AD_____

Award Number: W81XWH-08-1-0057

TITLE: The Development of Prostate Palpation Skills through Simulation Training May Impact Early Detection of Prostate Abnormalities and Early Management

PRINCIPAL INVESTIGATOR: Gregory Gerling

CONTRACTING ORGANIZATION: University of Virginia
                                             Charlottesville, VA 22904

REPORT DATE: May 2011

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
                         Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release; distribution unlimited

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)* 01-05-2011	2. REPORT TYPE Final	3. DATES COVERED *(From - To)* 1 MAY 2008 - 30 APR 2011

4. TITLE AND SUBTITLE		5a. CONTRACT NUMBER
The Development of Prostate Palpation Skills through Simulation Training May Impact Early Detection of Prostate Abnormalities and Early Management		
		5b. GRANT NUMBER W81XWH-08-1-0057
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Gregory Gerling		5d. PROJECT NUMBER
		5e. TASK NUMBER
E-Mail: gregory-gerling@virginia.edu		5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Virginia Charlottesville, VA 22904	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This is our second annual report. Our team has made good progress on our three year grant toward achieving aims. In the last year, we had 4 journal and 2 peer-reviewed conference papers accepted, and have others near completion. We have collected data from several human-subjects experiments and also tissue measurement experiments. We have subsequently analyzed that data via new procedures and algorithms developed, have worked on the development of a curriculum framework in which to embed the simulator in practice, and continued work in formulating an algorithm to allow computerized adaptive testing principles to be applied to reduce simulation exam duration. We will continue to work toward aim completion over the final year of the grant.

15. SUBJECT TERMS
No subject terms provided.

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U	UU	76	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Introduction

Our team is composed of the PI, Gregory J. Gerling, PhD, School of Engineering and two co-Is Reba Moyer Childress, MSN, FNP, School of Nursing and Marcus L. Martin, MD, School of Medicine. We have worked with graduate students (Angela Lee, William Carson, Leigh Baumgart, Ninghuan Wang, and Elmer Kim). We have also brought in two undergraduate students (Petheree Norman and Tashima Lambert). We have developed working collaborations with Tracey Krupski, MD, School of Medicine (Urology) and Randy Jones, School of Nursing – both of whom have particular expertise as relates to prostate cancer and disease. We have worked also in conjunction with O. John Semmes, PhD, Eastern Virginia Medical School (EVMS) and Beatriz Lopes, MD, University of Virginia, Autopsy Services.

The main topic of this work was to build and validate a physical-computerized simulator used to train physicans in performing the digitial rectal exam (DRE). This required an understanding of prostate tissue and also human perception limits.
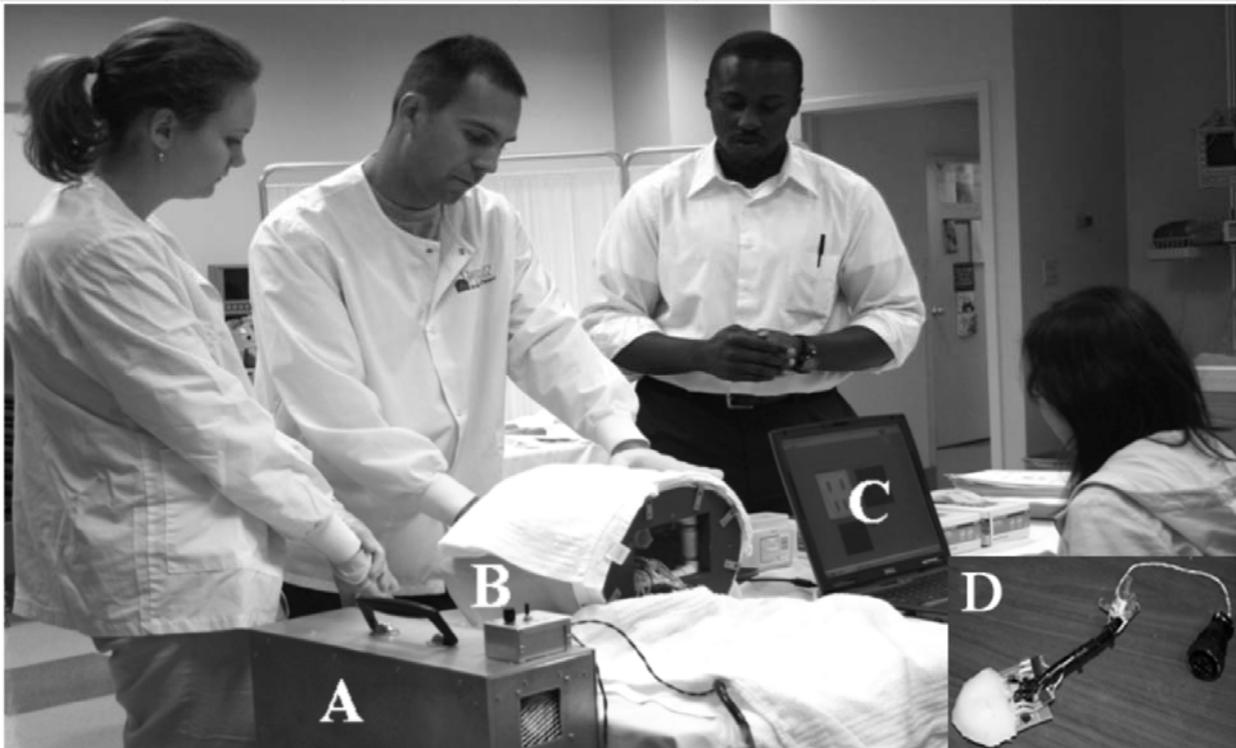
Figure 1. Main Components of Virginia Prostate Examination Simulator Apparatus: (A) electronics for automatic balloon inflation and sensor signal conditioning, (B) instrumented torso, (C) laptop and (D) instrumented prostate.

More specifically, the following series of aims and tasks were the subject of this project.

Aim 1. Determine distinct skill levels for discernment of palpable characteristics.
    Task 1.a) Characterize anatomical attributes and pathological stages of disease.
    Task 1.b) Determine the range of disease states that are palpable and simulate.

Task 1.c) Determine appropriate training scenarios to cover skill levels of various individuals.

Aim 2. Determine how contextual factors in the exam influence diagnosis decision-making.
    Task 2.a) Setup contextual scenarios.
    Task 2.b) Setup human-like aspects of standardized patient in simulated training environment.

Aim 3. Determine methods to customize performance assessment and training intervention.
    Task 3.a) Setup assessment based first on "up-down" or computerized adaptive testing (CAT) strategies.
    Task 3.b) Determine training interventions and levels of feedback.

Aim 4. Determine if applied finger techniques correlate with level of performance.
    Task 4.a) Correlate general aspects of technique with measures of assessment.
    Task 4.b) Correlate technique patterns of experts and novices with measures of performance assessment.
(Task 5) Plan for interaction with EVMS and U.Va. Biomaterials

Body

Aim 1 seeks to determine distinct skill levels for discernment of palpable characteristics.

Task 1.a is to characterize anatomical attributes and pathological stages of disease. Graduate student William Carson worked in this area. Following first year work to build the indenter and begun validating it with silicone-elastomer samples, we continued to collect data in the clinic with normal autopsied prostates at U.Va and cancerous prostates at EVMS. This data has since been analyzed, formed into a publication and accepted by the journal Medical Engineering and Physics. I attach below this abstract for this journal article listed in *Key Research Accomplishments* below.

Overall, the work in Task 1.a has sought to characterize the material properties of prostate tissue, removed post-surgery, and indented with a custom-built spherical indenter. The mechanical characterization of prostate tissue has not received much attention and is often disconnected from the clinic, where samples are readily attained. This work sought to inform the realistic design of artificial tissue – and also to relate material properties with disease states. We developed a spherical indenter to generate force-displacement data from *ex vivo* tissue, both whole mount and 5 mm cross-sections. Indentation velocity, depth, and sphere diameter, and four means of estimating elastic modulus (EM) were validated. EM was then estimated for 26 prostate specimens obtained from radical prostatectomy and 6 samples obtained from autopsy. Specimens were obtained in conjunction with Dr. Tracey Krupski (Urology, U.Va.) and Dr. O. John Semmes (Urology, Eastern Virginia Medical School). Prostatectomy prostates were also evaluated clinically upon digital rectal exam and pathologically post-extirpation. Overall, this work found that diseased prostate tissue is stiffer than normal tissue, stiffness increases with disease severity, and a large variability exists between samples, even though disease differences within a prostate are detectable. We compared these measurements with those of simulated prostate tissues and found that the two coincide, in terms of gross stiffness.

Background – The mechanical characterization of prostate tissue has not received much attention and is often disconnected from the clinic, where samples are readily attained. *Methods* – We developed a spherical indenter for the clinic to generate force-displacement data from *ex vivo* prostate tissue. Indentation velocity, depth, and sphere diameter, and four means of estimating elastic modulus (EM) were validated. EM was then estimated for 26 prostate specimens obtained via prostatectomy and 6 samples obtained from autopsy. Prostatectomy prostates were evaluated clinically upon digital rectal exam and pathologically post-extirpation. *Findings* – Whole-mount measurements yielded median EM of 43.2 kPa (SD = 59.8 kPa). Once sliced into cross-sections, median EM for stage T2 and T3 glands were 30.9 and 71.0 kPa, respectively, but not significantly different. Furthermore, we compared within-organ EM difference for prostates with (median = 46.5 kPa, SD = 22.2 kPa) and without (median = 31.0 kPa, SD = 63.1 kPa) palpable abnormalities. *Interpretation* – This work finds that diseased prostate tissue is stiffer than normal tissue, stiffness increases with disease severity, and large variability exists between samples, even though disease differences within a prostate are detectable. A further study of late-stage cancers would help to strengthen the findings presented in this work.

Figure 2. Portable indentation system and user interface built to make tissue measurements.

Figures 3 and 4 (below) represent the prostate specimen processing and the example collection of force versus displacement data.
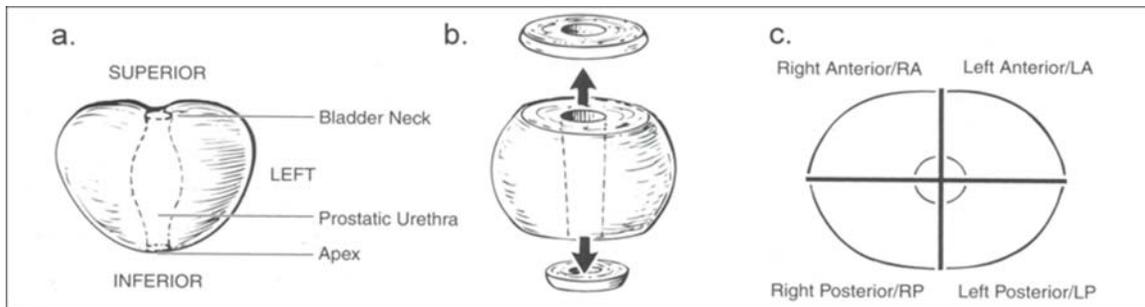
a) WM prostate on specimen plate; b) CS procedure; c) a single CS sample with its four quadrants labeled.

The Oliver-Pharr method for estimating EM relates Oliver-Pharr stiffness, S, and contact area, A, to reduced EM, E_r:

$$E_r = \frac{S\sqrt{\pi}}{2\sqrt{A}} \qquad E_r = \left[\frac{1-v_s^2}{E_s} + \frac{1-v_i^2}{E_i}\right]^{-1}$$

where S is the change in load with respect to indenter displacement at the moment of initial unloading. E_r is related to the EM of the sample, E_s, and indenter, E_i, and the Poisson ratios of the sample, v_s, and indenter, v_i.
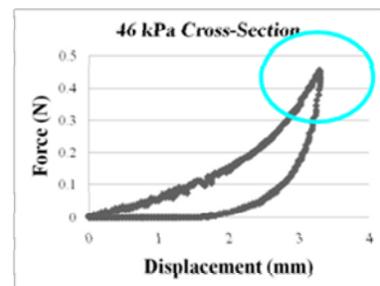
Table 1. EM for cross-section (CS) samples across pathological diagnosis

	EM (kPa)			
	Benign	*AC ≥ 10%*	*BPH*	*Prostatitis*
	n = 71	*n = 39*	*n = 35*	*n = 20*
Median	27.2	43.0	26.2	31.9
Mean	41.1	135.0	36.8	49.2
SD	41.3	240.6	37.0	52.8

Tables 2 and 3. EM for WM and CS samples across pathologic T stage

	EM (kPa) [WM]				EM (kPa) [CS]	
	T2	*T3*			*T2*	*T3*
	n = 83	*n = 6*			*n = 27*	*n = 10*
Median	42.1	59.9		**Median**	30.9	71.0
Mean	59.9	61.8		**Mean**	91.4	241.5
SD	48.8	31.4		**SD**	201.7	328.9

We have also written a second paper "Authenticating a high fidelity prostate exam simulator" and submitted it for review. It is presently under review. We are collected human-subjects data with urologists at UVa for this work. The main idea was to justify our hypothesis that tissue elasticity is indicative of carcinomous changes by correlating DRE findings with tissue elasticity and histopathology. Second, we sought to employ urologic surgeons to evaluate our prostate simulator in three ways: 1) authenticate that the elasticity of the simulated prostates accurately represents the range of normal prostate stiffness, 2) determine the range of nodule size reasonably palpable by DRE, and 3) discern what degree of elasticity difference within the same prostate suggests malignancy. These three overall results, respectively, are presented in the figures that follow.

Title: Authenticating a High Fidelity Prostate Exam Simulator

Objective: Despite continued controversy on the utility of digital rectal examination (DRE) in prostate cancer screening, all health care practitioners should be facile in its performance as it provides information on prostate size and nodularity, rectal pathology, and neurologic tone. By measuring the elastic properties of ex-vivo prostate specimens, we authenticated a novel prostate simulator for training of the DRE.   The purpose of this study was first to justify our hypothesis that tissue elasticity is indicative of carcinomatous changes by correlating DRE findings with tissue elasticity and

histopathology. Second, we employed urologic surgeons to evaluate our prostate simulator in three ways: 1) authenticate that the elasticity of the simulated prostates accurately represents the range of normal prostate stiffness, 2) determine the range of nodule size reasonably palpable by DRE, and 3) discern what degree of elasticity difference within the same prostate suggests malignancy.

Materials and Methods: Materials characterization, human-subjects experiments, histopathology, and chart abstraction of clinical history were used. To perform mechanical characterization of prostate tissue, 26 *ex vivo* prostatectomy specimens were evaluated using a custom-built and portable spherical indentation device that builds force-displacement curves to calculate tissue elastic modulus. To measure human-subject perception of prostatic state, the Virginia Prostate Examination Simulator (VPES) was employed and combines rubber-like synthetic materials with expandable balloons embedded within to simulate prostates and is controlled by a computer. Twelve urologic surgeons, naïve to the simulator, performed a methods of limits psychophysical test and completed a 30 minute qualitative assessment. In addition, histopathologic assessment was performed on both the entire prostate and sections from which the elastic modulus was calculated. Finally, chart abstraction identified the clinical parameters of interest. All portions of the study were IRB approved.

Results: From the materials characterization, the measurements of the 26 gross prostates and 40 cross-sections yielded 306 data points of which 84 comprised normal tissue and 41 consisted of adenocarcinoma. Within the same prostate, cancer was always stiffer. Of the 7 cases with an abnormal DRE, the DRE accurately identified adenocarcinoma in 85%. From the human-subjects experiments, the simulated prostates evaluated by urologists ranged in stiffness from 8.9-91 kPa, mimicking the range found on *ex vivo* analysis of 4.6-236.7 kPa. The urologic surgeons determined the upper limit of stiffness palpated as realistic for a healthy prostate was 59.63 kPa while the lower limit of stiffness was 27.1 kPa. Nodule size less than 7.5 mm was felt to be too small to reasonably palpate. In their experiments with palpable differences within a prostate, a background prostate of 34.1 kPa with less elastic nodules (85.1kPa) embedded within was felt to be the most realistic.

Conclusions: Based on materials characterization and histopathology of prostates and urologic expert opinion, our study confirms that a wide range of prostate tissue elasticity is deemed "normal." We found it is not the absolute elasticity of the nodule, but rather the relationship of the nodule to the background prostate elasticity that constitutes the critical tactile feedback. Urologic surgeons determined that within-prostate differential of approximately 50 kPa was most representative of what might be found clinically. Prostate simulator training may lead to greater familiarity with both the range of "normal" prostate elasticity, pertinent diagnostic cues, and diagnosis of prostate cancer.

Modified version of the psychophysical method of limits

Upper and Lower Average Stiffness Bounds
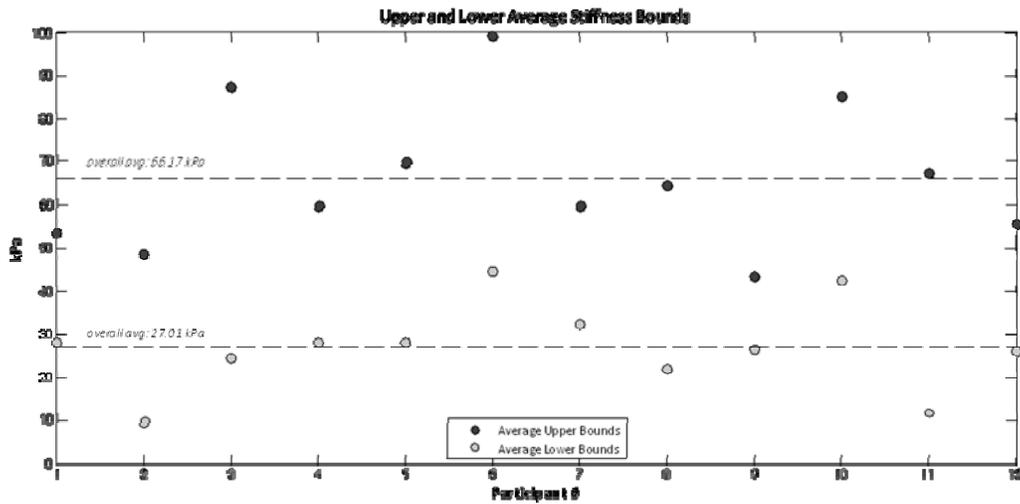
Figure 5. Upper and Lower Average Prostate Stiffness Bounds from human-subjects experiment with Urologists.

Scenario	Nodule Size	# of Detections	Majority Size Agreement	Majority Diagnosis
Prostate 2 Layout	15 mm	12	Within reasonable limits	Carcinoma
Prostate 2 Layout	10 mm	11	Within reasonable limits	Carcinoma
Prostate 2 Layout	7.5 mm	10	Within reasonable limits	Carcinoma
Prostate 2 Layout	5 mm	4	Too small	Normal
Prostate 2 Layout	4 mm	5	Too small	Normal
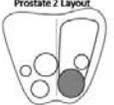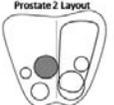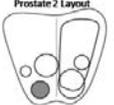
Figure 6. Nodule Size Limits of Detection (N=12 on each scenario)

kPa	Scenario	Diagnosis				
		Normal	Carcinoma	BPH w/o Cancer	BPH w/ Cancer	No Data
60.9	Prostate 1 Layout	4	3	2	0	3
34.1	Prostate 2 Layout	1	10	0	0	1
19.2	Prostate 3 Layout	1	9	0	0	2
60.9	Prostate 1 Layout	0	2	7	1	2
34.1	Prostate 2 Layout	0	2	7	0	3
19.2	Prostate 3 Layout	1	2	5	2	2
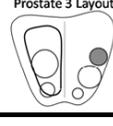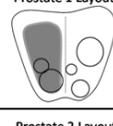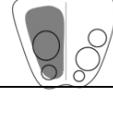
Figure 7. Prostate-Nodule Stiffness Differential for Cancer and BPH Scenarios

Task 1.b is to determine the range of disease states that are palpable and simulate those. Graduate student Leigh Baumgart worked in this area. Listed in *Key Research Accomplishments* below, this work has since been accepted by the Journal of Cancer Epidemiology. Its abstract is attached below.

Overall, this work has sought to understand the perceptible limits of the DRE, which are based on some unresolved combination of the size, depth, and hardness of abnormalities within a given prostate stiffness. This work seeks to inform the range of disease states that are palpable, from human sensory limits. Using a custom-built device similar to the VPES, an 18 participant human-subjects study was conducted that simulated the four aforementioned conditions. Within silicone-elastomers that mimic normal prostate tissue, only abnormalities of diameter greater than 4 mm (20 mm^3 in volume) were consistently detectable (above 75% of the time) at the shallowest depth (5 mm). In contrast, abnormalities located in simulated tissue of greater stiffness (82 kPa compared to 21 kPa) must be twice that volume. Overall, the study found that size and depth of abnormalities most influence detectability, while the relative hardness between abnormalities and tissue affects detectability for some size-depth combinations.

Background: Although the digital rectal exam (DRE) is a common method of screening for prostate cancer and other abnormalities, the limits of ability to perform this hands-on

exam are unknown. Perceptible limits are a function of the size, depth, and hardness of abnormalities within a given prostate stiffness. *Methods:* To better understand the perceptible limits of the DRE, we conducted a psychophysical study with 18 participants using a custom-built apparatus to simulate prostate tissue and abnormalities of varying size, depth, and hardness. Utilizing a modified version of the psychophysical method of constant stimuli, we uncovered thresholds of absolute detection and variance in ability between examiners. *Results:* Within silicone-elastomers that mimic normal prostate tissue (21 kPa), abnormalities of 4 mm diameter (20 mm^3 volume) and greater were consistently detectable (above 75% of the time) but only at a depth of 5 mm. Abnormalities located in simulated tissue of greater stiffness (82 kPa) had to be twice that volume (5 mm diameter, 40 mm^3 volume) to be detectable at the same rate. *Conclusions:* This study finds that the size and depth of abnormalities most influence detectability, while the relative stiffness between abnormalities and substrate also affects detectability for some size/depth combinations. While limits identified here are obtained for idealized substrates, this work is useful for informing the development of training and allowing clinicians to set expectations on performance.

Figures 8 and 9 show the high-level results of the psychophysical experiment. As can be observed, the abnormalities positioned in shallower depths are more readily detectable, as are abnormalities of larger size. A comparison of the figures also indicates that substrate stiffness also plays a role.
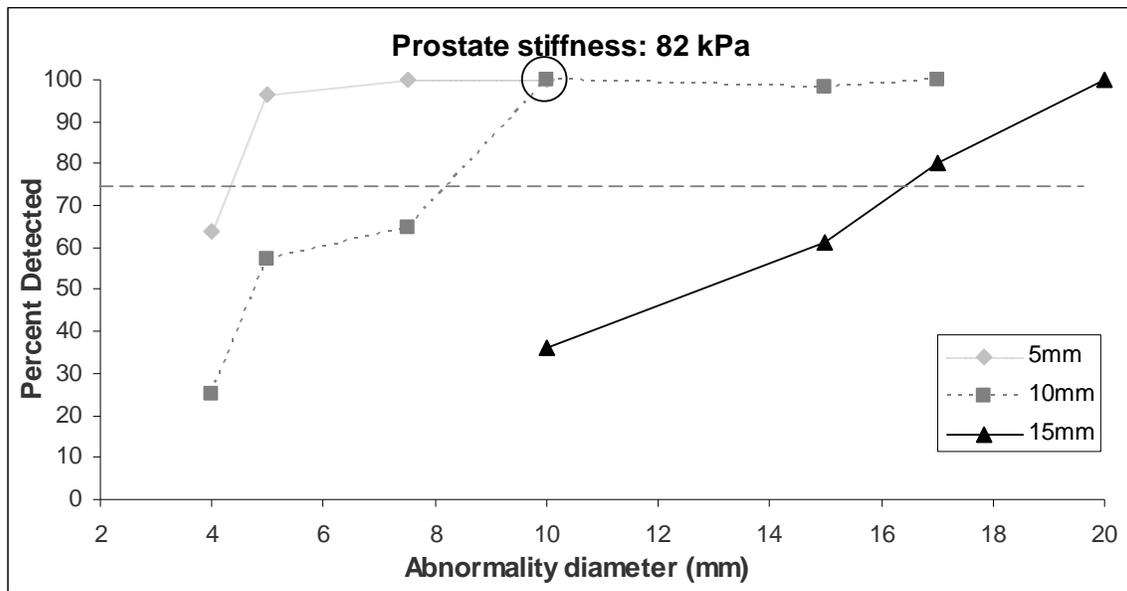
Figure 8. Psychophysical functions for the detectability of abnormalities of various diameters and depths for prostate stiffness of 82 kPa. Dotted line denotes 75% correct threshold.
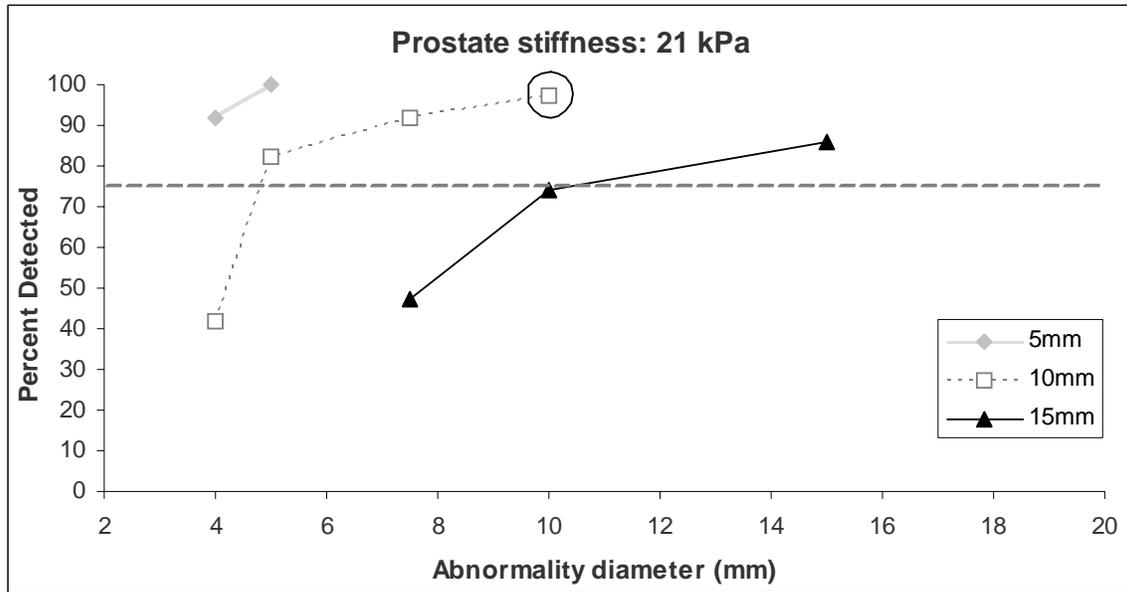
Figure 9. Psychophysical functions for the detectability of abnormalities of various diameters and depths for prostate stiffness of 21 kPa (more pliant).

Further work using logistic regression analyzed the relative importance of each factor and their interactions. This work formed the basis for a second publication, which was accepted for and presented at the Proceedings of the 2010 IEEE Haptic Interfaces for Virtual Environment and Teleoperator Systems.

> Softness discrimination and the detection of inclusions are important in surgery and other medical tasks. To better understand how the characteristics of an inclusion (size, depth, hardness) and substrate (stiffness) affect their tactile detection and discrimination with the bare finger, we conducted a psychophysics experiment with eighteen participants. The results indicate that within a more pliant substrate (21 kPa), inclusions of 4 mm diameter (20 mm³ volume) and greater were consistently detectable (above 75% of the time) but only at a depth of 5 mm. Inclusions embedded in stiffer substrates (82 kPa) had to be twice that volume (5 mm diameter, 40 mm³ volume) to be detectable at the same rate. To analyze which tactile cues most impact stimulus detectability, we utilized logistic regression and generalized estimating equations. The results indicate that substrate stiffness most contributes to inclusion detectability, while the size, depth, and hardness of the stimulus follow in individual importance, respectively. The results seek to aid in the development of clinical tools and information displays and more accurate virtual haptic environments in discrimination of soft tissue.

Task 1.c is to determine appropriate training scenarios to cover skill levels of various individuals. Reba Childress and Greg Gerling brought in a nursing student (Petheree Norman) along with an engineering student (Angela Lee) to help us develop the scenarios. Overall, we selected 3 instrumented prostates for use, down from the 10 we normally use. These cover the range of normal, prostatitis, carcinoma and BPH that we need. They also include a variety of

different carcinomas. With each of the three layouts, we chose 4 combinations of large balloons (row 1) to represent BPH and inflammation of the prostate, combinations of smaller balloons to represent carcinomas (row 2) and combinations of large and small balloons to represent prostatitis and inflammation with carcinoma.
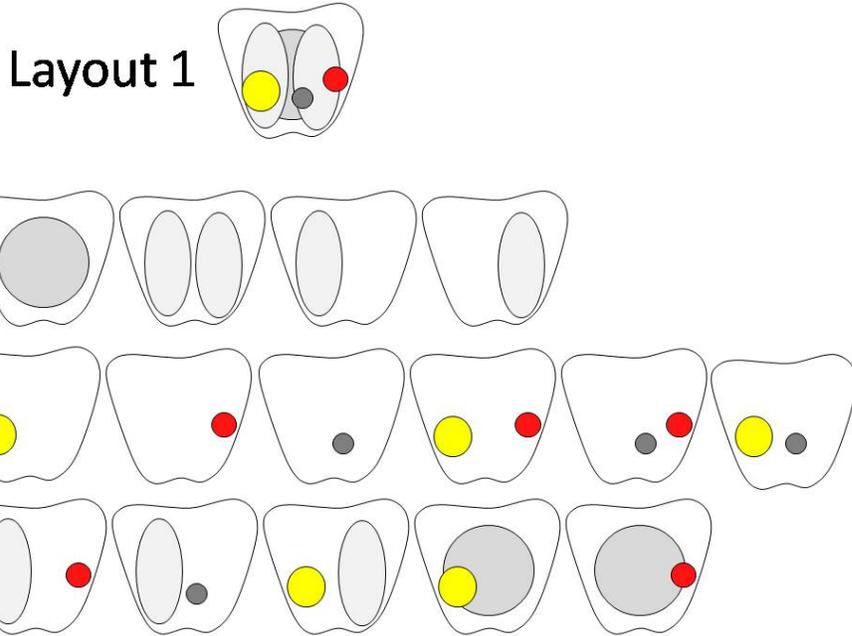
Figure 10. Layout of prostate 1
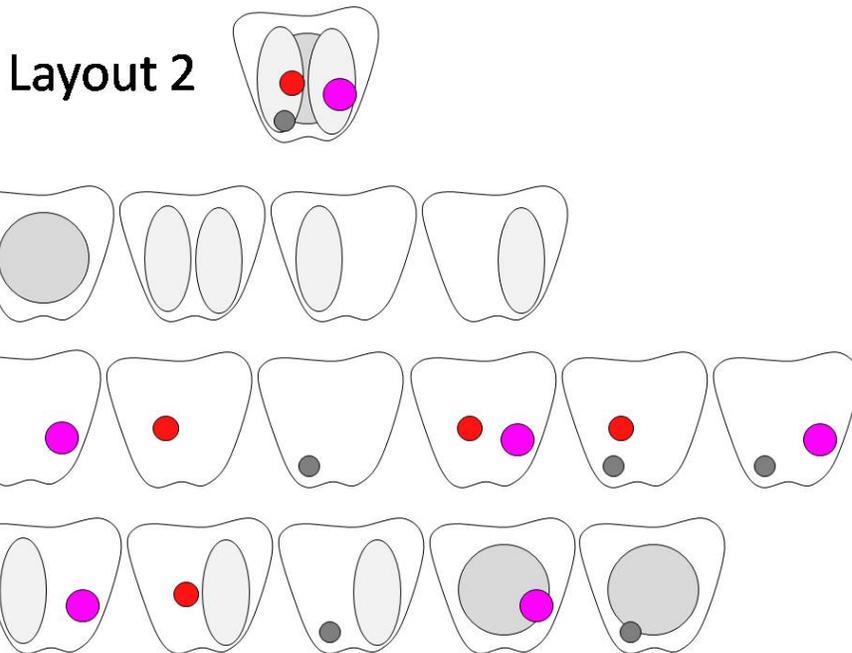
Figure 11. Layout of prostate 2
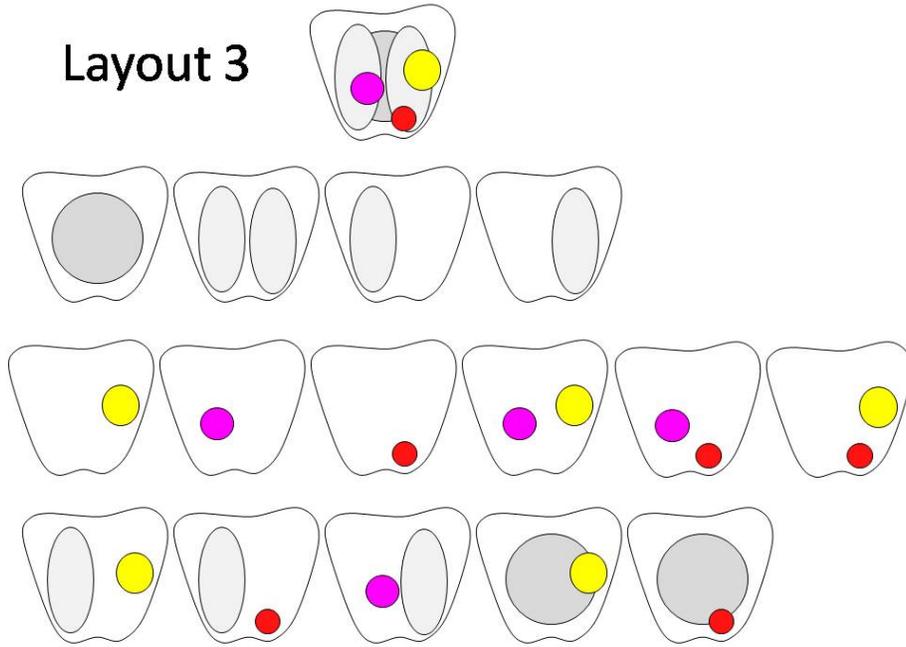
Figure 12. Layout of prostate 3

Aim 2 seeks to determine how contextual factors in the exam influence diagnosis decision-making.

Tasks 2.a and 2.b. is to setup contextual scenarios and user feedback. We used the detailed prostate tactile scenarios from Aim 1 to create the VPES Full Circle Interactive Learning Model to allow the integration of VPES simulator training into the Medical & Nursing Curricula. The figure below delineates the major training steps.
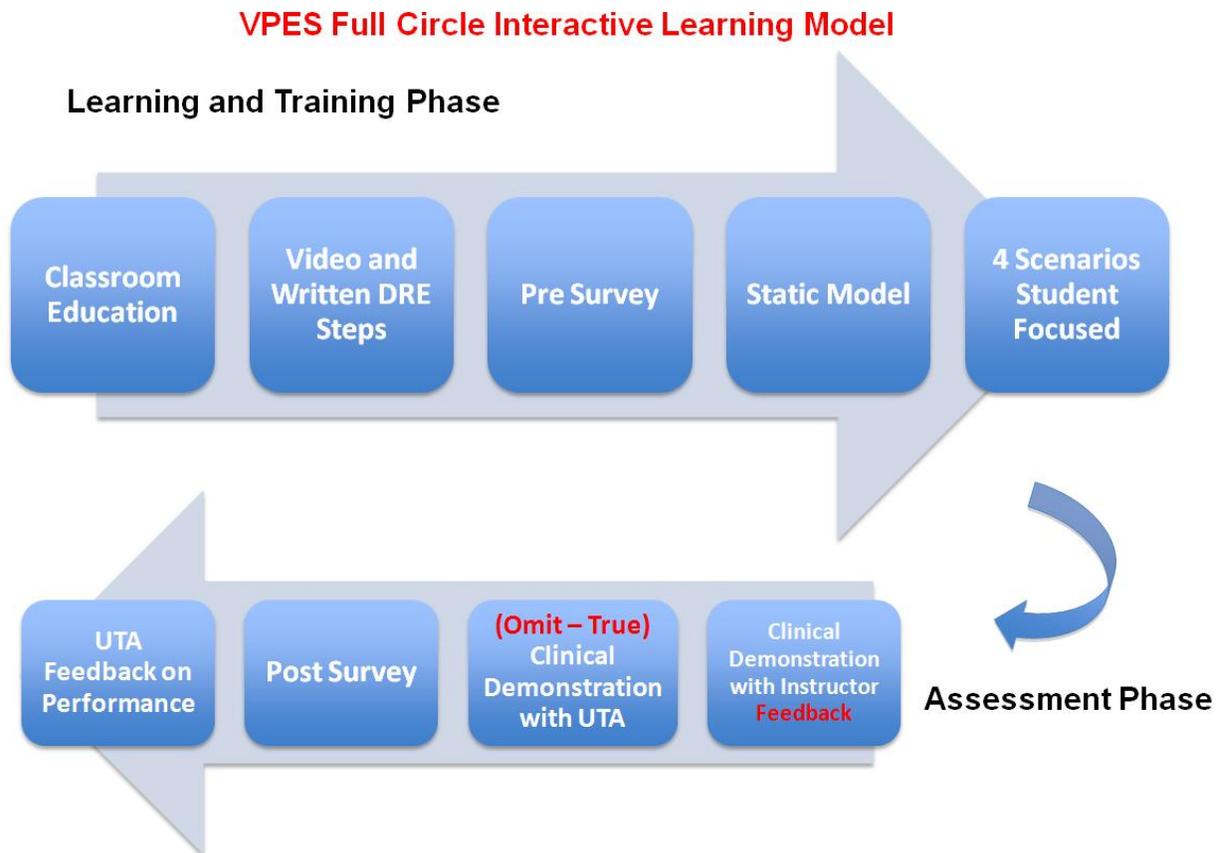
Figure 13. VPES Full Circle Interactive Learning Model

With this framework in places, we are developing scenarios for BPH, Prostatitis, Carcinoma and normal cases, following the example that can be observed in the 3 figures that follow.

