See It and Hear It: Multimodal Guidance in MR-Based Neurosurgical Simulation for Skill Retention

Pasquale Cascarano[*]
pasquale.cascarano2@unibo.it
University of Bologna
Bologna, Italy

Andrea Loretti
a.loretti@unibo.it
University of Bologna
Bologna, Italy

Luca Zanuttini
luca.zanuttini@studio.unibo.it
Institute of Neurological Sciences
Bologna, Italy

Daniele Giunchi
d.giunchi@bham.ac.uk
University of Birmingham
Birmingham, UK

Riccardo Bovo
r.bovo@greenwich.ac.uk
University of Greenwich
London, UK

Shirin Hajahmadi
shirin.hajahmadi2@unibo.it
University of Bologna
Bologna, Italy

Giacomo Vallasciani
giacomo.vallasciani2@unibo.it
University of Bologna
Bologna, Italy

Matteo Martinoni
m.martinoni@isnb.it
Institute of Neurological Sciences
Bologna, Italy

Gustavo Marfia
gustavo.marfia@unibo.it
University of Bologna
Bologna, Italy

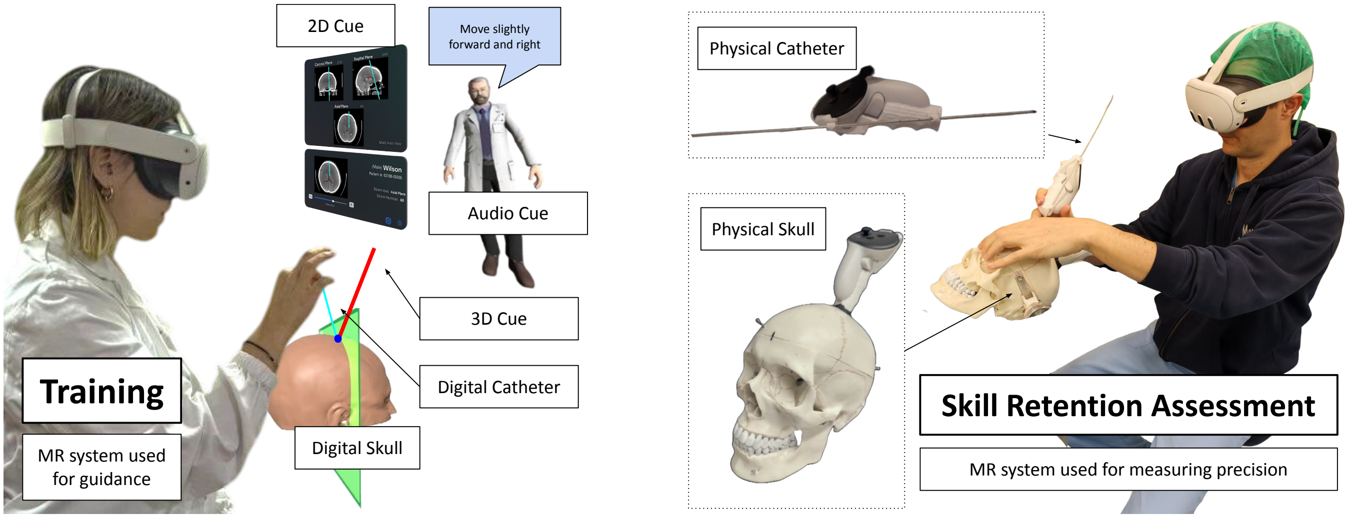


Figure 1: Mixed Reality system for training and skill retention assessment. On the left, during the Training phase, the user is guided by 2D, 3D, and audio cues to insert a digital catheter into a digital skull. On the right, during the Skill Retention Assessment phase, the user performs the same task on a physical skull using a physical catheter, however, unlike in training, the MR system is used solely to measure precision, without providing any guidance.

Abstract

External Ventricular Drain (EVD) placement is a complex neurosurgical task that requires identifying a target point within the brain and accurately positioning a catheter at the appropriate angle. While Mixed Reality (MR) technologies have seen limited adoption in the operating room, they offer significant potential for developing training systems that enhance skill acquisition and retention in unaided conditions. A current gap in research concerns the effectiveness of multimodal guidance systems that incorporate both visual and audio-based MR cues. In this paper, we present an MR-based simulator for EVD placement training and evaluate the impact of three MR-guided training modalities: (1) a baseline condition using only 2D CT scans and a 2D catheter projection; (2) a visual guidance modality incorporating a 3D trajectory overlay; and (3) an embodied-audio guidance modality featuring a virtual agent delivering spoken instructions and feedback. Participants underwent a

---

[*] Corresponding author.
This research was approved by the University of Bologna Ethics Committee (Approval number 0159749) on June 23, 2022. The study adhered to all ethical guidelines, and informed consent was obtained from all participants prior to their involvement.

VRST '25, Montreal, QC, Canada

ACM ISBN 979-8-4007-2118-2/2025/11
https://doi.org/10.1145/3756884.3766034

digital training phase using one of the three modalities, followed by an unaided EVD placement on a physical phantom with a real catheter to evaluate skill transfer and retention. Results indicate that both advanced MR modalities significantly improve procedural accuracy, execution speed and receive higher scores in usability and technology acceptance compared to the baseline. Notably, training with 3D visual trajectory guidance led to significantly higher unaided placement accuracy, indicating stronger skill retention. However, multimodal guidance demonstrated equivalent execution speed, while showing a trend toward lower overall cognitive load.

CCS Concepts

• **Computing methodologies → Mixed / augmented reality**.

Keywords

Mixed Reality, Neurosurgical Training, External Ventricular Drainage, Catherer Placement, Visual Guidance, Audio Guidance, Multimodal Guidance, Skill Retention.

ACM Reference Format:
Pasquale Cascarano, Andrea Loretti, Luca Zanuttini, Daniele Giunchi, Riccardo Bovo, Shirin Hajahmadi, Giacomo Vallasciani, Matteo Martinoni, and Gustavo Marfia. 2025. See It and Hear It: Multimodal Guidance in MR-Based Neurosurgical Simulation for Skill Retention. In *31st ACM Symposium on Virtual Reality Software and Technology (VRST '25), November 12–14, 2025, Montreal, QC, Canada.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3756884.3766034

1 Introduction

Neurosurgery has been at the forefront of adopting Mixed Reality (MR) technologies, offering a promising solution by providing risk-free simulators for medical training [5, 14, 23]. These simulators are characterized by the ability to recreate realistic anatomical environments and procedural workflows through highly detailed 3D models and interactive scenarios. By integrating visual and auditory guidances and feedback, they allow learners to practice critical decision-making and manual skills in a controlled setting [70, 31]. Due to the restricted use of MR headsets in the operating room [32], it is crucial that these training systems foster strong skill transfer and retention in unaided scenarios. This ensures that learners are able to perform surgical tasks accurately without relying on MR-based guidance during the procedures.

External Ventricular Drain (EVD) placement is a critical neurosurgical procedure that involves inserting a catheter into the brain's ventricular system to drain fluid and relieve intracranial pressure [24]. Following identification of the entry point, the surgeon advances the catheter toward the operation draining target, carefully setting the trajectory and depth to minimize the risk of complications. The task requires a high level of spatial awareness, anatomical knowledge, and manual precision [24]. Neuronavigation systems align preoperative imaging with the patient's anatomy using tracking technologies. Although neuronavigation systems can support accurate EVD placement [27, 1], they are not always available in all operating rooms and are often inaccessible in emergency settings [1]. As a result, freehand placement without any visual guidance remains the only practical option in many cases. Freehand placement is typically performed at the bedside and leads to catheter misplacement in up to 60% of cases, increasing the risk of patient complications [2].

Neurosurgical residents typically undergo standardized training under the supervision of experienced surgeons, often with the support of neuronavigation systems and following the "see one, do one, teach one" approach [33]. During these sessions, trainees receive real-time guidance through visual cues, such as observing the expert's catheter alignment or tracking its trajectory on the neuronavigation display, and auditory feedback from the supervising surgeon, which reinforces critical decisions and intraoperative adjustments. However, a variety of factors increasingly prevent intraoperative support, leading to reduced training opportunities. These include stricter regulations on resident working hours and growing concerns about patient safety [3, 52]. This highlights the need for MR training systems designed to improve residents' ability to perform freehand EVD placement without any form of intraoperative guidance.

