AMERICAN ACADEMY OF OTOLARYNGOLOGY–HEAD AND NECK SURGERY FOUNDATION

Original Research—Otology and Neurotology

Otolaryngology–Head and Neck Surgery
2015, Vol. 152(6) 1082–1088

DOI: 10.1177/0194599815570880
http://otojournal.org
SAGE

Developing Effective Automated Feedback in Temporal Bone Surgery Simulation

Sudanthi Wijewickrema, PhD[1], Patorn Piromchai, MD, MSc[1], Yun Zhou, MSc[2], Ioanna Ioannou[1], James Bailey, PhD[2], Gregor Kennedy, PhD[3], and Stephen O'Leary, MBBS, FRACS, PhD[1]

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

Received August 17, 2014; revised December 16, 2014; accepted January 13, 2015.

Abstract

Objective. We aim to test the effectiveness, accuracy, and usefulness of an automated feedback system in facilitating skill acquisition in virtual reality surgery.

Study Design. We evaluate the performance of the feedback system through a randomized controlled trial of 24 students allocated to feedback and nonfeedback groups.

Setting. The feedback system was based on the Melbourne University temporal bone surgery simulator. The study was conducted at the simulation laboratory of the Royal Victorian Eye and Ear Hospital, Melbourne.

Subjects and Methods. The study participants were medical students from the University of Melbourne, who were asked to perform virtual cortical mastoidectomy on the simulator. The extent to which the drilling behavior of the feedback and nonfeedback groups differed was used to evaluate the effectiveness of the system. Its accuracy was determined through a postexperiment observational assessment of recordings made during the experiment by an expert surgeon. Its usability was evaluated using students' self-reports of their impressions of the system.

Results. A Friedman's test showed that there was a significant improvement in the drilling performance of the feedback group, $\chi^2(1) = 14.450$, $P < .001$. The postexperiment assessment demonstrated that the system provided timely feedback (when trainee behavior was detected) 88.6% of the time and appropriate feedback (accurate advice) 84.2% of the time. Participants' opinions about the usefulness of the system were highly positive.

Conclusion. The automated feedback system was observed to be effective in improving surgical technique, and the provided feedback was found to be accurate and useful.

Keywords

automated feedback in surgery simulation, simulation-based surgical training, virtual reality temporal bone surgery

Apprenticeship has long been the backbone of surgical education, where an expert provides the trainee with feedback during supervised operative activity. More recently, competence-based training has played a significant role in response to calls for structured learning, reduced opportunities for operating room training, and community insistence on greater accountability of training programs. Within this context, simulation has emerged as an important training tool.[1,2] Virtual reality (VR) training environments are seen as advantageous because they allow repeated training in risk-free environments. They are particularly useful in domains such as surgery, where training resources are limited, participant numbers are high, and failure is either expensive or catastrophic. Similar approaches have been used in training for aviation,[3] health,[1] defense,[4] and emergency services.[5]

Simulation can be used to support the educational principle of "deliberate practice," the concept that in order for a novice to become an expert, he or she is required to undertake tasks with the explicit intent of improving his or her skills.[6] Deliberate practice calls for the individual to focus on a defined task, typically identified by a teacher, to improve particular aspects of performance; it involves repeated practice along with coaching and immediate feedback on performance.[7] Typically, the onus of providing feedback falls on human experts, and the need for them to

[1]Department of Otolaryngology, University of Melbourne, Australia
[2]Department of Computing & Information Systems, University of Melbourne, Australia
[3]Centre for the Study of Higher Education, University of Melbourne, Australia

Corresponding Author:
Sudanthi Wijewickrema, Deparment of Otolaryngology, University of Melbourne, Level 2, Royal Victorian Eye and Ear Hospital, 32, Gisborne Street, East Melbourne, VIC 3002, Australia.
Email: swijewickrem@unimelb.edu.au

oversee the training greatly limits the utility and application of VR training environments.

