

The Use of Stereoscopy in a Neurosurgery Training Virtual Environment

Nigel W John^{1*}, Nicholas I Phillips², Llyr ap Cenydd³, Serban R Pop¹, David Coope⁴,
Ian Kamaly-Asl⁵, Christopher de Souza⁶, Simon J Watt⁷

¹Department of Computer Science, University of Chester, United Kingdom

²Department of Neurosurgery, Leeds General Infirmary, United Kingdom

³School of Computer Science, Bangor University, United Kingdom

⁴Neurosurgery, Salford Royal NHS Foundation Trust, United Kingdom

⁵Neurosurgery, Royal Manchester Children's Hospital, United Kingdom

⁶School of Medicine, Cardiff University, United Kingdom

⁷School of Psychology, Bangor University, United Kingdom

* corresponding author:

Professor Nigel W John

Department of Computer Science

University of Chester

Thornton Science Park

Pool Lane, Ince

Chester CH2 4NU

United Kingdom

nigel.john@chester.ac.uk

Abstract

We have previously investigated the effectiveness of a custom built virtual environment in assisting training of a ventriculostomy procedure, which is a commonly performed procedure by a neurosurgeon and a core task for trainee surgeons. The training tool (called VCath) was initially developed as a low fidelity app for a tablet platform to provide easy access and availability to trainees. Subsequently we have developed a high fidelity version of VCath that uses a stereoscopic display to immerse the trainee in the virtual environment. This paper reports on two studies that have been carried out to compare the low and high fidelity versions of VCath, particularly to assess the value of stereoscopy.

Study 1 was conducted at the second annual boot camp organized for all year one trainees in neurosurgery in the UK. Study 2 was performed on lay people, with no surgical experience. Our hypothesis was that using stereoscopy in the training task would be beneficial. Results from Study 1 demonstrated that performance improved for both the control group and the group trained with the tablet version of VCath. The group trained on the high fidelity version of VCath with a stereoscopic display showed no performance improvement. The indication is that our hypothesis is false. In Study 2, six different conditions were investigated that covered the use of training with VCath on a tablet, a mono display at two different sizes, a stereo display at two different sizes, and a control group who received no training. Results from this study with lay people show that stereoscopy can make a significant improvement to the accuracy of needle placement. The possible reasons for these results and the apparent contradiction between the two studies are discussed.

Introduction

The use of virtual environments to help train surgical procedures is now well established and there are many examples of both research systems and commercial products that support a wide variety of surgical specialisms. For example, refer to Malone, Syed, Downes, D'Ambrosio, Quest, et al. (2010) who provide a review of simulation in neurosurgery. Our own work in developing training tools for neurosurgery has previously focused on low fidelity virtual environments using Web-based delivery (Phillips & John, 2000), and more recently tablet-based solutions (John, Phillips, ap Cenydd, Coope, Carleton-Bland, et al., 2015). We have demonstrated that it is not necessarily a requirement to use expensive hardware and software to produce an effective training tool for the ventriculostomy procedure, the insertion of an external ventricular drain (EVD) (Muralidharan, 2015). In our previous study, statistically significant changes in performance of selecting the burr hole entry point, the trajectory length and duration metrics for the group that used the tablet training tool (called VCath), together with a good indicator of improved normalized jerk (representing the speed and smoothness of arm motion), all suggest that there was a higher-level cognitive benefit to using the training tool, compared to the control group that only received conventional training.

Nevertheless new technologies continue to improve the fidelity of what can be achieved. We investigated whether the use of stereoscopy - so that the virtual patient in the training tool is rendered in three dimensions (3D), in life-size - could further improve the training of a ventriculostomy procedure. Stereoscopy could in principle be particularly beneficial in this context because it allows the user's visuo-motor system to construct a metric internal representation of the scene that can be used to guide decisions and hand actions (Howard and

Rogers, 2002). This paper reports on two different studies to compare two versions of VCath: our original tablet-based low fidelity version (VCath, 2012); and a new high fidelity version of VCath that utilizes the zSpace 200 (zSpace Inc., Sunnyvale, CA) - a 1080p HD, 120Hz stereographic three dimensional display with passive polarized eyewear, head tracking, and six degrees of freedom stylus interface (see Figure 1b).

Methods

In both versions of VCath, the introduction of a cannula into a defined body cavity of a virtual patient is broken down into a series of short tasks to determine the entry point, the trajectory, and the depth of insertion. Metrics are recorded to provide performance feedback to the trainee.



Figure 1. (a) The low fidelity iPad version of VCath; (b) High fidelity version of VCath running on the zSpace display (artistic impression used in figure to depict 3D effect).

The VCath app (Figure 1a) was developed for the 3rd generation iPad with a screen resolution of 2048×1536 pixels. All interaction was implemented using finger gestures: the camera view position and zoom can be changed; the positioning of the catheter, its orientation, and its insertion are manipulated by dragging a finger across the tablet screen. The application was coded in C++ and Objective-C using Apple's Xcode integrated development environment.

