

A Comparative Study of Haptic Stiffness Identification by Veterinarians and Students

Neil Forrest, Sarah Baillie, Patrick Kalita, and Hong Z. Tan, *Senior Member, IEEE*

Abstract— Palpation is an important clinical skill in both veterinary and medical health professions. The present study compares the ability of practicing veterinarians and veterinary students to identify the stiffness of virtual surfaces through palpation. An absolute identification paradigm was used where a force-feedback haptic device rendered virtual surfaces with 5 levels of stiffness within a “clinically relevant” range (0.2 – 0.5 N/mm). The mean information transfer was 0.97 bits (almost 2 perfectly identifiable stiffness levels) for 12 veterinarians and 0.58 bits (1 correctly identified level) for 14 veterinary students. Although the difference between the two groups was significant ($p < 0.001$), neither group was able to reliably identify more than 2 levels of stiffness, indicating that the success of veterinarians in clinical practice probably relies on additional properties such as size, shape and texture. Analyses of force vs. time and displacement vs. time recordings suggest that the superior performance of the veterinarians may be partially attributable to motor strategy. Specifically, veterinarians used a greater mean maximum force (2.0N) compared to students (1.6N) ($p < 0.05$). However, further studies are required to investigate motor strategy in more detail. The implications of our findings for veterinary education and quantitative skill assessment are discussed.

Index Terms—stiffness perception, stiffness identification, comparison of users, veterinary medicine, education, training

1 INTRODUCTION

IN both human and veterinary medicine, health professionals use palpation as part of many clinical examinations. When palpating a structure, the clinician uses the sense of touch to assess properties such as size, shape, texture and stiffness. The information gathered helps in the diagnostic process. Examples of palpation based examinations in human medicine include the detection of prostate and breast cancer and in veterinary medicine the diagnosis of pregnancy in several species.

Learning and teaching palpation is difficult, especially when the examination is internal and unsighted. Opportunities for trainees to practice on real patients are limited by ethical considerations and have been further reduced by rising student numbers. Additionally, the level of skill required is hard to quantify, which makes setting targets for students and assessing competence difficult. Simulators provide a potential solution to some of these issues and a number of medical and veterinary palpation simulators have been developed [2], [3], [4], [5], [6]. In the veterinary domain, The Haptic Cow [7] has been developed to teach palpation of the bovine reproductive tract. The increasing number of techniques being simulated is in-

dicative of the potential of haptics in this area, but training benefits need to be demonstrated before such simulators will be widely adopted. To this end, The Haptic Cow system has been proven to be effective at training veterinary students to locate the uterus in cows. It has been integrated into the undergraduate curriculum at the Faculty of Veterinary Medicine, University of Glasgow [8] and more recently at other veterinary schools in the UK.

In the present study we focus on one of the skills involved in palpation. When diagnosing the particular state of pregnancy in the cow, veterinarians feel for a reduction in the stiffness of the uterus associated with the presence of fetal fluid. Experienced veterinarians can estimate the gestation stage of a pregnant cow to within a few weeks or even days, an ability that untrained veterinary students do not possess until they have examined many cows. Palpation is an important skill in medical diagnosis in general, for example, when the clinician is identifying types of lump, e.g., abscess, cyst or tumor. More needs to be understood about the skills involved in palpation in order to maximize the training benefits that simulators offer. We are particularly interested in revealing the aspects of palpation that separate practicing veterinarians from veterinary students so that proper training modules can be developed to train more students in less time. To begin with, we examined a single element of palpation: judging stiffness. We sought to answer the following research question: *Is there a perceptual difference between experts and novices in terms of stiffness judgments?* By comparing the abilities of veterinarians (the “experts”) with those of veterinary students (the “novices”), we investigated if stiffness perception is affected by clinical practice. The results will be used to inform the design of future simulators. Also, by quantifying expert ability we can identify the level of skill that a student might need to achieve in

- N. Forrest is with the Royal Veterinary College, University of London, UK. E-mail: ndforrest@rvc.ac.uk.
- S. Baillie is with the Royal Veterinary College, University of London, UK. E-mail: sbaille@rvc.ac.uk.
- Patrick Kalita is with The MathWorks, Inc., Natick, MA. E-mail: patrick.kalita@gmail.com.
- H.Z. Tan is with the Haptic Interface Research Laboratory, Purdue University, West Lafayette, IN 47907. E-mail: hongtan@purdue.edu.
- Portions of this article reprinted with permission from [1], ©IEEE.

Manuscript received 22 March 2010.

order to be considered competent.

In designing our study, we were aware that the motor strategies associated with palpation may have an influence on the perceptual ability to judge stiffness. It is well known that people use different *Exploratory Procedures* (EPs) to interact with objects when seeking different types of haptic information [9]. It has also been shown, for example, that lateral scanning speed can influence perceived surface roughness using either bare fingers or rigid probes under both passive and active conditions [10], [11]. In fact, certain tasks, such as ascertaining the natural resonant frequency of an external dynamic system, require people to actively excite the system with appropriate movement patterns [12]. However, several studies have also suggested that, at least within the range of “natural movements,” people may be able to ignore the sensory input due to their own motor output [13] and that perception can remain unaffected by changes in hand movements [14]. To make the matter more complicated, our earlier studies have indicated that veterinarians often produce considerably different displacement and force magnitude patterns depending on the type and/or the stiffness of the object being palpated [15]. It is conceivable that veterinarians might bring this domain knowledge, as well as their own personal styles, to the task of palpating a surface, even when no clinical context is provided. Therefore, it was deliberately decided not to provide explicit instructions to the participants regarding the type of motions to use when judging stiffness. Instead, we decided to allow the participants to bring all their knowledge and experience (or lack thereof) to bear on stiffness identification. In this regard, our study uses controlled (clinically relevant) stiffness levels, but does not control force output patterns (i.e., motor strategy). We believe that it was important to observe the “natural” motor behavior of our participants before designing further studies that investigate the role of motor strategy in more depth.

Generally speaking, psychophysical studies can quantify stiffness perception in terms of detection, discrimination or identification [16]. In the case of detection, the ability to recognize the presence of a stimulus is measured as the Absolute Threshold, or the smallest detectable stimulus intensity. In the case of discrimination, the ability to discriminate between two stimuli is measured as the Difference Threshold, the just noticeable difference (JND), or the smallest change in the intensity of a stimulus that is noticeable. A third paradigm, absolute identification, estimates the participants’ ability to recognize stimulus values in isolation, i.e., without a reference or comparison value. In this case, given a particular type of stimulus, the maximum amount of information that the human sensory system can transmit, the information transfer (IT) or channel capacity, is determined experimentally (see [17] and [18] for a practical overview of conducting absolute identification experiments). The clinical task faced by practicing veterinarians, namely the assessment of the gestation stage of a pregnant cow, is closest in concept to the absolute identification paradigm.

