COMPUTER METHODS IN BIOMECHANICS AND BIOMEDICAL ENGINEERING: IMAGING & VISUALIZATION
https://doi.org/10.1080/21681163.2022.2154277

Audiovisual augmentation for coil positioning in transcranial magnetic stimulation

Laura Schütz[a,b], Emmanuelle Weber[a,c], Wally Niu[a,d], Bruce Daniel[d], Jennifer McNab[a,c], Nassir Navab[b*] and Christoph Leuze[a,d*]

[a]Wu Tsai Visualization Lab, Stanford University, Stanford, California, USA; [b]Chair for Computer Aided Medical Procedures and Augmented Reality, Department of Informatics, Technical University of Munich, Munich, Germany; [c]McNab Lab, Department of Radiology, Stanford University, Stanford, CA, USA; [d]Incubator for Medical Mixed and Extended Reality at Stanford, Department of Radiology, Stanford University, Stanford, CA, USA

ABSTRACT
Transcranial Magnetic Stimulation (TMS) is an effective non-invasive treatment method for major depressive disorder. Accurate placement of an electromagnetic coil on the patient's head during repetitive TMS is the key for stimulation of the desired brain regions and positive treatment outcome. Neuronavigation systems constitute the state-of-the-art method to accurately stimulate the appropriate brain region. Local separation of navigation information and the patient anatomy in combination with intricate visualisations and cumbersome setup limits the benefits and usability of this method. The present study addresses these problems by proposing an audiovisual Augmented reality (AR) system for coil positioning during TMS. The system sonifies and visualises translational and rotational differences between a target and the current instrument position using a minimalistic graphical user interface and auditory display. Effects of cross-modal integration on usability and targeting precision were shown in an experiment comparing audiovisual AR, audio AR and visual neuronavigation. Our approach revealed significant improvements in task time of all proposed AR conditions over neuronavigation ($p < 0.001$). Conversely, the neuronavigation system achieved significantly better targeting accuracy ($p < 0.001$). A purely auditory guidance achieved comparable performance as the audiovisual interface designs.

ARTICLE HISTORY
Received 19 October 2022
Accepted 19 November 2022

KEYWORDS
Multimodal interaction; augmented reality; surgical navigation; audiovisual; human computer interaction

1. Introduction

The human way of interacting with the world is multimodal in its nature, relying on a constant combination of information from multiple senses to understand the surrounding situation and make adequate decisions. (Turk 2014) Augmented reality (AR) blends the real world with virtual content, making multimodal interaction an obvious choice when designing interfaces. (Behringer et al. 2007) Multimodal feedback is more intuitive and efficient for users, increasing their task efficiency and providing overall more reliable results. (Oviatt 1999) The medical field is a prominent area of application for AR, as diagnostic and treatment relevant information can be superimposed on the patient (Cartucho et al. 2020). Along this development, the introduction of AR has the potential to both simplify procedures and improve user experience by conveying this wealth of information in an intuitive format. Numerous studies have shown the potential and advantages of incorporating AR in medical procedures (Bichlmeier et al. 2009; Leuze et al. 2018; Liebmann et al. 2019). Bichlmeier et al. (2009) created a 3D volumetric medical data viewer registered to patient anatomy for a more intuitive visualisation, while Liebmann et al. (2019) examined 3D augmented views for spine surgery using a Microsoft HoloLens to display target trajectories on the patient. However, the great majority of studies presenting AR solutions to assist medical professionals are limited to the visual modality. In this study, we examine the potential of combined audio and visual feedback to better guide the user while performing a highly demanding medical procedure such as TMS treatment. In the following paragraphs we review existing medical AR solutions using only visual feedback and those including visual and auditory feedback.

1.1. Visual medical AR

Research in medical AR can be grouped into applications for training and education, and medical planning and guidance applications. Training applications include work on training of hand-eye coordination for endotracheal intubations (Hamza-Lup et al. 2018) or the usage of ultrasound systems (Blum et al. 2009). Medical guidance applications include studies like Fuchs et al. (1998) who presented 3D visualisations for guidance during laparoscopic surgical procedures. Elmi-Terander et al. (2016) proved the accuracy of AR guidance for pedicle screw placement during spinal fusion surgery. Besides studies proving enhanced accuracy, others have demonstrated increased usability of medical navigation tasks when introducing AR. An article by Fischer et al. (2016) proved clear advantages of tool placement with AR over the conventional X-ray-based technique. Their intuitive AR display helped with precise and efficient tool guidance.

CONTACT Laura Schütz laura.schuetz@tum.de Chair for Computer Aided Medical Procedures and Augmented Reality, Department of Informatics, Technical University of Munich, Munich, Germany
*Shared senior author

Analogously, Heinrich et al. (2020) showed superior usability of AR over traditional monitor-based display guidance for the task of medical needle insertion by eliminating the mentally demanding hand-eye coordination.

