Providing Objective Feedback on Skill Assessment in a Dental Surgical Training Simulator

Phattanapon Rhienmora[1], Peter Haddawy[1],
Siriwan Suebnukarn[2], and Matthew N. Dailey[1]

[1] School of Engineering and Technology, Asian Institute of Technology,
Pathumthani, Thailand, 12120
{phattanapon.rhienmora,haddawy,mdailey}@ait.ac.th
[2] Faculty of Dentistry, Thammasat University,
Pathumthani, Thailand, 12121
ssiriwan@tu.ac.th

Abstract. Dental students devote several years to the acquisition of sufficient psychomotor skills to prepare them for entry-level dental practice. Traditional methods of dental surgical skills training and assessment are being challenged by the complications such as the lack of real-world cases, unavailability of expert supervision and the subjective manner of surgical skills assessment. To overcome these challenges, we developed a dental training system that provides a VR environment with a haptic device for dental students to practice tooth preparation procedures. The system monitors important features of the procedure, objectively assesses the quality of the performed procedure using hidden Markov models, and provides objective feedback on the user's performance for each stage in the procedure. Important features for characterizing the quality of the procedure were identified based on interviews with experienced dentists. We evaluated the accuracy of the skill assessment with data collected from novice dental students as well as experienced dentists. We also evaluated the quality of the system's feedback by asking a dental expert for comments. The experimental results show high accuracy in classifying users into novice and expert, and the evaluation indicated a high acceptance rate for the generated feedback.

Keywords: Dental surgical training, skill assessment, objective training feedback, virtual reality.

1 Introduction

Dental students obtain their surgical skills training from various sources. Traditional methods rely on practicing procedural skills on plastic teeth or live patients under the supervision of dental experts. However, the limitations of this approach include a lack of real-world challenging cases, unavailability of expert supervision, and the subjective manner of surgical skills assessment. With recent advances in virtual reality (VR) technology, VR simulators for medical and dental surgery have been introduced [1], [2]. The advantages of these simulators are that the students are able to practice

C. Combi, Y. Shahar, and A. Abu-Hanna (Eds.): AIME 2009, LNAI 5651, pp. 305–314, 2009.

procedures as many times as they want at no incremental cost and that the training can take place anywhere. The realism of these simulators has increased with the introduction of haptic devices that provide tactile sensations to the users [3], [4], [5], [6].

Skill assessment in traditional training is usually conducted by having an expert surgeon observe the procedure or only the final outcome. However, the level of detail of human expert assessment is limited. With VR simulators, many aspects such as data about the environment and the user's precise actions can be recorded during the simulation and analyzed further to provide fine-grained objective assessment and feedback. Unfortunately, existing dental simulators do not provide this functionality. There has been some work in other fields, however. Rosen et al. [7] present a technique for objective evaluation of laparoscopic surgical skills using hidden Markov models (HMMs). The models are based on force/torque information obtained from a surgical robot. Lin et al. [8] collected various measurements from the da Vinci surgical robot while an operator performed a suturing task. The aim of their study was to automatically detect and segment surgical gestures, which is a part of their ongoing research on automatic skills evaluation. As the da Vinci surgical robot does not provide haptic feedback, their research did not consider force applied during the operation.

To add more educational value, simulators should be able to provide objective feedback to users in order to reduce the time and effort required for instructors to supervise and tutor trainees using the system. Thus, incorporation of strategies for generating objective feedback with quality comparable to that of human tutors is essential to the development of an efficient, *intelligent* training simulator.

In this paper, we describe the first virtual reality dental training system to combine realistic haptic feedback with an objective dental performance assessment and feedback generation mechanisms. While the system currently simulates the tooth preparation procedure, many of the techniques and strategies implemented should generalize well to other medical and dental procedures.

2 VR Tooth Preparation Simulator

The graphical user interface of our simulator is illustrated in Fig. 1. Movement of a virtual dental tool is controlled by a haptic device stylus. The detailed development of our simulator is explained in [9]. Currently the system simulates only labial and incisal preparations in order to avoid conflating tool skills with indirect vision skills.

The tooth preparation procedure requires that 13 stages be performed on the incisal and labial surfaces including, 1) mid-incisal depth cut, 2) distal-incisal depth cut, 3) mesial-incisal depth cut, 4) incisal reduction, 5) mid-upper-labial depth cut, 6) distal-upper-labial depth cut, 7) mesial-upper-labial depth cut, 8) upper-labial reduction, 9) mid-lower-labial depth cut, 10) distal-lower-labial depth cut, 11) mesial-lower-labial depth cut, 12) lower-labial reduction, and 13) labial marginal preparation. Examples of simulated tooth preparation outcomes are shown in Fig. 2.

