Adaptive Virtual Reality Training to Optimize Military Medical Skills Acquisition and Retention

Ka-Chun Siu, PhD*; Bradley J. Best, PhD†; Jong Wook Kim, PhD‡; Dmitry Oleynikov, MD§; Frank E. Ritter, PhD‖

ABSTRACT The Department of Defense has pursued the integration of virtual reality simulation into medical training and applications to fulfill the need to train 100,000 military health care personnel annually. Medical personnel transitions, both when entering an operational area and returning to the civilian theater, are characterized by the need to rapidly reacquire skills that are essential but have decayed through disuse or infrequent use. Improved efficiency in reacquiring such skills is critical to avoid the likelihood of mistakes that may result in mortality and morbidity. We focus here on a study testing a theory of how the skills required for minimally invasive surgery for military surgeons are learned and retained. Our adaptive virtual reality surgical training system will incorporate an intelligent mechanism for tracking performance that will recognize skill deficiencies and generate an optimal adaptive training schedule. Our design is modeling skill acquisition based on a skill retention theory. The complexity of appropriate training tasks is adjusted according to the level of retention and/or surgical experience. Based on preliminary work, our system will improve the capability to interactively assess the level of skills learning and decay, optimizes skill relearning across levels of surgical experience, and positively impact skill maintenance. Our system could eventually reduce mortality and morbidity by providing trainees with the reexperience they need to help make a transition between operating theaters. This article reports some data that will support adaptive tutoring of minimally invasive surgery and similar surgical skills.

INTRODUCTION

About 4,300 physicians of the U.S. Army Medical Command continuously rotate through deployments across primary care, combat casualty care, and host nation care, with 2,800 individual deployments and an average deployment of 113 days. (Buller JL, Presentation given at the Medicine Meets Virtual Reality meeting, February 2011). The nature of required skills varies dramatically by deployment. For example, in-theater care for high-velocity wounds often requires procedures such as debridement, cauterization, and ligation, whereas usual surgical care in the civilian setting emphasizes procedures such as laparoscopic cholecystectomy and hernia repair.

This constant shift of required skills confronts surgeons with the need for skills different than those they are employing before deployment, but that they must somehow train or retrain to expertise before use. Although they are deployed, their previously sharp skills required in other theaters then may decay through disuse unless they are able to somehow train those skills as well. The enormous challenge posed by this problem is to understand and quantify the nature of surgical skill decay and develop a set of methodologies for training interventions to prevent that decay that minimizes training time, maximizes efficacy, and reduces mistakes during the initial portion of deployments.

The Department of Defense estimates a need to train 100,000 military health care personnel annually, representing a profound educational challenge.[1] The consequences of ineffective medical training are dire. In the United States, medical errors are estimated to result annually in at least 50,000 excess deaths and 1,000,000 avoidable injuries.[2] The military has long pursued the integration of simulation and robotic technologies into medical training and applications, and these techniques may provide leverage to address this issue.

Specifically, the use of these intelligent technologies for training military medical personnel can help measure skills, minimize errors, schedule training, and control the duration and expense of training. Medical personnel transitions, both when entering an operational area, and when returning to the civilian theater, are characterized by the need to rapidly reacquire skills that have decayed through disuse.

During this period of skill reacquisition, there is an increased risk of mistakes that may result in death and injury to patients. It is urgent that we increase the speed of reacquisition of surgical skills, while avoiding regaining the necessary skills on actual patients. The use of immersive virtual reality (VR) techniques coupled with metric-driven scheduling of training has the potential to dramatically reduce the cost of training, and the cost both in lives and dollars of errors and mistakes caused by lack of fluency in necessary techniques.

Although we are situating our solution within a particular surgical domain, our focus is on generalizing a theory of

*University of Nebraska Medical Center, 984220 Nebraska Medical Center, Omaha, NE 68198.

†Adaptive Cognitive Systems, 909 Harris Avenue, Suite 202D, Bellingham, WA 98225.

‡University of Central Florida, 4000 Central Florida Boulevard, Orlando, FL 32816.

§University of Nebraska Medical Center, 985126 Nebraska Medical Center, Omaha, NE 68198.

‖College of Information Sciences and Technology, Penn State University, University Park, PA 16802.