Most existing studies of stiffness perception have used

a discrimination paradigm. The results are often reported as the Weber fraction, i.e., the JND divided by the reference stiffness. Weber’s law states that this ratio is a constant, suggesting that JND is proportional to the reference stiffness. The Weber fraction is reported to be 23% for the elbow joint [19], 22% for a pinch grip between the thumb and forefinger [20], and 10% for unrestricted active probing using a PHANToM stylus [21]. One previous study of stiffness perception used an absolute identification paradigm [22]. It reports an information transfer of 1.46 bits over a stiffness range 0.2–3.0 N/mm for a group of college students and researchers with no clinical experience. This translates to the reliable identification of only 2–3 stiffness levels when stiffness is judged in isolation.

The present study follows the protocol of [22] with two important differences. Two groups of participants, experienced veterinarians and inexperienced veterinary students, were tested and their performance compared. In addition, the stiffness range was chosen to be “clinically relevant” to allow the practicing veterinarians to take advantage of their domain-specific knowledge and skills. Therefore, the present study was designed to assess the perceptual differences, if any, between experts and novices in a controlled yet clinically relevant experimental setting. The stiffness identification results, in terms of information transfer, were reported earlier in a preliminary form [1].

2 METHODS

2.1 Participants

Fourteen veterinary students (9F, 5M) and 12 practicing veterinarians (7F, 5M) participated in the experiment. The students (the “novices” in the present study) were in the 3rd year of the 5 year veterinary course at The Royal Veterinary College (RVC), University of London. They were at a stage in their course just prior to beginning clinical practical experience. The veterinarians (the “experts”) had been working in veterinary practice for between 4 and 24 years (mean: 11.4 years). The study was approved by the RVC Ethics Committee and all participants were volunteers who gave written consent to be involved.

2.2 Apparatus

A force-feedback haptic device (PHANToM Premium 1.5, SensAble Technologies, Woburn, MA, USA) was used in the experiment to render a virtual surface to which a variety of stiffness values were assigned. The participant interacted with the virtual surface using the middle finger inserted in the PHANToM thimble (Fig. 1). In the context of The Haptic Cow, veterinarians favored the use of the middle finger, as they judged it to provide a more realistic experience than using the index finger [8]. The haptic device was placed inside a box and concealed from view by a curtain. The participant was seated with the arm supported by a cushioned arm rest. The PHANToM rendered a stiff constraint (see 2.3 Stimuli) that restricted movement of the fingertip to the up-down dimension (y-axis). No restrictions were imposed on the range of vertical movements the participant could make. The participant

wore headphones to eliminate possible audible cues and distractions.

2.3 Stimuli

A horizontal virtual surface (in the x-z plane of the PHANToM workspace) was simulated with the haptic device. The elastic stiffness values of the virtual surface varied from 0.2 to 0.5 N/mm. This range was representative of stiffness values that would be commonly encountered by a veterinarian during palpation. This clinically relevant range was based on values previously selected by veterinarians to represent a range of tissue types (during the development of The Haptic Cow, a validated veterinary haptic palpation simulator) [23].

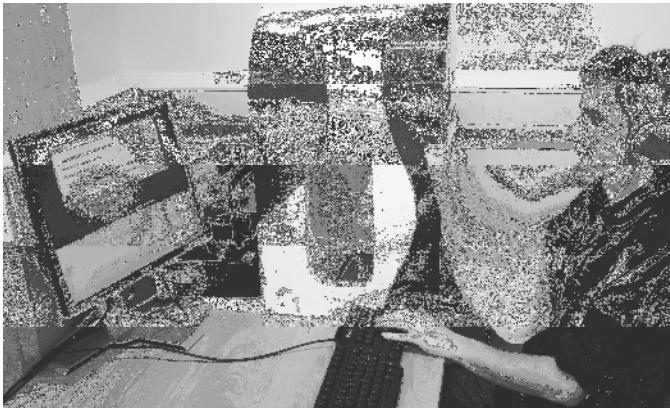


Fig. 1. The PHANToM Premium 1.5 and other apparatus as configured for the experiment. Shown on the computer screen are the instructions and a simple visualization of the virtual surface and haptic interaction point presented during the pre-experiment tutorial. No such graphical information was shown during the experiment.

Five different stiffness values were used in the present study. According to [18] the number of stimulus levels in an absolute identification experiment should be (1) higher than the expected best performance so that channel capacity can be estimated, and (2) as low as possible in order to minimize the number of trials required. In our earlier study on stiffness identification [22] where a wider range of stiffness values (0.2 – 3N/mm) was used, the best individual performance was an information transfer of 2.06 bits, or the correct identification of 4 stiffness levels. Since a smaller stiffness range (0.2 – 0.5 N/mm) was used in the present study, we expected the best performance to be less than 4 stiffness categories (see [24] for discussion on why information transfer increases with stimulus range for auditory intensity identification). Therefore, 5 stiffness levels were considered sufficient in the present study. With regard to the second consideration, it has been shown that a minimum of $5k^2$ trials are needed in order to obtain an unbiased estimate of information transfer (where k is the number of stimulus alternatives) [25]. With $k = 5$ in the present study, $5k^2 = 125$ trials, which was manageable. We chose to collect twice the minimum required number of trials per participant ($10k^2 = 250$) in keeping with our previous study on stiffness identification [22]. The 5 stiffness values were equally spaced on a logarithmic scale between 0.2 and 0.5 N/mm, based on

earlier studies suggesting that Weber's Law holds for stiffness discrimination (e.g., [20]).

The movement of the thimble was constrained to the up and down (y-axis) direction to make it easier for the participants to interact with the virtual surface. It also served to standardize the location within the haptic device's workspace at which each participant could make contact with the virtual surface. The latter was important because the characteristics of the haptic device are not uniform across the whole workspace. Preliminary testing revealed that the perceived stiffness of the virtual constraint needed to be larger than the highest stiffness level of the virtual surface. Otherwise the haptic interaction point would slip across the horizontal virtual surface while the participant tried to move it in the up-down direction. Such transverse movements would lead the participant to confuse the perceived stiffness of the constraint with the stiffness of the virtual plane. A Proportional-Derivative (PD) controller was implemented to achieve a sufficiently hard constraint without destabilizing the haptic device (stiffness=0.95 N/mm, damping=1.1 Ns/m).

The actual force levels the participants experienced depended on the penetration depth into the virtual plane and the constraint. The maximum force output of the haptic device was set at 5N to prevent the motors from overheating. Whenever the 5N output force was reached, a warning message was displayed to the participants instructing them to press more lightly on the virtual surface. This however was not treated as an error trial; the trial continued and the participant responded to the stiffness presented. In our previous study [15] in which veterinarians palpated virtual surfaces representing three different stages of bovine pregnancy, the maximum force used did not exceed 5N. The median peak force ranged between 1.62N and 2.48N depending on the stage of gestation.

