A feasibility study on measuring bound attention during planning of a complex liver surgery

Tim Schneider, Timur Cetin, Stefan Uppenkamp, Dirk Weyhe, Thomas Münder, Anke Reinschüssel, Daniela Salzmann, Verena Uslar

Submitted to: JMIR Formative Research
on: June 10, 2024

Table of Contents

A feasibility study on measuring bound attention during planning of a complex liver surgery

Tim Schneider[1*] MSc; Timur Cetin[1*] Dr rer nat; Stefan Uppenkamp[2] PD, Dr rer nat; Dirk Weyhe[1] Prof Dr Med; Thomas Münder[3] Dr; Anke Reinschüssel[3] Dr; Daniela Salzmann[1] Dr phil; Verena Uslar[1] Dr rer nat

[1]University of Oldenburg, University Hospital for Visceral Surgery, PIUS-Hospital Oldenburg, Oldenburg, Germany Oldenburg DE
[2]University of Oldenburg, Department of Medical Physics and Acoustics, Oldenburg, Germany Oldenburg DE
[3]University of Bremen, Digital Media Lab, Bremen, Germany Bremen DE
[*]these authors contributed equally

Corresponding Author:
Tim Schneider MSc
University of Oldenburg, University Hospital for Visceral Surgery, PIUS-Hospital Oldenburg, Oldenburg, Germany
Ammerländer Heerstraße 114-118
Oldenburg
DE

Abstract

Background: The integration of advanced technologies such as Augmented Reality (AR) and Virtual Reality (VR) in surgical procedures has garnered significant attention. However, the introduction of these innovations necessitates thorough evaluation in the context of human-machine interaction. Despite their potential benefits, new technologies can complicate surgical tasks and increase the cognitive load on surgeons, potentially offsetting the intended advantages. It is crucial to assess these technologies not only for their functional improvements but also for their impact on the surgeon's workload in clinical settings.

Objective: A surgical team nowadays has to increasingly deal with more advanced technologies like AR and VR, aiming at a reduction of surgical trauma and an increase of patient safety. However, each innovation needs to be evaluated in terms of human-machine-interaction. Even if the innovation seems to bring advancement to the field it is applied in, it might make the work more complicated and increase the surgeon’s workload rather than benefitting the surgeon.

Methods: This study aims to establish a method to determine the additional workload generated by using AR or VR glasses objectively in a clinical context for the first time. For this purpose, EEG signals were recorded in a passive auditory oddball paradigm while nine participants had to perform surgical planning of liver resection in three different conditions: using (1) AR glasses, (2) VR glasses, and (3) the conventional planning software on the computer. The electrophysiological results, i.e., the potentials evokes by the auditory stimulus, were compared with the subjectively perceived stress of the participants, as determined by the NASA-TLX questionnaire.

Results: The analysis of the EEG revealed a trend towards a lower amplitude of the N1 component as well as for the P3 component for AR condition at the central electrodes, which suggests a higher workload for the subjects while using AR glasses. In addition, EEG components in the VR condition did not reveal any noticeable differences in comparison to the EEG components of the conventional planning condition.

Conclusions: These results seem to indicate a lower stress level using VR glasses than AR glasses and suggests an advantage through the 3D visualization of the liver model. Furthermore, the fact that the subjectively determined results match the objectively determined results confirms the validity of the applied study design.

(JMIR Preprints 10/06/2024:62740)
DOI: https://doi.org/10.2196/preprints.62740

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.
2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?
✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

A feasibility study on measuring bound attention during planning of a complex liver surgery

Abstract

Background: The integration of advanced technologies such as Augmented Reality (AR) and Virtual Reality (VR) in surgical procedures has garnered significant attention. However, the introduction of these innovations necessitates thorough evaluation in the context of human-machine interaction. Despite their potential benefits, new technologies can complicate surgical tasks and increase the cognitive load on surgeons, potentially offsetting the intended advantages. It is crucial to assess these technologies not only for their functional improvements but also for their impact on the surgeon's workload in clinical settings.
Objective: A surgical team nowadays has to increasingly deal with more advanced technologies like AR and VR, aiming at a reduction of surgical trauma and an increase of patient safety. However, each innovation needs to be evaluated in terms of human-machine-interaction. Even if the innovation seems to bring advancement to the field it is applied in, it might make the work more complicated and increase the surgeon’s workload rather than benefitting the surgeon.
Methods: This study aims to establish a method to determine the additional workload generated by using AR or VR glasses objectively in a clinical context for the first time. For this purpose, EEG signals were recorded in a passive auditory oddball paradigm while nine participants had to perform surgical planning of liver resection in three different conditions: using (1) AR glasses, (2) VR glasses, and (3) the conventional planning software on the computer. The electrophysiological results, i.e., the potentials evoked by the auditory stimulus, were compared with the subjectively perceived stress of the participants, as determined by the NASA-TLX questionnaire.
Results: The analysis of the EEG revealed a trend towards a lower amplitude of the N1 component as well as for the P3 component for AR condition at the central electrodes, which suggests a higher workload for the subjects while using AR glasses. In addition, EEG components in the VR condition did not reveal any noticeable differences in comparison to the EEG components of the conventional planning condition.
Conclusions: These results seem to indicate a lower stress level using VR glasses than AR glasses and suggests an advantage through the 3D visualization of the liver model. Furthermore, the fact that the subjectively determined results match the objectively determined results confirms the validity of the applied study design.

Keywords: Workload measurement; Virtual Reality; Augmented Reality; EEG; ERP; AEP; Oddball-Experiment; NASA-TLX; surgical planning

Introduction

In modern surgery, new surgical techniques are increasingly being used. One example of such novel technologies are minimally invasive procedures that aim to reduce the trauma of surgery and increase patient safety [1]. However, while benefitting the patient, these surgical techniques are significantly more challenging – cognitively and physically – than traditional surgery for the surgical team, especially the surgeon, due to the limited visibility and physical strain. This, on the other hand, may jeopardize patient safety [2].

Another example of novel technologies applied in the field of surgery is the use of Augmented Reality (AR) and Virtual Reality (VR) simulations. These simulations allow 3D models of anatomical structures to be displayed on a 3D display, potentially facilitating examination and analysis. Glasses projecting those simulations may be used prior to a surgery as an aid for planning during the preoperative phase.

In order to use technologies such as AR or VR applications in everyday clinical practice or in the operating room, such technologies must be examined in advance with regards to safety and ergonomics, and especially for workplace strain [3]. So far, AR or VR glasses have generally been assessed for mental workload through subjective questionnaires, such as the NASA-TLX, or through measuring the user's quality of performance [4]. The present study aims to determine the mental workload objectively by using electrophysiological recordings as well as subjectively by using established questionnaires in order to be able to compare them afterwards. The study is not intended to evaluate or rate the different techniques used.

From a neuroergonomic perspective, it is advantageous to replace these subjective questionnaires with objective measurements. Subjective questionnaires rely on self-reporting, which is subject to various biases. Objective measurements, on the other hand, provide more accurate and reliable data by directly assessing neurophysiological and behavioral responses in real-time [5]. By minimizing reliance on subjective questionnaires, objective measurements enhance the scientific validity of neuroergonomic research, leading to more robust findings and practical implications which could lead to optimized human-machine interactions.