The design of the MR interfaces and the choice of the form of guidances in training systems is a crucial factor that can significantly impact skill transfer and retention. When users are presented with overly or poorly organized information, performance can suffer, and error rates may increase [9]. This is especially relevant in procedures that require high manual precision, such as freehand EVD placement. Visual cues, due to their rapid processing speed, are the most commonly used and include elements such as virtual 3D objects, strategic waypoints, animated overlays, and virtual hands [12, 37]. Auditory cues, although slower to process, can enhance spatial memory and are particularly useful in complex or noisy environments [21]. The use of virtual avatars in relation to audio cues has been shown to improve communication and mental immersion [8, 37].

The existing literature on MR-assisted EVD placement focuses primarily on systems intended for use in the operating room. In this context, MR-based solutions are preferred over Virtual Reality (VR) or non-immersive methods, as they integrate physical and virtual elements. Unlike VR, which isolates users from real instruments, MR enables manipulation of actual catheters and skull phantoms with spatially aligned digital guidance. However,previous studies have not specifically addressed the development of training tools specifically designed to improve performance in freehand procedures without MR-based assistance [11]. Furthermore, another research gap concerns the identification of which interface and guidance modalities most effectively support skill transfer and retention, particularly the role of multimodal interfaces that integrate visual cues with avatar-driven auditory feedback.

To address these gaps, the present study introduces an MR simulator for EVD placement and evaluates three distinct learning modalities. The first modality integrates axial, coronal, and sagittal CT slices into the user's field of view, along with a real-time projection of a virtual catheter overlaid on the imaging planes to support trajectory planning (Multiplanar modality). The second modality builds upon such a modality by introducing a 3D animated guide that demonstrates the correct spatial placement of the virtual catheter (Multiplanar–Spatial modality). The third modality extends the Multiplanar mode by incorporating multimodal cues with a virtual avatar that provides auditory feedback to guide catheter positioning (Multiplanar–Avatar modality). The MR simulator implemented is designed to be portable and low-cost. Unlike traditional MR systems for EVD placement that rely on complex tracking infrastructure, such as multiple infrared cameras and fixed setups [11], our solution does not require external tracking hardware. This not only significantly reduces the cost but also allows for flexible deployment, making it accessible in both academic and clinical settings.

The study included 40 medical residents with basic neuroimaging knowledge but without EVD placement experience. After completing the training session, all participants undergo a testing phase in which they are asked to perform the EVD procedure on a physical phantom skull using a real catheter, without any MR-based visual and/or audio assistance. A separate control group, which does not receive any training, completes only the testing phase. This allows for a comparative evaluation of performance and skill retention in the different conditions.

In this paper, our contributions are threefold. (i) System Design: we introduce a self-contained, mobile MR platform deployable on consumer-grade headsets (Meta Quest 3), enabling fully untethered training without external tracking systems or computing infrastructure. (ii) Avatar-Guided Training: we propose an instructional modality using a virtual avatar to deliver real-time, context-aware verbal feedback, simulating a clinical mentor. (iii) Skill Retention Study: we conduct a controlled user study (N = 40) assessing short-term EVD skill retention across three training modalities and a control group.

2 Related Work

This section reviews prior work on multisensory cue integration in MR and mechanisms of short-term skill retention, followed by literature on MR systems for EVD placement.

2.1 Multimodal Cues in MR precision tasks

In simulations requiring manual precision and trajectory memorization, managing sensory cues is crucial, as too many or unclear cues can reduce accuracy and raise error rates [9]. The existing literature [68, 67, 12, 43] highlights the importance of two categories of cues: visual (e.g., 3D objects, icons, animations, highlights, interactive holographic content) and auditory (e.g., step-by-step instructions and feedback, confirmation sounds, error alerts) [68, 26]. Avatars can serve as a form of multimodal guidance: assigning a voice to an embodied agent enhances communication and immersion compared to audio-only instructions [8], while audio remains valuable for preserving real-time interaction.

Building on prior work in surgical training [68, 26], our system integrates visual cues such as highlights, interactive 3D objects, and textual or spatial feedback. Differently from most approaches, we specifically investigate multimodality by introducing an avatar that provides spoken guidance.

A meta-analysis of 43 studies [9] showed that adding auditory to visual feedback improves reaction time and performance under normal load, but not error rates, and may hinder high-precision tasks. More recent evidence [48], however, indicates that combining verbal and visual feedback supports tasks requiring fine manual accuracy. The effectiveness of multimodal feedback thus remains unclear and, in surgical contexts such as EVD placement, largely unexplored.

2.2 Multimodal Cues and Skill Retention

The acquisition and retention of skills is a well-established field of study, grounded in foundational theories of learning stages [17, 38, 68, 48, 36]. The literature show that distributed practice yields more durable retention than massed practice, and that feedback is most effective when delivered progressively and with reduced frequency [68, 58].

In [38] the authors show that even skills acquired effectively through simulation-based training are subject to significant decay, if not reinforced, in particular complex tasks like surgery. This underscores that while simulation is a powerful tool for initial skill acquisition, regular refresher intervals are needed to ensure sustained competence [68]. Recent works [28, 58] show that modern training technologies, such as MR, provide immersive and adaptive learning environments that can mitigate skill decay.

To address these challenges, we developed a MR training system that is both portable and cost-effective. Unlike many existing training platforms [11], our system does not rely on cumbersome and expensive tracking hardware. This design not only enables effective training but also facilitates frequent and accessible practice in both academic and clinical settings.

For memorization tasks, visual cues often yield better long-term retention than auditory ones [41], though multimodal cues can enhance attention and retention [30]. In high-precision manual tasks, visual cues aid short-term retention [68], but the role of multimodal feedback remains unclear. Our study addresses this gap by examining their impact on short-term skill retention in EVD placement.

2.3 EVD placement in MR environments

Most MR research on EVD placement has focused on developing guidance tools to assist surgeons during the procedure [11]. In prior studies [63, 20, 40, 25, 65, 18], users performed the task with MR-based guidance, with no evaluation under unaided conditions. Our work differs by specifically examining user performance without MR assistance.

Visual cues in MR systems provide real-time guidance on catheter trajectory, angle, and depth [16, 7, 18, 59], often through virtual markers, anatomical overlays, diagnostic images, and feedback panels [25]. Furthermore,[19] triggers cues based on gaze tracking, while[7] found 3D aids more effective than 2D.Auditory cues (e.g. verbal instructions or sound effects) are far less common; in one of the few studies using them [13], surgeons emphasized their value for enhancing MR guidance in brain tumor resection.

Unlike previous work, our approach is the first in the EVD domain to implement a multimodal interface that combines visual cues with a voice-enabled avatar, delivering real-time verbal feedback to actively support the user during the procedure.

Most MR systems for EVD placement use Microsoft HoloLens with registration and outside-in tracking [18, 60, 20]. In contrast, our approach leverages Meta Quest 3's inside-out tracking, enabling autonomous localization without external references, greater portability, a wider field of view, improved rendering, and lower cost.

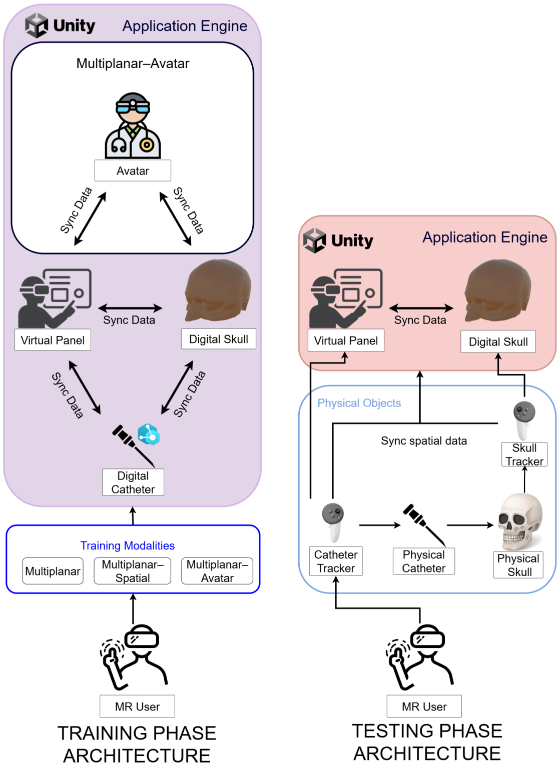


Figure 2: Training and Testing Architectures.