2.4 Procedures

The experiment used a one-interval five-alternative forced-choice absolute identification procedure. Prior to the experiment, the participants followed an automated tutorial on the computer. Computerized instructions described the correct operation of the haptic device and participants were able to feel an example virtual surface. A simple graphic visualization of the surface, haptic interaction point and virtual constraint were provided. The experiment itself consisted of a training session followed by a testing session. No graphical representation of the surface was provided during the training and testing sessions. In the training session participants learned to associate the five different stiffness levels of the virtual surface with the numbers 1 to 5. The softest surface was associated with the number 1 and the hardest with the number 5. The training program allowed the participant to press any number between 1 and 5 on the keyboard and then feel the corresponding stiffness via the haptic device (see Fig. 1). The participant was free to choose the order in which s/he experienced the stiffness levels and could revisit the same stiffness multiple times. The participant was limited to changing the stiffness level 20 times after which the testing session began.

During the testing session, on each trial, the participant was presented with a surface of a stiffness value randomly selected from the same five values experienced in the training session. The participant's task was to identify the stiffness of the surface and press the corresponding number key. No visual information was shown on the computer screen during palpation of the virtual surface. After a response was entered, the correct answer number was shown on the screen. A total of 250 trials were collected per participant. A 5-minute break was enforced after the first 125 trials to attempt to reduce the effect of fatigue on performance.

In both the training and testing sessions the participant was required to lift the haptic interaction point (i.e., the PHANTOM thimble) up from the virtual surface before the stiffness of the surface was changed. This prevented any sudden change in the force output of the haptic device. It also prevented the participants from using the sudden increase or decrease in force as a cue for identifying stiffness. The participants were aware that their finger movements were constrained to the up-down direction, but no specific instructions were given regarding the palpation technique to be used for stiffness identification.

For each trial, the computer recorded the level of stiffness that was presented to the participant (stimulus) and the number key pressed by the participant (response). During each trial, the software recorded the depth that the device penetrated into the virtual surface (displacement) as a function of time. For convenience, the reaction force that the device exerted was also recorded (although it could have been calculated from displacement and the stiffness of the surface using Hooke's law). Force and displacement were recorded in the y-axis only (since motion in any direction other than up and down was restricted by the PD controller). The force and displacement recordings were made at a sample rate of 100Hz.

2.5 Data Analysis

2.5.1 Information Transfer

For each participant, the recorded stimulus-response pairs were used as indices into a confusion matrix (5 rows representing the 5 stiffness levels, 5 columns representing the responses). Each cell in the confusion matrix accumulated the number of times that a specific stimulus-response pair occurred. The entries along the main diagonal correspond to the trials where the participant correctly identified the stimuli. For each participant, data from the first and second sets of 125 trials were combined to form one confusion matrix.

Equation (1) shows the formula for calculating infor-

mation transfer. By applying this equation to the confusion matrix, the amount of information communicated via the sensory system can be calculated [18]. In (1), k is the number of stimulus alternatives, n is the total number of trials, n_{ij} is the cell entry in the i -th row and j -th column of the confusion matrix, n_i is the sum of the entries in the i -th row, n_j is the sum of the entries in the j -th column, and IT denotes information transfer. The number of stiffness levels that the participants can identify without error can then be calculated as 2^{IT} .

$$IT = \sum_{j=1}^k \sum_{i=1}^k \frac{n_{ij}}{n} \log_2 \frac{(n_{ij} \cdot n)}{(n_i \cdot n_j)} \quad (1)$$

2.5.2 Force and Displacement Data Analysis

The force vs. time and displacement vs. time recordings were analyzed to determine whether motor strategy differed between veterinarians and students. Since it was previously shown that for veterinarians the stiffness of a surface can affect motor strategy [15], the recordings were first grouped by stiffness level. In this way the analysis could account for potentially different strategies used for different stiffness levels.

Using MATLAB (The MathWorks, Inc., Natick, MA, USA), a graphical interface was created to allow the force vs. time and displacement vs. time graphs for a particular participant and stiffness level to be selected and viewed interactively. By visually inspecting a small subset of the graphs using this interface, various metrics were identified which could then be measured and compared programmatically across the whole data set of 6500 recordings (26 participants \times 250 recordings per participant). The metrics were identified by visually comparing recordings from participants with a high IT to those with a low IT, and noticing the differences. The following 7 metrics were chosen, which could be calculated automatically for each recording:

Maximum Force and Maximum Displacement – the local maximum points in the force vs. time and displacement vs. time recordings, respectively (multiple values per recording).

Minimum Force and Minimum Displacement – the local minimum points in the same recordings.

Contact Time – The total length of time in a force vs. time recording during which force is non-zero (one value per recording).

Frequency – The frequency of the spectral peak with the largest amplitude (excluding 0 Hz) of the Fourier transform of an individual recording.

Power – The power of the same spectral peak.

For each metric, a MATLAB script was developed to extract the mean and standard deviation (SD) of the metric from the recordings. The mean and SD were calculated for each participant at each stiffness level. In the case of the metrics involving minimum or maximum points, the mean and SD were obtained by considering all local minimum or maximum points across all the recordings for a specific participant and stiffness level (i.e., there were multiple data points per recording). The mean and SD of contact time were calculated by considering a single value per recording. Likewise, power and frequency (which were obtained from the Fourier transform of an individual recording) contributed a single data point per recording.

The SD was calculated because it was hypothesized that, being more practiced, the veterinarians might be more consistent in their motor strategy, which might be reflected in a lower SD of a metric per participant.

For the mean and SD of each metric, a mixed measures ANOVA was conducted with stiffness level (1 to 5) as a within-subject variable and group (veterinarian or student) as a between-subject variable. For non-normally distributed data a Mann-Whitney test was used as the between-subjects test and a Friedman’s test was used to examine the within-subjects effect.

Tests were also performed to check for correlation between IT and the mean or SD of a metric (in this case averaged across all stiffness levels, i.e., one value per participant), first by plotting and examining a scatter graph and then by applying the appropriate statistical tests for linear correlation (Pearson’s correlation for normally distributed data and Spearman’s correlation for non-normal).

3 RESULTS

3.1 Information Transfer of Stiffness

Table 1 shows the information transfers estimated from the 250 trials per participant. The results for the 12 experienced veterinarians varied from 0.72 bits to 1.15 bits, with a mean of 0.97 bits and a SD of 0.14 bits. This corresponds to the identification of 2.0 levels of stiffness. The results for the 14 veterinary students varied from 0.05 bits to 0.78 bits, with a mean of 0.58 bits and a SD of 0.23 bits. This corresponds to the identification of 1.5 levels of stiffness. The best veterinarian could correctly identify 2.2 stiffness levels, but the best student could only identify 1.7 stiffness levels without error. The differences between the student and veterinarian groups can be clearly seen in Fig. 2, which compares the spread of information transfers calculated from the 12 veterinarians and 14 students. Shown in each boxplot are the smallest and largest values (the whiskers), the lower and upper quartiles (the bottom and top of the box, respectively) and the median (the line inside the box). Essentially, the veterinarians could correctly identify (almost) 2 stiffness levels without errors and the students could only identify 1 level. A Shapiro-Wilk test showed that IT was normally distributed for the veterinarians but not for the students. A Mann-Whitney test showed that the difference in IT between the veterinarians and students was highly significant ($p < 0.001$).

