye tracking: A comprehensive guide to methods and measures. OUP Oxford.

Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing?. *Journal of personality and social psychology, 79(6),* 995.

Kim, J., Palmisano, S. A., Ash, A., & Allison, R. S. (2010). Pilot gaze and glideslope control. *ACM Transactions on Applied Perception (TAP), 7(3),* 18.

Lefrancois, O., Matton, N., Gourinat, Y., Peysakhovich, V., & Causse, M. (2016). The role of Pilots' monitoring strategies in flight performance.

Lufthansa. (2016). Eye Tracking – Support for operations controllers. Retrieved from: https://www.lhsystems.com/article/eye-tracking-support-operations-controllers Retrieved on: 10-04-2018.

Merriënboer, J. J. van, & Kirschner, P. A. (2012). *Ten steps to complex learning: A systematic approach to four-component instructional design.* Routledge.

Microsoft. (2018). Microsoft Hololens. Retrieved from: https://www.microsoft.com/en-us/hololens Retrieved on: 09-03-2018.

O'Meara, P., Munro, G., Williams, B., Cooper, S., Bogossian, F., Ross, L., ... & McClounan, M. (2015). Developing situation awareness amongst nursing and paramedicine students utilizing eye tracking technology and video debriefing techniques: A proof of concept paper. *International emergency nursing, 23(2),* 94-99.

Oijen, J. van, & Voskamp, J. (2016*).* A-PiMod Intelligent Tool for Instructors to Assess Pilot Behaviour in the Simulator.

Price, M., Handley, K., Millar, J., & O'donovan, B. (2010). Feedback: all that effort, but what is the effect?. *Assessment & Evaluation in Higher Education, 35(3),* 277-289.

Proctor, R. W., & Van Zandt, T. (2008). *Human factors in simple and complex systems*. Boca Raton: CRC press.

Rogers, Y., Sharp, H., & Preece, J. (2009). *Interaction design: beyond human-computer interaction.* West Sussex, England: John Wiley & Sons.

Schrepp. (2017a). User Experience Questionnaire. Retrieved from: http://www.ueq-online.org/ Retrieved on: 09-03-2018.

Schrepp. (2017b). User Experience Questionnaire Handbook. Retrieved from: http://www.ueq-online.org/ Retrieved on: 09-03-2018.

Sheehy, K., Ferguson, R., & Clough, G. (2014). *Augmented education: bringing real and virtual learning together.* Springer.

Shlechter, T. M. (1992). Developing Automated Feedback Materials for a Training Simulator: An Interaction between Users and Researchers.

Wermeskerken, M., Litchfield, D., & Gog, T. (2018). What am I Looking at? Interpreting dynamic and static gaze displays. *Cognitive science, 42(1),* 220-252.

Wetzel, P. A., Anderson, G. M., & Barelka, B. A. (1998, October). Instructor use of eye position based feedback for pilot training. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 42(20),* 1388-1392. Los Angeles: SAGE Publications.

Appendix A

Iteration creations

During the **first iteration**, a number of visualisations were created. The following images show a simulator cockpit with two pilots and one instructor who is either wearing a HoloLens or holding a tablet.

Figure A1. Heatmap through HoloLens.

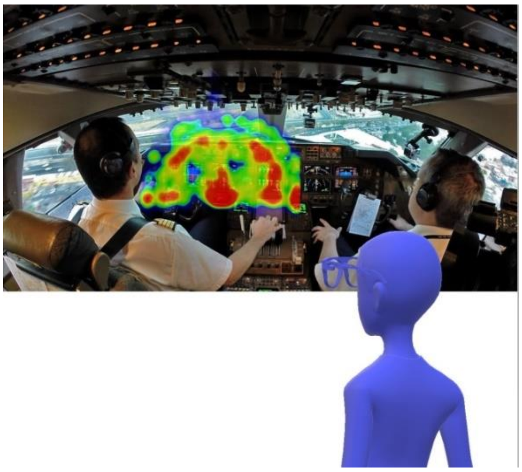


Figure A2. Different heatmap through HoloLens.

Figure A3. Scanpath through HoloLens.

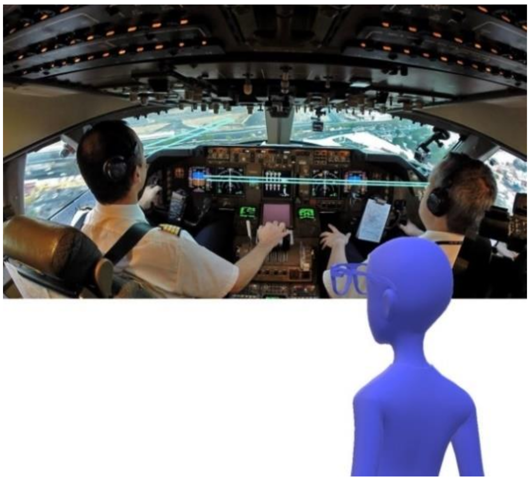


Figure A4. Eye gaze through HoloLens.

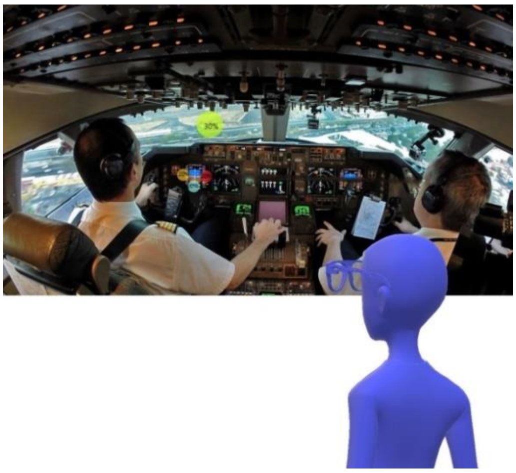


Figure A5. Areas Of Interest through HL.