Previous attempts at overcoming the need for expert supervision in VR training environments have mostly focused on end-of-task summative assessment.[8-10] While summative feedback may be constructive, it cannot replace meaningful feedback provided during training. The few researchers who have looked into the provision of real-time automated feedback in surgical simulation have provided only relatively simple forms of feedback. For example, Rhienmora et al[11] provided real-time feedback on individual metrics (force, position, and orientation) in a dental simulator by making comparisons with average expert values. Fried et al[12] quantitatively defined a range of errors for surgical performance (violation of tissue, violation of instrument tolerances, force patterns, etc) and provided real-time feedback by making comparisons with a database of metrics from prerecorded performances in an endoscopic sinus surgery simulator. Sewell et al[13] provided real-time feedback on bone removed with the correct/incorrect technique according to the currently selected metric (visibility, force, or removal region) in the form of colored voxels (3D points) in a temporal bone surgery simulator. Our previous work[14] provided automated feedback on force applied by trainees performing virtual temporal bone surgery.

Typically, these systems provided real-time feedback based on the analysis of individual metrics. However, surgical skill is multifaceted, and there exist complex interactions between various metrics that define it.[15,16] Moreover, feedback based on univariate analyses does not closely emulate the meaningful and nuanced advice that human experts provide during surgical training. We attempt to bridge this gap by introducing a system that provides real-time feedback on surgical technique based on multidimensional models of surgical expertise as applied to virtual temporal bone surgery. The system was trained to classify hand movements of surgeons as ''expert'' or ''trainee'' drilling behavior and to deliver feedback when trainee drilling was observed. The feedback consisted of advice on how to modify specific aspects of the drilling technique to better approximate expert behavior and warnings when trainees approached a critical anatomical structure with the drill.

In the study reported here, medical students undertook cortical mastoidectomy, the foundational operation on the temporal bone, within the virtual environment. After receiving standardized instructions on conducting the surgery, they were randomly allocated to receiving automated feedback or not. The main aim of the study was to determine whether participants receiving feedback significantly modified their drilling technique to approximate expert behavior and avoid injury to critical anatomical structures, compared with those receiving no feedback. A secondary aim was to determine whether the feedback given was appropriate and timely. Finally, the study aimed to determine the usability and usefulness of the feedback as assessed by participants' self-reports of their impressions of the system.

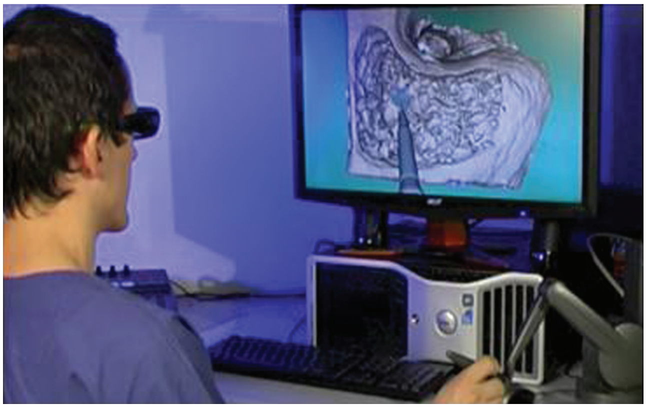

Figure 1. Melbourne University's temporal bone surgery simulator.

Methods

Test Platform

The simulation environment used in this research was the University of Melbourne VR temporal bone surgery simulator.[17] With this simulator, surgeons can practice otological operations such as mastoidectomy, middle ear surgery, and the approach to cochlear implantation. The simulator presents the trainee with 2 slightly offset images to produce the illusion of a 3D operating space, when viewed through 3D glasses (see **Figure 1**). Major anatomical structures that must be identified without injury during surgery, such as the facial nerve, sigmoid sinus, dura, ossicles, and the labyrinth, are represented in the virtual temporal bone. The surgeon interacts with the virtual temporal bone using a penlike haptic device (surgical drill) that provides force feedback in 3 dimensions.

Design of the Feedback System

To provide surgical technique feedback, we trained a classifier to recognize expert and trainee behavior using a previously collected data set of 16 performances provided by 7 experts and 11 performances provided by 6 trainees on the simulator. The training data consisted of a series of ''strokes'' identified in the continuous data stream output by the simulator during a surgical task. A stroke was defined as a set of points representing a continuous drilling motion. The end of a stroke was considered to be reached when there was no material being removed or when the direction of the trajectory showed an abrupt change.[18] For each stroke, metrics that represent surgical technique (duration, length, average speed, average acceleration, average force, straightness, median burr size, average magnification level, bone removal rate, and average distance to anatomical structures) were determined.[16] The inclusion of average distance to structures in the stroke metrics ensured that the drill location was considered when classifying strokes. A total of 15,455 and 20,779 strokes were obtained from the expert and trainee performances, respectively, to be used as training data for the classification model.