1.2. Audiovisual medical AR

Experiments in surgical sonification have been performed by Matinfar et al. (2018) who used automatic musical soundtrack modification to show the potential of auditory stimuli for conveying navigation information during medical procedures. Among bimodal approaches combining both the auditory and visual modality for medical guidance tasks are Bork et al. (2015) and Roodaki et al. (2017). Bork et al. (2015) introduced a dynamic multi-sensory AR system for locating occluded anatomy and used audio to improve localisation perception in 3D, which led to enhanced accuracy of needle placement when using auditory and visiotemporal guidance. Roodaki et al. (2017) utilised physical sound synthesis to evaluate the impact of auditory feedback on the performance of high-precision medical tasks, such as needle placement in ophthalmic microscope surgery. A sole audio guidance resulted in increased angle accuracy over a visual and audiovisual system. Black et al. (2017) also proposed an auditory feedback system for needle placement. They investigated the impact of audio and combined audiovisual feedback on the accuracy of navigated needle placement, task time, and subjective workload by adding audio feedback via headphones to an existing surgical navigation system. They adjusted the sound parameters of pitch for up-down and stereo panning for left-right movements for the needle position. Their study showed increased task time and workload for auditory feedback over audiovisual and visual navigation. Joeres et al. (2021) introduced an audiovisual AR system for laparoscopic procedures. Their auditory display utilised two sounds, running water to represent the instrument and a synthesised tone to represent the vessels. The water sound was modified in rhythm and pitch depending on structure density and distance to target. In our study we also evaluate the effects of an audiovisual system that is fully presented in AR and where auditory and visual stimuli are used to convey navigation information. Compared to Joeres et al. (2021) we aim to also separately evaluate all visual and audio feedback components and present them to the user independently. In doing so we evaluate the modalities and spatial parameters of distance and angle individually to understand cognition effects of cross-modal integration depending on the spatial information sonified.

1.3. Guidance systems in transcranial magnetic stimulation

The present study will investigate audiovisual augmentation for the task of coil positioning in transcranial magnetic stimulation (TMS). One important application of TMS therapy is depression treatment. During treatment an electromagnetic coil is close to the patient's head to stimulate brain regions that regulate mood and behaviour. Repeated sessions of repetitive TMS have been shown to improve depression symptoms (Loo and Mitchell 2005). Accurate coil placement using neuronavigation (Denslow et al. 2005) has been shown to improve treatment efficacy (Fitzgerald et al. 2009). However, current neuronavigation solutions are costly. Complex user interfaces require training and lengthy setup increases treatment time. This has hindered the widespread use of neuronavigation in clinical settings. As an alternative to neuronavigation some studies propose robot-guided TMS systems for precise localisation and treatment (Richter et al. 2011; Xiao et al. 2018). Although the use of robotic systems has great potential to increase targeting accuracy, they too suffer from high costs and complex setups. Therefore, most TMS procedures are still executed without navigation or guidance systems using a measuring tape and standard scalp measurements for targeting. While this is faster and easier to use, such a targeting approach does not account well for individual differences in head shape and brain anatomy leading to decreased targeting accuracy for some patients (Herwig et al. 2001).

To make TMS coil placement during treatment more intuitive and easier to use over neuronavigation and robotic guidance, we propose an AR approach using audiovisual augmentation. Unlike prior AR navigation solutions for TMS who only focused on visual anatomic display and tracking accuracy (Soeiro et al. 2016; Leuze et al. 2018; Sathyanarayana et al. 2020), we focus on the usability of a multimodal AR system. We will test whether sonifying the positional and rotational changes of the TMS coil with respect to the desired brain area will lead to quicker coil placement at the desired location. We furthermore compare unimodal visual, bimodal audiovisual or unimodal auditory guidance to test, which of these AR feedback modalities is most user-friendly and leads to quick and accurate coil placement. For the visual condition we will be using a standard neuronavigation system, the TMS Navigator by Localite[1]. For presenting the audio and audiovisual condition we are using custom software we developed for the Microsoft HoloLens 2[2]. We will explore the influence of two audiovisual AR conditions on the presentation of certain spatiotemporal information. Lastly, we will present a purely auditory display in AR and convey all navigational information for precise coil positioning via sound. We hypothesise that the audiovisual conditions will result in the best usability, followed by the audio and visual condition. We will be measuring placement accuracy at the time of impulse trigger, task completion time for guiding the coil to the target and usability measures.

2. Materials and methods

We present three AR guidance methods with varying visual and auditory feedback and compare the AR guidance to the usability and targeting speed of a state-of-art neuronavigation system. We used the Localite TMS navigator and a NDI optical tracking camera[3] to track the patient's head and the coil. The audiovisual and audio AR interfaces presented to the participants on the HoloLens 2 were developed in Unity[4].

Coil tracking

We recorded the coil position at each Unity frame with the HoloLens 2 integrated camera using the Vuforia computer

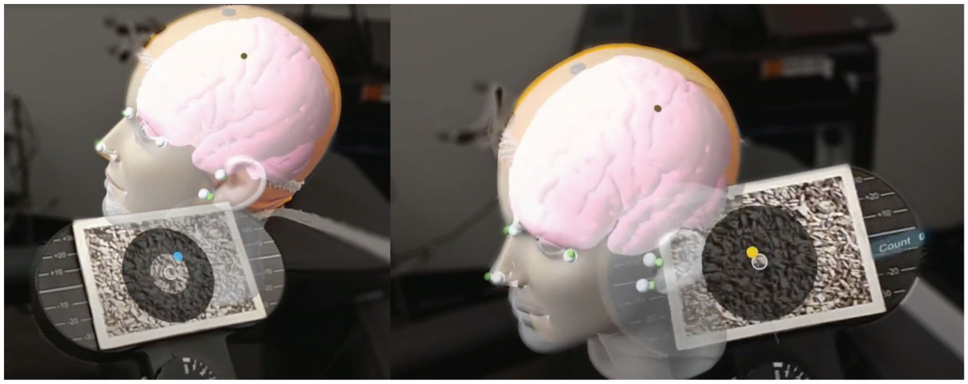

Figure 1. Audiovisual augmented reality guidance using HoloLens 2 for targeting a desired brain region for TMS. Left: Audiovisual Distance guidance – the visual display projected on the coil represents the current coil trajectory in relation to the pre-planned target trajectory. The distance to the target is sonified. Right: Audiovisual Angle guidance – the visual display shows the direction and distance to the target while the angle is sonified.

vision framework[5] and a Vuforia compatible image marker attached to the TMS coil.