Validation of these systems has been conducted on phantom heads [63, 20] and 3D-printed skulls [16], always with medical staff operating under MR-guided assistance. A key innovation of our study lies in evaluating training impact on skill retention in the absence of MR guidance.

3 Motivation and Design

In this section, we provide a detailed description of the training and testing systems and the design and objective of the present study. The system design and study protocol were developed in consultation with experienced neurosurgeons, who provided feedback on the clinical plausibility of the research workflow.

3.1 Training system design and architecture

The training system (as shown Figure 2, left panel) is a MR application and its architecture comprises two core components: the Application Engine and the Training Modalities. It was built with Unity3D (v2022.3.10f1) and the Meta SDK (v62.0.0) [47], and it was installed on the Meta Quest 3 headset, where interaction is provided through hand-tracking.

3.1.1 The Application Engine. The Virtual Panel (VP), Digital Skull (DS), and Digital Catheter (DC) are the three virtual components whose data visualization and synchronization are handled by the Application Engine. A real-time, bidirectional synchronization system is used to synchronize these parts. In addition to reacting to user input, every piece actively updates the others.

As an in-world interface, the VP gives the user real-time feedback and data visualization. Depending on the specific Training Modality interface chosen, different content will be shown. In general, important data including brain CT scans, trajectory angles, and insertion depth are displayed on the VP (see Figure 3). The design adopts a familiar layout [54], reflecting interfaces commonly encountered by professionals in the field [50]. Additionally, the VP includes green virtual panels that slice through the DS, displaying corresponding CT scan sections directly on the VP.

The DS represents the skull of an anonymous patient, allowing users to explore key anatomical landmarks such as Kocher's point and the foramen of Monro (blue and red points in Figure 4, respectively), which serve as the entry and target sites for the EVD procedure [24]. We used 3DSlicer software [50] to reconstruct the DS using a brain CT images dataset. The final 3D model was calibrated in Unity for accurate DICOM alignment and downsampled in Blender [49] to ensure hardware compatibility.

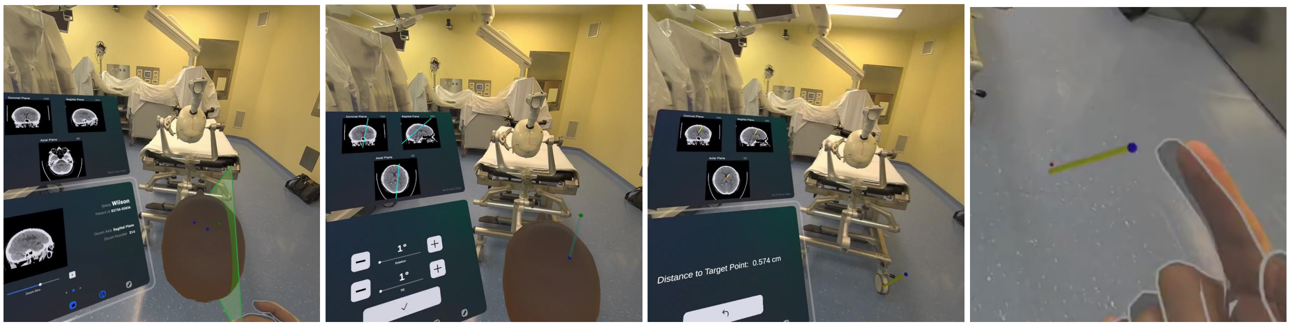
DC is the virtual surgical tool (blue stick in Figure 4). To improve the user's handling, a green sphere is fixed at the tip of the DC.

The user is guided through the key steps of EVD placement via an interactive workflow. Initially, the user selects one of three possible entry points on the DS, ideally identifying Kocher's point. If an incorrect point is chosen, the VP provides error feedback, encouraging anatomical recognition as required in real surgical settings. Once the correct entry point is confirmed, the virtual simulation begins. First, the target point is identified within the CT scans. The user can either drag green panels that appear on the DS or use the buttons on the VP to scroll the CT pictures in order to find the target spot indicated with a red dot (see Figure 3a). Then, the DC trajectory is defined using three degrees of freedom: Tilt, Rotation, and Depth. Tilt and Rotation can be adjusted via the VP or directly in 3D space, while Depth, requiring millimeter precision, is set exclusively through VP sliders [46] (see Figure 3b).

After confirming the trajectory, the system evaluates the trial and provides post-trial feedback, including a textual summary of the Euclidean distance from the target point and a 2D visualization showing the DC projections on the CT images (see Figure 3c). A 3D feedback is also provided, displaying both the target point and the DC (see Figure 3d). The user can click the "Retry" button on the VP to repeat the process, supporting iterative practice and gradual skill refinement. The user can restart the process by clicking "Retry" on the VP, enabling iterative practice and progressive skill improvement.

3.1.2 The Training Modalities. Multiplanar (MP), Multiplanar-Spatial (MP-S) and Multiplanar-Avatar (MP-A) are the three training conditions offered by the implemented system (see Figure 4).

Multiplanar Modality. The design of a neuron navigation system served as the model for the interface [27, 1]. The anatomical planes are shown simultaneously in the VP. The purpose of this interface is to facilitate the association between the 3D DS and the 2D CT slices on the VP [46]. The selected CT slices are retained as visual references in subsequent phases, providing contextual



(a) (b) (c) (d)

Figure 3: Key Steps and Feedback in MR Training. (a) User finds the target point on the DS. (b) DC trajectory is adjusted via VP sliders (Tilt, Rotation, Depth).(c) Post-trial feedback includes target distance and 2D CT projections. (d) 3D view displays DC (yellow stick), entry point (blue), and target (red).

support as the user proceeds to place the target following identification of the red anatomical marker. The user then chooses the parameters for rotation, tilt, and depth. This interface overlays a real-time 2D projection of the digital catheter onto the CT scan images, enabling continuous visual alignment during trajectory adjustment (see Figure 4).

Multiplanar-Spatial Modality. This modality extends the MP modality (see Figure 4). To help users select Tilt and Rotation settings, a 3D trajectory guide is incorporated as an additional visual aid. The ideal 3D insertion path, including the ideal angle, is graphically indicated by this red guide line. This enables participants to adhere to a predetermined trajectory that, if correctly followed, would result in a "perfect" placement. To initiate an animation, the user can press a specific button on the VP. This context-aware animation visually illustrates the movement necessary for the DC to align with the optimal insertion angle, starting from the device's current orientation. The VP's depth slider is then used to choose the depth component. The VP, in contrast to the earlier modalities, displays to the user the optimal depth component value to approach the inner target point.

Multiplanar-Avatar Modality. This modality builds on the MP interface (see Figure 4) by adding a 3D avatar placed in front of the user. The model, available in Sketchfab [1], was chosen as a balanced option between realism, resembling a doctor, and performance, with a lightweight 4MB size suitable for smooth rendering on the Quest 3. We included an avatar to enhance the sense of human presence and support user engagement during training. Prior work has shown that anthropomorphic digital agents can positively influence user trust and willingness to follow guidance in healthcare settings [57], making them particularly suited for trainer-trainee scenarios.

The Avatar was rigged and animated through Mixamo, a custom lip sync rigging and animation was made in Blender. All the animations and audio syncing was managed by the unity Animator tool and custom C# scripting. The avatar has awareness of what the user is doing and has the ability to speak through a series of pre-generated audio files, that will be used in the right context. The audio files were made by using an open source state-of-the-art generative speech model: "orpheus-3b-ft.gguf" with the use of tools such as Orpheus FASTAPI and LM-Studio. The choice to utilize pre-generated audio files, rather than a text-to-speech system in conjunction with an LLM was made because the Meta Quest 3's processing capabilities are not robust enough to run an LLM live and simultaneously maintain the required performance for the XR environment. Furthermore, we opted against relying on external asynchronous computation due to critical concerns regarding network security, especially when operating outside a secure clinical structure, and the potential for unacceptable latency impacting the real-time MR experience. The avatar helps to outline the steps of the operation and suggests adjustments in case the user is making major misjudgments in trajectory or depth choosing.