The open source OGRE (Object-Oriented Graphics Rendering Engine) provided all the graphical features required.

The zSpace version of VCath (Figure 1b) was developed using the Unity game engine (Unity Technologies, San Francisco, CA). Unity supports a variety of virtual reality peripherals for interacting with three-dimensional (3D) content and zSpace provide the ZSCore Unity Plugin to assist developers in creating content compatible with their display. The plugin supports a stereo camera mode and example scripts for manipulating the stylus interaction device. Figure 2 depicts the Unity scene being developed for VCath. In this high fidelity version of VCath the trainee wears polarized glasses to see in stereo 3D, and infrared sensors track the glasses so that head movements can be compensated for. The keyboard is used to translate and rotate the patient to the desired position and angle. The zSpace display is connected to a desktop PC with Intel Xenon E5-2609 2.4Ghz, 8Gb RAM and Nvidia Quadro 4000 graphics processing unit.

The combination of a large screen, and stereoscopic presentation, means that the zSpace version of VCath can present a life-size, stereoscopic 3D virtual patient that is perceptually similar in appearance to a real patient (unlike the small size ‘picture’ shown on the iPad version). The stylus is held similarly to a real world cannula and through it the trainee can manipulate the position and orientation of the virtual cannula with a 1:1 ratio. The trainee first defines an entry point on the virtual patient using the stylus as a laser pointer, and then uses the stylus to insert the cannula. Similar to the iPad version of VCath, once insertion has been detected the rotation of the cannula is locked so that it can only be translated along the insertion trajectory. While the trainee is free to withdraw the cannula and regain full rotational freedom, this counts as a separate insertion attempt. If the ventricles are punctured the stylus vibrates to provide momentary tactile feedback, and a particle system effect is used to give the appearance of

cerebrospinal fluid coming out of the end of the cannula. With the procedure complete the trainee is then shown their result metrics, and the patient is made semi-transparent so that they can see the trajectory of the cannula and internal anatomy (brain, skull and ventricles).

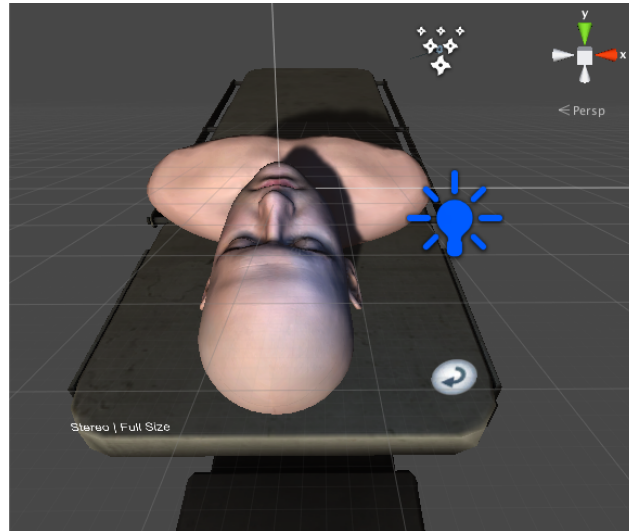


Figure 2. *Unity allows 3D content to be imported and edited (here the virtual patient), and provides a sophisticated scripting environment to control interaction. It has built in physics and lighting support and produces high quality graphical renderings.*

Study 1

A study using twenty-nine neurosurgery trainees took place at the second UK neurosurgical boot camp in September 2015 and followed the methodology of our iPad study the previous year – refer to John, Phillips, ap Cenydd, Coope, Carleton-Bland, et al., 2015, for full details. These trainees are the entire 2015 cohorts of the nationally appointed first year trainees and are the largest group of neurosurgical trainees of a similar level of experience available for the study in the UK. Note that the time constraints of the boot camp required all the data for the study to be collected in one day. The trainees participated in the study throughout the day as a part of a rotation with other (non related) training activities taking place. Their participation in the study was in a random order and as they became available each trainee was allocated to one of the

three test groups sequentially: the zSpace training group (9 members), iPad training group (10 members) and the Control group (10 members). It was not possible to distribute the trainees across the groups based on their current abilities, as this was unknown until they have started the study, and so a random distribution was used. All participants were between 26 and 36 years old (five females and 24 males). Three participants were left-handed and nobody was color blind. They were also all inexperienced in performing a ventriculostomy with the majority having performed zero or assisted in just one procedure. Three participants had assisted in between 5-10 procedures, which is still regarded as being inexperienced in overall surgical training terms. At this level any procedure will have been performed with a more senior surgeon guiding the placement of burr hole and trajectory of the catheter. To assess whether the trainees had functioning stereoscopic vision, their stereoacuity - the minimum depth difference that could be reliably discriminated on the basis of binocular stereopsis - was assessed using a standard stereo vision test (Randot; Stereo Optical Co. Inc., Chicago, IL). Most trainees had stereoacuity in the normal range (40 arcsec or better) but several (3 in the zSpace group, 2 in the iPad group, and 5 in the control group) had stereoacuity in the range 50-70 arcsec. Otherwise they had normal or corrected-to-normal vision, and no impairments to manual dexterity.