Figure A6. AOIs fixated through HoloLens.

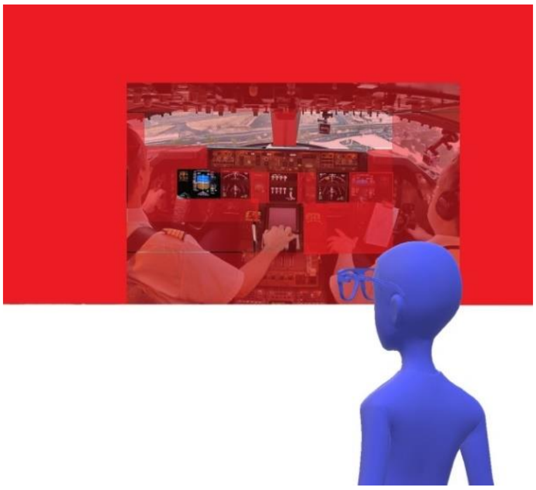


Figure A7. AOIs non-fixated through HL.

Figure A8. Heatmap through tablet.

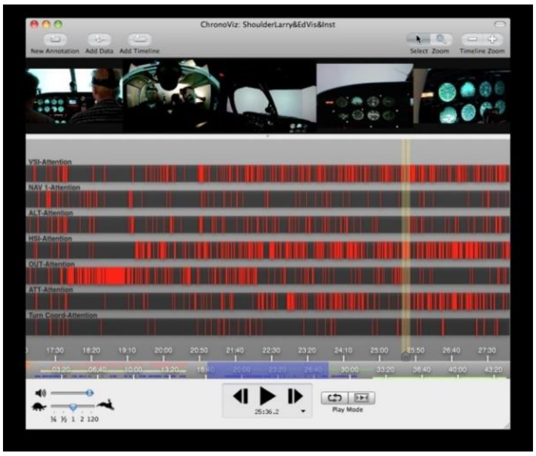


Figure A9. Graphical display of AOIs in time.

Figure A10. Foveation.

The scanpath, foveation and the AOIs as displayed in figure A5 were chosen to elaborate. Since the instructor was mostly interested in the scan ratio between the outside windows and the inside displays, this was taken into consideration to create in the application.

To gain insight in the dynamic possibilities of these concepts, the following YouTube videos were shown:

Cumulative heatmaps:
https://www.youtube.com/watch?v=3cwUxYvpmfw
https://www.youtube.com/watch?v=Qk4V_x6B7jY
Real-time (opacity) heatmaps:
https://www.youtube.com/watch?v=BY2fOW-5LI0
https://www.youtube.com/watch?v=K_hhFxg8vW0
Real-time scanpaths:
https://www.youtube.com/watch?v=SQxrsUXqKCM
https://www.youtube.com/watch?v=qUZeMH8hed0
https://www.youtube.com/watch?v=RpQVSmGvbMo
https://www.youtube.com/watch?v=nGGNjcwzofg
Non-preferred options:
https://www.youtube.com/watch?v=BPTyRucbfkg

During the **second iteration**, the visualisation concepts chosen from the first iteration were constructed in Unity, the programme used to create the HoloLens application. Meanwhile, new (2D) concepts were created to show the ratio of in/out-scan behaviour.

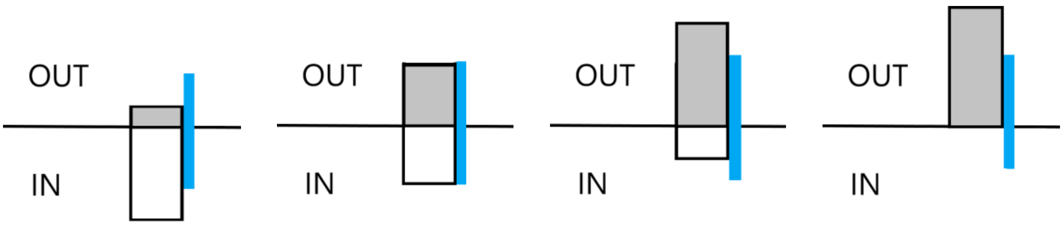


Figure A11. Dynamic graph showing the percentages of time looking at the outside windows or inside displays, compared to a benchmark in blue.