Head registration

We first performed a rigid registration of a 3D virtual head to a volunteer head. The TMS operator first placed five virtual fiducials onto anatomical landmarks of the volunteer's head (ears, nasian, eye corners, nose tip) using the Mixed Reality Toolkit's[6] hand tracking tool. Once the virtual fiducials were placed, the 3D virtual head including a mesh derived from the MNI T1 weighted scans (305 MRI) was registered onto the patient's head (Evans et al. 1992;1992b, Evans et al. 1993; Collins et al. 1994). The mesh was created using MeshLab (Cignoni et al. 2008) and the 3D scan of the head was edited in Blender[7]. The volunteer was seated in a treatment chair including a headrest, making it easy to keep the head steady during the experiment.

Visual and auditory cues

We used Unity User Interface elements to create a minimalistic visual display on the coil. Unity Audio Sources were used to modify sound samples in pitch and volume. The sound was played back using the HoloLens 2 stereo speakers. We designed two visually reduced interfaces displaying abstract 2D guidance (Figure 1). The two versions of the audiovisual guidance (Audiovisual Angle and Audiovisual Distance) map the real-time 3D Cartesian orientation of the coil to a polar coordinate system centred at the pre-planned target position.

The **Audiovisual Distance (AVD)** condition intuitively displays relative angular distance between the planned and the actual position of the coil on top of the image marker attached to the coil. The azimuth angle deviation is calculated using the dot product between the axis perpendicular to the target and the axis perpendicular to the coil pedal (Figure 2(a)). A circle element dynamically moves across the underlying black background in accordance with the live pose of the coil. By orienting the coil such that the circle-shaped element enters the central ring – outlined in white – the user attains the desired location. The correct positioning is indicated via additional visual feedback to the user by changing the colour of the circular element from blue to pink. The Audiovisual Distance condition is completed with a harmonic string sound at 440 Hz. Its volume is a linear function of distance between coil and target (Figure 2(b)). If the coil is far from the target, the volume is high. As the coil nears the target the volume is decimated and ultimately the sound goes silent once the minimal distance has been achieved.

The **Audiovisual Angle (AVA)** system uses the same principle and maps the direction and distance to the target onto the black circular background (Figure 2(c)). If the moving yellow

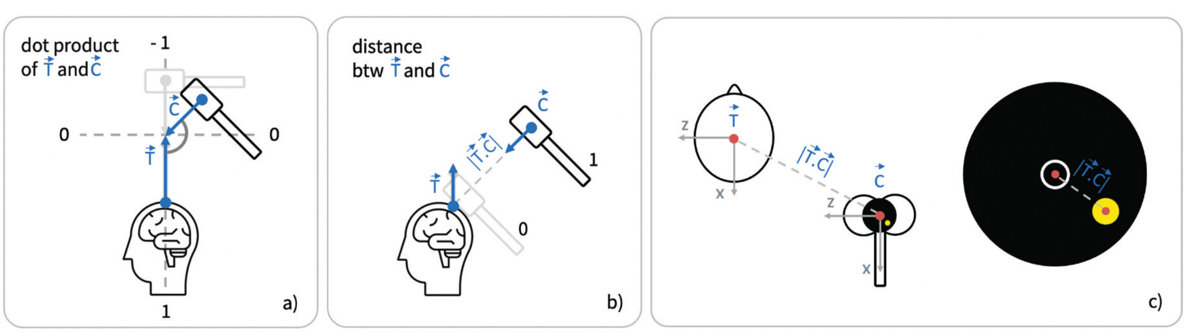

Figure 2. Sonification of angle and distance, a) the Audiovisual Angle condition modulates the frequency of the sine wave depending on the dot product between the target and coil vector, b) the Audiovisual Distance condition modulates the volume of the harmonic sound depending on the distance between the target and coil vector, c) the Audiovisual Angle condition displays the moving blue dot in relation to the inner white circle outline depending on the distance and direction of the coil to the target.

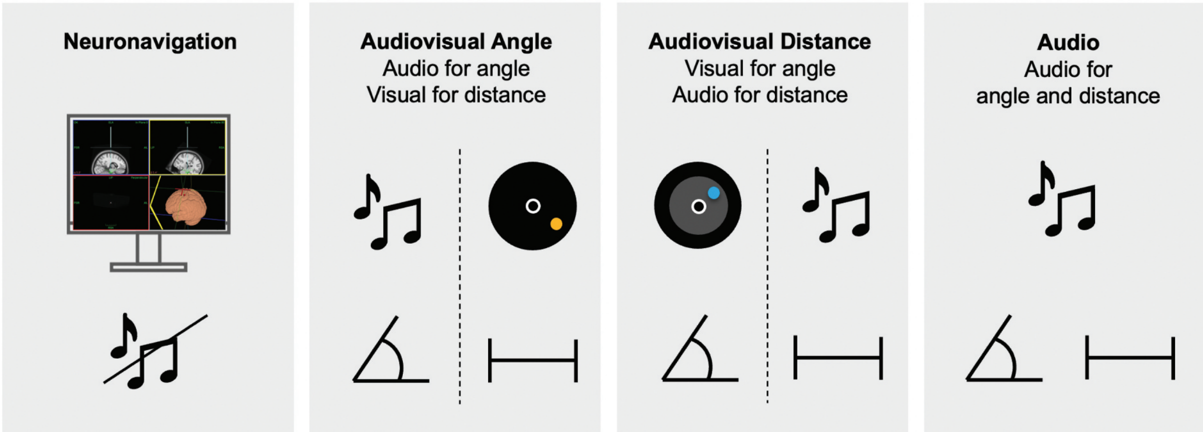

Figure 3. Four conditions that where presented to all participants in randomized order. From left to right: Neuronavigation only Visual (V)), Audiovisual Angle (AVA), Audiovisual Distance (AVD), only Audio (A).

circle enters the central white ring it turns pink. The visual feedback is accompanied by a dual tone sine wave of 250 Hz that is modulated in frequency depending on the angular difference between the coil and the desired target on the volunteer's head (Figure 2(a)). Based on the dot product between these two vectors the sound increases in pitch as the angle between planned target trajectory and coil increases and vice versa. Once the correct angle is achieved the sine wave sound goes silent.

