Flight Planning with Mixed Reality

Adrian Sarbach
D-BAUG, ETH Zurich
asarbach@ethz.ch

Styrmir Óli Thorsteinsson
D-INFK, ETH Zurich & University of Iceland
sthorsteinss@student.ethz.ch

Rohit Koonireddy
University of Zurich
rohit.koonireddy@uzh.ch

Abstract

This project, part of the mixed reality course offered at ETH Zurich in autumn 2023, studied the design of a mixed reality tool for flight planning, using HoloLens 2. Currently, pilots use a 2D medium (paper or digital), to plan their route in 3D space. This project explores the technical feasibility of employing mixed reality as a tool to help pilots prepare their flight. A particular focus was on providing information to pilots to aid their navigation in 3D.

Figure 1. How a mixed reality flight planning application could be used. Concept art.

1. Background

Aviation inherently takes place in 3D, or even four dimensions when taking into account time. Flight planning is done on a paper map, or within a computer/tablet application. In general, these two methods only show space in 2D, but neglect the third dimension. Extended reality (XR) allows for all three dimensions to be visualised, and can thus be a useful tool for flight planning.

XR, including augmented, mixed and virtual reality (AR, MR, and VR, respectively), has seen numerous use cases within the aviation domain. A recent publication gives an overview of existing literature of XR within aviation [1]. Here, a significant number of the surveyed papers focuses on training for pilots, while somewhat fewer results focus on maintenance, both training and for process guidance.

The focus on training is understandable, given how widespread simulators are for pilot training (which could be considered "VR"). Nevertheless, results from a number of domains show that XR does not necessarily lead to better performance than conventional training methods [2]. A number of mostly positive thoughts on how XR can be used for training was shown by Brown [3]. She also gives detailed explanations of the differences between AR, MR and VR.

Heads-up displays (HUD) have been used in military aviation for a considerable amount of time, and recently been introduced in civil commercial aviation as well. These displays are a form of MR [4] which are used in-flight. Katins et al. have recently published two papers MR in general aviation. Firstly one on pilot concerns, technical prerequisites, and use cases [5], and secondly on how MR can be used to support pilot workload [6].

Drawing spatial information conventionally takes place on a 2D medium - on the map, or within a mapping application. With aviation however taking place in the third dimension, the work by Kim et al. [7] can provide an insight on how potentially flight trajectories can be drawn in three dimensions.

In summary, XR is nothing new in aviation. AR has been in the cockpit for a significant amount of time. XR in all flavours has been tested for training purposes, with mixed results. Lastly, we have seen researchers contemplate on drawing in 3D in the form of sketches. What is missing is the connection between drawing in 3D, and using XR to represent a scale model of our world for flight planning.

2. Implementation

This section will first describe the data we used for our application. We will not only present the data we finally visualised, but also some possible extensions. Later, in three subsections, we will elaborate on how our implementation changed, from the first prototype the the final produce.

2.1. Data and Software

Throughout our project, we strove to use open data and open source software, where possible. All data we used is open, either public domain, or open government data. Software for processing the data was open-source. Namely, we used Python, QGIS [8], and Blender [9]. Non-open software includes Unity [10], Visual Studio, as well as initially Cesium [11] and MapBox SDK [12] for Unity. Finally, we used the Microsoft HoloLens Toolkit (MRTK) [13] and Vuforia [14] for their functionality with the HoloLens.

The visual basis for our application was 3D terrain data from the Federal Office of [15], with a resolution of 25 m.

We used swissNAMES3D [16], also from swisstopo, as basis for geographic name information. We wanted to visualise some geographic names to help the users, pilots, with their orientation. This dataset contains the names of points (e.g., mountain peaks), lines (e.g., rivers), and polygons (e.g., settlement boundaries), and includes their location and shapes. We split up the entire dataset into separate JSON files – one file for points, lines, and polygons each – using Python.

