The Effect of Fidelity: How Expert Behavior Changes in a Virtual Reality Environment

Ioanna Ioannou, BEng, BCom; Alex Avery, BSc; Yun Zhou, BEng, MSc; Jacek Szudek, MD, PhD; Gregor Kennedy, BA, PhD; Stephen O’Leary, MBBS, BMedSci, PhD

Objectives/Hypothesis: We compare the behavior of expert surgeons operating on the "gold standard" of simulation—the cadaveric temporal bone—against a high-fidelity virtual reality (VR) simulation. We aim to determine whether expert behavior changes within the virtual environment and to understand how the fidelity of simulation affects users' behavior.

Study Design and Methods: Five expert otologists performed cortical mastoidectomy and cochleostomy on a human cadaveric temporal bone and a VR temporal bone simulator. Hand movement and video recordings were used to derive a range of measures, to facilitate an analysis of surgical technique, and to compare expert behavior between the cadaveric and simulator environments.

Results: Drilling time was similar across the two environments. Some measures such as total time and burr change count differed predictably due to the ease of switching burrs within the simulator. Surgical strokes were generally longer in distance and duration in VR, but these measures changed proportionally to cadaveric measures across the stages of the procedure. Stroke shape metrics differed, which was attributed to the modeling of burr behavior within the simulator. This will be corrected in future versions.

Conclusion: Slight differences in drill interaction between a virtual environment and the real world can have measurable effects on surgical technique, particularly in terms of stroke length, duration, and curvature. It is important to understand these effects when designing and implementing surgical training programs based on VR simulation—and when improving the fidelity of VR simulators to facilitate use of a similar technique in both real and simulated situations.

Key Words: Virtual reality, temporal bone, fidelity, simulation, cadaver, expert, surgery, technique.

Level of Evidence: N/A

INTRODUCTION

Cadaveric temporal bone dissection forms the foundational training activity preparing surgical trainees for operative otological experience. It is in the temporal bone laboratory that surgeons become familiar with the surgical anatomy and drilling techniques required to safely undertake mastoid and middle ear surgery. There has been recent interest in high-fidelity surgical simulation as a method of training for temporal bone surgery, largely motivated by the increasing scarcity of cadaveric temporal bones.

Virtual-reality (VR) simulation environments present detailed three-dimensional (3D) renderings of the temporal bone, which can be operated on using force-feedback (haptic) devices that simulate an otological drill. In the context of VR simulation, the term fidelity refers to the degree of visual and haptic realism of the simulation. VR simulation has been demonstrated to have face validity, providing sufficient fidelity to allow trainees to learn surgical anatomy and approach. The performance measures recorded during VR temporal bone simulation are sensitive enough to detect differences in drilling behavior between experts, novices, and surgeons in residency programs. Past work has demonstrated that novices who are introduced to mastoidectomy in a simulator will exceed the performance of those undertaking more traditional forms of early surgical education (e.g., lectures, watching surgical videos, studying dissected temporal bones) when drilling their first temporal bone. In addition, we have demonstrated that self-directed learning in a VR-simulated environment leads to better drilling performance on cadaveric temporal bones compared to traditional training.

More recently, a larger multi-center study showed that residents undertaking a more prolonged period of unsupervised practice on a VR simulator performed as well as colleagues who practiced on real temporal bones over the same period. All of these observations support the introduction of VR temporal bone simulation into surgical training programs.

To be effective as a training tool, one of the features that a simulator must possess is content validity,
meaning that the simulator should teach what is intended. An important component of content validity, in the context of simulated surgery, is that drilling technique should replicate what is used in the operating room. Experienced surgeons frequently comment that drilling does not feel the same in VR as it does in live surgery or on a cadaver. This may in part be due to physical characteristics of the haptic devices used in surgical simulators. These devices typically provide 6 degrees of freedom for movement tracking and 3 degrees of freedom for force feedback. In real surgery, instruments are also capable of exerting torsional forces, which cannot be replicated by such haptic devices. Furthermore, haptic devices can provide only a limited magnitude of force feedback, usually less than what is encountered in reality. Although haptic devices that provide greater force and torsional feedback are available, these are prohibitively expensive; therefore, they are not practical in the training environment.