16

Figure 14. Patient information including age, race, history, etc. given to the student before the exam.

Figure 15. This information delineates an example of what the student should be finding and should be discussing with the teaching instructor upon debriefing.

VPES Presentation – Prostate CA

Figure 16. This is the user interface that the student interacts with to indicate aspects of the prostate that are found to be irregular upon the DRE.

We have also redesigned the feedback that the student receives in the exam from the simulator itself. Figure 17 shows the opening screen.
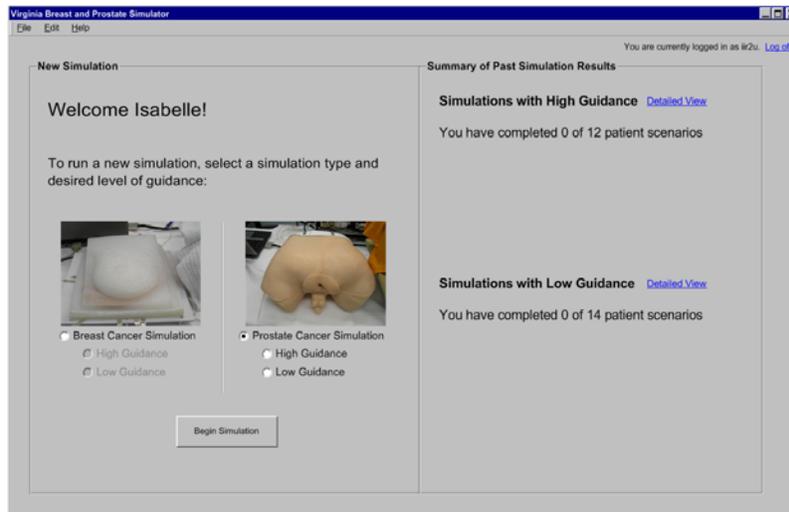
Figure 17: The screen when a user logs in for the first time.

In Figure 18, the user selects a patient scenario by clicking on the picture of the appropriate patient. Here, he or she may choose to perform the patient scenario or may view additional details about the patient. Another version of the interface abstracts the images of the prostates, if it is the case that the learner is selecting the scenarios, so not to provide the answers before the test begins.
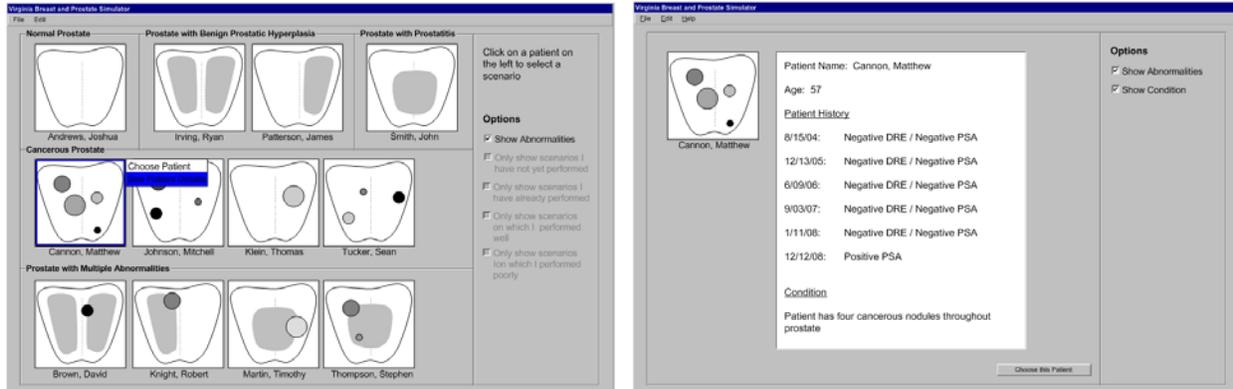
Figure 18. (left) Scenario selection with abnormalities shown and a specific patient selected and (right) Detailed view for patient Matthew Cannon

Once the simulated tumors are filled with water, it is time for the user to begin the exam. While palpating, the user will see the screen depicted in Figure 19. The user knows where he or she is palpating at a given time by looking at the red shading on the larger picture. The interface also displays a smaller picture that depicts all locations on the prostate that has already been palpated. The user knows how many abnormalities are being simulated by how many circles appear on the picture of the larger prostate and knows which abnormalities have been palpated because the circle becomes darker as the water pressure in the line spikes.
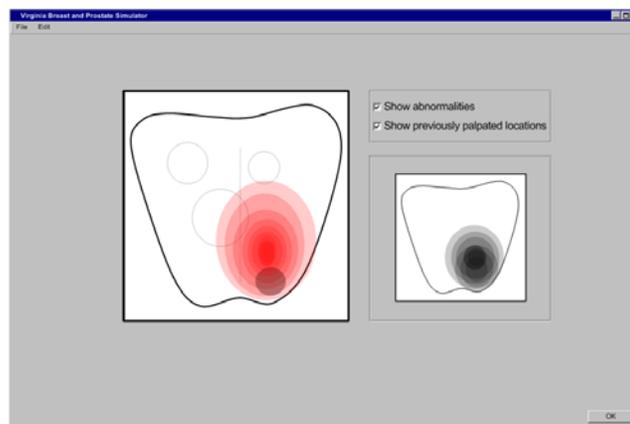
Figure 19: Screenshot of the feedback displayed to a user while he or she is palpating.

Three types of post-performance feedback are provided to a user after the simulation. The first, time elapsed, is depicted in Figure 20a, force exerted is depicted in Figure 20b, and percentage of the prostate palpated is shown in Figure 20c. Each of these prominently displays the

criterion for a successful run and a green check mark or a red X to indicate whether the criteria were met.
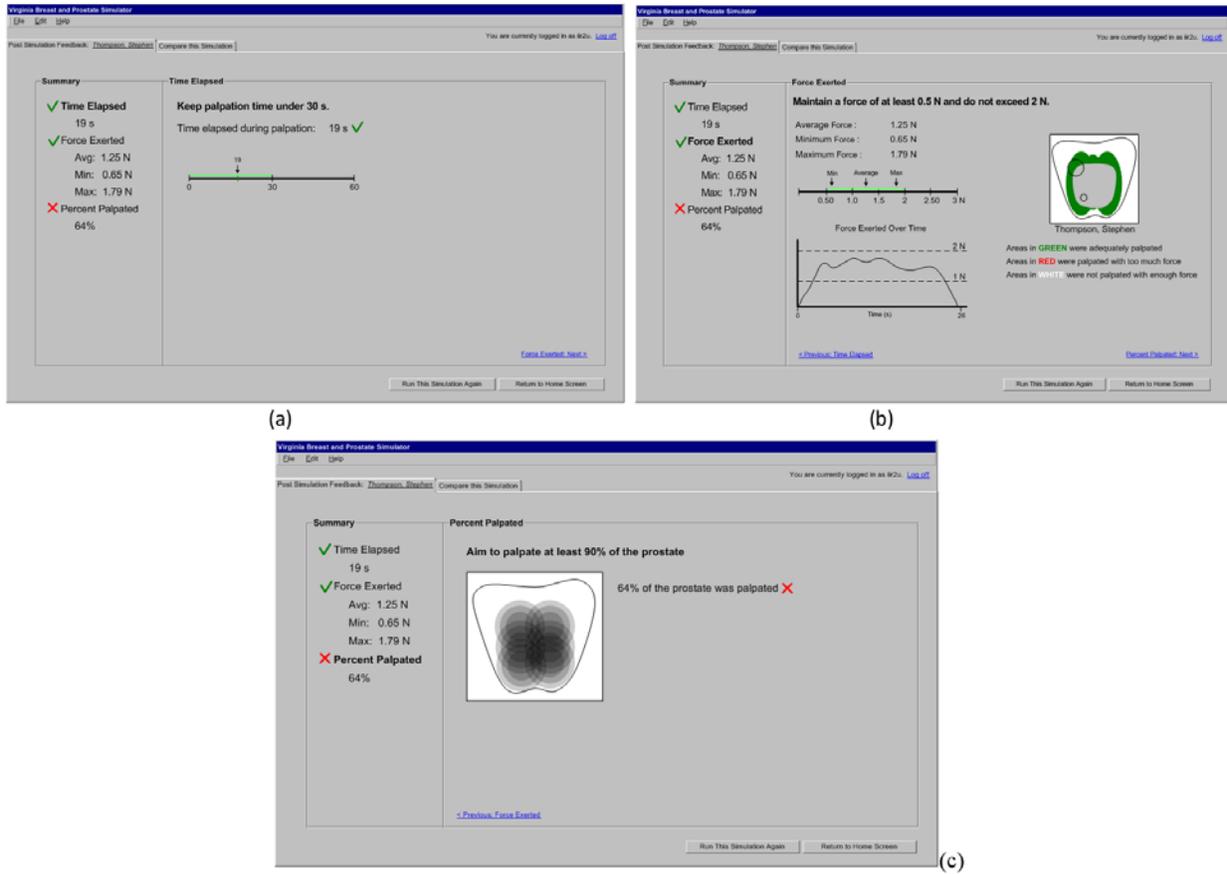
Figure 20: Screenshots of the post-performance feedback provided to the user

Aim 3. seeks to determine methods to customize performance assessment and training intervention.

Task 3.a is to setup assessment based first on "up-down" or computerized adaptive testing (CAT) strategies. Graduate student Angela Lee worked in this area.

This task sought to develop an efficient and accurate means of assessing the palpation skill of trainees. We seek to integrate computerized adaptive testing (CAT) with the VPES to provide proficiency estimates with fewer test items, thereby reducing testing duration. The main components in our CAT exam are to develop an item bank of prostate scenarios, implement the item response theory (IRT) and an item selection procedure, and determine the stopping criteria and scoring method. Using a three parameter logistic model, the developed computer algorithm will selectively choose subsequent prostate scenarios based on responses to previous scenarios. The three parameters that characterize each scenario are difficulty, item discrimination, and guessing parameters. To validate the CAT application, a set of two experiments will be conducted. The first hypothesis is that low performers will be differentiable from high performers. The second hypothesis is that the assessment made in experiment 2 will equal that of experiment 1 but be achieved within a reduced time period, of approximately 25 to 50%. The overall idea is presented in the figures below.

The four main components in the CAT as we have implemented are delineated in Figure 21.

Computerized Adaptive Testing (CAT) Model

1. Item Bank Development
Includes all possible test items of prostate scenarios
3 item parameters characterize each item:
   a: item discrimination
   b: item difficulty
   c: guessing

2. Item Response Theory
3 parameter logistic model, Θ: ability estimate

$$P(u_i = 1 | \Theta_j, a_i, b_i, c_i) = c_i + \frac{(1 - c_i)}{1 + e^{-1.7a_i(\Theta_j - b_i)}}$$

3. Item Selection Procedure
Subsequent items are tailored to the current ability estimate
Select item with maximum information

$$I(\Theta_j) = \frac{P'(\Theta_j)^2 d\Theta_j}{P(\Theta_j)(1 - P(\Theta_j))}$$

Item with maximum information

4. Stopping Criteria
Algorithm terminates when the ability level is estimated with sufficient precision

Figure 21: Four main steps in the CAT implementation (abstracted) as applied to VPES

To enable the CAT style of test administration, we had to build a new prostate torso apparatus that could hold 10 instrumented prostates, instead of the previous 3. Figure 22 shows the scenarios that each of the new 10 simulated prostates can offer.
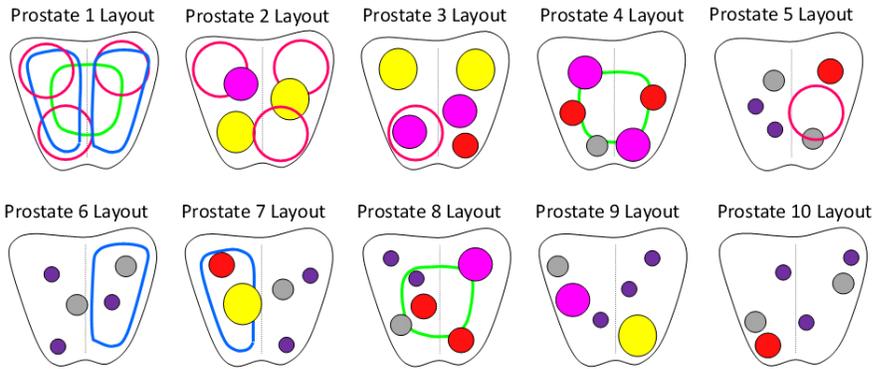
Figure 22: VPES Version 2.0 instrumented prostates and scenario generation, modified to accommodate the requirements of at least 200 scenarios for a CAT implementation.

The image sequence in Figure 23 shows an example iteration of CAT implemented with the VPES simulator. The participant is asked three questions. The first question is of medium difficulty and the participant answers correctly. Therefore, the next question is automatically selected to be of greater difficulty. This one is answered incorrectly which leads to an easy question. This process will help us to identify participant ability in fewer questions by reaching a stable ability state in few questions.
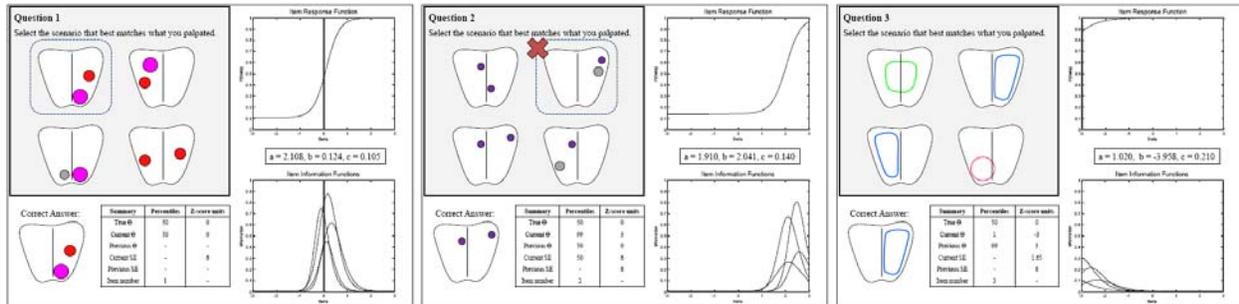
Figure 23: Example sequence of three questions administered to participants in a sequence

Listed below is the abstract from the thesis of Angela Lee. We are currently seeking to turn this work into a journal manuscript.

Title: Applying the Partial Credit Model to the Assessment of Clinical Hands-On Skills in a Part-Task Simulator

The training and assessment of palpation skills using patient simulators may improve hands-on skills and lead to earlier clinical diagnosis of cancer. The Virginia Prostate Examination Simulator (VPES) was built to improve skills in the digital rectal examination (DRE). The work herein seeks to address a new means of assessing performance in the simulated exam. First, a set of five performance factors relevant to

the DRE was developed (prostate enlargement, symmetry, median sulcus definition, nodule location, and nodule size), each discretized with three to five levels. Then, we applied a partial credit model (PCM) to i) determine an appropriate points allocation system for each prostate scenario where its five factors vary in difficulty, ii) identify the best scoring methodology for these factors, and iii) calibrate the factors of each prostate scenario to determine their difficulty measures. The PCM's use of polytomous scoring allows examinees to acquire points based on factor levels, therefore resulting in a more informative overall score, as compared to a simple correct or incorrect detection. A human subjects experiment (n = 13) was performed where each participant palpated 40 scenarios of graded difficulties and entered a total of 200 responses to a custom-built interface. The results of the experiment show that the best scoring method for the majority of the factors is dichotomous, where the original levels for each factor collapse to a binary score. The resulting scoring models and difficulty measures may serve as the foundation for a computerized adaptive test for the VPES. Such a CAT may provide a new means for clinical assessment of large numbers of students in short time periods in an automated way, potentially desirable for clinical board exams.

Task 3.b has been combined with Task 3.a above since the computerized adaptive test does adapt questions for learners of various levels.

Aim 4. seeks to determine if applied finger techniques correlate with level of performance.

Tasks 4.a and 4.b are to correlate general aspects of technique with measures of performance assessment and correlate technique patterns of experts and novices with measures of performance assessment. Graduate student Ninghuan "Miki" Wang worked in this area, along with undergraduate student Tashima Lambert. Two journal articles based on work done have been published and listed in *Key Research Accomplishments* below.

Overall, the study done with 16 resident physicians and 18 nurse practitioner students algorithmically defined a set of finger palpation techniques for the digital rectal exam (DRE) based upon past qualitative definitions of hands-on technique and evaluated performance between experts and novices. Four palpation techniques were defined: global finger movement, local finger movement, and average intentional finger pressure, and dominant intentional finger frequency. Streaming feedback from force and balloon sensors in the instrumented prostate provided the source data. With this information we sought to assess if certain techniques were prevalently used and correlated with greater performance accuracy. Although technique utilization varied, some elements clearly impacted performance. For example, those utilizing the local finger movement of vibration (i.e., firm pressure of varying intensity) were significantly better at detecting abnormalities. Also, the V pattern of global finger movement led to greater success and average finger pressure of greater magnitude was required to detect smaller, more deeply positioned abnormalities. We found that the quantified palpation techniques appear to account for examination ability at some level but not entirely for differences between experience levels.

Abstract for paper in IEEE Transactions on Information Technology in Biomedicine

This work seeks to quantify finger palpation techniques in the prostate clinical exam, determine their relationship with performance in detecting abnormalities, and differentiate the tendencies of nurse practitioner students and resident physicians. One issue with the digital rectal examination (DRE) is that performance in detecting abnormalities varies greatly and agreement between-examiners is low. The utilization of particular palpation techniques may be one way to improve clinician ability. Based on past qualitative instruction, this work algorithmically defines a set of palpation techniques for the DRE, i.e., global finger movement, local finger movement, and average intentional finger pressure, and utilizes a custom-built simulator to analyze finger movements in an experiment with two groups: 18 nurse practitioner students and 16 resident physicians. Although technique utilization varied, some elements clearly impacted performance. For example, those utilizing the local finger movement of vibration were significantly better at detecting abnormalities. Also, the V global finger movement led to greater success, but finger pressure played a lesser role. Interestingly, while the residents were clearly the superior performers, their techniques differed only subtly from the students. In summary, the quantified palpation techniques appear to account for examination ability at some level but not entirely for differences between groups.

As indicated in Figures 24 – 26, we have now setup algorithms to quantify finger palpation patterns and examined palpation patterns of medical resident physicians and nurse practitioner

students who identified that palpation via a particular method was associated with improved detection rates. In depth analysis of the palpation technique ascertained that global finger movement (GFP), local finger movement (LFP), and average intentional finger pressure (AIFP) were key components of this palpation technique (Fig. 24, Analysis Tool). In short, GFP is defined as the systematic movement of one's finger over the entire prostate (U, V, L, and Line patterns) while LFP is defined as palpation by finger movement within a single quadrant of the instrumented prostate or near a single abnormality. Three patterns are defined as tapping, vibration and sliding. Finally, we calculate AIFP as that applied over the duration of the exam in the vicinity of filled balloons.
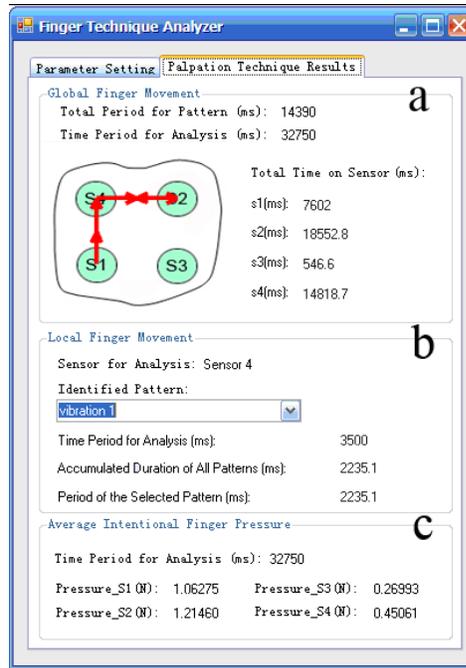
Figure 24: Palpation Technique Analyzer and Analysis Results, including a) Global Finger Pattern, b) Local Finger Pattern and c) Average Intentional Finger Pressure.

The continuous nature of the recording from force and balloon sensors (Fig. 25, left) allows for the quantification of these patterns.
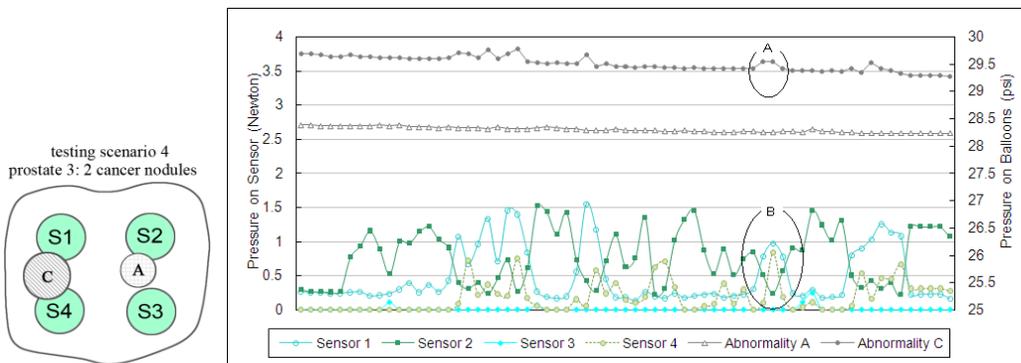
Figure 25: Example Plot of Force Sensor and Balloon Sensor Data for an Example Testing Scenario

With "local" finger pattern as an example, we show more formally the mathematical definition of the three local patterns (tapping, vibration and sliding), Fig. 26.
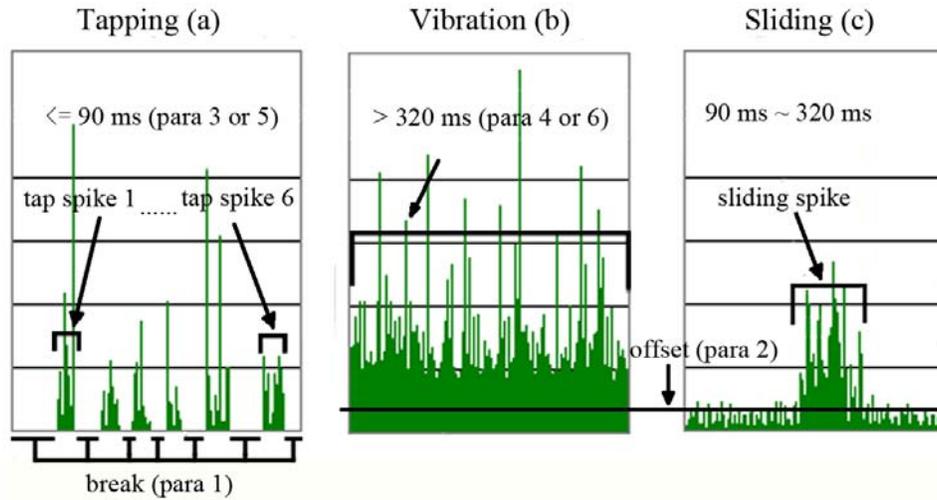
Figure 26: Three local finger palpation patterns of tapping, vibration and sliding.

The local tapping pattern is defined formally for the ith sensor L^i_{tap} in equation below, where the period between time j and k is a break with no spikes (para 1).

$$L^i_{tap} = ((\sum_{t=j}^{k} t_{S\_Spike^i_t} \le 90) \, \& \, ((t_{S\_Spike^i_j} - t_{S\_Spike^i_k}) < 48)), \; j>k$$

Visible in the data of Figure 26b, the vibration pattern is defined formally for the ith sensor in equation below as an examiner maintaining finger pressure above a certain value (para 2) on the prostate over a continuous time span of at least 320-msec.

$$L^i_{vibr} = ((\sum_{t=j}^{k} t_{S\_Spike^i_t} \ge 320) \, \& \, (S\_Spike^i_t > offset))$$

The sliding pattern comes in contrast to the tapping and vibration patterns where the examiner appears to transition from the global finger pattern to an intentional focus upon the local detection of a balloon. Visible in Figure 26c, the sliding pattern is defined formally for the ith sensor in equation 7 as an examiner maintaining pressure above a certain value (para 2) continuously for 90 to 320-msec.

$$L^i_{sl} = ((90 < \sum_{t=j}^{k} t_{S\_Spike^i_t} < 320) \, \& \, (S\_Spike^i_t > offset))$$

26

Abstract for paper in Journal of Simulation in Healthcare

Introduction: Prostate carcinoma (and other prostate irregularities and abnormalities) are detected in part via the digital rectal exam. Training clinicians to use particular palpation techniques may be one way to improve rates of detection. Methods: In an experiment of 34 participants with clinical backgrounds, we used a custom-built simulator to determine if certain finger palpation techniques improved one's ability to detect abnormalities smaller in size and dispersed as multiples over a volume. The intent was to test abnormality cases of clinical relevance near the limits of size perceptibility (i.e., 5 mm diameter). The simulator can present abnormalities in various configurations and record finger movement. To characterize finger movement, four palpation techniques were quantitatively defined (global finger movement, local finger movement, average intentional finger pressure, and dominant intentional finger frequency) to represent the qualitative definitions of other researchers. Results: Participants who used more thorough patterns of global finger movement (V and L) ensured the entire prostate was searched and detected more abnormalities. A higher magnitude of finger pressure was associated with the detection of smaller abnormalities. The local finger movement of firm pressure with varying intensity was most indicative of success and was required to identify the smallest (5 mm diameter) abnormality. When participants utilized firm pressure with varying intensity, their dominant intentional finger frequency was about 6 Hz. Conclusions: The use of certain palpation techniques does enable the detection of smaller and more numerous abnormalities, and we seek to abstract these techniques into a systematic protocol for use in the clinic.

We have since begun a study with collaborator Tracey Krupski to evaluate the finger patterns of expert urologist examiners. Testing with 12 urological clinicians has finished, seeking to determine if there is a standard palpation technique. If such a palpation technique does exist, the potential exists for a high yield training intervention. The analysis of this work will continue into the future.

Task 5 is to setup the interaction with EVMS and U.Va. Biomaterials. Task 5 was completed in year 1. We have IRB agreements in place at the University of Virginia and Eastern Virginia Medical School. These have also been approved by the IRB of the Department of Defense.

Key Research Accomplishments

We list several journal and conference publications, either already presented or published, currently under review, or to be submitted.

Peer-reviewed publications accepted
1. Baumgart, L.A., Gerling, G.J., and Bass, E.J. Characterizing the range of simulated prostate abnormalities palpable by digital rectal examination, Cancer Epidemiology, 34 (1): 79-84 2010

2. Wang, N., Gerling, G.J., Moyer Childress, R., and Martin M.L. Quantifying palpation techniques in relation to performance in a clinical prostate exam, IEEE Transactions on Information Technology in Biomedicine, 14(4): 1088-97 2010

3. Wang, N., Gerling, G.J., Moyer Childress, R., and Martin M.L. Using a prostate exam simulator to decipher palpation techniques that facilitate abnormality detection near clinical limits, Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare, 5(3):152-160 2010

4. Carson, W.C., Gerling, G.J., Krupski, T.L., Gundersen, C.A., Harper, J.C., and Moskaluk, C.A., Material characterization of *ex vivo* prostate tissue via spherical indentation in the clinic, Medical Engineering & Physics, 33(3): 302-309 2011

Peer-reviewed publications in progress
1. Gerling, G.J., Moyer Childress, R. and Martin, M.L., Designing a clinically functional simulator to support effective learning: features and cases. Planned submission to Medical Teacher.

2. Kowalik, C.G., Lee, A.J., Carson, W.C., Gerling, G.J., Harper, J.C., Moskaluk, C.A., and Krupski, T.L., Authenticating a High Fidelity Prostate Exam Simulator. Under review.

3. Lee, A.J., Schmidt, K., and Gerling, G.J. Applying the Partial Credit Model to the Assessment of Clinical Hands-On Skills in a Part-Task Simulator, in preparation for the Journal of the Human Factors and Ergonomics Society.

Conference papers and presentations (peer-reviewed)
1. Wang, N., Gerling, G.J., Moyer Childress, R., and Martin M.L. Characterizing finger palpation in the detection of prostate cancers and abnormalities. Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting, 2008, New York City, NY, pp. 813-817

2. Baumgart, L.A., Gerling, G.J., and Bass, E.J., Psychophysical detection of inclusions with the bare finger amidst softness differentials, Proceedings of the 2010 IEEE Haptic Interfaces for Virtual Environment and Teleoperator Systems, Boston, MA, pp 17-20 2010 (acceptance rate: 42%)

Student conferences (not rigorously peer-reviewed)

1. Baumgart, L.A., Gerling, G.J. & Bass, E.J., Characterizing the range of simulated prostate abnormalities palpable by digital rectal examination. Presented at the NLM Informatics Training Conference (July 23-24, 2009) Portland, Oregon

2. Lee, A.J., Gerling, G.J., Applying computerized adaptive testing to the Virginia prostate examination simulator, Works in Progress Abstract and Poster Presentation at the 10th Annual International Meeting on Simulation in Healthcare (Jan. 23-27, 2010) Phoenix, Arizona.

3. Baumgart, L.A., Gerling, G.J. and Bass, E.J., Characterizing the range of simulated prostate abnormalities palpable by digital rectal examination, Abstract and Poster Presentation for Academy of Distinguished Educators – 6th Medical Education Research Day and Poster Session (February 22-26, 2010) U.Va. School of Medicine

4. Lee, A.J., Gerling, G.J., Applying computerized adaptive testing to the Virginia prostate examination simulator, Abstract and Poster Presentation for Academy of Distinguished Educators – 6th Medical Education Research Day and Poster Session (Feb. 22-26, 2010) U.Va. School of Medicine

5. Gundersen, C.A., Gerling, G.J., Carson, W.C., Thomas, K.R., Harper, J., Moskaluk, C.A., Krupski, T.L., Assessing mechanical properties of benign and malignant prostate tissue. American Society of Clinic Oncology 2010, Chicago, Illinois (Published Abstract, Permanent Abstract ID: e15109)

6. Poster presentation at 2011 Academy of Distinguished Educators (ADE) medical education poster session March 7th-11th, 2011, entitled "Comparing Accuracy of Digital Rectal Exam Finger Patterns: Novice to Expert", George C. Bailey[1], Casey G. Kowalik[2], Angela J. Lee[3], N. Wang[3], Tashima Lambert[3], Gregory J. Gerling[3], and Tracey L. Krupski[4] *1. University of Virginia, School of Medicine, 2. Lahey Clinic Medical Center, Department of Urology, 3. University of Virginia, Department of Systems and Information Engineering, 4. University of Virginia, Department of Urology*

Students graduated
- Ninghuan Wang (Master of Science, May 2009)

- Angela Lee (Master of Science, May 2011)

- William Carson (Master of Science, May 2011)

Other grants written
Title: Connecting Digital Rectal Exam (DRE) Training with Improving Healthcare Provider Performance and Patient Safety using a Novel Prostate Simulator
Sponsor: AHRQ - NIH
Status: Declined, preparing resubmission

Congressionally Directed Research Program's Conference IMPACT on Prostate Cancer. Orlando, FL, 3/9-11, 2011. Presented two posters.

- Ninghuan Wang, Gregory J. Gerling, Tracey L. Krupski, Reba Moyer Childress, Marcus

L. Martin, "Quantifying palpation techniques in relation to performance in a clinical prostate exam, especially that facilitate the detection of abnormalities near clinical limits" to IMPaCT: Innovative Minds in Prostate Cancer Today

- William C. Carson, Gregory J. Gerling, Tracey L. Krupski, Casey A. Gundersen, Jeffrey C. Harper, and Christopher A. Moskaluk, "The mechanical characterization of ex vivoprostate tissue via spherical indentation in the clinic" to IMPaCT:Innovative Minds in Prostate Cancer Today

In terms of Public Service and Outreach, we have also presented to the public in several venues both in the popular press and with booths at the one-to-one level.

Coverage in Popular Press
- NBC 29 Interview: "Simulator Helps UVA Doctors Detect Cancer" (January 26th, 2009) Newscast available for viewing at: http://www.nbc29.com/global/story.asp?s=9730404

- Cavalier Daily Newspaper Article: "New Simulator Provides Unique Practice" (January 27th, 2009) Story available for reading at: http://www.cavalierdaily.com/news/2009/jan/27/new-simulator-provides-unique-practice/

Presented at a booth at the Charlottesville Community Health Fair, in conjunction with the 19th annual African-American Cultural Arts Festival in Booker T. Washington Park, Saturday, July 26, 2008, Attended and demonstrated the Virginia Prostate Examination Simulator, our research and posters, along with group informing the public about prostate and breast cancer.

Reportable Outcomes

Several papers have now been published. Those are listed above.

Conclusions

Our team made good progress on our three year grant toward achieving all aims. We have 7 journal papers either submitted or accepted, in addition to 2 conference proceedings, with other physical artifacts completed. We also had other presentations that were not peer reviewed. We have collected data from several human-subjects experiments and also tissue measurement experiments. We have recruited a group of students and have established collaborations with other researchers, in particular to gain access to tissue specimens. We have successfully built and validated a materials characterization procedure, a series of algorithms for detecting finger palpation patterns, formalized contextual feedback, have worked on the development of a curriculum framework in which to embed the simulator in practice, and formulated an algorithm to allow computerized adaptive testing principles to be applied to reduce simulation exam duration.

References

None applicable.

Appendices

All published journal and conference papers are attached as the following appendices.

Appendix A

Baumgart, L.A., Gerling, G.J., and Bass, E.J. Characterizing the range of simulated prostate abnormalities palpable by digital rectal examination, Cancer Epidemiology, 34 (1): 79-84 2010

Appendix B

Wang, N., Gerling, G.J., Moyer Childress, R., and Martin M.L. Quantifying palpation techniques in relation to performance in a clinical prostate exam, IEEE Transactions on Information Technology in Biomedicine, 14(4): 1088-97 2010

Appendix C

Wang, N., Gerling, G.J., Moyer Childress, R., and Martin M.L. Using a prostate exam simulator to decipher palpation techniques that facilitate abnormality detection near clinical limits, Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare, 5(3):152-160 2010

Appendix D

Carson, W.C., Gerling, G.J., Krupski, T.L., Gundersen, C.A., Harper, J.C., and Moskaluk, C.A., Material characterization of *ex vivo* prostate tissue via spherical indentation in the clinic, Medical Engineering & Physics, 33(3): 302-309 2011

Appendix E

Wang, N., Gerling, G.J., Moyer Childress, R., and Martin M.L. Characterizing finger palpation in the detection of prostate cancers and abnormalities. Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting, 2008, New York City, NY, pp. 813-817

Appendix F

Baumgart, L.A., Gerling, G.J., and Bass, E.J., Psychophysical detection of inclusions with the bare finger amidst softness differentials, Proceedings of the 2010 IEEE Haptic Interfaces for Virtual Environment and Teleoperator Systems, Boston, MA, pp 17-20 2010

SUPPORTING DATA

None applicable.