For the **Audio Only (A)** condition, both previously described sounds are combined and no visuals are displayed. The harmonic sound is a linear function of distance to the desired target and the dual tone sine wave is a linear function of the azimuth angle to the desired target. When both distance and angle values are below the target threshold, no more sound is audible. The decision to decrease instead of increase the volume and pitch as the coil nears the target is an attempt to put the user at ease when the correct position has been reached and to reduce the task load on the user.

3. User study

To answer the question whether audiovisual augmentation can improve the usability of medical guidance tasks, we performed a comparative study using four conditions (Figure 3). Specifically, we tested the following hypotheses:

- H1. Audiovisual augmentation (AVA, AVD) results in faster task completion time than a purely auditory (A) or visual display (V).
- H2. Audio mapping spatial distance information (AVD) is more intuitive than sonification of spatial angle (AVA) and will result in a lower mean task time and steeper learning curve.
- H3. Using audiovisual augmentation (AVA, AVD) reduces the cognitive load for the user when compared to the auditory display (A) and visual neuronavigation (V) condition.

Participants

A total of 16 (7 females and 9 males) volunteers participated in the study. Their mean age was 30 ± 11.2 *std* years. The participants consisted of students and researchers in Computer Sciences, Design, Music and Medicine who all consented to take part in the study. The participants can be separated into two groups, those experienced in using AR systems (8 participants) and novice users (8 participants) for whom the experiment was their first interaction with AR. None of the participants had previous experience with using the proposed guidance system. However, 4 participants indicated to be familiar with the TMS procedure, three were familiar with interpreting medical image data and two had even administered TMS treatment themselves before.

Procedure

Prior to the study the subjects were not exposed to a training session. This conscious decision was made to test a learnability effect when first using the provided interfaces. We performed the study in a dedicated room for TMS treatment containing a MagVenture TMS System. The setup consisted of the HoloLens 2 for AR navigation and the Localite neuronavigation system, which included NDI optical tracking cameras. A laptop was used to run a TCP server that received a trigger from the TMS system whenever a TMS pulse was executed and sent the trigger event data to the HoloLens 2 worn by the participant operating the TMS coil. An image marker was attached to the coil to enable tracking by the HoloLens 2 integrated camera. The TMS user was asked to go through all four conditions (Figure 3) in a randomised order. For each condition 12 targets - one at a time – were presented in a random order. The participant was instructed to move the coil to the virtual target using the visual and or auditory feedback of the respective condition. Once the user deemed to have reached the correct, pre-planned position, they could proceed to the next target by saying the keyword 'Next'. The MRTK Speech Input module registered the keyword and then presented the next target to

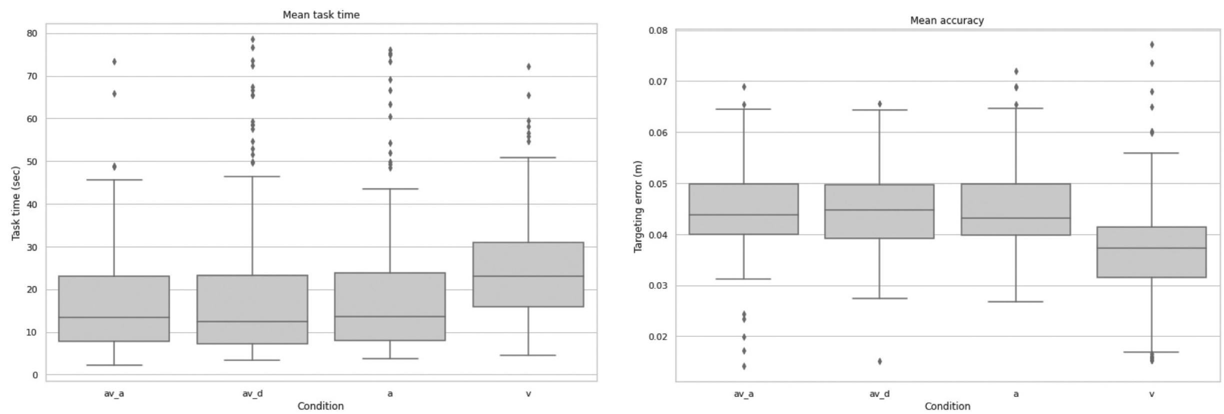

Figure 4. Box plots of task time and targeting accuracy as mean squared error.

the subject. For every trigger that the HoloLens 2 received, the coil position, the virtual head position and the current target position was stored in a file. Once the participants had gone through all four conditions, they were asked to complete a survey right away to assess subjective workload and usability measures. The study procedure took approximately one hour per participant.

Data Collection and Analysis

During the experiment the poses of the coil, target and virtual head were recorded for each trigger event to evaluate coil placement accuracy. The task completion time for each condition was also measured for each target individually to evaluate the ease of the user interface and navigation feedback. The time measurement started when the next target appeared on the virtual head or the neuronavigation desktop display and ended when the user pushed the trigger to signal accurate coil placement. The evolution of task time over the course of multiple trials was also used to investigate the learnability of the provided interfaces. Lastly, a survey comprising questions about demographic background and the NASA-TLX (Hart and Staveland 1988) was distributed to measure cognitive load and overall usability. All the above measures were compared across conditions. Adjustment for multiple comparisons was done using the Holm method.