In the training stage, algorithms were used to show the model being developed precollected strokes so that it can learn expert and trainee patterns of behavior from them.[16] Once the training was completed, the model was ready to be used in real-time to classify data according to the patterns learned. If a stroke with poor surgical technique was detected, advice on how to improve the performance was provided. To this end, the feedback system determined the best metric to provide advice on, such that surgical technique could approach the expert ideal. Thus, surgical technique feedback took the form of a suggestion, delivered via audio, to either increase or decrease a metric such as stroke length, stroke speed, stroke straightness, force, burr size, or magnification level.[16]

Proximity feedback (in the form of a verbal warning) was provided when the drill tip came within 5 mm of an anatomical structure. This distance threshold was based on the results of a previous study in which we showed that injury to structures can be significantly reduced by warning otology residents when a threshold of 5 mm was crossed.[19] The aim of this feedback was to make the trainees aware that they were nearing a structure and to remind them to exercise caution when drilling, so as to expose the structure without causing critical damage (eg, facial paralysis, intracranial injury, severe hemorrhage, or deafness).

To ensure that there exists a consistent and detectable pattern of behavior, surgical technique feedback was provided to the user only after n repetitions of the same trainee behavior was detected. To avoid overloading the user with feedback, processing of strokes was paused for a t period of time after feedback was presented. Further, within a T time period after feedback presentation, if the system generated the same feedback again, it was not presented to the user. In our trials, $n = 2$, $t = 5$ seconds, and $T = 10$ seconds were established to be optimal values for the system. **Figure 2** illustrates the workflow of the feedback system (see our previous work[20] for more details).

Experimental Setup

To evaluate the performance of the feedback system, 24 students were recruited (13 MBBS, 10 MD, and 1 PhD) to participate in a randomized controlled trial. This study protocol was approved by the Human Research Ethics Committee of the University of Melbourne (HREC No. 1135497). All participants had prior knowledge of the anatomy of the ear but had no surgical experience. They were shown a video tutorial on how to perform a cortical mastoidectomy, randomly allocated to feedback or nonfeedback groups, taught how to use the simulator, and after a familiarization period of approximately 5 minutes (during which the feedback group received automated feedback while the control group did not), asked to perform this procedure on the simulator twice. The performance of all participants was recorded using a continuous data stream from the simulator and through the use of screen capture software. At the end of the procedure, participants in the feedback group were interviewed to obtain their views on the system. **Figure 3** illustrates the design of the study.

To evaluate the effectiveness of the feedback in modifying stroke technique, the percentage of strokes classified as expert (using the behavior model discussed above) for the 2 groups was compared. Further analyses of how the stroke technique changed between the groups at different stages of the surgical procedure were also conducted.

A postexperiment evaluation carried out by an expert otologist assessed the accuracy of the feedback system on 3 error measures: (1) false-positive classifications, when feedback was provided while stroke technique was acceptable; (2) wrong feedback, when participants' technique was accurately classified as ''trainee'' but the content of the feedback was inaccurate; and (3) false-negative classifications, when feedback was not provided while stroke technique was unacceptable.

The metrics used to define surgical technique (stroke duration, stroke length, speed, acceleration, force, straightness of stroke, burr size, magnification level, bone removal rate, and distance to anatomical structures) were compared between groups using a Friedman's test to assess how they were affected by the feedback.

The amount of damage caused to anatomical structures was compared between the 2 groups in an attempt to evaluate the effectiveness of the proximity feedback. The damage caused was measured as the percentage of structure voxels (ie, voxels that make up critical anatomical structures) drilled when compared with the total number of voxels drilled during the procedure, as it shows the amount of damage relative to the extent to which the mastoid was drilled. Further, the end products of the procedures of all participants were evaluated by an expert otologist (blinded to the study groups) using the Welling Scale,[21] a validated method of assessment of the quality of a mastoidectomy that systematically scores exposure and injury of key surgical landmarks. A Friedman's test was performed on the resulting scores to identify differences in the performance of the 2 groups.