Efficacy of performing ventriculostomy for all three groups was assessed at the beginning and end of the study using a simulated insertion task. The ROWENA (Delta Surgical Ltd, Staffordshire, UK) neurosurgical physical model of a head consists of a moulded plastic base with internal skull anatomy including bone and dural layers. This was used, in conjunction with the Brainlab VectorVision (Brainlab AG, Feldkirchen, Germany) neuronavigation system. Participants used anatomical landmarks based upon their training to insert a catheter through the pre-formed right frontal burr-hole with the aim of positioning the tip of the catheter in a standard

location above the Foramen of Monroe (the channels that connect the paired lateral ventricles with the third ventricle at the midline of the brain). Only one attempt was allowed. The ventricular catheter was fitted with a Brainlab pre-calibrated stylet to enable the location and orientation of the catheter to be tracked but participants were blinded from the Brainlab tracking system throughout. Smoothness of movement of the catheter is derived from the magnitude of jerks in the recorded trajectory where jerk has the standard definition of change in acceleration over time. As this is dependent upon the length of the trajectory, it is normalized for analysis as previously described (Teulings, Contreras-Vidal, Stelmach and Adler, 1997). The metrics analyzed were: time taken to reach the end point (sec); assessed trajectory length (mm); assessed trajectory modulus (mm), the length of the straight line connecting the entry point to the final location; normalized jerk, the “smoothness” of the trajectory; and the final distance of the catheter tip from an idealized target (mm). The trajectory modulus equates to the depth of insertion, which for safety is typically limited to approximately 50mm. The difference between the actual trajectory length and its modulus gives an indication of how direct and decisive the individual trajectory was. The MATLAB software (MathWorks, Natick, MA) was used for statistical analyses.

A pre-training assessment was performed to establish the naïve skill levels of the participants. Two-sample Mann-Whitney tests were carried out between each pair of groups at this pre-training stage to determine if there were any significant differences between them. Only the assessed trajectory modulus was significant ($p = 0.0183$) when comparing the control group with the iPad group – refer to box plots in Figure 3. No significant differences were identified for any metrics when comparing the control and zSpace groups, or the iPad and zSpace groups.

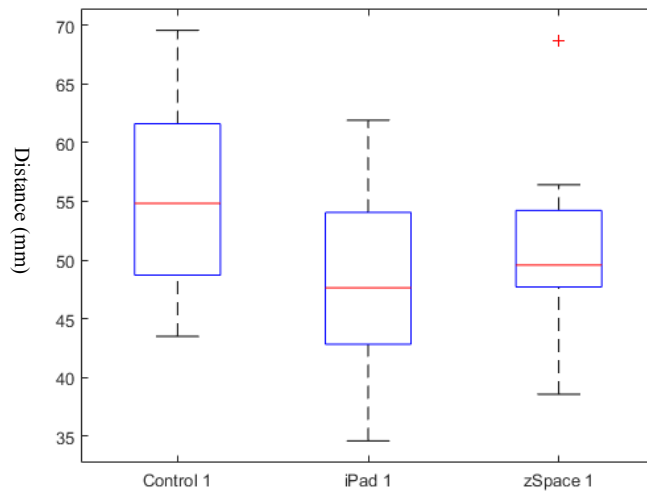
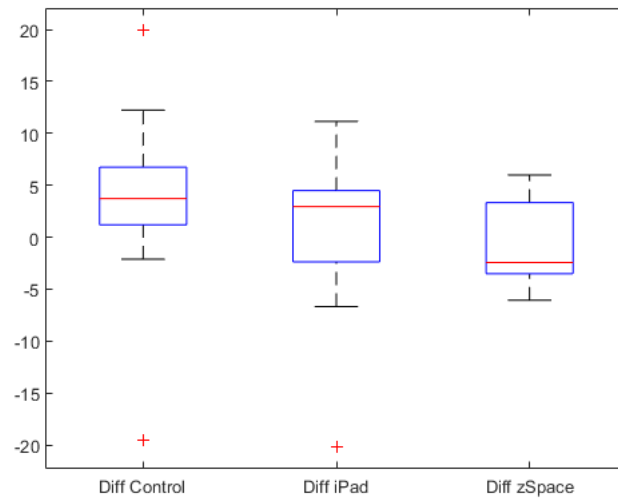


Figure 3. *At the pre-training assessment, a statistically significant difference was found between the Control group and the iPad group for the assessed trajectory modulus. The box plots for this metric for all three groups are presented above (the red cross represents an outlier). The Control group tends to use longer trajectories.*