3.2 Testing system design and architecture

The architecture of the MR testing system is shown in Figure 2 (right panel). Unlike the training system, in this setup the MR interface does not provide any guidance or feedback to the user. This design ensures that the system objectively assesses each user's ability to perform the procedure independently. Another difference is is the use of real physical components: a graduated catheter, referred as Physical Catherer (PC), and a silicone-based anatomical skull model referred to as Physical Skull (PS). To simulate the tactile resistance of brain tissue, an agar-based substance [62] was inserted into the cranial cavity of the PS. Additionally, since Kocher's point is fixed during testing, the PS is pre-perforated at the designated entry site. The MR system is used solely to measure user performance, thus minimizing human measurement error and enabling rapid and reliable assessment of each trial. Additionally, it is used to track the positions of both the PC and PS.

Due to Meta's limitations on direct camera access within the Meta Quest 3, it was difficult to track the PS and the PC accurately. This made it difficult to use computer vision techniques like those described in [53]. To overcome these limitations, we developed a custom tracking solution using Meta Quest 3 controllers, which provide a real-time and efficient spatial tracking via inertial sensors and infrared LEDs. The controllers were mounted on the PC and the PS with 3D-printed holders.

---

[1] https://sketchfab.com/3d-models/doctor-sketchfab-weekly-13-mar23-9c89a438a5e940e59a0f9a07c22d6ade

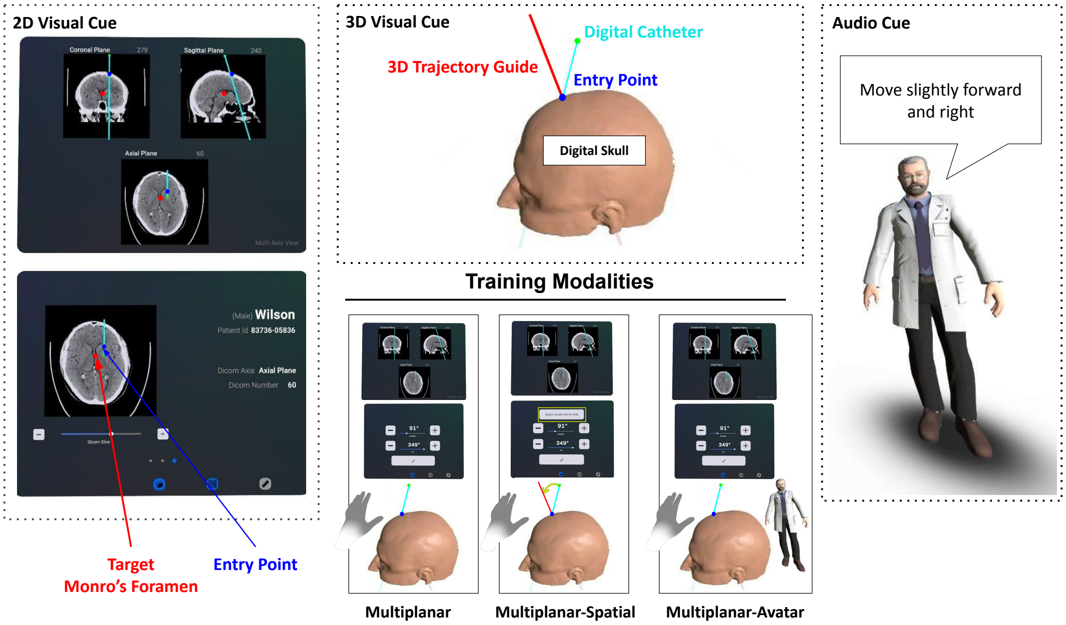


Figure 4: Overview of the Three Training Modalities. Left: 2D Visual Cue – Orthogonal CT slices (coronal, sagittal, axial) display the entry point (blue) and target (Monro's foramen, red), with a real-time 2D overlay of the digital catheter (DC). Center-top: 3D Visual Cue – A 3D trajectory guide (red) helps align the DC (cyan) with the optimal path. Audio Cue – A 3D animated avatar provides contextual spoken feedback. Center-Bottom: Training Modalities (a) Multiplanar: 2D Visual Cue (b) Multiplanar-Spatial: 2D + 3D Visual Cues (c) Multiplanar-Avatar: 2D Visual + Audio Cues.

The 3D-printed mount was designed to replicate the grip and handling of a neurosurgical catheter in clinical practice. The controller functions as a tracker, while the user manually operates the tracked PC during insertion. Pressing the "A" button on the controller ends the attempt, indicating that the user believes the target point has been reached.

The Application Engine is in charge of synchronizing and visualizing the VP and DS in real time. In order to offer consistent reference points created during the training phase, the same DS model utilized in the training system was rendered in the MR testing environment and registered on the PS. The insertion depth is also shown in real-time on the VP to assist the user during the EVD placement, as the graded catheter is hard to see properly while wearing the headset.

3.3 Study design

This section describes the study's experimental design, methodology, participants, and measurements.

3.3.1 Study Objective Question. We conducted a between-subjects experiment to address the following Research Questions (RQs):

RQ1: *How do usability, cognitive workload, and technology acceptance differ between visual-only and visual–auditory guidance modalities during EVD placement training?*

RQ2: *How do users' procedural accuracy and execution time differ between visual-only guidance conditions (Multiplanar and Multiplanar–Spatial) and the multimodal visual–auditory condition (Multiplanar–Avatar) during EVD placement training?*

RQ3: What is the effect of different MR training modalities, ranging from visual-only guidance (Multiplanar and Multiplanar–Spatial) to multimodal visual–auditory guidance (Multiplanar–Avatar), on skill retention in EVD placement, as measured by procedural accuracy and execution time when no guidance is available during testing?

3.3.2 Procedure and Task. Participants were randomly assigned to either the Experimental or Control Group. Experimental Group participants were further randomly assigned to one of three training modalities (MP, MP-S or MP-A). All of them completed a demographic and background questionnaire about age, gender, education, prior EVD knowledge, and exposure to immersive technologies. The study consisted of three phases: Familiarization, Training, and Testing. The Control Group completed only the Familiarization and Testing phases. During Familiarization, an experienced operator introduced EVD theory and practice [22], and both groups were trained on using the Meta Quest 3 headset to avoid bias from first-time MR exposure. In the Training Phase, Experimental Group participants completed five EVD placements using their assigned modality. In the Testing Phase, all participants performed five EVD insertions without MR guidance. The MR system was used solely for data collection. Experimental Group participants faced the same target anatomy as during training [24, 22]. Control Group participants could view CT scans using 3D Slicer on a standard monitor before the test. All participants completed a post-experiment questionnaire.

3.3.3 *Participants.* A total of 40 participants took part in the main experiment, ranging in age from 21 to 37 years (M = 27.75, SD = 3.59), including 19 females and 21 males. This sample size aligns with prior research in the field [11, 10]. Results from the background questionnaire indicated that all participants had a medical background, with very limited experience in EVD placement (M = 1.03, SD = 0.16), very limited familiarity with neuronavigation systems (M = 1.18, SD = 0.45), and limited experience in interpreting neuroimaging such as CT scans (M = 1.83, SD = 1.20). More precisely, the study enrolled 32 clinicians (residents and practicing specialists) and 8 advanced medical students (4th–6th year). In terms of handedness, 36 participants were right-handed, 3 left-handed, and 1 ambidextrous. Regarding familiarity with immersive technologies, participants had limited familiarity with Augmented, Mixed, or Virtual Reality (M = 1.75, SD = 0.98). Specifically, the vast majority (35 out of 40) had no prior experience using the Meta Quest 3 device employed in this study, and limited experience with other head-mounted displays (M = 1.63, SD = 0.93). The 40 participants were distributed as follows: Control Group (n = 10), Multiplanar (n = 10), Multiplanar-Spatial (n = 10), and Multiplanar-Avatar (n = 10).

3.3.4 *Measurements.* To evaluate the impact of the MR training system on procedural efficiency and accuracy, both objective and subjective measures were collected.

During the Training Phase, the system recorded: (a) number of attempts to locate the correct entry point; (b) 3D coordinates of operation points; and (c) Total Time (TT) for each insertion. In the Testing Phase, only (b) and (c) were recorded, as the entry point was predefined. Error Score (ES) was calculated as the Euclidean distance between the operation and target points. Average Error Score (AES) and Average Total Time (ATT) were computed across all trials and participants in both phases.