4. Results

4.1. Task time and learnability

Since a within-subject design was chosen for this study, a Friedman test was used to examine the effect of the four different conditions on the task time. The test showed that the kind of interface used led to statistically significant differences in task time ($p < 0.001$). Consequently, a paired Wilcoxon test with Holm correction was used to compare the four conditions. The test revealed that task completion time was significantly lower for both proposed audiovisual systems AVA ($p < 0.001$) and AVD ($p < 0.001$) as well as the audio condition A ($p < 0.001$) when compared to the purely visual condition presented on the external display (V). The tests further showed no significant differences between pairs of the three proposed AR conditions AVA, AVD, and A (Figure 4(a)). The task time's means and standard deviations are reported in Table 1.

To analyse the data for a learning effect, we plotted the task completion times of all participants over the course of the study for all targets and for the four conditions (Figure 5). A stark decline of task time was found indicating a strong learning effect for all four conditions. Condition AVA showed the most stable decline in task time. This result implies that all of the provided interfaces are easy to learn.

4.2. Targeting accuracy

The coil placement accuracy was recorded for the three conditions presented on the AR glasses as well as the neuronavigation system (Table 1). It was calculated as the mean squared error of the distances between the final coil positions and the target positions in the coordinate space of the navigation system. Statistical analysis using a Friedman test showed a significant difference between conditions ($p < 0.001$). A paired Wilcoxon test with Holm correction was chosen to compare between conditions. The tests showed no significant differences for pairs within the AR conditions (AVA, AVD, A). Pairwise comparisons between the neuronavigation (V) and AVA, AVD and A respectively all showed significantly better targeting accuracy of the neuronavigation system ($p < 0.001$). (Figure 4(b))

Table 1. Means and standard deviations (std) of task time in seconds, targeting error in metres and NASA-TLX ([0, 100]; lower is better) for all conditions: Audiovisual Angle (AVA), Audiovisual Distance (AVD), Audio Only (A), Neuronavigation (V).

	Time		Accuracy		NASA-TLX	
Cond	Mean	Std	Mean	Std	Mean	Std
AVA	16.88	12.58	0.0443	0.0083	43.08	9.18
AVD	18.42	17.03	0.0447	0.0076	48.45	11.41
A	18.70	15.40	0.0448	0.0078	47.31	12.41
V	25.70	12.86	0.0372	0.0109	53.65	6.59

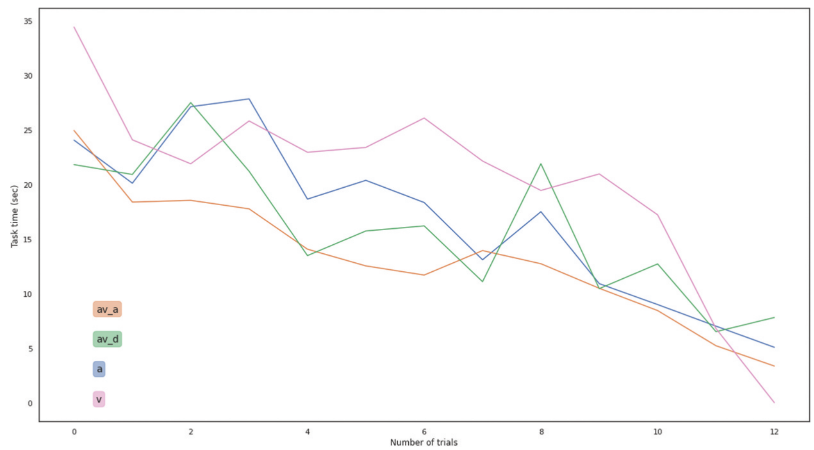

Figure 5. Learnability of the different conditions; plot of the average task time of all participants over all targets for each of the four conditions: Audiovisual Angle (av_a), Audiovisual Distance (av_d), Audio Only (a), neuronavigation (v).

4.3. Subjective workload

The obtained subjective workload data from the NASA-TLX questionnaire was analysed using a Friedman test. Results showed a significant difference between conditions. Hence a paired Wilcoxon test with Holm correction was used to compare between conditions. The pairwise comparisons did not yield significant differences. Overall Audiovisual Angle (AVA) performed best, followed by Audio (A) and Audiovisual Distance (AVD). (Table 1) The purely visual condition using the neuronavigation system received the worst rating for the overall workload as well as all six dimensions of the NASA-TLX (Mental Demand, Physical Demand, Temporal Demand, Performance, Effort and Frustration).

5. Discussion

The results have confirmed that audiovisual AR guidance results in faster task completion time than neuronavigation (H1). Contrary to what was expected the audio feedback (A) achieved the same time performance as the audiovisual systems. High learnability was reported for all interfaces promising easy adoption of these new systems into everyday routines. (Rafique et al. 2012)

The angle sound represented the angle between the two vectors perpendicular to the surface of the target and the coil respectively, thus encoding two rotational dimensions. Information on the third rotational dimension, the rotation around its own axis, was not presented to the user as it is secondary to the successful placement of the coil. This approach was chosen to abstract the complexity of the task in an attempt to reduce the cognitive load on the user. The mapping of this rich spatial information onto a linear scale revealed the unexpected rejection of the hypothesis (H2) that sonifying distance would be more easily processed by the user, as suggested by Bazilinskyy et al. (2016). Sonification of distance and angle information showing equivalent results raises the question if the abstraction of the angle onto a linear scale (high to low frequency) influenced the usability of the angle sound in a positive manner. Encoding all three rotational dimensions might help to improve the accuracy. However, depending on the chosen sound design the cognitive load and mental demand might increase.