AR and VR glasses allow for more accurate planning of invasive procedures through three-dimensional visualization of structural conditions. In this way, the medical team can prepare beforehand for possible anomalies and develop an adequate concept for a complex surgery that consider the patient's individual anatomy in a more accurate way than it is possible with traditional preparation by means of classical radiological images. This is especially useful in complicated interventions with poor direct visibility on site, such as laparoscopic interventions [6]. VR and AR may also be useful tools for inexperienced surgeons to train for interventions and gain more safety.

In multiple studies, the reduction of stress and cognitive burden during the use of VR and AR has been subjectively evaluated [7-9]. The results indicate that the use of VR and AR can reduce stress and cognitive burden for the user. AR and VR applications have been used to clarify anatomical situation and to learn practical tasks, thereby reducing the user's stress by training prior to surgery. Based on these previous studies, we expected that for this study, using AR and VR simulations be less taxing for the participants when preparing a surgery compared to the traditional method of using classical radiological images like MRT, CT, and X-Ray.

Ghani (2020) demonstrated that a passive oddball condition can be used to objectively determine mental workload in the participants EEG [10]. By combining subjective questionnaires and the objective evaluation of workload using EEG, we hypothesize in this study that the objectively determined results on mental workload will match the results from the subjective surveys using the NASA-TLX questionnaire [11].

Bernhardt et al. (2017), Nicolau et al. (2011) and Dixon et al. (2014) describe the significantly improved visualization possibilities through the use of AR and VR, especially in minimally invasive operations. Based on these results, we assumed that the use of AR and VR will make the mental visualization of the anatomical conditions easier than classical 2D-images due to its 3D-display and thereby leading to a better outcome, i.e., better performance in surgery [6, 12, 13].

In order to examine the three hypotheses of this work, a passive oddball experiment with continuous EEG monitoring was conducted with nine visceral surgeons as participants while they performed mock preoperative planning of a resection of a cancerous liver tumor using AR and VR headsets as well as traditional radiological images of an MRI and CT-Scan (resolution: 5 mm; Arterial as well as venous phase were available) on a standard computer screen.

The design of the EEG-task resembles a standard, passive auditory oddball experiment, in which a frequent auditory stimulus as well as a different, rare auditory stimulus are presented. The EEG-response to the target stimulus is analyzed as it is expected to produce two distinct potentials: besides the early, neuronal responses to an auditory stimulus (P1, N1, P2, N2, in the following called Auditory Evoked Potentials/”AEP”) also a later response to the oddball-modality, i.e., the switch from presenting a frequent to presenting a rare stimulus, is expected to manifest as the P3-potential (in the following called Event Related Potential/”ERP”). While the early AEPs are attributed to basic neuronal processing, mainly the activity of working memory, the later ERPs are attributed to a more complex, higher cognitive processing of the presented stimulus [21].

The degree of these potentials is used to draw conclusions about the objective workload in terms of attention required for a primary task: While the auditory stimuli are presented passively in the background, the subjects are focusing actively on a primary task, i.e., in our case, planning a surgery in different modalities. The idea that the EEG-potentials are more profound if the primary task is less demanding, i.e., not requiring so many neuronal resources, has been demonstrated multiple times: if the primary task is not demanding, more mental resources are allocated to the passive background task, eliciting higher EEG-potentials for AEPs as well as for ERPs. If the primary task is more complex, less mental resources are allocated to the secondary, passive background task, thus resulting in less pronounced EEG potentials [10, 21].

The objective workload found in the three different conditions will then be compared with the subjective workload measured through the NASA-TLX questionnaire. In addition, the surgeons are asked to perform simple tasks during the planning of the surgery, i.e., to measure the diameter of lesions, in order to assess the quality of their work.

Methods

Participants

Nine participants with prior experience in liver surgery were recruited for this study. It is worth noting that despite the relatively small sample size, all participants were carefully selected from a highly defined target group (surgeons with prior experience in liver surgery), which is a hallmark of the study's quality. Inclusion criteria were attending status (alternatively residents from the 4th year of training), the signed informed consent, inconspicuous coffee consumption, and no previous neurological damage. Exclusion criteria for participating subjects were overly pronounced caffeine consumption or consumption of psychoactive drugs/medication as well as mental and/or neurological disorders. Exclusion criteria were checked in advance through a standard questionnaire for EEG measurements.
All participants were male, right-handed individuals with an average age of 36.8 (± 5.59). Six of the participating surgeons were attending (two of them senior surgeons) and three were residents in the later stage of their training. On average, prior to the study, the participants had performed 43 liver surgeries as an assisting surgeon (± 65.63) and 16 (± 29.7) as a leading surgeon.
All participants were recruited from the medical staff of the University of Oldenburg’s department of

visceral surgery, located at PIUS-Hospital, Oldenburg. All participants gave their informed consent prior to the study. Conducting the study was approved by both, the medical ethical committee (reference number: 2020-150) of the University of Oldenburg as well as the staff council at the PIUS-Hospital.

Stimuli

The acoustic oddball stimuli in the background consisted of two sinusoidal tone pips with a duration of 50 ms and a rise/fall time of 5 ms each. The standard tone at frequency of 500 Hz had a probability of 88 % and the rarer target tone at a frequency of 1 kHz correspondingly a presentation probability of 12 %. Both tones were presented via loudspeakers (iLoud MTM, IK Multimedia, Modena, Italy connected to an amplifier [the t.amp E4-130, Thomann, Burgebrach-Treppendorf, Germany]) at a volume of 75 dB SPL with an inter-stimulus interval of one second. The tones were randomized with the constraint that any target tone was followed by a standard tone. The stimuli used were similar to the stimuli of Williams et al. (2005) [14]. The auditory stimuli were digitally created using MATLAB version R2021a (The MathWorks Inc., Natick, USA) and presented with the Presentation software (Version 23.1, Neurobehavioral Systems, Berkeley, USA).

Conditions

Three different conditions were tested for their impact on focused attention during the planning of liver surgery. The three conditions were: 1) AR headset, 2) VR headset, 3) PC - traditional radiological images. The VR setup consisted of the VIVE Pro headset (HTC, Taoyouan, Taiwan) including one controller and two tracking devices – so called lighthouses. For AR the Microsoft HoloLens 2 was used (Mircosoft, Redmond, USA). The software used for visualization in VR was a custom-built setup used previously in our department [15]. For AR setup a demonstrator was used which was developed as part of this project. For display of the traditional radiological images, the Xero Viewer (Agfa healthcare, Mortsel, Belgium) was used on a conventional PC. Screenshots and videos of the three conditions can be viewed in the supplements. In the AR and VR condition, the participants were presented with previously segmented 3D models in combination with the radiological images. In the PC condition, the participants were only presented with the radiological images. In all three conditions, the subjects were instructed to perform the planning of a liver resection using all the available functions in the respective conditions. The participants in both conditions were able to draw in the 3D model, set markers, and manually measure anatomical structures. In the VR condition, additional functionalities were available, such as volume cropping and placement of vascular clips. In both AR and VR conditions, participants were also capable of viewing or even superimposing the corresponding MRI images onto the 3D model. The control of the AR and VR conditions differed due to distinct hardware setups. The VR condition was controlled using dedicated controllers, whereas the AR condition was controlled through gesture-based interaction. In all three conditions, the participants were asked about the size and segment of the largest tumor, as well as the resectability of this tumor. Note that the goal was just to keep the participants engaged for at least 3 minutes in each condition, with one basic task (i.e., a diagnosis based on the images), which was the same for all the conditions, regardless of the respective handling or available functionalities. Evaluating the usability of the respective software and its functionalities was beyond the scope of this study, especially since both AR and VR set-ups were custom-build and under development.

