Effect of augmented visual performance feedback on the effectiveness of clinical breast examination training with a dynamically configurable breast model

Gregory J. Gerling
Dept. of Mechanical and Industrial Engineering
University of Iowa
Iowa City, IA U.S.A.
gregory-gerling@uiowa.edu

Geb W. Thomas
Dept. of Mechanical and Industrial Engineering
University of Iowa
Iowa City, IA U.S.A.
geb-thomas@uiowa.edu

Abstract - *Clinical breast examinations (CBE) are an important component in breast cancer screening. Inadequate training of tactile skills may contribute to a low and variable CBE sensitivity range. Based on previous investigations, we investigated adding visual feedback of pressure changes in balloons embedded in silicone under palpation to provide learners with positive confirmation of their performance, thus enhancing confidence in lump detection and improving sensitivity and specificity. The experiment tested the effects of visually augmented training by comparing training groups which receive either: a) no training (only verbal), b) palpation practice without feedback, or c) visually augmented feedback. The experiment utilized signal detection theory and included 18 medical students who completed a pretest, training, and a postest but found no significant differences between the training groups.*

Keywords: Haptic, Tactile, Training, Feedback.

1 Introduction

Clinical breast examinations (CBE) are an important component in breast cancer screening. To perform a CBE, a physician uses his fingertips to palpate a patient's breast toward her rib cage, feeling for tissue irregularities and tumors. Screening for breast cancer with CBE in addition to mammography reduces breast cancer mortality by 27% in women aged 45-64 [1]. However, many physicians receive inadequate training in tactile search of breast tissue, which may contribute to a low and highly variable CBE sensitivity range of 39-59%. While practitioners may underutilize CBE if they do not feel proficient in CBE performance [2], training can substantially improve performance and effective training tools can improve training success [3-5]. Commercial training models, breast-shaped silicone forms embedded with foreign objects representing lumps, are effective in improving lump detection [4]. Training with such models can improve detection of benign tumors in women's breasts, demonstrating a transfer of skills where skills learned with the models are successfully used in a clinical breast examination [5, 6]. After silicone model training, participants enjoy a 44-69% skill advantage over untrained participants [3]. These benefits come from studies using current breast models that typically provide only five lump conditions. The limited number of lump conditions can restrict extended practice because trainees quickly memorize the lump positions and further practice confounds memory with tactile skill. Also, studies with traditional silicone breast models also report an increase in false detections, [3, 4, 7] suggesting that training may simply increase clinicians willingness to diagnose more breast anomalies as lumps.

To address CBE training limitations, we developed and tested a breast simulator with 15 dynamically controllable lumps set to desired hardness within underlying rib and muscle structures, in a silicone breast model [8]. In a previous study of 48 medical students, we tested two hypotheses, that training with the dynamic breast model leads to: 1) higher lump detection without increasing false positives, compared with training with a static breast model; and 2) greater skill transfer to other breast models [9]. Results demonstrated that training with the dynamic model increased lump detection by 1.35 lumps compared to 0.60 lumps for a traditional breast model ($p=0.008$), reduced false positives by –0.70 lumps compared to +0.42 lumps ($p=0.028$), and demonstrated skill transfer with a 1.17 lump detection improvement on the traditional device compared to only a 0.17 lump detection improvement by traditional device trainees on the dynamic device ($p<0.001$) [9]. We then looked at measures of realism. Engineering material tests including stress/strain analysis of breast tissue and lump hardness were taken to ensure that the firmness of the dynamic and static models were similar. Additionally a subjective test with fifteen physicians showed the dynamic simulator was similar to the Mammatech breast models in simulating real tissue [8]. Findings demonstrated an advantage of the dynamic model over conventional models in training CBE tactile skills.

