Towards a field-compatible optical Spectroscopic device for cervical cancer screening in resource-limited settings: effects of calibration and pressure

Vivide Tuan-Chyan Chang,1,* Delson Merisier,2 Bing Yu,1 David K. Walmer,2,3 and Nirmala Ramanujam1
1Department of Biomedical Engineering, Duke University, Box 90281, Durham, North Carolina 27708, USA
2Family Health Ministries, Leogane, Haiti
3Department of Gynecology and Obstetrics, Duke University Medical Center, Box 3084, Durham, North Carolina 27710, USA
*vivide.chang@duke.edu

Abstract: Quantitative optical spectroscopy has the potential to provide an effective low cost, and portable solution for cervical pre-cancer screening in resource-limited communities. However, clinical studies to validate the use of this technology in resource-limited settings require low power consumption and good quality control that is minimally influenced by the operator or variable environmental conditions in the field. The goal of this study was to evaluate the effects of two sources of potential error: calibration and pressure on the extraction of absorption and scattering properties of normal cervical tissues in a resource-limited setting in Leogane, Haiti. Our results show that self-calibrated measurements improved scattering measurements through real-time correction of system drift, in addition to minimizing the time required for post-calibration. Variations in pressure (tested without the potential confounding effects of calibration error) caused local changes in vasculature and scatterer density that significantly impacted the tissue absorption and scattering properties of normal cervical tissues. Future spectroscopic systems intended for clinical use, particularly where operator training is not viable and environmental conditions unpredictable, should incorporate a real-time self-calibration channel and collect diffuse reflectance spectra at a consistent pressure to maximize data integrity.

©2011 Optical Society of America

OCIS codes: (170.6510) Spectroscopy, tissue diagnostics; (060.2280) Fiber design and fabrication; (170.4440) ObGyn.

References and links
1. PATH, “Cervical cancer prevention initiatives at PATH,” (PATH),
1. Introduction

Cervical cancer affects the lives of 500,000 women and results in more than 270,000 deaths worldwide annually [1]. A disproportionate burden of disease is borne by women living in low- and middle-income countries, where 85% of these cases occur [1,2]. Globally, the number of cervical cancer deaths is still rising, with estimates that the rates will increase by 25% over the next 10 years [3]. In contrast, the incidence of cervical cancer in developed countries has significantly decreased due to regular screening with a cytology-based approach – Papanicolaou (Pap) smear. An abnormal Pap is followed by colposcopically directed biopsy (2nd visit), and subsequent treatment (3rd visit) if pre-cancer or cancer is found [2,4]. Employing these methods requires multiple visits, as well as a centralized laboratory and skilled staff for processing and evaluation of cytology and pathology specimens [5]. Due to the lack of such resources, the benefits of established cervical cancer screening paradigms have yet to be realized in developing countries [6,7]. Studies have suggested that if a woman was evaluated for cervical cancer only once in her lifetime between the ages of 30 and 40, her risk of cancer would be reduced by 25 – 36% [8]. Thus, there is a compelling need for effective strategies to detect cervical disease (high grade cervical intraepithelial neoplasia (CIN 2 + ) or invasive carcinoma [9]) in resource-limited settings where multiple clinical visits are not feasible and centralized laboratories do not exist.

Guidelines have been written by the Alliance for the Prevention of Cervical Cancer (ACCP) on strategies for screening cervical cancer in resource-limited settings [1]. The most efficient and effective strategy for prevention of cervical cancer in low resource settings is to screen using HPV testing or visual inspection with acetic acid (VIA), followed by treatment of the pre-cancerous lesions using cryotherapy [8,10,11]. This should optimally be carried out in a single visit by physicians, nurses, or midwives. Until low cost HPV testing is realized, VIA, or VIA with low power magnification (VIAM), combined with cryotherapy appears to be the most viable option for reducing mortality associated with cervical cancer in the developing world. Unless there is a suspicion of invasive cervical cancer, the routine use of intermediate diagnostic biopsy (such as colposcopy) between screening and treatment is not recommended because it often leads to decreased programmatic coverage due to increased inefficiency (e.g., lack of access to a pathology laboratory) and the need for multiple visits (i.e., patient attrition with each additional clinic visit that is needed).