Experimental setup

A mobile EEG amplifier (Smarting mobile, mbt, Belgrade, Serbia) on a standard 10/20 EEG cap (Easycap, Woerthsee-Etterschlag, Germany) was connected to a recording software (smarting streamer V 3.4.2, mbt, Belgrade, Serbia) on a standard desktop computer running the Windows 10 operating system (Mircosoft, Redmond, USA). The two speakers were placed next to the recording computer. The subjects were placed two meters in front of the speakers. During the experiment, care was taken to place the subjects in the same marked position in all three conditions to ensure that the stimuli were presented to the subjects at the same sound level during all three conditions. A laptop was used to connect via miracast to the AR glasses to ensure that the instructor could follow what the participants were doing and seeing. It was placed in a position where it could not be seen by the participants to avoid distracting them. A second desktop computer was used to present the VR setup with the VR headset connected to this computer and the two VR lighthouses mounted on tripods to allow for movement of the subjects while wearing the VR headset. To ensure that the subjects could also remain in the same position during all three conditions, the traditional radiological images which were presented during the PC- condition were presented on a computer connected to the hospital information system that was placed on a mobile and height-adjustable table. The screen was positioned at eye level to prevent the EEG cap from slipping due to excessive head movements and to make sure that all conditions could be performed by the participants while standing (Fig. 1). Controllers were used to control the VR condition and were placed in the hands of the participants. For the AR condition, gesture control through the participants' hands was used for control. In advance, no standardized training was conducted with the participants. However, if the conditions were unknown to the participants or if the participants had any questions, they were extensively explained during a test run in which the participants could try out the controls and functions until they reported feeling confident in handling the technology.

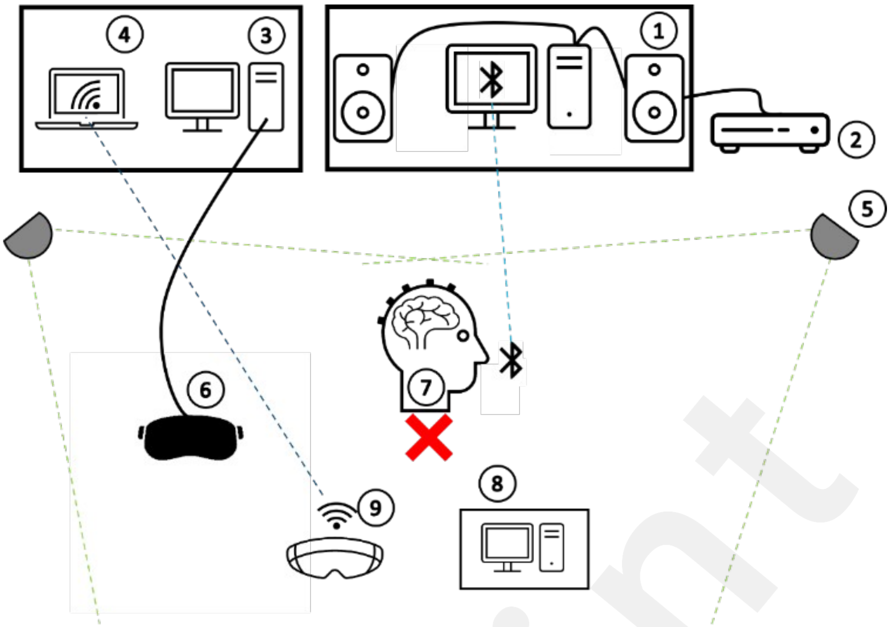

Figure 1: Experimental setup. 1: Main computer for recording EEG. 2: Amplifier. 3: Computer VR. 4: Computer AR. 5: Lighthouses for the VR headset. 6: VR headset. 7: EEG on marked position. 8: Computer for conventional surgical planning. 9: AR headset.

Procedure

Mobile EEG recordings [sampling rate 250 Hz, smarting mobi (mbt, Belgrade, Serbia)] with 24 electrodes (Fp1, Fp2, AFz, F7, F3, Fz, F4, F8, T7, C3, Cz, C4, T8, CPz, M1, M2, P7, P3, Pz, P4, P8, POz, O1, O2) were obtained with a standard EasyCap EEG recording cap (Easycap, Herrsching, Germany) with the designated reference located at a fronto-central position. Continuous EEG data was stored in .xdf format before processed for analysis. For AEPs, EEG were analyzed using EEGlab (V2020.0, open-source toolbox for Matlab R2020a MathWorks, Natick, Massachusetts, United States) on a standard PC running the Windows 10 operating system (Microsoft, Redmond, Washington, United States), while for ERP analysis, EEG was converted to Brain Vision Analyzer format using Matlab before ERP analysis was done using the Brain Vision Analyzer Software (Version 2.2.1.8266, Professional Edition, Brain Products, Gilching, Germany). EEG electrode impedances were adjusted below 10 kΩ using 70 % rubbing alcohol and abrasive conductive gel (Easycap Abalyte Hi-Cl, Woerthsee-Etterschlag, Germany) according to manufactures recommended protocol. After electrodes were properly placed, two minutes of EEG in resting state with eyes open/ eyes closed were recorded before actual testing was conducted. For this, one of the three conditions "AR", "VR" and "radiological images" was selected in randomized order. In each of the three conditions the participants had to plan two different surgeries with two different patient cases. The two patient cases were presented in randomized order as well. At the end of the session, the participants had performed six sets of surgical planning (two cases for each of the three conditions). The task for the participants was to perform the surgical planning for a procedure to remove cancerous tissue from a liver. The processing time for each condition (AR, VR, radiology) was ten minutes (five for each patient case). During the whole task, the auditory stimuli were presented to the subjects via the two speakers in the background. Before changing the condition, the subjective workload of the participants was determined using the NASA-TLX questionnaire. The NASA-TLX workload questionnaire has been used as a valid measure of subjective workload in a medical work environment [16].