A further open government dataset of the Federal Office of Civil Aviation were the aeronautical hazards of Switzerland [17]. Evidently, pilots want to avoid flying into hazards. Aeronautical hazards include towers, wind turbines, but also cables of cable cars and electricity transmission lines. Like geographic name information, we split up the hazards into separate JSON files for points and lines using Python.

Lastly, we used public domain data from OurAirports [18] to get data on airports and their runways. Two separate files contained data on airports and their runways, respectively. From their original format as CSV, we removed all airports (and runways) outside of Switzerland.

2.2. Application

After unsatisfactory results with Cesium and MapBox, we decided the visual base of our application would be a 3D terrain model we generate ourselves. Using GQIS and Blender, 3D terrain objects suitable for visualisation in Unity were generated from the DHM25 data. We created two 3D terrain objects, for two different study areas.

Our application knows two different modes: The first mode is similar to VR, where the 3D model, and everything placed on top of it, floats freely in space. We called this the "static map" mode. The base model is large, and allows for easier interaction. The second mode is a true AR mode. We called it "Vuforia" mode, where, using Vuforia, we recognise a paper map or printed satellite image of a given area, and align the 3D model belonging to that area on top of the 2D map/image. An image of the "static map" mode can be seen in Figure 2.

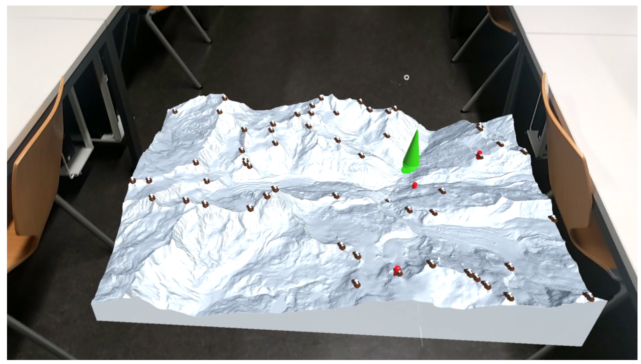

Figure 2. 3D terrain model floating in space ("static map" mode).

On top of the 3D terrain model, we placed both geographic and aeronautical points of interest (POIs). For our application, we limited ourselves to point features. We did not visualise line features (rivers, roads, cables, etc.) or polygons (settlement boundaries, no-fly zones, etc.). With distinct 3D icons, we visualised mountain peaks, aeronautical obstacles, and airports.

The user could interact with the visualised POIs, but not (directly) with the 3D terrain model. Interactions allowed the user to, through a "far interaction", to point at a specific POI, and upon a pinch gesture, additional information on that POI would be displayed. An indirect manipulation of the 3D terrain model was possible in the "Vuforia" mode, by moving and rotating the printout. By moving the printout, the 3D model moves along in space.

By showing the flat palm of one's hand, a menu would open. Within the menu, the user could toggle the POIs on and off. Additionally, the hand menu would allow to switch from the "static map" mode to the "Vuforia" mode. The appearance of the menu, floating over the 3D terrain model, is depicted in Figure 3.

One of the challenges were coordinate systems. Some data was in WGS 84 (latitude/longitude), others were in the Swiss coordinate system LV03. Finally, all coordniates needed to be transformed to Unity coordinates.

Figure 3. The hand menu.

In summary, we used a collection of publicly available data (pre-processed where necessary), together with MRTK and Vuforia, all in Unity, to create our application. The HoloLens recognised hand gestures for interactions and the menu, and can identify the physical base map in "Vuforia" mode with its cameras. All of this flows together in the HoloLens, as visualised in Figure 4.

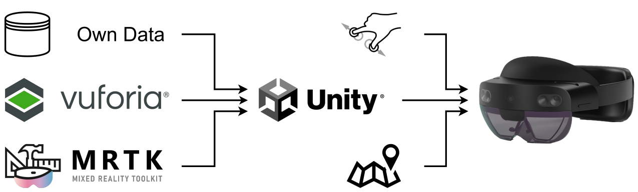

Figure 4. Summary of the flow of data and software within our application.