Part of this article was presented in poster format at the 2014 Military Health System Research Symposium, Fort Lauderdale, FL, August 19, 2014.

doi: 10.7205/MILMED-D-15-00164

skill decay to account for differential decay across skills that vary in their dependence on cognitive and psychomotor elements that will apply more broadly to medical skills. The particular domain we will focus on for preliminary work is laparoscopic surgery (LS), which we have chosen for its widespread use in the civilian theater, its affinity for study through simulation, and the precision of data that can be collected during its practice.

The skills required for LS, a minimally invasive surgery, transfer poorly from proficiency at open surgery on the same procedures, and the need for special training to acquire the fundamentals of LS has long been recognized.[3] Laparoscopic skills are difficult to acquire and maintain, however, because of both their technical complexity and the challenging physical environment in which they must be executed, which includes a spatially restricted monocular visual field, limited tactile feedback, and a confined working area. Training for LS in an operating theater, the traditional approach to surgical training, is both costly and time consuming,[4] limiting the time that surgeons can spend learning and practicing.

As a result, designers of medical curricula for minimally invasive surgery have turned to simulated environments in an effort to reduce or prevent the attrition of these critical surgical skills. Introductory LS training through simulation is now widespread, but has been primarily limited to novice laparoscopic surgeons practicing on simple psychomotor tasks, such as suturing. However, simulation training could be useful for training more complex tasks for surgeons with a wider range of experience, and for maintaining skills over periods of disuse. Few studies have examined durability of simulator-based training generally.[5,6] Moreover, little attention has been given to learning important cognitive skills or more complex tasks involving sequences of simple tasks.

To increase our understanding of learning cognitive skills during surgical training, we developed several simple surgical tasks (e.g., peg transfer [PT], needle passing [NP]) in our adaptive VR trainer for LS training.[7] These fundamental surgical tasks were used to train important basic surgical skills for complex and advanced tasks learning; for example, bimanual coordination, precision, and manipulations are the simple, but fundamental surgical skills for suturing.

Our adaptive training framework (Fig. 1) consists of three levels of design (modeling, comparison, and optimal training). A modeling methodology was developed including: (1) cognitive task analysis, to derive an ontology of the knowledge and skills to be measured and trained, (2) mathematical modeling, to determine the domain- and individual-specific variation in skill acquisition and attrition, and (3) cognitive modeling, to embed the specific model of skill attrition within a more general model of learner behavior, which can then be combined with the ontology derived from the task analysis.

At the second level of performance comparison, human subject data were collected from learners with varying levels of surgical experience, novices to expert surgeons. Our cognitive model at the first level is derived from and subsequently tested against the empirical human learning and forgetting data. We used the Adaptive Control of Thought/Rational (ACT-R) cognitive architecture[8,9] to model the learner's learning and forgetting. While these architectures are labeled "cognitive," cognitive architectures have been developed to provide complete processing models including the entire range of cognitive, perceptual, and psychomotor behavior,[10] and ACT-R in particular is an implementation of a unified theory of cognition.[11] As such, ACT-R includes distinct modules for perceptual processing (visual and auditory), motor behavior, memory, and skill acquisition. It is exactly