Data preprocessing of EEG-data

To retain the frequency components that are relevant for analyzing the EEG, the recordings were filtered with a bandpass filter (low cutoff 0.05 Hz, high cutoff 30 Hz). Continuous EEG data was divided in 1000 ms epochs (200 ms prior to 800 ms post stimulus). Data was categorized into one of two categories: standard stimulus (non-target, 0.5 kHz) and target stimulus (1 kHz). Manual artefact rejection was done for epochs showing fluctuations of more than ±100 mV and/or artefacts from eye blinks or eye movement. Averages were calculated for the epochs that were recorded during presentation of the target stimulus under each of the three conditions, i.e., one average was calculated for EEG response to the target stimulus during AR, one for the EEG response during VR and one for the EEG response during surgical planning using traditional radiological images. The averages were calculated for each participant for further statistical analyses. Baseline correction was conducted to prevent possible linear drifts in the recorded curves and make the grand averages across the participants of different conditions comparable. Data was re-referenced to electrodes T5 and T6 for

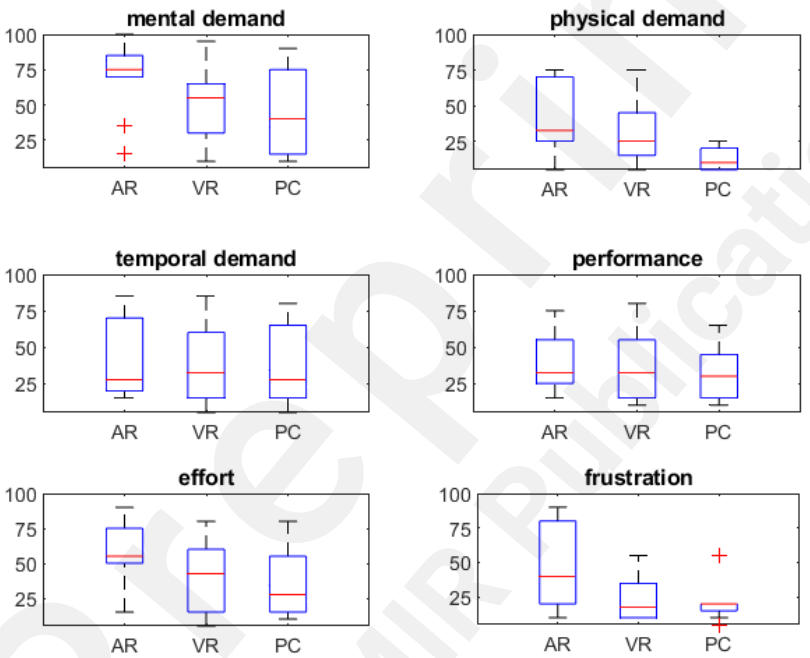

Figure 2: NASA-TLX Scores: AR = surgery planning using the Augmented Reality; VR = surgery planning using the Virtual Reality; PC = surgery planning using traditional radiological images on a desktop computer.

AEP analysis but left referenced to physical fronto-central reference for ERP analysis.

Results

Subjective Workload

The median physical demand score for the AR condition was 75 on the NASA-TLX scale, which is higher than the scores for the VR (NASA-TLX score 55) and PC (NASA-TLX score 25) conditions. When considering the medians for the perceived performance and temporal demand categories, the different conditions each yield the same NASA-TLX scores of 25 for temporal demand and 30 for perceived performance. Additionally, for the mental demand, effort, and frustration categories, the median score for the AR condition represents the highest value, with a NASA-TLX score of 30 for mental demand, 55 for effort, and 40 for the reported frustration of the participants (Fig. 2).

When examining the box plots, it is evident that the median for the assessment of mental demands of the AR headset (median NASA-TLX-Score: 75) was higher compared to both the VR (median NASA-TLX-Score: 55) headset and the PC (median NASA-TLX-Score: 40). Similarly, higher medians were observed for effort and frustration in the evaluation of NASA-TLX-Scores when using the AR (median NASA-TLX-Score: 55/40) headset compared to the VR (median NASA-TLX-Score: 43/18) headset or the PC (median NASA-TLX-Score: 28/20). In terms of physical demand, the median was lower in the PC (median NASA-TLX-Score: 10) condition compared to the AR (median NASA-TLX-Score: 33) or VR (median NASA-TLX-Score: 25) headset. Regarding the medians of NASA-TLX-Scores for temporal demands and performance.
A Friedman test was performed to test for significance as the data did not follow a normal distribution according to the Kolmogorov-Smirnov test. To proceed the correction for multiple testing with the Benjamini-Hochberg method using a web-based calculator [17]. Assuming a 95 % confidence interval, no significances were found.

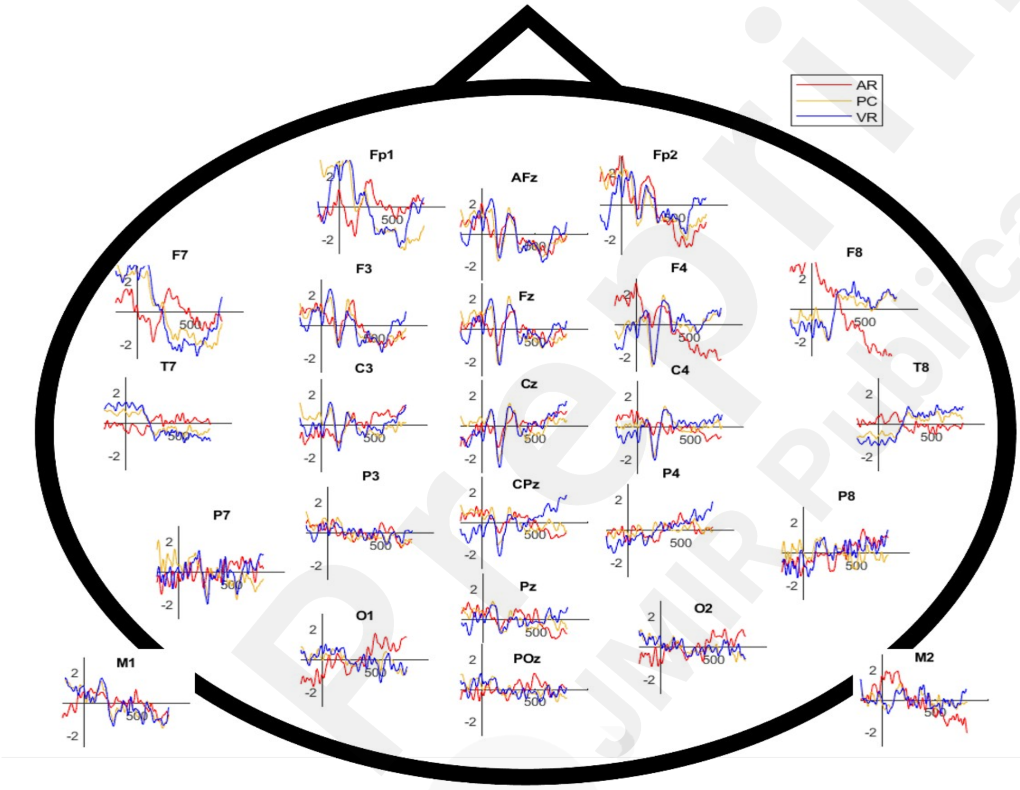

Figure 3: Event-related potentials as response to the target tone for all electrodes and all three conditions: AR = surgery planning using the augmented reality-setup; VR = surgery planning using the virtual reality setup; PC = surgery planning using traditional radiological images on a desktop computer.

Electrophysiological recordings – early components

AEPs were seen mainly on the electrodes of the central row and the midline (AFz to Pz and F3/C3 as well as F4/P4). The Grand Averages were calculated for all electrodes (Fig. 3).