The assumptions made in H3 were rejected as all provided interfaces achieved similar subjective workload ratings. Unlike hypothesised based on the work by Black et al. (2017), Anderson and Zahorik (2014) and Joeres et al. (2021) the purely auditory display has shown similar cognitive load as well as task time and targeting accuracy as the audiovisual conditions. The study by Joeres et al. (2021) showed worse results for the auditory display than the purely visual display. A possible explanation for this finding is that unlike Joeres et al. (2021) who presented a redundant multimodal system conveying the same information via both channels, we presented the two spatial parameters of distance and angle complementary via two distinct sensory channels for the audiovisual conditions and two distinct sounds for the audio condition that were easily distinguishable at all times. When using the audio AR an observed participant strategy was the placement of the coil onto the right location on the head, which silenced the harmonic distance sound followed by the adjustment of the angle using the sine wave sound. This clear separation between the spatial dimensions of distance and angle in the auditory domain might have influenced this outcome. Further investigation into the psychoacoustic effects underlying this audio navigation design might provide insights into audio cognition phenomena that are subject to this finding.

The non-significant results for some measures might be influenced by the diverse group of subjects who participated in the study. Some participants had a medical background which influenced their interaction with the neuronavigation system that presented Magnetic resonance imaging (MRI) information to the user. Other participants had never seen MRI images before but were expert users in AR. Yet other participants had never used AR or Virtual Reality before and never interacted with anatomical images. This diversity might have balanced out contradictory tendencies in results. In a future study a homogeneous group of the intended user of this application, the medical expert, should be recruited.

Unlike the usability measures, the neuronavigation system proved better in targeting accuracy than the AR conditions. A potential factor influencing this outcome is the implementation of the Unity application run on the HoloLens 2. While the frame rate was at 60 fps, the Vuforia image marker tracking for Unity caused a noticeable latency that resulted in participants overshooting targets, as the visuals and sound were not updated fast enough to reflect the movement of the coil at all times. In a future study this limitation should be resolved to create a tracking experience as steady as the one provided by the NDI Polaris Vega ST optical tracking system used by the Localite neuronavigation system. In improving the tracking experience the task time might decrease and the accuracy and usability of the AR application might increase. Such work would provide even more information about the viability of audiovisual AR guidance as an alternative to standard navigation solutions.

6. Conclusion

A multimodal augmented reality system for a medical guidance task was presented. All presented conditions provided navigation information in a complementary way. The bimodal audiovisual conditions presented in AR reserved either the distance or angle information for one sensory channel. The unimodal visual (neuronavigation) and unimodal audio AR condition conveyed both spatial dimensions of distance and angle via the same sensory channel. No clear distinction in task time, precision or usability was found between the three AR conditions (AVA, AVD, A). Whether the distance was sonified and the angle visualised, or vice versa did not influence the participants' performance of the guidance task. The purely auditory display achieved the same results as the audiovisual and visual conditions and has thus outperformed our expectations.

The study revealed a significant improvement in task time for all proposed audio and audiovisual AR guidance techniques over state-of-the-art neuronavigation. Our study has provided relevant findings for the future design of bimodal navigation applications. Overall, the application of audiovisual interactions in AR for medical guidance tasks has proven promising in reducing cognitive load and mental demand on the user. Furthermore, a strong learnability effect of the presented system was demonstrated, which would facilitate easy adoption into everyday clinical routine. Introducing multimodal AR guidance such as the audiovisual system for TMS treatment presented in this study has great potential to make precise and user-friendly navigation widely accessible as it reduces cognitive load, setup time and equipment cost over medical navigation systems.

Notes

1. Localite (https://www.localite.de/en/home/).
2. Microsoft HoloLens (https://www.microsoft.com/en-us/hololens).
3. NDI (https://www.ndigital.com/).
4. Unity (https://unity.com/).
5. Vuforia Engine Library (https://library.vuforia.com/).
6. MRTK2 for Unity (https://docs.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk2).
7. Blender (http://www.blender.org).

Acknowledgements

The authors wish to thank Christopher Cline at the Stanford Wu Tsai Neurosciences Building for his support with the TMS and Neuronavigation system setup and for providing the facilities for the study. The authors also wish to thank Jarrett Rosenberg for his advice on data analysis.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Laura Schütz is a graduate student in Design with a focus on healthcare innovation at the Stanford School of Engineering. Prior to her studies at Stanford she received a BA in Architecture and MS in Industrial Design from Technical University of Munich. For her work on interaction and user experience design for medical applications she received the IAUD International Design Award and the Universal Design Expert Award in 2020.

EmmanuelleWeber is a Postdoctoral Research Fellow at the Stanford University School of Medicine. She received her Ph.D. in Chemical Physics from École Normale Supérieure in Paris. Her research interests comprise the development of methods to study brain microstructures using Diffusion Magnetic Resonance Imaging and leveraging Mixed Reality devices to improve TMS targeting.

Wally Niu is a Ph.D. student in Bioengineering at Stanford University, with a prior BS degree in Biomedical Engineering, and minors in Computer Science, Robotics, and Entrepreneurship and Management from Johns Hopkins University. His primary interests lie in surgical robotics, human-robot interactions, augmented reality, and sensors.

Bruce Daniel is a Professor of Radiology and, by courtesy, of Bioengineering. He did his residency at the University of Michigan, before joining Stanford in 1995 as an NCI fellow and faculty member. His main areas of interest include breast MRI, and diagnostic body MRI techniques, as well as instrumentation for MRI-guided procedures. He has started and co-directed IMMERS, the Incubator for Medical Mixed and Extended Reality at Stanford where many projects centre around merging patients and their imaging data in augmented reality.