Next, we developed an electronic balloon inflation and pressure monitoring system. The system can inflate and monitor up to 8 balloons simultaneously and provides continuous pressure monitoring of each lump to a computer screen. The visual feedback indicates when a trainee has palpated a lump. In previous experiments, we observed that feedback to the trainee in the form of oscillating water pressure while the trainee's finger was near the balloon aided in lump localization. Once a trainee learned the feel of a lump, he or she subsequently was able to find similar lumps more easily and confidently. Thus, we were interested in ways to provide feedback to aid learning.

Several types of direct, immediate augmented performance feedback could potentially aid tactile training through the visual, auditory, or tactile senses. Feedback improves system accuracy, frequency response and stability, and provides for self-monitoring [10]. Feedback functions as a learning aid in providing outside information to the trainee regarding effective training to improve skills to a standardized level. Often before training, large, hard lumps near the surface can be palpated with ease. The goal is to gradually enhance training so that an examiner can find those lumps that are more difficult to detect (i.e., deep, soft, smaller lumps) by augmenting the trainee with feedback.

Beneficial training effects of adding user feedback have been reported in the literature. Akamatsu and MacKenzie [11, 12] studied a pointing task using a mouse-type device and the effect of adding tactile, auditory, and visual feedback individually and in combination. Vitense, et al report a multimodal interface using auditory, haptic, and visual feedback in a direct manipulation task finding that certain types of bimodal feedback can enhance performance while lowering self-perceived mental demand [13]. Real-time force and visual feedback are also provided to doctors in laproscopic and other medical procedures [14]. These authors emphasize the value of augmenting the trainee's senses with the appropriate information and recommend feedback in the development and advancement of multimodal interfaces.

2 Method

Based on our previous investigations, we believed that using visual feedback of the pressure changes in the lumps under palpation would provide learners with positive confirmation of their performance, thus enhancing confidence in lump detection and improving sensitivity and specificity. Our hypothesis is that augmenting the training experience with visual feedback will improve sensitivity and specificity.

2.1 Apparatus

The dynamic breast model inflates handmade balloons embedded in formed silicone with water to simulate lumps of controllable hardness. For this study, we used three cylinders oriented length-wise, each with an embedded balloon (approximate sizes 6 mm, 6 mm, and 8 mm). A thin piece of silicone (~1 mm) coated with baby powder covered the cylinders, Figure 1. Balloons were inflated to desired pressure with the electronic inflation and monitoring system. The balloons were undetectable when deflated. The readings from the continuous pressure monitoring system for each lump were visually displayed on a laptop computer. Trainee lump palpation increased the balloon pressure, providing visual performance feedback to the learner, Figure 2. The nominal fill pressure of the balloons was 20 psi and the sensor displayed pressure increases caused by mass palpation with a resolution of 0.1 psi, corresponding to a very light touch directly above the mass. Palpating near the mass also changed the pressure, so the subject's task was to adjust palpation position and pressure to maximize the pressure response. Maximum response occurred when the mass was directly palpated. In Figure 2, three balloons are inflated and the value of one mass changes with palpation.

For experimental pretest and posttest, silicone cylinders with embedded ball bearings were used, Figure 3. Six samples and a blank were used with different ball bearing locations and sizes.

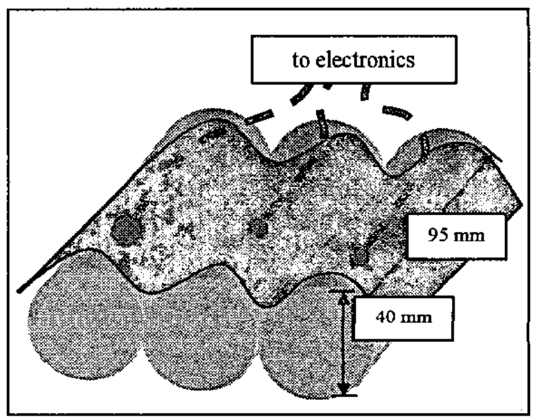


Figure 1. Training setup (3 silicone cylinders with embedded polyethylene balloons with filler tubes leading to the electronic filling and sensing device)