To gather qualitative information on the usability and usefulness of the feedback system, the participants' answers to the following interview questions were analyzed: (1) Did you pay attention to the feedback and notice it while you completed the task? (2) Did it assist you when you were completing the procedure or stages of it? (3) Was it unhelpful, irrelevant, or distracting when you were completing the procedure or stages of it? and (4) How could the provision of feedback by the system be improved or be made more useful?

Data analysis was performed using Matlab R2014a, and a confidence interval of 95% was considered when testing for significance ($P \leq .05$). Friedman's test (nonparametric 2-way analysis of variance)[22] was used for comparing performance as the data did not withstand the test for normality. Responses to the interview questions were grouped into the most common categories using qualitative coding.[23]

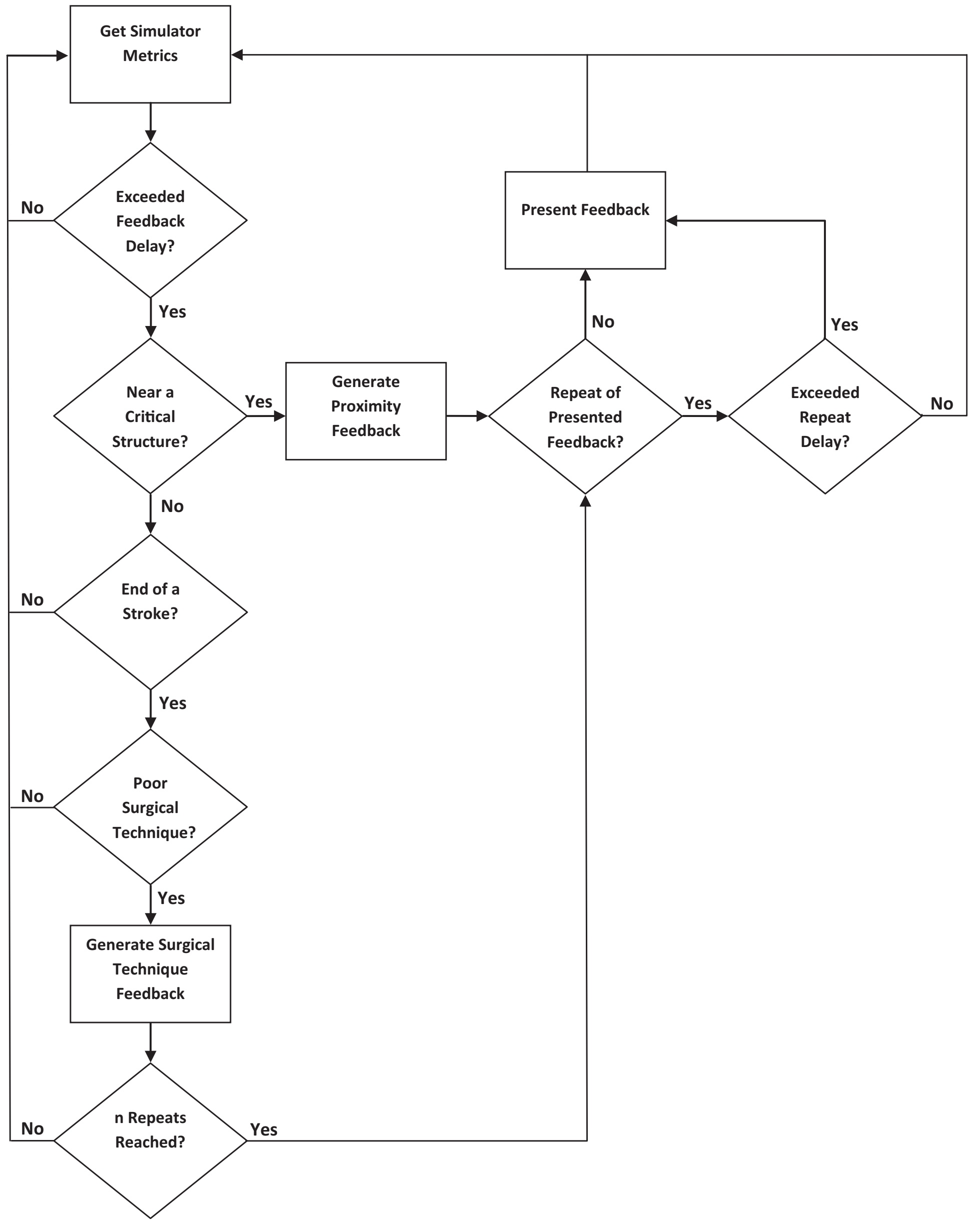

Figure 2. Design of the feedback system.