Characterizing the range of simulated prostate abnormalities palpable by digital rectal examination

Leigh A. Baumgart, Gregory J. Gerling *, Ellen J. Bass

Department of Systems and Information Engineering, University of Virginia, 151 Engineer's Way PO Box 400747, Charlottesville, VA 22904 USA

ARTICLE INFO

ABSTRACT

Background: Although the digital rectal exam (DRE) is a common method of screening for prostate cancer and other abnormalities, the limits of ability to perform this hands-on exam are unknown. Perceptible limits are a function of the size, depth, and hardness of abnormalities within a given prostate stiffness.
Methods: To better understand the perceptible limits of the DRE, we conducted a psychophysical study with 18 participants using a custom-built apparatus to simulate prostate tissue and abnormalities of varying size, depth, and hardness. Utilizing a modified version of the psychophysical method of constant stimuli, we uncovered thresholds of absolute detection and variance in ability between examiners.
Results: Within silicone-elastomers that mimic normal prostate tissue (21 kPa), abnormalities of 4 mm diameter (20 mm^3 volume) and greater were consistently detectable (above 75% of the time) but only at a depth of 5 mm. Abnormalities located in simulated tissue of greater stiffness (82 kPa) had to be twice that volume (5 mm diameter, 40 mm^3 volume) to be detectable at the same rate. *Conclusions:* This study finds that the size and depth of abnormalities most influence detectability, while the relative stiffness between abnormalities and substrate also affects detectability for some size/depth combinations. While limits identified here are obtained for idealized substrates, this work is useful for informing the development of training and allowing clinicians to set expectations on performance.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Prostate cancer has a high incidence rate (one in six for men in the U.S.), with an estimated 192,280 new cases in 2009. When diagnosed in an early and less aggressive stage, the five-year survival rate approaches 100% [1]. To promote early detection, the American Cancer Society advises that screening via the digital rectal examination (DRE) and prostate specific antigen (PSA) blood test be conducted concurrently. The DRE is important because the PSA tends to both over diagnose (65–75% of findings reported as false positives for PSA greater than 4.0 ng/l [2]) and miss cancerous tumors (15.2% of findings reported as false negatives for PSA less than 4.0 ng/l [3]). Although the DRE plays an integral role in early detection and is a skill clinicians are expected to learn, the perceptible limits surrounding this exam are unknown. Therefore, there is no basis from which to set reasonable expectations about clinical performance or to develop appropriate training.

When conducting a DRE, the clinician's task is to detect hard nodules that vary in size, depth and hardness or prostate enlargement that varies in volume change and stiffness. The former typically relate to carcinoma, the latter signal benign prostatic hyperplasia (BPH) or prostatitis [4]. The size, depth, and hardness of nodules and relative stiffness of a given prostate contribute to the perceptible range of abnormalities. At present, neither the thresholds of absolute detection nor variance in ability between examiners have been identified.

In contrast, palpable limits have been studied in terms of the clinical (CBE) and self (BSE) breast exams [5]. In two studies with rubber-like materials, abnormality size emerged as the major dimension affecting the detection of lumps [6,7]. In general, larger lumps in more shallow positions pose the least difficulty. However, simulated lumps as small as 3.0 mm diameter were detectable when embedded in breast-like materials (which is an order of magnitude more pliant than prostate tissue) [9–10]. Aside from the lump size findings, abnormality depth and hardness appear to have a minimal impact, whereas the stiffness of surrounding tissue may decrease one's ability to detect deeper lumps [11].

Hall et al. have shown that training on silicone models effectively increases exam performance on natural breast tissue [12]. Most of their training, and that prescribed by others for use with silicone models [13–17], takes place at the level of hands-on

skills. There is a focus on tactile skills because knowledge of disease and attitudes about domain are not strongly related to proficiency [18] and basic rules for diagnosing abnormalities as certain diseases are not difficult to master [19]. One major prerequisite for improving hands-on skills, however, is knowing the limits of tactile sensation, in this case with respect to the DRE.

When characterizing the limits on performance, the DRE differs from the breast exam in several key ways. First, a rectal wall is positioned between the finger and prostate, in addition to a glove and lubricant. Second, the clinician uses a single finger and is more constrained in search movement. Third, the prostate is a stiffer gland with less volume. Fourth, the exam typically takes place in less than 30 s in contrast to 2 min for the breast exam [7,20]. With these differences in mind, one common factor is that clinician performance in both exams does benefit from training.

In this work, the overall goal is to determine the perceptible limits of simulated abnormalities of various size, depth, and hardness within substrates of different stiffness when the examiner is constrained as with a DRE. In addition to determining the thresholds of absolute detection, we seek to determine the degree of variance in ability between examiners.

2. Methods

To analyze the limits of tactile perception in the DRE, we conducted a human-subjects experiment with 18 participants, using simulated prostates where abnormality size, depth, and hardness were varied within substrates of two stiffness levels. The objectives were to determine (1) the size of abnormalities detectable above 75% of the time at three discrete depths, (2) how substrate stiffness impacts the detectability of size/depth combinations, (3) if changes in abnormality hardness (consistent in objectives 1–2) impact detectability over size/depth combinations, (4) if some abnormalities require a minimum hardness be consistently detected, and (5) the variance in ability between participants.

2.1. Apparatus

An apparatus was built specifically for this study. The apparatus utilized silicone-elastomers to simulate the feel of prostate tissue and a rectal wall and employed a computer and electronics to control polyethelene balloons that simulated abnormalities. The computer also monitored the water pressure in the balloons and force on sensors embedded in the simulated tissue. The apparatus design is similar to that described in Ref. [19].

Twenty-three simulated prostates, 30 mm diameter and 20 mm tall, were mounted to a round platform that could be rotated so that the prostate under test was located beneath the examiner's finger. The idealized cyndrilical prostates did not include the surface undulations or an overall walnut shape, although the size was roughly the same as an actual prostate [19]. The platform containing the simulated prostates was housed within a structure that restricted access to and view of the simulated prostates. The examiner inserted his or her finger through an opening in the structure that was built of silicone-elastomer to mimic the rectal wall. The opening was angled at approximately 110° from the participant.

Each simulated prostate included a single polyethelene balloon embedded at one of three depths: 5, 10, and 15 mm. Balloons of seven volumes were used: 20, 40, 80, 200, 470, 1060, and 1770 mm^3 that correspond to diameters of 4.0, 5.0, 7.5, 10.0, 15.0, 17.0, and 20.0 mm, respectively. Preliminary studies found that balloons of 3.0 mm diameter were inconsistently detected in stiffer simulated prostates and were not included. The balloons were filled with water, thereby controlling hardness. Balloons could be

inflated to be hard, like a rock, but were not detectable when deflated. In this study, three hardnesses were used: 23, 27, and 31 durometers, type Shore A. These fall within the range used for simulated breast tumors [6,7]. Water pressure sensors (Honeywell, SenSym Pressure Sensor, Model SX100DD4) monitored the water pressure over time, which was logged by the computer.

In addition to factors of depth, size, and abnormality hardness, simulated prostates of two stiffness levels were used: 21 and 82 kPa. These have been evaluated via compression tests and fall into the measured range of prostate stiffness (mean elastic modulus = 44.20 kPa, SD = 25.89 kPa [8]). These stiffness values also fall in line with those deemed "realistic" in a subjective study with resident physicians and nurse practioner students [19]. Located in the backing of each prostate were four, laterally spaced pressure sensors (Flexiforce 0–1 lb, Tekscan, South Boston, MA) which logged the examiner's finger pressure over the simulated prostate.

2.2. Participants

Ten male and eight female participants (mean age = 20.4 years, SD = 1.38) were enrolled in the human-subjects study, approved by the IRB at the University of Virginia. No participant had prior clinical experience. A questionnaire also indicated that no participant had any remarkable prior experience working with his or her hands.

2.3. Experimental design

Using a modified version of the psychophysical method of constant stimuli [21], participants palpated the simulated prostates to determine the presence or absence of abnormalities. Typically the method of constant stimuli employs stimulus and blank trials presented in a randomized fashion where all stimulus combinations are presented an equal number of times. However, in the version we employed, we made three modifications to reduce participant fatigue. First, from all possible combinations (abnormality size, depth, hardness and substrate stiffness) only a subset of stimulus combinations were presented to participants (e.g., size 4 mm was used at 5 and 10 mm depth but not 15 mm depth). Pilot testing was conducted to remove combinations that were detectable 0% or 100% of the time. Second, the number of times that each stimulus combination was presented varied from two to four times depending on the difficulty of detecting the abnormality in the pilot study. Specifically, from the chosen subset of abnormalities, the most difficult to detect were presented four times, while the easiest to detect were presented two times. Third, due to hardware and time limitations, participants were presented with stimuli and blanks in one of six pre-determined random orders. Table 1 shows all stimulus combinations used and the number of times each was presented per participant in the experiment.

2.4. Procedure

Every participant participated in two experimental sessions, held on separate days for 90 min each. In session 1, each participant completed a 5 min pre-test questionnaire, a 5 min hands-on practice, and an 80 min hands-on experiment. During session 2, each participant completed a 5 min hands-on practice, an 80 min hands-on experiment and a 5 min post-test questionnaire. During sessions 1 and 2, participants palpated 96 simulated prostates, half of which contained an abnormality (the balloons were not inflated for the other half). Four participants returned for session 3, which was a 5 min hands-on practice and a 45 min hands-on experiment. Session 3

Table 1
Stimulus combinations and number of presentations per participant.

Abnormality dimensions			Number of presentations at 82 kPa stiffness	Number of presentations at 21 kPa stiffness
Size (diameter in mm)	Depth (mm)	Hardness (durometers, Shore A)		
4	5	31	4	2
4	10	31	2	4
5	5	23, 27, 31	3*, 3*, 3*	3*, 3*, 2*
5	10	23, 27, 31	3*, 3*, 3*	3*, 3*, 3*
7.5	5	31	3	–
7.5	10	31	3	2
7.5	15	31	–	4
10	5	31	2	–
10	10	23, 27, 31	4, 4, 4	4, 4, 2
10	15	23, 27, 31	–, –, 2	4, 4, 3
15	10	31	3	–
15	15	23, 27, 31	4, 4, 4	–, –, 2
17	10	31	2	–
17	15	31	3	–
20	15	31	2	–

* Stimulus combinations only presented in Session 3 and examined by four participants.

participants palpated 48 additional simulated prostates, 24 of which contained an abnormality.

Participants were given 20 s to examine a single prostate. At the conclusion of each exam, participants informed the proctor (via "yes" or "no") as to whether an abnormality was present. Verbal responses were recorded by the proctor on paper. Participants were given a 10 s break between subsequent examinations, a 60 s break after every 10–17 examinations and a 5 min break after every 32–42 examinations. During each exam, the proctor also monitored the pressure exerted on the prostates to ensure that finger pressure remained within 4 and 6 N and that the quadrants were palpated in the specified order. During the hands-on practice session, participants had been taught to utilize a consistent search technique. The technique was to move one's finger across the prostate surface in lines parallel with and then perpendicular to the participant's seated position. When traversing these line paths, participants used small, dime-sized circular motions.

3. Results

The results are summarized in five subsections corresponding with the five objectives.

3.1. Absolute threshold – abnormality size, depth, and substrate stiffness combinations

Scenarios across participants were combined to determine the size of abnormalities consistently detectable (above 75% of the time for all participants combined) at the three discrete depths for two levels of prostate stiffness. Deeper abnormalities were more difficult to detect for all abnormality sizes. For the stiffer substrate (82 kPa), the abnormality size threshold was 5 mm diameter at 5 mm depth, 10 mm diameter at 10 mm depth, and 17 mm diameter at 15 mm depth (Fig. 1). For the more pliant substrate (21 kPa), the abnormality size threshold was 4 mm diameter at 5 mm depth, 5 mm diameter at 10 mm depth, and 15 mm diameter at 15 mm depth (Fig. 2). Therefore, both increases to substrate stiffness and abnormality depth make similarly sized abnormalities harder to detect.

3.2. Effect of substrate stiffness on abnormality detection

The detection rate for larger abnormalities in stiffer substrates (Fig. 1) is similar to smaller abnormalities in more pliant substrates
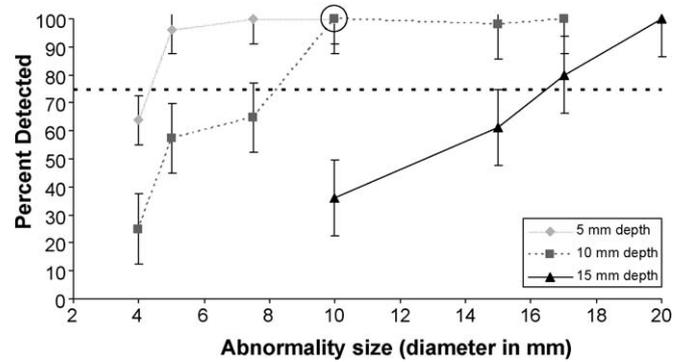
Fig. 1. Psychophysical functions for the detectability of abnormalities at 5, 10 and 15 mm depths as a function of abnormality size for prostate stiffness of 82 kPa. The dotted line denotes the 75% correct threshold. The circle denotes a size/depth combination expected to minimally surpass the 75% threshold when abnormality hardness is decreased, and is referenced in Fig. 3. Standard error bars are shown around the data points.
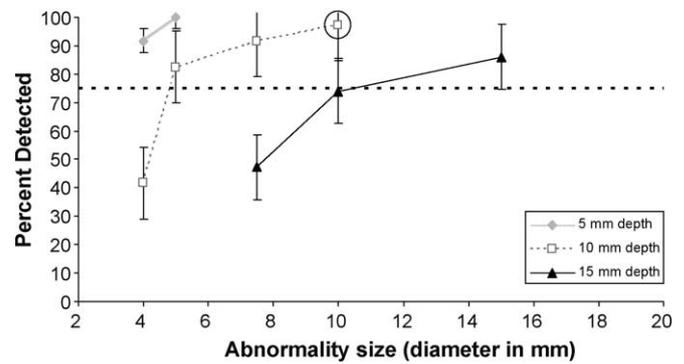
Fig. 2. Psychophysical functions for the detectability of abnormalities at 5, 10 and 15 mm depths as a function of abnormality size for prostate stiffness of 21 kPa. The dotted line denotes the 75% correct threshold. The circle denotes a size/depth combination expected to minimally surpass the 75% threshold when abnormality hardness is decreased, and is referenced in Fig. 3. Standard error bars are shown around the data points.

(Fig. 2), given equal depth. For example, at a depth of 15 mm, 17 mm abnormalities in the 82 kPa substrate were detected at approximately the same percentage (80%) as 10 mm abnormalities in the 21 kPa substrate (74%).

3.3. Effect of abnormality hardness on abnormality detection

To investigate the effect of abnormality hardness (held constant in objectives 1–2) on detectability, we analyzed specific size/depth combinations that were expected to minimally surpass the 75% threshold when abnormality hardness was decreased (see circled data points in Fig. 1 and Fig. 2). Fig. 3 illustrates the impact of varying the hardness of an abnormality that is 10 mm diameter and 10 mm depth, for both 21 and 82 kPa substrates. At the lowest abnormality hardness (23 durometers, type Shore A) the detectability rate in the stiffer substrate (82 kPa) drops close to the 75% threshold from the 100% detectable threshold when its hardness is greater. In contrast, the detectability rate in the more pliant substrate (21 kPa) was unaffected by changes to abnormality hardness.

3.4. Absolute threshold – abnormality hardness

Some size/depth combinations require a minimum abnormality hardness to be detected 75% of the time. Fig. 4 shows how abnormality hardness affects the detectability of four different
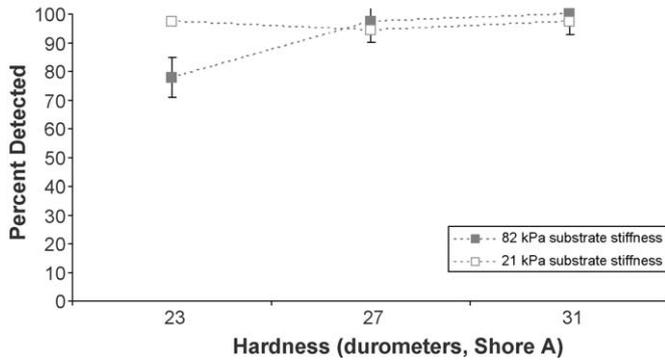
Fig. 3. Psychophysical functions for the detectability of abnormalities 10 mm diameter at 10 mm depth as a function of abnormality hardness for prostate stiffness levels of 82 and 21 kPa. Standard error bars are shown around the data points.

size/depth combinations in both substrates. In stiffer substrates (two top plots – 82 kPa), the decrease in detection rate with changes in the abnormality hardness is more evident. The detection rate decreased by 38% (Fig. 4a) and by 42% (Fig. 4b) when the hardness of the two abnormalities decreased from 31 to 23 durometers. In contrast for the 21 kPa simulated prostates, the detection rate exhibited a more gradual decrease (by 10% (Fig. 4c) and by 14% (Fig. 4d)) when abnormality hardness was decreased.

3.5. Variance in detection performance between examiners

Detection varied across stimulus combinations (abnormality size, depth, hardness and substrate stiffness) encountered by each

participant and also between participants. The hit rate for each participant was defined as the number of abnormalities detected by the participant divided by the total number of abnormalities presented to that participant. Participants' hit rates ranged from 95.8% to 48.5% (mean = 74.0%, SD = 13.0%), as indicated by the solid points in Fig. 5. Three (out of 18) participants were below one standard deviation of the mean hit rate. The distribution of participants' hit rates is shown in Fig. 6.

Furthermore, all stimulus combinations that were detected more than 90% of the time (from Figs. 1 and 2) were classified as "easy" stimulus combinations. All stimulus combinations that were detected less than 90% of the time were classified as "difficult" stimulus combinations. Participants' hit rates for "easy" stimulus combinations are indicated by the crosses (+) in Fig. 5, whereas hit rates for "difficult" stimulus combinations are indicated by the exes (x). The mean hit rate for all participants decreased (96.9% to 56.4%) while the standard deviation in hit rate increased (4.8% to 21.4%) from the "easy" to "difficult" stimulus combinations. Fig. 7 shows the hit rate distributions for these subsets of stimulus combinations.

The Pearson's product-moment correlation model was used to determine that there was no significant correlation between participants' hit rates and false alarm rates ($r = -0.174$, $p = 0.489$) for all stimulus combinations (Fig. 8).

4. Discussion

This study sought to characterize the range of prostate abnormalities that are palpable via digital rectal examination. Among the four factors (abnormality size, depth, hardness, and substrate stiffness), our overall finding was that at a depth of 5 mm, the majority of abnormalities can be detected
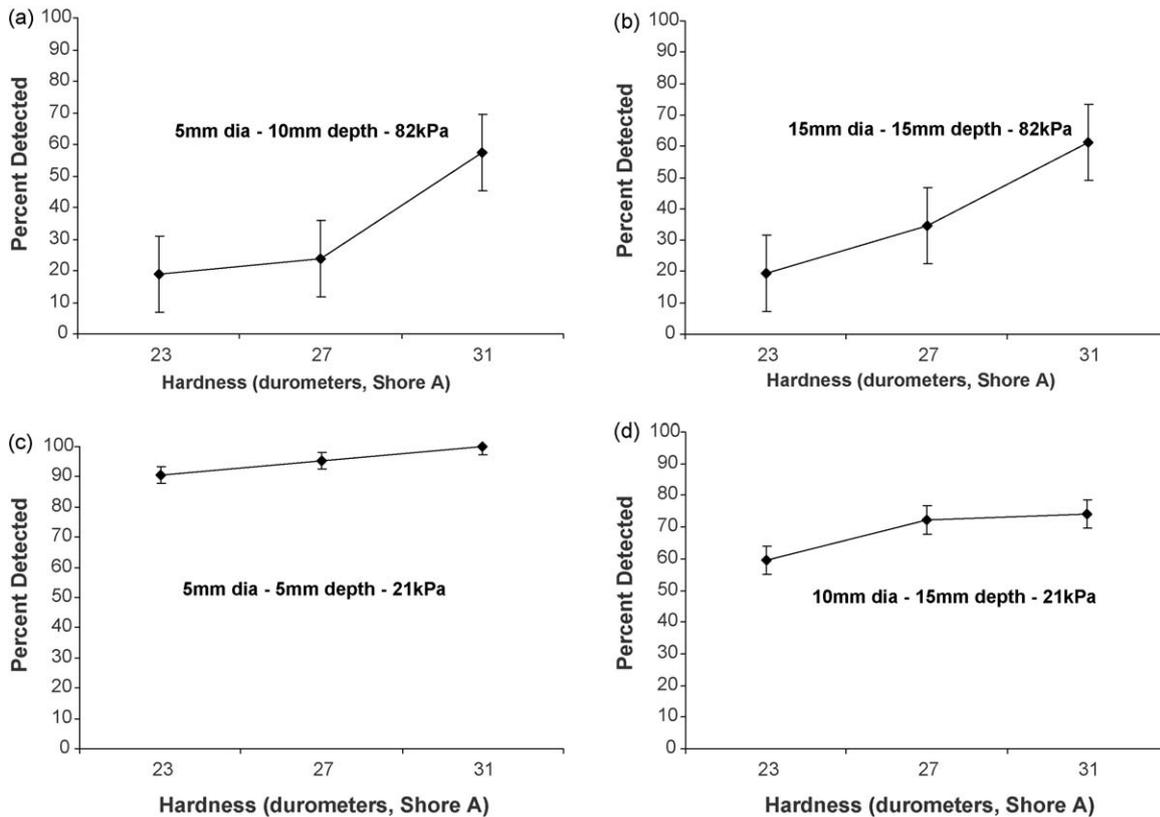
Fig. 4. Psychophysical functions for the detectability of four different size/depth combinations in two prostate stiffness levels as a function of abnormality hardness: (a) 5 mm diameter abnormality at a 10 mm depth embedded within prostate of stiffness 82 kPa, (b) 15 mm diameter abnormality at a 15 mm depth embedded within prostate of stiffness 82 kPa, (c) 5 mm diameter abnormality at a 5 mm depth embedded within prostate of stiffness 21 kPa, (d) 10 mm diameter abnormality at a 15 mm depth embedded within prostate of stiffness 21 kPa. Standard error bars are shown around the data points.
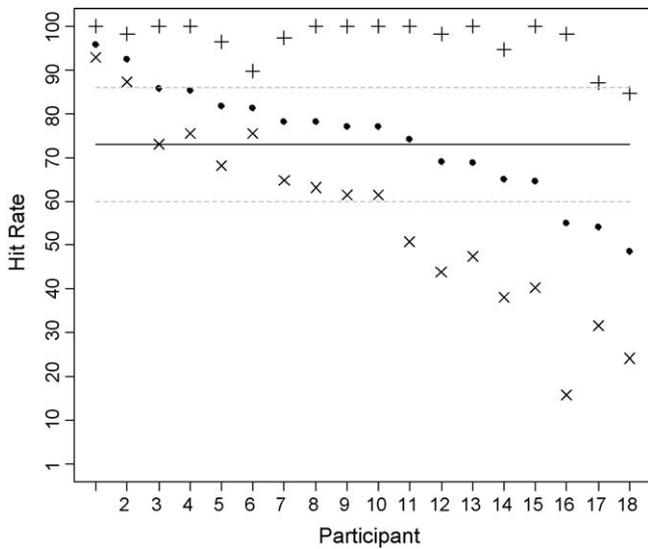
Fig. 5. Hit rates for all 18 participants. The solid points represent the hit rate for all stimulus combinations, the crosses (+) represent the hit rate for "easy" stimulus combinations, and the exes (x) represent the hit rates for "difficult" stimulus combinations. The solid line represents the mean hit rate for all stimulus combinations and the dashed lines represent one standard deviation above or below the mean hit rate for all stimulus combinations.
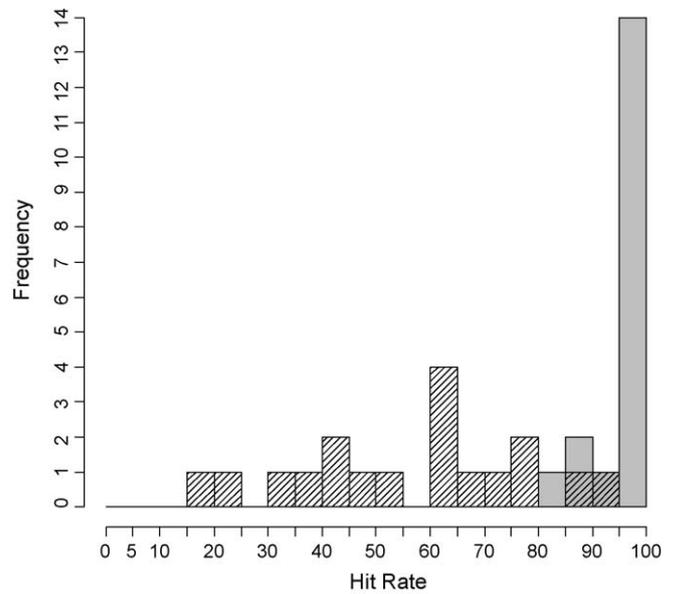
Fig. 7. Distribution of participants' hit rates for "easy" and "difficult" stimulus combinations. The "easy" stimulus combinations are indicated by the solid grey bars and the "difficult" stimulus combinations are indicated by the shaded bars.
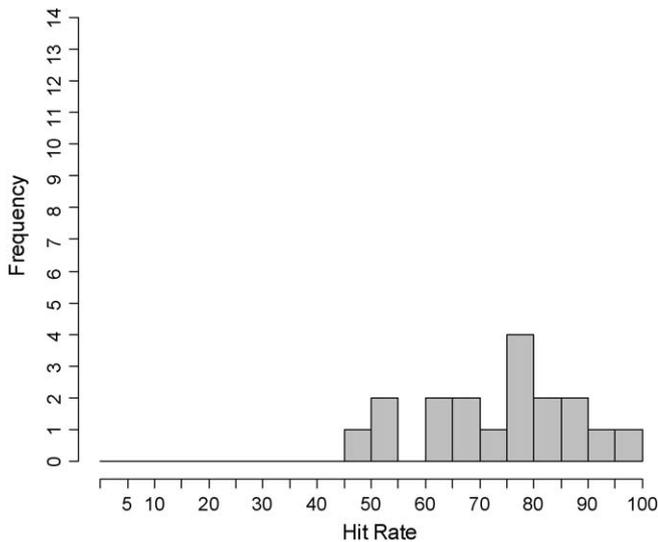
Fig. 6. Distribution of participants' hit rates for all stimulus combinations.
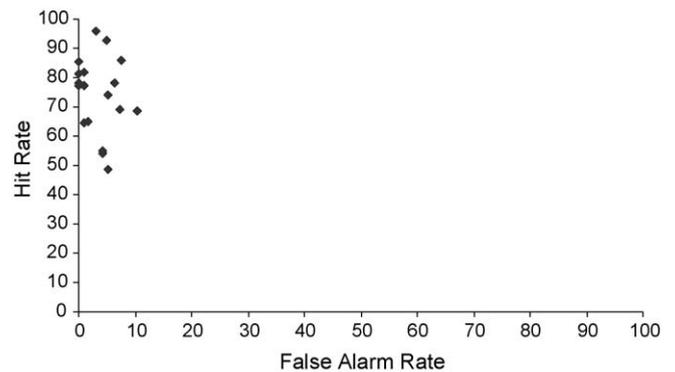
Fig. 8. Hit rate for all stimulus combinations versus false alarm rate for each participant.

independent of prostate stiffness or abnormality hardness or size. However, abnormalities located more deeply must be larger to be consistently detected. This is especially pertinent for the stiffer substrates, where abnormalities must be 10–17 mm diameter to be consistently detected at depths 10–15 mm. In more pliant substrates, detectable abnormalities at depths 10–15 mm are smaller (5–10 mm diameter). These results may aid in the development of training regimens so that scenarios utilize reasonable ranges per the four factors. Additionally the between examiner variance may also have implications in the clinic for a starting point in placing expectations on detection performance.

In comparison to simulated breast lumps, we found that the size of prostate abnormalities had be larger to be consistently detectable (5 mm compared to 3 mm) [6,7]. Additionally, our study found that depth played a larger role in abnormality detection. These findings are most likely due to the differences in substrate stiffness (21 or 82 kPa compared to 1–5 kPa).

Compared to size and depth, the hardness of an abnormality did not play as large a role, similar to results related to simulated breast tissue [6]. While hardness was not as prominent a factor overall, it did emerge as relevant when substrate stiffness and abnormality hardness were varied for the more deeply embedded (5–15 mm diameter and 10–15 mm depth) abnormalities. Detecting the hardness of an abnormality is most pertinent to detecting changes in prostate stiffness due BPH or prostititus, and therefore is relevant to larger and more deeply embedded tumors.

When we considered the performance of individual examiners, we uncovered that there were clearly a set of "easy" and "difficult" stimulus combinations. More effective training may need to be developed that reduces the number of "easy" training scenarios that yield high levels of detection performance and little variance between examiners. It is clearly the "difficult" scenarios that yield a lower level of detection performance and greater variability.

This experiment utilized simulated prostates in a simulated environment. The simulated prostates did not include fibrous tissue, varying surface texture, or undulations of the surface. Additionally, the experimental environment lacked any of the typical patient–clinician interactions. Ideally, future studies will augment our findings by using in vivo human prostate tissue with known abnormalities of various size, depth, and hardness. As a first

approximation, however, this study afforded a repeatable and controlled method and the results are valuable for gathering data on thresholds of detection and variance between examiners, similar to that done for the breast exam.

In relation, when analyzing the individual examiners, we also identified a low false alarm rate. This is likely because the simulated prostates did not include non-homogenous, fibrous tissue. The introduction of such elements would likely increase detection thresholds. That said, our false alarm finding does not indicate that better detectors are more willing to answer "yes" to improve their performance. Finally, participants were required to use the defined search technique. Other work indicates that search technique plays a role in detection performance [22] and should inform investigations, similar to that presented in here, to further quantify the impact of technique on improving the detection of smaller and deeper tumors.

Conflict of interest

None.

Acknowledgments

References

[1] American Cancer Society. Overview: Prostate cancer. 2009 [cited 03/10/09]; Available from: http://www.cancer.org/docroot/CRI/CRI_2_1x.asp?rnav=criov&dt=36.

[2] Smith DS, Humphrey PA, Catalona WJ. The early detection of prostate carcinoma with prostate specific antigen. Cancer 1997;80(9):1852–6.

[3] Thompson IM, Pauler DK, Goodman PJ, Tangen CM, Lucia MS, Parnes HL, et al. Prevalence of prostate cancer among men with a prostate specific antigen level <=4.0 ng per milliliter. N Engl J Med 2004;350(22):2239–46.

[4] Jarvis C. Physical Examination And Health Assessment, 4th ed., St. Louis, MO: Saunders, 2004.

[5] Güth U, Huang DJ, Huber M, Schötzau A, Wruk D, Holzgreve W, et al. Tumor size and detection in breast cancer: Self-examination and clinical breast examination are at their limit. Cancer Detect Prev 2008;32(3):224–8.

[6] Bloom HS, Criswell EL, Pennypacker HS, Catania AC, Adams CK. Major stimulus dimensions determining detection of simulated breast lesions. Percept Psychophys 1982;32(3):251–60.

[7] Fletcher SW, O'Mailey MS, Bunce LA. Physicians' ability to detect lumps in silicone breast models. J Am Med Assoc 1985;253(15):2224–8.

[9] Krouskop T, Wheeler T, Kallel F, Garra B, Hall T. Elastic moduli of breast and prostate tissues under compression ultrasound. Imaging 1998;20:260–74.

[10] Samani A, Bishop J, Luginbuhl C, Plewes D. Measuring the elastic modulus of ex-vivo small tissue samples. Phys Med Biol 2003;48:2183–98.

[11] McDermott MM, Dolan NC, Huang J, Reifler D, Rademaker AW. Lump detection is enhanced in silicone breast models simulating postmenopausal breast tissue. J Gen Intern Med 1996;11(2):112–4.

[12] Hall DC, Adams CK, Stein GH, Stephenson HS, Goldstein MK, Pennypacker HS. Improved detection of human-breast lesions following experimental training. Cancer 1980;46(2):408–14.

[13] Pilgrim CA, Lannon C, Harris PR, Cogburm W, Fletcher SW. Improving clinical breast examination training in a medical school. J Gen Intern Med 1993;8(12):685–8.

[14] McDermott MM, Dolan NC, Rademaker A. Effect of breast tissue characteristics on the outcome of clinical breast examination training. Acad Med 1996;71(5):505–7.

[15] Pennypacker H, Iwata M. MammaCare: a case history in behavioural medicine. In: Blackman D, Lejeune H, eds. Behaviour Analysis in Theory and Practice. Hillsdale, N.J.: Lawrence Erlbaum Assoc, 1991. pp. 259–288.

[16] Gerling GJ, Weissman AM, Thomas GW, Dove EL. Effectiveness of a dynamic breast examination training model to improve clinical breast examination (CBE) skills. Cancer Detect Prev 2003;27(6):451–6.

[17] Iannotti RJ, Finney LJ, Sander AA, De Leon JM. Effect of clinical breast examination training on practitioner's perceived competence. Cancer Detect Prev 2002;26(2):146–8.

[18] Lee KC, Dunlop D, Dolan NC. Do clinical breast examination skills improve during medical school? Acad Med 1998;73(9):1013–9.

[19] Gerling GJ, Rigsbee S, Childress RM, Martin ML. The design and evaluation of a computerized and physical simulator for training clinical prostate exams. IEEE Trans Syst Man Cybern A Syst Hum 2009;39(2):388–403.

[20] Campbell HS, Fletcher SW, Pilgrim CA, Morgan TM, Lin S. Improving physicians' and nurses' clinical breast examination: a randomized controlled trial. Am J Prev Med 1991;7(1):1–8.

[21] Gescheider GA. Psychophysics: the fundamentals: Lawrence Erlbaum Associates; 1997.

[22] Wang N, Gerling GJ, Childress RM, Martin ML. Characterizing Finger Palpation in the Detection of Prostate Cancers and Abnormalities. Proceedings of the Human Factors and Ergonomics Society. Annual Meeting 2008;813–7.

Quantifying Palpation Techniques in Relation to Performance in a Clinical Prostate Exam

Ninghuan Wang, Gregory J. Gerling, *Member, IEEE*, Reba Moyer Childress, and Marcus L. Martin

Abstract—**This paper seeks to quantify finger palpation techniques in the prostate clinical exam, determine their relationship with performance in detecting abnormalities, and differentiate the tendencies of nurse practitioner students and resident physicians. One issue with the digital rectal examination (DRE) is that performance in detecting abnormalities varies greatly and agreement between examiners is low. The utilization of particular palpation techniques may be one way to improve clinician ability. Based on past qualitative instruction, this paper algorithmically defines a set of palpation techniques for the DRE, i.e., global finger movement (GFM), local finger movement (LFM), and average intentional finger pressure, and utilizes a custom-built simulator to analyze finger movements in an experiment with two groups: 18 nurse practitioner students and 16 resident physicians. Although technique utilization varied, some elements clearly impacted performance. For example, those utilizing the LFM of vibration were significantly better at detecting abnormalities. Also, the V GFM led to greater success, but finger pressure played a lesser role. Interestingly, while the residents were clearly the superior performers, their techniques differed only subtly from the students. In summary, the quantified palpation techniques appear to account for examination ability at some level, but not entirely for differences between groups.**

Index Terms—**Biomedical measurements, human factors, medical decision making, simulation.**

I. INTRODUCTION

TO PROMOTE the early detection of prostate cancer, the American Cancer Society advises screening via both the prostate-specific antigen (PSA) blood test and the digital rectal exam (DRE) [1]. Clinicians palpate the prostate gland in conducting a DRE. Since there is a high false-positive rate (67% [2], [3]) for the PSA blood test, the DRE conducted concurrently is a valuable procedure used in the detection of cancer. Presently with the DRE, clinician performance in detecting tumors is variable (positive predictive value is 17%–34% [1] and agreement between examiners on diagnosis is low (21%–40% [4], [5]). Hands-on training is typically conducted for the DRE via standardized patients or physical simulators [6], [7] and improves

palpation skills, in general [8]. However, we do not understand exactly which elements of training improve performance.