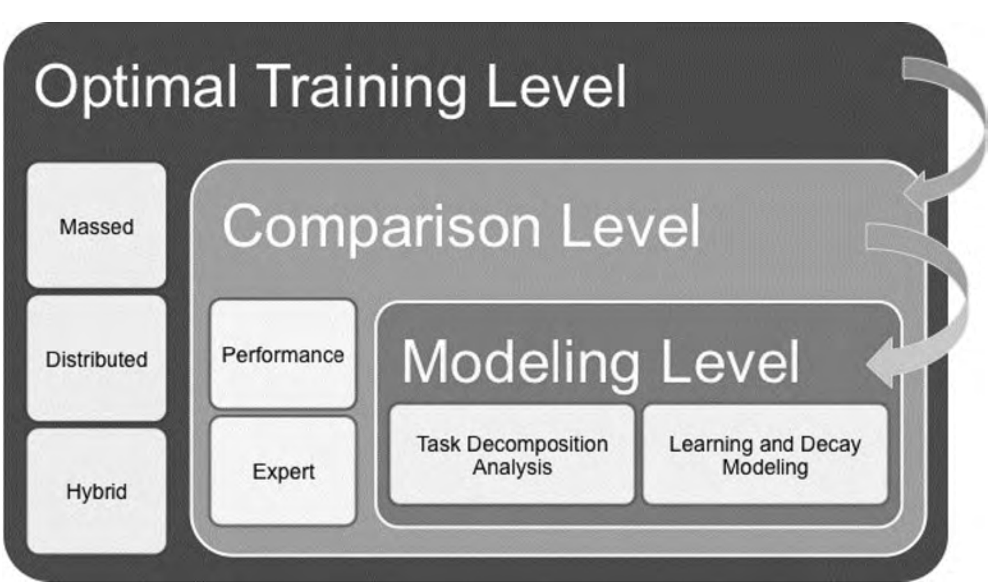

FIGURE 1. Adaptive training framework.

this breakdown that we intend to leverage in identifying the emphasis on cognitive and psychomotor elements within skills. The human data were used to test potential interventions for the third training level based on objective measures (kinematics and electromyography [EMG]).

We then implemented the third level of optimal training schedules for the cognitive and psychomotor tasks to be learned. Our adaptive VR trainer will incorporate a cognitive model of skill acquisition and retention. We based it on previous work on the user modeling of learning,[12] combining a hierarchical representation of surgical skills, where more complex skills and tasks are represented at higher levels, and training is driven by an estimate of individual skills and ability, as well as the dependence of the task on individual processing channels (e.g., cognitive, perceptual, and psychomotor). The model would tailor the training session to the level of complexity appropriate for the trainee at that moment in time and to predict and prescribe the course of training needed to produce a desired future level of competence, based on both demonstrated competence expected decay.

The Skill Retention Theory

As mentioned before, we can investigate surgical performance by considering skill acquisition and decay. A consensus understanding has been proposed, which specifies that there are continuous stages of learning.[13–16] Many theories propose a three-stage process of learning: (a) the first stage for acquiring declarative knowledge to perform a procedural task, (b) the second stage for consolidating the acquired knowledge, and (c) the final stage for tuning the knowledge toward overlearning. Based on this understanding, Figure 2 shows the three different stages of learning and forgetting, providing important insights about how forgetting would be different for the learners at each stage.[17]

The First Stage: Declarative

In the first stage of learning, skill acquisition occurs and simple training focused on skill acquisition may be adequate. For this first stage of learning and forgetting, knowledge in declarative memory degrades with lack of use, perhaps catastrophically as indicated by X's in Figure 2, leading to the inability to perform the task. In this stage, learning and forgetting are explained by the activation mechanism in ACT-R. With lack of use, the strength of declarative memory declines. Decreased memory strength leads to response time increasing and accuracy decreasing. In addition, the ACT-R theory explains that increases in working memory load leads to decrements of retrieval performance from memory based on the activation mechanism.[18] Thus, performance with this level of knowledge decreases with increased working memory load.

The Second Stage: Associative

In the second stage of learning, task knowledge is represented with a mix of declarative and procedural memory. With lack of use, the declarative knowledge can be forgotten, leading to missed steps. Procedural memory, on the other hand, is basically immune to decay. Forgetting slopes in this stage could vary by subtasks because mixed knowledge decays at different rates. In the first and second stage, catastrophic memory failure can occur because the declarative knowledge is not fully activated. In this mixed stage, training should be