The Experimental Group completed the System Usability Scale (SUS, Likert 1–5) for usability and learnability [39], and the NASA-TLX [29] (Likert 1–10) for workload. They also filled out the Technology Acceptance Model (TAM) questionnaire [15] (Likert 1–7), assessing perceived usefulness and ease of use.

In addition, we designed a custom Final Test Assessment (FTA) to capture participants' subjective perception of task difficulty during the unaided testing phase. The FTA included five 5-point Likert items covering perceived difficulty of each insertion. For example, one item asked: "How difficult was insertion Xth?".

3.3.5 *Tools for Statistical Analysis.* The analysis aimed to compare the Experimental and Control Groups, as well as training modalities within the Experimental Group. We first assessed normality using the Shapiro-Wilk test [51]. If normality was confirmed, a one-way ANOVA was followed by pairwise independent samples t-tests [34]. If normality was not met, we applied the Mann-Whitney U test [45] and the Kruskal-Wallis H-test [44]. A two-sided TOST procedure was used to assess statistical equivalence [35]. Finally, the ANCOVA [61] was used to analyze the effect of training modality on performance.

4 Results

This section presents the results of the statistical analysis for subjective and objective measures. In Table 1 we report the mean (M) and standard deviation (SD) for SUS, NASA-TLX, TAM and FTA. In Table 2 we report M and SD of the AES and ATT for the training and testing phases. A summary of significant equivalences of differences is reported in Table 3.

SUS questionnaire. All the modalities were given high Overall values (>70), indicating that the training system reached acceptable usability for all learning modalities [4]. Normality was confirmed for each of the constructs across all the learning modalities. The ANOVA test did not show any significant differences among the three modalities for Usability, Learnability and Overall. A pair-wise TOST analysis revealed significant equivalence in terms of Usability between the MP and MP-S modalities, and between the MP and MP-A modalities, but not between the MP-S and MP-A modalities.

NASA questionnaire. Normality was confirmed only for the Temporal dimension, with the Mental, Physical, Performance, Effort, and Frustration dimensions failing to meet the normality assumption. A one-way ANOVA on the Temporal dimension and Kruskal-Wallis H tests on the remaining dimensions did not reveal any significant differences among the three learning modalities for any of the NASA-TLX constructs. Finally, the TOST procedure did not reveal any significant equivalences between the MP, MP-S, or MP-A modalities for any of the workload dimensions.

TAM questionnaire. The assumption of normality was not met for the Usefulness construct, while it was confirmed for the Ease of Use construct across all modalities. A Kruskal–Wallis H test did not highlight significant differences for Usefulness or for Ease of Use among the three learning modalities. A pair-wise TOST test showed significant equivalence for the Ease of Use construct across all comparisons: MP vs. MP-S, MP vs. MP-A, and MP-S vs. MP-A. Moreover, for the Usefulness construct, significant equivalence was found between the MP-S and MP-A modalities.

FTA questionnaire. A Kruskal–Wallis H test conducted across all groups, including the control group, did not reveal any statistically significant differences in the overall perceived difficulty. The TOST analysis found statistically significant equivalence between MP and MP-A modality.

Table 1: Summary of Subjective Measures.

Variable	Construct	MP (M, SD)	MP-S (M, SD)	MP-A (M, SD)	Control (M, SD)
SUS	Usability	4.17, 0.35	4.03, 0.61	4.38, 0.43	–
	Learnability	3.90, 0.66	3.60, 0.74	3.45, 1.12	–
	Total	90.50, 9.41	86.00, 15.19	92.25, 11.27	–
NASA-TLX	Mental	5.90, 1.29	6.40, 1.43	6.30, 2.06	–
	Physical	3.00, 1.76	4.10, 1.60	4.10, 1.60	–
	Temporal	3.70, 1.16	3.50, 1.35	3.60, 1.71	–
	Performance	7.30, 1.16	7.70, 0.82	7.20, 1.69	–
	Effort	6.30, 2.00	5.70, 1.70	5.80, 2.10	–
	Frustration	3.00, 1.56	3.20, 2.25	2.50, 1.18	–
	Average	4.87, 0.88	5.10, 1.05	4.92, 1.13	–
TAM	Usefulness	4.26, 0.45	4.46, 0.55	4.43, 0.61	–
	Ease of Use	3.04, 0.32	3.12, 0.32	3.16, 0.26	–
FTA	Average	2.90, 0.38	3.04, 0.78	2.76, 0.74	2.70, 0.94

Accuracy analysis. The results of the Shapiro-Wilk test indicated that many of the variables violated the assumption of normality. The Kruskal–Wallis H-test revealed statistically significant differences between the three learning modalities (MP, MP-S and MP-A) during the Training phase. The Mann–Whitney U test showed that the MP-S modality resulted in significantly higher precision compared to both the MP and the MP-A modalities, suggesting that the 3D spatial guidance was the most effective during training.

For what concerns the Testing Phase, the four modalities (including the Control group) showed a significant difference. Pairwise comparisons using the Mann-Whitney U test showed that the MP-S modality was significantly more precise than all other groups, namely MP, MP-A and Control. Furthermore, both the MP-A and MP-S modalities were significantly more precise than the MP and Control groups.

Finally, we conducted an ANCOVA test using the number of training attempts as a covariate, the training modality as a between-subject factor and precision during the Testing Phase as the dependent variable. The analysis revealed a significant main effect of training modalities on operation precision, indicating that the type of training used had an impact on precision, independent of the number of attempts.

4.0.1 Time Analysis. A Kruskal-Wallis H test was conducted to compare the ATT on the training phase across the three learning modes (MP, MP-S and MP-A). The test was statistically significant, indicating that the total time required to complete the operation differed across modes. In particular, MP-S group was significantly faster than the MP-A group.

A Kruskal-Wallis H test was conducted to evaluate differences in the ATT for the Testing Phase across the MP, MP-S, MP-A, and Control groups. The analysis revealed a statistically significant difference within the data, suggesting that the type of aid used had a measurable effect on the operation time. Specifically, the MP group was significantly slower than the Control group. Moreover, MP, MP–S and MP–A are significantly equivalent.

Finally, we conducted an ANCOVA test using the number of training attempts as a covariate, the training modality as a between-subject factor, and the Total Time during the Testing Phase as the dependent variable. The training modality did not show a significant main effect on operation time, while the number of training attempts did have a significant impact.

Table 2: Summary of the AES [cm] and ATT [s].

Phase	Metric	MP	MP-S	MP-A	Control
Training	AES [cm]	0.60, 0.50	0.42, 0.27	0.59, 0.33	–
	ATT [s]	139.94, 78.56	125.61, 93.24	158.63, 88.84	–
Testing	AES [cm]	1.93, 1.76	1.28, 0.45	1.60, 0.79	2.29, 0.73
	ATT [s]	48.00, 21.75	46.60, 35.28	45.60, 32.86	33.47, 15.71

5 Discussion

In this section, we seek answers to the RQs.

RQ1 All the three training modes (MP, MP-S and MP-A) were found to be highly acceptable and usable to the participants. SUS scores were always more than the usual acceptance level for usability of 70 [4] without any significant differences between the groups and showing significant equivalence between MP and the other two modalities. These findings show that even highly advanced visualizations, such as animated 3D trajectory guidance or animated avatar, do not negatively impact perceived usability. Similarly, the TAM measures revealed high scores for Perceived Usefulness and Perceived Ease of Use for all modalities, and statistical equivalence between all pairs was determined based on Ease of Use dimension. Concerning the Usefulness, the MP-S and MP-A modalities were perceived equally, both with higher scores with respect to the MP. This is consistent with prior research showing that immersive training systems can maintain high usability and high perceived usefulness when interface design is intuitive and familiar [42, 46, 66]. Finally, the NASA-TLX did not show significant differences. Our findings seem to align with existing literature indicating that audio aids can be ineffective to cognitive load during high-precision and cognitive demanding tasks [9]. However, a qualitative analysis showed an interesting trend: the higher frustration score for the MP-S mode compared with the other two modalities, with neither statistically significant differences nor similarities. Interestingly, during the Testing Phase, MP and MP-A were perceived similarly, both receiving lower FTA scores than MP-S. Since the FTA measures perceived task complexity, this suggests that participants may have found MP-S more challenging, thus confirming the trend. This observations require further investigations in future studies.