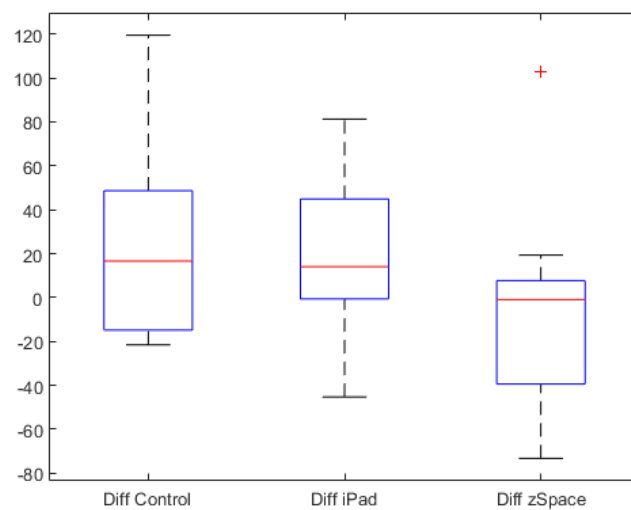
Following the pre-training assessment, all participants attended a general lecture on ventriculostomy where they were introduced to the study. At the end of the lecture the iPad and zSpace groups were allowed to familiarize themselves with the VCath application and then self-evaluate their performance over a twenty-minute training period using the inbuilt metrics. The Control group had no access to either of the VCath applications.

The three groups were then re-evaluated on the simulated insertion task to further establish their post-training skill level in hitting the ventricles. Our aim was to compare the change in performance of the trainees using the metrics obtained at the pre- and post-training simulated insertion tasks. Because the test procedure was identical to our previous study, the data from the Control and iPad groups from 2014 and 2015 were combined to give larger sample

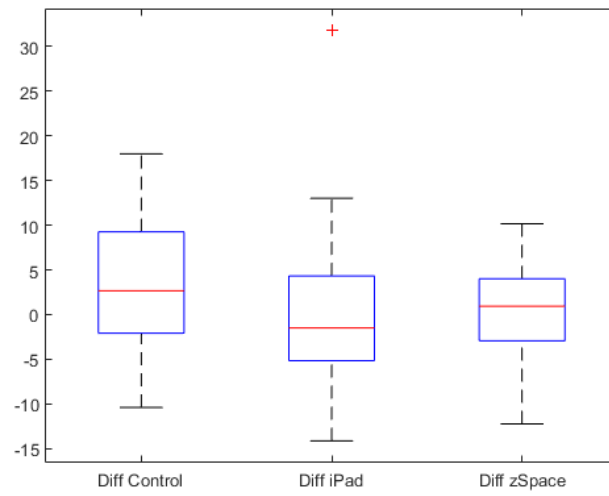
sizes of 20 and 21 respectively for these two groups. The metric differences for the pre-training minus the post-training for all three groups are summarized in the box plots in Figure 4.



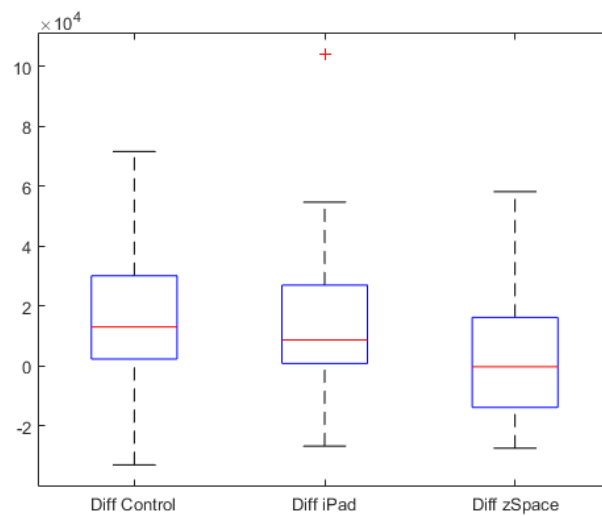
Time taken to reach the end point (sec)



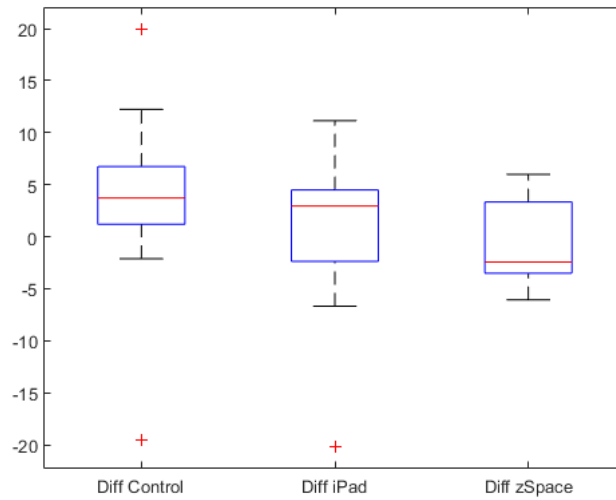
Assessed trajectory length (mm)