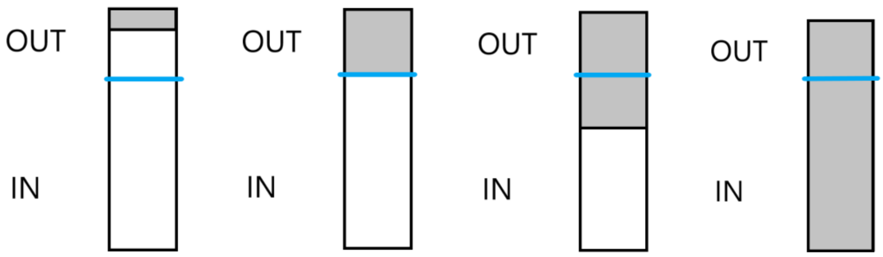
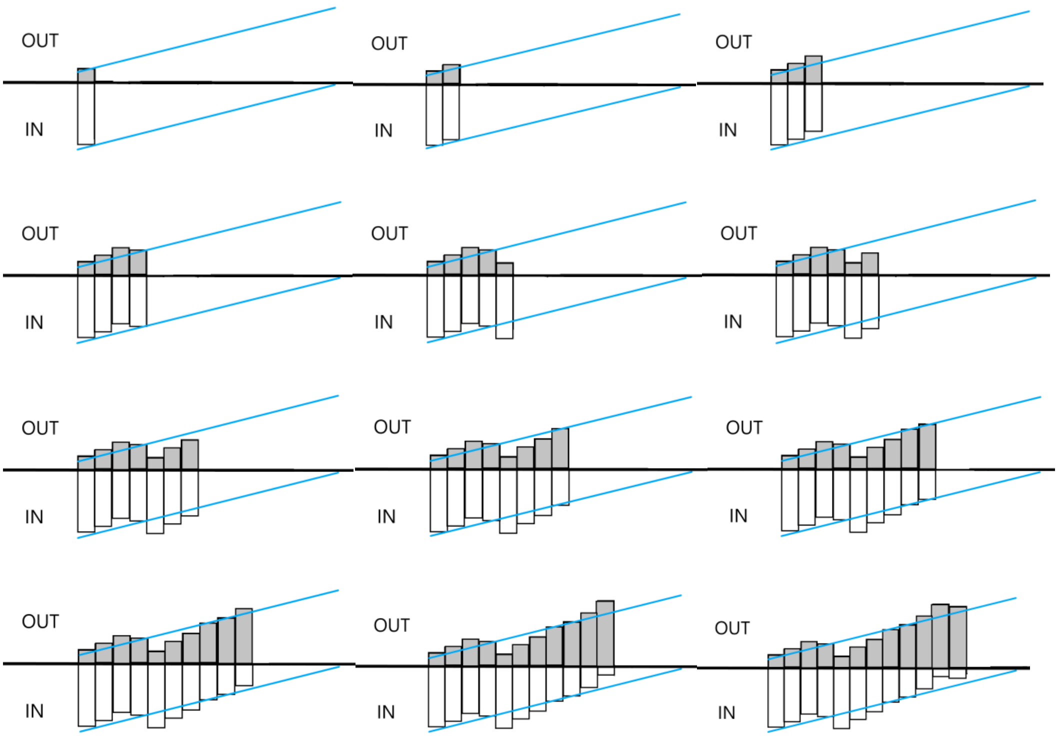


Figure A12. Dynamic graph showing the percentages of time looking at the outside windows or inside displays, compared to a benchmark in blue.



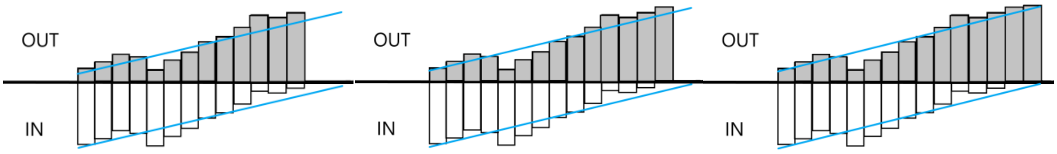


Figure A13. Cumulative dynamic graph from figure 11 without replacing old ratios.

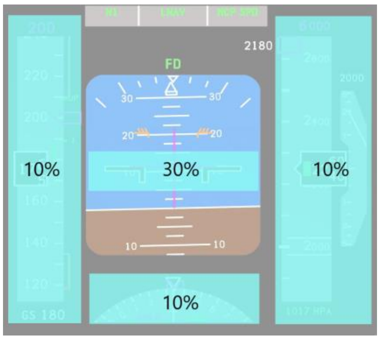


Figure A14. AOIs in PFD.

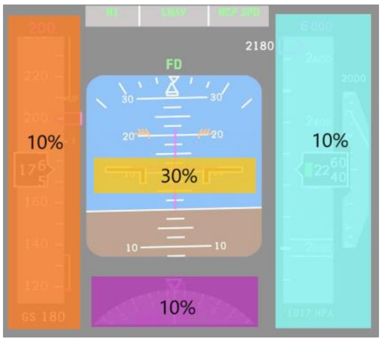
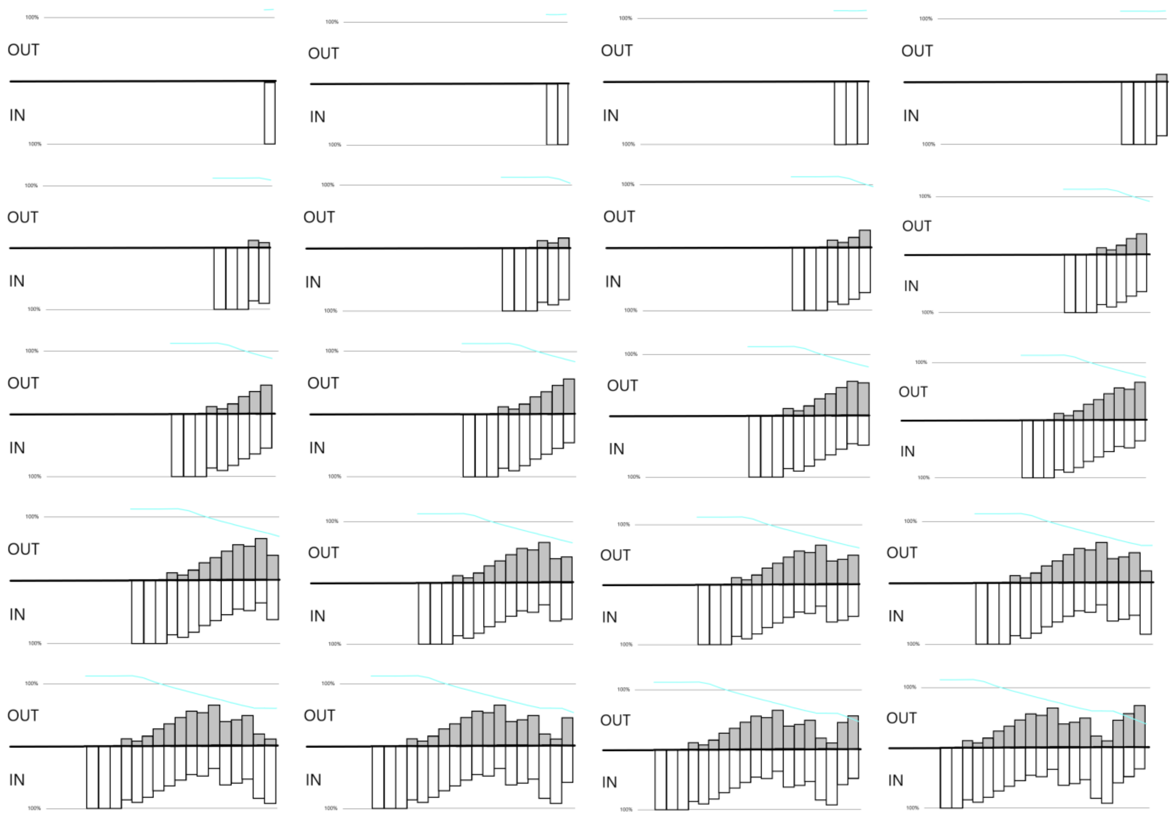