JenniferMcNab is an Associate Professor of Radiology at the Richard M. Lucas Center for Imaging at Stanford University. She is an MRI Physicist focused on the development of magnetic resonance imaging (MRI) contrast mechanisms and acquisition strategies that yield new and/or improved images of the in vivo human brain.

NassirNavab is director of the Computer Aided Medical Procedures (CAMP) research team at both Technical University of Munich and Johns Hopkins University. He has also secondary faculty appointments at both affiliated medical schools. He completed his Ph.D. at INRIA and University of Paris XI, France, and enjoyed two years of post-doctoral fellowship at MIT Media Laboratory before joining Siemens Corporate Research (SCR) in 1994. He received the SMIT Society Technology award in 2010 and the '10 years lasting impact award' of IEEE ISMAR in 2015. In 2012, he was elected as a Fellow of the MICCAI Society. He has acted as a member of the board of directors of the MICCAI Society, 2007-2012 and 2014-2017, and served on the Steering committee of the IEEE Symposium on Mixed and Augmented Reality (ISMAR) and Information Processing in Computer Assisted Interventions (IPACI). He is the author of hundreds of peer-reviewed scientific papers with more than 50000 citations and an h-index of 104 as of August 12, 2022. His current research interests include medical augmented reality, computer-assisted surgery, medical robotics, computer vision and machine learning.

ChristophLeuze is director of the Wu Tsai Visualisation Lab at Stanford University and founder of the Augmented Reality Startup Nakamir. He received his Ph.D. in Physics from Max Planck Institute for Human Cognitive and Brain Sciences & University of Leipzig. Dr. Leuze is an active member of the Incubator for Medical Mixed and Extended Reality at Stanford.

ORCID

Laura Schütz http://orcid.org/0000-0002-5534-3903
Emmanuelle Weber http://orcid.org/0000-0003-0277-394X
Bruce Daniel http://orcid.org/0000-0003-0475-5892
Jennifer McNab http://orcid.org/0000-0002-1763-0792

Nassir Navab http://orcid.org/0000-0002-6032-5611
Christoph Leuze http://orcid.org/0000-0003-4564-8014

References

Anderson PW, Zahorik P. 2014. Auditory/Visual distance estimation: accuracy and variability. Front Psychol. 5. doi:10.3389/fpsyg.2014.01097.

Bazilinskyy P, van Haarlem W, Quraishi H, Berssenbrugge C, Binda J, de Winter J. 2016. Sonifying the location of an object: a comparison of three methods. Ifac-PapersOnline. 49(19):531–536. doi:10.1016/j.ifacol.2016.10.614.

Behringer R, Christian J, Holzinger A, Wilkinson S. 2007. Some usability issues of augmented and mixed reality for e-health applications in the medical domain. In: Holzinger, A. (ed) HCI and usability for medicine and health careBerlin, Heidelberg: Springer; pp. 255–266.

Black D, Hettig J, Luz M, Hansen C, Kikinis R, Hahn H. 2017. Auditory feedback to support image-guided medical needle placement. Int J Comput Assist Radiol Surg. 12(9):1655–1663. doi:10.1007/s11548-017-1537-1.

Blum T, Heining SM, Kutter O, Navab N. 2009. Advanced training methods using an augmented reality ultrasound simulator. 2009 8th IEEE International Symposium on Mixed and Augmented Reality; Orlando, Florida, USA. p. 177–178.

Bork F, Fuers B, Schneider AK, Pinto F, Graumann C, Navab N. 2015. Auditory and visio-temporal distance coding for 3-dimensional perception in medical augmented reality. IEEE international symposium on mixed and augmented reality; Fukuoka, Japan. p. 7–12.

Cartucho J, Shapira D, Ashrafian H, Giannarou S. 2020. Multimodal mixed reality visualisation for intraoperative surgical guidance. Int J Comput Assist Radiol Surg. 15(5):819–826. doi:10.1007/s11548-020-02165-4.

Cignoni P, Callieri M, Corsini M, Dellepiane M, Ganovelli F, Ranzuglia G. 2008. MeshLab: an open-source mesh processing tool. In: Scarano V, Chiara R, and Erra U, editors Eurographics Italian Chapter Conference; Salerno, Italy. The Eurographics Association.

Collins D, Neelin P, Peters T, Evans A. 1994. Automatic 3-d intersubject registration of mr volumetric data in standardized talairach space. J Comput Assist Tomogr. 18(2):192–205. doi:10.1097/00004728-199403000-00005.

Denslow S, Bohning DE, Bohning PA, Lomarev MP, George MS. 2005. An increased precision comparison of tms-induced motor cortex bold fMRI response for image-guided versus function-guided coil placement. Cogn Behav Neurol. 18(2):119–126. doi:10.1097/01.wnn.0000160821.15459.68.

Elmi-Terander A, Skulason H, Söderman M, Racadio J, Homan R, Babic D, van der Vaart N, Nachabe R. 2016. Surgical navigation technology based on augmented reality and integrated 3d intraoperative imaging: a spine cadaveric feasibility and accuracy study. Spine. 41(21):E1303–1311. doi:10.1097/BRS.0000000000001830.

Evans A, Collins D, Mills S, Brown E, Kelly R, Peters T. 1993. 3d statistical neuroanatomical models from 305 mri volumes. Proc IEEE-Nuclear Science Symposium and Medical Imaging Conference; San Francisco, CA, USA.1813–1817.

Evans A, Collins D, Milner B. 1992. An mri-based stereotaxic atlas from 250 young normal subjects. Proc 22nd Annual Symposium, Society for Neuroscience. 18:408.

Evans A, Marrett S, Neelin P, Collins D, Worsley K, Dai W, Milot S, Meyer E, Bub D. 1992b. Anatomical mapping of functional activation in stereotactic coordinate space. NeuroImage. 1(1):43–63. doi:10.1016/1053-8119(92)90006-9.