In view of these considerations, it is not immediately apparent that the drilling technique used while drilling in VR will resemble that employed on a real bone. This could have significant ramifications for the early learning of surgical residents trained in VR, and could lead to the establishment of less than optimal psychomotor skills, which would need to be “unlearnt” when these residents progress to the operating room. Objective analyses of drilling technique within temporal bone simulators have focused on the measurement of surgical skill for the purposes of establishing construct validity and providing assessment. Very little work has been done that compares the drilling techniques of simulation with those of a real temporal bone to determine whether simulators replicate the correct drilling technique. Zirkle et al. tracked the hand movements of novice and experienced trainees drilling both a virtual and a cadaveric temporal bone, reporting the efficiency of hand movements, distance traveled, and time on task. These authors concluded that performance was “better” in the simulated environment because the values of the metrics were smaller (suggesting faster, more efficient drilling using fewer hand movements), but an alternate interpretation is that the drilling technique is actually different.

This study makes a direct comparison of the drilling motion of expert otologists undertaking a cortical mastoidectomy, posterior tympanotomy, and cochleostomy on cadaveric and virtual temporal bones. We reasoned that in order to identify differences across the real and simulated platforms, it was more appropriate to compare expert drilling behavior than that of trainees. In the latter group, surgical technique still is becoming established. To carry out this comparison, we recorded drill motion in both environments and used automated computer algorithms to generate a range of motion metrics with a view to identifying where the differences existed—and if possible, to identify why.

**MATERIALS AND METHODS**

An expert surgeon was defined as a qualified consultant ENT surgeon who practices in otology and cochlear implantation. After approval by the Royal Victorian Eye and Ear Hospital Research Ethics Committee, the hospital donated six cadaveric temporal bones to the study. Six expert surgeons were recruited, and each surgeon was asked to perform a single cortical mastoidectomy and cochleostomy on a cadaveric temporal bone.

Each bone was mounted to replicate the standard patient operative position and to enable the procedure to be performed as in the operating room. Electromagnetic motion-tracking sensors (Ascension TrakSTAR, Shelburne, VT) were attached to the bone and to the surgical drill to enable tracking of the 3D-position of the drill relative to the temporal bone. Figure 1 illustrates the experimental setup.

The motion tracking system recorded position measures in x, y, and z coordinates on a Cartesian plane, as well as orientation measures (azimuth, elevation, roll). In addition, information such as burr size was recorded manually, and the entire performance was captured on video.

Separately, the six expert surgeons were invited to perform a cortical mastoidectomy and cochleostomy on a standardized temporal bone within a VR temporal bone simulator. Of the original six participants, five returned to complete the simulation part of the study.

The temporal bone simulator used in this study was developed at the University of Melbourne. The simulator presents users with a virtual temporal bone generated from a segmented cadaveric micro-CT scan. The bone is rendered in 3D using the Nvidia 3D Vision kit (Santa Clara, CA). Haptic feedback is...
RESULTS

Algorithm adapted from Hall et al.15 A stroke was defined as a series of one-way analysis of variance (ANOVA) tests were conducted to investigate whether there were statistically significant differences between cadaver and simulator measures within each stage of the procedure. The metrics and stages with ANOVA P values less than 0.05 are denoted with asterisks in Figure 4.

Figure 4 demonstrates that total time was significantly greater on cadavers for the facial recess and cochleostomy stages, whereas drilling time was not significantly different for any stage. The number of burr changes was significantly different only in the facial recess stage, whereas median burr diameter was significantly larger on the simulator across all stages. The majority of the metrics characterizing stroke technique—strokes per second, stroke duration, stroke distance, stroke speed—were significantly different across two or more of the three stages. On the simulator, drilling strokes generally lasted longer, covered more distance, and were slower compared to cadaver strokes. Stroke shape metrics—percentage straight strokes, percentage round strokes—were significantly different in most stages and did not exhibit a similar pattern of change across stages. Participants tended to use more straight strokes and fewer round strokes on the