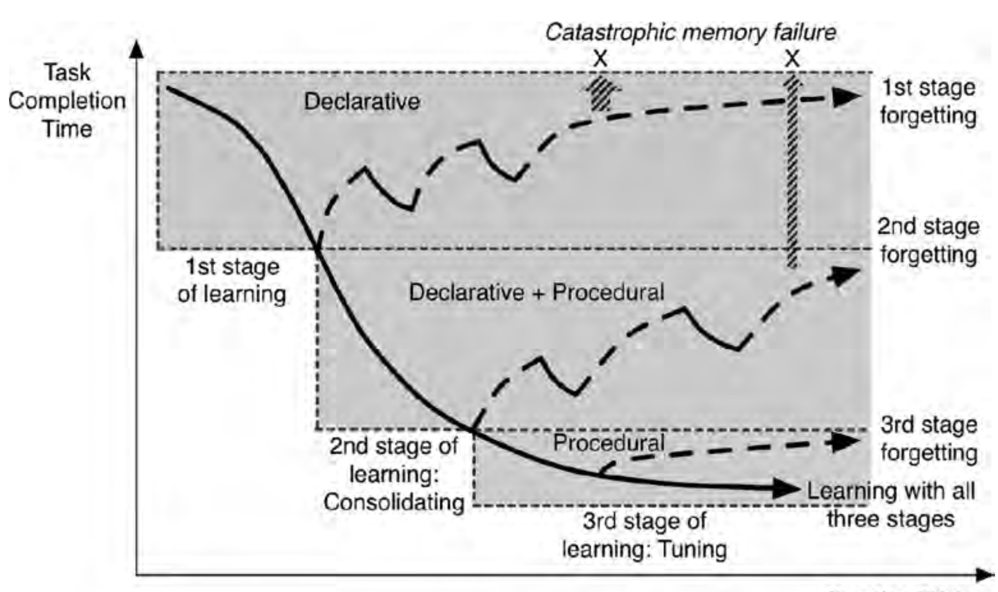

FIGURE 2. A theory of skill retention, showing the three stages of learning and forgetting. The solid line indicates a learning curve and the dashed line indicates a forgetting curve from each corresponding stage. At each stage, the learning and forgetting rates are different.

provided to keep declarative knowledge active and also to support further proceduralization.

The Third Stage: Procedural

In the third stage of learning, task knowledge is available in both declarative and procedural forms, but procedural knowledge predominantly drives performance. Practice will compile knowledge into procedural knowledge. We describe this type of task knowledge to be proceduralized skill. With lack of use, declarative knowledge may be degraded. Nevertheless, learners can still perform the task—if all the knowledge is proceduralized and thus not forgotten with time. Less well-known skills that are infrequently used, like recovery from unusual errors, may be degraded. This type of skill would require knowledge retrieval from declarative memory unless task knowledge is proceduralized. In this final stage, or to reach this final stage, practice for proceduralization should be provided. It also suggests that training should occur until trainees reach the crossing thresholds, noted as dashed horizontal lines for the stage thresholds in Figure 2.

To identify the possible solution to optimize the military medical skills acquisition and retention, the purpose of this project is to develop a methodology using our adaptive VR trainer to study surgical skill decay and test an intervention based on the integration of cognitive models of surgical skill decay within our adaptive training framework. This testing framework can predict the decay effect and maximize the training experience by monitoring the occurrence of mistakes during skill acquisition and retention.

METHODS

We collected data from 5 novices and 4 medical trainees. Each participant performed two basic surgical tasks (Figs. 3A and 3B), 5 times in three sessions: at baseline, 1 week after baseline, and 1 month after baseline. Kinematic data including time to task completion, total distance traveled, and average speed of both hands were recorded. Muscle effort of four muscles (upper trapezius, anterior deltoid, flexor carpi radialis, and extensor digitorum) was monitored using a wireless Trigno EMG system (Delsys, Boston, Massachusetts). Using methods from our previous studies,[19,20] raw EMG signals were recorded with a sampling rate of 2,000 Hz using EMG works acquisition software based on the manufacturer's recommendation, and were processed with a band-pass filter of 20–300 Hz and smoothed by a root-mean-square technique with a 150 ms moving window to compute the root-mean-square EMG data. To reduce the intersubject variation, the maximal voluntary contraction (MVC) was obtained from each muscle to normalize EMG signals.[21] The EMG data are presented as the percentage of MVC (%MVC).

RESULTS

All participants were able to complete the task. We examined their performance in terms of kinematics and EMG.