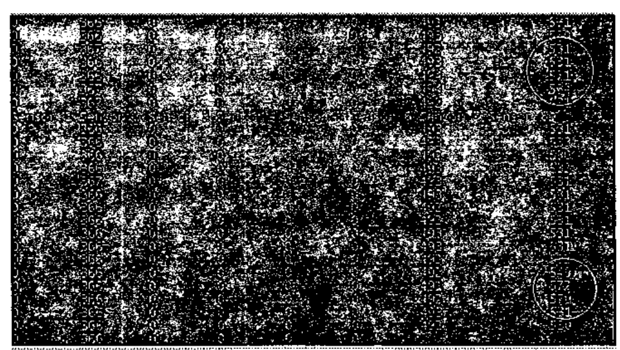
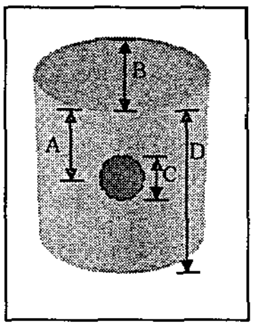


Figure 2. Computer screenshot showing three balloons inflated with one being palpated. The measure to the right has a circle where the pressure jumped briefly and a second circle with a more sustained palpation.

2.2 Participants and Protocol

The experiment tested the effects of visually augmented training by comparing training groups which receive either: a) no training (only verbal), b) palpation practice without feedback, or c) visually augmented feedback. The 18 participants were medical students at the University of Iowa. All participants signed informed consent documents for the study, which was approved by the University of Iowa Institutional Review Board. Participants were compensated $40 for their 3-hour involvement. The 8 women and 10 men were aged 23-36 with a mean age of 26. The 18 participants were assigned to three experimental cohorts, balanced by gender and year in medical school, factors that reflect prior opportunities to practice breast examination skills.



Sample	Blank	1	2	3	4	5	6
Dimension A Measure (mm)	None	40	20	40	20	40	20
Dimension B Measure (mm)	42	40	40	42	41	41	40
Dimension C Measure (mm)	None	4.8	4.8	8.2	8.2	11	11
Dimension D Measure (mm)	43	44	46	45	44	46	43

Figure 3: Silicone testing sample format with dimensions for the 6 samples and the blank

The experiment was conducted on two separate days. On the first day, participation included watching a 20-minute video and then performing a pretest. The video presented proper technique in performing a CBE with a silicone model. Following the video, participants verified their understanding of the video by reviewing CBE techniques with the research assistant. In the pretest, a single interval signal detection theory approach with 120 trials was used. The six samples were presented one by one in randomized order presented in Figure 4. In each trial, a trainee was allotted 15 seconds to palpate the sample and respond with a gradient choice to indicate confidence in the presence/absence of a lump. The confidence rating scale ranged from most confident in mass presence to most confident in mass absence, Figure 5. In previous experiments, participants often palpated with an unrealistically high pressure. In this experiment, the stimuli were placed on a digital scale and the research assistant verbally warned the participants if their palpation pressure exceeded 1500 grams. Before the pre-test began, participants were given practice on a sample to learn this maximum pressure value. During the trial, the research assistant monitored this scale reading.

The second day involved a training session, a break, and a posttest. Each 20-minute training session covered search pattern, finger pressure, part and number of fingers used, finger motion, nodularity effects, breast area coverage, and lump properties. Group specific training was provided, dependent on the random group placement. Training sessions for groups b and c employed the three cylinder setup described above, Figure 1. For group a, relevant material was read to the trainee but the trainee did not palpate the breast model. Group b participants palpated the cylinders without receiving visual feedback, and group c trainees palpated the cylinders while viewing the visual display of the pressure changes. The research assistant provided training according to detailed, written instructions describing the above techniques. After a 10-minute break, post-training scores were gathered following protocol instructions from the pretest.