Results

The results of the Friedman's test, after adjusting for the effects of repetition, showed that the percentage of expert strokes of the feedback group was significantly higher than that of the nonfeedback group, $\chi^2(1) = 14.450$, $P < .001$. A post hoc analysis of the data using a Bonferroni adjustment showed that there was no significant difference between the 2 repetitions. Given the lack of difference in stroke technique between simulation procedures, the data for each participant across the 2 repetitions were combined (averaged). The percentage of expert strokes in the 2 groups during different stages (at 10% intervals of completion) showed a consistent difference throughout the procedure (see **Figure 4**).

A total of 576 feedback messages were provided across the 2 repetitions of the participants in the feedback group. Thirty-nine feedback messages were determined to be false-positives, 52 messages were assessed as wrong feedback, and 69 instances were identified as false-negatives, where feedback should have been provided but was not. Therefore, timely feedback was provided by the system 88.6% of the time, and in 84.2% of these instances, it was accurate.

Of all the metrics used to define stroke technique, only the bone removal rates, when removing either solid cortical bone or porous trabeculated bone, were found to be significantly different between the 2 groups, $\chi^2(1) = 4.050$, $P = .044$, and $\chi^2(1) = 6.050$, $P = .014$, respectively, after adjusting for effects introduced by the repetitions. The bone

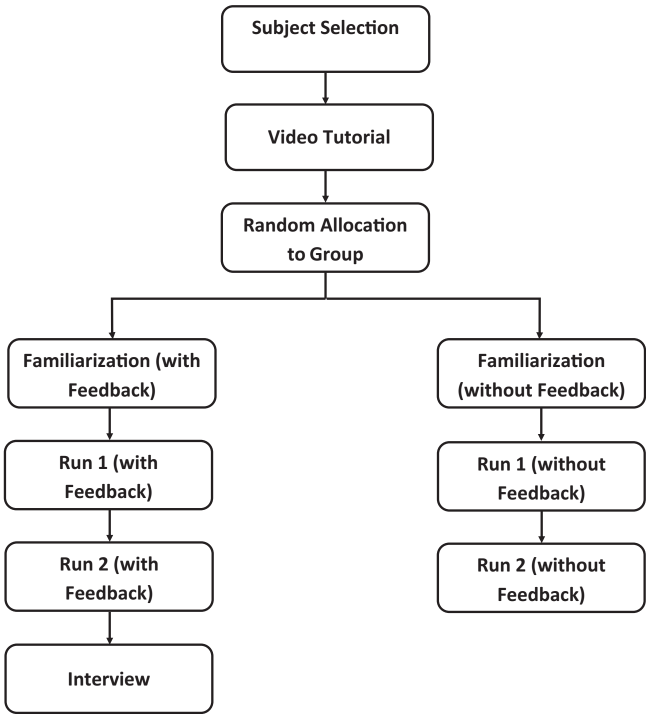

Figure 3. Study design.

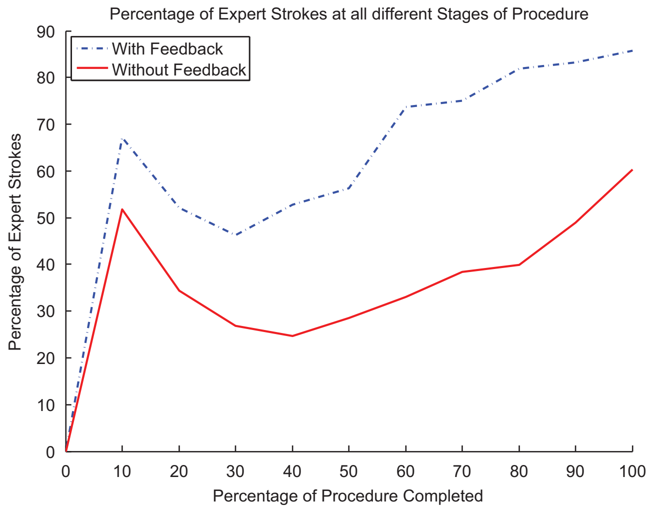

Figure 4. Comparison of surgical technique.

removal rates were found to be higher in the feedback group when compared with the control group.

