JOURNAL OF Surgical Simulation

Journal of Surgical Simulation (2018) **5**, 74–86
DOI: https://doi.org/10.1102/2051-7726.2018.0008

ORIGINAL ARTICLE

An examination of factors affecting performance on a patient-specific virtual reality-based ventriculostomy simulator

Ryan Armstrong,[a] Dayna Noltie,[b] Roy Eagleson[a,c] and Sandrine de Ribaupierre[a,b,d,*]

[a]*Biomedical Engineering Graduate Program, Western University, London, Ontario, Canada;* [b]*Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada;* [c]*Department of Electrical Engineering, Western University, London, Ontario, Canada;* [d]*Department of Clinical Neurological Sciences, Western University, London, Ontario, Canada*

[*]*Corresponding author at: London Health Sciences Centre, London, Ontario N6A 5W9, Canada. Email: Sandrine.deRibaupierre@lhsc.on.ca*

Date accepted for publication: 7 September 2018

Abstract

Background: The placement of an external ventricular drain is one of the most commonly performed neurosurgical procedures, and consequently, is an essential skill to be mastered for neurosurgical trainees. In this paper, we describe the development of a simulation environment to train residents on the acquisition of these skills. In addition, the environment was explored as a safe test-bed for the evaluation of novel techniques and factors that influence performance, particularly in regard to anatomic variations that occur clinically. **Methods:** Patient-specific simulation of free-hand ventriculostomies was provided with an integrated mobile augmented reality (AR) image-guidance system. Patient-specific cases represented a progression of burr-hole selection and ventricle targeting tasks. Seven residents and one expert neurosurgeon completed a number of targeting tasks with and without AR guidance. Novel performance metrics are presented and examined. **Results:** A strong correlation was found between expert-scored accuracy and subject experience (r = 0.93 with 95% confidence interval [0.90–0.95]), but this effect was not present with AR guidance. There was a significant difference in performance between cases classified by experts as simple compared with complex (P < 0.05). Expert subjective classification of difficulty was a stronger predictor of the challenge of a case than the chosen anatomic measurements. **Conclusion:** AR guidance showed slight task time improvement, but this was not significant. Objective measures of geometric accuracy show promise, but require further development.

Keywords: *surgical simulation; ventriculostomy; human performance; augmented reality*

Introduction

While certain professions have been using virtual reality (VR) for decades to train students,[1] the majority of surgical skills continue to be trained through the Halstedian paradigm,[2] whereby exposure to operative techniques is gained through apprenticeship programmes. In these programmes, trainees learn surgical skills by practicing on live patients. However, providing a virtual environment for junior trainees to learn basic procedures is becoming recognized as essential to increase patient safety and allow for standardized and accessible training.[3]

The placement of an external ventricular drain (EVD) is a procedure commonly performed in an emergency setting by junior residents to relieve intracranial pressure from hydrocephalus.[4] The insertion of a drain involves choosing an appropriate burr hole on the skull and guiding a catheter, without neuro-navigation, through the burr hole and intermediate brain matter into a lateral ventricle, in order to drain excess cerebrospinal fluid and relieve intracranial pressure. EVDs are used in various contexts that warrant immediate care, such as traumatic brain injury or acute hydrocephalus, whereas other ventriculostomy procedures, such as shunt placement, are intended for long-term management.[4] Although some cases involve prototypical anatomy that facilitates optimal placement of the catheter, traumatic brain injuries may cause significantly deformed anatomic features such as small or displaced ventricles,[5,6] thus making accurate placement a difficult task.

Since EVD placement is frequently performed in emergency situations by the bedside,[7] the use of state-of-the-art

This study was presented at the IEEE Virtual Reality Conference, Arles, France, 23-27 March 2015[33].

image-guidance systems is commonly precluded due to immobility of the patient and system, as well as the preparation required. The resident (or practicing surgeon) must review orthogonal slices of the patient's brain acquired from preoperative imaging (commonly axial computed tomography [CT] scans) to formulate a mental representation that would allow them to specify a trajectory to the ventricle relative to the skull of the patient, and based on known anatomic landmarks (eyes, ears, etc.) and their estimation of the 3D localization of the ventricles. Ultimately, the procedure is complex and requires sufficient manual dexterity, spatial processing, and repeated practice. This complexity is compounded by the large variation of anatomic features seen among patients, motivating the need for the development of diverse training scenarios.

Moreover, the insertion of an EVD is generally performed free-hand and without image guidance, which may contribute to the relatively high rate of malplacement.[8] Although free-hand ventriculostomy is a common neurosurgical procedure and generally considered safe, numerous reports in the literature have identified that there is room for improvement. Catheter malplacement involves the localization of the distal tip within an extraventricular space, or the progression of the tip through critical brain structures. Clinical studies have found misplacement rates to range from as low as 12.3% to as high as 60.1%, although we speculate that this large disparity in rates may be due to differences in the criteria used to define malplacement.[9–11] Extraventricular placement necessitates replacement, which not only causes additional damage to brain tissue but has been reported to result in increased complications and haemorrhaging.[12] Teaching institutions have a high proportion of drain placements performed by residents, likely inflating the rate of malplacement due to learning curve effects, an aspect that has not been thoroughly examined in the literature. There is also increasing clinical evidence indicating that anatomic variations affect outcome, but no work has examined creating models to predict risk based on anatomic measures.[10,11,13]

Numerous ventriculostomy simulators have been proposed or implemented, but for the sake of brevity, we only examine three vastly different modern implementations: the ImmersiveTouch,[14] University of Florida's ventriculostomy simulator,[15] and the NeuroTouch.[16] The ImmersiveTouch is an augmented reality simulator that uses a haptic tool to interact with the overlaid virtual scene, allowing a user to guide a catheter into the ventricular system. The tool and internal anatomy can be rendered to provide feedback. The simulator has been shown to be accurate and to improve clinical performance through practice and feedback.[17,18] A number of anatomically varying cases have been incorporated (normal, slit, and shifted ventricles) with initial data illustrating that such topological changes affect performance.[15] Yudkowsky et al.[18] examined a larger set of cases also illustrating the effect of patient-specific variation on targeting accuracy. The drawbacks of the platform are the cost, lack of physical presence, such as a mannequin, and lack of versatility in the targeting task. The first-generation ImmersiveTouch did not provide the capability of manually determining burr-hole location, but a recent version provides this capability.[19] In addition, there is no physical head for users to landmark, a feature desired by resident users of the system.[15] University of Florida's platform consists of a physical phantom to model tissue, along with a separate VR display to visualize the internal anatomy. Due to the physical nature of the simulator, inclusion of patient-specific cases is more challenging and less cost-effective than a strictly virtual system. Finally, the NeuroTouch is a haptic-enabled VR simulator that makes use of a full-sized mannequin head to simulate a number of neurosurgical procedures, including ventriculostomies.[20] The cost of the NeuroTouch is comparable with the cost of the ImmersiveTouch. Although the NeuroTouch supports burr-hole and trajectory selection tasks, it provides only a single case and does not yet allow for patient-specific cases. We have extended the functionality of the NeuroTouch to allow for inclusion of patient-specific cases with enhanced case preparation and feedback.