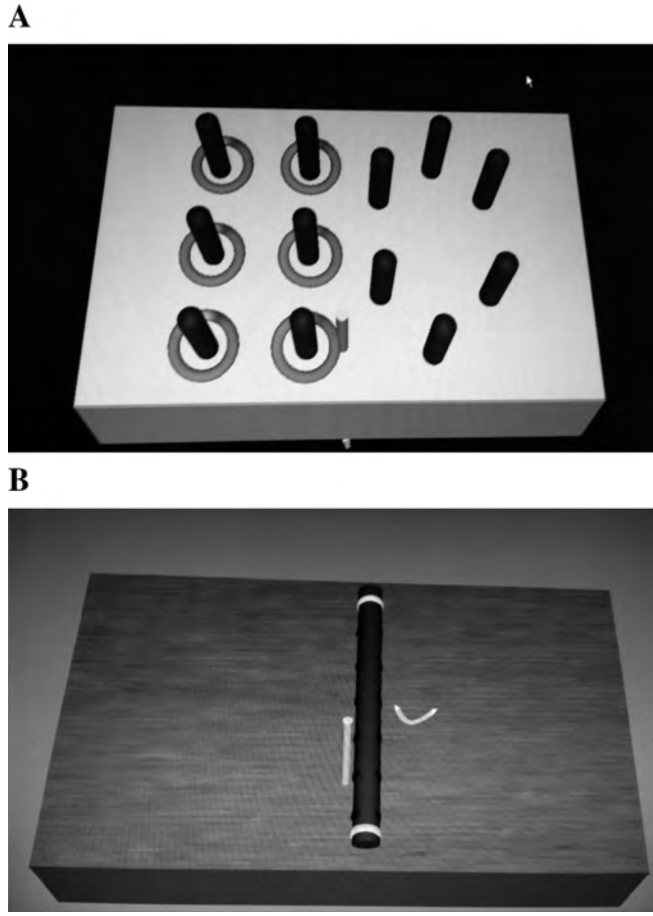

FIGURE 3. Virtual training tasks: (A) Peg transfer task and (B) Needle passing task.

Kinematics

On the PT task, participants completed tasks during the first session in 153.3 seconds on average, whereas the tasks during subsequent sessions were completed more quickly in 132.1 and 123.4 seconds, respectively. Similarly, collapsing across sessions, trial 1 was completed in 153.4 seconds while subsequent trials within the session were completed more rapidly (Fig. 4). The NP task showed a similar pattern, with participants completing the baseline session trials in 120.5 seconds on average, and more quickly in 96.1 and 91.7 seconds in subsequent sessions (Fig. 5).

Given the expectation that completion time during learning follows a power law, a statistical analysis was conducted using the log of completion time (LogTime). For task PT, we performed a regression of session and trial on LogTime, producing an R^2 of 0.13 (degrees of freedom [df] = 121). For the PT task, session 3 (at 4 weeks) was significantly faster than prior sessions ($p < 0.01$), whereas trials 3 and 4 were significantly faster than trials 1, 2, and 5 ($p < 0.05$). Similarly, for task NP, we performed a regression of session and trial on LogTime, producing an R^2 of 0.16 (df = 125). Unlike the PT task, for the NP task, both sessions 2 and 3

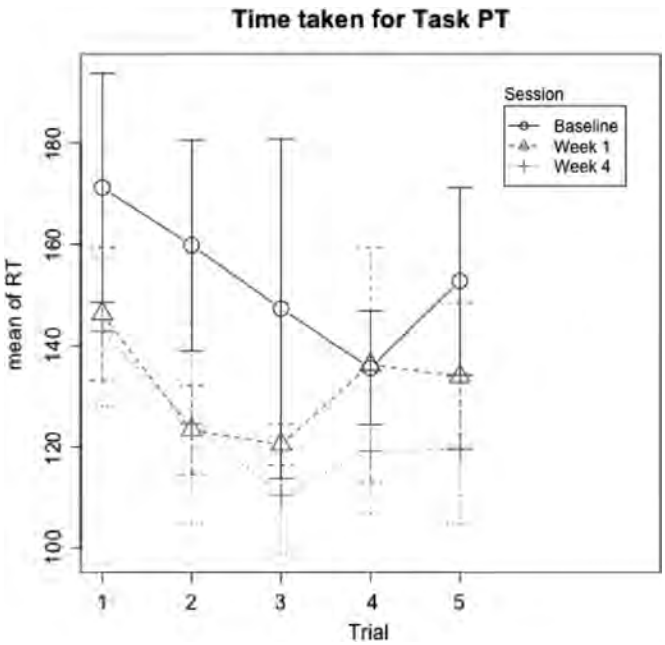

FIGURE 4. Time to task completion for the peg transfer (PT) task across trials (mean ± standard errors) (RT, reaction time).

(at 1 week and 4 weeks) were significantly faster than prior sessions ($p < 0.05$ and $p < 0.01$, respectively). Further, for the PT task, trials 3, 4, and 5 were significantly faster than trials 1 and 2 ($p < 0.05$, $p < 0.01$, and $p < 0.01$, respectively).