RQ2 During the training phase, no significant differences or equivalence were found in catheter placement precision or execution time when comparing MP and MP-A. Conversely, participants trained with the MP-S modality achieved significant higher accuracy compared to the other groups and lower execution time compared to MP-A. These findings align with prior research suggesting that multimodal systems cannot offer performance advantages, particularly for tasks with high cognitive demands [9]. Audio guidance, in particular, can reduce positioning precision and slow down execution [70]. Furthermore, these results are consistent with prior studies demonstrating that visual overlays, particularly when combining 2D imaging with 3D spatial cues, can improve performance in image-guided neurosurgical procedures [2, 11]. The ANCOVA test controlling for the number of attempts confirmed that the improvements during the training were due to the training modalities. This highlights the importance of studying the impact of different form of guidances [11, 9].

RQ3 The testing phase was specifically designed to evaluate short-term skill retention by removing all MR-based visual-audio aids and requiring participants to perform EVD placement using only physical instruments, thereby ensuring that performance reflected independent execution rather than reliance on guidance. The results show that prior exposure to training, regardless of the modality, led to a substantial improvement in unaided performance. Participants trained with MP-S retained the lowest error in terms of AES during the Testing Phase, outperforming also the Control group. This aligns with existing literature highliting that 3D visual cues can be effective for skill retention [68]. All the groups showed significant statistical difference in terms of accuracy especially if compared to the Control group with a significative lower error rate. This aligns with existing research highlighting the role of immersive environments in supporting transfer of motor skills [10, 28, 58]. In terms of execution speed, MP, MP-S, and MP-A were statistically equivalent. Interestingly, although auditory cues tended to slow down performance during training, skill retention in unaided testing scenarios resulted in comparable performance between the

Table 3: Summary of Statistical Results. ”=” indicates significant equivalence (TOST); ”<” / ”>” indicate significant differences (ANOVA, Kruskal-Wallis, Mann-Whitney), with ”<” meaning the first modality scored lower. ”None” means no significant result; ”-” indicates no comparison was performed. P-values are reported within round brackets.

	Training Phase					**Testing Phase**		
Comparison	**SUS**	**TAM**		**AES**	**ATT**	**FTA**	**AES**	**ATT**
	(Usability)	**Ease of Use**	**Usefulness**					
MP vs. MP-S	= (p=0.046)	= (p=0.005)	None	> (p=0.046)	None	None	> (p=1.5e-4)	= (p=0.017)
MP vs. MP-A	= (p=0.043)	= (p=0.004)	None	None	None	= (p=0.043)	> (p=0.037)	= (p=0.02)
MP-S vs. MP-A	None	= (p=0.001)	= (p=0.044)	< (p=0.005)	<	None	< (p=0.016)	= (p=0.03)
MP vs. Control	-	-	-	-	-	None	< (p=0.016)	> (p=3.7e-4)
MP-S vs. Control	-	-	-	-	-	None	< (p=1.2e-4)	None
MP-A vs. Control	-	-	-	-	-	None	< (p=2.3e-4)	None

visual-only and multimodal groups. Furthermore, the MP group was significantly slower when compared to the Control Group, as well as the other groups even if not significantly. This may reflect a more cautious, deliberate execution strategy, commonly observed after high-fidelity training, where accuracy is prioritized over speed [6]. Moreover, an ANCOVA controlling for the number of training attempts confirmed a significant effect of training modalities on retention performance, emphasizing that the type of training matters [56]. Although it suggested that while different training modalities significantly impacted skill retention in terms of accuracy, the contribution of different sensory cues is not equal across modalities with better outcomes when considering 3D visual cues. The literature on the impact of multimodal feedback in precision tasks has long been inconclusive. Our findings align more closely with those reported by [9]. The ANCOVA test indicated that execution speed depended primarily on the number of attempts. We attribute this to the training phase, where participants were instructed to focus on accuracy rather than speed. Indeed, all feedback emphasized placement precision, with no reference to execution time. Even when using the MP-S modality, which offered 3D visualization of the trajectory, participants still required a relatively long time on average [64].

6 Limitations and Future Works

This study demonstrated the effectiveness of the training system for EVD placement, however we acknowledge some limitations. The evaluation addressed only short-term retention, as testing followed immediately after training. Future work will examine long-term retention and how training duration affects skill acquisition. Additionally, to better investigate the trend highlighted when comparing cognitive load across modalities, in future studies we will design a within-subjects study, which offers greater sensitivity than the between-group approach used here [55]. Furthermore, the study relied solely on quantitative instruments (SUS, NASA-TLX, and TAM); subsequent studies will integrate qualitative methods (e.g., interviews or think-aloud protocols) to better capture user strategies, challenges, and experiences. Although the Testing Phase involved a physical catheter and skull phantom, we will enhance clinical realism by considering dynamic tissue deformation, bleeding, and the pressures of time-sensitive decision-making. These factors are known to influence user behavior, stress levels, and performance in real-world settings [69]. Moreover, the Training Phase was conducted on a single digital skull model. To address these issues, future developments will include the integration of multiple clinical cases and CT datasets, along with the 3D printing of anatomical models to enhance the realism of both the training and testing phases. Additionally, this study relied on controller-mounted sensors which, while practical and efficient, lack the precision of optical navigation systems used in operating rooms. With Meta’s recent release of headset camera access, we plan to integrate advanced computer vision for markerless tracking. Finally, hardware limitations of the HMD prevented integration of a large language model for the virtual avatar. Future hardware advancements may enable this, allowing a more comprehensive evaluation of the avatar’s impact on training.

7 Conclusions

This study introduced a portable, low-cost Mixed Reality simulator for EVD placement training, examining how visual and audiovisual guidance modalities affect usability, cognitive load, technology acceptance, and procedural accuracy, even without active MR guidance during execution. A user study with 40 participants showed that all three MR training modalities were rated highly in usability, usefulness, and ease of use. NASA-TLX scores revealed no significant workload differences. However, an interesting trend emerged: the multimodal group reported slightly lower frustration and perceived difficulty, suggesting a potential benefit of auditory cues in reducing subjective cognitive load. Moreover, participants trained with combined 2D and 3D visual cues (Multiplanar–Spatial) achieved significantly higher placement accuracy in unaided testing than both the baseline (Multiplanar), the audio-visual group (Multiplanar–Avatar), and the Control group. Statistical analysis confirmed that training modality influenced retention, with 3D visual guidance proving more effective for precision tasks than multimodal feedback. Execution time was longer for all trained groups compared to the Control group. Although the multimodal group required more time during training, all modalities showed equivalent performance in unaided testing. Further analysis indicated that speed improved mainly through repetition, suggesting that participants prioritized accuracy, which is critical in procedures like EVD, over speed.

References

[1] Ahmed AlAzri, Kelvin Mok, Jeffrey Chankowsky, Mohammad Mullah, and Judith Marcoux. 2017. Placement accuracy of external ventricular drain when comparing freehand insertion to neuronavigation guidance in severe traumatic brain injury. *Acta neurochirurgica*, 159, 1399–1411.

[2] Maryam Alizadeh, Yiming Xiao, and Marta Kersten-Oertel. 2024. Virtual and Augmented Reality in Ventriculostomy: A Systematic Review. en. *World Neurosurgery*, 189, (Sept. 2024), 90–107. doi:10.1016/j.wneu.2024.05.151.

[3] Sonal Arora, Louise Hull, Nick Sevdalis, Tanya Tierney, Debra Nestel, Maria Woloshynowych, Ara Darzi, and Roger Kneebone. 2010. Factors compromising safety in surgery: stressful events in the operating room. *The American Journal of Surgery*, 199, 1, 60–65.

[4] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual sus scores mean: adding an adjective rating scale. *Journal of usability studies*, 4, 3, 114–123.

[5] Michael Barrie, Jacob J Socha, Lauren Mansour, and Emily S Patterson. 2019. Mixed reality in medical education: a narrative literature review. In *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care* number 1. Vol. 8. Sage Publications Sage CA: Los Angeles, CA, 28–32.

[6] Anil Ufuk Batmaz, Michel de Mathelin, and Birgitta Dresp-Langley. 2016. Getting nowhere fast: trade-off between speed and precision in training to execute image-guided hand-tool movements. *BMC psychology*, 4, 1–19.