Paramount to a simulator's utility is descriptive performance metrics. By default, the NeuroTouch's performance evaluation is based on comparison with a gold standard burr-hole location and trajectory for the single simulated scenario. The ImmersiveTouch offers the greatest depth of evaluation and has been used to examine catheter depth, location in the ventricular system, and distance from the foramen of Monro.[18] Although these measures are useful, they are not fully descriptive of the approach and may prove insufficient when examining unique approaches resulting from unique anatomic variation. A mixed-reality simulator established by Hooten et al.[15] examined numerous clinically functional measures, such as damage to eloquent structures, but used only distance from the foramen of Monro as the standard for tip localization. To date, there has not been a significant effort to expand the traditional repertoire of performance metrics in modern physical simulators.

In addition to providing a safe training environment, simulators allow us to test novel techniques and tools prior to clinical validation. Thus, augmented reality (AR) guidance systems are gaining traction within research and clinical settings. As early as 2002, an AR guidance system for needle biopsies (a task similar to ventriculostomies) was evaluated using phantom models.[21] Although image-guidance systems exist for ventriculostomy procedures, they have not gained widespread acceptance by surgeons, presumably due to added setup time and complexity.[22] Previously, our group developed a mobile augmented reality image-guidance tool for use in ventriculostomies that runs on consumer hardware, requires minimal setup time, and does not impede the procedural workflow. The system also has the potential to simulate tasks currently constrained to the ImmersiveTouch or NeuroTouch systems at a comparably minor cost. The technical aspects and a more thorough review of related AR technology can be found in Kramers et al.[23]

In this paper, we present our module for a mixed-reality part-task simulation of burr-hole and catheter trajectory selection for EVD placement using patient-derived simulation scenarios. The simulator encompasses the spatial components of EVD placement but does not simulate catheter extension, which is largely a haptic skill. Although existing ventriculostomy simulators have been deployed at various institutions,[17,24] the versatility of existing implementations and their metrics to evaluate task performance are somewhat limited. Existing metrics are often narrow in scope and do not account for much of the possible variation in patient anatomy and surgical approach. Using our simulation platform, we investigated the impact of patient-specific anatomic variation by identifying important factors that influence performance. Performance in the strictest sense was analysed with Fitts's methodology in mind,[25] taking into consideration both speed and accuracy of the task, as well as the natural trade-off between the two measures. Although the simulated part tasks are not exclusively psychomotor, and hence not exclusively within the purview of Fitts's law, the framework provides a basis for analysing trade-offs in speed and accuracy. We derived extensions to these objective metrics that are gaining acceptance in the literature and applied them to characterize the user performance over a set of unique simulation scenarios derived from clinical imaging data.

Finally, we used the platform to evaluate a mobile AR image-guidance system targeting deployment in an intensive care unit setting. The guidance system overlays ventricle topology onto the surgical field, allowing for real-time targeting. The intent of the study was to examine the impact of AR guidance on task performance and additionally the influence of user experience and well-defined anatomic variations. These are all factors that influence the spatial reasoning processes inherent to this task. We expect experts to outperform novice subjects, although this effect may diminish with the use of novel guidance technology.

Methods

This section describes two aspects of our work: the technical development of the system and the design of user experiments, including relevant metrics and models.

System description

The technical requirements involved incorporation of patient-specific anatomy into a mixed-reality environment that simulates the clinical process of preoperative planning, selection of a burr-hole location, and catheter trajectory. As this was intended to be a training activity, the system was to provide immediate descriptive feedback to the user in the form of 3D renderings. A physical mannequin head was required for user landmarking as well as for use with image-guidance systems; it provides a common frame of reference for aligning scenes and allows clinicians to perform familiar landmarking approaches to guide their trajectory and burr-hole selection.

The simulator module focused on a sub-task of the overall ventriculostomy, in which the participant must localize a desired burr-hole location and indicate the trajectory to the lateral ventricles. This phase of the task isolates the initial spatial reasoning component of the trajectory estimation crucial for targeting the ventricles. This clinical phase of the task requires no exploratory tactile navigation through the brain, and so the haptic tool was used as an input manipulandum (not for force feedback at this stage). The burr-hole location estimation and angle-of-entry estimation are perceptual motor in nature. There are three components to each patient-derived case, which proceed in the following order:

(1) The user examines either segmented axial slices of the case prior to the task, or they use the AR guidance tool.

(2) The user uses the mechanically tracked pointing tool to indicate a burr-hole location as well as a trajectory into the ventricles on the mannequin head.

(3) The user is shown feedback on their approach through a rendering of their trajectory within the head.

All scenarios used were derived from volumetric magnetic resonance and CT imaging data. Primary features of interest were the lateral and third ventricles; these features were segmented using a combination of manual and semi-automatic techniques. Live-wire and region growing techniques were used to speed up the segmentation process,[26,27] but the final contours were inspected manually. Although automatic segmentation algorithms exist,[28,29] they are not robust over strong anatomic variations, particularly within the third ventricle and with low-resolution clinical images. Rather than viewing the raw medical images, users are only shown the ventricular system within an outline of the mannequin head. Fig. 1 presents a view of the slice-based interface.

Our module is integrated into the NeuroTouch, a VR surgical simulator developed by the National Research Council of Canada, providing simulation modules for various neurosurgical procedures.[16,30] Although the NeuroTouch contains a module for simulating burr-hole placement and trajectory selection, the simulator does not allow for incorporation of custom scenarios and feedback. To use the simulator's interface, we registered our custom scenarios to the coordinate system of the platform's mannequin. A neurosurgeon provided manual registration to align the virtual ventricles within the visual space of the mannequin head.

In addition to incorporating our custom scenarios, we integrated a mobile AR image-guidance system previously developed by our group.[23,31] The system tracks the position of an image-based marker (a coloured cube with QR codes) mounted to the mannequin and overlays 3D renderings of the ventricles corresponding to the current scenario. The scenario is chosen manually on the device based on the experimental ordering. The system is portable, lightweight, low cost, and smoothly operable on consumer hardware. The tracking is implemented using the VuForia API with rendering performed using OpenGL. Although the marker is intended to rest on the nasion to allow for seamless registration, the mannequin head is visualized to facilitate manual alignment between the physical and virtual scenes if there is misalignment with the tracking system. Although there is potential for using the AR display as a low-cost training tool, our aim is to use the simulator platform to evaluate the use of the AR tool as a clinical aid through image guidance. In a previous study, the interface was demonstrated to improve user performance in generalized targeting tasks involving the alignment of the trajectory with the longest axis of ellipsoids.[32,33] The current study extends evaluation of the platform into a specialized surgical domain requiring precise targeting of ventricles rather than ellipsoids. Fig. 2 depicts the NeuroTouch with the guidance system.

In order to provide users with meaningful feedback on their performance, we developed a feedback module, using the open-source 3D modelling and animation suite Blender. The module takes the user output from the NeuroTouch and renders the user's trajectory through a transparent rendering of the mannequin with the internal ventricles corresponding to the scenario. To illustrate the ideal target, the right anterior horn is highlighted in red. Performance metrics are not provided through the feedback module; only a visual illustration of the user's trajectory in relation to the target is shown. An example can be seen in Fig. 3.