It may be that performance improves as clinicians become more proficient in their hands-on technique. Palpation techniques, for example, in the clinical breast exam (CBE), include global finger movement (GFM) for the search of abnormalities over the entire breast, local finger movement (LFM) for the palpation of small areas of the breast, and levels of finger pressure for adequate search through the tissue depth [9]–[11]. A range of studies have shown that examiners, who utilize these components in a specified fashion, increase mean lump-detection rates at a statistically significant level [12], [13], in some cases doubled from pre- to posttest [9], [14]. In addition, skills learned on silicone models transfer to the detection of human lesions [14]. Together, these results tend to indicate that improved technique correlates with improved performance, although the majority of studies focus upon the CBE, rather than the DRE.

Among the three palpation techniques typically cited for the CBE, often discussed is the technique of GFM. Specific GFMs are the vertical strip, concentric circle, and radial spoke [15]. For example, in utilizing 1) the vertical strip, one begins at one outside quadrant and continues in overlapping vertical strips; 2) the concentric circle, one originates at outermost top of the breast and moves in concentric circles until terminating at nipple; and 3) the radial spoke, one begins at the perimeter of the conical part of the breast and converges in a vector toward the nipple. Study results indicate that use of the vertical strip leads to an increased detection of abnormalities [10], [16], as it may enable a more efficient and thorough coverage of tissue. Others have shown that search technique plays a lesser role [17]–[19], although these studies evaluated breast self-exam (BSE) rather than CBE.

Additionally for the technique of LFM, the literature advises palpation with the finger pads of the middle three fingers in dime-sized circular motions [15], [20], [21]. Others suggest using two or three fingers in circular motions with varying pressure, rolling tissue between two fingers, sliding the fingers over the surface of the breast, or employing the patterns in combination [22]. While such strategies are widely promoted and utilized in practice, one of the only empirical comparisons of circular and sliding patterns found that neither resulted in more lumps being detected in breast models, although sliding movement was associated with increased false-positive reports [19]. Finally, for the technique of finger pressure, the application of three magnitudes (light subcutaneous, medium midlevel, and deep-chest wall) advocates increased pressure for each concentric circle [9], although little evidence supports its effectiveness.

Studied widely for the CBE, few techniques have been introduced in relation to the DRE, although typically mentioned is the need to search its entire volume.

Few researchers have sought to quantify hands-on techniques. This is due, in part, to the lack of sensor arrays to track finger motion, lack of algorithms and tools to analyze finger motion, and lack of devices to present a variety of scenarios, although several simulators are moving in this direction [6], [23], [24]. Such analysis might allow us to determine if certain techniques are consistently utilized or positively correlated with performance over a range of test scenarios. Work in this direction has begun for the female pelvic exam, utilizing a simulator (E-Pelvis) that can track finger motion [25]. The results of Pugh's work introduced three techniques (number of critical areas touched, frequency at which areas were touched, and maximum pressure exerted) that were correlated with ability to diagnose simulated findings. More broadly, analysis of force sensor data has sought to classify human behaviors and knowledge, e.g., stride patterns in human gait [26] and characteristics underlying activities of walking, running, and cycling [27], though many examples in the literature span medical and athletic domains.

In this paper, our overall goal is to quantify hands-on techniques that characterize human performance in the palpation of prostate abnormalities via a custom-built simulator. In doing so, we seek to determine if certain techniques are related to superior performance.

II. METHODS

To analyze palpation technique and its link with performance, we conducted an experiment with two groups: group 1 with 18 nurse practitioner students and group 2 with 16 resident physicians, using the Virginia Prostate Examination Simulator (VPES), described later. Utilizing algorithms and software developed here, the objectives were to determine:

1) if performance in detecting abnormalities differs between the two groups;
2) the characteristic patterns that underlie the palpation techniques of GFM (e.g., U, V, L, or Line), LFM (e.g, tapping, vibration, or sliding), and average intentional finger pressure (AIFP);
3) for each palpation technique, if the identified patterns are consistently employed;
4) if their utilization differs between the two groups;
5) if certain techniques and patterns yield better performance in the number of abnormalities detected.

A. Apparatus: VPES

The design of the VPES utilizes rubber-like materials to simulate the feel of tissue while using a computer to reconfigure test scenarios and pressure sensors to record finger pressure for immediate feedback and postperformance review [6]. See previous reference for in-depth detail on the simulator.

Three instrumented prostate models with accurate size and stiffness are attached to a tracking system internal to a posterior torso. Their dimensions are 5.5 cm (transverse, width dimension) by 5.0 cm (longitudinal, length dimension). Four to six polyethylene balloons are embedded in each instrumented prostate and filled with water to simulate palpable abnormalities. Deflated balloons are not palpable. In this study, we utilized the VPES to simulate normal and abnormal prostate conditions, including prostatitis (enlarged and boggy inflammation) and carcinoma (small and firm isolated nodule) by utilizing balloons in various configurations of size and location, but similar hardness ($\sim$30 Shore A durometers). Balloons of increasing size are named from "A" to "E" and are positioned in different locations for each prostate. The size of the smallest balloon "A" is 0.5 by 0.5 cm, while that of the largest balloon "E" is 3.0 by 1.5 cm.

Using water pressure and force sensors, respectively, the VPES captures the finger pressure employed on inflated balloons and on the prostate. A change in water pressure denotes that a user has palpated an abnormality and this information is logged for postexamination analysis. Accompanying the water pressure sensors, force sensors monitor forces exerted by the examiner's finger on the prostate. Four force sensors (Tekscan, South Boston, MA, Flexiforce 0-1 lb) are embedded at the base of each instrumented prostate and record the location and magnitude of applied finger pressure.

B. Experiment

1) Participants: Eighteen nurse practitioner students and 16 resident physicians participated in the human–subjects experiment, approved by the Institutional Review Board at the University of Virginia. Among the nurse practitioner students (group 1) were 4 males and 14 females, ages 23–47 ($M = 33.90$, standard deviation (SD) $= 8.56$), while among the resident physicians (group 2) were six males and ten females, ages 26–32 ($M = 28.70$, SD $= 1.70$). The major difference between two groups is that resident physicians have much more clinical experience in DRE. In general, group 2 had performed more DREs in the clinic (*median* $= 15$) than group 1 (*median* $= 0$). Based on their recollections, all resident physicians had performed at least nine clinical exams, while only five nurse practitioner students had ever performed a clinical exam, within the preceding 18 months.

2) Basic Setup: The overall procedure included the experience questionnaire, a hands-on orientation, and a hands-on testing session. In the 5-min orientation, the participant palpated the VPES in three scenarios. After each scenario, the proctor pointed out simulated abnormalities on a printed figure. During the 20-min testing session, each participant was given 30 s to palpate each of five scenarios. The participants were asked to palpate the entire prostate, as they would in a clinical exam. After each 30-s testing session, the participant identified any abnormalities to the proctor stating, in particular, their size and location. Utilizing two proctors, the first proctor facilitated paper work, hands-on sessions, and recording the verbal report of detections (on a standardized score sheet, where data were the size, location, and hardness of an abnormality). While this proctor was blind to scenario, the second proctor controlled the order of scenarios presented. The proctors trained in a preliminary experiment.

3) Statistical Analysis: We generated logistic regression models by employing the GENMOD procedure in SAS,

Version 9.1, which uses generalized estimating equations for repeated measurements. Factors were GFM × LFM × AIFP × group × (abnormality simulated or testing scenario). For the technique variables, there were four levels of GFM, three levels of LFM, and one continuous variable of average AIFP. These are described algorithmically in *high-level algorithms*. Group was a between-subjects factor: group 1 and group 2. Abnormality simulated was a within-group factor consisting of six simulated abnormalities in four testing scenarios (scenarios 2–5). While abnormality simulated applied to LFM and AIFP, the within-group factor for GFM was the four test scenarios (2–5).

In regards to performance (P), the dependent variable is a correct detection, valued 1 or 0. The performance variables applied to the six abnormalities simulated (LFM/AIFP) or four test scenarios (GFM). Note that the normal prostate (scenario 1) was not analyzed because no abnormality was simulated.

Given these variables, we formed a relationship between performance and technique in (1), where α_1, α_2, α_3, α_4 are coefficients associated with each factor. Note that coefficient α_4 relates to group, where there are multiple subdifferences, including accumulated experience and disciplinary training. In addition, we use β to represent uncertainty related to other cognitive elements, at the level of tactile object recognition, which certainly impact proficiency, but which are difficult to decouple from physical techniques. We know that knowledge of disease and attitudes about the domain are not strongly related to proficiency [28] and the basic rules for diagnosing abnormalities, as certain diseases, are not difficult to master [6]

$$P = \alpha_1 \text{GFP} + \alpha_2 \text{LFP} + \alpha_3 \text{AIFP} + \alpha_4 \text{group} + \beta. \quad (1)$$

In an analysis of sample size for a regression analysis, we found that at least 144 participants (12 technique combinations × 6 abnormalities simulated × 2 groups) would be required. This is well more than the 34, we employ here. Therefore, we present the results here as preliminary and as the rationale for entering into a larger scale (and more costly) study. In addition, because the participant number is small in comparison to the number of explanatory variables in the model, it is possible that our model overfits the data. For this reason, we also investigated the explanatory variables separately.

4) Prostate Scenarios Simulated: Each of the eight total scenarios utilized one of three instrumented prostates in either normal or abnormal (prostatitis or carcinoma) states. Orientation scenarios 1–3 and testing scenarios 1–5 were used (see Fig. 1). Note that orientation scenario 3 and testing scenario 1 are not shown because they are normal cases, with no balloons. The choice of simulated abnormalities was made in consultation with clinicians, to include sizes and locations typical of tumors and prostatitis. Moreover, the simulated order of the testing scenarios was varied across participants.

C. Data and Data Collection

Three sources of data collected in the testing session were analyzed, which are as follows: 1) finger pressure from force sensors; 2) balloon pressure; and 3) participant reports to proctor of abnormalities palpated. The finger pressure data were
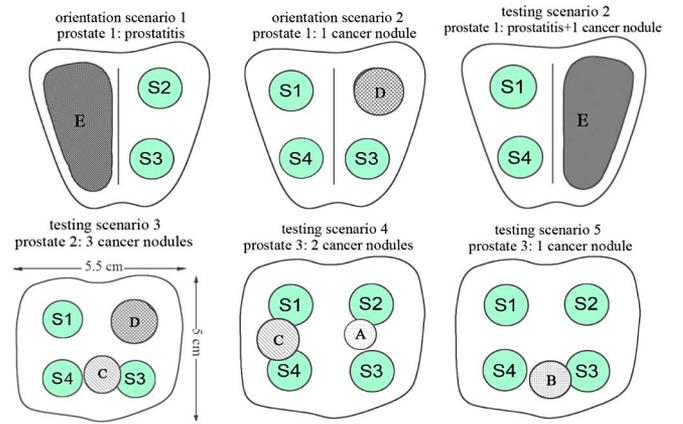
Fig. 1.   Simulated orientation and testing scenarios. "S1–S4" denote positions of force sensors and letters "A–E" denote sizes of balloons.

collected from the four force sensors at a rate of 1000 samples per second, set by the resolution of the A/D converter, over the ∼30 s exam. The data were converted from volts (V) to Newtons (F) in the following equation:

$$F = 1.7744 \times V. \quad (2)$$

Balloon sensor data were sampled from the water pressure transducers attached to balloons and at a rate of 3 samples per second. The only use of the balloon sensor data was to help to determine the interval over which to analyze LFM, i.e., when the examiner palpated a balloon.

The third source of data was the paper record maintained by the proctor, indicating if a participant reported palpating any abnormalities.

D. High-Level Algorithms

The analysis focuses upon three palpation techniques: GFM, LFM, and AIFP. A preliminary analysis of the data had identified four GFMs (U, V, L, and Line) and three LFMs (tapping, vibration, and sliding) [29]. Algorithms were set up to mathematically characterize these techniques and inform the design of an analysis tool.

1) Global Finger Movement: GFM is defined as the systematic movement of one's finger over the entire prostate. Four patterns are defined as U, V, L, and Line with deviations in the orientation for each pattern.

The algorithm to identify the GFM follows three main steps: 1) eliminate low-magnitude noise of the circuit and any noise initially produced by a balloon inflated over the sensor; 2) discretize the resultant continuous data into binary values (0 or 1) every 16 ms for each sensor, to create a state for the group of four sensors (e.g., 0 0 0 0, 0 1 1 0, or 1 1 1 1); and 3) recognize the GFM as the sequence of sensor states. First, the two types of noise are eliminated that are produced from the hardware circuit (*noise*, valued at 0.005 N) and that are produced from the initial forces of a balloon on the ith sensor (tol^i). The following equation (3) is thereby employed for the ith sensor data at time t (s_t^i):

$$s_t^i = s_t^i - \text{noise} - tol^i. \quad (3)$$
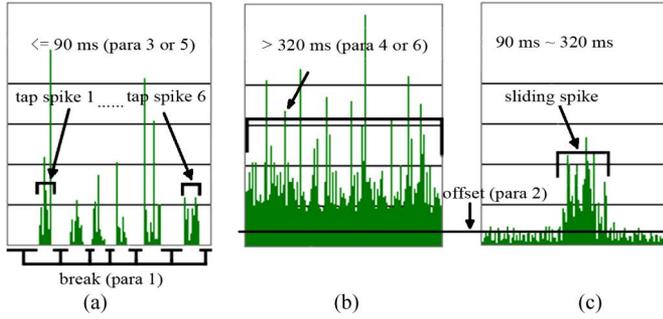
Fig. 2. Three patterns of LFM. Identifying characteristics (parameters 1–6) marked in the panels accompany algorithmic descriptions in the text. (a) Tapping. (b) Vibration. (c) Sliding.

Then, the ith sensor data are discretized (bit_t^i) via (4), if it is larger than $g\_threshold$, set to indicate the intentional pressure exerted by an examiner's finger. The $g\_threshold$ is set at 0.05 N, i.e., minimum AIFP

$$\text{bit}_t^i = 1 \qquad \text{if } s_t^i > g\_\text{threshold}, \qquad \text{otherwise } \text{bit}_t^i = 0. \tag{4}$$

The resultant bit values for sensors 1–4 are used to identify a sensor state at time t (state_t). Sixteen states were defined in advance as 4-bit numbers, ranging from 0 0 0 0 to 1 1 1 1. The leftmost bit stands for sensor 1 and the rightmost bit stands for sensor 4. The bit is set as 1 if the sensor is triggered and 0 otherwise. For example, a state defined as 0 1 1 1 specifies a point in time whereby sensors 2, 3, and 4 are triggered simultaneously, while sensor 1 is untouched.

The completion of this procedure typically results in a high frequency of sensor states, i.e., one for every 16 ms time period. Sequential time periods at the same state indicates a static point, i.e., lack of finger movement. Between the static points are transition vectors, i.e., movements of the finger from one state to another. The static points and transition vectors are conceptually similar to fixations and saccades in vision, although they happen on a different time scale (static point = accumulated duration of at least 16 ms up to 30 s, vectors = 16 ms duration; where 16 ms is set by the resolution of the A/D converter). Therefore, GFM is made up of transition vectors between static points.

2) Local Finger Movement: LFM is defined as palpation by finger movement within a single quadrant of the instrumented prostate or near a single abnormality. Three patterns are defined as tapping, vibration, and sliding (see Fig. 2).

Another set of experiments was conducted to identify the parameters that objectively characterize these three patterns. Pressure was exerted on each of the four force sensors for three iterations, applying the quantitatively defined local techniques following the order of tapping, vibration, and sliding. Parameters were adjusted until the analysis tool recognized the applied pattern. After averaging the parameter values for three iterations, the following parameters distinguish the patterns:

1) the break between two periods of continuous spikes;
2) the offset to recognize maintained finger pressure for vibration and sliding;
3) the longest period for the tapping pattern;
4) the smallest period for the vibration pattern;
5) the smallest period for the sliding pattern is the value of which equals parameter 3;
6) the longest period for the sliding pattern is the value of which equals parameter 4.

Based on these parameters, the three algorithms for LFM were quantitatively defined. Note that a spike in the sensor data is defined as $S\_\text{Spike}_t^i = (s_t^i > l\_\text{threshold})$, i.e., pressure exerted above a threshold ($l\_\text{threshold} = g\_\text{threshold}$), set to indicate the intentional finger pressure on the ith sensor. In addition, t_{S_Spike} represents the time when a spike ($S\_\text{Spike}_t^i$) appears. Therefore, different from the analysis of GFM, the continuous data from sensors were not compiled into binary states.

A tapping pattern is identified as an examiner striking briefly on the defined area, before quickly pulling up his finger, and doing so in an intermittent fashion. As shown in data of Fig. 2(a), the tapping pattern is the duration of continuous spikes not longer than 90 ms (parameter 3), followed by a sequence of no spikes for less than 48 ms (parameter 1). Given the sampling rate of the A–D converter (16 ms per sample), up to three spikes could appear in a time span of 48 ms. The tapping pattern is defined formally for the ith sensor L_tap^i in the following equation, where the period between time j and k is the break with no spikes (parameter 1) and $j < k$:

$$L_\text{tap}^i = \sum_{t=j}^{k} t_{S\_\text{Spike}_t^i} \le 90 \qquad t_{S\_\text{Spike}_j^i} - t_{S\_\text{Spike}_k^i} < 48. \tag{5}$$

A vibration pattern is identified as an examiner maintaining finger contact with a single area, but shifting the weight of the finger (e.g., proximally to distally, side to side, or in a circular motion). A shown in data of Fig. 2(b), the vibration pattern is defined formally for the ith sensor in (6) as an examiner maintaining finger pressure above a certain value (parameter 2) on the prostate over a continuous time span of at least 320 ms (parameter 4). The frequency of oscillation is not considered

$$L_\text{vibr}^i = \sum_{t=j}^{k} t_{S\_\text{Spike}_t^i} \ge 320 \qquad S\_\text{Spike}_t^i > \text{offset}. \tag{6}$$

A sliding pattern is identified as an examiner moving the finger smoothly over a defined area without stopping. The observation of a sliding pattern indicates a lack of finger pause. This comes in contrast to the tapping and vibration patterns, where the examiner appears to transition from the GFM to an intentional focus upon the local detection of a balloon. As shown in Fig. 2(c), the sliding pattern is defined formally for the ith sensor in the following equation as an examiner maintaining pressure above a certain value (parameter 2) continuously for 90–320 ms (parameters 5 and 6):

$$L_\text{sl}^i = 90 < \sum_{t=j}^{k} t_{S\_\text{Spike}_t^i} < 320 \qquad S\_\text{Spike}_t^i > \text{offset}. \tag{7}$$

3) Average Intentional Finger Pressure: We calculate AIFP as that applied over the duration of the exam in the vicinity of filled balloons. If a filled balloon is located directly over a force

TABLE I
LOGISTIC REGRESSION (LR) MODELS WITH ALL FOUR FACTORS: (a) SCORE
STATISTICS FOR EACH FACTOR IN THE OVERALL MODEL. (b) LR MODEL WITH
ALL FOUR FACTORS AS PREDICTORS

(a)

Source	DF	Chi-Square	p-value
GFM	3	4.11	0.2499
LFM	2	19.14	<.0001
Pressure	1	0.71	0.3981
Group	1	8.44	0.0037

(b)

Coefficient	Estimate	Standard Error	95% Confidence Interval		p-value
Intercept	-1.4338	0.5569	-2.5253	-0.3423	0.01
GFM_U	0.041	1.1559	-2.2245	2.3065	0.9717
GFM_V	0.7531	0.4226	-0.0752	1.5814	0.0747
GFM_L	0.6935	0.3082	0.0894	1.2977	0.0245
GFM_Line	0	0	0	0	-
LFM_Tap	-0.0122	0.5629	-1.1155	1.0911	0.9827
LFM_Vib	2.2073	0.4174	1.3892	3.0254	<.0001
LFM_Slid	0	0	0	0	-
Pressure	0.1749	0.1866	-0.1908	0.5407	0.3486
Group_1	-1.0803	0.3319	-1.7308	-0.4298	0.0011
Group_2	0	0	0	0	-

sensor, the finger pressure measured by that force sensor alone is used in calculating the average. Otherwise, if a filled balloon is located between two sensors, the finger pressure from both sensors is averaged.

Further, as a simple average is heavily weighted by the lack of any finger pressure, AIFP (s_{avg}^i) was computed in the following equation, for sensors $i = 1$–4, where s_t^i is force sensor data at time t for sensor i, T is the time interval, and the parameter $p\_$threshold ($= g\_$threshold) indicates intentional finger pressure:

$$s_{\mathrm{avg}}^i = \frac{1}{T} \sum_{t=1}^{T} s_t^i, \qquad \text{subject to } s_t^i > p\_\text{threshold}. \qquad (8)$$

III. RESULTS

This preliminary data were analyzed in accordance with the five objectives. Recall that objectives 1 and 5 involve performance, while objectives 2 and 3 relate to technique only, independent of performance. We first note the overall results of a logistic regression model setup to determine the impact of the explanatory variables (GFM, LFM, AIFP, and group) on performance. As evident in Table I(a), LFM and group both significantly impact performance ($p < 0.0001$ and $p = 0.0037$, respectively). This is discussed in details in Section III-E.

A. Performance Differences Between Groups (Objective 1)

In a trend, which held across simulated abnormalities, group 2 detected more abnormalities than did group 1 (see Table II). To test for a statistical difference, a logistic regression model was used with group as the categorical predictor in (9), where p is the probability that the abnormality was successfully detected.

TABLE II
PERFORMANCE IN DETECTING SIMULATED ABNORMALITIES

Testing Scenario	Balloon	Correct Detectors Group 1 (Novice, N=18)	Correct Detectors Group 2 (Experienced, N=16)
1	None	10	12
2	E	11	14
3	C	5	8
	D	5	7
4	A	3	4
	C	6	9
5	B	11	13

TABLE III
LOGISTIC REGRESSION (LR) MODELS WITH EACH FACTOR SEPARATELY.
(a) LR MODEL WITH GROUP AS THE PREDICTOR. (b) LR MODEL WITH GFM
AS THE PREDICTOR. (c) LR MODEL WITH LFM AS THE PREDICTOR. (d) LR
MODEL WITH AIFP AS THE PREDICTOR

(a)

Coefficient	Estimate	Standard Error	95% Confidence Interval		p-value
Intercept	0.2938	0.2083	-0.1145	0.702	0.1584
Group 1	-0.7849	0.3037	-1.3802	-0.1896	0.0098
Group 2	0	0	0	0	-

(b)

Coefficient	Estimate	Standard Error	95% Confidence Interval		p-value
Intercept	-0.73	0.278	-1.2748	-0.1852	0.0086
GFM_U	0.7523	0.5935	-0.4109	1.9155	0.2049
GFM_V	1.1733	0.3943	0.4006	1.9461	0.0029
GFM_L	0.7279	0.3138	0.104	1.3518	0.0222
GFM_Line	0	0	0	0	-

(c)

Coefficient	Estimate	Standard Error	95% Confidence Interval		p-value
Intercept	-1.0807	0.3151	-1.6983	-0.463	0.0006
LFM_Tap	-0.3037	0.5015	-1.2867	0.6793	0.5549
LFM_Vib	2.1079	0.3622	1.3979	2.8179	<.0001
LFM_Slid	0	0	0	0	-

(d)

Coefficient	Estimate	Standard Error	95% Confidence Interval		p-value
Intercept	-0.6013	0.312	-1.2129	0.0102	0.054
Pressure	0.3093	0.1485	0.0183	0.6003	0.0372

The reference level is group 2, with coefficient 0

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1(\text{group}\_1). \qquad (9)$$

Table III(a) shows that the estimated coefficient for group 1 is negative (-0.7849) and significant ($p = 0.0098$). In addition, the score statistic for group in this model has p-value 0.0159 (not displayed in Table III(a)). Therefore, being a participant in group 1 decreased the odds of successful detection by a factor of 0.46 [$\exp(\beta_2)$] relative to being in group 2.

TABLE IV
EMPLOYED GFM PER PARTICIPANT (NO.) PER TESTING SCENARIO (T.S)

No.	T.S 2	T.S 3	T.S 4	T.S 5	No.	T.S 2	T.S 3	T.S 4	T.S 5
				GFM for Group 1					
1	V	Line	Line	Line	10	V	Line	V	Line
2	Line	V	V	U	11	V	V	Line	V
3	V	V	Line	V	12	V	V	L	Line
4	V	V	V	L	13	Line	Line	Line	L
5	V	V	V	V	14	V	U	L	V
6	V	L	Line	L	15	Line	Line	Line	U
7	L	V	L	L	16	V	Line	Line	V
8	V	Line	Line	V	17	V	V	V	Line
9	V	Line	L	L	18	Line	L	Line	Line
				GFM for Group 2					
1	V	Line	V	L	9	Line	Line	Line	Line
2	Line	V	V	Line	10	Line	Line	Line	L
3	L	Line	Line	V	11	L	Line	U	Line
4	V	V	V	Line	12	V	V	L	L
5	Line	Line	Line	V	13	Line	Line	L	Line
6	L	V	L	L	14	Line	V	L	Line
7	Line	Line	L	V	15	Line	L	L	V
8	V	V	V	L	16	Line	L	L	L

TABLE V
EMPLOYED LFM PER PARTICIPANT (NO.) PER ABNORMALITY (T.S 2E—T.S 5B) (1 = TAPPING, 2 = VIBRATION, 3 = SLIDING, "-" = UNTOUCHED, AND TEXTS HIGHLIGHTED INDICATE EMPLOYING A CONSISTENT PATTERN)

No.	T.S 2E	T.S 3C	T.S 3D	T.S 4A	T.S 4C	T.S 5B	No.	T.S 2E	T.S 3C	T.S 3D	T.S 4A	T.S 4C	T.S 5B
						LFM for Group 1							
1	2	-	1	-	1	-	10	2	1	3	2	2	3
2	2	2	2	1	1	2	11	2	2	1	1	1	1
3	1	2	3	1	1	1	12	2	2	2	3	1	3
4	2	2	2	3	2	2	13	2	1	1	2	1	2
5	2	2	2	2	2	2	14	2	3	3	1	1	1
6	2	2	2	3	2	1	15	2	1	1	3	2	2
7	3	2	1	1	2	1	16	2	3	1	2	3	1
8	2	1	1	2	1	2	17	2	2	1	1	1	2
9	2	2	1	2	2	2	18	1	1	1	1	-	-
						LFM for Group 2							
1	2	1	2	2	3	1	9	2	2	1	2	-	2
2	3	2	1	2	2	1	10	2	2	1	1	3	2
3	1	1	1	3	3	1	11	2	2	2	2	2	2
4	3	2	3	3	1	1	12	2	2	2	3	3	1
5	2	3	1	2	3	2	13	2	2	1	2	2	1
6	3	2	1	3	3	2	14	2	2	3	2	2	1
7	2	2	1	1	2	2	15	3	2	2	3	2	1
8	2	2	2	2	2	2	16	2	2	1	3	1	1

B. Identified Patterns Within Palpation Techniques (Objective 2)

For GFM, by a visual inspection of Table IV, all participants utilized either U, V, L, or Line patterns. For LFM, most participants utilized either tapping, vibration, or sliding (see "1," "2," and "3" in Table V), although in a few cases abnormalities went untouched ("–"). Note that we dropped six observations from occurrences, where sensors were not touched (indicated by dash). In each case, the participant failed to find the abnormality because his or her finger never moved over the abnormality. Three of these observations came from one nurse practitioner student in group 1 and two others came from another participant in group 1. The sixth came from a resident physician in group 2. Since these cases produced certain failure in GEN-MOND when the repeated measures were included to address dependence between observations from the same participant, they were dropped from the dataset. Thus, 198 observations (34 participants × 6 abnormalities, minus the 6 dropped cases) were left for analysis of LFM and AIFP.

For AIFP, participants exerted 0.05–5.29 N of pressure on force sensors in vicinity of that abnormality.

C. Consistency of Pattern in Palpation Techniques (Objective 3)

For GFM, no participant utilized the same pattern across his or her five testing scenarios (within-subjects, i.e., rows in Table IV). Nor was any pattern consistently utilized by all trainees over a single testing scenario (between-subjects, i.e., columns in Table IV). We can also test the probability of choosing an inconsistent examiner from the population of examiners (P_{GFM}) if we assume that the 34 subjects are representative of prostate examiners and that the choice of dominant GFM is independent of the scenario. In doing so in a one-sided significance test (where $\alpha = 0.05$) of H_0: $P_{\mathrm{GFM}} = X$ and $H\alpha$: $P_{\mathrm{GFM}} > X$, where all 34 subjects are inconsistent in their predominant GFM, the fact that H_0 is true means that P_{GFM} is larger than 0.9160 (given by X^{34}).

For LFM, only two participants utilized a single pattern across all six presented abnormalities (within-subjects, i.e., highlight rows in Table V), and no single pattern was consistently utilized by all trainees for a simulated abnormality (between-subjects, i.e., a column in Table V).

For AIFP, SD and mean were examined as metrics to evaluate the dispersion and consistency of AIFP in the vicinity of the simulated abnormality. Across the six simulated abnormalities, the SD range of AIFP for each participant is 0.16–2.34 N. To further test the inconsistency, we used a Mantel–Haenszel statistic to see if mean AIFP was different between participants, adjusting for abnormality. The result was significant ($p < 0.0001$), which indicated that no participant employed consistent finger pressure across the six simulated abnormalities. Nor did all trainees exert similar finger pressure on a given simulated abnormality. The SD range of AIFP by all 34 participants on each abnormality is 0.73–1.08 N. Here, the Mantel–Haenszel statistic was also significant ($p < 0.0001$), thus indicating that mean AIFP varied between abnormalities across the 34 participants.

D. Difference in Techniques Utilized by Groups 1 and 2 (Objective 4)

For GFM, both group 1 and group 2 employed similar patterns. The Fisher's exact test was conducted to see if utilized GFM was independent of group. The p-values for each scenario (testing scenario 2–5) were 0.0160, 0.9399, 0.3083, and 0.6284, respectively. However, we conducted an additional test using the Bonferroni approach to validate preliminary significance for scenario 2. Because the probability of a type I error increases with a small sample size, the level of significance was set equal to desired significance level (0.05) divided by number of tests (4). At ($p < 0.0125$), there were no statistically significant differences between the groups for any testing scenario.

For LFM, both group 1 and group 2 employed similar patterns. The Fisher's exact test was conducted to see if utilized LFM was independent of group. The p-values for individual abnormalities were 0.3538, 0.4590, 1.000, 0.2007, 0.0222, and 0.4200, respectively, for test scenario (T.S) 2, abnormality E, T.S 3C, T.S 3D, T.S 4A, T.S 4C, T.S 5B. Similarly, with the desired significance level (0.0083) by the Bonferroni approach, none of the tests were significant.

For AIFP, group 1 and group 2 employed different AIFP across the six abnormalities. A multivariate analysis of variance (MANOVA) was conducted to see if group was a significant predictor of AIFP across the six abnormalities, doing a single analysis with the six finger pressure numbers as six dependent variables and group as a single independent variable. Using the Wilks' Lambda test statistic, the results were significant ($p = 0.0002$), indicating the effect of group on finger pressure. The MANOVA results also included individual ANOVA tests on each pressure variable separately. The p-values for individual abnormalities (aforementioned paragraph) were 0.0780, 0.1010, 0.1725, 0.0235, 0.0016, and < 0.0001, respectively, based on the assumption of equal variance for the two groups. Therefore, AIFP significantly impacted the detection of only the last three, the smallest, abnormalities (A and C in testing scenario 4 and B

in scenario 5). Greater finger pressure was utilized by group 2 on these three scenarios.

E. Impact of Technique on Detection Performance (Objective 5)

Technique was then analyzed relative to performance. First, a logistic regression model was run with all explanatory variables (GFM, LFM, AIFP, and group), (10), where p is the probability that an abnormality was successfully detected. GFM_variables are indicators for three GFMs: U, V, and L. LFM_variables are indicators for two LFMs: tapping and vibration. *Pressure* is AIFP. *Group* is a dichotomous variable that identifies the examiners: group 1 or group 2. The reference levels for GFM, LFM, and group variables are line, sliding, and group 2, respectively,

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1(\text{GFM\_U}) + \beta_2(\text{GFM\_V})$$
$$+ \beta_3(\text{GFM\_L}) + \beta_4(\text{LFM\_Tap})$$
$$+ \beta_5(\text{LFM\_Vib}) + \beta_6(\text{pressure})$$
$$+ \beta_7(\text{group}). \quad (10)$$

Table I(a) shows that both LFM and group are significant ($p < 0.0001$ and $p = 0.0037$, respectively). According to Table I(b), the effect of LFM comes from LFM_Vib, where the odds of success, using (10), increase by an estimated factor of 9.20 ($\exp(\beta_5 - \beta_4)$, where $\beta_5 = 2.2073$ and $\beta_4 = -0.0122$ when one uses vibration instead of tapping and of 9.09 [$\exp(\beta_5)$] when one uses vibration instead of sliding. The overall effect of GFM is not significant, but the coefficients suggest that the V pattern (0.75) and an L pattern (0.69) give better performance than a U pattern (0.04) or a Line pattern (0.00).

Second, each of GFM, LFM, and AIFP were analyzed separately as a predictor of performance. In analysis of GFM, a logistic regression was employed to see which defined GFM (U, V, L, and Line) leads to better performance across all six abnormalities. Therefore, the statistical model can be expressed in the following equation, where the variables are as defined earlier:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1(\text{GFM\_U}) + \beta_2(\text{GFM\_V})$$
$$+ \beta_3(\text{GFM\_L}). \quad (11)$$

The score statistic for GFM in this model has a p-value of 0.0461 (not displayed in table), so is a significant predictor when no other effects are taken into account. But as we saw earlier, its effect is not significant in the presence of other explanatory variables. Since the latter analysis had a small number of observations ($N = 34$) in relation to the seven explanatory variables, it is possible that the results were affected by overfitting. In this situation, with inconsistent results, we cannot draw a firm conclusion regarding the significance of GFM on performance. On the other hand, the parameter estimates and significance levels from (11) are consistent with the results of the overall model (10), in that the V and L patterns are superior to U and Line patterns. Among all GPMs, those who employed the V pattern performed the best. For LFM, the following equation was used

with the variables defined earlier:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1(\text{LFM\_Tap}) + \beta_2(\text{LFM\_Vib}). \quad (12)$$

The results of this regression are consistent with those from the overall regression. The score statistic for LFM in this model has p-value less than 0.0001, from Table I(b), as before. Therefore, use of the vibration pattern increases the odds of successful detection, using (12) and Table III(c), by a factor of 8.23 [$\exp(\beta_2)$] relative to use of sliding, and by a factor of 11.15 ($\exp(\beta_2 - \beta_1)$, with $\beta_2 = 2.1079$ and $\beta_1 = -0.3037$) relative to use of tapping.

For AIFP, the following equation was used with variables defined earlier:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1(\text{pressure}). \quad (13)$$

The results of this regression are also consistent with those from the overall regression. The score statistic for AIFP in this model has p-value 0.0810, which is much lower than in the overall model, but still falls short of significance. The coefficient *pressure* is positive in both regressions, indicating that greater pressure is associated with the more successful detection of abnormalities.

IV. Discussion

Overall, while technique utilization varies both between participants on the same simulated abnormality and within each participant across abnormalities, some elements of technique clearly impact performance. This assertion is supported by six main findings from this pilot experiment, which are as follows.

1) Group 2 (resident physicians) performed significantly better than group 1 (nurse practitioner students) in detecting abnormalities.
2) Finger techniques (GFM, LFM, and AIFP) were not consistently utilized by single participants across abnormalities or by all participants within an abnormality.
3) Both group 1 and group 2 employed similar GPF and LFM, but different AIFP.
4) The V and L patterns led to greater success in detecting abnormalities.
5) The vibration pattern led to greater success in detecting abnormalities.
6) Higher AIFP led to a greater success in detecting some of the abnormalities.

A. Techniques

Neither nurse practitioner students, nor resident physicians employed a consistent set of techniques. The odds of a participant consistently employing the same set of techniques over six simulated abnormalities is quite low, and so are the odds that the same techniques are being utilized by all participants over the same abnormality. The fact that the technique is not consistent suggests that the technique utilized may be affected by size, location, and stiffness of the presented abnormality, although this is not studied as part of this paper.