1	No Sig.	31	Sig. 5	61	No Sig.	91	Sig. 5
2	Sig. 1	32	No Sig.	62	Sig. 1	92	No Sig.
3	No Sig.	33	No Sig.	63	Sig. 3	93	No Sig.
4	No Sig.	34	No Sig.	64	No Sig.	94	No Sig.
5	Sig. 3	35	Sig. 3	65	No Sig.	95	Sig. 5
6	No Sig.	36	No Sig.	66	Sig. 1	96	Sig. 1
7	Sig. 5	37	Sig. 5	67	No Sig.	97	No Sig.
8	Sig. 1	38	No Sig.	68	Sig. 2	98	Sig. 6
9	No Sig.	39	Sig. 4	69	No Sig.	99	Sig. 4
10	Sig. 6	40	No Sig.	70	Sig. 6	100	No Sig.
11	Sig. 4	41	Sig. 3	71	No Sig.	101	No Sig.
12	No Sig.	42	Sig. 6	72	No Sig.	102	Sig. 2
13	Sig. 4	43	No Sig.	73	Sig. 3	103	Sig. 5
14	Sig. 2	44	No Sig.	74	Sig. 6	104	No Sig.
15	No Sig.	45	Sig. 2	75	No Sig.	105	No Sig.
16	No Sig.	46	Sig. 6	76	No Sig.	106	Sig. 1
17	Sig. 3	47	No Sig.	77	Sig. 3	107	Sig. 3
18	Sig. 1	48	No Sig.	78	Sig. 6	108	No Sig.
19	No Sig.	49	Sig. 5	79	No Sig.	109	No Sig.
20	No Sig.	50	Sig. 1	80	No Sig.	110	Sig. 2
21	Sig. 2	51	No Sig.	81	Sig. 4	111	No Sig.
22	No Sig.	52	Sig. 6	82	No Sig.	112	Sig. 2
23	Sig. 2	53	No Sig.	83	Sig. 4	113	Sig. 5
24	No Sig.	54	Sig. 2	84	Sig. 1	114	No Sig.
25	Sig. 4	55	Sig. 3	85	No Sig.	115	Sig. 4
26	No Sig.	56	No Sig.	86	No Sig.	116	No Sig.
27	Sig. 3	57	No Sig.	87	Sig. 4	117	Sig. 4
28	Sig. 6	58	Sig. 1	88	Sig. 2	118	Sig. 6
29	No Sig.	59	Sig. 5	89	Sig. 5	119	No Sig.
30	No Sig.	60	No Sig.	90	No Sig.	120	No Sig.

Figure 4: Trial order for both pretest and posttest for all participants

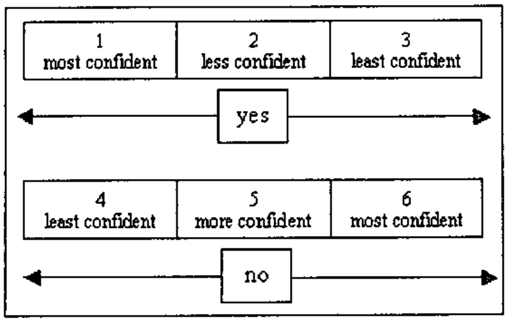


Figure 5: Confidence rating scale used in the experiment

2.3 Statistical Methods

Signal detection theory with a single interval model was used in the experiment and analsysis was done with the Rscore software [15]. With a single interval model, one stimulus is presented in each trial and the observer judges whtere the stimulus contains the signal and gives a confidence rating. An equal number of signal and blank (no signal) conditions are presnted. Thus, in our case, of the 120 trials, 60 were blank and 10 signal conditions for each of the six masses were used. The k-alternative forced choice model was also considered, but the single interval model was preferred because the simulus is judged in isolation rather than allowing for a comparison. Lack of comparison is somewhat more similar to a real breast exam where comparisons are generally unavailable.