Figure A15. AOIs in PFD in colour.

The instructor was really enthusiastic about the graphs, except the one from figure A12. Instead of displaying a benchmark, he would rather see the altitude of the airplane. Concerning the AOI displays, he preferred one colour over different colours.

During the **third iteration**, the dynamic graph was adapted and created in Unity.



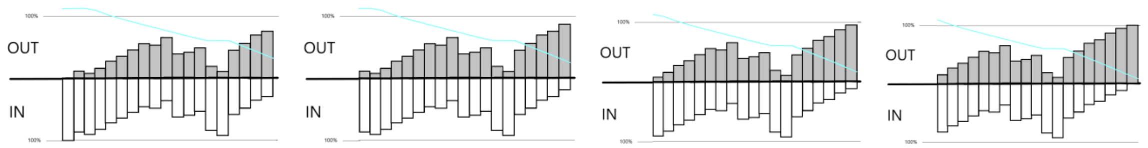


Figure A16. Dynamic graph showing the percentages of time looking at the outside windows or inside displays, including a blue line which represents the airplane altitude.

The concept of the **third** iteration, which was also the first creation, is shown in Figure A17:

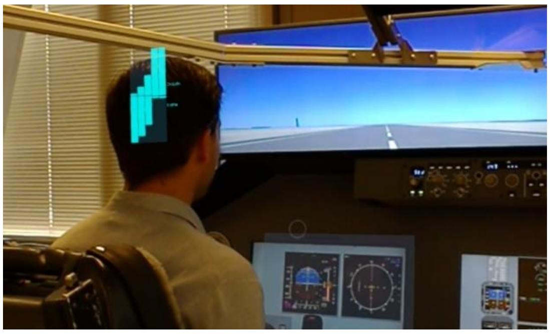


Figure A17. FOV of the HoloLens of the first creation of application as tested in pilot experiment. The dynamic graph can be found in the upper left corner of the image, the raw fixation location of the pilot is visualized as a white circle in the bottom-centre of the image.

Appendix B
Creation of HoloLens application

After the conceptualization and design process, we started creating the application. The eye tracker was connected with a USB to the simulator-screen which shows the PFD and ND. A Software Development Kit (SDK) was used to control the eye tracker, including eye calibration and data gathering. The eye tracking data consists of a timestamp, fixation location and a connotation informing if the pilot is looking at the screen or not. This data is saved as a .bse-file. This extension is created by NLR to synchronize data of different instruments. This data arrives in Unity, where it is being used as input for the visualisation. The simplest visualisation looks like a coloured dot and represents the coordinates of the pilot’s fixations in real-time. To keep it real-time, the frequency or refreshment rate of all instruments need to be equal, or higher than the former instrument in the data flow. Since every object in the HoloLens can be placed in the real world, it is important to attribute a location where the data will be displayed while wearing the HoloLens. The data need to be assigned to a location in Unity and in the real world. The locations in Unity can be set with coordinates relating to the camera position. The locations in the real world can be set when starting the application by using a marker that Unity can recognize. In Unity, a 3D model of the simulator is created, including the marker at the same relative position. Unity does not need to be installed on the data house computer; it works as a standalone application. To save the data for later use, the data from different instruments need to be synchronized in time. This includes the simulator input. The complete application works simultaneously with the simulation.

With the available equipment, it is impossible to track the fixation locations of the pilot outside the boundaries of the simulator-screen. The remote eye trackers can only be set to one specific infrared frequency, which results in an interference of reflections when using multiple eye trackers for the same person. A mobile or head-mounted eye tracker is not desirable in this situation, since it has a huge impact on the comfort of the pilots.

Appendix C

Questionnaires

Needs questionnaire

Universiteit Utrecht

Hallo! Ik ben Jeanine Vlasblom en hoop door middel van uw hulp met deze enquête mijn masterscriptie te kunnen versterken.