Trajectory modulus (mm)



Normalized jerk, the “smoothness” of the trajectory



The final distance of the catheter tip from an idealized target (mm)

Figure 4. Box plots representation of the metric differences. Red crosses are outliers

A normal distribution cannot be assumed (particularly from randomly assigning subjects from a small but highly varied population). Probability plots were therefore made from the differences between the pre- and post-training metric values for each group. This showed that assessed duration and normalized jerk did follow a normal distribution and so a two-tailed t -test was used when comparing differences between groups for these two metrics. For all other metrics, the nonparametric Mann–Whitney test was used. For the Control and iPad groups the sample size was large enough to also apply a z -test. The p -values from all three statistical tests are listed in Tables 1, 2 and 3.

Table 1. *p*-values from Two-Tailed *t*-test (Hypothesis is mean values of Group 1 = mean values of Group 2, alternative hypothesis values are not equal)

Group 1 – Group 2	Time to reach end point	Normalised jerk
Control - iPad	0.6319	0.8930
Control - zSpace	0.1133	0.2081
iPad - zSpace	0.1884	0.3006

Table 2. *p*-values from Two-Tailed Mann-Whitney Test (Hypothesis is mean values of Group 1 = mean values of Group 2, alternative hypothesis values are not equal)

Group 1 – Group 2	Assessed trajectory length	Assessed trajectory modulus	Final distance from target
Control - iPad	0.9273	0.1792	0.3280
Control - zSpace	0.0942	0.4367	0.0251
iPad - zSpace	0.0635	0.7173	0.2051

Table 3. *p*-values from z-Test comparing Control and iPad groups

Group 1 – Group 2	Time to reach end point	Normalised jerk
Control - iPad	0.4978	0.8316

The only statistically significant result was that the Control group was closer than the zSpace group in the final distance to target.

Finally, a one-sample Wilcoxon Signed Rank test was then performed within each group to compare the mean difference of the metric values between the tests at pre- and post-training –

see Table 4. Apart from assessed trajectory modulus, both the Control and iPad groups show a significant improvement between tests for all the other metrics. There is no improvement for any metric in the zSpace group.

Table 4. Wilcoxon Signed Rank test (The null hypothesis asserts that the *medians* of the two samples are identical)

	Time to reach end point	Assessed trajectory length	Assessed trajectory modulus	Normalised Jerk	Final distance from target
Control (pre) vs Control (post)	0.0013	0.0304	0.1084	0.0040	0.0057
iPad (pre) vs iPad (post)	0.0024	0.0071	0.8484	0.0190	0.1592
zSpace (pre) vs zSpace (post)	0.8203	0.7344	0.9102	0.9102	0.7344

In summary we observed that there was one metric at the pre-test stage where the Control group were significantly different but overall the three groups were at a similar starting point in terms of their skill level. At the post-training assessment there is improvement demonstrated within both the Control and iPad groups, but not from the zSpace group. The only significant result when comparing the differences between groups was for the final distance from target in the Control/zSpace comparison. The zSpace group has performed worse than the Control group who had no extra training.

Study 2

The second study was carried out with non-medical participants and more time was available for its completion. Forty-eight undergraduate psychology students from Bangor University were recruited. The study was designed to isolate the two main factors that differentiate the iPad and

zSpace versions of VCath: size of screen (small vs. life-size virtual patient) and the presence or absence of stereoscopic presentation. This resulted in six conditions:

1. Control group (who just read the literature on the procedure)
2. iPad group
3. zSpace with full size screen and stereo (i.e. life-size)
4. zSpace full size screen, without stereo
5. zSpace with half size screen (similar size to iPad) with stereo
6. zSpace with half size screen (similar size to iPad) without stereo

Criteria for inclusion were having normal stereoscopic vision and no reported impairments to manual dexterity. Participants were randomly allocated to conditions, resulting in eight participants per group. However, the simulated insertion task with the Brainlab system used in Study 1 was not available and so the final comparison test for the six groups was performed on the high fidelity zSpace version of VCath.

All participants received an introduction and written explanation on how to successfully perform the ventriculostomy procedure. This was comprised of an explanation of how to navigate around anatomical landmarks to optimise the needle insertion point, in order to accurately penetrate the skull, and successfully catheterise the lateral ventricle. Land marks and terms unfamiliar to the participant were fully explained in laymans terms and participants were encouraged to ask questions if further clarity was needed. They were then given 15 minutes to practise the procedure with the exception of the Control Group who used this time to just read information about the procedure.