No significant differences were observed either between groups or between participants' repetitions with respect to the percentage of structure voxels damaged or the scores obtained using the Welling Scale.[21]

Participants' responses to the interview questions were highly positive. Eight participants (out of 12) indicated that they paid attention to the feedback when completing the task, while 11 found the feedback to be helpful. Five students said they thought some of the feedback was contradictory or wrong. Five participants also found some feedback to be unclear. A more detailed analysis of the results is given in **Table 1**.

Discussion

The results of the study indicate that the surgical technique feedback offered by the system was effective in guiding drilling behavior toward an ''expert'' ideal, demonstrating that it can help shape better surgical technique. These results are consistent with those of previous studies on automated feedback within surgical simulators,[13,24] although none were conducted on feedback of the sort provided in this study. The major difference was that in our study, the focus was on surgical technique as a multivariate behavior model rather than on the individual metrics that define it. For example, Sewell et al[13] provided real-time feedback on bone removal (based on individual metrics such as visibility, force, and removal region) and showed that the group that received feedback maintained better visibility while drilling. Judkins et al[24] presented feedback on parameters such as speed, grip force, and relative phase in virtual laparoscopy surgery and observed that it improved performance with respect to these metrics.

Whether the observed modification of drilling behavior is in effect a move toward improved performance is largely dependent on how close the expert ideal is to actual expert performance and if such an ideal can in fact be defined for a given task. It has been shown previously that expert and trainee behavior can be differentiated on a VR temporal bone simulator, and the core metrics that define this difference have also been identified.[25] The multivariate behavior models developed using these metrics have been shown to accurately classify expert and trainee performance,[16] indicating that for our application, an expert ideal can be defined. These results are further validated by the outcome of the postexperiment analysis that assessed the accuracy of the feedback system.

The need for using multivariate over univariate drilling behavior to define the ideal is that there are interactions between individual metrics that are not expressed in a univariate model (eg, the relationship between force and proximity to structures identified in Wan et al[26] as an element of competency in the evaluation of cortical mastoidectomy). There is a risk that if individual dimensions of expert drilling behavior were presented to trainee surgeons, they may concentrate on these at the exclusion of others, running the risk that other aspects of surgical technique are compromised. For example, novices aspiring to emulate the faster drilling rates of experts may ignore other factors such as force, orientation of the drill, and proximity to structures, thereby risking injury to critical anatomy.

Of the core metrics used to define surgical technique (ie, stroke duration, stroke length, speed, acceleration, force, straightness of stroke, burr size, magnification level, bone removal rate, and distance to anatomical structures), only bone removal rate was significantly different between the groups. This demonstrates that greater efficiency in drilling

Table 1. Analysis of Participant Responses.

Topic	Common Responses	n/12
Did you pay attention to the feedback and notice it while you completed the task?	Yes	8
	Yes, but they ignored some of the feedback or thought it to be less important	3
	Paid attention especially to distance feedback	2
Did it assist you when you were completing the procedure or stages of it?	Feedback was helpful/useful	11
	Some of the feedback was helpful	1
	Some feedback was useful initially but not later	1
	Feedback was not helpful	1
Was it unhelpful, irrelevant, or distracting when you were completing the procedure or stages of it?	Feedback was not distracting	2
	Feedback was distracting	1
	Some of the feedback was distracting	2
	Distance feedback was a little distracting	1
	Repetitive feedback was distracting	1
	Some feedback was contradictory/thought to be wrong	5
	Some feedback was unclear (ie, wasn't sure how to follow it)	5
How could the provision of feedback by the system be improved or be made more useful?	Indicate how close structures are when giving warning	1
	Combine distance and stroke feedback	2
	More feedback on burr size	2
	Provide procedural feedback (ie, where to drill)	3
	Fix zoom feedback	1
	Give more specific feedback	2
	Less frequent distance feedback	1
	Provide feedback more visually	1