3. User Study

In this section, we will present first the methodology, and then the results of our user study with $n = 13$ participants. Even though the target audience if private pilots, all our participants were laypeople.

3.1. Method

Our user study was composed of four parts, which we will describe in the next few paragraphs.

First, we collected *demographics* of our participants. This included age and gender, but also familiarity with the Microsoft HoloLens 2, rated on a three-point scale from beginner, through intermediate, to advanced.

Secondly, we asked a number of *application-specific questions*, including both rating questions (on a 10-point scale) and free-text questions. Free text questions were on the positive and negative features of our application. Rating scales were used for the intuitiveness of the user interface and navigation, comfort of gestures and interactions, quality of visualisations, alignment with reality, and responsiveness of our application.

The two last parts of the questionnaires were standard usability questionnaires. The first of the two was *Systen Usability Scale (SUS)* [19], which has been evaluated positively [20], and used in hundreds if not thousands of studies. The second of the two standard questionnaires was the *short version of the User Experience Questionnaire (UEQ)* [21]. We chose the short version over the full UEQ [22] in order to keep the survey short for our participants.

We considered applying the NASA Task Load Index (TLX) [23] as part of our survey too. However, we found several arguments against it: First, our study participants are laypeople, and not experts in the field that our application would be used. Furthermore, our application does not yet allow for a a "meaningful" flight planning task. Lastly, the full NASA TLX, including weighting of the individual dimensions, is a rather lengthy survey, which we wanted to refrain from for our participants. Using only the raw TLX scores is a frequent shortcut to address the length problem. This process, however, is frowned upon by the developers of the TLX [24]. In the end, we decided against using the NASA TLX as part of our user study.

3.2. Results

Of our participants, 6 were male, and 7 were female. Average age was 25.3 years old, with a standard deviation of 3.1 years. Minimum age was 21, and maximum age was 32. On a three-point scale, 9 participants were beginners with the HoloLens 2, 2 had intermediate experience, and 2 were advanced HoloLens 2 users.

The overall experience with our application was rated 5.8 on average on a 10-point scale (higher is better), with a standard deviation of 1.6. The minimum rating was 3, and the highest obtained rating was 8. Figure 5 shows the results of more fine-grained application specific questions. These were rated on a scale from 1 to 5, higher being better.

An overview of the answers given in the free text answers can be seen in Table 1. The most frequently mentioned aspect was a difficulty with pinching the POIs to display additional information about them (5 mentions). With four mentions each, being able to switch between the "static map" and "Vuforia" modes as well as the 3D visualisation were lauded. The interaction with the hand menu was criticised by four participants.

SUS scores range from 50 to 87.5, with a mean of 66.2 and a standard deviation of 12.0.

The overall scores of the short UEQ test range from -0.75 to 2.25, with an average score of 0.83 (standard deviation: 0.86). These results are shown in Figure 6. In the category "pragmatic quality" of the short UEQ, scores range from 0.0 to 2.0, with a mean of 0.87 (standard deviation: 0.57) For the "hedonic quality", the lowest score is -2.25, and

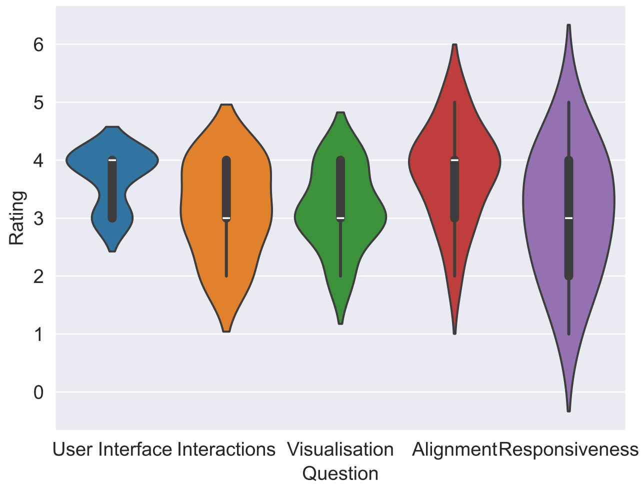

Figure 5. Application specific questions, rated on a scale from 1 to 5. Higher is better. Values < 1 and > 5 are an artefact of the KDE underlying the visualisation.