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total time</td>
<td>Total time taken to complete stage</td>
</tr>
<tr>
<td>Drilling time</td>
<td>Time spent drilling during stage</td>
</tr>
<tr>
<td>Number of burr changes</td>
<td>Count of burr changes during stage</td>
</tr>
<tr>
<td>Median burr diameter</td>
<td>Median diameter of burrs used during stage (mm)</td>
</tr>
<tr>
<td>Strokes/sec</td>
<td>Mean number of strokes/sec</td>
</tr>
<tr>
<td>Stroke duration</td>
<td>Mean stroke duration (sec)</td>
</tr>
<tr>
<td>Stroke distance</td>
<td>Mean stroke length (mm)</td>
</tr>
<tr>
<td>Stroke speed</td>
<td>Mean stroke speed (mm/sec)</td>
</tr>
<tr>
<td>% straight strokes</td>
<td>Percentage of strokes with approximately straight shape†</td>
</tr>
<tr>
<td>% round strokes</td>
<td>Percentage of strokes with approximately round shape‡</td>
</tr>
</tbody>
</table>

*A stroke was considered approximately straight if the ratio of the stroke displacement to stroke distance was above 0.85.

†Stroke roundness was evaluated by calculating the stroke centroid (i.e., the average of all points in the stroke) and the distance of each stroke point from the centroid. If the standard deviation of centroid distance was below 0.45, the stroke was considered to be approximately round in shape.
simulator compared to cadavers. These differences are summarized in Table II.

Our results suggest that participants spent the same amount of time removing bone material in both environments, but their drilling technique was different on the simulator compared to cadavers. However, the graphs in Figure 4 clearly show that the pattern of change across stages in some metrics was very similar on cadavers and the simulator, even if their absolute value was different (e.g., 4c, 4d, 4g, 4h). To determine whether simulator and cadaver metrics changed in a similar manner across stages, a two-way ANOVA test was performed for each metric, with stage and condition (i.e., simulator or cadaver) as within-subjects variables. Mauchly’s sphericity test and the Greenhouse-Geisser correction were used when necessary. The result of each ANOVA test for the stage x condition interaction is shown in Figure 4 above the respective graph for each metric.

When the interaction between stage and condition was considered, it was found that only two metrics showed a statistically significant interaction: Figure 4i
percentage straight strokes \([F(2,8) = 13.65; P = 0.003]\) and Figure 4j percentage round strokes \([F(2,8) = 33.18; P = 0.000]\). For the majority of metrics, this suggests that the pattern of change across stages in both simulator and cadaver was similar, although the absolute values may have been different.

**DISCUSSION**

We have observed that, although the metrics describing surgical strokes differed in magnitude between cadaveric and VR temporal bone drilling, experts varied their technique in similar (and predictable) ways across the later stages of an intact canal wall and cochleostomy procedure. This suggests that in adapting their technique to the simulated environment, experts retained many of the principles that were applied to the cadaveric temporal bone.

Many of the differences in metrics can be accounted for by the inherent dissimilarities in the operationalization of nondrilling activities between the cadaveric and VR environments. The measure of total time taken provides evidence that the procedure took less time on the VR simulator than on a cadaveric temporal bone. However, the time spent drilling was similar across the two environments, suggesting that the surgical component of the task was similar. Nondrilling time in cadaveric dissections was significantly greater because it included time spent changing burrs and suction irrigators, as well as adjusting the magnification of the microscope. By contrast, burr and magnification changes in the virtual environment were almost instantaneous. Streamlining the nondrilling components of temporal bone surgery resulted in a quicker completion of virtual operations, and also ensured that the user’s focus remained trained on the drilling task.16-18 This could be beneficial in early training because it would allow trainees to fully concentrate on the surgical task at hand rather than on less critical matters such as switching burrs.

Another observed difference was that experts changed burrs more often when operating in the VR environment compared to the cadaver operation. The facial recess stage accounted for the majority of variation between the two environments. This difference can be attributed to the fact that it was much easier to change burrs in VR. When participants opened the facial recess on the simulator, they took advantage of the speed with which they could change burrs, and they did so more than twice as often as on cadavers. Although limiting the number of burr changes in VR could result in a behavior that is more similar to the cadaveric environment, it is unlikely that this would result in better training. Burr diameter also differed across the two environments, but the diameter changed proportionally across stages. This is not unexpected; simulator burr sizes were not calibrated to correspond precisely to the burr sizes used on cadavers. The fact that burr size changed proportionally across stages suggests that experts were making the same relative adjustments in the simulator as on cadavers.