In order to statistically analyze the amplitudes of each participant, in the following the peaks and their corresponding latencies were determined for all conditions for the central electrodes Fz, Cz, and Pz (Fig. 4. (b) - (d)). To statistically verify the differences in amplitude of the grand averages, the work of Handy (2005) was used as a guide [18]. The intervals for P1, N1, and P2 were marked accordingly (Fig. 4.(a)). The resulting peaks and latencies of each participant were then subjected to an rmANOVA (SPSS version 28.0.1.0, IBM, Armonk, USA) with the factors of patient case,

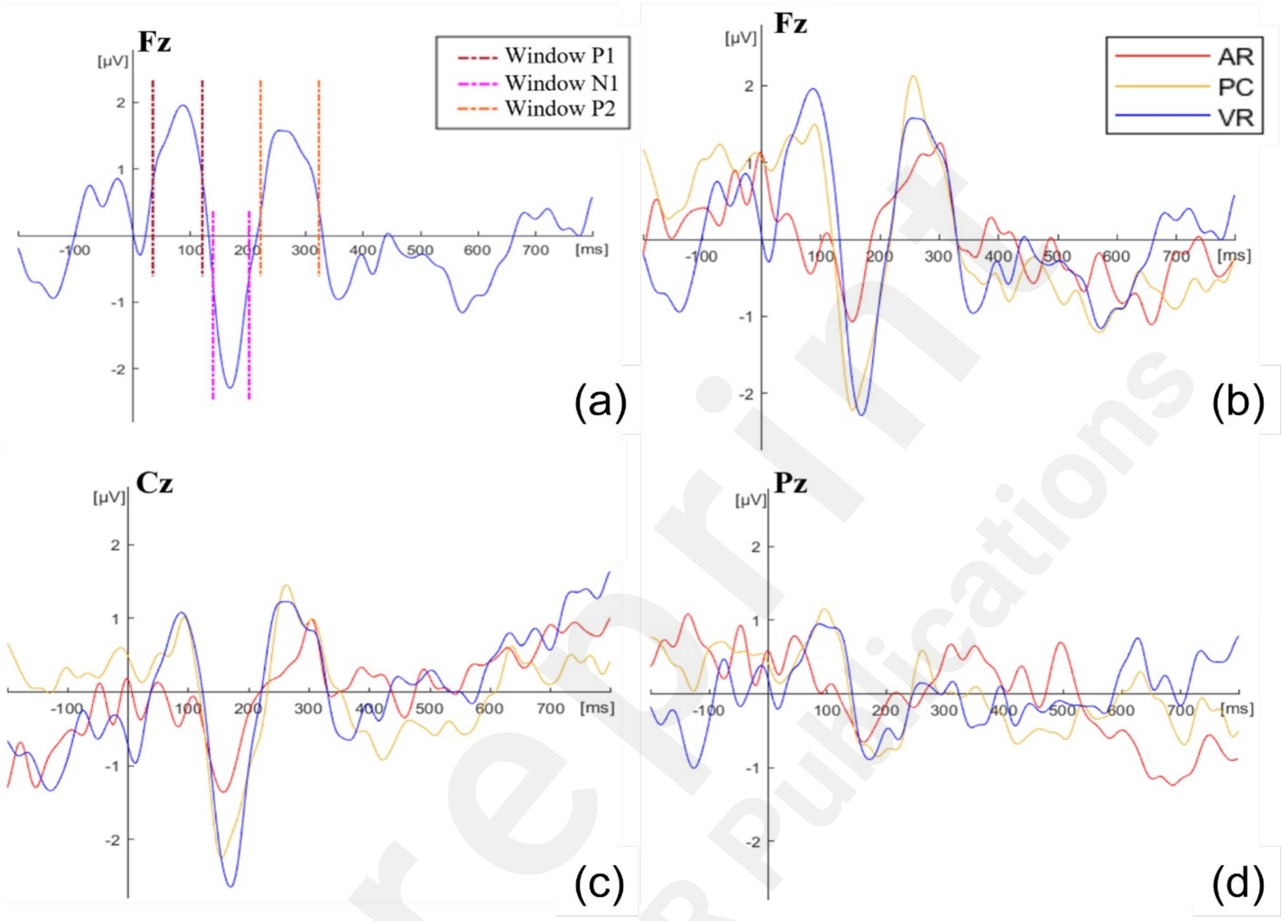

Figure 4: Grand-Averages of Event-related potentials as response to the target tone for statistically analysed electrodes under all three conditions: AR = surgery planning using the augmented reality-setup; VR = surgery planning using the virtual reality setup; PC = surgery planning using traditional radiological images on a desktop computer. (a): Temporal windows for the event-related components P1, N1 and P2, exemplarily shown for electrode position Fz; (b): Grand-Averages of electrode Fz; (c): Grand-Averages of electrode Cz; (d): Grand-Averages of electrode Pz.

condition, electrode, and ERP components. A significant difference was found between the ERP components ($F(1.718; 13.743) = 46.754$; $p < 0.001$).

The tendencies observed in the grand averages were not confirmed by the statistical analysis. No significant differences were found between patient cases, conditions, and electrodes. To determine the latencies of the AEP components (P1, N1, and P2) that occurred, the approach of Handy (2005) was also used [18]. The latency of the identified peaks within defined intervals was determined and noted for each participant. Subsequently, an rmANOVA was performed with the factors of patient case, condition, electrode, and ERP components to identify significant differences in the temporal occurrence of the components.

The tests of within-subject effects yielded significance for the interaction of condition, electrode, and ERP components ($F(3.401; 27.211) = 6.747$; $p = 0.001$). The results of the post-hoc tests (pairwise comparisons) are shown in table 1: According to the pairwise comparisons, the P1 component at the

Fz electrode is significantly earlier (75.3 ms ± 25.8) in the VR condition than when using traditional radiological images (99.4 ms ± 28.6). The Cz electrode also shows a significant difference in the P1 component between the VR condition (111.1 ms ± 9.9) and the AR condition (80.4 ms ± 32.1). The P2 component in the AR condition (290.6 ms ± 23.5) occurs significantly later than when using traditional radiological images (265.3 ms ± 13.1). Significant differences were found for the P1 component at the Pz electrode, with the VR condition (62.8 ms ± 24.8) occurring significantly earlier than the AR condition (89.1 ms ± 14.6) and when using traditional radiological images (92.8 ms ± 28.4). The N2 component did not show any significant interactions.

Table 1: Significant results of the post-hoc tests (pairwise comparisons): AR = surgery planning using the Augmented Reality-setup; VR = surgery planning using the Virtual Reality setup; PC = surgery planning using traditional radiological images on a desktop computer.

Elec.	ERP comp.	Cond. (I)	Cond. (J)	Means of latency [ms] I	Means of latency [ms] J	Differences Between means (I-J)	Sig.
Fz	P1	VR	**PC**	**75.3**	**99.4**	**-24.1**	**0.013**
Cz	P1	VR	**AR**	**111.11**	**80.4**	**30.71**	**0.036**
		AR	**PC**	**290.7**	**265.3**	**25.4**	**0.027**
Pz	P1	VR	**AR**	**62.9**	**89.1**	**-26.2**	**0.040**
			PC	**62.9**	**92.9**	**-30**	**0.041**

Electrophysiological recordings – P3

For the P3 component of the ERP, the peak amplitudes in the time range of interest (300 – 400 ms post stimulus) were calculated. Using all electrodes, a grand average for all three conditions (radiology, VR, AR) was calculated (Fig. 5). A clear trend was seen with the classical radiological approach eliciting the lowest amplitude of P3 (2,45µV ±2,26 µV), VR showing the second largest P3 with 3,08µV (±2,80 µV) and the AR condition provoking the largest P3 with 6,20 µV (±6,84 µV). Standard deviations are high and no statistical significance was found using a rmANOVA (SPSS version 28.0.1.0, IBM, Armonk, USA) with the factors of patient case, condition and P3 amplitude (F(1.94;10,58)=1.469; p=0,262).