Experimental design

In total, 14 scenarios were created from 13 unique patients. The additional scenario was created from a case with

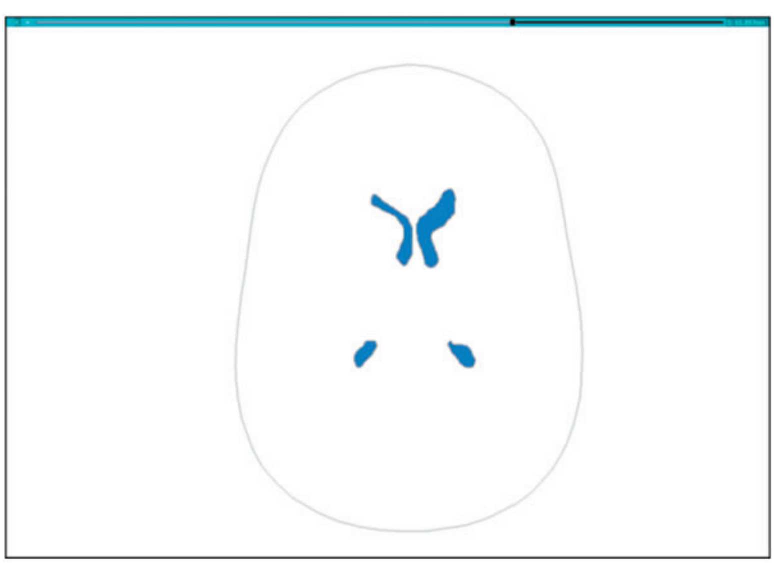

Figure 1. A screenshot of the slice-based user interface implemented using Slicer. Unlike conventional radiographs, the right side of the image corresponds to the right side of the head.

significant midline shift, which we modified by rotating it along the midline. To inform our initial arrangement of scenarios, an expert neurosurgeon classified each scenario into four distinct levels of difficulty based on the size of the ventricles, as well as on their shape and localization in space due to external mass effects, as sometimes seen in traumatic brain injury. The levels were simple (large ventricles seen in hydrocephalus), mild-moderate (enlarged ventricles, but smaller than the simple cases), moderate (ventricle of normal size with no deformation), and complex (ventricles of normal size, but with added deformation due to local mass effect). For each case, the total ventricular volume, Evan's ratio (a ratio of the maximal width of the anterior horns on an axial plane to the maximal width of the head, which has been used as a measure of the severity of hydrocephalus[34]), ipsilateral frontal horn volume, and midline shift were measured to examine how these traits affect performance of the procedure. An illustration of Evan's ratio can be seen in Fig. 4.

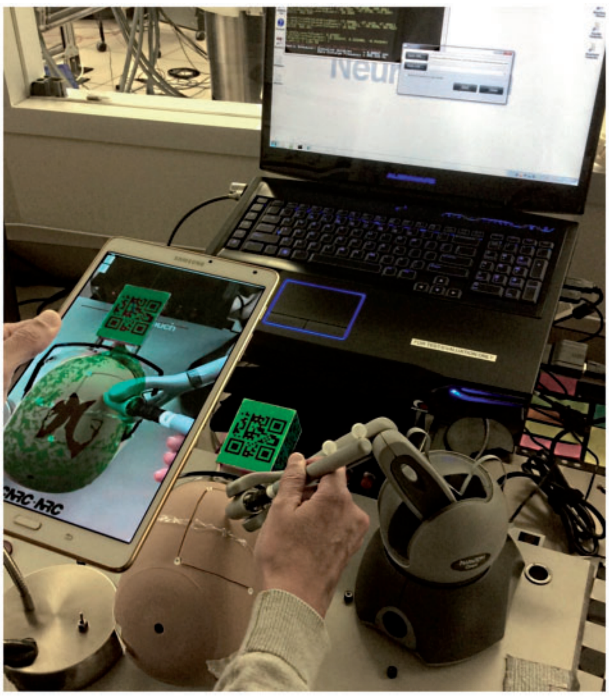

Figure 2. The custom NeuroTouch module is used in conjunction with the AR guidance tool. The ventricles can be seen overlaid on the mannequin head in addition to the green head mesh, which provides a reference for alignment of the AR tracking.

Experiments consisted of a progression of scenarios utilizing all patient-derived cases. The participants consisted of seven residents and one expert neurosurgeon and were categorized based on years of residency (PGY1–PGY6 in this study). The expert surgeon was placed in a distinct category. Ethical approval was obtained for the study, and participants provided written consent prior to enrolling in the experiment. An overview of subject experience and simulation scenarios is depicted in Fig. 5.

As EVD procedures are commonly performed in residency, this is a reasonable measure of exposure to the procedure, and we therefore considered residency experience to coincide with procedure experience. Each user was tasked with performing all 14 scenarios without AR guidance in a set order, and each scenario with AR guidance after a brief practice period with the AR device. Because it could be argued that the user would experience a learning curve, we randomized the order of the different scenarios as well as whether the user would be using the AR first or last. All users were divided into two groups with either a simple to complex ordering or a pseudo-random ordering of cases, to account for any possible learning effects of the ordering. Users were further subdivided into performing the AR guidance component first or doing the tasks without AR guidance first (Fig. 6). In addition to recording the burr-hole location and trajectory of each task, the entire task was also

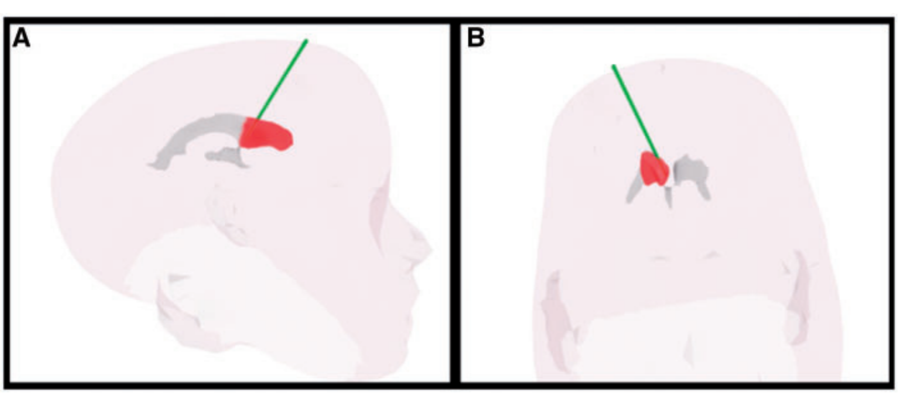

Figure 3. Two renderings providing feedback. The green line depicts the trajectory of the user, while the area highlighted in red indicates the target area (right anterior horn of the ventricles). (A) Right sagittal view. (B) Coronal view from the front. More detail regarding this interface can be found[33].

timed from the beginning of the pre-task image exploration to the final selection of trajectory.