The sensitivity of VIA/VIAM is better than, if not at least as good as, Pap smear [8,12]; but, unlike Pap smear, it does not require specimen collection or processing. VIA/VIAM also allows for surveillance of the entire cervix, thus enabling visualization of areas to biopsy and/or treatment. However, as shown in Table 1, the inherent issue is that the specificity of VIA/VIAM is lower than that of Pap smear. In a screen and diagnose paradigm (similar to what is done in the developed world), this would lead to a large number of unnecessary biopsies. In a screen and treat paradigm (as recommended by the ACCP for the developing world), this would result in the over-treatment of patients that have no disease, which is a major issue as it would negatively impact programmatic success. Improving the specificity of VIA/VIAM should have a significant and positive impact on the programmatic success of cervical cancer screening in resource-limited settings. One potential strategy to achieve this is to leverage the underlying absorption and scattering properties obtained via quantitative reflectance spectroscopy to improve the specificity with which VIA/VIAM can be used for cervical cancer screening in resource-limited settings (Table 1). Pilot cervical tissue optical spectroscopy studies have been reviewed by Thekkek et al. in 2008 [13] and the review shows that this technology can achieve sensitivities and specificities in the range of 83 – 92% and 80 – 90%, respectively, thus having the potential to address the limited specificity of VIA or VIAM.
Table 1. Sensitivity and Specificities of Current Screening and Diagnostic Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Sample Size</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIA [13]</td>
<td>1093 – 55,000</td>
<td>70 – 77%</td>
<td>49 – 86%</td>
</tr>
<tr>
<td>VIAM [14]</td>
<td>5378</td>
<td>85%</td>
<td>69%</td>
</tr>
</tbody>
</table>

The primary sources of intrinsic absorption and scattering contrast in cervical neoplasia are increased absorption associated with neovascularization in the stroma [15–17], increased epithelial scattering with cell proliferation, and decreased stromal scattering with collagen degradation [18–20] (dominant source of multiply backscattered light in the visible wavelengths [21]). Using immunohistochemical stains for vascular endothelial cells, Abulafia et al. [22] and Dellas et al. [15] have demonstrated that neovascularization, quantified through microvessel density, is positively correlated with increasing cervical dysplastic grades. Concomitant increase in epithelial scattering with neoplastic development (through increased nuclear-to-cytoplasmic ratio) has been observed by Collier et al. [23] through confocal microscopy. Brummer et al. [24] and Talvensaari-Mattila et al. [25], using antibodies against activated matrix metalloprotenases (MMP) that degrade collagen, have found weak to strong expression of MMP-1 and MMP-2 in CIN 2 + but negative staining in CIN 1 and normal tissues, suggesting that decreased stromal scattering could be attributed to the degradation of collagen. Thus, optical spectroscopy may provide a cost-effective and accurate alternative to traditional screening and diagnostic methods by measuring physiological and morphological alterations of the dysplastic cervix [26–31].

Our group recently employed a fiber-optic spectrometer and a fast, scalable Monte Carlo (MC) model [32,33] to elucidate the underlying sources of absorption and scattering contrast in 39 patients at the Duke University Medical Center (DUMC) [34]. The probe was designed to interrogate both the epithelium and stroma. A significant increase in total hemoglobin content ([total Hb]) was observed in CIN 2 + compared to CIN 1 and normal cervical tissues. The increase in neovascularization was also validated independently with immunohistochemical staining of endothelial cells, which demonstrated that microvessel density (representative of neovascularization) was statistically higher in CIN 2 + tissues compared to CIN 1 and normal tissues [31]. These results are consistent with those observed by other groups who carried out quantitative spectroscopy of the cervix [35,36]. Our group also observed [34] a decrease, albeit not statistically significant ($P = 0.06$), in the wavelength-averaged reduced scattering coefficient ($<\mu_s>$) in CIN 2 + from CIN 1 and normal tissues, which may in part reflect the degradation of collagen in the stroma. Hornung et al. [37] observed a decreasing trend in scattering ($P = 0.16$) in CIN 2 + using near infrared spectroscopy, and Georgakoudi et al. [38] also observed a decreasing trend in the scattering slope (350 – 750 nm) in CIN versus normal tissues. Interestingly none of the studies noted statistical significance.