There was a significant difference between the IT calculated for the first 125 trials compared with the second 125 trials (mixed measures ANOVA, $p < 0.05$). The students’ mean IT increased from 0.6 to 0.66, while the veterinarians’ increased from 0.98 to 1.07. There was no significant interaction between group and time, suggesting that both students and veterinarians improved in a similar way.

The stimulus-response confusion matrices are shown in Table 2a for the students and Table 2b for the veterinarians. Rows S1-S5 denote the five stimulus levels, and columns R1-R5 the five response labels. The rows labeled n_j show the number of times each response label was used. The entries along the main diagonals are the correct responses whereas all other entries are errors. A visual

TABLE 1
INFORMATION TRANSFER FOR STIFFNESS IDENTIFICATION

Students	Information Transfer (bits)	Veterinarians	Information Transfer (bits)
S1	0.26	V1	1.15
S2	0.70	V2	1.13
S3	0.46	V3	0.76
S4	0.44	V4	0.72
S5	0.05	V5	1.01
S6	0.77	V6	1.04
S7	0.49	V7	0.99
S8	0.76	V8	0.92
S9	0.78	V9	0.99
S10	0.78	V10	0.83
S11	0.77	V11	1.14
S12	0.73	V12	0.99
S13	0.69		
S14	0.46		
Mean	0.58	Mean	0.97
Std. Dev.	0.23	Std. Dev.	0.14

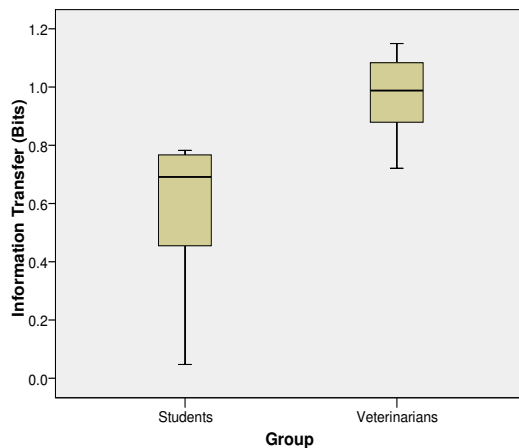


Fig. 2. Information transfer for the student and veterinarian groups. Boxplots show minimum and maximum (whiskers), upper and lower quartile (top and bottom of box) and the median (line in box).

TABLE 2
CONFUSION MATRICES FOR (A) STUDENT AND (B) VETERINARIAN GROUPS

		Response				
		R1	R2	R3	R4	R5
Stimulus	S1	375	194	80	29	15
	S2	155	267	213	77	21
	S3	34	137	223	181	67
	S4	21	46	178	270	216
	S5	13	26	84	218	360
n_j		598	670	778	775	679

(a) Students

		Response				
		R1	R2	R3	R4	R5
Stimulus	S1	461	137	14	4	0
	S2	142	285	153	16	0
	S3	14	108	338	131	10
	S4	2	11	167	313	124
	S5	0	3	36	150	381
n_j		619	544	708	614	515

(b) Veterinarians

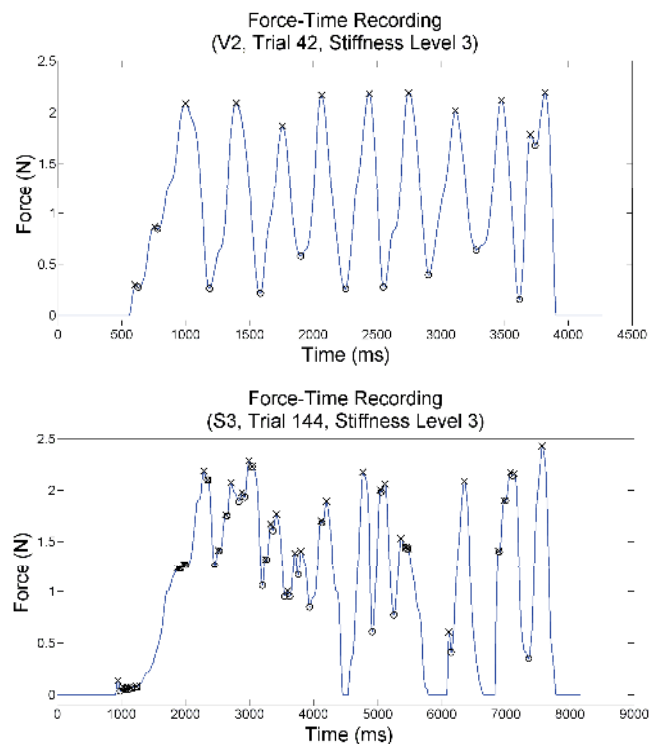


Fig. 3. Example force vs. time recordings from veterinarian participant V1 (top) and student S3 (bottom), both presented with stiffness level 3. Maximum and minimum points are marked with 'x' and 'o' respectively.

inspection indicates that there are a lot more errors for the students that are further away from the main diagonal line than for the veterinarians. This means that the veterinarians made “smaller” errors than the students (i.e., identifying a level 1 stiffness as 2, but not 5), which is consistent with the difference in IT for the two groups. Note also that there is no systematic response bias for either group as indicated by the consistent number of times each response label was used (see the rows labeled n_j with the total number of times each response label was used).

For the veterinarians, there was a significant positive correlation between the number of years of experience working in practice and IT score (Pearson’s correlation: 0.71, $p < 0.01$).

3.2 Force/Displacement Recordings

Visual inspection of the force vs. time and displacement vs. time recordings revealed examples of very different individual strategies, which helped determine the metrics that were used to look for patterns across the entire data set. For example, Fig. 3 shows the force vs. time recordings for veterinarian V2 (IT=1.13) and student S3 (IT=0.46). Visually, it appears that V2 has a more consistent strategy, by observing the spread of minimum forces (marked ‘o’) and maximum forces (marked ‘x’) used. This is reflected by a lower SD in both the minimum and maximum points compared to S3 (for these recordings SD of maximum points is 0.61 for V2 and 0.72 for S3; SD of minimum points is 0.44 for V2 and 0.68 for S3). This difference between V2 and S3 is still apparent when all their recordings are considered, with the SD of minimum and maximum points at each stiffness level being consistently higher for S3 than for V2 (SD of all maximum points, averaged over all stiffness levels, is 0.83 for V2 and 0.98 for S3; SD of minimum points is 0.48 for V2 and 0.67 for S3).