Table 1. Summary of positive and negatively mentioned features from free text answers. The # columns indicates the number times an element was mentioned. Not all comments of all participants are listed in this table.

Positive	#	Negative	#
Mode switching	4	Pinching/tapping POIs	5
3D visualisation (on a 2D physical map)	4	Hand menu interaction	4
Interactivity	2	Map size	2
Visualisation of POIs	2		

the highest score is 2.5. Mean and standard deviation of "hedonic quality" are 0.79 and 1.38, respectively.

4. Discussion

We conducted a user study with 13 participants, with an (almost) equal number of male and female participants, and participants aged from 21 to 32, so relatively young and "digitally native."

The answers to our custom questions were mostly rated between 3 and 4 (on a scale from 1 to 5), with distributions looking mostly normal (see Figure 5). However, we did not conduct an explicit test for normality. The rating for the question on "responsiveness" was answered very differently by different participants, with answers spreading the full spectrum of possible ratings.

Scores of the SUS range from 50 to almost 90, with the average just shy of 70. Based on [20], 50 corresponds to an "ok" score, and 85 is excellent. A score of 68, according to [25], is an average score, roughly at the 50 % quantile.

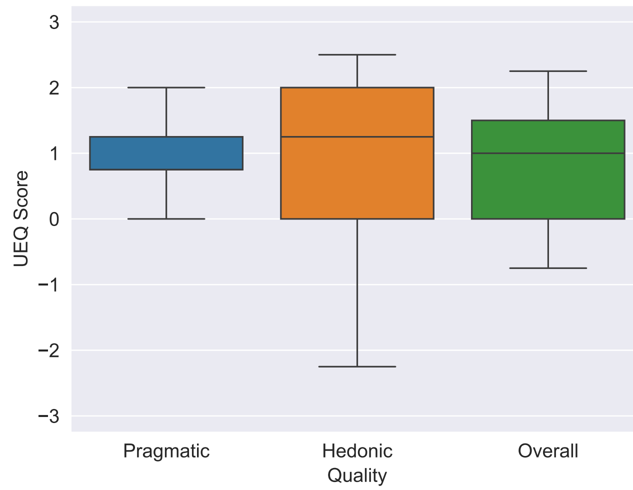

Figure 6. The results of the short UEQ questionnaire.

This corresponds to the result we obtained. Comparing the experience level to the obtained SUS scores, we observe an interesting pattern: Beginners, while contributing the most answers in our study, also gave the widest range of answers. Intermediate level HoloLens users gave the worst ratings, while the experienced users gave above-average ratings. This is visualised in Figure 7.

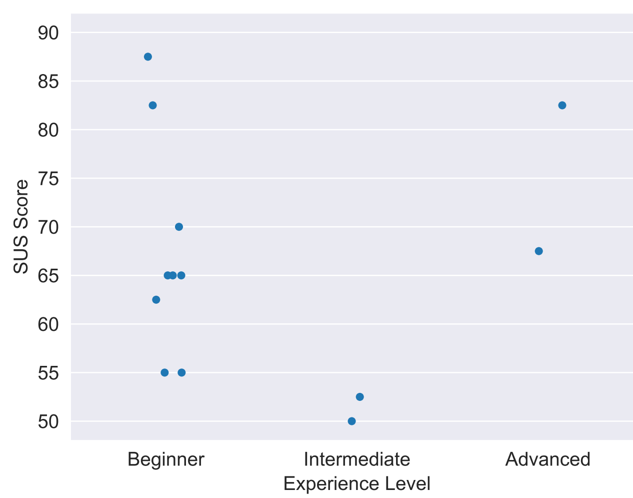

Figure 7. Comparing the experience level with the SUS score.