Fischer M, Fuerst B, Lee SC, Fotouhi J, Habert S, Weidert S, Euler E, Osgood G, Navab N. 2016. Preclinical usability study of multiple augmented reality concepts for k-wire placement. Int J Comput Assist Radiol Surg. 11(6):1007–1014. doi:10.1007/s11548-016-1363-x.

Fitzgerald PB, Hoy K, McQueen S, Maller JJ, Herring S, Segrave R, Bailey M, Been G, Kulkarni J, Daskalakis ZJ. 2009. A randomized trial of RTMs targeted with mri based neuro-navigation in treatment-resistant depression. Neuropsychopharmacology. 34(5):1255–1262. doi:10.1038/npp.2008.233.

Fuchs H, Livingston MA, Raskar R, Keller K, Crawford JR, Rademacher P, Drake SH, Meyer AA. 1998. Augmented reality visualization for laparoscopic surgery. In: International Conference on Medical Image Computing and Computer-Assisted Intervention; Cambridge, MA, USA. Springer. p. 934–943.

Hamza-Lup FG, Rolland JP, Hughes C. 2018. A distributed augmented reality system for medical training and simulation. arXiv preprint arXiv:181112815.

Hart SG, Staveland LE. 1988. Development of nasa-tlx (task load index): results of empirical and theoretical research. In: Peter, A., Meshkati, Najmedin, (eds). Advances inpsychology. Vol. 52. Amsterdam, Netherlands: North-Holland; pp. 139–183.

C, Heining SM, Feuerstein M, Navab N. 2009. The virtual mirror: a new interaction paradigm for augmented reality environments. IEEE Trans Med Imaging. 28(9):1498–1510. doi:10.1109/TMI.2009.2018622.

Heinrich F, Schwenderling L, Joeres F, Lawonn K, Hansen C. 2020. Comparison of augmented reality display techniques to support medical needle insertion. IEEE Trans Vis Comput Graph. 26(12):3568–3575. doi:10.1109/TVCG.2020.3023637.

Herwig U, Padberg F, Unger J, Spitzer M, Schönfeldt-Lecuona C. 2001. Transcranial magnetic stimulation in therapy studies: examination of the reliability of “standard” coil positioning by neuronavigation. Biol Psychiatry. 50(1):58–61. doi:10.1016/S0006-3223(01)01153-2.

Joeres F, Black D, Razavizadeh S, Hansen C. 2021. Audiovisual AR concepts for laparoscopic subsurface structure navigation. In: Graphics Interface 2021.

Leuze C, Yang G, Hargreaves B, Daniel B, McNab JA. 2018. Mixed-reality guidance for brain stimulation treatment of depression. 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct); Munich, Germany. IEEE. p. 377–380.

Liebmann F, Roner S, von Atzigen M, Scaramuzza D, Sutter R, Snedeker J, Farshad M, Fürnstahl P. 2019. Pedicle screw navigation using surface digitization on the Microsoft hololens. Int J Comput Assist Radiol Surg. 14(7):1157–1165. doi:10.1007/s11548-019-01973-7.

Loo CK, Mitchell PB. 2005. A review of the efficacy of transcranial magnetic stimulation (tms) treatment for depression, and current and future strategies to optimize efficacy. J Affect Disord. 88(3):255–267. doi:10.1016/j.jad.2005.08.001.

Matinfar S, Nasseri MA, Eck U, Kowalsky M, Roodaki H, Navab N, Lohmann CP, Maier M, Navab N. 2018. Surgical soundtracks: automatic acoustic augmentation of surgical procedures. Int J Comput Assist Radiol Surg. 13(9):13. doi:10.1007/s11548-018-1827-2.

Oviatt S. 1999. Mutual disambiguation of recognition errors in a multimodal architecture. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; New York, NY, USA: Association for Computing Machinery. p. 576–583.

Rafique I, Weng J, Wang Y, Abbasi MQ, Lew P, Wang X. 2012. Evaluating software learnability: a learnability attributes model. International Conference on Systems and Informatics (ICSAI); Yantai, China. p.2443–2447.

Richter L, Ernst F, Schlaefer A, Schweikard A. 2011. Robust real-time robot–world calibration for robotized transcranial magnetic stimulation. Int J Med Rob Comput Assisted Surg. 7(4):414–422. doi:10.1002/rcs.411.

Roodaki H, Navab N, Eslami A, Stapleton C, Navab N. 2017. Sonifeye: sonification of visual information using physical modeling sound synthesis. IEEE Trans Vis Comput Graph. 23(11):2366–2371. doi:10.1109/TVCG.2017.2734327.

Sathyanarayana S, Leuze C, Hargreaves B, Daniel B, Wetzstein G, Etkin A, Bhati MT, McNab JA. 2020. Comparison of head pose tracking methods for mixed-reality neuronavigation for transcranial magnetic stimulation. In: Medical imaging 2020: image-guided procedures, robotic interventions, and modeling. Vol. 11315. Houston, Texas, United States: International Society for Optics and Photonics; p. 113150.

Soeiro J, Cláudio AP, Carmo MB, Ferreira HA. 2016. Mobile solution for brain visualization using augmented and virtual reality. In: 20th International Conference Information Visualisation (IV); Lisbon, Portugal. p. 124–129.

Turk M. 2014. Multimodal interaction: a review. Pattern Recognit Lett. 36:189–195. doi:10.1016/j.patrec.2013.07.003.

Xiao Z, Ruan Q, Wang X. 2018. Cortexbot: 3d visual fusion of robotic neuronavigated tms system. In: IEEE International Conference on Cyborg and Bionic Systems (CBS); Shenzhen, China. p. 406–411.