B. Technique and Performance

Correct detectors employed a characteristic subset of techniques. Following from their definition in (1), the relationship of the α's in (14) indicates that a participant's LFM (α_2) was a significantly better predictor of performance than GFM (α_1), which was a slightly better predictor than AIFP (α_3) (see Table I(a)):

$$\alpha_2 \gg \alpha_1 > \alpha_3. \quad (14)$$

LFM using vibration was significantly linked with improved detection ($p < 0.0001$). Further, in examining the data for a single participant, who made correct and incorrect detections, seems to indicate the use of vibration only when in the vicinity of inflated balloons, thus indicating an engaged search of a subarea of the prostate. This focused, local search is likely prompted by an initial encounter with a suspected abnormality in the global search. In contrast, a lack of initial recognition may lead to the continuance of the sliding pattern, where the examiner is never drawn out of his or her global search.

Usage of the V GFM was the best predictor of correct detections when compared with the other three patterns. The V pattern may enable a more thorough search, although these experiments used an artificial simulator and need to be confirmed in experiments with actual human prostate. Even though the U pattern covers more area, its usage does not relate to better performance over the V pattern. At this point, however, the significance of the GFM remains in some question because of the fact that only six participants employed the U pattern. While extent of being thorough is perhaps not as difficult with the prostate as in the larger volume of the breast, it may become important in search of dispersed and small tumors. Another idea is that a strategic examiner might utilize the V and L patterns to compare the two halves of the prostate, although further data are needed to support this.

The level of finger pressure appears to impact performance less than other factors. However, the use of greater AIFP is associated with the more successful detection of some abnormalities. In a contradictory findingthough, note that group 2 exerted a significantly lower average pressure than group 1 across several abnormalities. Perhaps, group 2 more clearly understands the dynamics of patient interaction, where higher finger pressure would create tension with the patient in conducting the exam. It could also be the case that the size, location, and stiffness of individual tumors may require utilizing greater finger pressure, which averaging across abnormalities can negate, but which experienced examiners selectively employ.

C. Technique and Group 1/Group 2 Performances

The resident physicians (group 2) performed at significantly higher levels than nurse practitioner students (group 1). However, while correct detectors in both groups tended to utilize LFM of vibration, GFM of V and higher finger pressure, these metrics do not fully explain the performance difference between the two groups.

According to the logistic regression in Table I(a), being in group 1 is the major factor accounting for the difference ($p =$

0.0037). This does not, however, delineates a specific element of technique, but may be due to an inherent difference between two groups, such as accumulated years of clinical experience or the nature of training in each discipline. It is to note that more of group 1 (11 of 18 or 61.1%) than group 2 (6 of 16 or 37.5%) had practiced with silicone models. That practice would have come from silicone models other than the VPES. Silicone models apart from the VPES are not realistic enough, do not generate enough practice cases, or do not provide the right type of feedback [6]. Therefore, accumulated clinical experience, disciplinary training, and practice background (α_4), together with cognitive objective recognition or another unknown factor (β) may contribute to our current inability to entirely account for performance differences between groups.

V. STUDY LIMITATIONS

The study methodology has a few limitations. First, the notion of an "experienced examiner" was applied to residents, who certainly have more clinical experience than students. However, it is more likely that there are levels of expertise following from student to resident, and ultimately, attending urologist. A future study should be extended to the latter group. Second, the sample size (as noted in Section II) was small, which means that it is possible that the regression models overfit the data. Third, the techniques defined herein link back to accepted palpation strategies of the clinical literature, but other techniques—currently unknown—might more adequately describe to what clinicians are doing. For example, clinicians may conduct an initial, overview search to find gross irregularities before focusing in on a GFM that will lead to a thorough subsequent search.

ACKNOWLEDGMENT

REFERENCES

[1] American Cancer Society, "Screening and Early Detection of Prostate Cancer," Atlanta, GA: American Cancer Society, 2008.

[2] R. Etzioni, D. F. Penson, J. M. Legler, D. di Tommaso, R. Boer, P. H. Gann, and E. J. Feuer, "Overdiagnosis due to prostate-specific antigen screening: Lessons from US prostate cancer incidence trends," *J. Nat. Cancer Inst.*, vol. 94, pp. 981–990, 2002.

[3] S. Weinmann, K. E. Richert-Boe, S. K. Van Den Eeden, S. M. Enger, B. A. Rybicki, J. A. Shapiro, and N. S. Weiss, "Screening by prostate-specific antigen and digital rectal examination in relation to prostate cancer mortality—A case-control study," *Epidemiology*, vol. 16, pp. 367–376, 2005.

[4] D. S. Smith and W. J. Catalona, "Interexaminer variability of digital rectal examination in detecting prostate-cancer," *Urology*, vol. 45, pp. 70–74, 1995.

[5] C. G. Roehrborn, S. Sech, J. Montoya, T. Rhodes, and C. J. Girman, "Interexaminer reliability and validity of a three-dimensional model to assess prostate volume by digital rectal examination," *Urology*, vol. 57, pp. 1087–1092, 2001.

[6] G. J. Gerling, S. Rigsbee, R. M. Childress, and M. L. Martin, "The design and evaluation of a computerized and physical simulator for training clinical prostate exams," *IEEE Trans. Syst., Man, Cybern. A: Syst., Humans*, vol. 39, no. 2, pp. 388–403, Mar. 2009.

[7] A. K. Madan, S. Aliabadi-Wahle, A. M. Babbo, M. Posner, and D. J. Beech, "Education of medical students in clinical breast examination during surgical clerkship," *Amer. J. Surg.*, vol. 184, pp. 637–640, 2002.

[8] H. S. Campbell, M. McBean, H. Mandin, and H. Bryant, "Teaching medical-students how to perform a clinical breast examination," *Acad. Med.*, vol. 69, pp. 993–995, 1994.

[9] D. Saslow, J. Hannan, J. Osuch, M. H. Alciati, C. Baines, M. Barton, J. K. Bobo, C. Coleman, M. Dolan, G. Gaumer, D. Kopans, S. Kutner, D. S. Lane, H. Lawson, H. Meissner, C. Moorman, H. Pennypacker, P. Pierce, E. Sciandra, R. Smith, and R. Coates, "Clinical breast examination: Practical recommendations for optimizing performance and reporting," *CA Cancer J. Clin.*, vol. 54, pp. 327–344, 2004.

[10] K. J. Saunders, C. A. Pilgrim, and H. S. Pennypacker, "Increased proficiency of search in breast self-examination," *Cancer*, vol. 58, pp. 2531–2537, 1986.

[11] H. Pennypacker and M. Iwata, "MammaCare: A case history in behavioural medicine," in *Behaviour Analysis in Theory and Practice*, D. Blackman and H. Lejeune, Eds. Hillsdale, NJ: Lawrence Erlbaum Assoc, 1991, pp. 259–288.

[12] M. M. McDermott, N. C. Dolan, A. Rademacher, and C. Pilgrim, "Effect of breast-tissue characteristics on the outcome of clinical breast examination training," *Acad. Med.*, vol. 71, pp. 505–507, 1996.

[13] C. Pilgrim, C. Lannon, P. R. Harris, W. Cogburn, and S. W. Fletcher, "Improving clinical breast examination training in a medical school," *J. General Intern. Med.*, vol. 8, pp. 685–688, 1993.

[14] 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. 408–414, 1980.

[15] H. S. Pennypacker and C. A. Pilgrim, "Achieving competence in clinical breast examination," *Nurse Pract. Forum*, vol. 4, pp. 85–90, 1993.

[16] S. W. Fletcher, M. S. O'Malley, J. L. Earp, T. M. Morgan, S. Lin, and D. Degnan, "How best to teach women breast self-examination," *Ann. Intern. Med.*, vol. 120, pp. 772–779, 1990.

[17] M. E. Murali and K. Crabtree, "Comparison of two breast self-examination palpation techniques," *Cancer Nurs.*, vol. 15, pp. 276–282, 1992.

[18] D. N. Rutledge, "Sensitivity and specificity of lump detection in breast models," *Amer. J. Prev. Med.*, vol. 8, pp. 314–318, 1992.

[19] E. Atkins, L. J. Solomon, J. K. Worden, and R. S. Foster, "Relative effectiveness of methods of breast self-examination," *J. Behav. Med.*, vol. 14, pp. 357–367, 1991.

[20] M. A. Trapp, T. E. Kottke, R. A. Vierkant, J. S. Kaur, and T. A. Sellers, "The ability of trained nurses to detect lumps in a test set of silicone breast models," *Cancer*, vol. 86, pp. 1750–1756, 1999.

[21] V. A. Clarke and S. A. Savage, "Breast self-examination training: A brief review," *Cancer Nurs.*, vol. 22, pp. 320–326, 1999.

[22] W. I. Goodson, "Clinical breast examination," *West J. Med.*, vol. 164, pp. 355–358, 1996.

[23] B. Panchaphongsaphak, R. Burgkart, and R. Riener, "Three-dimensional touch interface for medical education," *IEEE Trans. Inf. Technol. Biomed.*, vol. 11, no. 3, pp. 251–263, May 2007.

[24] A. M. Tahmasebi, K. Hashtrudi-Zaad, D. Thompson, and P. Abolmaesumi, "A framework for the design of a novel haptic-based medical training simulator," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 5, pp. 658–666, Sep. 2008.

[25] C. M. Pugh and P. Youngblood, "Development and validation of assessment measures for a newly developed physical examination simulator," *J. Amer. Med. Inf. Assoc.*, vol. 9, pp. 448–460, 2002.

[26] S. Bamberg, A. Y. Benbasat, D. M. Scarborough, D. E. Krebs, and J. A. Paradiso, "Gait analysis using a shoe-integrated wireless sensor system," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 4, pp. 413–423, Jul. 2008.

[27] J. Parkka, M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Peltola, and I. Korhonen, "Activity classification using realistic data from wearable sensors," *IEEE Trans. Inf. Technol. Biomed.*, vol. 10, no. 1, pp. 119–128, Jan. 2006.

[28] K. C. Lee, D. Dunlop, and N. C. Dolan, "Do clinical breast examination skills improve during medical school?," *Acad. Med.*, vol. 73, pp. 1013–1019, 1998.

[29] N. Wang, G. J. Gerling, R. M. Childress, and M. L. Martin, "Characterizing finger palpation in the detection of prostate cancers and abnormalities," in *Proc. Hum. Factors Ergon. Soc., Annu. Meeting*, 2008, pp. 813–817.

Ninghuan Wang was born in Shanghai, China. She received the Bachelor's degree in software engineering from Shanghai Jiaotong University, Shanghai, China, and the M.S. degree from the Department of Systems and Information Engineering, University of Virginia, Charlottesville.

She is currently working full-time as a Software Development Engineer in Test (SDET) in Microsoft Corporation, Fargo, ND. Her research interests include data/pattern analysis and performance prediction by using medical simulators for training and improving doctors/nurses in their clinical examination.

Reba Moyer Childress received the B.S. and M.S. degrees in nursing from the School of Nursing, University of Virginia, Charlottesville, in 1979 and 1992, respectively, where she completed the Family Nurse Practitioner certificate program in 1991.

She is an Assistant Professor of nursing with the School of Nursing, University of Virginia, where she is the Director of Clinical Simulation Learning Center.

Gregory J. Gerling (S'03–M'05) received the Ph.D. degree from the Department of Mechanical and Industrial Engineering, The University of Iowa, Iowa City.

He has been with the Department of Systems and Information Engineering, University of Virginia, Charlottesville, since fall of 2005. Before returning to graduate school, he had industry experience in software engineering at Motorola, NASA Ames Research Center, and Rockwell Collins. His major research interests include haptics, human factors/ergonomics, computational neuroscience, biomechanics, and human–computer interaction. The application of his research seeks to advance neural prosthetics/robotics, aid people whose sense of touch is deteriorating, and improve human–robot interfaces, particularly in medicine.

Marcus L. Martin was born in Covington, VA. He received the B.S. degree in pulp and paper technology and the B.S. degree in chemical engineering from the North Carolina State University, Raleigh, in 1970 and 1971, respectively, and the M.D. degree from the Eastern Virginia Medical School, Norfolk, in 1976, where he was a member of the charter class and the first African American graduate. He completed his emergency medicine residency training at the University of Cincinnati, Cincinnati, OH, in 1981.

He is currently with the University of Virginia (U.Va), Charlottesville, where he is a Professor and the immediate past Chair in the Department of Emergency Medicine, the Assistant Dean of the School of Medicine, and the Associate Vice President for Diversity and Equity. He established the Emergency Medicine Center for Education, Research and Technology and the Life Saving Technique course for medical students at U.Va using computerized human patient simulation. He is working on several research projects with systems engineers, developing computerized patient simulators.

Using a Prostate Exam Simulator to Decipher Palpation Techniques that Facilitate the Detection of Abnormalities Near Clinical Limits

Ninghuan Wang, MS;

Gregory J. Gerling, PhD;

Tracey L. Krupski, MD, MPH;

Reba Moyer Childress, MSN, FNP-BC;

Marcus L. Martin, MD

Introduction: Prostate carcinoma (and other prostate irregularities and abnormalities) is detected in part via the digital rectal examination. Training clinicians to use particular palpation techniques may be one way to improve the rates of detection.
Methods: In an experiment of 34 participants with clinical backgrounds, we used a custom-built simulator to determine whether certain finger palpation techniques improved one's ability to detect abnormalities smaller in size and dispersed as multiples over a volume. The intent was to test abnormality cases of clinical relevance near the limits of size perceptibility (ie, 5-mm diameter). The simulator can present abnormalities in various configurations and record finger movement. To characterize finger movement, four palpation techniques were quantitatively defined (global finger movement, local finger movement, average intentional finger pressure, and dominant intentional finger frequency) to represent the qualitative definitions of other researchers.
Results: Participants who used more thorough patterns of global finger movement (V and L) ensured that the entire prostate was searched and detected more abnormalities. A higher magnitude of finger pressure was associated with the detection of smaller abnormalities. The local finger movement of firm pressure with varying intensities was most indicative of success and was required to identify the smallest (5-mm diameter) abnormality. When participants used firm pressure with varying intensities, their dominant intentional finger frequency was about 6 Hz.
Conclusions: The use of certain palpation techniques does enable the detection of smaller and more numerous abnormalities, and we seek to abstract these techniques into a systematic protocol for use in the clinic.
(*Sim Healthcare* 5:152–160, 2010)

Key Words: Haptics, Simulation, Nursing, Medical, Human factors, Assessment, Palpation, Clinical prostate examination.

Prostate cancer is one of the most common causes of death in American men, with an estimated 189,280 new cases during 2009.[1] Although the digital rectal examination (DRE) complements the prostate-specific antigen blood test as a valuable screening examination,[2] its variable positive predictive value (17%–34%[1]) and low agreement on diagnosis between examiners (21%–40%[3,4]) seems to indicate a need for enhanced clinical examination training. Concurrently, hands-on training is being increasingly emphasized and conducted in medical and nursing schools via standardized patients or physical simulators[5,6] and has a positive effect on clinical performance.[7] In addition to the training devices and protocols being designed as training aids, we also need to better understand how the employment of particular techniques impact performance.

Hands-on techniques have been studied for the clinical breast examination and include global finger movement (GFM) for the search of abnormalities over the entire breast, local finger movement (LFM) for the palpation of small areas of the breast, and levels of finger pressure for adequate search through the tissue depth.[8–10] A range of studies show that certain techniques increase clinical breast examination effectiveness by increasing the mean lump detection rates at a statistically significant level,[11,12] sometimes doubled from pre- to post-test.[8,13] Very few studies have assessed specific hands-on techniques relative to the DRE. However, Balkissoon et al[14] found a DRE simulator useful in determining that students as opposed to experienced clinicians used different palpation techniques.

Palpation techniques have begun to be quantitatively defined by building on qualitative definitions. With respect to the female pelvic examination, the E-Pelvis simulator was used to track the motion of the finger over force sensors embedded in silicone. Three characteristic palpation practices emerged (number of critical areas touched, frequency at which areas were touched, and maximum pressure exerted), which correlated with ability to diagnose simulated find-

Figure 1. Main components of Virginia Prostate Examination Simulator apparatus, including (A) torso, (B) laptop, (C) automatic balloon inflation, (D) instrumented prostate, and (E) internal track system.

ings.[15] In addition, our group has sought to characterize palpation techniques for the clinical prostate examination using a custom-built simulator that can present abnormalities in various configurations and record finger movement.[16] That work quantified three palpation techniques [GFM, LFM, and average intentional finger pressure (AIFP)] based on the qualitative, clinical definitions. A human-subjects study revealed that one's pattern of LFM (firm pressure with varying intensity, in particular) strongly impacted his ability to detect abnormalities, and that more thorough patterns of GFM led to greater success. Differences between novices and experts were also evaluated.

This work investigates more specifically whether the use of certain finger palpation techniques improves one's ability to detect abnormalities smaller in size (5, 10, and 20 mm) and dispersed as multiples (one vs. two) over a volume. Abnormality size and examiner thoroughness are factors particularly relevant to clinical detection. Furthermore, we used test cases that link both with clinical findings and limits of human tactile detection. Because of the notable impact that the LFM of firm pressure with varying intensity played previously on abnormality detection, this work introduces a fourth palpation technique, dominant intentional finger frequency (DIFF), which analyzes the varying intensity.

METHODS

A human-subjects experiment was conducted with 34 participants using the Virginia Prostate Examination Simulator (VPES). Postexperiment analysis was conducted to evaluate four objectives: (1) whether those participants who detect more than one abnormality in a prostate apply a more thorough GFM (U pattern that covers four sensors compared with a V or L pattern that covers three sensors and Line pattern that covers only two sensors), (2) whether the LFM pattern of firm pressure with varying intensity is required to detect the smallest (5-mm diameter) of abnormalities, (3) whether greater AIFP will be required to detect relatively smaller abnormalities (~20, 10, and 5 mm), and (4) whether a higher DIFF will be required to detect relatively smaller abnormalities.

Apparatus: VPES

The VPES (Figs. 1A–C) uses rubber-like materials to simulate the feel of tissue, a computer, valves, and a pump to create test scenarios by inflating with water, small embedded balloons, and pressure sensors to record finger pressure.[5] Three instrumented prostates (Fig. 1D) with dimension of 55 mm (transverse, width dimension) by 50 mm (longitudinal, length dimension) are attached to a track system (Fig. 1E) internal to a posterior torso. The stiffness of the instrumented prostates, evaluated via compression tests, yielded an elastic modulus of ~55 kPa.

Each prostate was embedded with four to six polyethylene balloons. Balloons are filled with water to simulate palpable abnormalities, whereas deflated balloons are not palpable. With this feature, both normal and abnormal prostate conditions including prostatitis (enlarged and boggy inflammation) and carcinoma (small and firm isolated nodule) can be simulated by varying the configurations of balloons in size and location. Balloons of different sizes are positioned in unique locations for each instrumented prostate. Four spherical balloons (5, 7.5, 10, and 15 mm diameter) were used along with one elongated (E) balloon (30-mm length × 15-mm width × 10-mm height).

Using water pressure and force sensors, the VPES captures the finger pressure used on both inflated balloons and the entire prostate. The finger pressure is recorded for postexam analysis. Water pressure sensors (SenSym Pressure Sensor, Model SX100DD4, Honeywell, Golden Valley, MN) on each balloon are sampled at a rate of three samples per second. Force sensors (Tekscan, South Boston, MA, Flexiforce 0–1 lb range) embedded in each instrumented prostate were used to record the location and magnitude of applied finger pressure during palpation at 1000 samples per second over the ~30 seconds time interval of the examination.

Example Data Collection

Force sensor data were used to characterize finger movements, whereas the balloon sensor data were used to determine the interval over which local palpation techniques were analyzed (ie, LFM, AIFP, and DIFF). Example data from assessment scenario 3 are plotted in Figure 2. Spikes in the balloon pressure data indicate if a simulated abnormality was palpated. In this example, the 5-mm balloon went untouched over the 30-second period, whereas the 10-mm balloon was palpated multiple times (upward spike; circle "A" in Fig. 2). Moreover, when this balloon was palpated, the magnitude of finger pressure for nearby sensors 1 and 4 increases while that applied to sensor 2 decreases (circle "B").

Figure 2. Example plot of force (lower four lines) and balloon (upper two lines) sensor data for assessment scenario 3. Force data have been smoothed to improve its display for this example.
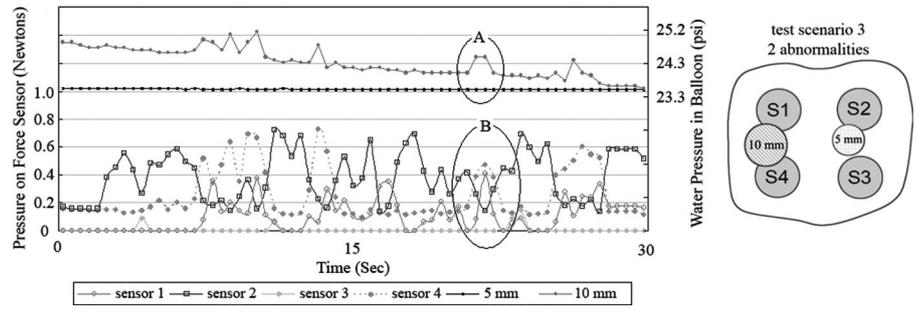
High-Level Algorithms

DIFF is introduced in this work. The other three palpation techniques—GFM (U, V, L, and Line), LFM (repetitive tapping, firm pressure with varying intensity, and sliding), and AIFP— have been presented in greater detail elsewhere[16] and therefore are detailed here only at a high level.

Global Finger Movement

GFM is defined as the systematic movement of one's finger over the entire prostate in search of abnormalities. Four patterns are defined as U, V, L, and Line (Fig. 3) with deviations in orientation recognized.

The algorithm to identify the GFM follows two main steps. First, the initial pressure on the kth sensor data at time t (S_t^k) is discretized (bit$_1^k$) into binary values (0 or 1) every 16 milliseconds for each sensor to create a state for the group of four sensors (eg, 0000, 0110, or 1111) via Eq. (1). The $g\_threshold$ is set at 0.05 N to indicate intentional pressure exerted by an examiner's finger.

$$\text{bit}_t^k = 1 \text{ if } S_t^k > g\_threshold, \text{ otherwise bit}_t^k = 0; \quad (1)$$

The resultant bit values for sensors 1 to 4 are used to identify a sensor state at time t (state$_t$). The states range from 0000 to 1111. The leftmost bit stands for sensor 1 and the rightmost bit for sensor 4. A bit is set as 1 when a sensor is triggered and 0 otherwise. For example, a state defined as 0111 specifies a point in time whereby sensors 2, 3, and 4 are triggered simultaneously, while sensor 1 is untouched.

As noted in the example in Figure 2, the identified GFM by the analysis tool (written in the C# programming language) is displayed (Fig. 4A), where the "L" GFM was identified and is shown with lines and arrows. Although a participant may use more than one GFM pattern during an examination, the pattern classified is the single one used for the longest duration.

Local Finger Movement

LFM is defined as palpation by finger movement within a single quadrant of the instrumented prostate or near a single abnormality. An LFM is typically used in the palpation of small areas to isolate suspected abnormalities. The three patterns are defined as repetitive tapping, firm pressure with

varying intensity, and sliding. These are presented below and shown with their corresponding parameters in Figure 5.

The LFM pattern of repetitive tapping (Fig. 5A) is defined as an examiner striking briefly on a small area in an intermittent fashion. With the pattern of firm pressure with varying intensity (Fig. 5B), an examiner maintains his or her finger contact with a single area but shifts the weight of the finger (eg, proximally distally or laterally medially). A sliding pattern (Fig. 5C) is identified when an examiner moves the finger smoothly over a defined area without stopping. Often examiners locally isolate an abnormality, with a repetitive tapping or firm pressure with varying intensity, after their global search hints at a suspected area. In contrast, the sliding pattern likely represents the continuous movement of one's finger across a sensor (ie, low-magnitude spikes observed when the finger is far from the sensor's center and high-magnitude spikes when the finger passes over the sensor's center). The sliding pattern may represent failure to notice an abnormality in one's global search. Figure 4B indicates the identification of the pattern of firm pressure with varying intensity by the analysis tool on sensor 4 lasting for 2.24 of the 3.50 seconds analysis period over which balloon C was palpated.

Average Intentional Finger Pressure

AIFP is defined as the pressure applied on each of the four force sensors over the duration of the examination. Force greater than the $p\_threshold$ (0.05 N) is averaged for each sensor (k) given the analysis period (T), Eq. (2).

$$S_{\text{avg}}^k = \frac{1}{T}\sum_{t=1}^{T} S_t^k, \text{ subject to } S_t^k > p\_threshold \quad (2)$$

As shown in Figure 4C, the AIFP for the example was computed over the entire examination for sensors 1 to 4 as 0.45, 1.21, 0.27, and 1.06 N, respectively.

Dominant Intentional Finger Frequency

DIFF is used in the palpation of a simulated abnormality. This analysis is done when a participant's LFM has been defined as firm pressure with varying intensity. DIFFs were
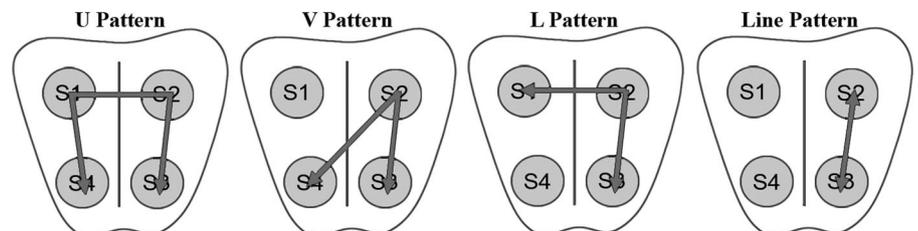
Figure 3. Defined patterns of global finger movement.
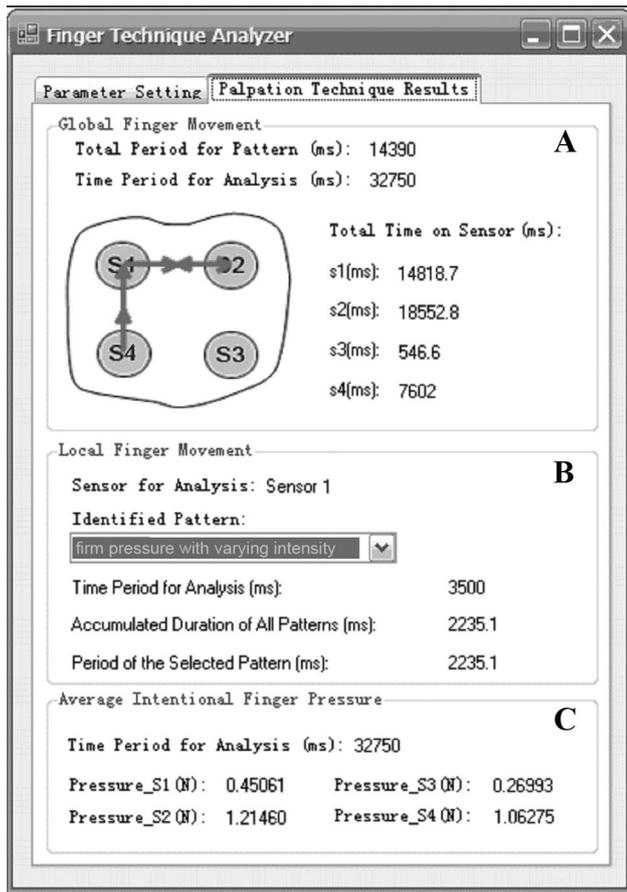
Figure 4. Palpation technique analyzer with example results for (A) global finger movement, (B) local finger movement, and (C) average intentional finger pressure.

identified using the Fourier transform (MATLAB, version 6.5). In Eq. (3), force data from each of the sensors over the time span of an intentional palpation (S_t^k) are converted to a discrete Fourier transform of the sensor signal [transform(k)] for each of the k force sensors. The variable N represents the total number of discrete force values collected by the particular force sensor ($k = 1, 2, 3,$ and 4) under analysis.

$$\text{transform}(k) = \sum_{i=1}^{N} s_t^k e^{\frac{-2\pi i}{N} \times (j=1)} \qquad (3)$$

Once the data has been converted to a discrete Fourier transform [transform(k)], the power at each of the various frequencies is obtained to give the power spectrum [power(k)], Eq. (4).

$$\text{power}(k) = e^{\text{transform}(k) \times \text{conj (transform}(k))/N} \qquad (4)$$

An analyzed frequency range of 5 to 32 Hz was setup, with the nth frequency (freq$_n$) in that range calculated via Eq. (5). The lowest (lf) and highest (hf) finger frequencies are set to 5 and 32 Hz. This particular frequency range has physiological implications, because slowly and rapidly adapting mechanoreceptors in fingertip skin are most sensitive to vibration in the 5 to 30 Hz[17] range. The sampling frequency (sf) of 64 Hz was double the Nyquist frequency (hf).

$$\text{freq}_n = \text{sf} \times (j-1)/N \ \& \ \text{freq}_n \ \epsilon \ [\text{lf, hf}],$$
$$n = 1, 2 \ldots (N+1) \qquad (5)$$

Finally, the dominant frequency (DIFF) is identified as the frequency of greatest power where Y is the vector of power(k), Eq. (6).

$$S^k\_\text{DIFF} = \text{freq}_n \ \& \ n = \text{position} (Y, \max(\text{power}(k))) \qquad (6)$$

In an example analysis of DIFF, a set of data from force sensor 4 and a balloon sensor are shown in Figure 6A. From that dataset, a subsection of the force sensor data, over a period of intentional balloon palpation, were isolated in Figure 6B. Then, from this dataset, a series of finger frequencies were extracted (Fig. 6C). The frequency (5.59 Hz) with the largest power number (3.41) is identified as the DIFF.

Experiment
Participants

Thirty-four participants, including 18 nurse practitioner students and 16 resident physicians, were enrolled in the human-subjects experiment that was approved by the Institutional Review Board at the University of Virginia. There were 10 men and 24 women, ranging in age from 23 to 47 (mean = 31.47, SD = 6.79) years. Sixteen participants had performed at least 10 DREs in the clinic, another 5 had performed at least
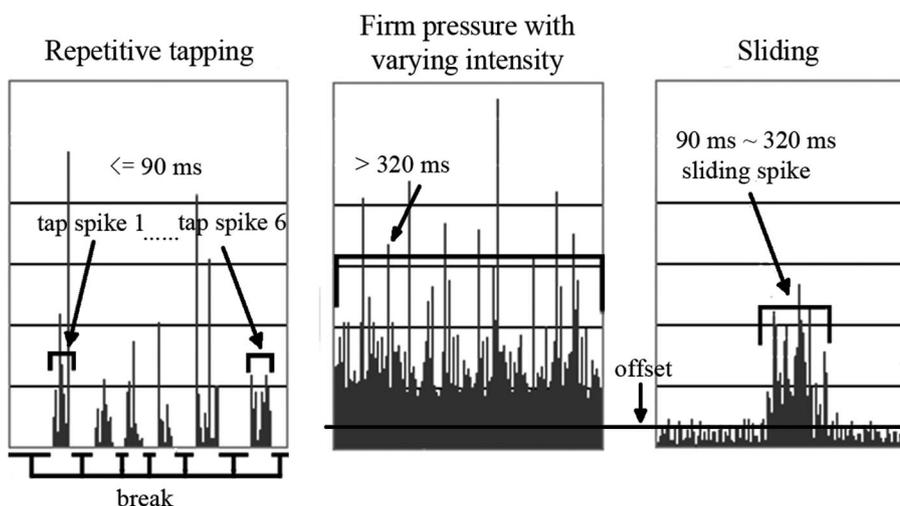
Figure 5. Three patterns of local finger movement. Identifying characteristics are overviewed in each panel.

© 2010 Society for Simulation in Healthcare **155**
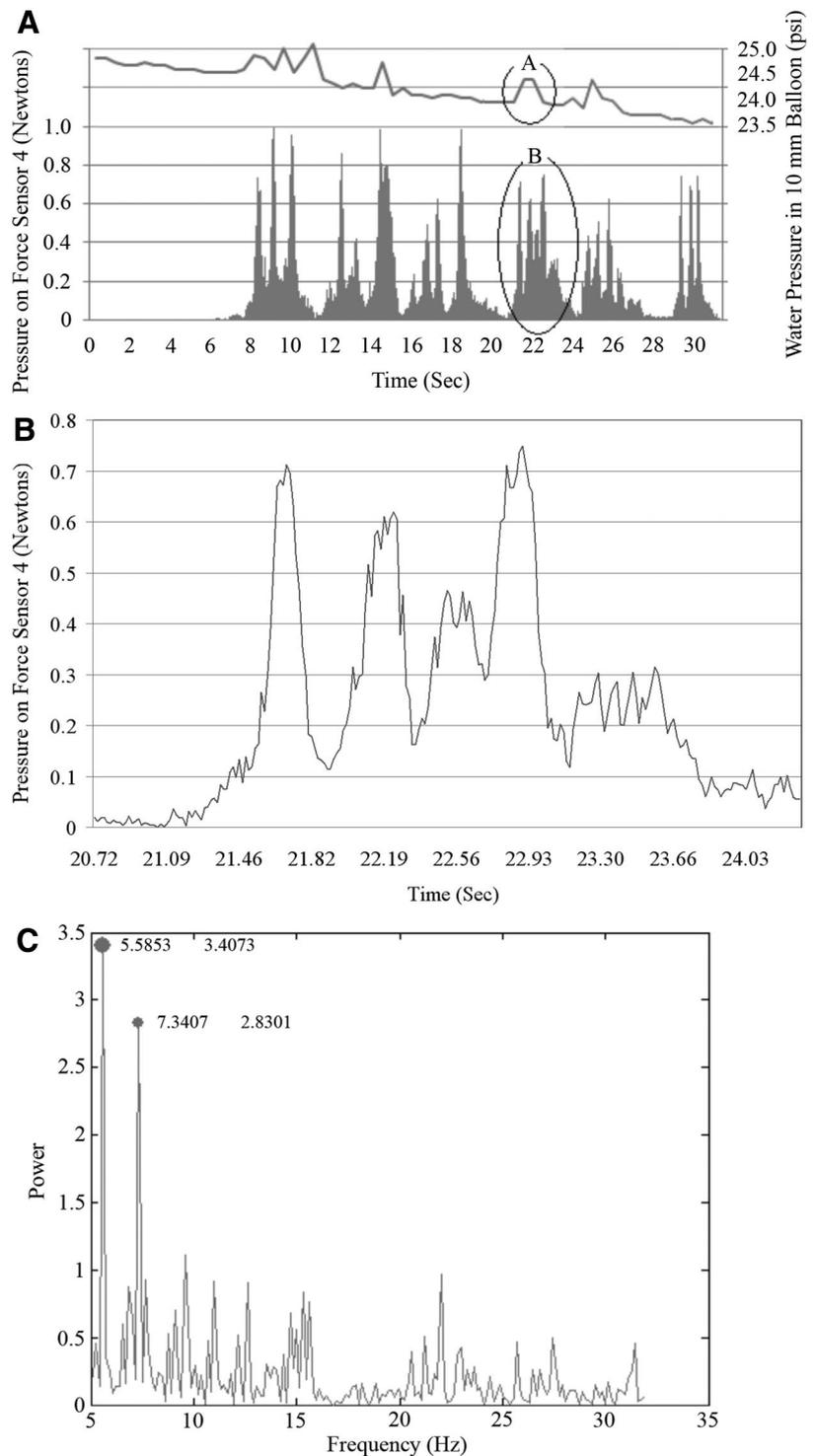
Figure 6. A, Finger pressure upon force sensor 4 (below spikes) and balloon sensor for the 10-mm abnormality (above line) in assessment scenario 3. Circle "B" marks the section of force data, per the corresponding intentional palpation of circle "A", which is analyzed further part B. B, Force sensor 4 data from circle "B" in part A. C, Two prominent frequencies are identified for force sensor 4 in frequency range 5 to 32 Hz, with 5.58 Hz being the DIFF.

one, and 13 had performed none. Seventeen participants had practiced with silicone models on a prior occasion.