We obtained a score of around 0.8 for both pragmatic and hedonic quality when conducting the short UEQ. Correspondingly, the overall UEQ score is also around 0.8. This, according to the short UEQ evaluation too (available at `www.ueq-online.org`), is just at the border between a neutral and a positive evaluation. As for the SUS, we analysed wanted to know if there was a correlation between HoloLens experience and the given ratings. In this case, we again hae a wide range or ratings for the beginners. We have slightly below average ratings from the itermediate level users. The most interesting group of respondents are the two “experienced” HoloLens users, which have a spread of 2.5 for the overall UEQ score. The very low score of -2.25 for the hedonic quality is more than two standard deviations away from the mean, so one could consider removing the data of this participants. We however did not remove this participant’s data. The correlation between experience and rating for the UEQ is shown in Figure 8.

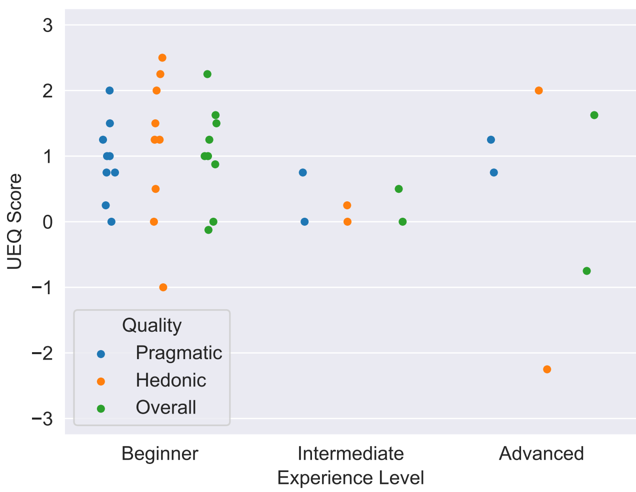

Figure 8. Comparing the experience level with the UEQ scores.

The results of the free text answers were mainly in line with our expectations. The goal of our application was to create a useful tool for pilots. While we tested the application with laypeople, a frequently mentioned benefit was the 3D visualisation on top of a 2D map, which helps to understand spatial relations. Negative aspects mentioned were mainly the challenges of interaction with the POIs (pinching/tapping) and the interaction with the hand menu. This may come from the majority of users not being experienced with the HoloLens interactions. An aspect that, for us surprisingly, was not mentioned, was the lack of colour in the 3D terrain objects. While we don’t know the reasons for this, we see two obvious ways the 3D terrain model could be coloured: Either in the form of a map, or like satellite imagery. The later, in particular, would allow the pilots to in advance see how space beneath and around them looks in real life.

5. Conclusion and Outlook

In our project, we developed an application that can help private pilots with their orientation and navigation. A 3D model of an area can be projected in space, or attached to a physical base map, with aeronautical and geographic POIs attached to the corresponding locations on the 3D model.

The user study returned that our application has an ok to good usability and user experience for both SUS and the short UEQ. A key criticism were the interactions with the POIs and the hand menu, so these could be improved. From our developer point of view, we see the lack of colour on the 3D terrain model, as well as a missing flight path drawing function, as the most important missing features. Naturally, having not only point features, but also having lines and polygons would be desirable. Additional “wishlist” capabilities are a zooming function, to see a smaller or larger area, and adding air space structures [26]. Lastly, having dynamic real-time content would be very interesting. This would include weather in the air (including forecasts), as well the location of aircraft in space, potentially including estimated trajectories.