Het doel van de enquête is om een inventarisatie te maken van de behoeftes vanuit instructeurs voor simulatie trainingen. De vragenlijst zal ongeveer 5 minuten van uw tijd in beslag nemen.

Alle gegevens worden geanonimiseerd en de ruwe data zal niet aan derden beschikbaar worden gesteld. Bij publicatie van de data wordt er zorgvuldig op gelet dat de gegevens niet individueel herkenbaar zijn. Als u tijdens de enquête besluit deze niet af te willen maken, kunt u op elk gewenst moment stoppen.

Alvast hartelijk bedankt!

Bevestig uw deelname.

Ik bevestig bij deze dat ik goed ben geïnformeerd over het doel en de werkwijze van het onderzoek en de bovenstaande uitleg begrijp en accepteer. Mijn deelname aan het onderzoek is geheel vrijwillig en ik zal de vragenlijst nauwkeurig en serieus invullen.

→

Universiteit Utrecht

Bent u een...

Man

Vrouw

Wat is uw leeftijd?

Hoeveel jaren vliegervaring heeft u?

Voor welk type vliegtuig leidt u (toekomstige) piloten op?

Hoeveel jaar ervaring heeft u in totaal met het trainen van piloten?

Wat is het gewenste kijkpatroon van een piloot tijdens de nadering en landing?

→

Universiteit Utrecht

Hoe beoordeelt u de huidige technologische hulpmiddelen aanwezig bij een simulatie training? (Denk aan instructor screen, etc)

Vindt u de beschikbare informatie tijdens de huidige simulatie trainingen voldoende om goede feedback te kunnen geven?

- Altijd
- Meestal wel
- Meestal niet
- Nooit

Zou *het zien waar de piloten kijken tijdens de simulatie training* ondersteuning kunnen bieden aan de debriefing sessie? (meerdere antwoorden mogelijk)

- Nee
- Ja, want dan heb ik als instructeur meer informatie om feedback te kunnen geven.
- Ja, want dan kan ik de piloten in opleiding direct laten zien waar ze hebben gekeken.
- Ja, want...

Zou *informatie over de cognitieve belasting van de piloot* ondersteuning kunnen bieden aan de debriefing sessie? (meerdere antwoorden mogelijk)

- Nee
- Ja, want dan heb ik als instructeur meer informatie om feedback te kunnen geven.

Ja, want dan kan ik de piloten in opleiding direct laten zien waar ze hebben gekeken.

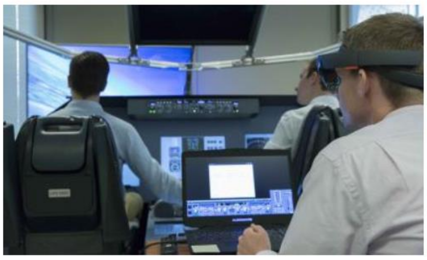
Ja, want...

→

Universiteit Utrecht

Op dit moment zijn wij bezig met het ontwikkelen van een applicatie om de oogbewegingen van de piloot tijdens een simulatie training te tonen aan de instructeur. Op de volgende pagina's ziet u verschillende voorbeelden van hoe deze informatie gepresenteerd kan worden. We zijn hierbij uitgegaan van een opstelling waarin de instructeur een HoloLens draagt (zie foto). Hierdoor ziet hij alles wat men zonder HoloLens ook ziet, maar dan met extra, virtuele informatie over de werkelijkheid heen geprojecteerd.

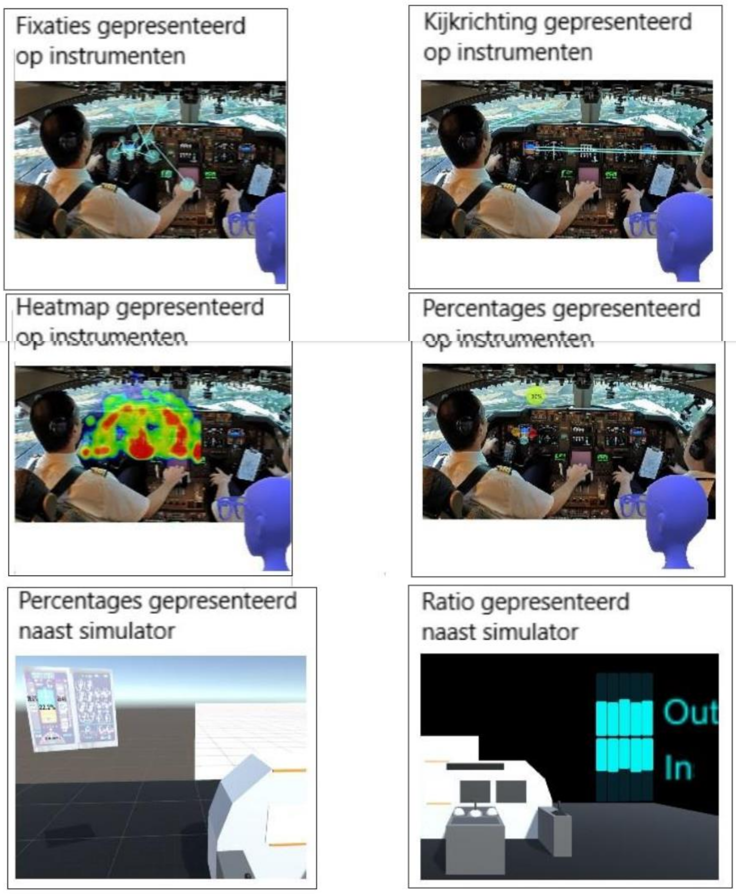
Het kijkgedrag van de piloot kan op verschillende manieren gepresenteerd worden. Bij een deel van de opties worden de oogbewegingen geprojecteerd op de locatie waar de piloot kijkt: in de simulator. Bij een ander deel van de opties wordt berekend waar de piloot de afgelopen 5 seconden heeft gekeken: de percentages kijkgedrag binnen het PFD en de ratio 'kijkend naar buitenwereld' - 'kijkend naar instrumenten'. Deze percentages worden naast de simulator geprojecteerd.