In order to evaluate user performance, suitable metrics must be used that relate performance on the simulator to the user's ability to perform the procedure clinically under similar conditions. A naive approach would involve selecting a single correct burr-hole and catheter tip location and scoring based on deviation from this trajectory. From a geometric perspective, there are a number of possible trajectories from the skull to the ventricles. Clinically and particularly when dealing with anatomic variations, there is generally no single path that can be identified as "correct" for a successful placement. Instead, there are many possible variations that would result in a functional placement without risking damage to eloquent tissue. Traditionally, through the didactics of training, Kocher's point is sought as a desired burr-hole location from which a trajectory can be estimated leading to the entrance of the ipsilateral foramen of Monro.[35] We were able to render this path (from Kocher's point to the ipsilateral foramen of Monro) and provide it as a reference when providing feedback to users, but deviation may still result in a perfectly scored (and clinically functional) placement. In their clinical evaluation, Kakarla et al.[11] used a 3-grade scoring system that took into account the general tip location, functionality of the drain, and damage to eloquent tissue. The difficulty with this scoring approach, however, is that it does not discern between different manners of successful and unsuccessful placements. For example, there is no difference between a trajectory that misses the ventricles by a small margin compared with one that is not at all close. Yudkowsky et al.[18] examined whether the tip was successfully placed in the ipsilateral ventricle and how far it was from the foramen of Monro. Abnormal anatomic variations, however, often necessitate catheter placements that do not target the foramen of Monro area. Hooten et al.[15] developed a compound score that factored time, distance to multiple attempts and passage through critical structures. It is the most expansive scoring metric we have encountered but relies on a number of seemingly arbitrary values summed to a final score.

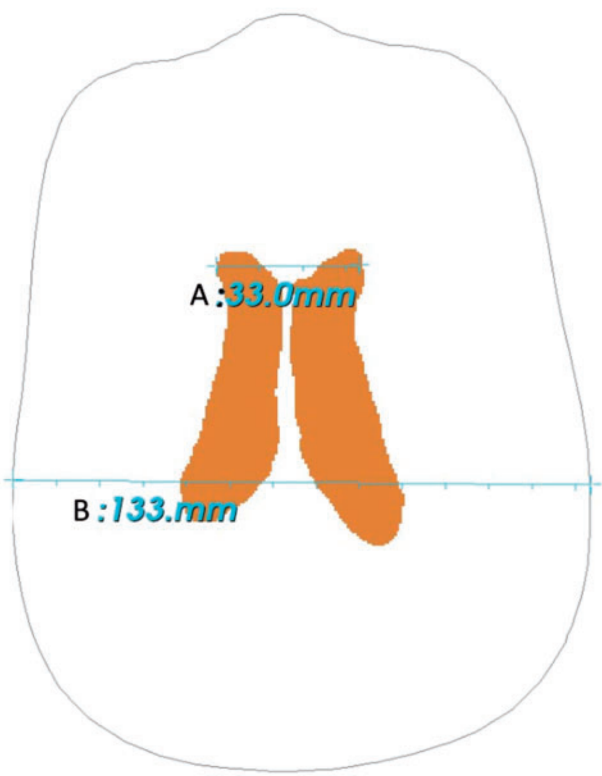

Figure 4. Measures taken along an axial slice to determine Evan's ratio. In this case, the ratio A/B produces Evan's ratio.

Our approach uses a hybrid methodology, where trajectories are both scored by a blinded expert and measured against a

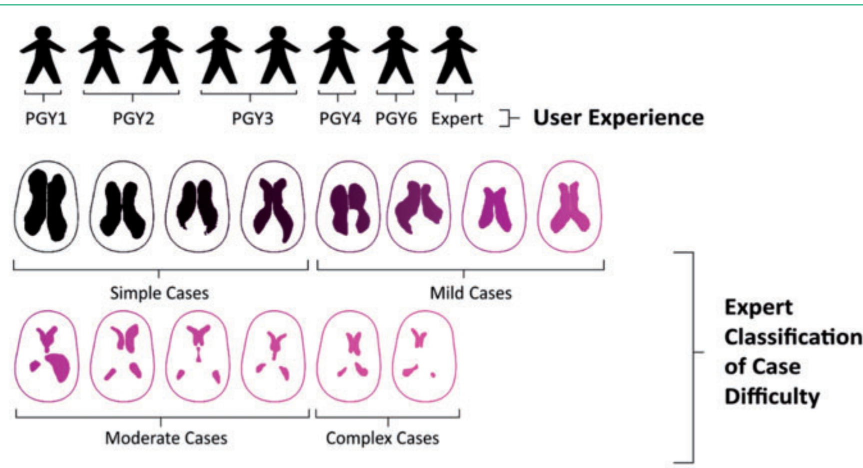

Figure 5. Summary of the user study illustrating the experience of all eight users and showing cross-sections of each case to convey the difficulty.

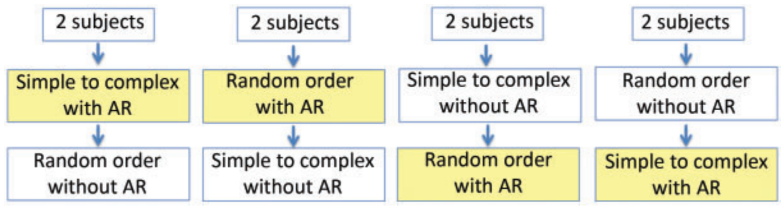

Figure 6. Flowchart of the experimental design.

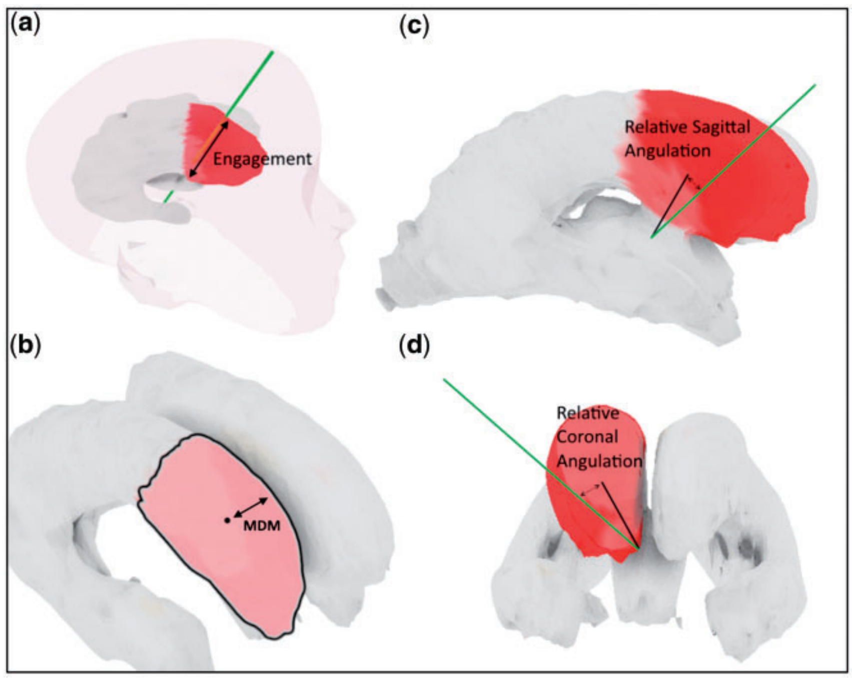

Figure 7. An illustration of the geometric accuracy metrics. Engagement (a) is the length of the trajectory intersecting the anterior horn. The MDM (b) is the minimum distance between the trajectory point and the edge of the shape created by projecting the 3D anterior horn onto the viewing plane. The relative sagittal (c) and coronal (d) angulations are the angles the trajectory makes with the foramen of Monro's respective path.