From the signal detection theory analysis, d' pretest and d' posttest values were generated. Using these values, analysis of variance (ANOVA) techniques were used to compare the data and determine differences among control, no feedback, and feedback training groups. The split-split-plot design includes dependent variables (within-subjects dimension) of 2 whole plots (tests: pretest and posttest) with 6 blocks (masses: 1,2,3,4,5,6), Table 1. Each subject completes both pretest and posttest with all six masses. The independent variable (between-subjects dimension) includes the 3 subplots (training types: no training, training without feedback, training with visual feedback) for the 18 sub-subplots (subjects: 1-6 for no training, 7-12 for training without feedback, 13-18 training with visual feedback). The statistical model for the split-split-plot design is:

$$
\begin{aligned}
y_{ijkh} = {} & \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \gamma_k + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} \\
& + (\tau\beta\gamma)_{ijk} + \delta_h + (\tau\delta)_{ih} + (\beta\delta)_{jh} + (\tau\beta\delta)_{ijh} \\
& + (\gamma\delta)_{kh} + (\tau\gamma\delta)_{ikh} + (\beta\gamma\delta)_{jkh} + (\tau\beta\gamma\delta)_{ijkh} + \varepsilon_{ijkh}
\end{aligned}
$$

$$
\text{where} \left\{ \begin{aligned}
&\tau_i = \text{tests: factor A (whole plot blocks)} \\
&\beta_j = \text{training: factor B (main treatments)} \\
&\gamma_k = \text{masses: factor C (subplot treatments)} \\
&\delta_h = \text{participants per training group:} \\
&\qquad \text{factor D (sub-subplot treatments)}
\end{aligned} \right\} \left\{ \begin{aligned}
&i = 1, 2 \\
&j = 1, 2, 3 \\
&k = 1, 2, 3, 4, 5, 6 \\
&h = 1, 2, 3, 4, 5, 6
\end{aligned} \right\}
$$

Because participants were randomly distributed into the training groups (and thus the major unit of participants came from the same population of medical students), the experiments was crossed with respect to treatments rather than nested within a treatment. Between the three experimental training groups, the experiment was analyzed to find differences in lumps detected based on type of training.

Table 1: Split-Split Plot Design Layout

			Test											
			Pretest (Mass)						Posttest (Mass)					
			1	2	3	4	5	6	1	2	3	4	5	6
Training	No Training	Subj. 1												
		Subj. 2												
		Subj. 3												
		Subj. 4												
		Subj. 5												
		Subj. 6												
	Training without feedback	Subj. 7												
		Subj. 8												
		Subj. 9												
		Subj.10												
		Subj.11												
		Subj.12												
	Training with visual feedback	Subj.13												
		Subj.14												
		Subj.15												
		Subj.16												
		Subj.17												
		Subj.18												

3 Results

Using the Rscore program, each participant's pretest and posttest was analyzed. After review of the analysis, several χ^2 values exceeded a very liberal 34.805 with 18 degrees of freedom. All subjects who exceeded this value were removed from the analysis. We then proceeded to find the d' pretest and d' posttest values, suggesting that their response pattern did not match the pattern predicted by the signal detection theory. Further analysis revealed that many of these subjects did not make use of the confidence range provided, but instead repeatedly used a few of the response categories.

Using the d' values for the 9 qualifying participants (3 per training group) and using the general linear model in

SAS for the split-split plot design, it was found that the p-value for differences in training group were not significant (p=0.402), Figure 6. Gender, participants within each training group, and test order were not significant. The test plot (pretest and posttest) is shown, Figure 7. There was a significant difference (p=0.0611) between the mass samples. This can be seen in Figure 8. While masses 2-6 show nearly equal d' values, it should be noted that mass 1 was much more difficult for the participants to distinguish clearly.