was achieved by the feedback group without evidence of increased damage to structures or reduced scores in the end-product analysis. However, previous studies[25,27] show that other independent metrics such as force are also useful in differentiating the experience levels of surgeons. This implies that the changes in the surgical technique of participants who received feedback were not large enough to be detected by a univariate analysis of individual metrics. It is also probable that, as rank beginners with no training in operative surgery, the participants were not able to build enough competence to reach a higher skill level. This outcome is consistent with the educational notion of deliberate practice, which indicates that the number of hours spent in practice is an important determinant of the level of expertise.[28] The observations that proximity warnings did not significantly reduce anatomical structure damage, and that the end product assessment scores were not significantly different between groups, complement the above observations and also suggest that expertise has more dimensions that have not been addressed in this study. Indeed, a comprehensive surgical training program would require feedback to be provided in dimensions beyond psychomotor skills (which was the main focus of this study), such as advice on where to drill and how to proceed at certain points of the procedure.

According to the outcomes of this study, surgical drilling behavior of medical students could be improved by providing automated real-time feedback. It will be important in future studies to test whether similar benefits are seen with surgical residents, who arguably may have better a priori knowledge of surgery. While we have demonstrated that the feedback changed drilling behavior at the time that it was presented, we have not here addressed whether there is retention of this learning and ultimately improved drilling behavior in either cadaveric or operating room environments. However, studies in other fields (eg, in laparoscopy[29,30]) suggest that skills learned through simulation-based training do successfully transfer to the operating room, and it would be of interest to determine whether the same is found when automated feedback is integrated into the simulation environment.

In conclusion, the results of this study suggest that trainees could be guided in the ''right'' direction when learning to handle a drill in a surgical procedure and indicate that the dream of a self-guided simulation-based surgical training system for temporal bone surgery is attainable. Such a training platform would not only reduce the burden placed on expert instructors but would also assist in producing a better class of surgeons.

Acknowledgments

The authors would like to thank the Asian Office of Aerospace Research and Development for funding the development of the feedback system and Cochlear Ltd for providing grants to assist with its testing.

Author Contributions

Sudanthi Wijewickrema, design and implementation of feedback system, implementation of classifiers, study design, data collection, data analysis, writing of the manuscript, final approval, agreement to be accountable; **Patorn Piromchai**, assessment of data, writing of the manuscript, final approval, agreement to be accountable; **Yun Zhou**, implementation of classifiers, writing of the manuscript, final approval, agreement to be accountable; **Ioanna Ioannou**, implementation of classifiers, writing of the manuscript, final approval, agreement to be accountable; **James Bailey**, implementation of classifiers, design of feedback system, writing of the manuscript, final approval, agreement to be accountable; **Gregor Kennedy**, study design, writing of the manuscript, final approval, agreement to be accountable; **Stephen O'Leary**, study design, writing of the manuscript, final approval, agreement to be accountable.

Disclosures

Competing interests: None.

Sponsorships: None.

Funding source: The Asian Office of Aerospace Research and Development and Cochlear Ltd provided grants to conduct this research.