number of geometric standards. Our accuracy measures are extended from previous work and derived from Muirhead et al.[36] Although Muirhead et al. used their measures to determine the optimal trajectories based on landmarking approaches, we adapted their work to evaluate user performance of free-hand trajectories chosen by users. The paper describes four measures of accuracy: engagement, relative sagittal angulation, relative coronal angulation, and error margin. Engagement is the length of the line segment that is created by the intersection of the surgical trajectory and the anterior ventricle horn, the partition of which is described by Lind et al.[37] The relative coronal and sagittal angulations are the angles between the trajectory and the coronal and sagittal components of the foramen of Monro. Finally, the error margin is the smallest angle the trajectory can be deviated for the midpoint of the intra-horn trajectory segment to no longer be contained in the anterior horn. This measure fails to account for the various ways the topology can vary; in some cases, the calculated midpoint can be placed outside the anterior horn and yet still be a more viable surgical route than a trajectory that produces a large error margin. We have adapted this measure to what we call the minimum distance to miss (MDM), which is the distance between the trajectory point to the edge of the shape produced by projecting the 3D anterior horn onto the observer's 2D view plane from the chosen burr-hole location. This helps relate the task to Fitts's model by examining the anterior horn as a target area and adapts the error margin into a measure that better differentiates between functional and non-functional drain placements. Fig. 7 illustrates the measures outlined.

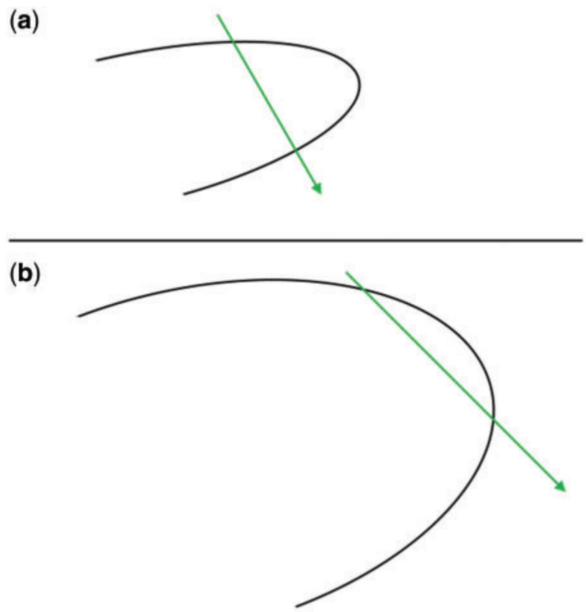

Figure 8. In this illustration, trajectories are shown intersecting ellipses. Although the trajectory in (a) is further from the right-most boundary of its target ellipse than (b), the engagement measure is still larger for (b) as a result of its size.

While these metrics provide a strong foundation to compare performance between users on similar cases, they fail to scale with ventricle size and topology in order to compare the performance of a single user among patient scenarios. Consider the engagement measure. A perfect trajectory on a case with a small anterior horn may result in an engagement measure that is in fact shorter than a near miss on a case with much larger anterior horns. This is illustrated in Fig. 8.

In order to account for the relative size of scenario ventricles, we further adapted the engagement and MDM into the relative engagement and relative minimum distance to miss. Each of the original measures in these cases were simply scaled by the volume of the anterior horn, allowing for comparison between cases.

Expert ratings were based on a 7-point Likert scale referred to as the index of accuracy (IA), and ranged from 1 (a perfect trajectory) to 7 (a very poor trajectory). Expert ratings were considered to be the gold standard against which novel geometric measures were evaluated. Example ratings are illustrated in Fig. 9.

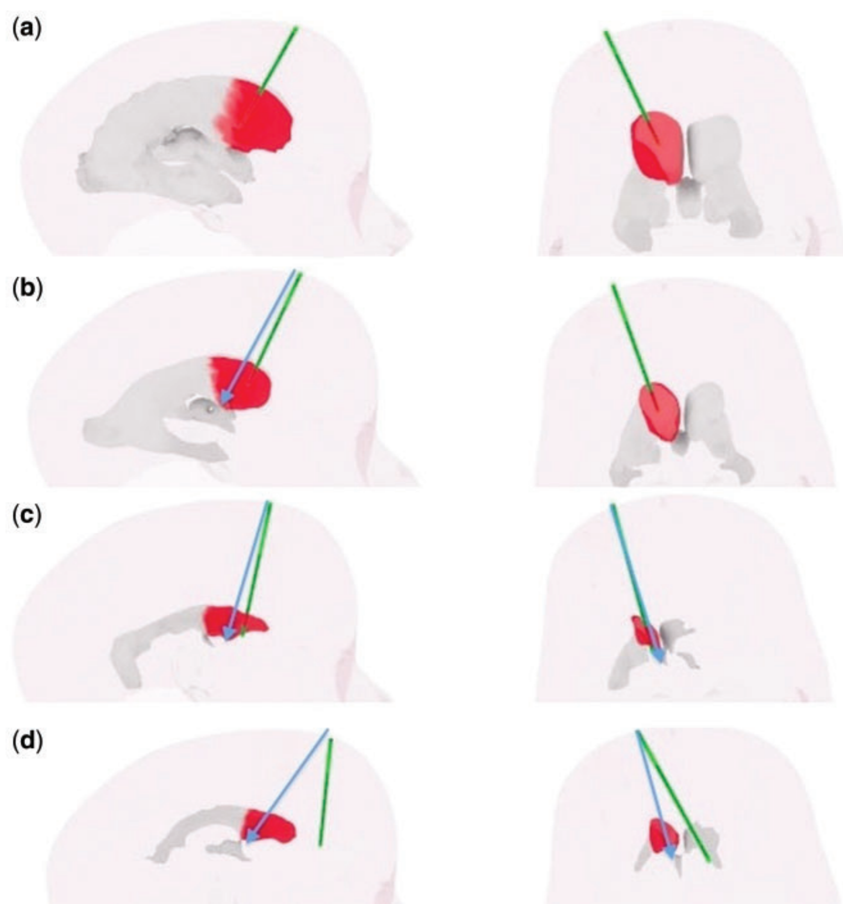

Figure 9. Illustration of different user-guided trajectories. A right sagittal view is shown on the left, while a coronal view from the front is shown on the right. The green line indicates the user s trajectory, while the blue line depicts the ideal trajectory. (A) Index of accuracy (IA) of 1, with a perfect trajectory. Note that the blue and green trajectories are perfectly coincident. (B) IA of 2 with a small angle change in one plane. (C) IA of 4 with an angle modification in both planes, with the trajectory extending beyond the ventricular boundary. (D) IA of 6, with a completely missed trajectory. More detail regarding this interface can be found[33].