However, analyzing all participants, tests found no significant difference between the SD of minimum or maximum points of veterinarian and student recordings (mixed measures ANOVA for maximum points and Mann-Whitney test for minimum points). There was also no correlation with IT (Pearson’s correlation for maximum points and Spearman’s correlation for minimum points).

Similarly, there was no significant difference between the mean of minimum points for veterinarians and students (Mann-Whitney tests). When considering the mean of maximum points however, there was a significant difference between veterinarians and students (mixed measures ANOVA, $p < 0.05$). There was no significant correlation between mean minimum or maximum point and IT (Pearson’s correlation for maximum points and Spearman’s correlation for minimum points).

Fig. 4 shows the mean maximum point in the force vs. time recordings (top graph) and the displacement vs. time recordings (bottom graph) for the student and veterinarian groups, for each of the 5 stiffness levels presented. The graphs show that veterinarians have a higher mean maximum force (and displacement) than students. The error bars show standard error. In the case of force (top graph), the mean maximum point remains roughly constant across all stiffness levels (at a mean of 2.0N for veterinarians and 1.6N for students). The mixed measures ANOVA showed that there was no significant difference between the mean maximum force used at each stiffness level. In contrast, for displacement (bottom graph), the mean maximum point decreases as the surface increases in stiffness. A mixed measures ANOVA showed a significant difference between stiffness levels ($p < 0.001$). Post-hoc t-tests revealed that all stiffness levels were significantly different from each other (all $p < 0.05$, with Bonferroni correction). For displacement, the mixed measures ANOVA showed significant interaction between group (i.e., student or veterinarian) and stiffness level ($p < 0.05$), suggesting that the way in which veterinarians change their behavior with stiffness level is different from that of

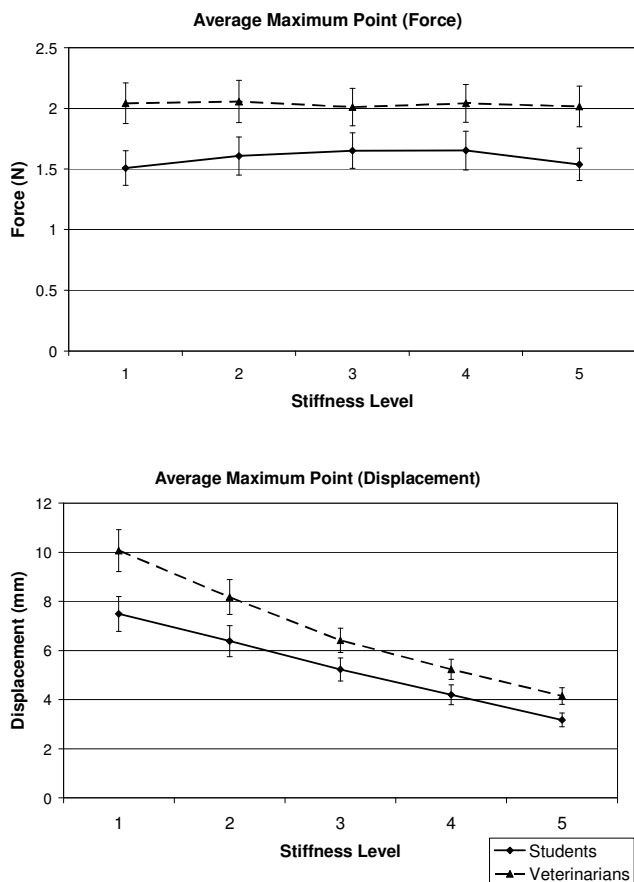


Fig. 4. Average maximum point in force (top) and displacement (bottom) recordings. Error bars show standard error.

students. This can be observed in Fig. 4 (bottom graph) as the divergence of the two lines at the lower end of the stiffness range.

The frequency domain analysis showed no significant difference between veterinarians and students for either the mean or SD of the frequency or the power of the largest frequency component (excluding 0Hz, mixed measures ANOVA for mean frequency and Mann-Whitney test for all others).

In terms of the total time spent in contact with the virtual surface per trial, veterinarian V1 took much more time than the other participants (7.2 seconds when averaged for all stiffness levels, almost 2 times as long as the next slowest participant V2 who took 3.8 seconds, and almost 3 times as long as the mean for all participants of 2.5 seconds). V1 was the best performing participant with the highest IT (1.15). We might therefore suspect that contact time could be positively correlated with IT. However, there was no significant difference between the mean or SD of contact time between veterinarians and students (Mann-Whitney tests) and there was also no significant correlation between IT and the mean or SD of contact time (Spearman's correlation).

Since V1 is clearly an outlier in terms of contact time, the same analysis was run again with V1 excluded. This time there was a significant difference in mean contact time between the veterinarians and students, with students spending more time in contact with the surface on

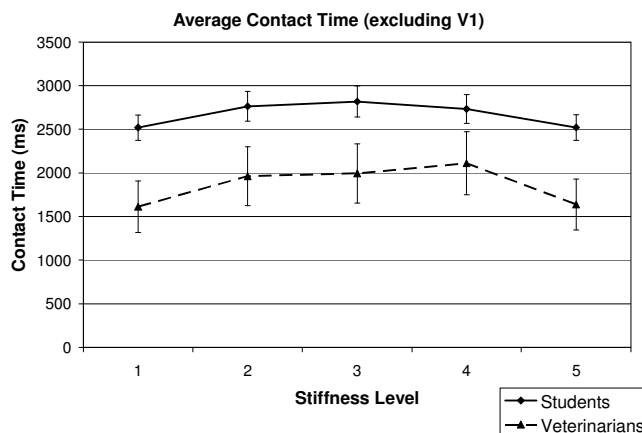


Fig. 5. Average time spent in contact with the surface per trial. Error bars show standard error.

average than veterinarians (mixed measures ANOVA, $p < 0.05$). Students had a mean contact time of 2.7 seconds with the surface per trial, compared to veterinarians at 1.9 seconds. Fig. 5 shows mean contact time per stiffness level for veterinarians and students (with V1 excluded), and error bars showing standard error. As well as showing that students spent more time in contact with the surface on average than the veterinarians, the graph also shows that both groups spent more time in contact with middle stiffness levels than the hardest or softest stiffness levels. The mixed measures ANOVA showed that there was a significant difference in the mean contact time between stiffness levels ($p < 0.001$). Post-hoc t -tests showed a significant difference in contact time between stiffness level 1 compared to levels 2, 3 and 4, and between level 5 compared to 2, 3, and 4 ($p < 0.05$ for all, with Bonferroni correction).

4 DISCUSSIONS

The present study measured the haptic perceptual abilities of veterinarians and veterinary students when identifying the stiffness of a virtual surface. The veterinarians were significantly better at the task, being able to identify more values within a set range. Furthermore, within the veterinarian group, the number of years of experience in clinical practice a participant had was positively correlated with stiffness identification performance. These findings indicate that stiffness perception in the context of veterinary medicine is a learned clinical skill, i.e., with clinical experience the skill of assessing stiffness improves.