References

1. Hammoud MM, Nuthalapaty FS, Goepfert AR, et al. To the point: medical education review of the role of simulators in surgical training. *Am J Obstet Gynecol*. 2008;199:338-343.
2. Chen C-H, Jeng M-C, Fung C-P, et al. Psychological benefits of virtual reality for patients in rehabilitation therapy. *J Sport Rehabil*. 2009;18:258-269.
3. Howard CE. Simulation & training: expecting the unexpected. *Mil Aerosp Electron*. 2011;22:12-23.
4. Cosma D, Stanic M-P. Implementing a software modeling-simulation in military training. *Rev Academ Fortelor Terestre*. 2011;16:204-215.
5. Syi S, Shih C-L. Modeling an emergency medical services system using computer simulation. *Int J Med Inform*. 2003;72:57-72.
6. Ericsson KA, Krampe RT, Tesch-Römer C. The role of deliberate practice in the acquisition of expert performance. *Psychol Rev*. 1993;100:363-406.
7. Cox M, Irby DM, Reznick RK, et al. Teaching surgical skills—changes in the wind. *N Engl J Med*. 2006;355:2664-2669.
8. Macke T, Rosen J, Pugh C. Data Mining of the E-Pelvis Simulator Database: a quest for a generalized algorithm for objectively assessing medical skill. *NextMed/MMVR14*. 2005;355-360.
9. Megali G, Sinigaglia S, Tonet O, et al. Modelling and evaluation of surgical performance using hidden Markov models. *IEEE Trans Biomed Eng*. 2006;53:1911-1919.
10. Kerwin T, Wiet G, Stredney D, et al. Automatic scoring of virtual mastoidectomies using expert examples. *Int J Comput Assist Radiol Surg*. 2012;7:1-11.
11. Rhienmora P, Haddawy P, Suebnukarn S, et al. Intelligent dental training simulator with objective skill assessment and feedback. *Artif Intell Med*. 2011;52:115-121.
12. Fried MP, Satava R, Weghorst S, et al. Identifying and reducing errors with surgical simulation. *Qual Saf Health Care*. 2004;13(suppl 1):i19-i26.
13. Sewell C, Morris D, Blevins NH, et al. Providing metrics and performance feedback in a surgical simulator. *Comput Aided Surg*. 2008;13:63.
14. Kennedy G, Ioannou I, Zhou Y, et al. Mining interactions in immersive learning environments for real-time student feedback. *Aust J Educ Tech*. 2013;29(2).
15. Sewell C, Morris D, Blevins NH, et al. Validating metrics for a mastoidectomy simulator. *Stud Health Technol Inform*. 2007;125:421-426.
16. Zhou Y, Bailey J, Ioannou I, et al. Constructive real time feedback for a temporal bone simulator. *Med Image Comput Comput Assist Interv*. 2013;16:315-322.
17. O'Leary SJ, Hutchins MA, Stevenson DR, et al. Validation of a networked virtual reality simulation of temporal bone surgery. *Laryngoscope*. 2008;118:1040-1046.
18. Hall R, Rathod H, Maiorca M, et al. Towards haptic performance analysis using K-metrics. *Haptic and Audio Interaction Design*. 2008;50-59.
19. Zhao YC, Kennedy G, Yukawa K, et al. Improving temporal bone dissection using self-directed virtual reality simulation results of a randomized blinded control trial. *Otolaryngol Head Neck Surg*. 2011;144:357-364.
20. Wijewickrema S, Ioannou I, Zhou Y, et al. A virtual reality temporal bone surgical simulator with automated real-time feedback for effective learning of surgical technique. *NextMed/MMVR21*. 2014;462-468.
21. Butler NN, Wiet GJ. Reliability of the Welling scale (WS1) for rating temporal bone dissection performance. *Laryngoscope*. 2007;117:1803-1808.
22. Seigel S. *Nonparametric Statistics for the Behavioural Sciences*. New York: McGraw-Hill; 1956.
23. Miles MB, Huberman AM. *Qualitative Data Analysis: An Expanded Sourcebook*. Thousand Oaks, CA: SAGE; 1994.
24. Judkins TN, Oleynikov D, Stergiou N. Enhanced robotic surgical training using augmented visual feedback. *Surg Innov*. 2008;15:59-68.
25. Zhao YC, Kennedy G, Hall R, et al. Differentiating levels of surgical experience on a virtual reality temporal bone simulator. *Otolaryngol Head Neck Surg*. 2010;143:S30-S35.
26. Wan D, Wiet GJ, Welling DB, et al. Creating a cross-institutional grading scale for temporal bone dissection. *Laryngoscope*. 2010;120:1422-1427.
27. Agus M, Giachetti A, Gobbetti E, et al. A haptic model of a bone-cutting burr. *Stud Health Technol Inform*. 2003;94:4-10.
28. Ericsson KA, ed. The acquisition of expert performance. In: *The Road to Excellence: The Acquisition of Expert Performance in the Arts and Sciences, Sports, and Games*. Mahwah, NJ: Lawrence Erlbaum Associates; 1996;1-50.
29. Korndorffer JR Jr, Dunne JB, Sierra R, et al. Simulator training for laparoscopic suturing using performance goals translates to the operating room. *J Am Coll Surg*. 2005;201:23-29.
30. Torkington J, Smith S, Rees B, Darzi A. Skill transfer from virtual reality to a real laparoscopic task. *Surg Endosc*. 2001;15:1076-1079.