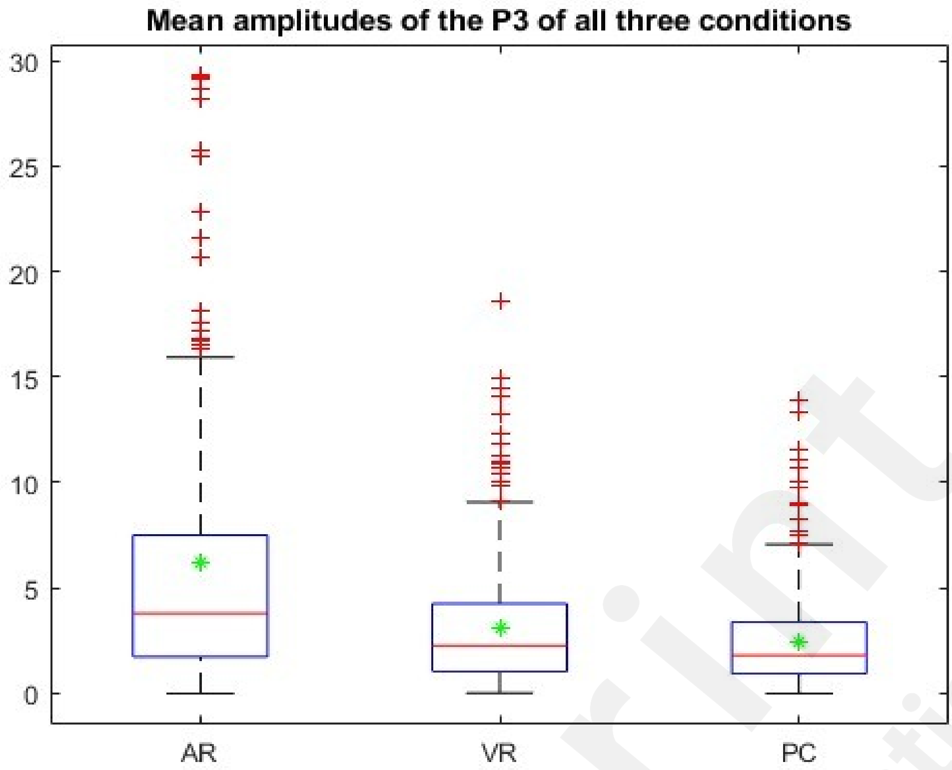

Figure 5: Amplitude of the P3-component of the ERP in all three condition from AR approach (left) with the highest P3-Amplitude 6,20 µV (±6,84 µV), the VR-condition (center) showing an amplitude of 3,08µV (±2,80 µV) and the classical approach (right) showing the lowest P3-amplitude with average of 2,45µV(±2,26 µV). There were no statistical significances found (F(1.94;10,58)=1.469; p=0,262).

Post-hoc Power analysis

Due to the small sample size of N=9, a post-hoc power analysis was conducted to determine the effect size and assess whether a larger sample could yield different results. To ascertain the effect size, partial eta-squared (η^2=0.114) was calculated for the interaction between different experimental conditions using SPSS version 28.0.1.0 (IBM, Armonk, USA). In order to identify the sample size at which significant results might occur, a web-based calculator was employed [17], yielding a required sample size of 101 participants for the determined η^2, a statistical power of 0.8, and a significance level (α = .05).

Discussion

The aim of this was to establish a method to determine the additional workload generated by using AR or VR glasses objectively in a clinical context for the first time. For this purpose, two runs of planning a liver surgery were performed with each of the three conditions. In the background, a passive oddball experiment was conducted, which was supposed to lead to clear early AEP components (P1, N1, and P2) as well as the P3 component of ERP in the EEG recordings. The work of Ghani et al. (2020) was used as a guide [10].

Subjective Workload

NASA-TLX scores show a tendency to higher scores, i.e. a higher subjective workload, for AR planning especially in terms of “mental demands” as well as “frustration”. This does not reach a level of statistical significance if compared to VR or traditional approach using radiological images, which may be due to the sample size being too small. The observed tendencies can may possibly be attributed to the more complex and unfamiliar gesture controls of the AR headset. In addition, a poor Internet connection may have resulted in higher latencies, which could have potentially affected gesture control and thus tended to result in higher workloads when using the AR demonstrator.
Higher physical demands in the AR and VR setup compared to the traditional approach was expected. and a tendency could be observed in the subjective ratings as the stationary use of a mouse on a desktop PC required comparably less physical effort than using an AR or VR headset.

Objective Workload

The expectation that using an AR or VR setup is generally easier for a user when planning surgery due to its 3D projection of the anatomical situation was not confirmed.

EEG data indicates a trend with more pronounced early AEPs when using VR and classical radiological images compared to AR, thus hinting at the use of AR resulting in the highest workload for the surgeon while using VR does not increase the workload when compared to classical radiological images. Differences are especially visible in the ERPs P1 and N1 component: While no P1 component was found for the AR condition, the N1 component showed a significant difference for the AR condition compared to radiological images as well as using of the VR setup on the level of grand averages. This goes in line with the finding of Ghani et al. (2020) that indicate that the N1 component becomes less pronounced with the increased difficulty of the task which was given due to the more complicated gesture control of the AR-headset and the little opportunity to familiarize with the gestures needed to handle the 3D objects in AR [10].

We would like to emphasize that the significances at the electrode Pz should be treated with caution, as the amplitudes are generally lower here. The results of the electrode Pz are clearly noisier, which may be due to the position of the electrode, as this may be obstructed by the fastening of the glasses, which are attached to the head near the position of electrode Pz. These results should therefore be treated with caution.

When looking at the significances of the latencies of the different early AEP components, it is noticeable that there are occasional significances that do not follow a comprehensible structure. Additionally, the significances at the electrode Pz should be treated with caution, as the amplitudes are generally lower here. The results of the electrode Pz are clearly noisier, which may be due to the position of the electrode, as this may be obstructed by the fastening of the glasses, which are attached to the head near the position of electrode Pz.

Concerning the event-related P3 wave in response to the target stimulus of the passive oddball paradigm, the pattern found is different from that seen in the early potentials: a trend from a low P3 in the radiology condition towards a high P3 in the AR condition can be seen. The studies conducted by Chao et al. (2017), Küçük et al. (2016), and Thees et al. (2020), asserting that AR and VR headsets result in reduced cognitive load, could thus only be partly substantiated in the present investigation [7-9].