Procedure

Participants performed three tasks over 40 minutes: (1) the experience questionnaire, (2) the 5-minute hands-on orientation, and (3) the 20-minute hands-on assessment session. Two proctors jointly facilitated this process. The first proctor facilitated the paperwork, hands-on orientation, and assessment sessions and recorded the participant's verbal report of any detected abnormalities (on a standardized score sheet in which data were the size, location, and hardness of an abnormality). The second proctor configured the scenarios and controlled the order of the scenarios presented to each participant, because the first proctor was blind to the scenario of abnormalities being presented. The proctors had been trained in a preliminary experiment. During the orientation and assessment sessions, a total of seven scenarios in either normal or abnormal (prostatitis or carcinoma) states were simulated using three instrumented prostates. Orientation scenario 1 simulated a normal prostate, orientation scenario 2 filled a large, elongated balloon covering sensors 2 and 3, and orientation scenario 3 filled a 15-mm balloon covering
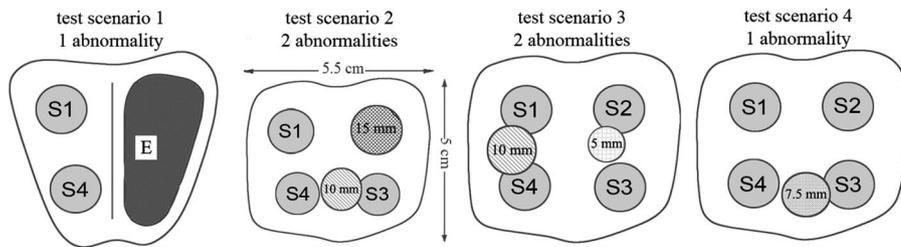
Figure 7. Four assessment scenarios. Balloon E dimensions are 3.0-mm length × 1.5-mm width × 10-mm height.

sensor 2. Immediately after each orientation scenario, the proctor pointed out the simulated abnormalities on a printed figure. No instructions on techniques to use were provided during the orientation nor were participants allowed to view the simulator's immediate feedback display. The orientation scenarios were not included in the analysis.

Assessment Session and Scenarios

Assessment scenarios 1 to 4 are shown in Figure 7 and are presented in a randomized order. Over the 20-minute assessment session, each participant was given 30 seconds, which is an appropriate duration for the clinical examination, to palpate each of four scenarios and was instructed to report any abnormalities found over the entire prostate. After palpating within each scenario presented, the participant identified any abnormalities to the proctor, indicating their approximate size and location. To get at the limits of clinical detectability, we used 5-mm balloons as the smallest size. A previous psychophysical experiment had shown that this size was detected at reasonable perceptual limits, being detected 55% of the time at a 10-mm depth and 95% of time at the 5-mm depth.[18] In contrast, 4-mm balloons at the 5-mm depth were detected <75% of the time.

Dependent and Independent Variables

The dependent variables were based on performance, ie, the correct or incorrect detection of a simulated abnormality. The four independent variables were the quantified palpation techniques, ie, GFM, LFM, AIFP, and DIFF.

Analysis Methods

Logistic regression, t tests, and confidence intervals were used in the analysis of the four objectives. Statistical Analysis System, version 9.1, software was used. The data were grouped into three datasets: dataset 1, overall regression model (independent variable impact on performance); dataset 2, objective 1 (abnormality number and GFM); and dataset 3, objectives 2 to 4 (abnormality size and LFM, AIFP, and DIFF). Dataset 1 contained 198 observations (34 participants × all 6 abnormalities). Note that the independent variable DIFF was not included in the regression model, because it is only used when analyzing a particular LFM (firm pressure with varying intensity) and therefore is not assigned a value for other LFMs. Dataset 2 comprised 68 observations (34 participants × 2 assessment scenarios; ie, scenarios 2 and 3). Because only GFM was analyzed, we considered just the two scenarios with more than one abnormality. Dataset 3 comprised 102 observations (34 participants × 3 abnormalities) using the following three abnormalities: E in assessment scenario 2, 15 mm in assessment scenario 2, and 5 mm in assessment scenario 3. Because the three abnormalities were

located in the right side of the prostate, the impact of lateral location was minimized.

RESULTS

The data were analyzed in accordance with the four study objectives. Results related to the overall logistic regression model and the four study objectives are detailed in the subsections.

Overall Logistic Regression Model

The logistic regression model [Eq. (7)] was used to determine those independent variables that most significantly impact performance.

$$\log\left(\frac{p}{1-p}\right) = \beta_o + \beta_1(Global\_U) + \beta_2(Global\_V)$$
$$+ \beta_3(Global\_L) + \beta_4(Local\_Tapping)$$
$$+ \beta_5(Local\_Vibration) + \beta_6(AIFP) \quad (7)$$

Among the variables, P represents the probability that the abnormality was detected. Global_variables are indicators for the three GFM patterns (U, V, and L) with Line as the reference level. Local_variables are indicators for the two LFM patterns (repetitive tapping and firm pressure with varying intensity) with sliding as the reference level. AIFP is a continuous variable.

The results indicate that LFM was significantly more important in abnormality detection than GFM or AIFP (Table 1).

Objective 1: Do Those Participants Who Detect More Than One Abnormality in a Prostate Apply a More Thorough GFM Pattern?

Participants who used a more thorough GFM pattern (ie, V or L that covers three sensors) performed better in detecting both simulated abnormalities than those using the Line pattern (ie, that cover only two sensors). In assessment scenario 2, among participants who detected both abnormalities, 4 of 5 (80.00%) used either a V or L pattern; among participants who detected just one abnormality, 10 of 15 (66.67%) used a V or L pattern (Table 2). In contrast, among participants who failed to find an abnormality, 10 of 14 (71.43%) had used the Line pattern. The same trend held in

Table 1. Score Statistics for Each Factor in the Overall Model

Source	df	χ^2	P
Global	3	3.56	0.3134
Local	2	19.98	<0.0001
Pressure	1	0.76	0.3841

157

Table 2. Number of Correct and Incorrect Detections per Pattern of Global Finger Movement

| | No. Detectors | | | |
| | Correct | | | |
GFM Pattern Utilized	Both	One	Incorrect	Sum
Test scenario 2 abnormalities				
U	0	1	0	1
V	3	8	3	14
L	1	2	1	4
Line	1	4	10	15
Sum	5	15	14	34
Test scenario 3 abnormalities				
U	0	1	0	1
V	2	3	4	9
L	1	6	4	11
Line	0	6	7	13
Sum	3	16	15	34

assessment scenario 3. The V and L patterns were most prominently used by those who detected both abnormalities (100%). The Line pattern again yielded worse performance, with no correct detections of both abnormalities and 7 of 15 incorrect detectors of abnormalities having used this pattern. Utilization of the U pattern, whereby all four force sensors were touched, was identified only twice.

Objective 2: Is the LFM Pattern of Firm Pressure With Varying Intensity Required to Detect the Smallest Abnormalities?

Firm pressure with varying intensity was the only pattern of LFM that led to the detection of the smallest (5 mm) abnormality, ie, 7 of 7 (100%) in Table 3. In addition, this pattern was used consistently across the other sizes of abnormalities, ie, 22 of 25 (88%) for the E balloon and 9 of 12 (75%) for the 15-mm balloon.

Objective 3: Is a Greater Magnitude of Finger Pressure (AIFP) Required to Detect Relatively Smaller Abnormalities?

Correct detectors of abnormalities used higher AIFP as the size of the abnormality became smaller, but the difference between abnormalities was not statistically significant. The mean values of AIFP applied by correct detectors increased 1.91 N (SD = 0.98), 2.02 N (SD = 0.68), and 2.51 N (SD = 1.36) as abnormality size increased for E, 15, and 5 mm. However, when AIFP was compared, the differences were not statistically significant: $P = 0.6975$ [$t(29.26) = -0.39$] between the E and 15 mm abnormalities, $P = 0.4359$ [$t(7.70) = -0.82$] between the 15 and 5

mm abnormalities, and $P = 0.3404$ [$t(7.61) = -1.02$] between the E and 5 mm abnormalities.

Objective 4: Is a Higher Finger Frequency (DIFF) Required to Detect Relatively Smaller Abnormalities?

Participants did not increase their finger frequency when localizing small-sized abnormalities, but we do find that participants tend to use frequencies near 6 Hz consistently. The mean DIFF values applied by the 38 correct detectors on abnormalities E size, 15 and 5 mm were 5.78 (SD = 0.53), 5.76 (SD = 0.68), and 6.16 (SD = 0.91) Hz, respectively (Table 4). Although these values are not significantly different, the DIFF range does increase as the size of the simulated abnormality decreases. DIFF was also evaluated between correct and incorrect detectors of abnormalities. The mean DIFF values used by the 12 incorrect detectors on abnormalities E size, 15 and 5 mm were 6.51 (SD = 1.00), 6.32, and 6.82 (SD = 3.29) Hz, respectively. The DIFFs were nearly equal between correct and incorrect detectors, although slightly higher and more variable for incorrect detectors.

DISCUSSION

This study sought to determine whether the use of certain finger palpation techniques improves one's ability to detect abnormalities smaller in size and dispersed as multiples over a volume. Abnormality size and examiner thoroughness are factors particularly relevant to clinical detection. The results show that utilization of the LFM pattern of firm pressure with varying intensity seems to be required to detect the smallest abnormalities (5-mm diameter) and that the dominant palpation frequency associated was approximately 6 Hz. In addition, those who used the V or L GFM patterns, where more area is palpated, found a greater number of abnormalities in a single prostate. Comparing the utility of the three palpation techniques, the logistic regression model indicated that LFM was a significantly better predictor of performance than GFM or AIFP.

Those who used the LFM pattern of firm pressure with varying intensities detected more abnormalities of all sizes, especially the smallest size. The dominant frequency used was approximately 6 Hz across participants, consistently at or lower than 10 Hz, and increased only slightly as the abnormality size decreased. This frequency lies in a range that elicits a response from both slowly adapting (Merkel cells) and rapidly adapting (Meissner's corpuscles) receptors in fingertip skin.[17] These receptors assist in resolving spatial differences, which may enhance one's spatial detection and discrimination. In addition, because skin mechanoreceptors are about

Table 3. Number of Correct (and Incorrect) Detections per Pattern of Local Finger Movement

| | LFM Pattern Used | | | | |
	Repetitive Tapping	Firm Pressure With Varying Intensity	Sliding	None Detected	No. Correct (Incorrect) Detectors
Test scenario 1, E	0 (3)	22 (4)	3 (2)	0 (0)	25 (9)
Test scenario 2, 15 mm	1 (18)	9 (1)	2 (3)	0 (0)	12 (22)
Test scenario 3, 5 mm	0 (9)	7 (7)	0 (10)	0 (1)	7 (27)

Table 4. Mean DIFF for Correct (and Incorrect) Detectors

	No. Participants	Mean (Hz)	SD	95% Confidence Interval	Min (Hz)	Max (Hz)
Correct detector						
Test scenario 1, E	22	5.78	0.53	5.57–5.99	5.26	6.90
Test scenario 2, 15 mm	9	5.76	0.68	5.32–6.20	5.30	7.53
Test scenario 3, 5 mm	7	6.16	0.91	5.32–7.05	5.37	7.80
Incorrect detector						
Test scenario 1, E	4	6.51	1.00	6.11–6.91	5.66	7.81
Test scenario 2, 15 mm	1	6.32	—	—	—	—
Test scenario 3, 5 mm	7	6.82	3.29	4.39–9.26	5.41	14.26

10 times more responsive to dynamic stimuli than to static,[19] varying one's finger on the substrate may enhance his or her ability to detect smaller objects. A further analysis of DIFF also showed that incorrect detectors of abnormalities used a frequency of approximately 6 Hz similar to correct detectors. This seems to be a preferred frequency. This finding also indicates that the utilization of the 6 Hz frequency does not guarantee detection, although the number of correct detectors (22, 9, and 7; total = 38) was greater than the number of incorrect detectors (4, 1, and 7; total = 12) across the three abnormalities (E size, 15, and 5 mm).

Although the correlation of GFM and AIFP with correct detection was not as strong as for LFM, both measures reveal interesting insights. The use of the V or L GFM patterns, where greater area is palpated, increases the number of abnormalities detected in a single prostate, compared with the Line pattern. This links with literature on the qualitative employment of palpation techniques, whereby it is consistently acknowledged that a main factor in tumor palpation is that the examiner systematically covers the entire area.[8–10,15] As for AIFP, although the finger pressure used did monotonically increase for correct detectors of incrementally smaller tumors, the magnitude differences were not statistically significant. This was in part because of a wide variance in finger pressure used between examiners.

Clinical Applicability of the Palpation Techniques

Although several studies have evaluated genitourinary skills training,[20,21] we found few studies to date that analyze the finger palpation techniques used during a DRE. In one in particular, however, the authors found that medical students were more likely to supinate and pronate their finger as opposed to sweeping across the gland as did the experienced clinicians.[14] This is not only an inadequate DRE but will also miss the lateral aspects of the prostate where prostate cancer is most likely to occur. Poststudy, one-on-one interviews with the participants found that the videos and textbooks the students used to learn the DRE did not adequately address all the palpation maneuvers. Similarly, medical students at the University of California, Irvine, have ranked standardized patient digital rectal examinees as more valuable than either didactic lectures or tutorials of abnormal genitourinary exams.[21]

Taken together, these studies suggest that the current education of the DRE in medical school and residency can be improved. At present, in our institution, we teach medical and nurse practitioner students to systematically search the entire prostate but not how to do so. Resident DRE training falls into the apprenticeship category with attending urologists performing a rectal examination after the resident and describing what was palpated. This discussion centers more on clinical diagnosis rather than palpation techniques, perhaps to the resident's detriment. Furthermore, this method provides no feedback to the resident as to what his or her finger actually palpated but rather only what the attending urologist felt the resident should have palpated. Such instruction differs from offering specific techniques as in the breast examination (eg, the use of a global coverage pattern of vertical strip or radial spoke), where global (and local) techniques increase rates of detection.[13]

The finding that particular LFM patterns (firm pressure with varying intensity) were linked with the detection of the smallest lesions is provocative. Although the analysis of patterns here is our first effort to determine how to instruct learners, the logical next step is to discern whether experienced urologists use these same palpation patterns in the clinical arena. We are pursuing these experiments. If repetitive patterns are uncovered among experienced urologists, then these palpation techniques can be systematically taught and may improve DRE education for all health care practitioners. Also of note is that from the perspective of future training, we seek to tighten links to what will be encountered in the clinic. Toward that end, we sought here to use sets of tumors and abnormalities of a variety of sizes, depths, hardness, etc. that lie, as reported elsewhere,[18] within a reasonable level of detectability.

Study Limitations

There are some limitations to the study and algorithmic methods. First, our sample population included residents and nurse practitioner students. Although this group has some clinical experience, a further study should be extended to experienced, attending urologists to further validate these outcomes. Second, the finger palpation techniques that were quantified here relate directly to the clinical literature, but there may yet be other techniques, currently unknown, which could also be of interest. Third, we used two proctors in the study, one to record participants' answers and the other to operate the simulator. Although recording in writing a participant's verbal answers might tend to introduce bias, we used a standardized score sheet, trained the proctors, and ensured that the proctor to which the participant reported was blind to the abnormality being presented. This was done because it was awkward for the participants to write while

conducting multiple sequential exams with a glove and surgical lube.

REFERENCES

1. American Cancer Society. Screening and Early Detection of Prostate Cancer. Atlanta, GA: American Cancer Society; 2008.

2. Weinmann S, Richert-Boe KE, Van Den Eeden SK, et al. Screening by prostate-specific antigen and digital rectal examination in relation to prostate cancer mortality: —a case-control study. *Epidemiology* 2005; 16:367–376.

3. Roehrborn CG, Sech S, Montoya J, Rhodes T, Girman CJ. Interexaminer reliability and validity of a three-dimensional model to assess prostate volume by digital rectal examination. *Urology* 2001;57:1087–1092.

4. Smith DS, Catalona WJ. Interexaminer variability of digital rectal examination in detecting prostate-cancer. *Urology* 1995;45:70–74.

5. Gerling GJ, Rigsbee S, Childress RM, Martin ML. The design and evaluation of a computerized and physical simulator for training clinical prostate exams. *IEEE Trans Syst Man Cybern A Syst Hum* 2009; 39:388–403.

6. Madan AK, Aliabadi-Wahle S, Babbo AM, Posner M, Beech DJ. Education of medical students in clinical breast examination during surgical clerkship. *Am J Surg* 2002;184:637–640.

7. Campbell HS, McBean M, Mandin H, Bryant H. Teaching medical-students how to perform a clinical breast examination. *Acad Med* 1994;69:993–995.

8. Saslow D, Hannan J, Osuch J, et al. Clinical breast examination: practical recommendations for optimizing performance and reporting. *CA Cancer J Clin* 2004;54:327–344.

9. Saunders KJ, Pilgrim CA, Pennypacker HS. Increased proficiency of search in breast self-examination. *Cancer* 1986;58:2531–2537.

10. Pennypacker H, Iwata M. MammaCare: a case history in behavioural medicine. In: Blackman D, Lejeune H, eds. *Behaviour Analysis in Theory and Practice*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1991: 259–288.

11. McDermott MM, Dolan NC, Rademacher A, Pilgrim C. Effect of breast-tissue characteristics on the outcome of clinical breast examination training. *Acad Med* 1996;71:505–507.

12. Pilgrim C, Lannon C, Harris PR, Cogburn W, Fletcher SW. Improving clinical breast examination training in a medical school: a randomized controlled trial. *J Gen Intern Med* 1993;8:685–688.

13. Hall DC, Adams CK, Stein GH, Stephenson HS, Goldstein MK, Pennypacker HS. Improved detection of human breast lesions following experimental training. *Cancer* 1980;46:408–414.

14. Balkissoon R, Blossfield K, Salud L, Ford D, Pugh C. Lost in translation: unfolding medical students' misconceptions of how to perform a clinical digital rectal examination. *Am J Surg* 2009;197:525–532.

15. Pugh CM, Youngblood P. Development and validation of assessment measures for a newly developed physical examination simulator. *J Am Med Inform Assoc* 2002;9:448–460.

16. Wang N, Gerling GJ, Moyer Childress R, Martin ML. Quantifying palpation techniques in relation to performance in a clinical prostate exam. *IEEE Trans Inf Technol Biomed* (in press).

17. Kandel ER, Schwartz JH, Jessell TM. *Principles of Neural Science*. 4th ed. New York, NY: McGraw-Hill Professional; 2000:437.

18. Baumgart LA, Gerling GJ, Bass EJ. Characterizing the range of simulated prostate abnormalities palpable by digital rectal examination. *Cancer Epidemiol* 2010;34:19–29.

19. Johnson KO, Yoshioka T, Vega-Bermudez F. Tactile functions of mechanoreceptive afferents innervating the hand. *J Clin Neurophysiol* 2000;17:539–558.

20. Dakum NK, Ramyil VM, Abgo S, Ogwuche E, Makama BS, Kidmas AT. Digital rectal examination for prostate cancer: attitude and experience of final year medical students. *Niger J Clin Pract* 2007; 10:5–9.

21. Kaplan AG, Kolla SB, Gamboa AJR, et al. Preliminary evaluation of a genitourinary skills training curriculum for medical students. *J Urol* 2009;182:668–673.

Material characterization of *ex vivo* prostate tissue via spherical indentation in the clinic

William C. Carson[a], Gregory J. Gerling[a,*], Tracey L. Krupski[b], Casey G. Kowalik[b], Jeffrey C. Harper[c], Christopher A. Moskaluk[c,d]

[a] Department of Systems and Information Engineering, University of Virginia, United States
[b] Department of Urology, University of Virginia Health System, United States
[c] Department of Pathology, University of Virginia Health System, United States
[d] Department of Biochemistry and Molecular Genetics, University of Virginia, United States

ARTICLE INFO

ABSTRACT

Background: The mechanical characterization of prostate tissue has not received much attention and is often disconnected from the clinic, where samples are readily attained.
Methods: We developed a spherical indenter for the clinic to generate force–displacement data from *ex vivo* prostate tissue. Indentation velocity, depth, and sphere diameter, and four means of estimating elastic modulus (EM) were validated. EM was then estimated for 26 prostate specimens obtained via prostatectomy and 6 samples obtained from autopsy. Prostatectomy prostates were evaluated clinically upon digital rectal exam and pathologically post-extirpation.
Findings: Whole-mount measurements yielded median EM of 43.2 kPa (SD = 59.8 kPa). Once sliced into cross-sections, median EM for stage T2 and T3 glands were 30.9 and 71.0 kPa, respectively, but not significantly different. Furthermore, we compared within-organ EM difference for prostates with (median = 46.5 kPa, SD = 22.2 kPa) and without (median = 31.0 kPa, SD = 63.1 kPa) palpable abnormalities.
Interpretation: This work finds that diseased prostate tissue is stiffer than normal tissue, stiffness increases with disease severity, and large variability exists between samples, even though disease differences within a prostate are detectable. A further study of late-stage cancers would help to strengthen the findings presented in this work.

© 2010 IPEM. Published by Elsevier Ltd. All rights reserved.

1. Introduction

The digital rectal exam (DRE), a palpation-based diagnostic, is part of the recommended clinical screening protocol for prostate cancer [1]. Training practitioners for the DRE can improve performance and is typically done with standardized-patients or part-task and human-patient simulators [2,3]. However, most current simulators are palpably inaccurate and this is partially due to a lack of literature on the mechanical properties of the prostate. Better understanding a tissue's stiffness, viscoelastic and vibratory responses may also impact the further development and validation of non-invasive characterization technologies, especially since we know that bulk tissue properties systematically vary between normal and diseased organs (e.g., in the liver [4]).

Materials characterization of biological tissues, such as bone, cartilage, liver, kidney, and other organs [5–8], is typically done via tensile testing, ultrasound, whole-organ compression, aspiration, or indentation testing—each modality has its benefits and limitations. Tensile testing is often the preferred and simplest form of materials characterization [9], but the thin cross-sections it requires destroy the original sample. Moreover, tensile tests do not subject the sample to compressive forces similar to those applied in the DRE. Ultrasonic imaging is a non-invasive method that has been used successfully with soft tissues [10,11]. Another technique is whole-organ compression, which precludes one from precisely gathering data from regions-of-interest within an organ. Finally, indentation testing is minimally damaging and, if necessary, can be performed on organs *in vivo* [12,13], like recent work using aspiration-based material characterization devices [14,15]. Although indentation tests are typically not as accurate as uni-axial tensile tests for estimating material properties, reasonable engineering estimates can be achieved for these parameters [16].

At least four groups have analyzed prostate tissue, in various states of health, using ultrasonic elastography and strain mapping methodologies [17–20]. In this work, we used a spherical inden-

* Corresponding author at: Department of Systems and Information Engineering, University of Virginia, P.O. Box 400747, 151 Engineer's Way, Charlottesville, VA 22904, United States. Tel.: +1 434 924 0533; fax: +1 434 982 2972.
E-mail address: gregory-gerling@virginia.edu (G.J. Gerling).
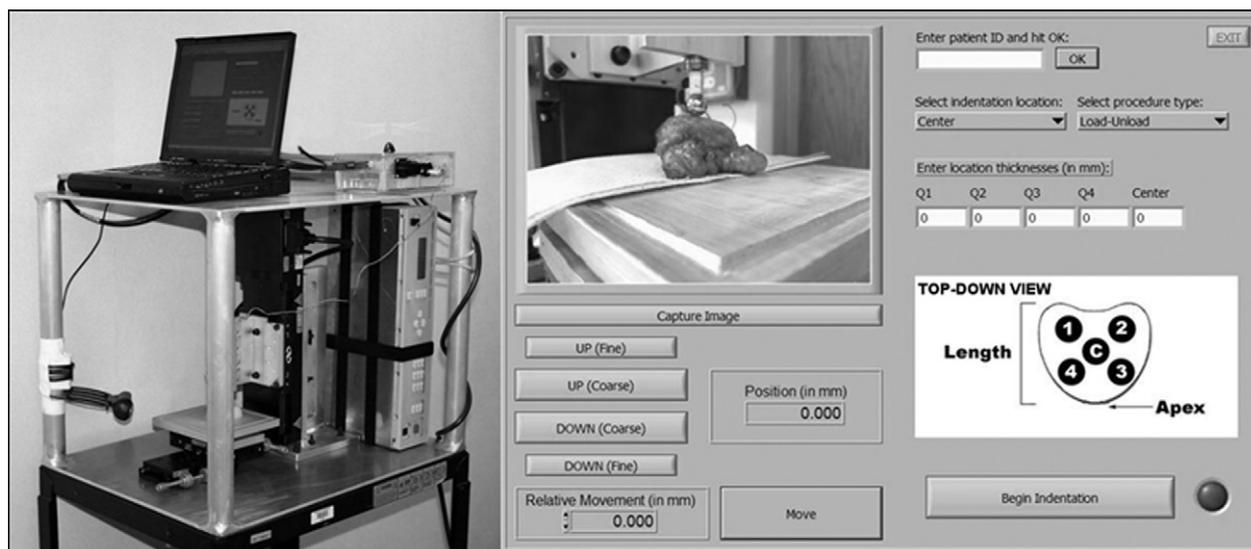
Fig. 1. Indenter rig (left) and user interface (right).

tation technique to estimate elastic modulus (EM) for the prostate in both whole-organ and cross-sectioned forms. The rationale for using this approach was that: (1) the technique does not damage the tissue prior to histopathologic sectioning; (2) the interface between sphere and sample matches the interface between fingertip and prostate during the DRE; and (3) the data collection process takes less than 20 min, as diseased prostate tissue is pathologically evaluated shortly after extirpation. EM was compared between healthy and diseased prostates gathered post-autopsy and post-prostatectomy, respectively. Though calculating EM is not always straightforward for biological samples, it is a convenient metric to compare tissue stiffness across specimens. Furthermore, although family history, prostate-specific antigen (PSA) levels, and the DRE help detect prostate cancer, ultimately the diagnostic result is a histopathologic one. We measured the elastic modulus of organ confined (T2) and non-organ confined tumors (T3) for comparison to clinical and histopathologic findings.

2. Methods

To compare the elasticity of normal and diseased tissues, we designed and built a spherical indentation rig to process whole-organ and cross-sectioned *ex vivo* prostate tissue samples. From 26 prostates removed via radical prostatectomy and 6 removed at autopsy, we obtained EM values as converted from load versus displacement measurements. These prostates were also clinically evaluated upon DRE and pathologically evaluated post-operation. The data were analyzed according to three research questions:

(1) What are the ranges of EM: (a) between and within prostate glands and (b) between prostates gathered via autopsy and prostatectomy?
(2) Considering cross-sectioned samples, how does stiffness relate to the pathologic T stage?
(3) Within a single prostate diagnosed with carcinoma, is one indentation site stiffer than an adjacent location and does this differential relate to a DRE report of firmness or nodularity?

Since a novel method is used for calculating EM in soft tissues, we evaluated its robustness beforehand in tests with silicone-elastomer phantoms (see Appendix).

2.1. Apparatus: actuation, recording hardware, and data collected

The actuation sled, recording hardware and user interface are shown in Fig. 1. The indenter tip is a 12 mm diameter AISI E52100 steel ball mounted to an aluminum standoff (1.27 cm length, 0.64 cm diameter). The standoff is attached to a load cell (Honeywell, Model Sensotec 11, Columbus, OH) with a 44 N maximum load capacity. The load cell is mounted to an aluminum sled which is driven by a motorized linear stage (Newport, Model ILS100, Mountain View, CA) with 100 mm travel and 50 mm/s maximum velocity. It is controlled by a motion controller (Newport, Model ESP300) with 0.0001 mm positioning accuracy. A laptop computer (Lenovo, Model X61 ThinkPad, Morrisville, NC) is used for data acquisition (National Instruments, Model DAQCard-6036E, Austin, TX) and for operating the graphical user interface (National Instruments, Lab-View 8.5 Professional). We used a 24-bit, 1000 sample/sec A–D converter and a high-quality instrumentation amplifier to digitize analog load cell signals for the National Instruments I/O card on the computer. Samples are positioned on an aluminum plate attached to two low-profile linear stages (Newport, Model 443) mounted in an X–Y configuration with 50 mm travel and 5 turn/mm positioning resolution.

Two classes of data are logged, force on the load cell and position of the linear stage (1×10^{-6} N and 0.0005 mm resolutions, respectively). To smooth the recorded force versus displacement data from the load cell, the Smoothing Filter Coefficients Virtual Instrument in LabView was implemented with a 20-sample half-width. This is a moving average filter of the form:

$$MA(f(x_0)) = \frac{\sum_{i=-m}^{m} f(x_i)}{2m+1},\tag{1}$$

where force, f, is a function of displacement, x, for a given time epoch, i, and m is the half-width of the filter. Since the experimental data follow a generally monotonically increasing trend across time, the moving average filter positively biases the resultant data. However for these data, this bias is negligible. Fig. 2 shows an example of smoothed force versus displacement data collected from a cross-sectioned prostate sample of average EM.

2.2. Equations for estimating EM

EM is calculated for soft tissues from load versus displacement data using a technique originally used to characterize hard
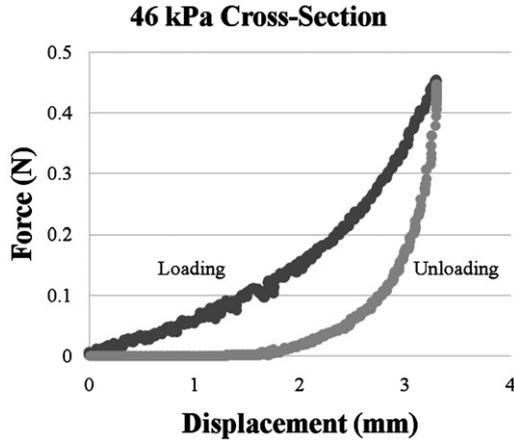
Fig. 2. Example force versus displacement data taken from a cross-section sample.

materials. The Appendix details the specifics of the preliminary experiments used to validate the procedure with soft tissues. We refer to the following equation as the "Oliver–Pharr method" for relating Oliver–Pharr stiffness, S, and contact area, A, to reduced EM, E_{red} [21]:

$$E_{red} = \frac{S\sqrt{\pi}}{2\sqrt{A}}. \tag{2}$$

Contact area is simplified to $2\pi r^2$, the surface area of a hemisphere, where r is the radius of the indenter. Justification for this simplification can be found in Appendix A.1.2. Oliver–Pharr stiffness is given by the following equation [21]:

$$S = \frac{dP}{dh}, \tag{3}$$

where dP/dh is the change in load, P, with respect to the change in indentation depth, h, at the initial instant of unloading in a load–unload indentation procedure. To approximate S, we used the slope from a linear fit applied to the first 20 *unloading* data points from the force versus displacement data gathered by the indenter for all specimens—whole-mount and cross-sectioned, biological and synthetic. Trial and error experimentation led us to the use of 20 data points for the slope estimate, as it minimized error due to jittery load cell data and downward bias from too many samples in the time-dependent linear estimate. Finally, E_{red} is related to both the EM of the indenter, E_i, and the EM of the sample, E_s [22]:

$$E_{red} = \left[\frac{1 - v_s^2}{E_s} + \frac{1 - v_i^2}{E_i} \right]^{-1}, \tag{4}$$

where v_s and v_i are Poisson's ratios for the sample and indenter, respectively. Note that in this work, values for E_i and v_i are 210 GPa and 0.30, respectively, for an AISI E52100 steel ball bearing. Based on internal organ material property data in Fung [5,23] and ranges

of Poisson's ratios used in Hu's modeling of the prostate [24], we set v_s to 0.47 as an estimate for the Poisson's ratio of prostate tissue. We set v_s to 0.499 for the nearly incompressible, polydimethylsiloxane-based silicone-elastomers. Since E_i is about 10^4 times greater than E_s, the inclusion of the second term contributes minimally to the reduced elastic modulus of the system. Nonetheless, Eq. (4) was presented in this form for completeness.

The Oliver–Pharr method with simplified contact area, described above, was selected from the possible candidates because it best minimized EM variance in the preliminary experiments with silicone-elastomers. This method was also desirable because it was insensitive to indenter diameter and sample thickness, it was highly sensitive to sample stiffness and small stiffness differentials, and through validation with finite element (FE) analysis, it produced reasonable EM estimates. For support of these findings, see Figs. 5 and 6 and Tables 7 and 8 in Appendix.

2.3. Experiment

After obtaining approval from the Institutional Review Board, patients undergoing radical prostatectomy consented to allow the material properties of their prostates to be measured. There were three parts to the experiment: the measurement of the material properties on both whole-mount and cross-sectioned samples, pathological evaluation, and chart abstraction to document the clinical indications.

2.3.1. Material properties of radical prostatectomy specimens

Once dissected free of investments (i.e., the vas deferens, seminal vesicles, urethra, and musculature) and removed from the patients, prostates were transferred from the operating room to pathology within 20 min of extirpation (autopsies were conducted within 24 h after patient deaths and tissue samples were harvested within 1 h after the procedures began). After weighing and marking surgical margins, spherical indentation measurements were obtained from intact and then cross-sectioned specimens.

In the whole-mount procedure, the prostates were placed on top of an aluminum plate, in an orientation of anterior-side down, such that their apices faced the operator. The samples were neither constrained at their periphery nor affixed to the plate. The geometry of the excised prostate is such that when placed anterior-side down, a substantial portion of the organ's surface area rests flat on the specimen plate and the indenter contributes to indentation into the tissue, not whole-organ deformation. Each sample was indented to 30% of its total thickness in the z-direction (i.e., 9 mm indentation for 30 mm thick whole mount samples and 2.4 mm indentation for 8 mm thick cross sections), at a rate of 0.1 mm/s and in 2–4 of the following quadrants—superior right, superior left, inferior left, and inferior right (see Fig. 3). We consistently indented in the inferior left and right quadrants, as these locations are most readily accessible in the DRE; however, measurements were made in a greater breadth of quadrants in the first few experimental
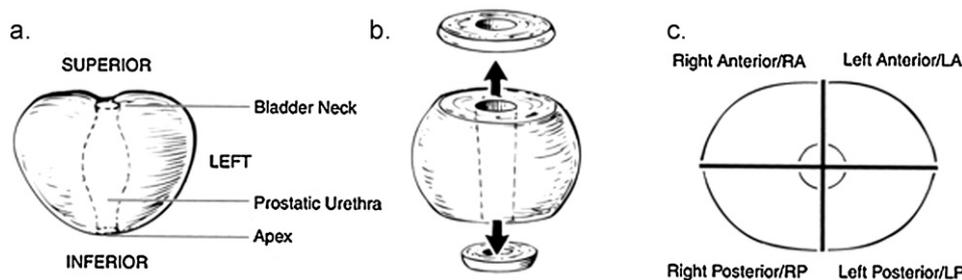
Fig. 3. (a) Whole-mount prostate on specimen plate; (b) cross-sectioning procedure; (c) a single, cross-sectioned sample with its four quadrants labeled. *Note*: views (a and c) are top-down and indentation occurs in the z-direction into the page. Adapted from [30].

Table 1
EM across whole-mount prostatectomy and autopsy samples.

	EM (kPa)		
	Prostatectomy $n = 92$	Autopsy $n = 18$	All $n = 110$
Median	43.5	42.5	43.2
Mean	60.2	69.6	61.8
SD	46.9	105.3	59.8

Table 2
EM for cross-section samples across pathologic diagnosis.

	EM (kPa)			
	Benign $n = 71$	AC $\geq$ 10% $n = 39$	BPH $n = 35$	Prostatitis $n = 20$
Median	27.2	43.0	26.2	31.9
Mean	41.1	135.0	36.8	49.2
SD	41.3	240.6	37.0	52.8

runs. Each load–unload procedure took between 6 and 10 min, with approximately 15 min being spent taking all measurements of a single sample.

After the whole-mount measurements were completed, the cross-section procedure was performed with prostate specimens serially sectioned into 3–4 slices between 0.5 and 1.0 cm thick. These sections were placed on the specimen plate, in an orientation with the inferior-side down and posterior-side closest to operator. Four measurements were taken from each slice; thus, approximately 20 measurements were possible for each prostate—4 from whole mount and 16 from 4 slices of 4 quadrants each. *En face* tissue samples from each quadrant that had been indented were then procured for histologic processing. The cross-sectional procedure was performed only with prostates acquired via prostatectomy, as there were no gross differential diagnoses done for autopsy samples.

2.3.2. Pathologic and histologic evaluation of prostate tissue

The remainder of the specimen was processed and sampled for routine histologic examination by the clinical staff of the university hospital's pathology department. The *en face* tissue samples were processed into formalin-fixed paraffin blocks from which hematoxylin and eosin-stained histologic sections were made. The histologic sections were examined by a single Board-Certified Anatomic Pathologist. The sections were scored for the presence of normal tissue, benign prostatic hyperplasia (BPH), inflammation and carcinoma. If carcinoma was present, the Gleason score and the percentage of cross sectional area occupied by cancer were recorded. Each prostate was given a pathologic T stage and assessed for margin positivity and percentage of gland affected by adenocarcinoma using standard techniques [25,26].