Following training, a final test for all participants was performed using the zSpace with full size screen and stereo. In a 2-minute time frame, participants were required to successfully

catheterise the simulated patient and were limited to a maximum of three attempts at the needle insertion before it was classified as a fail (it has been suggested that in clinical practice following three failed attempts a trainee would abort the procedure) (Greenberg, 2010). Three metrics were recorded: accuracy of burr hole position; accuracy of needle placement (how far away from ideal position); and number of attempts to achieve a puncture. It should be noted that the size of the ventricles was larger than that used in the pre- and post-training task in Study 1, and so a successful puncture should be easier to achieve in this study. Nine participants failed to successfully complete this test - five from the zSpace full-size non-stereoscopic group, two from the zSpace half-size non-stereoscopic group, and one each from the zSpace half-size stereoscopic and iPad groups.

Figure 5 summarizes the errors across all conditions for the three metrics being examined. The burr hole position error was similar with the full-size stereoscopic modality performing best (smallest error) and the iPad the worst, though the difference was marginal (range=0.50-1.13cm). The needle placement error was more varied in outcome (range=1.28-3.12cm) again with the full-size stereoscopic having the smallest error. Full-size non-stereoscopic (2.51cm, SEM \pm 0.42) and literature (3.12cm, SEM \pm 0.44) performed noticeably worse than the other training conditions with the next closest average being an error of 1.99cm (half-size stereoscopic). The number of insertion attempts was the most variable of the measures (range=1.63-5.75) with full-size stereoscopic again performing the best (least number of insertion attempts). The iPad, half-size non-stereoscopic and full-size non-stereoscopic variants were noticeably less successful than the other three alternatives.

In summary, five scenarios for the platform used to run VCath were explored. The statistical analysis carried out was based on the metrics collected by the VCath software, as

Brainlab metrics were not available. The only statistically significant differences found related to the accuracy of needle placement where the participants using the full-size stereoscopic zSpace simulator outperformed all other categories, specifically a Mann-Whitney U test showed a statistically significant superiority over the control group (p -value=0.04). The iPad group performed well in comparison with the other groups but not significantly so. There was also a noticeable difference between all simulator-trained participants compared to the control group, who performed the worst for this metric. This suggests that a theoretical understanding is inadequate at instructing this aspect of the task in comparison to a practical and simulated process. All groups were comparable in accuracy of burr hole placement. There was a small improvement in performance for the two groups using stereoscopy, but it is not significant.

The number of insertion attempts was the most variable of the metrics for Study 2 with the zSpace full-size with stereo group again performing the best (least number of insertion attempts). The iPad, and zSpace without stereo (both full and half size screen) groups were noticeably less successful than the other three alternatives but this was not statistically significant.