→

Universiteit Utrecht

Op welke manier zou u het kijkgedrag van de piloot willen zien? Klik op de afbeeldingen om aan te geven of u het kijkgedrag zo zou willen zien (groen), of dubbelklik wanneer deze optie u niet effectief lijkt (rood).



→

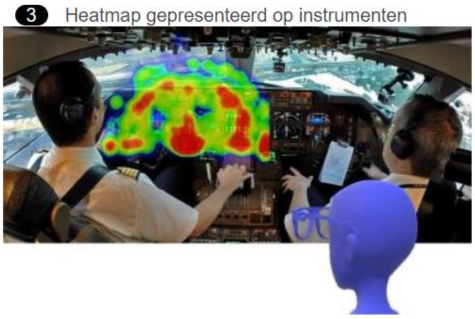
Universiteit Utrecht

Op welke manier zou u het kijkgedrag van de piloot het liefst willen zien? Sleep de voorbeelden op volgorde met 1 als beste optie en 6 als slechtste optie.

1 Fixaties gepresenteerd op instrumenten

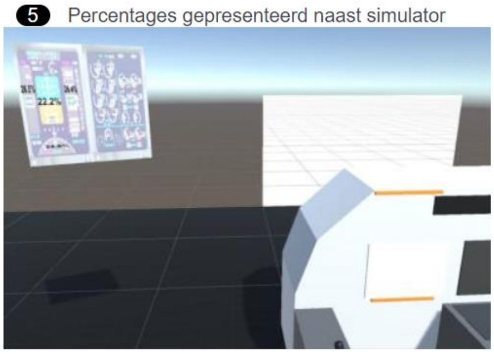
2 Kijkrichting gepresenteerd op instrumenten

3 Heatmap gepresenteerd op instrumenten

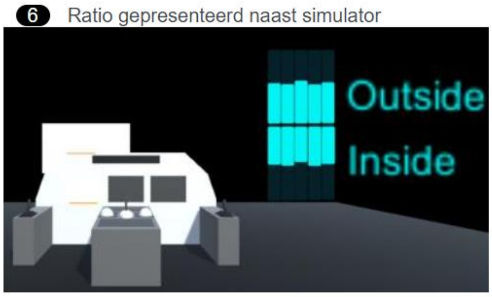


4 Percentages gepresenteerd op instrumenten

5 Percentages gepresenteerd naast simulator



6 Ratio gepresenteerd naast simulator



→

Universiteit Utrecht

Hartelijk dank voor uw deelname! Voor meer informatie kunt u mailen naar jeaninevlasblom@gmail.com.

Heeft u nog overige opmerkingen?

→

Universiteit Utrecht

Bedankt voor uw tijd om aan deze enquête deel te nemen.
Uw antwoord is geregistreerd.

Instructors' experience HoloLens:

Universiteit Utrecht

Hallo! Ik ben Jeanine Vlasblom en hoop door middel van uw hulp met deze enquête mijn masterscriptie te kunnen versterken.

Het doel van de enquête is om een inventarisatie te maken van de behoeftes vanuit instructeurs voor simulatie trainingen. De vragenlijst zal ongeveer 5 minuten van uw tijd in beslag nemen.

Alle gegevens worden geanonimiseerd en de ruwe data zal niet aan derden beschikbaar worden gesteld. Bij publicatie van de data wordt er zorgvuldig op gelet dat de gegevens niet individueel herkenbaar zijn. Als u tijdens de enquête besluit deze niet af te willen maken, kunt u op elk gewenst moment stoppen.

Alvast hartelijk bedankt!

Bevestig uw deelname.

Ik bevestig bij deze dat ik goed ben geïnformeerd over het doel en de werkwijze van het onderzoek en de bovenstaande uitleg begrijp en accepteer. Mijn deelname aan het onderzoek is geheel vrijwillig en ik zal de vragenlijst nauwkeurig en serieus invullen.

→

Universiteit Utrecht

Bent u een...

- Man
- Vrouw

Wat is uw leeftijd?

Hoeveel jaren vliegervaring heeft u?

Voor welk type vliegtuig leidt u (toekomstige) piloten op?

Hoeveel jaar ervaring heeft u in totaal met het trainen van piloten?

Wat is het gewenste kijkpatroon van een piloot tijdens de nadering en landing?