2.3.3. Chart abstraction to interpret clinical findings

We performed chart abstraction to interpret clinical findings from the DREs and pathologic gross differential diagnoses. While the procedures for reporting gross differential diagnosis are specified in *M.3.2*, DRE reports typically involve narrative statements which address prostate shape, size, stiffness, texture, symmetry, and nodularity. A trained urologist abstracted these narratives into "abnormality-positive" or "abnormality-negative" for both the right and the left lobes of all prostates.

3. Results

The dataset, consisting of 32 excised prostates from 26 radical prostatectomies and 6 autopsies, was analyzed in accordance with the three research questions.

3.1. EM for whole-mount and cross-section samples

3.1.1. Whole-mount measurements

With respect to the whole-mount specimens, the EM (median, mean and standard deviation) for prostatectomy, autopsy and combined samples are presented in Table 1. The median EM for prostatectomy samples is slightly greater than that for autopsy samples, while the mean EM is greater for autopsy samples than it is for prostatectomy samples, which reflects the contribution of

Fig. 4. EM versus percent adenocarcinoma for cross-section samples.

the outliers but also follows Tay et al. [12], since organ stiffness increases with the length of time after death. However, these differences were not statistically significant. In Table 1, each indentation is considered as one sample and each prostate was indented 2–4 times, therefore *n* represents the total number of samples and not participants.

3.1.2. Cross-section data

From each cross-section, one indentation measurement was made per quadrant. Based on histologic findings (benign, adenocarcinoma in at least 10% of sample, BPH, or prostatitis), the median, mean, and standard deviation of the EM are shown in Table 2. Samples with no histologic data, adenocarcinoma in less than 10% of sample, high-grade prostatic intraepithelial neoplasia (HGPIN), or composed of smooth muscle were not included.

From these data, the adenocarcinoma group exhibits the greatest median and mean EM, followed by the prostatitis, benign, and BPH groups, but these differences are not statistically significant. Furthermore, we plotted EM versus percentage adenocarcinoma in Fig. 4 and found there is no relationship between the two variates.

3.2. Relationship between stiffness and pathologic T stage

One prostate was given a T0 stage, 23 were given T2 stages, and two were given T3 stages. With the prostatectomy samples grouped by T2 and T3 pathologic stage, the EM (median, mean and standard deviation) for whole-mount measurements are shown in Table 3. Both the median and the mean EM for the stage T3 samples are greater than those for the stage T2 samples, with the T3 median of 59.9 kPa and the T2 of 42.1 kPa, but again these differences are insignificant.

Table 3
EM for whole-mount samples from stage T2 and T3 prostates.

	EM (kPa)	
	T2 $n = 83$	T3 $n = 6$
Median	42.1	59.9
Mean	59.9	61.8
SD	48.8	31.4

Note: $n_{T2} + n_{T3} < n_{prostatectomy}$ since one prostate (three samples) was stage T0.

Table 4
EM for cross-section samples from stage T2 and T3 prostates.

	EM (kPa)	
	T2 $n=27$	T3 $n=10$
Median	30.9	71.0
Mean	91.4	241.5
SD	201.7	328.9

Table 5
Maximum within-prostate EM difference for whole-mount samples with and without palpable abnormalities.

	Max EM difference within prostate (kPa)	
	Palpable $n=5$	Not palpable $n=14$
Median	46.5	31.0
Mean	47.7	58.1
SD	22.2	63.1

As with the whole mount samples, we grouped the cross-section samples by clinical stage, given the samples had at least 10% adenocarcinoma. As Table 4 indicates, the T3 median of 71.0 kPa is larger than the T2 median of 30.9 kPa, but this difference is not significant. The mean EM for stage T3 samples is twice the mean of stage T2 samples.

3.3. Within prostate EM variability and relationship to DRE findings

We calculated the maximum difference in EM within prostates for those which had DRE-palpable abnormalities and those which did not in Table 5. In comparing the maximum EM difference, the median difference in EM for prostates with palpable abnormalities is greater than that for prostates without palpable abnormalities. Although counterintuitive, the mean difference in EM for prostates with palpable abnormalities is less than that for prostates without palpable abnormalities, but the standard deviations for both groups are large and the difference between these means is insignificant.

A further analysis was then done to consider within-prostate stiffness differences. Few prostate specimens yielded samples which are uniformly malignant. For a given prostate, we compared the EM of malignant and benign samples to examine across-disease stiffness variability within an organ. As shown in Table 6, there appears to be a relationship between malignancy and EM within-prostate, although more samples are required for confirmation.

A paired t-test was performed and yielded a p-value of 0.05. However, because the data violate the "normally distributed" and

Table 6
Comparison of within-prostate EM for malignant and benign cross-section samples.

Sample, slice, ANT/POS	EM (kPa)		Difference	% Difference
	Malignant	Benign		
11, 1, ANT	17.4	9.4	8.0	46.2
15, 1, ANT	264.6	20.8	243.8	92.1
15, 1, POS	68.5	19.2	49.4	72.0
15, 2, POS	920.8	28.7	892.1	96.9
16, 3, ANT	70.3	63.8	6.5	9.3
29, 2, POS	17.6	13.3	4.3	24.5
30, 1, POS	102.9	44.3	58.6	56.9
32, 1, POS	978.3	17.4	960.9	98.2
32, 1, ANT	7.1	30.6	−23.5	−331.6
Median	–	–	49.4	–
Mean	–	–	244.5	–
SD	–	–	394.9	–

"equal variance" assumptions required to infer from the results of a t-test, we must be cautious with this result.

4. Discussion and conclusions

In reviewing the results on the whole-mount level, we find few significant differences between the EM for normal and diseased tissues, because across-sample variability is large. The mean EM for our benign samples from autopsy was 69.6 kPa—Krouskop et al. reported approximately 65 kPa [19], Murayama et al. reported 10–15 kPa [17], and Parker et al. reported only 2.15 kPa [18]. The mean EM for diseased samples was 60.2 kPa, while Krouskop et al. reported 230 kPa. Again, we believe the mean EM for autopsy samples is greater than that for prostatectomy samples because of a change in tissue stiffness due to patient death, as seen in Tay et al. [12]. There are also no systematic differences between the EM for prostates with and without palpable abnormalities at the whole-mount level. However, it is important to note that a spherical indentation procedure in which a large sphere is used and few indentations are performed can only reveal information about a prostate's overall firmness. Detection of DRE-specific disease cues, such as nodularity, texture, and size, must rely upon a set of metrics other than simple EM and force versus displacement estimations. Finally, there is an insignificant difference in EM between pathologic stage T2 and T3 for whole-mount samples, though we only have 6 samples with stage T3 disease. We hypothesize that since prostate cancer is detected earlier due to the PSA test, it is difficult to gain access to prostates with more advanced disease.

Although the mean cross-section sample differences in Table 2 were not statistically significant, the results follow Murayama and Phipps [20] in their conclusions that stiffness differences exist between prostate glands with and without carcinoma. The mean EM for benign samples was 41.1 kPa, 135.0 kPa for samples with adenocarcinoma in over 10% of the gland, and 36.8 kPa for samples with BPH. Of note is that Krouskop et al. reported a similar mean of 38 kPa for samples with BPH. We found that cross-section samples from stage T3 (non-organ confined) tumors were stiffer than samples from stage T2 (organ confined) tumors, with medians of 71.0 and 30.9 kPa, respectively. However, these differences were insignificant. There are insignificant EM differences for cross-sectioned samples if one compares malignant and benign regions within the same prostate. It is this type of within-prostate comparison upon which a DRE most likely relies. However, we only have 9 cases in which this situation arises, again likely due to cancers now being caught earlier because of PSA screening.

Large variance in the measurements is evident in the study. However, we conducted a thorough validation of the indentation and EM calculation techniques used here in the preliminary experiments with silicone elastomers. Based on those results, we believe we can justify that the large variances in EM observed for whole-mount and cross-section *ex vivo* prostate tissue samples are due to large differences in tissue stiffness, as opposed to measurement error. Other recent work [7] also affirms that accurate characterization of the mechanical properties of soft biological tissues can be accomplished through small indentation devices.

Finally, we note that we used a procedure to acquire the prostates to be measured within 20 min upon removal from the body. Although this minimized tissue degradation, others have made significant efforts to keep tissues perfused both before and during the measurement procedure. It has been shown that perfused and non-perfused livers have different mechanical properties [27]; however, the prostate is a small, fibrous gland which secretes comparatively less fluid. Furthermore, our diseased organs were cauterized upon extirpation and tested immediately thereafter. As such, we hypothesize that the *ex vivo* mechanical properties pre-

sented herein closely approximate *in vivo* conditions encountered during the DRE.

5. Study limitations

There are limitations to the experiment and mathematical methods. Our selection of a relatively large 12 mm sphere somewhat limited the spatial precision of our force versus displacement data. However, this size sphere was validated in preliminary experiments and allowed us to characterize the material properties of the prostate at large, yielding a higher probability of contacting a tumor, given our 20 min constraint in which to complete testing. For cross-section samples, a load cell with lower total range, perhaps ∼10 N, may have yielded more accurate data in lower load ranges. Nonetheless, the use of a high-fidelity 24-bit A–D converter allowed us to gather data which was still highly accurate over the load cell's full scale. It would also have been ideal to perform uniaxial testing on at least one prostate for comparison with our spherical indentation findings. Finally, most of the prostates studied here had low-grade cancers, most likely due to advances in diagnosing prostatic disease. Sampling more prostates with advanced stage disease may have further strengthened our findings of stiffness differences between normal and diseased tissue.

Acknowledgments

Appendix A.

A.1. Indenter validation with silicone-elastomer specimens

Validation of the indentation procedure and the means for calculating EM were done with silicone-elastomer specimens. The silicone-elastomers were employed because their properties are controllable and they are mechanically similar to soft biological tissues. The spherical indentation parameters, indentation velocity and indenter tip diameter, were validated for the load–unload procedure in conjunction with controlled changes to the stiffness and thickness of the silicone-elastomer samples. Given those independent variables, the outcome variables were EM, calculated using four disparate methods in Section A.1.2.

A.1.1. Independent variables

The first set of independent variables included the stiffness and thickness of the silicone-elastomer samples (BJB Enterprises, TC-5005, Tustin, CA). We employed three levels of stiffness, determined by the percentage of silicone mixture component C, in which a greater percentage yields a more pliant sample. These stiffness levels, 25, 75, and 85%, correspond with those thought to reasonably estimate prostate stiffness, as determined by psychophysical tests with clinicians [2]. For each level of stiffness, there were two levels of sample thickness: 20 and 30 mm.

The second set of independent variables relates to the spherical indentation procedure and include indentation velocity and indenter tip diameter. There were two levels for indentation velocity, 0.01 mm/s and 0.1 mm/s. Previous work by Cox et al. with bio-artificial muscle has shown that a suitable lower-bound to velocity is 0.01 mm/s with a 2 mm indenter [28]. However, an indenter driven at 0.01 mm/s requires more than 20 min to complete one load–unload procedure on a whole-mount sample, an impractical timeframe for the clinical setting. The difference in EM estimates for data collected at indenter speeds of 0.01 mm/s and 0.1 mm/s were approximately one percent. Furthermore, preliminary tests at velocities over 0.1 mm/s yielded poor estimates of EM. There were six levels for indenter tip diameter: 6 mm, 8 mm, 10 mm, 12 mm, 16 mm, and 24 mm.

A.1.2. Dependent variables

The outcome variables were EM values calculated using four disparate methods. The first is the "compressive strain" method:

$$E_{red} = \frac{p_m}{\varepsilon_{est}} = \frac{P/A}{h/t}, \tag{5}$$

where p_m is mean contact pressure or indentation stress, given by load per unit contact area, and ε_{est} is a rough estimate for compressive strain, based on indentation depth and sample thickness at the point of indentation, t.

The second method is the "indentation strain" method, and closely resembles the first, but uses the notion of indentation strain, or ε_{ind}, which is given by indenter contact radius, a, divided by nominal indenter radius, r [29]:

$$E_{red} = \frac{p_m}{\varepsilon_{ind}} = \frac{P/A}{a/r}. \tag{6}$$

Fischer-Cripps proposed that indentation strain scales with elastic strains for a given specimen, thus providing an analog to the stress-strain relationship found in uniaxial tensile testing [29]. The third method was derived from Fischer-Cripps' analysis of Hertz's contact equations:

$$E_{red} = \frac{3rP}{4a^3}. \tag{7}$$

The fourth method, Oliver and Pharr's stiffness to elastic modulus relation, is described in Section 2. We take indenter displacement, h, from Eq. (3) and subtract from it substrate surface deflection at the contact perimeter, h_s, to obtain contact depth, h_c [21]:

$$h_c = h - h_s, \tag{8}$$

where h_s is given by:

$$h_s = \frac{\kappa P}{S}. \tag{9}$$

The geometric constant, κ, of a spherical indenter is 0.75. For a spherical indenter, contact radius at any given contact depth is given by:

$$a = \sqrt{2rh_c - h_c{}^2}. \tag{10}$$

However, since the indenter has an EM much greater than that of the samples and the loads applied in this experiment were small enough to preclude macrodeformation of the steel tip, we made the assumption that the area of contact between the two objects was not of a conventional "Hertzian circle", but of a partial sphere. Therefore, for each displacement step, contact surface area between the specimen and the load applied by the indenter is given by:

$$A = 2\pi r^2[1 - \cos(\alpha(\cdot))], \tag{11}$$

where $\alpha(\cdot)$ is the function for the generating angle of the partial sphere of contact, with respect to the axis of compression. This generating angle is related to nominal radius of the indenter tip, contact depth, and contact radius:

$$\alpha(\cdot) = \arctan\left(\frac{a}{r - h_c}\right). \tag{12}$$

Combining Eqs. (8) through (12) yields:

$$A = 2\pi r^2 \left[1 - \cos\left(\arctan\left(\frac{\sqrt{2r(h - (\kappa P/S)) - (h - (\kappa P/S))^2}}{r - (h - (\kappa P/S))} \right) \right) \right]. \tag{13}$$

The use of Eq. (13) for contact area constitutes "complex" calculi in the results below, while the use of $2\pi r^2$ for contact area of the surface area of a hemisphere, constitutes "simple" calculi (which can be justified since we are using a large deformation framework with a rigid indenter and an extremely pliant substrate).

A.1.3. Results of silicone-elastomer tests

The procedure for testing the silicone-elastomer samples described in Section A.1.1 was that in each configuration of sample stiffness, sample thickness, and indenter diameter, one load–unload trial was performed per sample. These data were analyzed using ANOVA.

Fig. 5 and Table 7 contain visual and numerical summaries of the EM estimates from the tests. EM calculated with the complex Oliver–Pharr method (i.e., what would have been Method 4C) are not shown in Fig. 5A because they were hundreds of times larger than other values. Table 8 summarizes the ANOVA performed on the EM estimates using the simple Oliver–Pharr method with % Part C, indenter diameter, and sample thickness as factors.

Two important results arise from the data in Fig. 5 and Table 7. First, the simple Oliver–Pharr method minimizes the variance in EM across stiffnesses. Second, from the boxplot in Fig. 5B, we see that EM estimates across sample stiffnesses are well-defined, as the boxplots do not overlap. This suggests the indenter rig can gather data which yield unique EM across both relatively large stiffness differentials, 25% Part C versus 85% Part C, and small stiffness differentials, 75% Part C versus 85% Part C.

These effects are due to the insensitivity of the simple Oliver–Pharr method to indenter diameter and sample thickness compared to the other calculi. In Table 8, we performed ANOVA on the EM calculated using this method with sample stiffness, indenter diameter, and sample thickness as factors. As expected, sample stiffness accounted for an overwhelming proportion of variation in the data, approximately 99.12%, and yielded an F-statistic greater than 4093, which is significant at any practical level. Furthermore, indenter diameter accounted for 0.02% of data variation and was insignificant at the $\alpha = 0.05$ level. This result suggests that the indentation procedure and calculation method are robust to different indenter sizes, which allows us to compare the EM of prostates collected in this experiment. Sample thickness accounted for 0.05% of variation in the data and was significant at the $\alpha = 0.05$ level. However, although this factor was statistically significant, its practical significance is suspect, considering it accounted for so lit-

Fig. 5. (a) Results of silicone-elastomer validation study. EM (y-axis) is plotted against silicone cross-linker percentage and method of calculating EM (where number–letter designations represent the method and contact area calculation complexity: e.g., 1C is complex method 1). (b) A closer view of boxplots for the simplified Oliver–Pharr method, 4S.

Table 8
ANOVA table of simplified Oliver–Pharr method, 4S, for silicone-elastomer validation tests.

	Df	Sum of Sq.	Mean Sq.	F value	Pr(F)
% Part C	2	71149.84	35574.92	4093.48	0
Diameter	1	13.67	13.67	1.57	0.21
Thickness	1	34.85	34.85	4.01	0.05
Residuals	67	582.27	8.69	–	–

tle variation in the data and, for comparison, 0.81% of variation in the data was unexplained.

To ensure our EM estimates were reasonable, we created a FE model in ABAQUS 6.6 (SIMULIA, Providence, RI) which simulated spherical indentation by a 12 mm diameter indenter into a 20 mm thick, 85% Part C elastomer sample. This elastomer model used 2,160 CAX4RH elements (4-node, bilinear axisymmetric quadrilateral elements with hourglass distortion control) 0.5 mm × 0.5 mm in size and was constrained on the bottom to replicate the sample resting on the specimen plate during indentation. From our experimental data, we had estimated that such a sample had an EM of 25.3 kPa. In our FE analysis, we used two models to replicate this

Table 7
Summary of EM estimates from silicone-elastomer validation tests.

		Method 1 (kPa)		Method 2 (kPa)		Method 3 (kPa)	Method 4 (kPa)	
		Complex	Simple	Complex	Simple	–	Complex	Simple
25% Part C	Mean (SD)	144.38 (79.11)	133.47 (76.68)	46.29 (10.83)	44.00 (21.33)	181.88 (52.81)	23127.51 (10057.10)	94.68 (4.40)
	Coeff. of Variation	0.55	0.57	0.23	0.48	0.29	0.43	0.05
75% Part C	Mean (SD)	53.91 (33.72)	50.06 (26.18)	17.29 (23.74)	17.38 (7.51)	75.70 (16.88)	8161.82 (2286.40)	32.95 (2.59)
	Coeff. of Variation	0.63	0.52	1.37	0.43	0.22	0.28	0.08
85% Part C	Mean (SD)	41.57 (16.28)	36.43 (19.88)	14.33 (7.39)	12.67 (5.80)	54.11 (13.47)	5805.10 (1653.61)	23.95 (1.18)
	Coeff. of Variation	0.39	0.55	0.52	0.46	0.25	0.28	0.05

Fig. 6. FE analysis output for a silicone-elastomer specimen, noting the fit of linear and hyperelastic modeling parameters.

scenario: an Ogden hyperelastic model and a simple linear elastic model. Our Ogden hyperelastic model had parameters $\mu = 8.4$ kPa and $\alpha = 2$, since shear modulus, μ, is related to Young's modulus, Y, by:

$$\mu = \frac{Y}{2(1+\nu)}, \tag{14}$$

and $\alpha = 2$ is appropriate for near-incompressible, rubber-like materials. Our linear elastic model had Young's modulus $Y = 25.3$ kPa and Poisson's ratio $\nu = 0.499$. In Fig. 6 below, we overlay experimental and FE output force versus displacement data.

A visual inspection of the plot demonstrates that although the exact material properties of the simulated and experimental substrates differ, indentation tests yield very similar force versus displacement relationships.

To further validate the spherical indentation estimates, we compared EM values obtained in our spherical indentation experiments with Young's moduli from uniaxial tensile tests performed on a uniaxial testing machine (MTS Systems Corporation, Eden Prairie, MN). The three samples were 25%, 75%, and 85% Part C elastomers; they had lengths and cross-sectional areas of 0.0495 m and 0.001662 m^2, 0.0475 m and 0.001662 m^2, and 0.048 m and 0.001385 m^2, respectively. At 30% strain, uniaxial testing yielded Young's moduli of 116.8 kPa, 28.8 kPa, and 21.9 kPa for 25%, 75%, and 85% Part C elastomers, respectively. In comparison, the spherical indentation tests had yielded similar EM values of 94.7 kPa, 33.0 kPa, and 24.0 kPa, respectively.

These validation tests suggest our method for estimating EM in thick silicone-elastomer samples produces values close to Young's moduli generated using uniaxial tensile testing and commercial FE analysis software.

Conflict of interest

The authors have no conflicts of interest to disclose.

References

[1] Smith RA, Cokkinides V, Eyre HJ. Cancer screening in the United States: a review of current guidelines, practices, and prospects. CA Cancer J Clin 2007;57(2):90–104.

[2] Gerling GJ, Rigsbee S, Childress RM, Martin ML. The design and evaluation of a computerized and physical simulator for training clinical prostate exams. IEEE Trans Syst Man Cybern A Syst Hum 2009;39(2):388–403.

[3] Madan AK, Aliabadi-Wahle S, Babbo AM, Posner M, Beech DJ. Education of medical students in clinical breast examination during surgical clerkship. Am J Surg 2002;184:637–40.

[4] Mazza E, Nava A, Hahnloser D, Jochum W, Bajka M. The mechanical response of human liver and its relation to histology: an in vivo study. Med Image Anal 2007;11(6):663–72.

[5] Fung YC. Biomechanics: mechanical properties of living tissues. New York: Springer Verlag; 1993.

[6] Yamada H. Strength of biological materials. Baltimore: Williams and Wilkins; 1970.

[7] Egorov V, Tsyuryupa S, Kanilo S, Kogit M, Sarvazyan A. Soft tissue elastometer. Med Eng Phys 2008;30:206–12.

[8] Gerard JM, Ohayon J, Luboz V, Perrier P, Payan Y. Non-linear elastic properties of the lingual and facial tissues assessed by indentation technique: application to the biomechanics of speech production. Med Eng Phys 2005;27:884–92.

[9] Sharpe Jr WN, Bin Y, Vaidyanathan R, Edwards RL. Measurements of Young's modulus, Poisson's ratio, and tensile strength of polysilicon. In: Tenth Annual International Workshop on Micro Electro Mechanical Systems. 1997. p. 424–9.

[10] Chen EJ, Novakofski J, Jenkins WK, O'Brien Jr WD. Young's modulus measurements of soft tissues with application to elasticity imaging. IEEE Trans Ultrason Ferroelectr Freq Control 1996;43(1):191–4.

[11] Ophir J, Cespedes I, Garra B, Ponnekanti H, Huang Y, Maklad N. Elastography: ultrasonic imaging of tissue strain and elastic modulus in vivo. Eur J Ultrasound 1996;3(1):49–70.

[12] Tay BK, Jung K, Srinivasan MA. In vivo mechanical behavior of intra-abdominal organs. IEEE Trans Biomed Eng 2006;53(11):2129–38.

[13] Pailler-Mattei C, Bec S, Zahouani H. In vivo measurements of the elastic mechanical properties of human skin by indentation tests. Med Eng Phys 2008;30:599–606.

[14] Schiavone P, Chassat F, Boudou T, Promayon E, Valdivia F, Payan Y. In vivo measurement of human brain elasticity using a light aspiration device. Med Image Anal 2009;13:673–8.

[15] Nava A, Mazza E, Furrer M, Villiger P, Reinhart WH. In vivo mechanical characterization of human liver. Med Image Anal 2008;12:203–16.

[16] Herbert EG, Pharr GM, Oliver WC, Lucas BN, Hay JL. On the measurement of stress-strain curves by spherical indentation. Thin Solid Films 2001;398–399:331–5.

[17] Murayama Y, Omata S, Yajima T, Qiyu P, Shishido K, Peehl DM, et al. High resolution regional elasticity mapping of the human prostate. Conf Proc IEEE Eng Med Biol Soc 2007:5802–5.

[18] Parker KJ, Huang SR, Lerner RM, Lee Jr F, Rubens D, Roach D. Elastic and ultrasonic properties of the prostate. Proc IEEE Ultrason Symp 1993:1035–8.

[19] Krouskop TA, Wheeler TM, Kallel F, Garra BS, Hall T. Elastic moduli of breast and prostate tissues under compression. Ultrason Imaging 1998;20(4):260–74.

[20] Phipps S, Yang THJ, Habib FK, Reuben RL, McNeill SA. Measurement of tissue mechanical characteristics to distinguish between benign and malignant prostatic disease. Urology 2005;66:447–50.

[21] Oliver WC, Pharr GM. An improved technique for determining hardness and elastic modulus using load and displacement sensing indentation experiments. J Mater Res 1992;7:1564–83.

[22] Johnson KL. Contact mechanics. Cambridge University Press; 1985.

[23] Fung YC. Biomechanics: motion, flow, stress, and growth. New York: Springer Verlag; 1990.

[24] Hu Y, Morgan D, Ahmed HU, Pendsé D, Sahu M, Allen C, et al. Modelling prostate gland motion for image-guided interventions. Lect Notes Comput Sci 2008;5104:79–88.

[25] Greene FL, Page DL, Fleming ID, Fritz A, Balch CM. AJCC cancer staging handbook. 6th ed. New York: Springer-Verlag; 2002.

[26] College of American Pathologists. Reporting on cancer specimens; 2009.

[27] Kerdok AE, Ottensmeyer MP, Howe RD. Effects of perfusion on the viscoelastic characteristics of liver. Med Eng Phys 2006;39:2221–31.

[28] Cox M, Gawlitta D, Driessen N, Oomens C, Baaijens F. The non-linear mechanical properties of soft engineered biological tissues determined by finite spherical indentation. Comput Methods Biomech Biomed Eng 2008;11:585–92.

[29] Fischer-Cripps AC. The Hertzian contact surface. J Mater Sci 1999;34:129–37.

[30] Lester SC. Manual of surgical pathology. 2nd ed. Boston: Churchill Livingstone; 2006.

Characterizing Finger Palpation in the Detection of Prostate Cancers and Abnormalities

Ninghuan Wang [a], Gregory J. Gerling [a], Reba Moyer Childress [c], and Marcus L. Martin [c]

[a] Systems and Information Engineering, [b] Nursing, [c] Emergency Medicine, University of Virginia, USA

Prostate cancer is detected in part via the digital rectal examination. Training for this hands-on exam is limited, in particular, by feedback given trainees with respect to finger palpation. This work characterizes finger palpation technique exerted by participants, using a simulator that can electronically record finger pressure on the prostate gland and abnormalities. In principal, we analyze 1) global finger movement, 2) local finger movement and 3) average finger pressure. The analysis determined that participants utilize four patterns of global finger movement (U, V, L and line), three patterns of local finger movement (tapping, vibration, smooth movement), and distinct finger pressures (in Newtons). This analysis also determined that participants who utilized certain techniques were better able to detect presented abnormalities.

INTRODUCTION

To promote early detection of prostate cancer, screening via two diagnostic tools is advised, the prostate specific antigen (PSA) blood test and the digital rectal examination (DRE) (American Cancer Society, 2007). The DRE is the clinical palpation of the prostate gland. Alarmingly, the level of clinician performance with the DRE is variable (positive predictive value is 17-34% (AMA, 2000)) and agreement between-examiners on diagnosis is low (21-40% (Hennigan, Franks, Hocken, & Allenmersh, 1990; Saslow et al., 2004)). Hands-on training is used to improve these skills, typically via either standardized patients or physical simulators. Training with physical simulators, in particular, helps to improve the detectability of abnormalities (Pilgrim, Lannon, Harris, Cogburn, & Fletcher, 1993). While physical simulators mimic abnormal cases, current models present few practice cases (4-5) and offer little feedback on examiner performance.

The Virginia Prostate Examination Simulator (VPES) was designed to address both issues (Rigsbee, Gerling, Childress, & Martin, 2007). It combines the use of rubber-like materials to simulate the feel of tissue while using a computer to generate scenarios and record data for immediate feedback and post-performance review of exam performance.

Of interest to this work is the VPES's capability to acquire data on the finger pressure exerted on the instrumented prostate. In palpation exams, clinicians are often taught to utilize proper technique. In clinical breast exams, the total exam time, search pattern utilized and finger palpation technique are thought to influence an examiner's accuracy (Barton, Harris, & Fletcher, 1999; Bickley, 2002). Finger pressure has also been examined for the pelvic exam (Pugh & Youngblood, 2002). Here, we analyze finger palpation utilized in the DRE and determine if underlying characteristics impact abnormality detection.

METHODS

To inform the analysis, we utilized data from our previous experiment with 25 medical students. Our three analyses of finger palpation considered: 1) global finger movement, 2) local finger movement and 3) average finger pressure. Additionally, we considered how patterns employed within each (of 1-3) impacted one's ability to detect abnormalities.

Apparatus

The VPES (Fig.1, A-E) utilizes three instrumented prostates (Fig.1, F) of accurate size and stiffness attached to a track system (Fig.1, G) internal to a constructed posterior section (Rigsbee, Gerling, Childress, & Martin, 2007). The simulator includes three features not in current physical simulators: multiple and reconfigurable scenarios of graded difficulty, technique and performance feedback, and physiologically accurate anatomy.
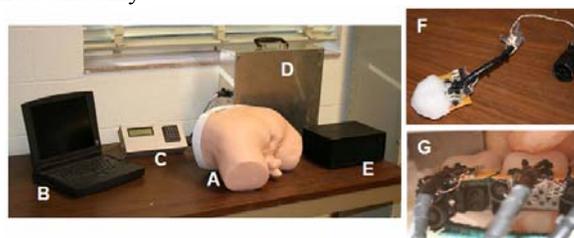
Figure 1: Main Components of VPES Apparatus, including (A) torso, (B) laptop, (C-D) automatic balloon inflation, (E) pressure sensor signal conditioning, (F) instrumented prostate and (G) internal track system.

To enable multiple and reconfigurable scenarios of graded difficulty, 4 to 6 balloons are embedded in each instrumented prostate. Basically, water-filled balloons are used to simulate palpable abnormalities. Attached to the balloons, sensors monitor the water pressure. This mechanism helps enable the performance feedback as the water pressure can denote whether or not an examiner has palpated an abnormality.
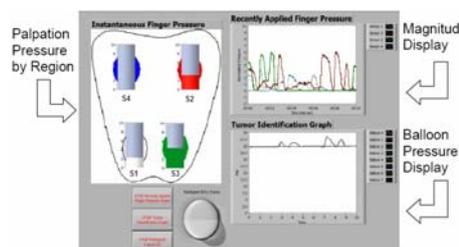
Figure 2: Real-time Visual Feedback Display

For example, when an examiner applies finger pressure to a water-filled balloon, the change in water pressure is displayed on the feedback screen (Figure 2, "Balloon Pressure Display," where the spike indicates that a balloon has been pressed). While the display offers information in the training mode, the data is also logged to a file for post-exam analysis.

To enable the feature of technique feedback, force sensors are embedded in each instrumented prostate. The four Flexiforce force sensors (Tekscan, South Boston, MA, 0-1 lb range) record the location and magnitude of applied finger pressure during palpation. The instantaneous magnitude of finger pressure exerted on the prostate is visually displayed via colored tanks that correspond to pressure sensor locations in four regions of the prostate (Figure. 2, "Palpation Pressure by Region"), and via continuous pressure plots in an analog view (Figure. 2, "Magnitude Display"). Finger force is also logged to a file for post-exam analysis.

Analysis Techniques

Overview and Rationale. In each analysis, characteristics of finger palpation technique were related with ability to detect abnormalities. In analysis 1, global finger movement was examined. We define global finger movement as tracking an examiner's finger in procession from one of the four quadrants of the prostate to another. A similar idea applies in the clinical breast exam, where global finger patterns (e.g., radial, clock, and vertical strip) are taught, ensuring that the entire breast is examined (Bennett et al., 1990). In analysis 2, local finger movement was considered. We define local finger movement as palpation by the finger within a single quadrant or near a single abnormality. Similarly, superior technique in a breast exam entails the use of the pads of the 2nd, 3rd, and 4th fingers held together, making dime-sized circles. In analysis 3, we computed average finger pressure employed near an abnormality.

Participants and Procedure of Experiment. All 25 subjects were novice medical students in their 3rd or 4th year. The overall procedure was a series of pretests, training session, and posttest-2 always conducted on the VPES (scenario described in the following section). Only the posttest-2 results are examined here. Also, while 36 subjects participated in the prior experiment, the participant number reduced to 25 due to a technical issue. This smaller dataset is believed to broadly characterize the entire subject pool.

Experimental Testing Scenario of Posttest-2. For the posttest-2 scenario analyzed, a carcinoma disease state was simulated via the inflation of two balloons (denoted by circles "1" and "2" in Figure 3).
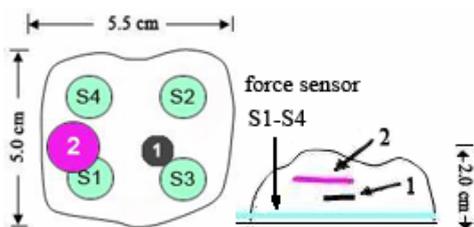
Figure 3: Balloon and Sensor Layouts in a Simulated Prostate: Top and Side views. S1-S4 are force sensors and 1, 2 are balloons.

Balloon 1 is sized 0.5 cm × 0.5 cm and balloon 2 is sized 1.0 cm × 1.0 cm. Both of them were inflated to 25 PSI to generate Shore A durometer hardness of approximately 25.54 and 29.92, respectively. Moreover, balloon 1 is positioned slightly deeper than balloon 2. See (Hall, Roehrborn, & McConnell, 1996; Jarvis, 2004) for details on the prostate gland and disease states, including carcinoma.

Data and Data Collection. The analyses utilized three sources of data: a) finger pressure from force sensors ("S1"-"S4" in Figure 3), b) balloon pressure ("1" and "2" in Figure 3), and c) participant reporting with the experiment's proctor to indicate if an abnormality was palpated, its size and location.

Force sensor data (a) were collected from the four sensors at a rate of 60 samples per second over the duration of the exam (approximately 30 seconds). Then, the sensor output (s_t^i), in Newtons for the ith sensor at time t, was normalized.

$$s_t^i = \left\| s_t^i \right\| \qquad (1)$$

Additionally, we eliminated data below 1.50 Newton, as these were below a significant magnitude of exerted pressure, and so were set to zero, via equations 2 and 3.

$$\text{if } s_t^i < 1.50, \ s_t^i = 0 \qquad (2)$$

$$\text{delete } i^{th} \text{ data if } \bigcap_{i=1}^{4} s_t^i = 0 \qquad (3)$$

Balloon sensor data (b) were collected from water pressure transducers attached to the two balloons inflated for the experimental scenario above. The data were sampled at a rate of 3 samples per second. The data were converted from volts to psi and then converted from psi (p) into Shore A durometer hardness (h), via equation 4 for balloon 1 and equation 5 for balloon 2, which were achieved by our measurement.

$$h_{balloon1} = 0.3067p + 17.867 \qquad (4)$$

$$h_{balloon2} = 0.3767p + 20.5 \qquad (5)$$

The third source of data was collected via paper records charting by the proctor (c). Following the exam, a participant indicated if he or she palpated any abnormalities, specifically its location and size. These data are used to determine if any balloon had been palpated. While we can also determine if a tumor has been palpated by looking at the water pressure data, it is possible that a participant can press on a balloon and not know it, or vice versa. Therefore, using the paper charts gives us greater confidence in the correctness of participant responses.

An example set of data collected from force and balloon sensors are plotted together in Figure 4. Spikes in the data ($Spike_b$) indicate if a balloon was palpated. Equation 6 is used to determine if the current balloon pressure h_{cur} has exceeded an increase of 3% compared to the previous measurement h_{pre}.

$$Spike_b = ((h_{cur} / h_{pre}) \geq 1.03) \qquad (6)$$

Hence, whenever balloon 2 is palpated (an upward spike; circle "A" in Figure 4), it is evident the magnitude of finger pressure for sensors 1 and 4 increases while that applied to sensor 2 decreases (circle "B"). This makes sense given the layout of the sensor numbers and balloon 2 in Figure 3.
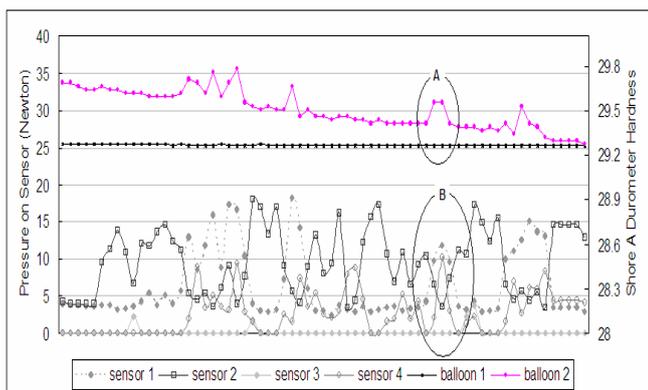
Figure 4: Plot for Sensor and Balloon Data

EXPERIMENTS AND RESULTS

Global Finger Movement (Analysis 1)

For analysis 1, the goal was to determine if any patterns of global finger movement are applied systematically. In our analysis of the data, four main patterns (U, V, L an d line) were identified and defined (Table 1). Deviations within each pattern were also defined so that, for example, the V pattern could be conducted in a variety of orientations.