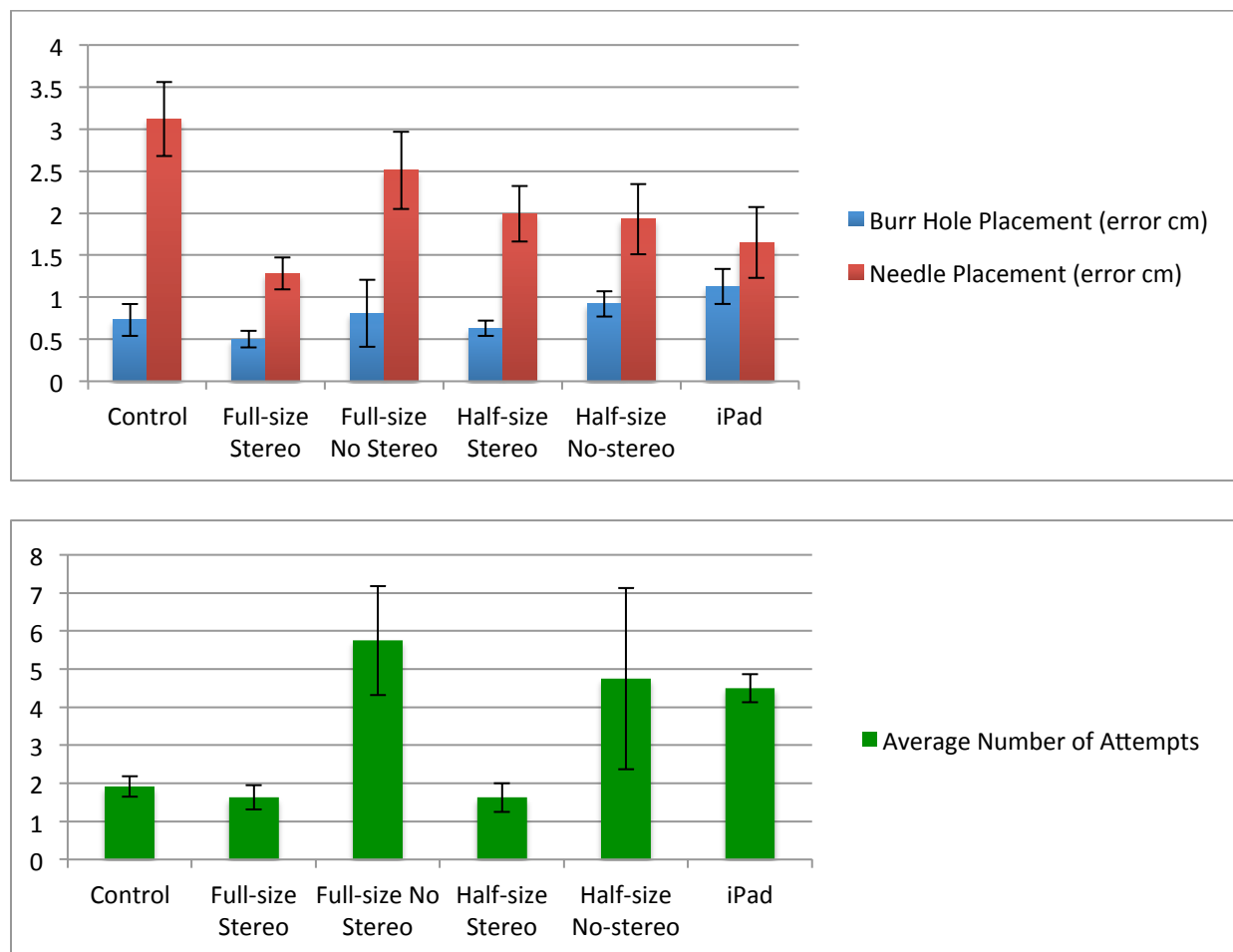


Figure 5. For each of the six groups, these graphs summarize the accuracy of burr hole position (represented by error in cm); accuracy of needle placement (by how far away from ideal position in cm); and average number of attempts to achieve a puncture. Error bars represent ± 1 standard error.

Discussion and Conclusions

In Study 1, the improvement in metric scores within the iPad group was anticipated and in line with the results of our previous study (John, Phillips, ap Cenydd, Coope, Carleton-Bland, et al., 2015). However, the improvement within the Control group and the lack of improvement within

the zSpace groups were unexpected. The participants with poor stereo acuity could be a factor but they were distributed throughout the three groups with more of them being in the Control group. In terms of the Control group getting better on the post-training assessment, then the didactic lecture given to all participants should result in some improvement. A possible further explanation is that a degree of gamesmanship occurs with candidates doing repeat assessments who make assumptions of what a good performance is and adjust their approach to the task accordingly. This would usually be a faster attempt with less hesitation and even if this were less accurate it generally would improve several of the metrics. However, our initial hypothesis that the use of a training environment using stereoscopy would improve performance of this task has been proven to be false in this study.

Previous studies have compared outcomes from a low and high fidelity training tool, particularly for laparoscopy tasks. For example, Reznick & MacRae (2006) divided participants between a high-fidelity video endoscopic urology system and a simple bench model. Results showed that both groups demonstrated the same level of improvement in performing the task. Tan, Marlow, Field, Altree, Babidge, et al. (2012) conducted one of the largest studies with 228 participants. They found that whereas training on both a low and high fidelity laparoscopy training tool did improve the participants' scores from their baseline, the low fidelity trainees were better able to transfer their skills to new settings.

Unger, Tordon, Pisa & Hochman (2016) carried out a study to investigate the importance of stereoscopy in a haptics-enabled simulator for training temporal bone surgery. They used two groups of ten first year medical students who had to practice three mastoidectomy procedures on the simulator, one group with stereoscopy and the other without stereoscopy. Following training, participants were assessed on their performance of an actual mastoidectomy

on a single 3D-printed bone model. No significant differences in the performance of the two groups were found. Outside of medical training applications, psychology experiments have also provided evidence that human observers do not necessarily have accurate perceptions of 3D metric structure (for example Todd & Norman, 2003), and that while binocular vision is often important in perceiving 3D space and guiding hand movements, it does not generally have a privileged or critical role compared to other, non-binocular sources of depth information (Hillis, Watt, Landy & Banks, 2004; Greenwald & Knill, 2009; Keefe, Hibbard & Watt, 2011; Held, Cooper & Banks, 2012).

Our results seem to concur with the above findings and we have found no strong evidence of an additional training benefit through the use of high fidelity simulators and stereoscopy. This may partly be explained by the novelty of the zSpace 3D display distracting the participants in the short time available for this study and they did not spend enough time focused on skills acquisition. The participants in Study2 had far longer to acclimatize to the simulation environment. Also, perhaps a limitation of high fidelity simulators is the difficulty in imparting skills that are transferable to a real world setting. It maybe that the zSpace group in Study 1 would show improvement at completing the simulated task (like the “non-medical” subjects in Study 2) despite not improving when asked to perform the task on the physical model in the Brainlab set up. In contrast, we hypothesize that the iPad version of VCath gives the trainees a framework to understand the task and tools to help visualize what they are doing without directly imposing a technique that they will replicate on a patient.