→

Universiteit Utrecht

Wat vond u van de applicatie in de **HoloLens**? Ga af op uw eerste instinct.

onplezierig	○	○	○	○	○	○	○	plezierig
onbegrijpelijk	○	○	○	○	○	○	○	begrijpelijk
creatief	○	○	○	○	○	○	○	fantasieloos
makkelijk te leren	○	○	○	○	○	○	○	moeilijk te leren
waardevol	○	○	○	○	○	○	○	waardeloos
vervelend	○	○	○	○	○	○	○	prikkelend
oninteressant	○	○	○	○	○	○	○	interessant
onvoorspelbaar	○	○	○	○	○	○	○	voorspelbaar
snel	○	○	○	○	○	○	○	langzaam
origineel	○	○	○	○	○	○	○	conventioneel
belemmerend	○	○	○	○	○	○	○	ondersteunend
goed	○	○	○	○	○	○	○	slecht
complex	○	○	○	○	○	○	○	eenvoudig
afstotend	○	○	○	○	○	○	○	begeerlijk
gebruikelijk	○	○	○	○	○	○	○	nieuw
onaangenaam	○	○	○	○	○	○	○	aangenaam
vertrouwd	○	○	○	○	○	○	○	niet vertrouwd
motiverend	○	○	○	○	○	○	○	demotiverend
volgens verwachtingen	○	○	○	○	○	○	○	niet volgens verwachtingen
inefficiënt	○	○	○	○	○	○	○	efficiënt
overzichtelijk	○	○	○	○	○	○	○	verwarrend
onpragmatisch	○	○	○	○	○	○	○	pragmatisch
ordelijk	○	○	○	○	○	○	○	rommelig
aantrekkelijk	○	○	○	○	○	○	○	onaantrekkelijk
sympathiek	○	○	○	○	○	○	○	onsympathiek
conservatief	○	○	○	○	○	○	○	innovatief

→

Universiteit Utrecht

Deze vragen gaan over de applicatie in de **HoloLens**

	Helemaal niet mee eens	Niet mee eens	Enigszins mee oneens	Noch eens noch oneens	Enigszins mee eens	Mee eens	Helemaal mee eens
De HoloLens **ondersteunde** mij in het feedback geven (tijdens de debriefing)	○	○	○	○	○	○	○
Ik kon me dankzij de HoloLens beter focussen op de **oorzaak** van het vlieggedrag dan in de normale trainingsopstelling	○	○	○	○	○	○	○
Ik kon dankzij de HoloLens **gerichter** feedback geven dan in de normale trainingsopstelling	○	○	○	○	○	○	○
Ik voelde mij dankzij de HoloLens **zekerder** in het geven van feedback dan in de normale trainingsopstelling	○	○	○	○	○	○	○
Ik vond de HoloLens **onhandig** in deze simulatie-training	○	○	○	○	○	○	○
De HoloLens gaf mij extra **werklast**	○	○	○	○	○	○	○
Ik kon de **oorzaak** van het gedrag van de piloot beter herkennen met de HoloLens dan in de normale trainingsopstelling	○	○	○	○	○	○	○
De HoloLens **verbetert de training** in het algemeen	○	○	○	○	○	○	○
Het **zien waar de piloot in opleiding kijkt** tijdens de simulatie heeft toegevoegde waarde	○	○	○	○	○	○	○
Het kijkgedrag van de piloot is op een **intuitieve** manier **gevisualiseerd**	○	○	○	○	○	○	○
Ik kon de gemaakte **fouten goed herkennen**	○	○	○	○	○	○	○

→

Universiteit Utrecht

Deze interventie (mits geperfectioneerd) zou van **toegevoegde waarde** kunnen zijn bij...

- Complete beginners
- Semi-beginners
- Gevorderden
- Anders, namelijk...
- Niemand

Deze interventie (mits geperfectioneerd) zou van **toegevoegde waarde** kunnen zijn bij...

- Nadering en landing
- Noodgevallen / alarm situaties
- Samenwerking tussen piloot en co-piloot
- Anders, namelijk...
- Geen enkel trainingsonderdeel

Zou u deze interventie **gebruiken** tijdens een training?

- Ja
- Nee

Zou u deze interventie **aanraden** aan anderen om te gebruiken in trainingen?

- Ja
- Nee

→

Universiteit Utrecht

Hoe kan deze interventie worden **verbeterd**? (meerdere antwoorden mogelijk)

- Geen HoloLens gebruiken, maar een tablet.
- Geen HoloLens gebruiken, maar een...
- De data op een andere manier visualiseren, bijvoorbeeld ...
- De grafiek helemaal weghalen uit het beeld van de HoloLens.
- De grafiek vastzetten in de linkerbovenhoek van het gezichtsveld in plaats van naast de simulator
- De fixatielocaties (op het PFD/ND) helemaal weghalen uit het beeld van de HoloLens.
- De instructeur tijdens de simulatie zelf laten kiezen wat hij wil zien in de HoloLens.
- De applicatie toepassen op een andere trainingsituatie, bijvoorbeeld ...
- Informatie over de cognitieve belasting van de piloot toevoegen aan het beeld in de HoloLens.
- Informatie over de mate van aandacht van de piloot toevoegen aan het beeld in de HoloLens.
- Informatie over de simulator-input toevoegen aan het beeld in de HoloLens, bijvoorbeeld de onderdelen...
- Anders, namelijk...

Heeft u nog andere opmerkingen, tips of (positieve / negatieve) ervaringen?

→

Universiteit Utrecht

Hartelijk dank voor uw deelname!

Bij interesse in de uitkomsten van het onderzoek kunt u hier uw emailadres achterlaten.

→

Pilots' experience:

Universiteit Utrecht

Hoe veel officiële trainingen heeft u gevolgd?

- Ik heb nooit een officiële training of opleiding tot piloot gevolgd.
- Ik ben begonnen aan een opleiding, maar heb (nog) geen Private Pilot License.
- Ik heb mijn Private Pilot License.

Hoe heeft u de **simulatie ervaren**, wetende dat de instructeur mee kon kijken met waar u hebt gekeken?

Had u het idee dat de instructeur feedback kon geven op de **oorzaak** van uw handelingen **dankzij de interventie**?