Table 1: Finger Pattern Categories

As noted in the previous example concerning Figure 4, when the exerted pressure in b alloon 2 increases, the force exerted on sensors 1 and 4 increases while sensor 2 decreases. This happens each time balloon 2 is palpated. As time progress, when the water pressure in balloon 2 returns to its average value, the force exerted on sensor 2 increases again while sensors 1 an d 4 d ecreases. Furthermore, since t he participant does not exert pressure on sensor 3 nearly as often and because the water pressure in balloon 1 does not indicate a spike, we could infer that the participant's finger traversed back and forth among sensor 2, sensor 1, and sensor 4 in a "V(a)" pattern.
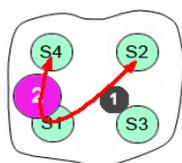
Figure 5: V(a) Global Finger Pattern

The number of participants who employed particular patterns, along with the number of participants in each respective category who detected either or both abnormalities, is shown in Table 2, where rate is th e percentage of those participants who successfully reported at least one balloon.

Table 2: Statistics of Patterns of Global Finger Movement Employed and Abnormality Detections Based on Pattern Employed.

Pattern	Number of participant	Number of participant to detect abnormalities (rate=%)

	Balloon 1	Balloon 2	Both
U pattern	11	1(9.09%)	4(36.36%)	6(54.55%)
V pattern	5	0	4(80%)	1(20%)
L pattern	3	2(66.66%)	1(33.33%)	0
Line pattern	5	1(20.00%)	0	0
Other	1	0	0	0

Table 2 results demonstrate that 11 participants exerted a U pattern, 5 applied a V pattern and 3 used an L pattern (19 of 25). These three patterns were most associated with the detection of an ab normality. People who used one of three patterns achieved 100% rate to report at least one abnormality. Second, those who employed either the U o r V p attern were the only to detect both abnormalities. Third, those who employed the U pattern detected the greatest number of both abnormalities (6 of 11). Four th, of t he 6 participants who used either a line or other pattern, only one detected an abnormality. Finally, balloon 2 was more easily detected than balloon 1 (16 people found balloon 2, wh ile 11 people identified balloon 1).

Local Finger Movement (Analysis 2)

For analysis 2, the goal was to determine if any patterns of local finger movement were applied systematically, where local finger movement is defined as palpation by finger movement within a sin gle quadrant or near a single abnormality. In an analy sis of local finger movement, we identified three patterns as tapping, vibration and smooth. Figure 6 displays characteristic features of each pattern, where data originate from force sensors alone.
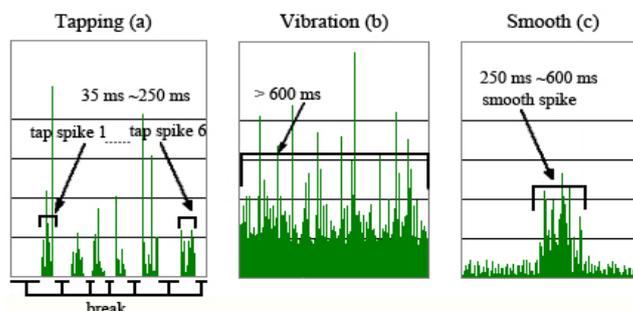
Figure 6: Examples of the Three Patterns of Local Finger Movement

Objective criteria were developed to identify these three patterns. A spike in th e local finger pattern is d efined as $Spike_s = (s^i_t > 0)$, i.e., pressure exerted above 0 N.

1) A tapping pattern is defined when an examiner strikes briefly on the defined area, before quickly pulling up hi s finger, and doing this in an in termittent fashion. Quantitatively, this pattern is identified (Figure 6(a)) by a period of continuous spikes (i.e., spikes of > 0 N) of 35-250 msec, followed by a sequence of no spikes (i.e., spikes of = 0N) that occurs for a period of at least 35 msec. Within a time span of ~35 msec, up to two spikes could have appeared, given sufficient finger pressure was exerted.

2) A vibration pattern is defined when an ex aminer maintains pressure on the prostate over a continuous time span of at least 600 msec. Th e pressure may not necessarily be constant, but rather exhibit oscillatory features (Figure 6(b)). The participant's behavior therefore is likely that of

maintaining finger contact with a single area but shifting the weight of the finger (e.g., proximally-distally or laterally-medially). Exact frequencies of oscillations are not analyzed.

3) A smooth pattern is identified when an examiner's finger moves smoothly through a defined area. It is similar to the tapping pattern, except that the period of spikes in either case differs (duration of 250-600 msec for the former compared to 35-250 msec for the latter). Moreover, in the smooth pattern, there is no break between these periods of spikes, rather a sustained, low magnitude pressure. The smooth pattern likely represents the continuous movement of one's finger across a sensor (low magnitude spikes observed when the finger is far from the sensor's center, higher magnitude spikes finger passes over the sensor's center). With tapping or vibration patterns, the examiner likely transitions out of the global finger movement mode due to the local detection of a balloon. In contrast, the smooth pattern may represent an examiner having not detected a balloon to focus more closely upon.

Given these patterns, we compared data over the period of the entire exam (approximately 30 sec) when a participant's finger was located in the vicinity of a balloon. First, given the above criteria, we inferred each examiner's dominant pattern. Second, by looking at his charted responses, the relationship between pattern and performance in reporting abnormalities was established. Since balloon 1 is located between sensors 2 and 3, while balloon 2 is placed near sensors 1 and 4, we separated sensors into two pairs (S2-S3) and (S1-S4). Then, we developed seven "combined" patterns (Table 3). Each of them is made up of two patterns used near one pair of sensors (e.g., a smooth-vibration pattern indicates a smooth pattern for sensor 2 and vibration pattern for sensor 3 in Pair 1, or vise versa).

All 25 participants palpated the area over both pairs of sensors. The results are shown in Table 3, where "B1" and "B2" indicate the number of detections for each respective balloon, and "success rate" is the percentage of participants who successfully reported a balloon given a combined pattern.

Table 3: Statistics on Combined Patterns of Local Finger Movement Employed and Rate of Abnormalities Detected per Pattern.

Patterns	Pair 1 (S2-S3)			Pair 2 (S1-S4)		
	# patterns	# B1 detections	Success Rate	# patterns	# B2 detections	Success Rate
Smooth-Vibration	14	11	78.57%	11	11	100%
Vibration-Vibration	0	0	0	2	2	100%
Vibration-Untouched	3	0	0	5	3	60%
Smooth-Smooth	4	0	0	1	0	0%
Smooth-Untouched	3	0	0	4	0	0%
Tapping-Tapping	1	0	0	1	0	0%
Untouched-Untouched	0	0	0	1	0	0%

Table 3 indicates that near balloon 1 (the smaller and deeper of the two abnormalities), 14 of 25 participants used a smooth-vibration pattern on sensors 2 and 3, with a detection success rate of 78.57%. Of the 11 participants who utilized another technique, none detected balloon 1. Near balloon 2 (the larger abnormality located nearer the surface), 11 of 25 participants employed the smooth-vibration pattern

respectively on sensors 1 and 4, with a detection success rate of 100%. Those who employed patterns of vibration-vibration demonstrated a detection success rate of 100% while the vibration-untouched pattern produced a 60% rate. Furthermore, across both balloons, participants who did not employ a vibration pattern did not detect an abnormality.

Average Finger Pressure (Analysis 3)

Sub-methods. Similar to Analysis 2, the sensors were separated into two pairs (Pair 1 = sensors 2 and 3, which reflect the pressure near balloon 1 and Pair 2 = sensors 1 and 4, which reflect the pressure near balloon 2.

First, average finger pressure for each sensor s^i_{avg} was computed using equation 7, for sensors i = 1, 2, 3, 4, where s^i_t is the finger pressure data (in Newtons) at time t for sensor i.

$$s^i_{avg} = \frac{1}{T} \sum_{t=1}^{T} s^i_t \text{, subject to } s^i_t \neq 0 \tag{7}$$

Then the average pressure for each pair was computed.

$$Pair_1 = avg(s^2_{avg}, s^3_{avg}) \tag{8}$$

$$Pair_2 = avg(s^1_{avg}, s^4_{avg}) \tag{9}$$

Results. The average finger pressure for Pair 1 is shown in Figure 7 as 5.065 N.
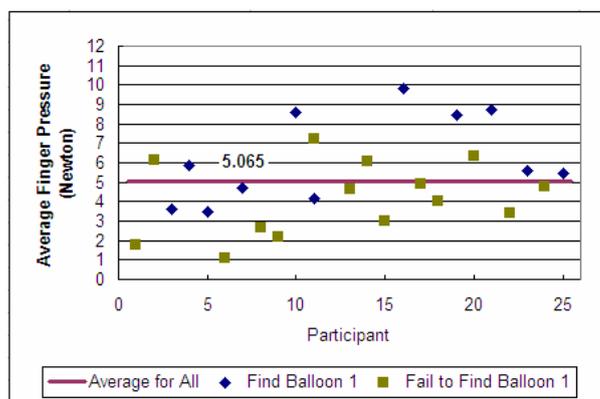
Figure 7: Average Finger Pressure per Participant near Balloon 1 (Pair 1)

Of the 11 participants who applied more than 5.065 N near balloon 1, seven (63.64%) found balloon 1. In contrast, among the remaining 14 participants who applied pressure below 5.065 N, only 4 (28.57%) found balloon 1. Moreover, those who applied greater than 8 N (4 participants) scored a 100% success rate. Additionally, a t-test was run to compare finger pressure exerted between two groups: those who successfully detected balloon 1 and those who failed to find it. The resultant p-value was 0.020. This significant difference indicates the finger pressure employed near balloon 1 significantly contributed to its detectability.

Similarly, the average finger pressure for Pair 2 is shown in Figure 8 as 3.890 N. Thirteen of 15 (86.67%) who exerted over 3.890 N near balloon 2 found the abnormality, while only 3 out of 10 (30%) located it when applying pressure below 3.890 N. Furthermore, the average finger pressure for most participants ranged from 3.50 to 5.0 N. Similar to the

analysis of Pair 1 data, a t-test compared the pressure exerted by those who succeeded in detecting balloon 2 and those who failed to detect balloon 2. The resultant p-value was 0.269. This is not a significant difference. However, nearly 75% of those who exerted a greater magnitude of finger pressure found balloon 2.
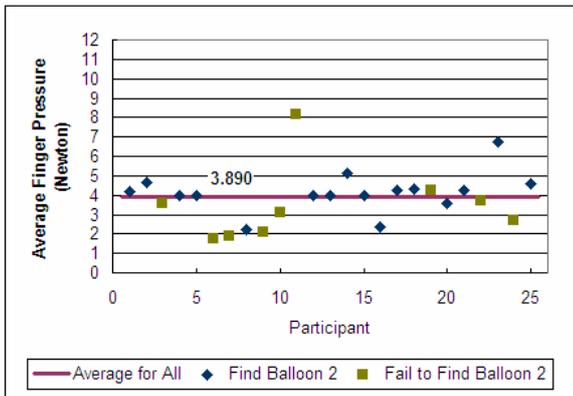
Figure 8: Average Finger Pressure per participant near Balloon 2 (Pair 2)

A final t-test compared finger pressure exerted on Pair 1 with finger pressure exerted on Pair 2, regardless of successful or unsuccessful balloon detection. The p-value obtained is 0.033. This finding indicates that the finger pressure used to detect balloon 1 and that used to detect balloon 2 were statistically different. That is, greater pressure was used to locate balloon 1.

DISCUSSION

In addition to characterizing patterns, the work here positively relates these with ability to detect abnormalities. Under our selected testing scenario, the results indicate that:

1) For global finger movement, examiners who employed a U, V or L pattern correctly reported a greater number of abnormalities (100% success rate) than those who exerted a line pattern (20%). Among those three patterns, the U pattern yielded the highest percentage of success (54.55%) in detecting both of abnormalities. This may be explained by the fact that those who utilize this pattern tend to palpate a greater amount of surface area, leading to a higher possibility of detection.

2) For local finger movement, among examiners who successfully identified at least one abnormality, all employed a derivative of the vibration pattern. When comparing the results for pairs 1 and 2, the addition of the smooth pattern to the vibration pattern enhances the rate of successful detection. We believe that the addition of the smooth pattern, rather than two vibration patterns, indicates that the trainee is transiting from global finger movement to local finger movement. In other words, the trainee has initially identified something in his global search and is now transitioning into a closer evaluation of an anomaly via a local search.

3) For average finger pressure, greater exerted pressure may be required to detect deeper and smaller balloons. For example, those who found balloon 1 (smaller, deeper balloon) utilized greater pressure (5.065 N) than those who found

balloon 2 (3.890 N). And the difference between pressures applied in the vicinity of balloon 1 versus balloon 2 is significant (p = 0.033 between Pair 1 and Pair 2). Size and depth of the balloon may be the most reasonable factors to explain the need for increased finger pressure because balloon 1 is $1/4^{th}$ the size of balloon 2 and nearly twice the depth. Moreover, a greater pressure on Pair 1 increases the possibility of detecting balloon 1 (p = 0.020). However, for Pair 2, greater pressure did not boast such a contribution to identify balloon 2 (p = 0.269). This may indicate that pressure may not be a dominant factor in detecting larger and shallower balloons, while it would be for smaller and deeper balloons.

CONCLUSION

In this paper, we analyzed three techniques of finger palpation: global finger movement, local finger movement and average finger pressure. Additionally, we examined the correlation between the utilization of certain palpation techniques and performance in detecting abnormalities. We find that participants who utilize certain techniques (a U global pattern, a vibration-smooth local pattern and greater finger pressure) are better able to detect presented abnormalities. However, our future work will examine if these trends remain over a breadth of scenarios.

REFERENCES

AMA. (2000). *Screening and Early Detection of Prostate Cancer* from http://www.ama-assn.org/ama/pub/category/13604.html#combined_use.

American Cancer Society. (2007). Prostate Cancer. 2007, from http://www.cancer.org/docroot/CRI/CRI_2_1x.asp?dt=36

Barton, M. B., Harris, R., & Fletcher, S. W. (1999). Does This Patient Have Breast Cancer? The Screening Clinical Breast Examination: Should It Be Done? How? *JAMA, 282*, 1270-1280.

Bennett, S. E., Lawrence, R. S., Angiolillo, D. F., Bennett, S. D., Budman, S., Schneider, G. M., et al. (1990). Effectiveness of Methods Used to Teach Breast Self-Examination. *American Journal of Preventive Medicine, 6*(4), 208-217.

Bickley, L. S. (2002). *Bates' Guide to Physical Examination & History Taking* (8th ed.). Philadelphia: Lippincott Williams & Wilkins.

Hall, M. C., Roehrborn, C. G., & McConnell, J. D. (1996). Is screening for prostate cancer necessary in men with symptoms of benign prostatic hyperplasia? *Semin Urol Oncol, 14*(3), 122-133.

Hennigan, T. W., Franks, P. J., Hocken, D. B., & Allenmersh, T. G. (1990). Rectal Examination in General-Practice. *British Medical Journal, 301*(6750), 478-480.

Jarvis, C. (2004). *Physical Examination & Health Assessment* (4th ed.). Mo: Saunders.

Pilgrim, C., Lannon, C., Harris, R. P., Cogburn, W., & Fletcher, S. W. (1993). Improving Clinical Breast Examination Training in a Medical-School - a Randomized Controlled Trial. *Journal of General Internal Medicine, 8*(12), 685-688.

Pugh, C. M., & Youngblood, P. (2002). Development and validation of assessment measures for a newly developed physical examination simulator. *Journal of the American Medical Informatics Association, 9*(5), 448-460.

Rigsbee, S., Gerling, G. J., Childress, R. M., & Martin, M. L. (2007). A computerized physical simulator for training clinical prostate palpation skills. *Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference*, 2223-2227.

Saslow, D., Hannan, J., Osuch, J., Alciati, M. H., Baines, C., Barton, M., et al. (2004). Clinical Breast Examination: Practical Recommendations for Optimizing Performance and Reporting. *CA Cancer J Clin, 54*(6), 327-344.

Psychophysical Detection of Inclusions with the Bare Finger amidst Softness Differentials

Leigh A. Baumgart, Gregory J. Gerling, and Ellen J. Bass

Department of Systems and Information Engineering, University of Virginia, USA

ABSTRACT

Softness discrimination and the detection of inclusions are important in surgery and other medical tasks. To better understand how the characteristics of an inclusion (size, depth, hardness) and substrate (stiffness) affect their tactile detection and discrimination with th e bare finger, we conducted a psychophysics experiment with eighteen participants. The results indicate that within a more pliant substrate (21 kPa), inclusions of 4 mm diameter (20 mm^3 volume) and greater were consisten tly detectable (above 75% of th e time) but only at a depth of 5 mm. Inclusions embedded in stiffer substrates (82 kPa) had to be twice that volume (5 mm diameter, 40 mm^3 volume) to be detectable at the same rate. To analyze which tactile cues most impact stimulus detectability, we utili zed logistic regression and genera lized estimating equations. The r esults indicate that substrate stiffness most contributes to inclusion d etectability, while the s ize, depth, and hardness of the stimulus f ollow in ind ividual importance, respectively. The results seek to aid in the development of clinical tools and infor mation displays and more accu rate virtual haptic environments in discrimination of soft tissue.

KEYWORDS: Softness discrimination, inclusion detection, psychophysics, tactile perception, medical simulation.

1   INTRODUCTION

Inclusion detection, or the detection of one bo dy (a stimulus) within another, is important in many environments, particularly clinical surgery (e.g., cauterization during surgery, the removal of cysts using surgical scissors) and palpation exams (e.g. breast and prostate cancer screenings). It is necessary to understand how the characteristics of the substrate in which the stimulus is embedded and dimensions of the stimulus itself affect tactile perception. This knowledge can inform the developm ent of appropriate clin ical tools and information displays [1-2].

With the adv ent of m inimally invasive surgery, teleoperation, and virtual-environment applicati ons, there is a growing interes t in how to pres ent human operators with natural tactile and haptic experiences, particularly those experienced through a sheath o r other form of constraint [3-6] . Many of these devices involve the interaction of a rigid body (probe or indenter) with an elastic body (tissue). In thes e cases, such as during laparos copic surgery, the operator feels the shaft of the in strument and therefore uses cues

such as its angle, the force at its tip, and any vibrations in the shaft. For this class of inter action, psychophysical research has focused on tex ture perception or discrimination of absolu te stiffness [7] in order to dete rmine how to build sim ulated environments or appropriate virtual feedback devices.

Another class of inter actions involves th e bare finger and another elastic bod y (tissue). In fact, some robotic devices are now being developed to incorp orate tactile sensors that m imic human perception [8]. However, there has been less focus on the underlying psychophysical experiments, modeling, and simulation that surround bare finger in teraction with soft objects . Most research in this area has focused on the discrimination of softness [9-10], roughness [11], and objects through a glove [12]. Other analysis of perception with the bare finger surrounds the modeling of neural r esponses (both single-units and populations o f mechanoreceptors in fingertip skin) [13-14].

The objective of the work presented h ere is to conduct a psychophysical experiment to characterize the factors that surround inclusion detection. The overall goal is to und erstand how varying the characteristics of th e substrate and inclusion affect tactile perception in a constrained environment.

2   METHODS

We conducted an experiment using a modified method of constant stimuli with 18 participants and inclusions that varied in size, depth, and hardness that were placed within s ubstrates of tw o stiffness levels. The participants' task was to palpate th e substrate and indicate if the stimulus is present. The objectives of the study were to determine a) the size of inclusions consistently detected at three discrete depths, b) how substrate stiffness impacts detectability, c) how inclusion hardness impacts detectability, and d) how each o f the vari ables differentially impact inclusion detection.

2.1   Apparatus

An apparatus was built speci fically for th is study. The main components were a hole to constrai n the finger, a thin sheath, and twenty-three cylindrical substrates (30 mm diameter and 20 mm tall) mounted onto a round platform. The substrates were made of silicone-elastomers of two stiffness levels (21 and 82 kPa) tha t simulate the fe el of interior human prostate tissue and other internal organs. The platform was housed w ithin a structure that restricted viewing of th e substrates and included the finger hole angled at appro ximately 110 d egrees from the participan t. A sheath made of silicone-elastomer (2 mm thick, 180 kPa) was attached to the hole. This constrained the finger at the knuckle and extended beyond the length o f the fing er. The platform of substrates could be rotated so th at only the substrate und er test was located beneath the sheath.

Each cylindrical substrate in cluded a single poly ethylene balloon embedded at one of three depths: 5, 10, and 15 mm. Balloons of seven volumes were used: 20, 40, 80, 200, 470, 1060, and 1770 mm^3 that correspond to diameters of 4.0, 5.0, 7.5, 10.0, 15.0, 17.0, and 20.0 mm, respectiv ely. The balloons were filled with water, thereby controlling hardness of the stimulus. Balloons

Gregory J. Gerling, Ph.D.
The University of Virginia
Department of Systems and Information Engineering
151 Engineer's Way, P.O. Box 400747
Charlottesville, VA 22904
gg7h@virginia.edu

could be inflated to be hard, lik e a rock, but we re not detectable when deflated. In this study, three hardness levels were used: 23, 27, and 31 durometers, type Shore A.

2.2 Participants

Ten male and eight female participants (mean age = 20.4 years, SD = 1.4) were enrolled in the human-subjects study, approved by the IRB at t he University of Virginia. A dem ographic questionnaire indicated that no participant h ad any remarkable prior experience working with his or her hands.

2.3 Procedure

Using a modified version of the ps ychophysical method of constant stimuli [15] , participants palpated the substrates to determine the d etectability of s timuli. Typically the method of constant stimuli employs stimulus and noise only trials presented in a randomized fashion wher e all stimulus combinations are presented an eq ual number of times. However, here we made three modifications to reduce the number of trials and thereby participant fatigue. First, only 31 of the possible 42 combination s of independent variable levels (stimulus size, depth, hardn ess and substrate stiffness) was presented to participants (Table 1) as pilot testing identified combinations that were a lways or never detectable and therefore were r emoved from th is study. Second, the number of times that each stimulus combination was presented varied depending on the difficulty of detecting the stimulus in the pilot study. The most difficult to detect were presented four times, while the easiest to detect were presented twice and the rest were presented three times (Table 1). Third, to address potential trial order issues, particip ants were presented with stimuli and noise trials in one of six random orders.

Every participant completed two, 90-min experimental sessions, held on separate days. Before the first experimental trial, participants had a h ands-on practice session with f eedback concerning whether or not the s ubstrate contained an inclus ion. They were ins tructed to achieve a cons istent search technique: move one's finger across th e substrate surface in lin es parallel with and then p erpendicular to the participant's seated position . When traversing these line paths, participants were ins tructed to use small, dime-sized circular motions. During train ing and the experimental trials, the proctor monitored the pressure exerted on the substrates to ensure that finger pressure remained within 4 and 6 N and that the quadrants were palpated in the specified order. When participants deviated from the specified pressure or palpation order, they were reminded of the appropriate technique.

During the ex perimental trials, participants palpated 192 substrates, half of which con tained a stimulus (the balloons wer e not inflated for the others). Pa rticipants were given 20-sec to examine a single substrate. After each trial, participants informed the proctor (via "y es" or "no") as to whether a s timulus was present and th e proctor noted the response. P articipants were given a 10-sec break between subsequent tr ials, a 60-sec break after every 10-17 trials and a 5 -minute break after ev ery 32-42 trials.

Table 1. Stimulus trials per participant.

Stimulus dimensions			Number of trials at 82 kPa stiffness	Number of trials at 21 kPa stiffness
Size (dia mm)	Depth (mm)	Hardness (durometers, Shore A)		
4	5	31	4	2
4	10	31	2	4
5	5	31	3	2
5	10	31	3	3
7.5	5	31	3	-
7.5	10	31	3	2
7.5	15	31	-	4
10	5	31	2	-
10	10	23, 27, 31	4, 4, 4	4, 4, 2
10	15	23, 27, 31	-, -, 2	4, 4, 3
15	10	31	3	-
15	15	23, 27, 31	4, 4, 4	-, -, 2
17	10	31	2	-
17	15	31	3	-
20	15	31	2	-

2.4 Data Analysis

In seeking to identify general relationships between th e independent variables (stimulus size, dep th, hardness an d substrate stiffness) and inclusion detectability, graphical analysis was used to address objectives a-c. Logistic regression was used to address objectiv e d. For th e logistic regression, we used generalized estimating equations or GEE, [16]. Substrate stiffness was considered both a fixed and random effect and stimulus size, depth, and hardness as quantitative and continuous variables. Size was coded as 1 -7 (4 mm – 20 mm); depth was coded as 1 (15 mm), 4 (10 mm), and 7 (5 mm); hardn ess was coded as 1 (23 durometers), 4 (27 durometers), and 7 (31 d urometers); and substrate stiffness was coded as 1 (82 kPa) and 2 (21 kPa).

For logistic regr ession, one response per partic ipant for each combination of independent variables was needed. Ther efore, if participants detected at le ast half of the s timuli for a s pecific combination of variables, the response was con sidered a "yes" (coded as 1) and if less than half of the stimuli were detected, the response was considered a "no" (coded as 0).

Statistical analyses were conducted using R . For the logistic regression results, the *geeglm* function for GEE from the *geepack* package was used.

3 RESULTS

In general, deeper and smaller stimuli were m ore difficult t o detect at both s ubstrate stiffness levels and stimuli were more likely to be de tected in the 21 k Pa substrate compared to the 82 kPa substrate. To help illustr ate the im pact of s ize and d epth on the percent detected, Figure 1 depicts the relationship betw een detection percentage and stimulus size at each of the three depths for the 82kPa stiffness (top) and the 21kPa stiffness (bottom).

Figure 3. Detection percentage as a function of stimulus hardness for a 15 mm diameter, 15 mm deep stimulus in the 82 kPa substrate. Standard error bars are shown around data points.

Figure 4. Detection percentage as a function of stimulus hardness for a 10 mm diameter, 15 mm deep stimulus in the 21 kPa substrate. Standard error bars are shown around data points.

Figure 1. Detection percentage as a function of stimulus size and depth at two stiffness levels – 82 kPa (top) and 21 kPa (bottom). Standard error bars are shown around data points.

The detection rate for larger inclusions in stiffer substrates is similar to smaller inclusions in more pliant substrates, given equal depth (Figure 1). For ex ample, at a depth of 15 mm, 17 mm inclusions in the 82 kPa substrate were detected at approximately the same percentage as 10 m m inclusions the 2 1 kPa substrate (80% and 75% respectively).

The impact of h ardness can be more readily seen for particular size-depth combinations. Figure 2 illustrates the result of varying the hardness of a 10 mm diameter stimulus at a 10 mm depth, for both the 21 and 82 kPa substrates. At the lowest hardness (23 durometers, type Shore A) the percent d etected in the s tiffer substrate (82 k Pa) drops to 75% from 100%. In con trast, the percent detected in th e more pliant substrate (2 1 kPa) was not affected by changes to stimulus hardness. Similarly, in the other two cases where hardness was varied, the percent detected was impacted more in the 82 kPa substrate compared to the stimulus in the 21 kPa substrate (Figures 3 and 4).

Figure 2. Detection percentage as a function of stimulus hardness for a 10 mm diameter, 10 mm deep stimulus for both substrate stiffness levels. Standard error bars are shown around data points.

The following logistic regr ession model was used to determine how the independent variables impacted detection of the stimulus, where p is the percent of stimuli detected.

$$\log(p/(1-p)) = \qquad\qquad (1)$$
$$\beta_0 + \beta_1(\text{stiffness}) + \beta_2(\text{size}) + \beta_3(\text{depth}) + \beta_4(\text{hardness})$$

Table 2 d epicts the ou tput of the model using stiffness as a random effect. The output ind icates that all variables significantly impacted detection of the inclusion. In addition, the values of the coefficients indicate that stiffness of the substrate most contributes to detectability, then si ze, then depth and finally stimulus hardness. The results ar e presented using α = 0 .05 for significance.

Table 2. First order coefficients for the main effects model.

	Coefficient	SE	Wald St.	Pr(>|W|)
Intercept	-10.41	1.51	47.33	**5.99e-12**
Stiffness	2.16	0.41	28.33	**1.02e-07**
Size	1.29	0.17	58.62	**1.91e-14**
Depth	1.09	0.14	60.74	**6.55e-15**
Hardness	0.18	0.065	7.60	**0.0058**
Correlation Parameter	0.12	0.067		

Tables 3-4 show the model out puts from a sub- group analysis with substrate stiffness considered as a fixed effect. From the sub-group model o utputs, we can see that all stimulus variables significantly impacted detection of the stimulus in the 82 kPa substrate. However, in the more pliant substrate (21 kPa) stimulus, inclusion hardness was not a significant factor in detection. Further, we see in the 82 kPa model, the coefficients for size and depth are comparable, while in the 21 kP a model, s ize has a greater impact on detection.

Table 3. First order coefficients for each stimulus dimension in sub-group model for 82 kPa stiffness.

	Coefficient	SE	Wald St.	Pr(>|W|)
Intercept	-8.43	1.32	41.07	**1.5e-10**
Size	1.25	0.17	54.89	**1.3e-13**
Depth	1.13	0.16	47.71	**4.9e-12**
Hardness	0.24	0.08	8.92	**0.0028**
Correlation Parameter	0.07	0.08		

Table 4. First order coefficients for each stimulus dimension in sub-group model for 21 kPa stiffness.

	Coefficient	SE	Wald St.	Pr(>|W|)
Intercept	-7.50	1.87	16.10	**6.0e-05**
Size	1.75	0.40	18.68	**1.5e-05**
Depth	1.24	0.20	37.70	**8.2e-10**
Hardness	0.15	0.11	1.76	0.18
Correlation Parameter	0.20	0.34		

An analysis of two factor interaction effects was completed in two steps. First intera ction effects to consider were identified. Each two-way interaction was added one at a time to the m ain effects model. Using this method, the size-depth, depth-stiffness, and hardness-stiffness interactions effects were significant. However, when the statistical analysis was conducted with these three interaction effects along with the main effects, only the size-depth and stiffn ess-hardness interactions were signi ficant (Table 5). The values of the coefficients indicate that stiffness of th e substrate still most contributes to detectability, then size, then depth, then stimulus hardness , and then the size-depth an d stiffness-hardness interactions.

Table 5. First order coefficients and selected interactions.

	Coefficient	SE	Wald St.	Pr(>|W|)
Intercept	-11.20	2.10	28.46	**9.57e-08**
Stiffness	3.54	0.87	16.57	**4.70e-05**
Size	0.90	0.20	20.32	**6.53e-06**
Depth	0.72	0.27	7.20	**0.007**
Hardness	0.54	0.17	10.75	**0.001**
Size-Depth	0.18	0.05	13.39	**0.0002**
Stiffness-Hardness	-0.23	0.09	7.05	**0.008**
Stiffness-Depth	-0.06	0.11	0.31	0.57
Correlation Parameter	0.12	0.08		

4 DISCUSSION AND CONCLUSION

This study sought to understand the effects of substrate stiffnes s and inclusion characteristics on tactile perception in a constrained environment, representing a clinical setting. Through an elasti c sheath, participants palpated 31 different combinations of substrate stiffness and stimulus size, depth, and hardness.

Within a m ore pliant substra te (21 kPa), inclus ions of 4 mm diameter (20 mm^3 volume) and greater were detected above 75% of the time, but only at a depth of 5 mm. Inclusions embedded in stiffer substrates (82 kPa) h ad to be twice that volume (5 mm diameter, 40 mm^3 volume) to be detected at the same rate.

Substrate stiffness most impacted inclusion detection in all cases while the interaction of stiffness and hardness play ed a role in detection, particularly in the less pliant substrates (82 kPa). However, compared to size and depth, stimulus hardness did not play as large a role in detection and was not a signifi cant factor in the model of the more pliant substrates (21 kPa), similar to results found in [1]. This could be, in part, because th e hardness of the stimuli was not varied over a large enough range (only 23 to 31 durometers, Shore A). To give s ome perspective on the hardness of various objects in durometers, Shore A, a rub ber band is 40, a rubber shoe heel is 70, and a sho pping cart wheel is 100. It might be appropriate to go up to 60 durometers, Shore A in further studies in addition to including more substrate stiffness levels.

The results of this stud y may aid in the development of clinical tools and infor mation displays and more accu rate virtual haptic environments in discrimination of soft tissue.

REFERENCES

[1] Bloom HS, Criswell EL, Pennypacker HS, Catania AC & Adams CK. Major stimulus dimensions determining detection of si mulated breast lesions. *Perception and Psychophysics*, 32(3):251-260, 1982.

[2] Fletcher SW, O'Mailey MS & Bunce LA. Physici ans' ability to detect lumps in s ilicone breast models. *Journal of the American Medical Association*, 253(15):2224-2228, 1985.

[3] Lederman, SJ & Klatzky, RL. Haptic identification of co mmon objects: Effects of constraining the manual exploration process. *Perception and Psychophysics*, 66(4):618-628, 2004.

[4] Jansson, G & Monaci, L. Identification of real objects un der conditions similar to th ose in haptic displays: providing spatially distributed information at the contact areas is more important than increasing the number of areas. *Virtual Reality*, 9:243-249, 2006.

[5] Tan, HZ. Haptic interfaces for virtual environments: Perceived instability and force constancy in ha ptic sensing of virtual surfaces. *Canadian Journal of Experimental Psychology*,61(3):265-275, 2007.

[6] Wagner, CR, Lederman, SJ & Howe, RD. A tactile s hape display using RC ser vomotors. *Proceedings of the 10th Symposium On Haptic Interfaces for Virtual Environments & Teleoperator Systems*, Orlando, FL, 354-356, 2002.

[7] Klatzky, RL, Lederman, SJ, Hamilton, C, Grindley, M & Swendsen, RH. Feeling textu res through a probe: Effects of probe and sur face geometry and explor atory factors. *Perception and Psychophysics*, 65(4):613-631, 2003.

[8] Dargahi, J & Naja rian, S. Human tactile perception as a standard for artificial tactile sensing—a review. *International Journal of Medical Robotics and Computer Assisted Surgery*, 1(1):23–35, 2004.

[9] Srinivasan, MA & LaMotte, RH. Tactual Discrimination of Softness. *Journal of Neurophysiology*, 73(1):88-101, 1995.

[10] Friedman, RM, Hester, KD, Green, BG, & LaMotte, RH. Magnitude estimation of soft ness. *Experimental Brain Research*, 191:133–142, 2008.

[11] Yoshioka, T, Gibb, B, Dorsch, AK, Hsiao, SS, & Johnson, KO. Neural coding mechanisms underlying perceived roughness of finely textured surfaces. *Journal of Neuroscience*, 21(17):6905-6916, 2001.

[12] Gibson, GO & Craig, JC. Tactile sp atial sensitivity and anisotropy. *Perception and Psychophysics*, 67(6):1061-1079, 2005.

[13] Freeman, AW & Johnson, KO. Cutaneous mechanoreceptors in macaque monkey: temporal discharge patterns evoked by vibration, and a receptor model, *Journal of Physiology*, 323:21-41, 1982.

[14] Wheat, HE & Goodwin, AW. Tactile discrimination of e dge shape: limits on spatial re solution imposed by parameters of the per ipheral neural population, *Journal of Neuroscience*, 21:7751-7763, 2001.

[15] Gescheider GA. *Psychophysics: The fundamentals (3rd ed.).* Mahwah, NJ: *L. Erlbaum Associates*, 1997.

[16] Zeger, SL & L iang, KY. Longitudinal data analysis for discrete and continuous outcomes, *Biometrics*, 42:121-130, 1986.