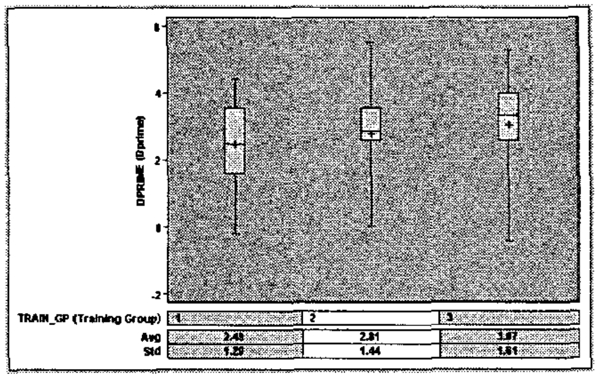


Figure 6: Box plot showing training groups (1=control, 2=no feedback, and 3=visual feedback training)

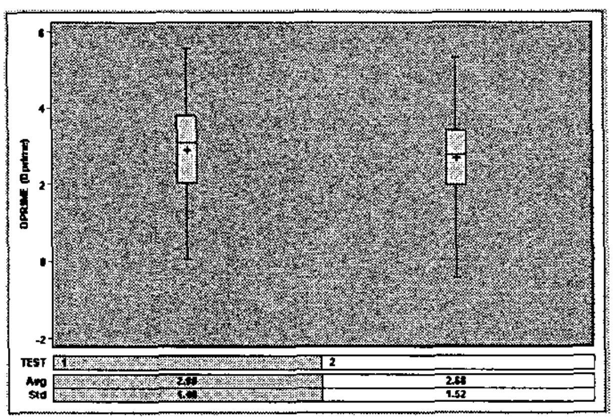


Figure 7: Box plot showing test order differences (1=pretest, 2=posttest)

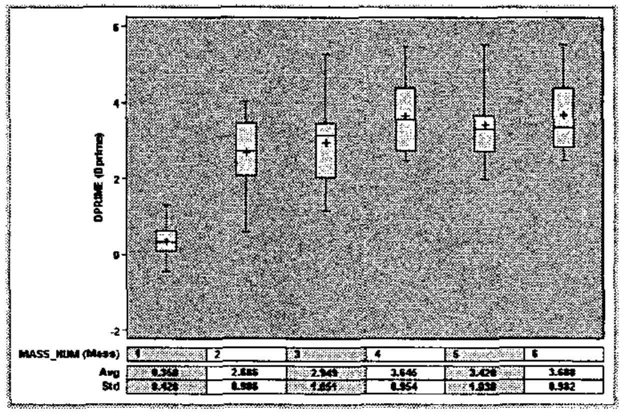


Figure 8: Box plot showing 6 sample masses

4 Discussion

While there were no significant results present in the study, a trend did exist for slightly greater improvement in the visual feedback group. If there is in fact a trend in this data, it is not significant. Because we dismissed half of the sample (from 18 down to 9 participants), due to statistical analysis issues discussed above, we would need to increase the sample size in a future study to confirm any such trend. The high χ^2 value, which indicates that the signal detection theory model did not account for all of the variance in these participants, is troubling. We believe this is in part due to the low number of trials (120) which should be increased to between 300-1000.

In the future, following from Figure 7, a better selection of samples in terms of ball bearing size and location is needed. While masses 2-6 were very easy for many of the participants to locate confidently, mass 1 was extremely difficult for almost all participants to find. Better selection distribution would include a gradient of easier to more difficult embedded masses in a more representative fashion. Also, we would like to determine those places where lumps are on the limits of detectability.

Additionally and suprisingly, few participants made comments about the 1500 gram pressure limit. The research assistant observed that participants quickly learned the 1500 gram pressure limit and would work their way to it without surpassing it, even on the following day. They seemed to learn the pressures without viewing the scale.