- Ja
- Hij kon wel feedback geven op de oorzaak van mijn handelingen, maar even goed als wanneer hij mijn oogbewegingen niet kon zien
- Nee, hij kon de oorzaak niet achterhalen

Hoe heeft u de **EEG** meting ervaren?

onplezierig	○	○	○	○	○	○	○	plezierig
goed	○	○	○	○	○	○	○	slecht
afstotend	○	○	○	○	○	○	○	begeerlijk
onaangenaam	○	○	○	○	○	○	○	aangenaam
aantrekkelijk	○	○	○	○	○	○	○	onaantrekkelijk
sympathiek	○	○	○	○	○	○	○	onsympathiek

Heeft u nog andere opmerkingen, tips of (positieve / negatieve) ervaringen?

→

Appendix D

Personal contributions

From constructing the idea of this experiment to designing, creating and testing the tool, I have been actively engaged in this project.

The first weeks I have been focusing on the construction of the experiment. I planned brainstorm sessions with several experts to create the first ideas about the tool and how to measure its effect. While exploring this measurement, I discovered that there is no validated measurement instrument to analyse the focus of the feedback, the tools used by an instructor to support their feedback, the quality of feedback, nor the effect of feedback. To measure the experienced support, I created a small questionnaire. Since this project turned out to be a product design process, I decided to also measure its user experience.

To create the application as designed, two major things needed to be taken care of: the dataflow between all instruments and the visualisation of the data. I attended every meeting with the project team when discussing the best data flow. To help the project team with the visualisation, I created all 2D visualisations which are shown in Appendix A. Next to that, I created several 3D visualisations. In Unity I created a 3D, scaled model of the simulator used in this experiment (see Figure 19). To create visualisations of the eye tracking data, I created a script with fake eye tracking data, since I couldn’t use the eye tracker at that moment. I coded these scripts in C# and attached them to several objects. The fixation locations are displayed by an orange dot, as can be seen in Figure D1 on the outside window screen. The ratio between looking to the outside windows and inside displays are displayed in a dynamic graph, see Figure D1.

In my spare time I contributed to other projects by helping several colleagues and peer-interns with their subjects, which included creating a training about sleep deprivation, EEG, processing big data in Excel, and designing information displays in a pilot’s Head-Up Display.

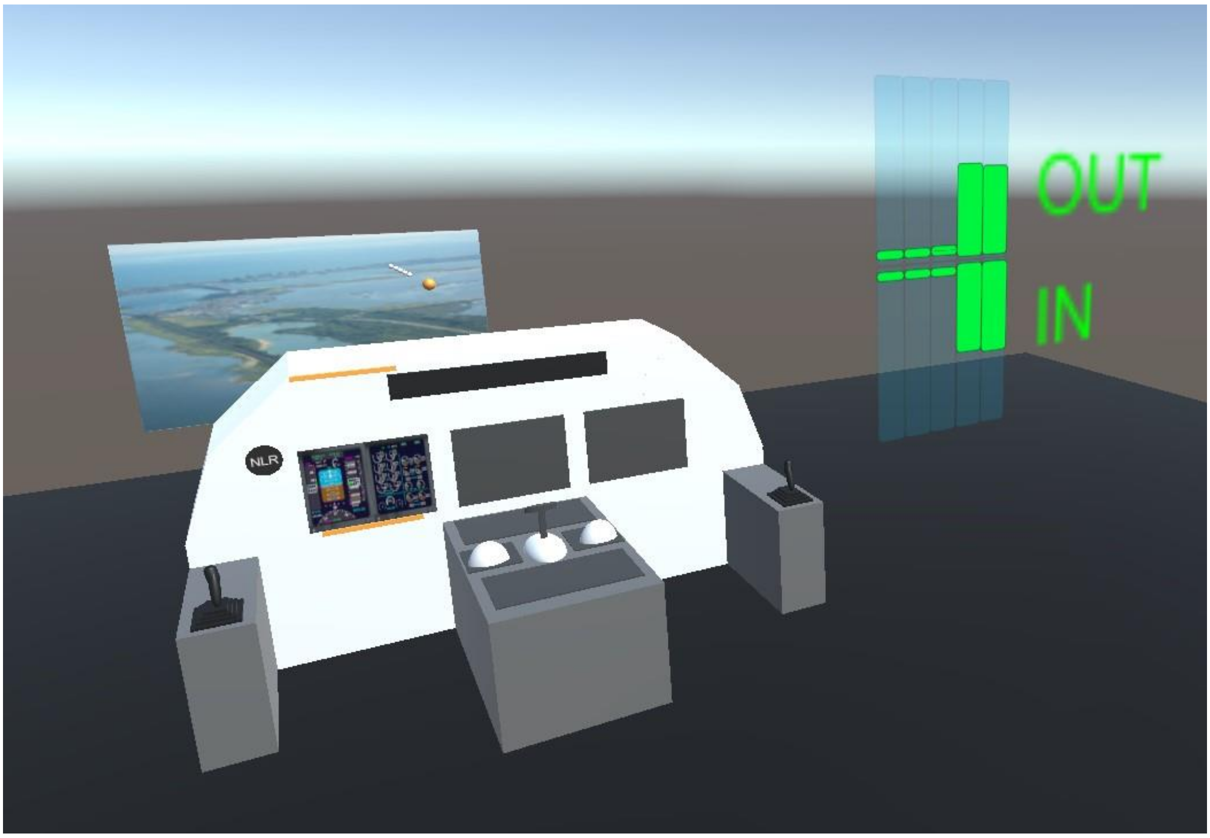


Figure D1. 3D model of the APERO-simulator created in Unity, including orange visualisation of the raw eye tracking data and dynamic graph which updates every couple of seconds.