We tested the ability of our simulator to produce outcomes that reflect operator skill. Ten simulated preparation outcomes completed by five students and five experienced dentists were shown to another expert, who was not aware of the nature of the experiment, who was asked to assign outcome scores based on errors found in Incisal, Labial-incisal, Labial-gingival, and Marginal. The maximum score was 16. The

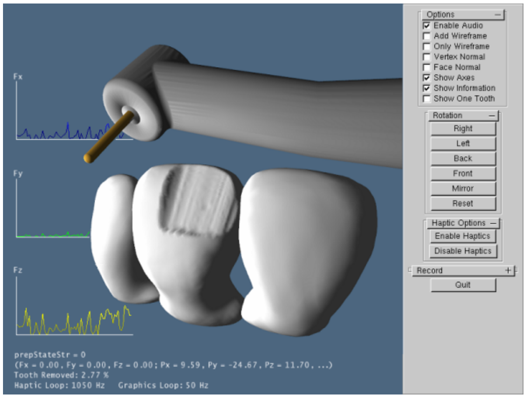

Fig. 1. Graphical user interface of our VR tooth preparation simulator

Fig. 2. Example of two outcomes of tooth preparation on the labial and incisal surfaces: an expert outcome *(left)*; a novice outcome *(right)*

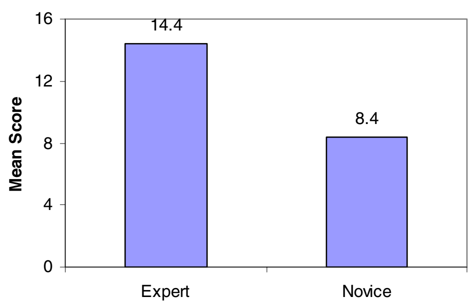

Fig. 3. Mean scores for simulated tooth preparation performed by expert *(left)* and novice *(right)*. Mean expert score = 14.4, SD = 0.89; mean novice score = 8.4, SD = 1.14.

experts' mean score (14.4) was significantly difference than the novices' mean score (8.4) ($p < 0.05$) as shown in Fig. 3. This result indicates that the simulator captures the important aspects of the physical environment.

3 Objective Assessment of Dental Surgical Skills

The current means for evaluating clinical performance and skill acquisition during training are limited to measurement of task completion time and number of errors or a subjective evaluation by an expert [10]. The aforementioned measures do not characterize the operator's movements (e.g., position, orientation, or speed). While speed is closely related to task completion time, faster is not necessarily better; the speed–accuracy trade-off is a well-known phenomenon in motor control, in which speed increases cause decreases in accuracy and vice versa [11]. More accurate movements may take more time to complete. Therefore, additional objective measures are needed to quantify surgical performance improvements and differentiate between expert and novice surgeons.

Based on interviews with experienced dentists, we hypothesized that important features for distinguishing experts from novices in dental surgery are tool movement

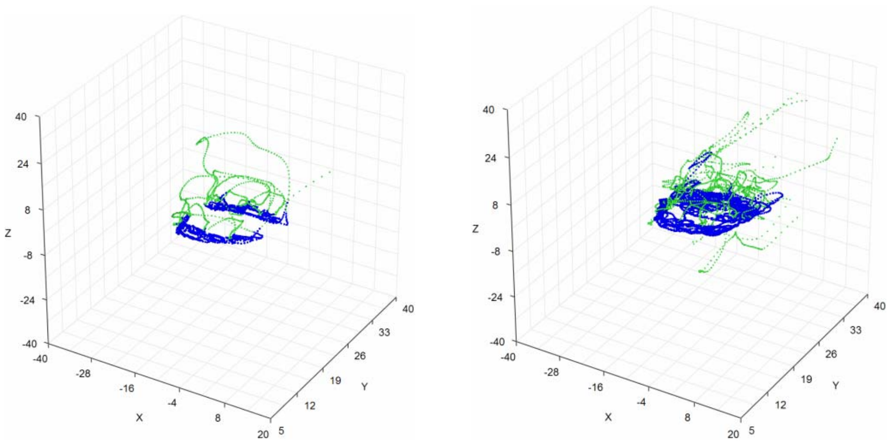

Fig. 4. Example of tool paths of an expert *(left)* and a novice *(right)*. *Darker* paths indicate cutting operation.