[7] Mohamed Benmahdjoub, Abdullah Thabit, Marie-Lise C. Van Veelen, Wiro J. Niessen, Eppo B. Wolvius, and Theo Van Walsum. 2023. Evaluation of AR visualization approaches for catheter insertion into the ventricle cavity. en. *IEEE Transactions on Visualization and Computer Graphics*, 29, 5, (May 2023), 2434–2445. doi:10.1109/TVCG.2023.3247042.

[8] Nico Brettschneider, Jens Herder, Jeroen de Mooij, and Bektur Ryskeldiev. 2019. *Audio vs. Visual Avatars as Guides in Virtual Environments*. Hochschule Düsseldorf.

[9] Jennifer L Burke, Matthew S Prewett, Ashley A Gray, Liuquin Yang, Frederick RB Stilson, Michael D Coovert, Linda R Elliot, and Elizabeth Redden. 2006. Comparing the effects of visual-auditory and visual-tactile feedback on user performance: a meta-analysis. In *Proceedings of the 8th international conference on Multimodal interfaces*, 108–117.

[10] Isidro III Mendoza Butaslac, Yuichiro Fujimoto, Taishi Sawabe, Masayuki Kanbara, and Hirokazu Kato. 2022. Systematic review of augmented reality training systems. *IEEE Transactions on Visualization and Computer Graphics*, 29, 12, 5062–5082.

[11] Ali Buwaider, Victor Gabriel El-Hajj, Alessandro Iop, Mario Romero, Walter C Jean, Erik Edström, and Adrian Elmi-Terander. 2024. Augmented reality navigation in external ventricular drain insertion—a systematic review and meta-analysis. en. *Virtual Reality*, 28, 3, (July 2024), 141. doi:10.1007/s10055-024-01033-9.

[12] Andrés-Marcelo Calle-Bustos, Jaime Juan, Francisco Abad, Paulo Dias, Magdalena Méndez-López, and M.-Carmen Juan. 2021. Visual vs Auditory Augmented Reality for Indoor Guidance. en. In *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. SCITEPRESS - Science and Technology Publications, Online Streaming, — Select a Country —, 85–95. doi:10.5220/0010317500850095.

[13] João Cartucho, David Shapira, Hutan Ashrafian, and Stamatia Giannarou. 2020. Multimodal mixed reality visualisation for intraoperative surgical guidance. *International journal of computer assisted radiology and surgery*, 15, 819–826.

[14] Kirsty Clarke, Aws Al-Mukhtar, Lina Alim, Amr Nimer, Emma Hatfield, Sanjeeve Sabharwal, and Sadie Syed. 2024. Using mixed reality simulation to improve junior medical trainees' preparedness to manage high-acuity trauma. *BMJ Open Quality*, 13, 2, e002575.

[15] Fred D Davis. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319–340.

[16] César F. Domínguez-Velasco, Isaac E. Tello-Mata, Gerardo Guinto-Nishimura, Adriana Martínez-Hernández, Víctor Alcocer-Barradas, Juan S. Pérez-Lomelí, and Miguel A. Padilla-Castañeda. 2023. Augmented reality simulation as training model of ventricular puncture: Evidence in the improvement of the quality of punctures. en. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 19, 5, (Oct. 2023), e2529. doi:10.1002/rcs.2529.

[17] Alfred O Effenberg, Ursula Fehse, Gerd Schmitz, Bjoern Krueger, and Heinz Mechling. 2016. Movement sonification: effects on motor learning beyond rhythmic adjustments. *Frontiers in neuroscience*, 10, 219.

[18] Sangjun Eom, Seijung Kim, Joshua Jackson, David Sykes, Shervin Rahimpour, and Maria Gorlatova. 2024. Augmented Reality-based Contextual Guidance through Surgical Tool Tracking in Neurosurgery. en. *IEEE Transactions on Visualization and Computer Graphics*, 1–16. doi:10.1109/TVCG.2024.3390680.

[19] Sangjun Eom, Tiffany Ma, Neha Vutakuri, Alexander Du, Zhehan Qu, Joshua Jackson, and Maria Gorlatova. 2024. Did you do well? real-time personalized feedback on catheter placement in augmented reality-assisted neurosurgical training. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 1208–1209.

[20] Sangjun Eom, Tiffany S. Ma, Neha Vutakuri, Tianyi Hu, Aden P. Haskell-Mendoza, David A. W. Sykes, Maria Gorlatova, and Joshua Jackson. 2024. Accuracy of routine external ventricular drain placement following a mixed reality–guided twist-drill craniostomy. en. *Neurosurgical Focus*, 56, 1, (Jan. 2024), E11. doi:10.3171/2023.10.FOCUS23615.

[21] Shuo Feng, Xinjing He, Weiping He, and Mark Billinghurst. 2023. Can you hear it? Stereo sound-assisted guidance in augmented reality assembly. en. *Virtual Reality*, 27, 2, (June 2023), 591–601. Publisher: Springer Science and Business Media LLC. doi:10.1007/s10055-022-00680-0.

[22] Alexander C Flint, Vivek A Rao, Natalie C Renda, Bonnie S Faigeles, Todd E Lasman, and William Sheridan. 2013. A simple protocol to prevent external ventricular drain infections. *Neurosurgery*, 72, 6, 993–999.

[23] Juan P Giraldo et al. 2025. Advances in neurosurgical education: literature review of mixed-reality simulation models and novel mixed-reality spine prototype. *Journal of Neurosurgery: Spine*, 1, aop, 1–14.

[24] 2023. Anatomy and Physiology. en. In *Greenberg's Handbook of Neurosurgery*. (10th Edition ed.). Mark S. Greenberg, (Ed.) Thieme, (Apr. 2023). ISBN: 978-1-68420-504-2. doi:10.1055/b000000751.

[25] Ronny Grunert, Dirk Winkler, Johannes Wach, Fabian Kropla, Sebastian Scholz, Martin Vychopen, and Erdem Güresir. 2024. IMAGINER: improving accuracy with a mixed reality navigation system during placement of external ventricular drains. A feasibility study. en. *Neurosurgical Focus*, 56, 1, (Jan. 2024), E8. doi:10.3171/2023.10.FOCUS23554.

[26] Payal Guha, Jason Lawson, Iona Minty, James Kinross, and Guy Martin. 2023. Can mixed reality technologies teach surgical skills better than traditional methods? A prospective randomised feasibility study. en. *BMC Medical Education*, 23, 1, (Mar. 2023). Publisher: Springer Science and Business Media LLC. doi:10.1186/s12909-023-04122-6.

[27] Hartmut K Gumprecht, Darius C Widenka, and Christianto B Lumenta. 1999. Brainlab vectorvision neuronavigation system: technology and clinical experiences in 131 cases. *Neurosurgery*, 44, 1, 97–104.

[28] Martin Guy, Jean-Marie Normand, Camille Jeunet-Kelway, and Guillaume Moreau. 2023. The sense of embodiment in virtual reality and its assessment methods. *Frontiers in Virtual Reality*, 4, 1141683.

[29] Sandra G. Hart and Lowell E. Staveland. 1988. Development of nasa-tlx (task load index): results of empirical and theoretical research. In *Human Mental Workload*. Advances in Psychology. Vol. 52. Peter A. Hancock and Najmedin Meshkati, (Eds.) North-Holland, 139–183. doi:https://doi.org/10.1016/S0166-4115(08)62386-9.

[30] Guan Huang, Chao Chen, Yahan Tang, Haohua Zhang, Rui Liu, and Li Zhou. 2024. A study on the effect of different channel cues on learning in immersive 360° videos. *Frontiers in Psychology*, 15, 1335022.

[31] Hisashi Ishida, Juan Antonio Barragan, Adnan Munawar, Zhaoshuo Li, Andy Ding, Peter Kazanzides, Danielle Trakimas, Francis X Creighton, and Russell H Taylor. 2023. Improving surgical situational awareness with signed distance field: a pilot study in virtual reality. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 8474–8479.

[32] Talha Khan, Edward G Andrews, and Jacob T Biehl. 2024. Future of surgical mixed reality: cutting-edge or cutting too deep? In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 203–207.

[33] Sandra V Kotsis and Kevin C Chung. 2013. Application of the "see one, do one, teach one" concept in surgical training. *Plastic and reconstructive surgery*, 131, 5, 1194–1201.