The sources of optical contrast measured using quantitative optical spectroscopy can be affected by operator bias and unpredictable variations in system throughput due to the environment in which the technology is used. Both of these issues are particularly relevant to implementation of the technology in a resource-limited setting in the developing world. Operator bias can influence quality control in routine calibration measurements and placement of the probe on the cervix with reliable pressure, and environmental conditions such as humidity and temperature and the care with which the system is handled can impact the stability in the throughput of the system during its operation. Automated accountability of these potential confounders of contrast will increase the reliability of quantitative optical spectroscopic measurements in clinical studies, particularly in resource-limited settings. Another important consideration in the development of analytical instruments for use in resource-limited settings is power consumption. Roblyer et al. [30] conducted a pilot study...
using a multispectral digital colposcope to identify neoplastic cervical tissues in a clinical study in Ibadan, Nigeria and pointed out that high operating requirements, in particular, power, rendered the device inappropriate for use in low-resource settings. Building a system that can be powered off a laptop battery would eliminate this additional variable when conducting studies in resource-limited environments.

Our group has designed a first-generation fiber-optic diffuse reflectance spectroscopy system that attempts to address several of the issues presented above. The system consists of a specialized self-calibration probe coupled to a light emitting diode (LED) and miniature USB fiber-based spectrometers, making the system highly portable and self-contained without the need for external power supply [39]. Tissue spectra are divided by the simultaneously acquired self-calibration (SC) channel spectra and a probe-dependent correction factor [40,41] to correct for real-time variations in throughput of the instrument due to lamp warm up, bending of the fiber optic probe, and errors that may arise from non-real time calibration. We [41] have also demonstrated that the SC feature (real-time) is superior to conventional calibration with a Spectralon® puck (non real-time), particularly with respect to extraction of the wavelength-averaged reduced scattering coefficient ($\langle \mu_s' \rangle$) with errors of 2.1 ± 1.1% for SC vs. 12.5 ± 6.1% for puck calibration in tissue-mimicking phantom studies. This has clinical implications in that accurate extraction of scattering contrast may better resolve the statistically insignificant, but decreased scattering trends seen previously by our group and other groups in CIN 2 + vs. CIN 1 and normal tissues [34,35,37].

The objective of the work presented in this paper is to test the feasibility of implementing the technology described above in a pilot clinical study to measure cervical tissue optical properties in a resource-limited clinic in Leogane, Haiti. The first goal was to compare the effect of real-time calibration to that of post-measurement calibration on the extraction of optical properties and determine if the results observed were consistent with that carried out independently on synthetic tissue phantoms in the lab. The second goal was to evaluate the effect of pressure without the confounding effects of potential calibration errors to essentially determine the contribution of this source of operator bias on the extraction of cervical tissue optical properties. The effect of self-calibration was consistent between the phantom and clinical tissue measurements. In the phantom studies, the calibration method used had no statistically significant impact on the extraction of the absorption coefficient (which essentially reflects total hemoglobin concentration, or [total Hb]) and this was also the case in the clinical measurements. However, compared to self-calibration, calibration with the puck resulted in over estimation of $\langle \mu_s' \rangle$ in the phantom studies. Disparities between the results from the two calibration methods in the phantom studies (puck calibration resulted in higher $\langle \mu_s' \rangle$ values) were recapitulated in the clinical measurements, i.e., the extracted $\langle \mu_s' \rangle$ showed higher median values and greater variance when calibrated with a puck post-measurement in the cervical tissue studies. The results from the clinical studies also showed that when the potential confounding effects of calibration errors were eliminated using the self-calibration strategy, pressure significantly affected the accurate extraction of [total Hb] and $\langle \mu_s' \rangle$, but not Hb saturation.