Conversely, the results from Study 2 show that stereoscopy can make a significant difference to the accuracy of needle placement. This result must be treated with caution, however, as the lack of the baseline test used in Study 1 meant that the full-size stereo condition

was used for the post-training assessment. Therefore improvement was shown by the group who were trained and then tested in the same stereoscopic environment. Note that some of the other conditions were as good as this same-same test. The fact that Study 2 participants were not from a medical background should not necessarily influence the results as we aim to assess the acquisition and conceptualisation of a skill that, while medically relevant, is largely independent of medical knowledge. It is possible that extended use of the zSpace version of VCath for the neurosurgical trainees would also support this finding. Repeated use of VCath over many days could also be particularly important for the iPad, which has easy accessibility.

In conclusion, the educational benefit of our low-fidelity simulator appears to be as good as, if not better, than our high fidelity version in improving procedure skill. Any additional benefit from the use of stereoscopy has not been shown. This is consistent with other medical training studies. The crucial factor for performing a successful cannulation is to have a 3D appreciation of what you are trying to do. Commonly this is achieved by dividing the procedure into a series of formal steps (although some surgeons will do this informally, subconsciously). The metrics we have used are aimed at capturing these steps. These steps are qualitatively different, and sequential, and as such cannot really be thought of as being more or less important than one another, or summing to a weighted-average performance. The practical methodology for performing the task in patients involves measuring from anatomical landmarks for the burr hole placement then lining up the catheter along two anatomical vectors. These can all be achieved as two-dimensional tasks and so the addition of stereoscopy to the simulation may not provide any added benefit in developing the skills needed for this procedure. The target ventricles are not normally visible and so in any case once the entry point has been selected and insertion begins, stereoscopy will cease to have any effect on the outcome. With training simulators

predominantly aimed towards inexperienced junior doctors, the use of low fidelity models issues a compelling argument towards the use of widely accessible inexpensive provision to supplement their formal education.

References

- Greenberg, M. (2010). *Handbook of Neurosurgery*. 7th ed. New York: Thieme Publishers, 208.
- Greenwald, H. S., & Knill, D. C. (2009). A comparison of visuomotor cue integration strategies for object placement and prehension. *Visual Neuroscience*, 26, 63–72.
- Held, R. T., Cooper, E. A., & Banks, M. S. (2012). Blur and disparity are complementary cues to depth. *Current Biology*, 22(5), 426-431.
- Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, 4, 967-992.
- Howard, I., & Rogers, B. (2002). *Seeing in depth: Vol. 2. Depth perception*. Thornhill, Ontario, Canada: University of Toronto Press.
- John, N., Phillips, N., ap Cenydd, L., Coope, D., Carleton-Bland, N., Kamaly-Asl, I. & Gray, W. (2015). A tablet-based virtual environment for neurosurgery training. *Presence: Teleoperators and Virtual Environments*, 24(2), 155-162.
- Keefe, B. D., Hibbard, P. B., & Watt, S. J. (2011). Depth-cue integration in grasp programming: No evidence for a binocular specialism. *Neuropsychologia*, 49, 1246-1257.
- Malone, H, Syed, O, Downes, M, D'Ambrosio, A., Quest, D. & Kaiser, M. (2010). Simulation in neurosurgery: a review of computer-based simulation environments and their surgical applications. *Neurosurgery*, 67(4), 1105-1116.
- Muralidharan, R. (2015). External ventricular drains: Management and complications. *Surgical neurology international* 6, no. Suppl 6: S271.

- Phillips, N. & John, N. (2000). Web-based Surgical Simulation for Ventricular Catheterisation. *Neurosurgery: Official Journal of the Congress of Neurosurgical Surgeons*, 46(4), 933-937.
- Reznick R. & MacRae H. (2006). Teaching Surgical Skills — Changes in the Wind. *The New England Journal of Medicine*. 355 (1), 2264-2269.
- Tan S., Marlow N., Field J., Altree M., Babidge W., Hewett P. & Maddern G. (2012). A randomized crossover trial examining low-versus high-fidelity simulation in basic laparoscopic skills training. *Surgical endoscopy*, 26(11), 3207-3214.
- Teulings, H., Contreras-Vidal, J., Stelmach, G., & Adler, C. (1997). Parkinsonism reduces coordination of fingers, wrist, and arm in fine motor control. *Experimental Neurology*, 146(1), 159-170.
- Todd, J. & Norman, J. (2003). The visual perception of 3-D shape from multiple cues: are observers capable of perceiving metric structure? *Perception & Psychophysics*, 65(1), 31-47.
- Unger, B., Tordon, B., Pisa, J., & Hochman, J. (2016). Importance of Stereoscopy in Haptic Training of Novice Temporal Bone Surgery. *Studies in health technology and informatics* 220, 439-446.
- VCath (2012). Available at: <https://itunes.apple.com/gb/app/vcath/id568887198>. Retrieved May 2016.