Results

Expert rating and performance

The accuracy of each trajectory was scored by a blinded expert for both AR and non-AR tasks. Distributional differences between user metric variance (time and IA) was not seen applying a non-parametric Leven's test ($P > 0.15$). To examine for the presence of user difference, a Kruskal-Wallis test was run across users for time and IA. Differences between users were seen across each measure ($P < 0.001$ and $P < 0.05$). Using the resulting chi-squared values, effect size estimates were determined, suggesting that differences between users accounts for 57.35% of the variation in time and 6.46% of the variation in IA.

Figure 10 reports the mean IA and time by subject without AR guidance. Interestingly, there is large task time variance among users but low intra-user variance among tasks. Time measured encompassed the overall visual inspection, tool positioning, and final angle selection (as opposed to just the selection task).

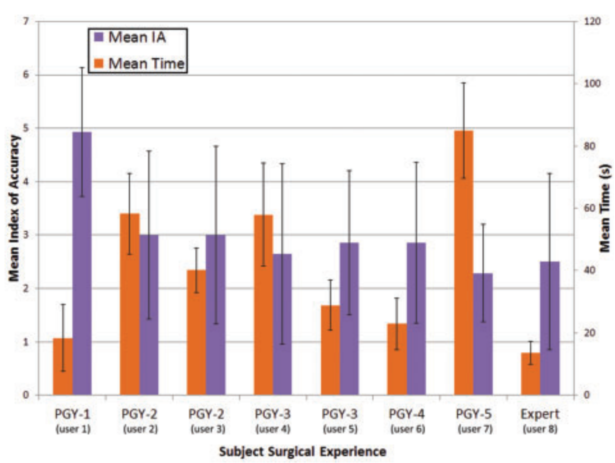

Figure 10. Mean time and IA for each user without AR guidance. The users are numbered and their class of experience is indicated. For both time and IA, a lower score is an indicator of higher performance. Standard deviations are visualized on the graph.

There was a strong inverse correlation between years of experience among residents and IA ($r = 0.93$ with 95% confidence interval [0.90–0.95]). As can be seen in Table 1, the overall accuracy of all subjects was significantly better for simple cases (mean IA = 2.43 ± 1.27) than for complex cases (mean IA = 4.5 ± 1.67) based on expert classification ($P < 0.05$ using a two tailed t test; neither group was found to deviate from normality using Kolmogorov-Smirnov test where each $P > 0.15$). For performance of all non-AR-guided cases compared with anatomic measures, see Table 1.

Overall, a significant difference was not seen to differentiate accuracy for AR-guided or non-guided targeting; however, users performed somewhat faster overall using AR guidance, although the difference did not reach significance. When analysing the individual AR results, the first five cases showed poor accuracy compared with later cases, although this was still not significant compared with the non-AR approach. Experience was not found to be predictive of performance using the AR system, as seen in Fig. 11.

Table 2 presents the performance results for AR guidance scenarios in relation to anatomic measures.

Geometric measures of accuracy

There was a significant difference in the engagement measures of all users between simple and complex cases ($P < 0.05$ and $U = 82$ using a Mann-Whitney U test as samples were found to be non-normal using the Kolmogorov-Smirnov test with $P > 0.15$). No significant findings resulted from examining the additional objective metrics in this phase. No trend was observed relating any geometric measure of accuracy with experience.

Anatomic variation

Weak correlations were seen between total ventricle volume and IA for AR and non-AR, as well as Evan's ratio and IA for both AR and non-AR targeting. Examining midline shift, the mean shift was 4.03 ± 5.73 mm. Only two cases exhibited midline shift outside the first standard deviation;

Table 1. Performance of users for all cases (without AR guidance) by difficulty classification with anatomic measures

Difficulty classification	Mean volume (mL)	Mean Evan's ratio	Mean task time (s)	Mean IA
Simple ($n = 4$)	212.94 ± 69.11	0.43 ± 0.054	38.22 ± 26.20	2.43 ± 1.27
Mild-moderate ($n = 4$)	240.77 ± 241.53	0.39 ± 0.20	37.62 ± 24.02	2.79 ± 1.52
Moderate ($n = 4$)	33.76 ± 0.26	0.26 ± 0.03	39.26 ± 28.24	3.33 ± 1.53
Complex ($n = 2$)	14.81 ± 0.00	0.21 ± 0.00	40.13 ± 23.36	4.5 ± 1.67

Values are means ± standard deviation.

these constituted the cases classified as complex. When comparing the accuracy of targeting ventricles with midline shift outside 1SD with those within, accuracy was found to be non-normal for each group (Kolmogorov-Smirnov test $P < 0.01$), but a significant difference in IA was seen for AR ($P < 0.05$ and $U = 421$ using the Mann-Whitney U test) and non-AR ($P < 0.05$ and $U = 345$ using the Mann-Whitney U test) targeting. The direction of midline shift was not seen to affect accuracy or performance.

Discussion

In this paper, we present a module to train residents to place EVDs in patient-specific scenarios implemented on a surgical simulation platform that is already deployed and in use at a number of teaching institutions. In addition, we integrated a previously developed AR image-guidance system to evaluate its efficacy in guiding EVD placement. Expert ratings were used as a gold standard for accuracy and a number of novel and adapted metrics were evaluated. We observed a correlation between surgical experience and task performance, providing evidence that measures used may discriminate between experience levels (a form of construct validity). A high correlation was seen between experience and IA, indicating that IA may be an appropriate measure to assess skill. Time itself was not predictive of experience. Indeed, although the subjects were instructed to consider the speed of their targeting, small individual differences teased out in this study are likely not of clinical focus or relevance; accuracy is surely given the full weighting of importance by users. An interesting trend to note is the high inter-subject variability of task time compared with the relatively low intra-subject variability (see Figs 9 and 10). The likely explanation is that each subject has a preferred preparation and targeting approach that characterizes their response, although it is interesting to note that all subjects were trained at the same institution.

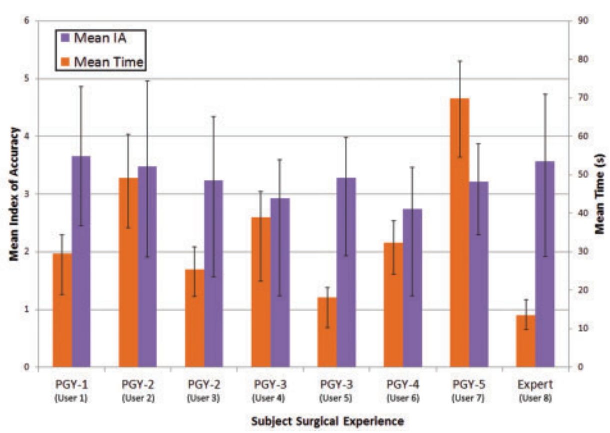

Figure 11. Mean time and IA for each user with AR guidance. The users are numbered and their class of experience is indicated. For both time and IA, a lower value is an indicator of higher performance. Standard deviations are visualized on the graph.