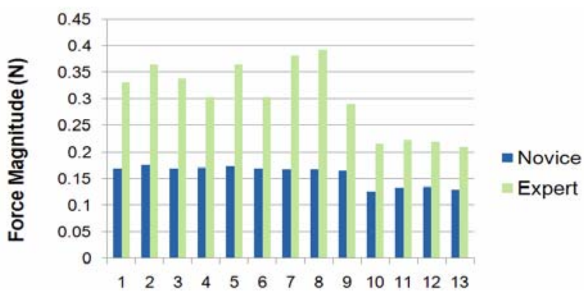

Fig. 5. An example of average force applied by an expert and a novice during 13 stages of simulated tooth preparation

(position and orientation of the tool) and applied force during a procedure. We can visualize these features by plotting tool movement of an expert and a novice in three dimensions as shown in Fig. 4 and the average magnitude of the force applied by an expert and a novice over time as shown in Fig. 5.

It can be clearly seen from Fig. 4 that expert and novice performance in tool movement is different. Expert movement, especially while cutting, is more consistent than novice movement. From Fig. 5, the force used at each stage of the procedure for experts and novices are also different. These measures can provide a foundation for quantifying surgical expertise and can be used for objective skill assessment.

3.1 Experiment

The main objective of this experiment was to test the ability of a machine learning technique, the hidden Markov model, to recognize and classify an observed procedure as novice or expert, based on a set of recorded important features.

Five novices (forth-year dental students, ages 20-22 years) and five experts in prosthodontics (ages 35-45 years) were recruited to participate in this study. All participants were right-handed.

Their task was to perform a tooth preparation on the upper left central incisor with the simulator. Experts and novices performed five trials of the task. The last trial was used for data analysis.

3.2 Evaluation

When a user performs tooth preparation on our simulator, all of the data relevant to the user's actions are monitored and recorded to a file. This data includes all important features mentioned previously as well as the active status of the drill and the indices of the vertices being cut on the tooth surface. We manually labeled the preparation stage transitions in order to facilitate later evaluation of automatic stage segmentation strategies.

After collecting the data from all participants, we developed discrete hidden Markov models (HMMs) to classify procedure sequences as novice or expert. In our model, the hidden states are the thirteen stages of tooth preparation. The observed feature set includes force calculated during the simulation as well as positions and orientations of the dental tool. Stage labels were not used in training HMMs. Since we use discrete HMMs, we first converted the feature vectors into symbols using the k-means clustering algorithm with $k = 13$. After training, we calculated the probability and log likelihood of test sequences under the novice and expert HMMs. If the log likelihood of the test sequence under the novice HMM is greater than that under the expert HMM, the system classifies the test sequence as a novice sequence; otherwise, the system classifies it as an expert sequence.

3.3 Result and Discussion

We used a different k-means for every cross validation fold and the same k-means for the novice and expert model in the same fold. For each fold, we trained the novice HMM with four novice and four expert sequences. To determine the accuracy of the

Table 1. Average log likelihood results for expert and novice performance sequences

	Log likelihool for Expert HMM	Log likelihool for Novice HMM
Expert Performance	-3.574×10^{3}	-2.229×10^{6}
Novice Performance	-6.272×10^{5}	-3.494×10^{3}

method, after training the two HMMs in each fold, we fed the test novice and expert data to each model. The average log likelihood of all sequences across all five folds for the two HMMs is shown in Table 1.

For every cross validation fold, the log likelihood of a test sequence for its corresponding HMM was higher than that for another HMM. The result demonstrated the ability of HMM to distinguish between novice and expert performance with 100 percent accuracy. However we note that the number of participants (ten) was relatively small.

4 Strategies for Objective Feedback Generation

The stage of the tooth preparation procedure and its unique force/position/orientation characteristic are the basis of our feedback generation mechanism. The average position, orientation, force, and main axis for force direction differ between procedure stages. In stage 1) (Fig. 6), for example, force and tool movement is mostly in the minus Y direction, while in stage 5) (Fig. 6) they progress mostly in the minus Z direction. These characteristics can be observed by the simulator and compared to a gold standard in order to generate useful feedback. Examples of our feedback strategy considering applied force for stage 1) and 5) are shown in Table 2.

Fig. 6. Examples of stage 1) mid-incisal depth cut *(left)*, stage 5) mid-uppper-labial depth cut *(middle)* and stage 9) mid-lower-labial depth cut *(right)*

Table 2. Examples of feedback generated in stages 1) and 5) considering only applied force. Subscript *e* indicates the expert average value (out of five experts) with one standard deviation while *n* indicates the current novice value. The full table considers every feature (force, position, and orientation) and covers all 13 stages for each novice.