Future work includes refining the dynamic training model, measuring its educational efficacy, and exploring training transfer to clinical experience. Recently we received a grant from the National Board of Medical Examiners and we are currently working with the Ontario Breast Screening Program (OBSP) in Ontario, Canada. This project will refine the existing prototype dynamic silicon breast examination simulator, and test its effectiveness in assessing clinical breast examination skill with a group of clinical breast examination specialists in the OBSP. The project will evaluate two hypotheses, 1) that performance with the simulator correlates with performance on clinical breast exams in practice and 2) retesting with the simulator accurately measures performance improvement over time. Comparing performance of the simulator with clinical performance as measured through records documenting the practitioner's recent clinical sensitivity and specificity will test the ability of the device to measure clinical skill.

5 Conclusions

Several types of direct, immediate augmented performance feedback could potentially aid tactile training through the visual, auditory, or tactile senses. In the study, we were interested in ways to provide feedback to aid the learning process, in particular through the addition to visual feedback of pressure changes in palpated masses during training. While the study provided insights, we had no conclusive evidence of the efficacy of such feedback.

References

[1] S. M. Moss, "16-year mortality from breast cancer in the UK Trial of Early Detection of Breast Cancer," *Lancet*, vol. 353, pp. 1909-14, 1999.

[2] J. E. Korn, "The clinical breast examination: still important," *The Medical Journal of Allina*, vol. 7, pp. 1-4, 1998.

[3] S. E. Bennett, R. S. Lawrence, D. F. Angiolillo, S. D. Bennett, S. Budman, G. M. Schneider, A. R. Assaf, and M. Feldstein, "Effectiveness of methods used to teach breast self-examination," *American Journal of Preventive Medicine*, vol. 6, pp. 208-17, 1990.

[4] H. S. Campbell, S. W. Fletcher, C. A. Pilgrim, T. M. Morgan, and S. Lin, "Improving physicians' and nurses' clinical breast examination: a randomized controlled trial," *American Journal of Preventive Medicine*, vol. 7, pp. 1-8, 1991.

[5] D. C. Hall, C. K. Adams, G. H. Stein, H. S. Stephenson, M. K. Goldstein, and H. S. Pennypacker, "Improved detection of human breast lesions following experimental training," *Cancer*, vol. 46, pp. 408-14, 1980.

[6] V. A. Clarke and S. A. Savage, "Breast self-examination training: a brief review," *Cancer Nursing*, vol. 22, pp. 320-6, 1999.

[7] K. C. Lee, D. Dunlop, and N. C. Dolan, "Do clinical breast examination skills improve during medical school?," *Academic Medicine*, vol. 73, pp. 1013-9, 1998.

[8] G. J. Gerling, G. W. Thomas, and A. M. Weissman, "Dynamic Simulator Technical Description," University of Iowa, Iowa City Technical Report R03-01, 2003.

[9] G. J. Gerling, G. W. Thomas, A. M. Weissman, and E. L. Dove, "Dynamic Simulator for Training Clinical Breast Examination," presented at Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting, Baltimore, MD, 2002.

[10] P. Emanuel and E. Leff, *Introduction to feedback control systems*. New York: McGraw-Hill, 1979.

[11] M. Akamatsu, I. S. MacKenzie, and T. Hasbrouc, "A comparision of tactile, auditory, and visual feedback in a pointing task using a mouse-type device.," *Ergonomics*, pp. 816-827, 1995.

[12] M. Akamatsu and I. S. MacKenzie, "Movement characteristics using a mouse with tactile and force feedback," *International Journal of Human-Computer Studies*, vol. 45, pp. 483-493, 1996.

[13] H. S. Vitense, J. A. Jacko, and V. K. Emery, "Multimodal feedback: an assessment of performance and mental workload," *Ergonomics*, vol. 46, pp. 68-87, 2003.

[14] J. Rosen, B. Hannaford, M. P. MacFarlane, and M. N. Sinanan, "Force Controlled and Teleoperated Endoscopic Grasper for Minimally Invasive Surgery - Experimental Performance Evaluation," *IEEE Transactions on Biomedical Engineering*, vol. 46, pp. 1212-1221, 1999.

[15] L. O. Harvey, "RscorePlus," 5.3.4 ed. Boulder, CO, 2002.