Our results can be compared to those from our previous stiffness identification experiment where a larger stiffness range was used (0.2 – 3.0 N/mm as opposed to 0.2 – 0.5 N/mm used in the present study) [22]. As expected, our information transfer estimates for both groups (0.97 and 0.58 bits for veterinarians and students, respectively) were lower than the information transfer obtained with what can be considered non-experts in [22] (1.46 bits). The difference is most likely due to the differences in the stiffness range used in the two studies. There were also two additional differences in the methodologies of

the two experiments that preclude a direct comparison of results. Firstly, the haptic devices were different in the two studies, and the previous study used a stylus interface whereas the present study used a thimble interface. Secondly, the previous study prescribed the use of a tapping technique, while the present study allowed participants to use any method they desired. The possible influence of motor strategy on stiffness perception is an interesting and important issue, which is discussed later.

One might argue that the (almost) 2 levels of perfectly-identifiable stiffness levels achieved by the experienced veterinarians in the present study are not very impressive. Indeed, within a clinically-defined stiffness range, a practicing veterinarian would be expected to identify a number of different states of bovine pregnancy from not pregnant to several stages of early pregnancy. However, in the experiment the veterinarians barely identified two levels of stiffness across a slightly wider range. The ability of veterinarians to perform better when assessing pregnancy in a cow as compared to identifying stiffness values in the current study is probably related to the diagnosis depending on changes in other properties, such as size and shape, in conjunction with stiffness. This can be investigated with a future study using a multidimensional paradigm, where the participants are required to simultaneously identify the size, shape and stiffness of an object (for examples, see [26], [27]). Additionally, the veterinarian used one finger with the haptic device whereas during the real task s/he can use multiple digits. The ability to combine component skills and make diagnostic judgments is also important in the development of expertise. Therefore, as well as considering skills in isolation, the other factors that create the clinician should also be considered in our future work.

The nature of the simulated stimuli may also have contributed to the measured IT being lower than expected. Specifically, the stimuli used in the present study provided stiffness cues in terms of force/displacement ratio only, with no tactile information on surface deformation. A recent study showed that people are able to perceive the hardness of rubber specimens (specified by Young's modulus) directly based on surface deformation, without having to calculate the ratio of force/displacement changes [28]. An earlier study by the same authors suggested that the kinesthetic channel contributes just one quarter of the information used to assess stiffness, with cutaneous cues providing the rest [29]. Indeed, with deformable surfaces, tactile information alone is sufficient for discrimination, while for rigid surfaces, both tactile and kinesthetic information is required ([30]; see also [31] for an integrated haptic system that uses a fingertip contact area display to enhance softness discrimination). Visual information can also influence compliance perception [32]. We hasten to point out, however, that human and animal organs filled with non-compressible fluids lie somewhere between objects with rigid and deformable surfaces. Furthermore, physicians usually perform palpations wearing a glove and there are usually other anatomical structures such as the abdominal or rectal wall between the gloved hand and the palpated organ. We

thus think it unlikely that tactile information on local deformation could have provided much information for stiffness judgment in the clinical context considered in the present study.

A comparison of the first and second 125 trials showed that IT for both groups improved over the course of the experiment, suggesting learning effects. If this was the result of adjusting to the simulated environment and/or experimental procedure, then we would expect that the measured IT would be an underestimate of the veterinarians' performance in the real task. Since both veterinarians and students improved in a comparable way (there was no significant interaction between group and time) and veterinarians performed better than students in both the first and second half, it appears that the veterinarians' superior performance was due to a superior ability which they acquired during clinical practice, not a better ability to learn during the experiment.

What might explain the difference in performance between the veterinarian and student groups who participated in the present study? One might also ask whether the difference is due to a peripheral mechanism (that the veterinarians have more sensitive fingers) or a central mechanism (that the veterinarians have developed better sensory-motor strategies and can use the sensory information from their fingers more effectively).

The analysis of the force vs. time and displacement vs. time recordings suggested that veterinarians used a higher maximum force than the students (2.0 N on average for the veterinarians, compared to 1.6 N for the students). This finding is supported by the experience of some veterinary educators (personal communications), who find that students are reluctant to use a sufficient force during examinations – either due to lack of confidence or for fear of harming the patient. It was also found that, for both veterinarians and students, the mean maximum force per trial did not vary significantly between stiffness levels, but instead, the displacement decreased as stiffness increased (Fig 4). This finding was consistent with the force-constancy hypothesis proposed by [33] which states that when stroking a surface consisting of regions of different stiffness levels, in order to perceive its topology, people try to maintain a consistent level of force. Therefore, force-constancy also predicts a decrease in penetration depth (i.e., displacement in the present study) when stiffness level increases. In the present study, the surfaces of different stiffness levels are presented completely separately (in time), and the interaction is perpendicular to the surface (with the aim of identifying stiffness) rather than parallel to it (with the aim of identifying topology, as in [33]). However, our results are in keeping with the principal of the force-constancy hypothesis, i.e., that people try to apply a consistent force when exploring a surface.

With the notable exception of participant V1 (also the participant with the highest IT), on average, veterinarians spent significantly less time in contact with the virtual surface per trial than the students (1.9 seconds on average, compared to 2.7 seconds for the students). However, it seems more likely that this is a function of the veteri-

narians' confidence in their ability rather than part of a motor strategy that contributes to their ability. It was also shown that participants spent significantly more time in contact with a surface, per trial, when presented with a stiffness in the middle of the stimulus range, compared to a stiffness at the high or low end (Fig. 5). This could be explained by supposing that the participants were more confident at identifying the hardest and softest stiffness levels and therefore needed to spend less time feeling them. This is supported by the stimulus-response matrices (Table 2), which show higher percent-correct scores for S1 and S5 than for the other stimuli, indicating that participants were most successful at identifying the hardest and softest stiffness levels.

Given the highly significant difference in IT between veterinarians and students, we might have expected to detect a more obvious difference in motor strategy. However, with the exceptions mentioned above, the present study's analyses of motor behavior showed no other significant differences between the veterinarians and students for the metrics measured. The present study suggests that any differences in motor strategy between veterinarians and students are likely to be subtle, and a study involving larger numbers of participants is required in order to identify other significant differences and confirm the present findings. Future studies need to take into account how clinical knowledge can influence motor strategy and, if not controlled, potentially confound comparisons made between participants.