Stage	F_x (N)	F_y (N)	F_z (N)	Feedback
1_e 1_n	0.103±0.037 0.026	0.480±0.047 0.164	0.106±0.023 0.091	*"Force in minus Y direction should be 3 times higher"*
5_e 5_n	0.040±0.014 0.028	0.038±0.019 0.019	0.237±0.053 0.129	*"Force in minus Z direction should be 2 times higher"*

We generate feedback for position and orientation with the same strategy. For example, Fig. 7 shows a stage in which a novice's tool orientation was too different from that of the expert. In this case the feedback generated was *"try to lower the degree of rotation around X axis."* For states in which the operator does well, we generate a compliment such as "*well done.*"

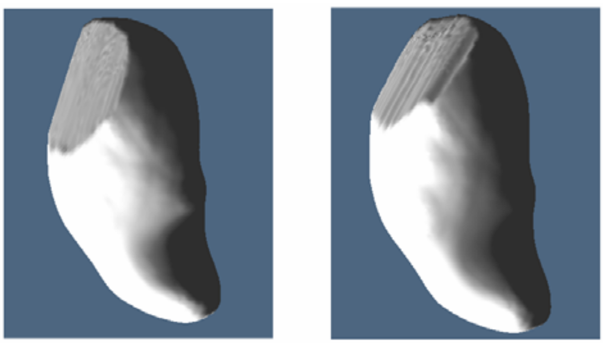

Fig. 7. Example of a difference in tool orientation between expert *(left)* and novice *(right)*

4.1 Experiment

The main objective of this experiment was to test the overall acceptability of the training feedback generated by the simulator.

The simulator loaded data files of all five novices collected during the previous experiment described in section 3.1 and replayed the procedure, one novice at a time. During playback, the system observed the characteristics of each stage, computed statistical results, compared them with those of expert performances, and then generated and displayed the tutoring feedbacks on the screen. An expert examined both the replay of the novice procedure and the feedback generated by the system. The corresponding force values in three axes were also plotted on the screen during replay to aid understanding of how the forces were applied by the operator.

During the experiment, a total of 65 tutoring feedback messages were generated. The expert was asked to rate the acceptability of each feedback message on a scale of 1-5, where 1 implied unacceptable, 2 implied not quite acceptable, 3 implied not sure, 4 implied close to acceptable and 5 implied acceptable.

4.2 Result and Discussion

Please see Fig. 6 as a reference for the desired outcomes of stages 1), 5) and 9).

From Table 3, during stage 5), where the main force should be applied in the minus Z direction, the average force applied by a user in this direction was not within one standard deviation from the expert mean (0.184 N – 0.290 N). Since the novice's average force was around half that of the expert, the generated feedback, *"Force in minus Z direction should be 2 times higher"*, was rated as *acceptable* (score 5).

Table 3. Part of the expert evaluation form for stages 1), 5) and 9). Subscript *e* indicates the expert average value (out of five experts) with one standard deviation while *n* indicates the current novice value. The full evaluation form contains all 65 cases and shows all features (force, position, and orientation) considered in the feedback generation mechanism.

Stage	F_x (N)	F_y (N)	F_z (N)	Feedback	Acceptability
1_e 1_n	0.103±0.037 0.026	0.480±0.047 0.164	0.106±0.023 0.091	*"Force in minus Y direction should be 3 times higher"*	4
5_e 5_n	0.040±0.014 0.028	0.038±0.019 0.019	0.237±0.053 0.129	*"Force in minus Z direction should be 2 times higher"*	5
9_e 9_n	0.064±0.024 0.108	0.035±0.019 0.115	0.285±0.033 0.159	*"Force in minus Z direction should be 2 times higher"*	3

For stage 9), however, even though the situation in minus Z direction was almost the same as in stage 5), the feedback (*"Force in minus Z direction should be 2 times higher"*) was rated as *not sure* (score 3). The expert noticed that, during this stage, the force value in X and Y were quite high although they should have been close to zero. There might be two causes for this behavior; either the novice did not know the main direction of the force in this stage (minus Z) or he/she knew but could not control the tool to move in the right direction. The expert suggested giving a tutoring hint such as "*Do you know that minus Z should be the main direction of force in this stage?*" This kind of hint would be especially useful in online training as the system can observe a novice's reaction after the feedback is given. Note that even though we have not yet applied this strategy, the system was capable of detecting the behavior as forces in X and Y (0.108 N and 0.115 N respectively) were also higher than one standard deviation from the expert means.