The practicalities discussed above cannot account for the difference in stroke technique between the two environments. Strokes in the VR environment were generally longer in duration and distance. As a result, stroke speed and frequency were lower. This discrepancy suggests that something was different about the way that bone was removed in the VR environment, which requires further elaboration. Possible causes include differences in bone removal rate, tactile (haptic) feedback, or burr behavior.

Although general stroke metrics such as distance and duration changed proportionally across stages, the metrics pertaining to stroke shape did not. In the VR environment, the percentage of straight strokes
decreased across the operation stage, whereas the percentage of round strokes increased. On cadavers, this trend was observed in the opposite direction but was not as pronounced. Overall, cadaveric dissection featured more round strokes, whereas simulator dissection featured more straight strokes. This difference in stroke shape seems likely due to different burr behavior in the simulator.

We speculated that in real surgery a cutting burr has grooves that favor efficient bone removal when spinning in a clockwise direction; therefore, surgeons will tend to orient their drill strokes to take advantage of this, lifting the burr between strokes to drill in one direction. However, the simulated burr removes material with the same efficiency in all directions. Thus, efficient drilling technique on cadavers involves lifting the burr at the end of each stroke to return it in position for the next stroke, whereas on the simulator it is more efficient not to lift the burr and instead to continue drilling in the opposite direction. In other words, the drilling algorithm of the simulator resembled more that of a polishing (or diamond) burr than that of a toothed cutting burr. Figure 4 illustrates the two drilling techniques. It is possible that the differences in stroke shape metrics can be attributed to this effect because lifting the burr could produce a more rounded stroke shape. A sample test of several strokes from each technique on the VR simulator confirmed that this was the case. The explanation suggests that the VR environment would benefit from more realistic burr interaction models that reproduce either cutting or polishing drill properties.

The outcomes of this study are consistent with the findings of Zirkle et al., who also found that trainees exhibited longer and straighter strokes in the virtual environment during a cortical mastoidectomy. Although this might be due to the drilling algorithms in the simulator, as suggested above, the possibility that the haptic technology may be dictating drilling behavior in the VR environment should also be entertained. As outlined in the introduction, affordable haptic devices provide a lower magnitude of force feedback than can be generated in real drilling, and they do not generate force during torsional movements. Until further research is done, it is reasonable to assume that drilling in the real and virtual environments is not identical, and some drilling behaviors that might be viewed as “better” in VR—such as the generation of longer, straighter stroke—should instead be interpreted as a product of the simulation environment.

Although differences between VR and cadaveric drilling have been identified, our observation that changes in technique were largely proportional across stages in both environments suggests that the simulator may still be useful for training in drill handling, provided that trainees have an existing frame of reference from cadaver-based training and are made aware of how the training environments differ. There is evidence to support this notion, given that time on a simulator has been shown to produce an equivalent or better performance than does traditional teaching when assessed on a real temporal bone. The results of this study also suggest that VR training cannot and should not completely replace temporal bone dissection.

We propose that the simulator is an excellent environment for training in surgical anatomy and approach, providing standardized exposure to anatomical variation and objective assessment. However, time still will need to be spent dissecting cadaveric bones in order to learn how a real drill behaves prior to commencement of training in the operating room. It is also worth noting that the behavior of a real drill can vary considerably depending on the model, servicing status of the handpiece, and the sharpness and torque of the burr. Therefore, it is beneficial that ENT residents are accustomed to performing temporal bone surgery with drills of varying performance characteristics, of which the VR drilling simulation could be considered as one.

CONCLUSION

We have examined the differences in surgical technique employed by a group of expert ENT surgeons in performing cortical mastoidectomy and cochlceotomy on cadavers compared to a VR temporal bone simulator. Our results suggest that, although the drilling component of the task was completed in similar time, the technique employed in the two environments was not identical. Slight differences in the modeling of burr interactions within the simulator caused measurable variation in stroke characteristics such as duration, distance, speed, and shape.

This study demonstrates the importance of undertaking detailed analyses of surgical performance in both simulated and real drilling environments. Understanding the differences between real and simulated environments not only establishes the validity of VR simulation for training, but also helps prioritize the technical improvements that may be made in the simulated environment to increase its fidelity and content validity—enabling VR surgical simulation to be employed as a valuable supplement to the cadaveric temporal bone training.

BIBLIOGRAPHY

2. George AP, De R. Review of temporal bone dissection teaching: how it was, is and will be. J Laryngol Otol 2010;124:119–125.