Based on our data, we conclude that the use of the passive oddball paradigm is a suitable tool to

measure objective workload between the radiology, AR and VR-setups in planning a complex surgery and there seem to be measurable differences in neuronal processing of 2D screen images when compared to VR and AR images. However, the answer to the question why the different modalities of EEG components – early AEP vs. ERP – indicated different findings seems complex and rooted in different, complex neuronal and psychological processes responsible for producing these potentials. For a better understanding of the different psychological processes behind the processing of 2D screen images vs. VR and AR images, besides from an increased no. of subjects, a more in-depth analysis of the P3 potential is needed. In our analysis we only looked at the grand average of the peak amplitude. However, also the location of the most pronounced P3 potentials, as well as the dynamics, i.e. the distribution of P3 peak latencies over the different electrode localizations and the change thereof over time, needs to be investigated. Current research shows that the P3 potential originates as a combination of different potentials, P3a and P3b, that might be involved in different neuropsychological processes. This also needs to be addressed in further studies.
A simple, but feasible explanation for the different findings between early AEPs and P3-ERPs might be the different mental processes they are attributed too: while the complex, cognitive processing of AR pictures – represented in the P3-potential – might be less demanding for the user and thus be in line with the findings of Chao et al. (2017), Küçük et al. (2016), and Thees et al. (2020), the difficult or less familiar navigation in the AR and – to a lesser extent – VR environment may lead to a higher demand in basic neuronal tasks, i.e. visuo-motor-coordination, that may be represented more by the early neurological potentials then by the late cognitive potentials. While this is speculation at this point, we believe that our study should be seen as a proof-of-principle for further studies with more subjects in a more complex setting, addressing this issue.

Concerning the results of the subjective and the objective workload measurements, one can say that the use of a passive oddball paradigm is a suitable tool to objectively measure the workload for medical personnel when evaluating the workload of VR and AR in a simulation context. This conclusion may be drawn as the results of the objective, electrophysiological measurements go in line with the results of the subjective questionnaire (NASA-TLX) for perceived workload. Thus, the work of Ghani et al. (2020) could be confirmed, and it can be assumed that the developed study design is suitable to evaluate mental workload in everyday clinical practice in an objective way [10]. This new workload measurement provides the advantage over subjective questionnaires that the results are much more comparable and subjective parameters such as personal preferences do not affect the measurement of mental workload.

The use of a mobile EEG also proved to be very useful in this study. Due to the high degree of mobility provided by the possibility of using portable smartphones as recording devices, it enables recordings to be made in situations close to the operation without restricting the participants freedom of action. The validity of corresponding mobile EEGs has already been described by Wascher et al. (2021). Furthermore, the work of El Basbasse et al. (2023), among others, has already confirmed that reliable data can be derived using mobile EEG recordings, including the implementation of the easyCap system, in conjunction with VR headsets [19, 20]. The results of this study support this work. Moreover, our results demonstrate that the study design employed in conjunction with mobile EEG technology enables the investigation of additional workload induced by various medical and clinical devices. Notably, even head-worn devices cannot be unequivocally excluded from this measurement approach.

The AR setup was perceived and measured as the most difficult of the three setups. As mentioned, his may be due to the complexity of its controls. While the classical approach was well-known to the participants, controlling the VR setup seems to be more intuitive than using the AR setup. However, it is crucial to consider that AR headsets represent a relatively nascent technology, and gesture recognition is still in its developmental stages, with anticipated advancements in future iterations of AR headsets. It seemed that the participants were overwhelmed with handling without a controller or a corresponding control element, as gesture control was a new experience for most of the participants. This difficulty may be overcome by a more accurate training of the users prior to working with the AR setup. We believe that our work might be used as a tool to see how well the users are trained in using the AR setup and may be used to investigate generally the need for training when navigating in an AR and VR environment even besides its uses in a medical setting. Additionally, more complex, subconscious explanations might also be considered: The VR approach might benefit from the possibility of exclusively displaying the anatomical model while the AR setup might strain the participants mental capabilities more due to the simultaneous display of the computer-generated 3D model as well as the real-world environment. In addition, aspects such as the different displays of the HoloLens and the VIVE Pro could contribute to differences in the participants load. However, this requires more detailed investigations and it must be considered in this discussion that the two-dimensional viewing habits that have been trained over decades are completely abandoned. New three-dimensional viewing habits have to be re-learned in AR and VR, as well as voice and gesture control.

This study concerned only the workload of the three conditions. It should be made clear that there are situations in which a higher workload is acceptable when seen with the benefits that the approach would yield. While the VR setup might be a good way of planning a surgery due to the lack of visibility into the real environment, the AR setup might be the more suitable approach during surgery, i.e., to project radiological images as 3D models directly into the patients' situs during a procedure to localize, for example tumors or blood vessels. In these situations, however, technologies such as AR glasses can offer significant advantages such its see-through capabilities, so that a slightly higher workload, as observed in this study, can be accepted if parameters such as patient safety are increased through the use of such technologies.

Interpretation of our results should be conducted with caution as the number of only nine participants is very low, thus resulting in limited statistical power. Our study should be seen as a feasibility study on the possibility of how to measure objective workload while using augmented and/or virtual reality in a surgical simulation setting. However, we believe the study still shows valuable data as experienced surgeons in a clinical setting used real-life patients' cases with state-of-the-art equipment for augmented- and virtual reality presentation.

To give an outlook beyond the scope of this study, AR and VR can be very useful technologies in the clinical context in our opinion, when used in an appropriate task, i.e., VR could be more appropriate in a teaching setting or for preoperative planning, whereas AR has immense potential beside preoperative planning, in using at the operating room due to the open view on the real environment and its sterility for example to virtual overlay anatomical structures on a patient's body during surgery. Whether the great benefit of such technologies outweighs the possible additional workload of the surgeons should thus be investigated in further studies.
Furthermore, investigating the impact of training, the continuous development of software and hardware, and the increasing familiarity of the younger generation of physicians with these new technologies on measures of workload would be of great interest. Such research could provide valuable insights into the potential benefits and challenges posed by integrating advanced technologies like AR and VR in medical practice. Further exploration in this area may also facilitate the design of tailored training programs and technological interventions to optimize the utilization of these innovations and improve overall healthcare outcomes.

Conclusion

We conclude that the use of a passive oddball paradigm is a feasible approach to measure the workload during surgical planning and thereby probably also during other medical applications when evaluating the strain, the use of AR or VR technology puts on the users' cognitive capabilities. The evaluated study design, using an oddball paradigm, proved to be a good opportunity to determine workplace strain objectively. In future studies, it may replace subjective questionnaires and enhance the significance of the measured workplace strain achieved by the objectivity provided.

Furthermore, we consider that AR or VR approaches to displaying radiological images are not inherently better than the classical approach and that especially the use of AR technology before the first use seems to require a certain amount of training, possibly due to the novelty of gesture control. However, we are confident, that with future developments in AR technology and gesture recognition, those issues will become irrelevant in the future. In any case, the differences in workload between AR and VR highlight the need for intuitive user interaction and ease of use for any new technology, especially when applied in a medical setting.

In conclusion, this paper demonstrates the feasibility of utilizing a mobile EEG setup in conjunction with a passive Oddball paradigm to assess the potential additional cognitive load imposed by emerging technologies such as AR and VR headsets in a clinical setting.