We observed better performance for simple cases compared with more challenging cases based on expert classification, which seemed to outperform simple anatomic measures such as ventricular volume and midline shift in determining the difficulty of a given case. This indicates that the expert is relying on additional cues when making judgments relating to the difficulty. This was evident when examining more difficult cases; these cases contain significant mass effects, leading to deformations that are not entirely captured by current metrics. The shortcoming of a global volume measurement, for example, is that it is not descriptive of ventricle topology. An ideal approach would not examine anatomic measures independently but strive towards a composite model. Further investigation is warranted.

A learning curve was seen with the use of AR guidance, which is expected when incorporating novel guidance techniques.[38] Following the initial learning curve, the AR system seemed to provide some benefit in terms of speed and accuracy, but we failed to find a significant difference compared with the non-AR approach. Because the effects of surgical experience diminished with the use of the AR approach, it is reasonable to conclude that AR guidance was more helpful for less experienced residents. Importantly and unlike in a clinical setting, we did not augment the procedure with the AR but replaced the

Table 2. Performance of users for all cases (with AR guidance) by difficulty classification

Difficulty classification	Mean volume (mL)	Mean Evan's ratio	Mean task time (s)	Mean IA
Simple ($n = 4$)	212.94 ± 69.11	0.43 ± 0.054	29.37 ± 26.20	3.39 ± 1.75
Mild-moderate ($n = 4$)	240.77 ± 241.53	0.39 ± 0.20	36.46 ± 24.02	2.75 ± 1.43
Moderate ($n = 4$)	33.76 ± 0.26	0.26 ± 0.03	32.81 ± 28.24	3.00 ± 1.15
Complex ($n = 2$)	14.81 ± 0.00	0.21 ± 0.00	34.65 ± 23.36	4.12 ± 2.10

Values are means ± standard deviation.

traditional preoperative planning of examining the patient's images. It may be that the combination of approaches allows for the greatest perspective, and therefore superior performance. Such an approach may also better facilitate learning with the AR tool, as it would pair the novel technique within a familiar context.

The geometric measures of accuracy presented in this study show promise (primarily the engagement) but fail to fully align with expert ratings and expert classification of case difficulty. As with anatomic measures, geometric metrics may be most effective when pieced together into a composite measure.

One limitation of this study is the small sample size, although this is often expected of surgical simulation research. We intend to continue recruiting more participants to improve the scope of the data to test additional hypotheses. We will continue to examine the relationship between experience and performance as well as the role of AR guidance in improving user performance. Although AR guidance did not augment performance as significantly as predicted, the study was limited in that subjects did not receive extensive training on the platform (especially important considering the learning curve), and the implementation is under constant development. We also hope to further examine various objective metrics of ventricular geometry that have an impact on performance. The safety of the simulation environment allows us to examine the intersecting effects of unique/rare anatomic variations and user surgical experience. By determining the role that these variables play in the difficulty of a case, we can make clinically relevant predictions regarding accuracy that could inform the preoperative planning process, ultimately improving patient outcome. Future work will involve investigations into predictive models that can be validated in the safety of the simulation environment.

As with all training approaches, the true test of a curriculum involves a rigorous application of the gold standards of evaluation. Additional research will examine the outcome of extended training in the environment, particularly concerning skill transfer into clinical settings. By providing diverse patient-derived scenarios, we are able to expose trainees to a wide range of possible cases prior to their clinical experiences.

Conclusion

This preliminary study indicates that the system presented has potential in training residents to find the ideal entry point and trajectory for the placement of an EVD on unique and varied cases. It is also a suitable platform to study the efficacy of new technologies (namely our AR system) in providing guidance for ventriculostomies and related neurosurgical procedures.

Acknowledgements

The authors wish to acknowledge funding from the Ontario government (OGS), and NSERC. Thank you to Marcus Lo for reviewing and formatting the manuscript.

Conflict of interest

None declared.

References

1. Quest DO. Naval aviation and neurosurgery: traditions, commonalities, and lessons learned. J Neurosurg 2007; 107: 1067–1073. https://doi.org/10.3171/JNS-07/12/1067.
2. Sealy WC. Halsted is dead: time for change in graduate surgical education. Curr Surg 1999; 56: 34–39. https://doi.org/10.1016/S0149-7944(99)00005-7.
3. Satava RM. Surgical education and surgical simulation. World J Surg 2001; 25: 1484–1489. https://doi.org/10.1007/s00268-001-0134-0.
4. Roitberg Z, Khan N, Alp MS, Hersonskey T, Charbel FT, Ausman JI. Bedside external ventricular drain placement for the treatment of acute hydrocephalus. Br J Neurosurg 2001; 15: 324–327. https://doi.org/10.1080/02688690120072478.
5. Yadav YR, Parihar V, Pande S, Namdev H, Agarwal M. Endoscopic third ventriculostomy. J Neurosci Rural Pract 2012; 3: 163. https://doi.org/10.4103/0976-3147.98222.
6. Farin A, Aryan HE, Ozgur BM, Parsa AT, Levy ML. Endoscopic third ventriculostomy. J Clin Neurosci 2006; 13: 763–770. https://doi.org/10.1016/j.jocn.2005.11.029.
7. Schödel P, Proescholdt M, Ullrich O-W, Brawanski A, Schebesch K-M. An outcome analysis of two different procedures of burr-hole trephine and external ventricular drainage in acute hydrocephalus. J Clin Neurosci 2011; 19: 267–270. https://doi.org/10.1016/j.jocn.2011.04.026.
8. Huyette DR, Turnbow BJ, Kaufman C, Vaslow DF, Whiting BB, Oh MY. Accuracy of the freehand pass technique for ventriculostomy catheter placement: retrospective assessment using computed tomography scans. J Neurosurg 2008; 108: 88–91. https://doi.org/10.3171/JNS/2008/108/01/0088.
9. Saladino A, White JB, Wijdicks EF, Lanzino G. Malplacement of ventricular catheters by neurosurgeons: a single institution experience. Neurocrit Care 2009; 10: 248–252. https://doi.org/10.1007/s12028-008-9154-z.