For stage 1), the generated feedback, *"Force in minus Y direction should be 3 times higher"*, was *close to acceptable* (score 4). The expert commented that a novice could

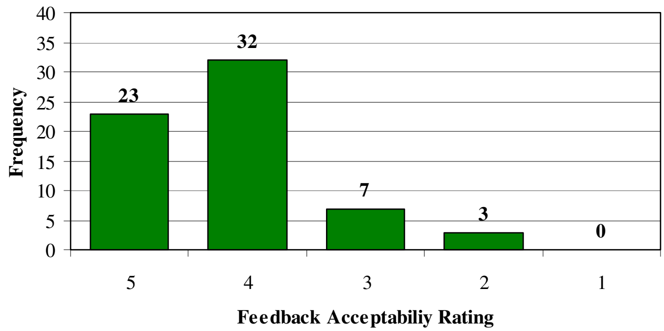

Fig. 8. Distribution of feedback acceptability ratings for 65 generated feedback messages. The average score was 4.154.

accidentally damage a tooth in this stage if he/she tried to applied force too much, therefore, suggested feedback could possibly be only *"2 times higher"* (instead of 3).

The acceptability ratings for all 65 training feedback messages generated by the system are shown in Fig. 8. The average score assigned by the expert for the generated feedback was 4.154 out of 5.

5 Conclusion

In this paper, we describe a mechanism for providing objective feedback on skill assessment using our dental training simulator. After a procedure is done, the simulator is able to classify the performance of a particular operator as novice-level or expert-level based on the force applied, tool position, and tool orientation using a hidden Markov model. Moreover, the simulator can later generate tutoring feedback with quality comparable to the feedback provided by human tutors. Additional tutoring strategies will be explored in the future work. The evaluation results are promising and prove the applicability of the simulator as a supplemental training and performance assessment tool for dental surgical skills.

Acknowledgments. This research was funded by grant NT-B-22-MS-14-50-04 from the National Electronics and Computer Technology Center, Thailand. The authors would like to thank Prabal Khanal for his contribution to the haptics software, and Kan Ouivirach for his HMM implementation. We also would like to thank Gajananan Kugamoorthy for many useful discussions and comments.

References

1. Gorman, P.J., Meier, A.H., Rawn, C., Krummel, T.M.: The future of medical education is no longer blood and guts, it is bits and bytes. Am. J. Surg. 180(5), 353–356 (2000)
2. Roberts, K.E., Bell, R.L., Duffy, A.J.: Evolution of surgical skills training. World J. Gastroenterol. 12(20), 3219–3224 (2006)

3. Kim, L., Hwang, Y., Park, S.H., Ha, S.: Dental training system using multi-modal interface. Computer-Aided Design & Applications 2(5), 591–598 (2005)
4. Yau, H.T., Hsu, C.Y.: Development of a dental training system based on point-based models. Computer-Aided Design & Applications 3(6), 779–787 (2006)
5. Yau, H.T., Tsou, L.S., Tsai, M.J.: Octree-based virtual dental training system with a haptic device. Computer-Aided Design & Applications 3, 415–424 (2006)
6. Wang, D., Zhang, Y., Wang, Y., Lu, P.: Development of dental training system with haptic display. In: 12th IEEE International Workshop on Robot and Human Interactive Communication, pp. 159–164 (2001)
7. Rosen, J., Solazzo, M., Hannaford, B., Sinanan, M.: Task decomposition of laparoscopic surgery for objective evaluation of surgical residents' learning curve using hidden Markov model. Comput. Aided Surg. 10(1), 49–61 (2002)
8. Lin, H.C., Shafran, I., Yuh, D., Hager, G.D.: Towards automatic skill evaluation: detection and segmentation of robot-assisted surgical motions. Comput. Aided Surg. 11(5), 220–230 (2006)
9. Rhienmora, P., Haddawy, P., Dailey, M.N., Khanal, P., Suebnukarn, S.: Development of a dental skills training simulator using virtual reality and haptic device. NECTEC Technical Journal 8, 140–170 (2006)
10. Hashizume, M., Shimada, M., Tomikawa, M., Ikeda, Y., Takahashi, I., Abe, R., Koga, F., Gotoh, N., Konishi, K., Maehara, S., Sugimachi, K.: Early experiences of endoscopic procedures in general surgery assisted by a computer-enhanced surgical system. Surg. Endosc. 16(8), 1187–1191 (2002)
11. Rose, D.J.: A multilevel approach to the study of motor control and learning. Allyn & Bacon, Boston (1997)