Acknowledgements

We would like to express our gratitude to Kilian Gloy for his contributions to the planning of the conducted study. His expertise and dedication have played a significant role in shaping the research design and ensuring its successful implementation. Our thanks also go to Fraunhofer MEVIS, Bremen, namely Andrea Schenk and Christian Schumann for providing us with segmented liver models used in this study. This study was conducted within the scope of the VIVATOP project and has thus received funding from the German Ministry of Education and Research (BMBF) in the grant program "Gesundes Leben" (healthy living), grant number 16SV8079.

Conflicts of Interest

No conflicts of interest exist except for co-author Dr. Daniela Salzmann, who was employed in the involved research group at the University Clinic for Visceral Surgery at the University of Oldenburg during the project, and subsequently transitioned to apoQlar (Hamburg, Germany), a company also involved in the project for the development of the AR demonstrator.

References

1. Dick J, Gebhard F. Zukunft der intraoperativen Bildgebung und Navigation. In: OP-JOURNAL; 2015; 30th edition (03), p. 188–195. DOI: 10.1055/s-0034-1383340.
2. Bogdanova R, Boulanger P, Zheng B. Depth Perception of Surgeons in Minimally Invasive Surgery. In: Surgical innovation; 2016; 23th edition (5), p. 515–524. DOI:

10.1177/1553350616639141.
3. Kramme R (Hg.). Verfahren - Systeme - Informations-verarbeitung. In: medizintechnik; 4th edition. Berlin, Heidelberg: Springer Berlin Heidelberg. 2011.
4. Funk J, Schmidt L. Evaluation of an Augmented Reality In-struction for a Complex Assembly Task. In: i-com; 2021; 20th edition (1), p. 63–72. DOI: 10.1515/icom-2021-0005.
5. Rosanne O, Albuquerque I, Cassani R, Gagnon J-F, Tremblay S and Falk TH. Adaptive Filtering for Improved EEG-Based Mental Workload Assessment of Ambulant Users. In: Front. Neurosci.; 2021; 15th edition; DOI: 10.3389/fnins.2021.611962
6. Bernhardt S; Nicolau S, Soler L, Doignon C. The status of augmented reality in laparoscopic surgery as of 2016. In: Medical image analysis; 37th edition; 2017. p. 66–90. DOI: 10.1016/j.media.2017.01.007.
7. Thees M, Kapp S, Strzys MP, Beil F, Lukowicz P, Kuhn J. Effects of augmented reality on learning and cognitive load in university physics laboratory courses. In: Computers in Human Behavior; 2020; 108th edition, p. 106316. DOI: 10.1016/j.chb.2020.106316.
8. Küçük S, Kapakin S, Göktaş Y. Learning anatomy via mobile augmented reality: Effects on achievement and cognitive load. In: Anatomical sciences education; 2016; 9th edition; (5), p. 411–421. DOI: 10.1002/ase.1603.
9. Chao CJ, Wu SY, Yau YJ, Feng WY, Tseng FY. Effects of three-dimensional virtual reality and traditional training methods on mental workload and training performance. In: Hum. Factors Man; 2017; 27th edition (4), p. 187–196. DOI: 10.1002/hfm.20702.
10. Ghani U, Signal N, Niazi IK, Taylor D. A novel ap-proach to validate the efficacy of single task ERP paradigms to measure cogni-tive workload. In: International journal of psychophysiology: official journal of the International Organization of Psychophysiology; 2020; 158th edition; p. 9–15. DOI: 10.1016/j.ijpsycho.2020.09.007.
11. Hart S, Stavenland LE. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In Human Mental Workload. Elsevier, Chapter 7, 139–183.
12. Nicolau S, Soler L, Mutter D, Marescaux J. Augmented reality in laparoscopic surgical oncology. In: Surgical oncology; 2011; 20th edition (3), p. 189–201. DOI: 10.1016/j.suronc.2011.07.002.
13. Dixon BJ, Daly MJ, Chan H, Vescan A, Witterick JJ, Irish JC. Augmented real-time navigation with critical structure proximity alerts for endoscopic skull base surgery. In: The Laryngoscope; 2014; 124th edition (4), S. 853–859. DOI: 10.1002/lary.24385.
14. Williams LM, Simms E, Clark CR, Paul RH, Rowe D, Gordon E. The test-retest reliability of a standardized neurocognitive and neurophysiological test battery: "neuromarker". In: The International journal of neuroscience; 2005; 115th edition; p. 1605–1630. DOI: 10.1080/00207450590958475.
15. Reinschluessel AV, Muender T, Salzmann D, Döring T, Malaka R and Weyhe D. Virtual Reality for Surgical Planning – Evaluation Based on Two Liver Tumor Resections. In: Front. Surg; 2022; 9th edition; DOI: 10.3389/fsurg.2022.821060
16. Colligan L, Potts HWW, Finn CT, Sinkin RA. Cognitive workload changes for nurses transitioning from a legacy system with paper documentation to a commercial electronic health record. In: International Journal of Medical Informatics; 2015; 84th edition (7), p. 469–476. DOI: 10.1016/j.ijmedinf.2015.03.003
17. Hemmerich W. StatistikGuru: Rechner zur Adjustierung des α-Niveaus. Retrieved from https://statistikguru.de/rechner/adjustierung-des-alphaniveaus.html. 2016.
18. Handy TC. (Hg.). Event-related potentials. A methods handbook. Cam-bridge, Mass.: MIT Press (A Bradford book). 2005. ISBN:9780262083331
19. Wascher E, Reiser J, Rinkenauer G, Larrá M, Dreger FA, Schneider D, et al. Neuroergonomics on the Go: An Evaluation of the Potential of Mobile EEG for Workplace

Assessment and Design. In: Hu-man factors; 2021; 187208211007707. DOI: 10.1177/00187208211007707.
20. El Basbasse Y, Packheiser J, Peterburs J, Maymon C, Güntürkün O, Grimshaw G, Ocklenburg S. Walk the plank! Using mobile electroencephalography to investigate emotional lateralization of immersive fear in virtual reality. In: R. Soc. Open Sci.; 2023; 10th edition. DOI:10.1098/rsos.221239
21. Polich J. Updating P300: an integrative theory of P3a and P3b. Clin Neurophysiol. 2007 Oct;118(10):2128-48. doi: 10.1016/j.clinph.2007.04.019. Epub 2007 Jun 18. PMID: 17573239; PMCID: PMC2715154.

Abbreviations

AR: Augmented Reality
AEP: Auditory Evoked Potential
EEG: Electroencephalography
ERP: Event Related Potential
VR: Virtual Reality

Supplementary Files

Multimedia Appendixes

Video part 1 of VR-Condition.
URL: http://asset.jmir.pub/assets/ecc2a7731aa60cdcb612fbdc269d293e.mp4

Video part 2 of VR-Condition.
URL: http://asset.jmir.pub/assets/1e953d4721155f2c5cafba5debbe8bb8.mp4

Screnshot of PC-Condition.
URL: http://asset.jmir.pub/assets/3f689deb53ed59878fadeb8a1fa768cf.png

Video of AR-Condition.
URL: http://asset.jmir.pub/assets/80124e8c1ec0dda60b0acf560c0048bf.mp4