We perceive two elements on which the user study could be significantly improved: First, not asking laypeople to participate, but actual pilot licence holders. Second, having a task that the participants need to fulfil, which allows for a quantifiable measure of performance, and potentially using the NASA TLX. Using the full UEQ would be a possible extension, to obtain a more accurate UEQ measurement.

All in all, with a steep learning curve for Unity and HoloLens, and limited access to “professional” participants (pilots), we nevertheless got valuable feedback on our application.

References

[1] B. Torrence and J. Dressel, ‘Critical review of extended reality applications in aviation’, in *Virtual, Augmented and Mixed Reality: Applications in Education, Aviation and Industry*, ser. Lecture Notes in Computer Science, J. Y. C. Chen and G. Fragomeni, Eds., vol. 13318, Springer International Publishing, 2022, pp. 270–288, ISBN: 978-3-031-06014-4, DOI: 10.1007/978-3-031-06015-1_19.

[2] A. D. Kaplan, J. Cruit, M. Endsley, S. M. Beers, B. D. Sawyer and P. A. Hancock, ‘The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: A meta-analysis’, *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 63, no. 4, pp. 706–726, 24 Feb. 2020, DOI: 10.1177/0018720820904229.

[3] L. J. Brown, ‘Professional reflection – mixed reality to augment the next generation of aviation professionals’, in *Engaging the Next Generation of Aviation Professionals*. Routledge, 7 Feb. 2020, ch. 2.8, pp. 163–180, ISBN: 9780429287732, DOI: 10.4324/9780429287732-17.

[4] J. Blundell, J. Huddlestone, C. Collins, S. Scott, R. Sears and A. Plioutsias, ‘Envisioning mixed realities on the flight deck’, in *Advances in Intelligent Systems and Computing*, N. Stanton, Ed., Springer International Publishing, 2020, pp. 493–499, ISBN: 978-3-030-50943-9, DOI: 10.1007/978-3-030-50943-9_62.

[5] C. Katins, S. S. Feger and T. Kosch, 'Pilots' considerations regarding current generation mixed reality headset use in general aviation cockpits', in *Proceedings of the 22nd International Conference on Mobile and Ubiquitous Multimedia*, (Vienna, Austria, 3–6 Dec. 2023), ser. MUM '23, ACM, Dec. 2023, pp. 159–165, ISBN: 979-8-4007-0921-0, DOI: 10.1145/3626705.3627785.

[6] C. Katins, S. S. Feger and T. Kosch, 'Exploring mixed reality in general aviation to support pilot workload', in *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, (Hamburg, Germany, 23–28 Apr. 2023), ser. CHI '23, ACM, Apr. 2023, pp. 1–7, ISBN: 978-1-4503-9422-2, DOI: 10.1145/3544549.3585742.

[7] K. G. Kim, J. Krukar, P. Mavros *et al.*, '3D sketch maps: Concept, potential benefits, and challenges', in *15th International Conference on Spatial Information Theory (COSIT 2022)*, vol. 240, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022, ch. 14, 14:1–14:7, DOI: 10.4230/lipics.cosit.2022.14.

[8] QGIS Project, *QGIS desktop user guide/manual (QGIS 3.16)*, Apr. 2022, [Online]. Available: https://docs.qgis.org/3.16/en/docs/user_manual/index.html (visited on 2 Nov. 2023).

[9] Blender Foundation, *Blender 4.0*, comp. software, version 4.0, 14 Nov. 2023, [Online]. Available: https://www.blender.org/download/releases/4-0/ (visited on 4 Jan. 2024).

[10] Unity Technologies, *Unity 2021 LTS*, comp. software, version 2021.3.10fa, 16 Sep. 2022, [Online]. Available: https://unity.com/releases/editor/qa/lts-releases?version=2021.3.

[11] Cesium GS, Inc, *Cesium for Unity*, comp. software, 2023, [Online]. Available: https://cesium.com/platform/cesium-for-unity/ (visited on 5 Nov. 2023).

[12] Mapbox, *Maps SDK for Unity*, comp. software, version 2.1.1, 15 Oct. 2019, [Online]. Available: https://www.mapbox.com/unity.