[34] Daniël Lakens. 2013. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and anovas. *Frontiers in psychology*, 4, 863.

[35] Daniël Lakens. 2017. Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. *Social psychological and personality science*, 8, 4, 355–362.

[36] Daniel M Landers. 1983. The effects of mental practice on motor skill learning and performance: a meta-analysis. *Journal of sport psychology*, 5, 1.

[37] Shahid Latif, Hagen Tarner, and Fabian Beck. 2022. Talking Realities: Audio Guides in Virtual Reality Visualizations. en. *IEEE Computer Graphics and Applications*, 42, 1, (Jan. 2022), 73–83. Publisher: Institute of Electrical and Electronics Engineers (IEEE). doi:10.1109/mcg.2021.3058129.

[38] Camille Legoux, Richard Gerein, Kathy Boutis, Nicholas Barrowman, and Amy Plint. 2021. Retention of critical procedural skills after simulation training: a systematic review. *AEM Education and Training*, 5, 3, e10536.

[39] James R Lewis and Jeff Sauro. 2009. The factor structure of the system usability scale. In *Human Centered Design: First International Conference, HCD 2009, Held as Part of HCI International 2009, San Diego, CA, USA, July 19-24, 2009 Proceedings 1*. Springer, 94–103.

[40] Ye Li et al. 2019. A wearable mixed-reality holographic computer for guiding external ventricular drain insertion at the bedside. en. *Journal of Neurosurgery*, 131, 5, (Nov. 2019), 1599–1606. doi:10.3171/2018.4.JNS18124.

[41] Katie Lindner, Greta Blosser, and Kris Cunigan. 2009. Visual versus auditory learning and memory recall performance on short-term versus long-term tests. *Modern Psychological Studies*, 15, 1, 6.

[42] Cristian A Linte et al. 2013. On mixed reality environments for minimally invasive therapy guidance: systems architecture, successes and challenges in their implementation from laboratory to clinic. *Computerized Medical Imaging and Graphics*, 37, 2, 83–97.

[43] Abel J Lungu, Wout Swinkels, Luc Claesen, Puxun Tu, Jan Egger, and Xiaojun Chen. 2021. A review on the applications of virtual reality, augmented reality and mixed reality in surgical simulation: an extension to different kinds of surgery. *Expert review of medical devices*, 18, 1, 47–62.

[44] Patrick E McKight and Julius Najab. 2010. Kruskal-wallis test. *The corsini encyclopedia of psychology*, 1–1.

[45] Patrick E McKnight and Julius Najab. 2010. Mann-whitney u test. *The Corsini encyclopedia of psychology*, 1–1.

[46] Daniel Mendes, Fabio Marco Caputo, Andrea Giachetti, Alfredo Ferreira, and Joaquim Jorge. 2019. A survey on 3d virtual object manipulation: from the desktop to immersive virtual environments. In *Computer graphics forum* number 1. Vol. 38. Wiley Online Library, 21–45.

[47] Meta. 2023. *Meta SDK for Unity: Hand tracking documentation.* Retrieved from https://developer.meta.com.

[48] Arsalan Moinuddin, Ashish Goel, and Yashendra Sethi. 2021. The role of augmented feedback on motor learning: a systematic review. *Cureus*, 13, 11.

[49] Tony Mullen. 2011. *Mastering blender*. John Wiley & Sons.

[50] Steve Pieper, Michael Halle, and Ron Kikinis. 2004. 3d slicer. In *2004 2nd IEEE international symposium on biomedical imaging: nano to macro (IEEE Cat No. 04EX821)*. IEEE, 632–635.

[51] Nornadiah Mohd Razali, Yap Bee Wah, et al. 2011. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics*, 2, 1, 21–33.

[52] Richard K Reznick and Helen MacRae. 2006. Teaching surgical skills—changes in the wind. *New England Journal of Medicine*, 355, 25, 2664–2669.

[53] Faith C Robertson et al. 2021. Frameless neuronavigation with computer vision and real-time tracking for bedside external ventricular drain placement: a cadaveric study. *Journal of neurosurgery*, 136, 5, 1475–1484.

[54] Paola Salomoni, Catia Prandi, Marco Roccetti, Lorenzo Casanova, Luca Marchetti, and Gustavo Marfia. 2017. Diegetic user interfaces for virtual environments with hmds: a user experience study with oculus rift. *Journal on Multimodal User Interfaces*, 11, 173–184.

[55] Jeff Sauro and James R Lewis. 2016. *Quantifying the user experience: Practical statistics for user research*. Morgan Kaufmann.

[56] Luc Sels. 2002. 'more is not necessarily better': the relationship between the quantity and quality of training efforts. *International Journal of Human Resource Management*, 13, 8, 1279–1298.

[57] Andrea Sestino and Alfredo D'Angelo. 2023. My doctor is an avatar! the effect of anthropomorphism and emotional receptivity on individuals' intention to use digital-based healthcare services. *Technological Forecasting and Social Change*, 191, 122505.

[58] Lucas Siqueira Rodrigues, Timo Torsten Schmidt, Johann Habakuk Israel, Stefan Zachow, John Nyakatura, and Thomas Kosch. 2024. Comparing the effects of visual, haptic, and visuohaptic encoding on memory retention of digital objects in virtual reality. In *Proceedings of the 13th Nordic Conference on Human-Computer Interaction*, 1–13.

[59] Simon Skyrman et al. 2021. Augmented reality navigation for cranial biopsy and external ventricular drain insertion. en. *Neurosurgical Focus*, 51, 2, (Aug. 2021), E7. doi:10.3171/2021.5.FOCUS20813.

[60] Giuseppe Emmanuele Umana et al. 2021. Multimodal simulation of a novel device for a safe and effective external ventricular drain placement. *Frontiers in neuroscience*, 15, 690705.

[61] Gerard JP Van Breukelen. 2013. Ancova versus change from baseline in nonrandomized studies: the difference. *Multivariate Behavioral Research*, 48, 6, 895–922.

[62] Jesse A. M. Van Doormaal, Tim Fick, Ernest Boskovic, Eelco W. Hoving, Pierre A. J. T. Robe, and Tristan P. C. Van Doormaal. 2025. Development and Validation of a Neurosurgical Phantom for Simulating External Ventricular Drain Placement. en. *Journal of Medical Systems*, 49, 1, (Jan. 2025), 1. doi:10.1007/s10916-024-02133-4.

[63] Frederick Van Gestel et al. 2021. The effect of augmented reality on the accuracy and learning curve of external ventricular drain placement. en. *Neurosurgical Focus*, 51, 2, (Aug. 2021), E8. doi:10.3171/2021.5.FOCUS21215.

[64] Teodóra Vékony, Claire Pleche, Orsolya Pesthy, Karolina Janacsek, and Dezso Nemeth. 2022. Speed and accuracy instructions affect two aspects of skill learning differently. *npj Science of Learning*, 7, 1, 27.

[65] Martin Vychopen, Fabian Kropla, Dirk Winkler, Erdem Güresir, Ronny Grunert, and Johannes Wach. 2025. IMAGINER 2—improving accuracy with augmented realIty navigation system during placement of external ventricular drains over Kaufman's, Keen's, Kocher's and Frazier's point. en. *Frontiers in Surgery*, 11, (Jan. 2025), 1513899. doi:10.3389/fsurg.2024.1513899.

[66] Brody Wells, Nanjia Wang, and Frank Maurer. 2024. Technical challenges in building cross reality applications for analyzing 3d medical images. In *2024 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 152–155.

[67] Fei Wu and Jerald Thomas. 2020. Comparison of Audio and Visual Cues to Support Remote Guidance in Immersive Environments. en.

[68] Benjamin Zendejas, Ryan Brydges, Stanley J Hamstra, and David A Cook. 2013. State of the evidence on simulation-based training for laparoscopic surgery: a systematic review. *Annals of surgery*, 257, 4, 586–593.

[69] Yan-Bang Zhou, Shun-Jie Ruan, Kun Zhang, Qing Bao, and Hong-Zhi Liu. 2024. Time pressure effects on decision-making in intertemporal loss scenarios: an eye-tracking study. *Frontiers in Psychology*, 15, 1451674.

[70] Tim Ziemer. 2023. Three-dimensional sonification as a surgical guidance tool. *Journal on Multimodal User Interfaces*, 17, 4, 253–262.