We fit a preliminary predictive performance model to these data using the power law of learning and forgetting to make predictions, using the regression model of session and trial on LogTime for comparison. The initial model for the PT task produces an R^2 of 0.72 and a coefficient of variation of 0.09, indicating the model accounts for the majority of the variance in the human data. The preliminary model of the NP task produces an R^2 of 0.78 and a coefficient of variation of 0.15, similarly indicating that the model predicts the majority of the variance in the human performance data. Graphs of the model predictions are presented in Figures 6 and 7.

Qualitatively, the NP and PT task models predict the appropriate range of variation, with both models predicting the greatest speedup during the first session and decay in learning across sessions that effectively rewind this learning. Further, they capture the appropriate range of performances with the PT model spanning 171 to 114 seconds, and the NP model ranging from 142 to 90 seconds, in line with the human performance.

EMG

The EMG data were further analyzed within task by converting EMG data (%MVC) to a z-score measure, thereby controlling for individual differences in overall muscle activation. The EMG z-scores were averaged to provide an overall indication of muscle activation during task performance. Because of the normalization transformation, the z-score means for tasks PT and NP are both 0, enabling the collapsing of the two tasks. We performed a regression analysis of session and trial on the z-score of the EMG muscle activation, and found that session 3 (week 4) was characterized by statistically significantly less muscle activation ($p < 0.001$, df = 255) than sessions 1 and 2 (weeks 0 and 1).

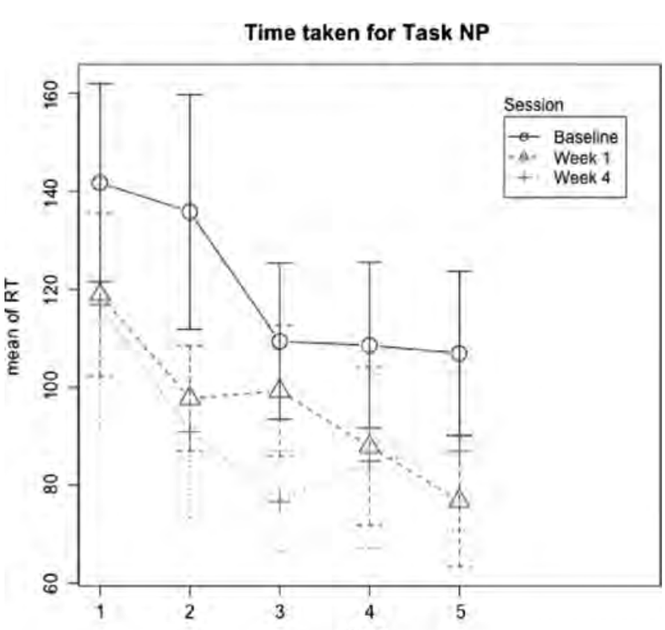

FIGURE 5. Times to task completion for needle passing (NP) task across trial (mean ± standard errors) (RT, reaction time).

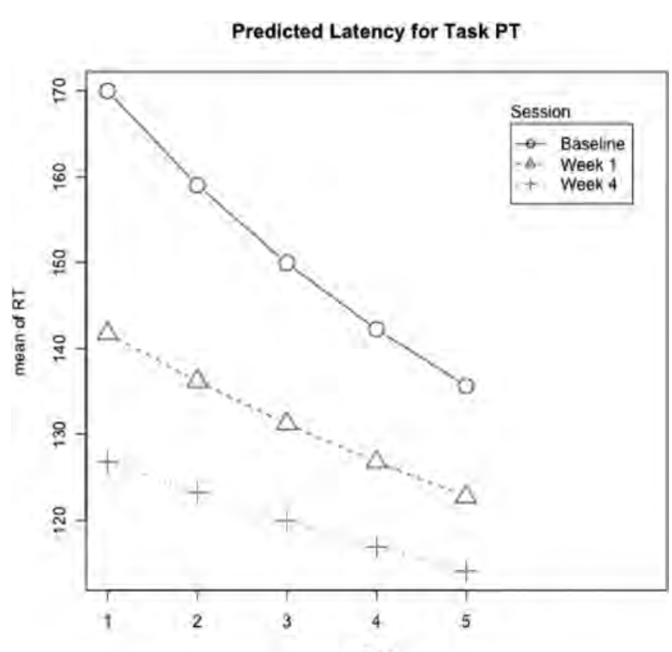

FIGURE 6. Performance model predictions for the peg transfer (PT) task (RT, reaction time).