As mentioned previously, the mean maximum force used did not vary with stiffness level (Fig. 4), which suggests that unlike our previous study [15], veterinarians did not use a consistent domain-specific assumption about the surface they were feeling that affected the force they used. In the previous study, the veterinarians were told to assume that the virtual surfaces of different stiffness levels represented different clinical cases (stages of bovine pregnancy). As a result participants adapted their motor strategy in a consistent way when feeling different stiffness levels, which included using different levels of maximum force. In the present study, no instruction was provided to the participants regarding what the different stiffness levels they were presented with might represent. This could explain the wide variety of different motor strategies observed when visually inspecting the force/displacement vs. time recordings, as individual participants may have adopted different clinical assumptions, which could have led them to choose different strategies. All participants in the previous study were farm animal veterinarians, whereas the present study included both farm animal and small animal veterinarians, which probably added to the variety of different motor behaviors observed. Discussions with some of the veterinarians in the present study, after the experiment, indicated that they were trying to associate a clinical context with the stiffness values, but that they were using, in some cases, quite different contexts. For example, one veterinarian imagined that the highest stiffness represented bone, which can tolerate high forces, compared to another veterinarian who associated the same stiffness

value with a tight "hard" full bladder, which requires a gentle touch. The variety of strategies used may have contributed to the difficulties of identifying significant differences between the motor strategies of the veterinarian and student groups.

The findings from the present study have important implications for veterinary education in the sense that students clearly need to improve their skills of stiffness perception above the level that is innate or has been acquired during other manual tasks. The present study's comparison of motor strategies adopted by veterinarians and students hints that the maximum force used may have an influence on the perceptual ability to identify stiffness. If this is the case, it highlights the need to teach students how to use an effective and safe level of force to perform clinical examinations that involve palpation.

The progress of the novice along the path to clinical competence will involve repeated deliberate practice [34]. The boxplots in Fig. 2 show a much wider spread of information transfer values observed in the student group than in the veterinarian group. The plots could suggest that with training the poorest performing students can reach an "expert" level and that it would be interesting to follow these students, re-testing them at intervals throughout their education, to look for trends in IT over time. It is also possible that those who find such manual skills difficult to master never reach the practicing veterinarian population, perhaps choosing to pursue other career options. Also, testing final year veterinary students would reveal what level of expertise in stiffness perception is developed during their student education compared with the ongoing development of expertise acquired during professional practice.

The present study is only the beginning of many exciting studies where psychophysical methods are used to gain a better understanding of palpation. By quantifying expert ability, student training can be improved and targets set. Also, using the same methods, student ability can be monitored and assessed throughout the learning process. It would also be interesting to undertake further work to investigate other component skills, such as perceiving subtle differences in size or texture, to identify those skills that characterize the expert. This would then in turn provide metrics against which to assess competence and target training. Our research approach can be generalized to the analysis, training and assessment of other medical tasks, or in general any manual task, where experts attain a superior level of performance after an extended period of time on the job.

ACKNOWLEDGMENT

The authors thank the veterinarians and students at the Royal Veterinary College and the George Veterinary Group, Malmesbury, for their help with this study. N.F. was partially supported by a Student Exchange Fellowship from the IEEE Technical Committee on Haptics in 2007.