10. Toma AK, Camp S, Watkins LD, Grieve J, Kitchen ND. External ventricular drain insertion accuracy: is there a need for change in practice? Neurosurgery 2009; 65: 1197–1201. https://doi.org/10.1227/01.NEU.0000356973.39913.0B.
11. Kakarla UK, Chang SW, Theodore N, Spetzler RF, Kim LJ. Safety and accuracy of bedside external ventricular drain placement. Neurosurgery 2008; 63: ONS162-ONS167. https://10.1227/01.neu.0000335031.23521.d0.
12. Arabi Y, Memish ZA, Balkhy HH, Francis C, Ferayan A, Al Shimemeri A, et al. Ventriculostomy-associated infections: incidence and risk factors. Am J Infect Control 2005; 33: 137–143. https://doi.org/10.1016/j.ajic.2004.11.008.
13. Ehtisham A, Taylor S, Bayless L, Klein MW, Janzen JM. Placement of external ventricular drains and intracranial pressure monitors by neurointensivists. Neurocrit Care 2009; 10: 241–247. https://doi.org/10.1007/s12028-008-9097-4.
14. Luciano C, Banerjee P, Lemole G, Charbel F, Charbel F. Second generation haptic ventriculostomy simulator using the ImmersiveTouch™ system. Stud Health Technol Informatics 2005; 119: 343–348.
15. Hooten KG, Lister JR, Lombard G, Lizdas DE, Lampotang S, Rajon DA, et al. Mixed reality ventriculostomy simulation: experience in neurosurgical residency. Neurosurgery 2014; 10: 576–581. https://doi.org/10.1227/NEU.0000000000000503.
16. Delorme S, Laroche D, DiRaddo R, Del Maestro RF. NeuroTouch: A physics-based virtual simulator for cranial microneurosurgery training. Neurosurgery 2012; 71: ONS32–ONS42.
17. Banerjee PP, Luciano CJ, Lemole GM Jr, Charbel FT, Oh MY. Accuracy of ventriculostomy catheter placement using a head- and hand-tracked high-resolution virtual reality simulator with haptic feedback. J Neurosurg 2007; 107: 515–521. https://doi.org/10.3171/JNS-07/09/0515.
18. Yudkowsky R, Luciano C, Banerjee P, Schwartz A, Alaraj A, Lemole GM Jr, et al. Practice on an augmented reality/haptic simulator and library of virtual brains improves residents' ability to perform a ventriculostomy. Simul Healthcare 2013; 8: 25–31. https://doi.org/10.1097/SIH.0b013e3182662c69.
19. Alaraj A, Charbel FT, Birk D, Tobin M, Luciano C, Banerjee PP, et al. Role of cranial and spinal virtual and augmented reality simulation using immersive touch modules in neurosurgical training. Neurosurgery 2013; 72: A115–A123. https://doi.org/10.1227/NEU.0b013e3182753093.
20. Choudhury N, Gélinas-Phaneuf N, Delorme S, Del Maestro R. Fundamentals of neurosurgery: virtual reality tasks for training and evaluation of technical skills. World Neurosurg 2013; 80: e9–e19. https://doi.org/10.1016/j.wneu.2012.08.022.
21. Rosenthal M, State A, Lee J, Hirota G, Ackerman J, Keller K, et al. Augmented reality guidance for needle biopsies: an initial randomized, controlled trial in phantoms. Med Image Anal 2002; 6: 313–320. https://doi.org/10.1016/S1361-8415(02)00088-9.
22. O'Neill BR, Velez DA, Braxton EE, Whiting D, Oh MY. A survey of ventriculostomy and intracranial pressure monitor placement practices. Surg Neurol 2008; 70: 268–273. https://doi.org/10.1016/j.surneu.2007.05.007.
23. Kramers M, Armstrong R, Bakhshmand SM, Fenster A, de Ribaupierre S, Eagleson R. A mobile augmented reality application for image guidance of neurosurgical interventions. Am J Biomed Eng 2013; 3: 169–174. https://doi.org/10.5923/j.ajbe.20130306.05.
24. Breimer GE, Bodani V, Looi T, Drake JM. Design and evaluation of a new synthetic brain simulator for endoscopic third ventriculostomy. J Neurosurg Pediatr 2015; 15: 82–88. https://doi.org/10.3171/2014.9.PEDS1447.
25. Fitts PM. The information capacity of the human motor system in controlling the amplitude of movement. J Exp Psychol 1954; 47: 381. https://doi.org/10.1037/h0055392.
26. Hamarneh G, Yang J, McIntosh C, Langille M. 3 D live-wire-based semi-automatic segmentation of medical images. Proc SPIE 2005; 5747: 1597–1603. https://doi.org/10.1117/12.596148.
27. Adams R, Bischof L. Seeded region growing. IEEE Trans Pattern Anal Mach Intell 1994; 16: 641–647. https://doi.org/10.1109/34.295913.
28. Schnack H, Hulshoff Pol H, Baaré WFC, Viergever M, Kahn R. Automatic segmentation of the ventricular system from MR images of the human brain. Neuroimage 2001; 14: 95–104. https://doi.org/10.1006/nimg.2001.0800.
29. Schönmeyer R, Prvulovic D, Rotarska-Jagiela A, Haenschel C, Linden DE. Automated segmentation of lateral ventricles from human and primate magnetic resonance images using cognition network technology. Magn Reson Imaging 2006; 24: 1377–1387. https://doi.org/10.1016/j.mri.2006.08.013.
30. Jiang D, Hovdebo J, Cabral A, Mora V, Delorme S. Endoscopic third ventriculostomy on a microneurosurgery simulator. Simulation 2013; 89: 1442–1449. https://doi.org/10.1177/0037549713491519.
31. Kramers M, Armstrong R, Bakhshmand SM, Fenster A, de Ribaupierre S, Eagleson R. Evaluation of a mobile augmented reality application for image guidance of neurosurgical interventions. Stud Health Technol Inform 2014; 196: 204–208. https://doi.org/10.3233/978-1-61499-375-9-204.
32. Armstrong R, Wright T, de Ribaupierre S, Eagleson R. Augmented reality for neurosurgical guidance: an objective comparison of planning interface modalities. In: Zheng G, Liao H, Jannin P, Cattin P, Lee SL, editors. Medical Imaging and Augmented Reality. MIAR Lecture Notes in Computer Science, vol. 9805. Cham: Springer; 2016. pp. 233–243. https://doi.org/10.1007/978-3-319-43775-0_21.
33. de Ribaupierre S, Armstrong R, Noltie D, Kramers M, Eagleson R. VR and AR simulator for neurosurgical training. 2015 IEEE Virtual Reality (VR), 2015, Arles, 2015, pp. 147–148. https://doi.org/10.1109/VR.2015.7223338.

34. Shprecher D, Schwalb J, Kurlan R. Normal pressure hydrocephalus: diagnosis and treatment. Curr Neurol Neurosci Rep 2008; 8: 371–376. https://doi.org/10.1007/s11910-008-0058-2.
35. Kanner A, Hopf N, Grunert P. The "optimal" burr hole position for endoscopic third ventriculostomy: results from 31 stereotactically guided procedures. Minim Invasive Neurosurg 2000; 43: 187–189. https://doi.org/10.1055/s-2000-11374.
36. Muirhead WR, Basu S. Trajectories for frontal external ventricular drain placement: virtual cannulation of adults with acute hydrocephalus. Br J Neurosurg 2012; 26: 710–716. https://doi.org/10.3109/02688697.2012.671973.
37. Lind CR, Tsai AM, Law AJ, Lau H, Muthiah K. Ventricular catheter trajectories from traditional shunt approaches: a morphometric study in adults with hydrocephalus. J Neurosurg 2008; 108: 930–933. https://doi.org/10.3171/JNS/2008/108/5/0930.
38. Shamir R, Joskowicz L, Shoshan Y. An augmented reality guidance probe and method for image-guided surgical navigation. Proc 5th IEEE Int Symp on Robotics and Automation, San Miguel Regla, Mexico, 2006. p. 25–28.