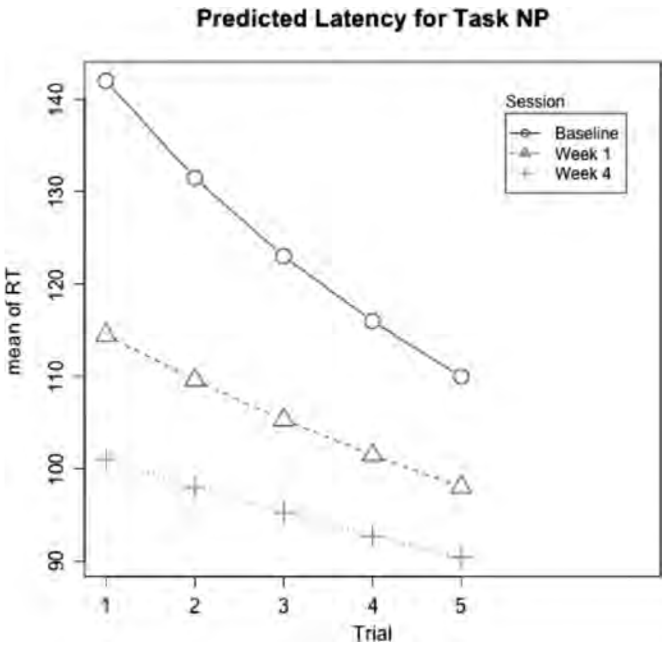

FIGURE 7. Performance model predictions for the needle passing (NP) task (RT, reaction time).

Further, trials 3, 4, and 5 had significantly less muscle activation than trials 1 and 2 ($p < 0.05$, df = 255).

DISCUSSION

In the course of working through practice sessions on the NP and PT tasks, participants clearly demonstrated both learning and forgetting, demonstrating statistically significant patterns. Participants exhibited speedup during skill acquisition and decay of that acquired skill during periods of disuse. Participants also showed less decay due to disuse as the skill became more practiced. Although the pattern of reaction time could be expected to fit a logarithmic regression model, our efforts to fit such a model only resulted in an R^2 of 0.16 and 0.13, thus failing to account for the majority of the variance in participant completion times.

These patterns were much more successfully predicted through a model employing the Power Laws of Learning and Forgetting, encapsulated within the ACT-R cognitive architecture. This modeling, although preliminary, captures the majority of the variance available in the data set, with an R^2 of 0.72 for the PT task, and an R^2 of 0.78 for the NP task. Thus, the ACT-R based model captured substantially more variance than a logarithmic regression model.

The data appear to suggest that there is also a fatigue effect at work, and participants slow down after three or four trials of either the PT or NP task, though this pattern did not reach statistical significance within the small sample we evaluated. Although we have previously modeled such slowdowns,[22] we have not yet attempted to apply a fatigue component to our modeling work. We would, however, expect to capture even more of the variance in the human performance through such a mechanism.

The EMG analysis confirms that the muscle activation required for the task was decreasing, and thus the procedural aspects of the task were becoming simpler. That is, the low-level motor learning appeared to be durable, but the task performance still shows decrements that are not accounted for by the durability of the motor skill. Our theory, however, accounts for the forgetting that takes place at the declarative and procedural levels as well, and predicts this time course of change over skill disuse.

Our future work will use further data analysis at the motor, declarative, and procedural levels to make holistic performance predictions. These research efforts will eventually help to address the maintenance of surgical skills, especially for experienced surgeons, and combat surgery. More importantly, the use of our adaptive VR trainer could measure medical skills in military medical personnel, minimize errors, schedule training, and potentially control the duration and expense of military medical training.

While preliminary, this work demonstrates the ability to accurately predict the acquisition and decay of surgical skill within a VR training system by leveraging the laws of learning and forgetting embedded within a cognitive architecture. Thus, this study serves as a validation for both the VR platform and the cognitive modeling paradigm. Given these components, we expect to be able to plan surgical training remediation with the aim of making the best use of available training time and avoiding costly mistakes because of skills that fall into disuse.