[13] Mixed Reality Toolkit Organization, *Mixed reality toolkit 3*, comp. software, version 3, 7 Sep. 2023, [Online]. Available: https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk3-overview/ (visited on 4 Jan. 2024).

[14] PTC, Inc., *Vuforia engine*, comp. software, version 10.19.3, 15 Nov. 2023, [Online]. Available: https://developer.vuforia.com/ (visited on 4 Jan. 2024).

[15] Federal Office of Topography (swisstopo), *DHM25*, 2023, [Online]. Available: https://www.swisstopo.admin.ch/en/geodata/height/dhm25.html (visited on 4 Jan. 2024).

[16] Federal Office of Topography (swisstopo), *swissNAMES3D*, 2023, [Online]. Available: https://www.swisstopo.admin.ch/en/geodata/landscape/names3d.html (visited on 2 Nov. 2023).

[17] Federal Office of Civil Aviation (FOCA), *Swiss air navigation obstacle data*, 2011, [Online]. Available: https://opendata.swiss/en/dataset/luftfahrthindernisdaten-schweiz (visited on 2 Nov. 2023).

[18] OurAirports, *Open data downloads*, Oct. 2023, [Online]. Available: https://ourairports.com/data/ (visited on 2 Nov. 2023).

[19] J. Brooke, 'SUS: A 'Quick and dirty' usability scale', in *Usability Evaluation in Industry*, P. W. Jordan, B. Thomas, B. A. Weerdmeester and I. L. McClelland, Eds., Taylor & Francis, 11 Jun. 1996, ch. 21, pp. 189–194, ISBN: 0748403140, DOI: 10.1201/9781498710411-35.

[20] A. Bangor, P. T. Kortum and J. T. Miller, 'An empirical evaluation of the system usability scale', *International Journal of Human-Computer Interaction*, vol. 24, no. 6, pp. 574–594, Jul. 2008, ISSN: 1044-7318, DOI: 10.1080/10447310802205776.

[21] M. Schrepp, A. Hinderks and J. Thomaschewski, 'Design and evaluation of a short version of the user experience questionnaire (UEQ-S)', *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no. 6, pp. 103–108, 11 Sep. 2017, ISSN: 1989-1660, DOI: 10.9781/ijimai.2017.09.001.

[22] B. Laugwitz, T. Held and M. Schrepp, 'Construction and evaluation of a user experience questionnaire', in *HCI and Usability for Education and Work*, ser. Lecture Notes in Computer Science, A. Holzinger, Ed., vol. 5298, Springer Berlin Heidelberg, 2008, pp. 63–76, ISBN: 978-3-540-89349-3, DOI: 10.1007/978-3-540-89350-9_6.

[23] S. G. Hart and L. E. Staveland, 'Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research', in *Human Mental Workload*, ser. Advances in Psychology, P. A. Hancock and N. Meshkati, Eds., vol. 52, Elsevier, 1988, pp. 139–183, ISBN: 978-0-444-70388-0, DOI: 10.1016/s0166-4115(08)62386-9.

[24] S. G. Hart, 'NASA-Task Load Index (NASA-TLX); 20 years later', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 9, pp. 904–908, Oct. 2006, DOI: 10.1177/154193120605000909.

[25] J. Brooke, 'SUS: A retrospective', *Journal of Usability Studies*, vol. 8, no. 2, pp. 29–40, Feb. 2013, [Online]. Available: https://uxpajournal.org/sus-a-retrospective/ (visited on 26 Aug. 2023).

[26] A. Sarbach, T. Weber, K. Henggeler, L. Lutnyk and M. Raubal, 'Evaluating and comparing airspace structure visualisation and perception on digital aeronautical charts', *AGILE: GIScience Series*, vol. 4, pp. 1–13, Jun. 2023, DOI: 10.5194/agile-giss-4-12-2023.