REFERENCES

- [1] N. Forrest, S. Baillie, and H.Z. Tan, "Haptic Stiffness Identification by Veterinarians and Novices: A Comparison," *Proc. World Haptics Conf. (WHC '09)*, pp. 646-651, 2009.
- [2] C.M. Pugh, W.L. Heinrichs, P. Dev, S. Srivastava, and T.M. Krummel, "Use of a Mechanical Simulator to Assess Pelvic Examination Skills," *J. Am. Med. Assoc.*, vol. 286, no. 9, pp. 1021-1023, 2001.
- [3] Y. Kuroda, M. Nakao, T. Kuroda, H. Oyama, and M. Komori, "Interaction Model Between Elastic Objects for Haptic Feedback Considering Collisions of Soft Tissue," *Comput. Methods Programs Biomed.*, vol. 80, no. 3, pp. 216-224, 2005.
- [4] G. Burdea, G. Patounakis, V. Popescu, and R.E. Weiss, "Virtual Reality-Based Training for the Diagnosis of Prostate Cancer," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 10, pp. 1253-1260, 1999.
- [5] M.O. Alhalabi, V. Daniulaitis, H. Kawasaki, and T. Hori, "Medical Training Simulation for Palpation of Subsurface Tumor Using HIRO," *Proc. World Haptics Conf. (WHC '05)*, pp. 623-624, 2005.
- [6] R.L. Williams II, M. Srivastava, R.R. Conatser Jr., and J.N. Howell, "Implementation and Evaluation of a Haptic Playback System," *Haptics-e: The Electronic J. for Haptics Research*, vol. 3, no. 3, <http://www.haptics-e.org/>, 2004.
- [7] S. Baillie, A. Crossan, S. Brewster, D. Mellor, and S. Reid, "Validation of a Bovine Rectal Palpation Simulator for Training Veterinary Students," *Stud. Health Technol. Inform.*, vol. 111, pp. 33-36, 2005.
- [8] S. Baillie, D. Mellor, S. Brewster, and S. Reid, "Integrating a Bovine Rectal Palpation Simulator into an Undergraduate Veterinary Curriculum," *J. Vet. Med. Educ.*, vol. 32, no. 1, pp. 79-85, 2005.
- [9] S.J. Lederman and R.L. Klatzky, "Hand Movements: A Window into Haptic Object Recognition," *Cogn. Psychol.*, vol. 19, no. 3, pp. 342-368, 1987.
- [10] S.J. Lederman, "Tactile Roughness Perception: Spatial and Temporal Determinants," *Can. J. Psychol.*, vol. 37, no. 4, pp. 498-511, 1983.
- [11] S.J. Lederman, R.L. Klatzky, C.L. Hamilton, and G.I. Ramsay, "Perceiving Roughness via a Rigid Probe: Psychophysical Effects of Exploration Speed and Mode of Touch," *Haptics-e: The Electronic J. for Haptics Research*, vol. 1, no. 1, <http://www.haptics-e.org/>, 1999.
- [12] A. Israr, Y. Li, V. Patoglu, and M.K. O'Malley, "Passive and Active Discrimination of Natural Frequency of Virtual Dynamic Systems," *IEEE Trans. Haptics*, vol. 2, no. 1, pp. 40-51, 2009.
- [13] S.J. Blakemore, D. Wolpert, and C. Frith, "Why Can't You Tickle Yourself?," *NeuroReport*, vol. 11, no. 11, pp. 11-16, 2000.
- [14] H.Z. Tan, S. Yang, Z. Pizlo, P. Buttolo, and M. Johnston, "Manual Detection of Spatial and Temporal Torque Variation through a Rotary Switch," *IEEE Trans. Haptics*, vol. 1, no. 2, pp. 96-107, 2008.
- [15] S. Baillie, A. Crossan, N. Forrest, and S. May, "Developing The 'Ouch-O-Meter' to Teach Safe and Effective Use of Pressure for Palpation," *EuroHaptics 2008*, M. Ferre, ed., LNCS 5024, Berlin: Springer-Verlag, pp. 912-917, 2008.
- [16] G.A. Gescheider, *Psychophysics: The Fundamentals*, 3rd ed., Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1997.
- [17] W.R. Garner, *Uncertainty and Structure as Psychological Concepts*, New York: Wiley, 1962.
- [18] H.Z. Tan, "Identification of Sphere Size using the PHANTom: Towards a Set of Building Blocks for Rendering Haptic Environment," *Proc. 6th Int'l Symp. Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pp. 197-203, 1997.
- [19] L.A. Jones and I.W. Hunter, "A Perceptual Analysis of Stiffness," *Exp. Brain Res.*, vol. 79, no. 1, pp. 150-156, 1990.
- [20] H.Z. Tan, N.I. Durlach, G.L. Beauregard, and M.A. Srinivasan, "Manual Discrimination of Compliance Using Active Pinch Grasp: The Roles of Force and Work Cues," *Percept. Psychophys.*, vol. 57, no. 4, pp. 495-510, 1995.
- [21] G. De Gerssem, "Kinaesthetic Feedback and Enhanced Sensitivity in Robotic Endoscopic Telesurgery," PhD dissertation, Katholieke Universiteit Leuven, 2005.
- [22] S.A. Cholewiak, H.Z. Tan, and D.S. Ebert, "Haptic Identification of Stiffness and Force Magnitude," *Proc. Symp. Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pp. 87-91, 2008.
- [23] S. Baillie, A. Crossan, S. Brewster, and S. Reid, "Preliminary Development and Evaluation of a Bovine Rectal Palpation Simulator for Training Veterinary Students," *Cattle Pract.*, vol. 11, no. 2, pp. 101-106, 2003.
- [24] N.I. Durlach and L.D. Braida, "Intensity Perception. I. Preliminary Theory of Intensity Resolution," *J. Acoust. Soc. Am.*, vol. 46, no. 2, pp. 372-383, 1969.
- [25] G.A. Miller, "Note on the Bias of Information Estimates," *Information Theory in Psychology: Problems and Methods*, H. Quastler, ed., pp. 95-100, 1954.
- [26] W.M. Rabinowitz, A.J.M. Houtsma, N.I. Durlach, and L.A. Delhorne, "Multidimensional Tactile Displays: Identification of Vibratory Intensity, Frequency, and Contactor Area," *J. Acoust. Soc. Am.*, vol. 82, no. 4, pp. 1243-1252, 1987.
- [27] H.Z. Tan, W.M. Rabinowitz, and N.I. Durlach, "Analysis of a Synthetic Tactile System as a Multidimensional Tactile Display," *J. Acoust. Soc. Am.*, vol. 86, no. 3, pp. 981-988, 1989.
- [28] W.M. Bergmann Tiest and A.M.L. Kappers, "Cues for Haptic Perception of Compliance," *IEEE Trans. Haptics*, vol. 2, no. 4, pp. 189-199, 2009.
- [29] W.M. Bergmann Tiest and A.M.L. Kappers, "Kinaesthetic and Cutaneous Contributions to the Perception of Compressibility," *EuroHaptics 2008*, M. Ferre, ed., LNCS 5024, Berlin: Springer-Verlag, pp. 255-264, 2008.
- [30] M.A. Srinivasan and R.H. LaMotte, "Tactile Discrimination of Softness," *J. Neurophysiol.*, vol. 73, no. 1, pp. 88-101, 1995.
- [31] E.P. Scillingo, M. Bianchi, G. Grioli, and A. Bicchi, "Rendering Softness: Integration of Kinaesthetic and Cutaneous Information in a Haptic Device," *IEEE Trans. Haptics*, vol. 3, no. 2, pp. 109-118, 2010.
- [32] M. Kuschel, M.D. Luca, M. Buss, and R.L. Klatzky, "Combination and Integration in the Perception of Visual-Haptic Compliance Information," *IEEE Trans. Haptics*, preprint.
- [33] S. Choi, L.A. Walker, H.Z. Tan, S. Crittenden, and R. Reifenger, "Force Constancy and its Effect on Haptic Perception of Virtual Surfaces," *ACM Trans. Appl. Percept.*, vol. 2, no. 2, pp. 89-105, 2005.
- [34] R.L. Kneebone, "Twelve Tips on Teaching Basic Surgical Skills Using Simulation and Multimedia," *Med. Teach.*, vol. 21, no. 6, pp. 571-575, 1999.



Neil Forrest received a Bachelor's degree in Computing Science from the University of Glasgow, UK in 2006. Until 2008, he worked at the Royal Veterinary College, University of London, UK, where he helped to develop haptic training simulators for veterinary students. He is currently a software development consultant specializing in haptic technology. His research interests include haptic skills training, haptic interaction design and haptic prototyping and modeling.



Sarah Baillie received her Bachelor's degrees in Anatomical Science (1984) and Veterinary Science (1986) from the University of Bristol, UK. She worked as a veterinarian in general practice for 20 years, latterly while retraining in computing science. She earned her Master (2003) and Doctorate (2006) from the Department of Computing Science, University of Glasgow. She is currently a Senior Lecturer in Veterinary Education at The Royal Veterinary College, University of London. She is the creator of

the Haptic Cow – a veterinary simulator now widely used in teaching in the UK. Her research interests include validation of haptic simulators in veterinary and medical training, psychophysics and other areas of veterinary education including peer-assisted learning, work-based learning, assessment and professionalism. She recently won the inaugural 'Most Innovative Teacher of the Year' at the Times Higher Education Awards 2009 for her work with haptic simulators and is one of Resource Centre UK's 'Women of Outstanding Achievement in Science, Engineering and Technology (SET)' for 2010.



Patrick Kalita received a Bachelor's degree from Rose-Hulman Institute of Technology (2006) and a Master's degree from Purdue University (2008), both in Electrical Engineering. At Purdue University he performed data analysis for a study on simulated textures rendered on a haptic device. He currently works on computer graphics at The MathWorks, Inc.



Hong Z. Tan received her Bachelor's degree in Biomedical Engineering (in 1986) from Shanghai Jiao Tong University and earned her Master and Doctorate degrees (in 1988 and 1996, respectively), both in Electrical Engineering and Computer Science, from the Massachusetts Institute of Technology (MIT). She was a Research Scientist at the MIT Media Lab from 1996 to 1998 before joining the faculty at

Purdue University. She is currently an associate professor of electrical and computer engineering, with courtesy appointments in the school of mechanical engineering and the department of psychological sciences. She is an associate editor of *Presence*, *ACM Transactions on Applied Perception* and *IEEE Transactions on Haptics*. Tan served as the founding chair of the IEEE Technical Committee on Haptics from 2006-2008. She was a recipient of the National Science Foundation CAREER award from 2000-2004. Her research focuses on haptic human-machine interfaces in the areas of haptic perception, rendering and multimodal performance. She is a senior member of the IEEE.