CONCLUSIONS

Based on the learning theory and this and similar results, our adaptive VR training system can improve the capability to interactively assess the level of skills learning and decay, optimize skill relearning across levels of surgical experience, and positively impact skill maintenance. Our training system could eventually reduce patient injury and morbidity by providing trainees with the reexperience they need before operating in a new theater.

ACKNOWLEDGMENTS

This study was supported by a NASA Nebraska Space Grant and EPSCoR (KCS), a Nebraska Research Initiative grant (KCS and DO), and grants from ONR No. W911QY-07-01-0004, N000141010401, and N00015-01-2275 (FER).

REFERENCES

1. 2008 Annual Report of the TATRC, U.S. Army Medical Research and Materiel Command, Telemedicine and Advanced Technology Research Center (TATRC), 2008. Available at http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA510903&Location=U2&doc=GetTRDoc.pdf; accessed March 1, 2016.
2. Committee on Quality of Health Care in America, Institute of Medicine; Kohn LT, Corrigan JM, Donaldson MS (editors): To Err is Human: Building a Safer Health System. Washington, DC, National Academies Press, 2000.

3. Soper NJ, Fried GM: The fundamentals of laparoscopic surgery: its time has come. Bull Am Coll Surg 2008; 93(9): 30–2.
4. Bridges M, Diamond DL: The financial impact of teaching surgical residents in the operating room. Am J Surg 1999; 177: 28–32.
5. Stefanidis D, Korndorffer JR Jr, Sierra R, Touchard C, Dunne JB, Scott DJ: Skill retention following proficiency-based laparoscopic simulator training. Surgery 2005; 138: 165–70.
6. Stefanidis D, Korndorffer JR Jr, Markley S, Sierra R, Heniford BT, Scott DJ: Closing the gap in operative performance between novices and experts. J Am Coll Surg 2007; 205: 307–13.
7. Garcia-Ruiz A, Gagner M, Miler JH, Steiner CP, Hahn JF: Manual vs robotically assisted laparoscopic surgery in the performance of basic manipulation and suturing tasks. Arch Surg 1998; 133(9): 957–61.
8. Anderson JR: How Can the Human Mind Occur in the Physical Universe? New York, Oxford University Press, 2007.
9. Anderson JR, Bothell D, Byrne MD, Douglass S, Lebiere C, Qin Y: An integrated theory of mind. Psychol Rev 2004; 111(4): 1036–60.
10. Newell A: You can't play 20 questions with nature and win: projective comments on the papers of this symposium. In: Visual Information Processing, pp 283–308. Edited by Chase WG. New York, Academic Press, 1973.
11. Newell A: Unified Theories of Cognition. Cambridge, Harvard University Press, 1990.
12. Park SH, Suh IH, Chien JH, et al: Modeling surgical skill learning with cognitive simulation. Stud Health Technol Inform 2011; 163: 428–32.
13. Fitts PM: Perceptual-motor skill learning. In: Categories of Human Learning, pp 243–85. Edited by Melton AW. New York, Academic Press, 1964.
14. Anderson JR: Acquisition of cognitive skill. Psychol Rev 1982; 89(4): 369–406.
15. Rasmussen J: Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering. New York, Elsevier, 1986.
16. VanLehn K: Cognitive skill acquisition. Annu Rev Psychol 1996; 47: 513–39.
17. Kim JW, Ritter FE, Koubek RJ: An integrated theory for improved skill acquisition and retention in the three stages of learning. Theor Issues Ergon 2013; 14(1): 22–37.
18. Anderson JR, Reder LM, Lebiere C: Working memory: activation limitations on retrieval. Cogn Psychol 1996; 30(3): 221–56.
19. Siu KC, Suh IH, Mukherjee M, Oleynikov D, Stergiou N: The impact of environmental noise on robot-assisted laparoscopic surgical performance. Surgery 2010; 147: 107–13.
20. Suh IH, Mukherjee M, Schrack R, et al: Electromyographic correlates of learning during robotic surgical training in virtual reality. Stud Health Technol Inform 2011; 163: 630–4.
21. Soderberg GL, Knutson LM: A guide for use and interpretation of kinesiologic electromyographic data. Phys Ther 2000; 80: 485–98.
22. Gonzalez C, Best B, Healy AF, Kole JA, Bourne LE: A cognitive modeling account of simultaneous learning and fatigue effects. Cogn Syst Res 2011; 12: 19–32.