In summary, our work demonstrates that the differences between self-calibration and post-calibration effects on the extraction of absorption and scattering properties in synthetic tissue phantoms is recapitulated in cervical tissue measurements in vivo with profound effects on $\langle \mu_s' \rangle$, and when the potential confounding effects of calibration are eliminated through self-calibration, the effects of pressure on cervical tissues can be independently assessed showing a significant impact of pressure on the measurement of both [total Hb] and $\langle \mu_s' \rangle$. Thus, a robust platform for quantitative optical spectroscopy should include an automated feature that has the ability to account for system throughput variations due to operator bias and environmental instability and minimize the variations in probe pressure on the tissue surface and thus, the measured spectral intensities. These requirements are particularly
important when conducting clinical studies in resource-limited settings where operator inconsistency and environmental variables are difficult to control for.

2. Method

2.1 Instrumentation

The portable spectroscopic system consists of: an LED illumination module, a self-calibration fiber-optic probe, two spectrometers, and a laptop computer for control and power (Fig. 1). The LED source was a cool white, high-power LED (XR-E, Cree, Durham, NC) with outputs between 400 and 700 nm. The LED was coupled to the source optical fiber via a collimating lens (XLamp 7090, Cree, Durham, NC) and a fiber optic collimator (FOC-010-006-V Mightex, Toronto, ON) aligned for maximum output. The LED, collimating lens, and fiber optic collimator were housed in an aluminum enclosure (custom machined) for protection and ease of handling. The LED was driven through a current regulated driver (LuxDrive 2008B PowerPuck, Randolph, VT) and was powered using the 5V supply from the universal serial bus (USB) port available in a laptop PC. The mean power consumption of the portable spectroscopic system is 4.5 W.

Fig. 1. (a) The portable spectroscopic system consists of an ultrabright white LED module, a spectrometer for tissue sensing, a spectrometer for self-calibration, and a fiber optic probe to deliver and collect diffuse reflectance from 450 – 600 nm from the cervix in vivo. All fibers are 200/220 µm in core/cladding diameter with a numerical aperture (NA) of 0.22. (b) The distal end in contact with tissue consists of a central collection fiber encircled by a ring of 6 illumination fibers with a center-to-center separation of 622 µm. (c) Light delivered to Spectralon® coating and collected via self-calibration collection fiber is used to account for drifts in system throughput in real time [41].

Two fiber-optic USB spectrometers were used for this clinical study: one for the tissue sensing channel and another for the self-calibration channel to account for system drift in real time. Spectra were acquired concurrently on both spectrometers using a custom LabVIEW® (Natick, MA) control software, with integration times ranging from 50 to 500 ms. Three repeated scans were acquired at each site for quality control and a scan was discarded if it deviated from other two scans by more than 10%. The spectrometer used for tissue sensing,
HRS-VIS-025 (Mightex, Toronto, ON), and the spectrometer for self-calibration, USB-4000 VIS-NIR (Ocean Optics, Dunedin, FL), have spectral resolutions of 0.4 nm and 1.5 nm, respectively. Two spectrometers with different specifications were used due to logistical constraints.

Two-hundred-µm (core diameter) fibers (Polymicro Technologies, Phoenix, AZ) were used for illumination (x7), diffuse reflectance detection (x1), and self-calibration detection (x1). Light was launched to seven fibers in a closed packed structure, of which six were used to illuminate the cervix and one for self-calibration illumination. The self-calibration illumination fiber was terminated at a stainless steel rod coated with spectrally flat Spectralon® (LabSphere, North Sutton, NH) inside the ferrule at the distal end of the fiber-optic probe, as shown in Fig. 1(c). The reflected light from the spectrally flat rod was coupled back to the spectrometer via a self-calibration collection fiber to record drifts in system throughput in real-time. Details of the self-calibration probe construction can be found in [40,41]. Yu et al. [41] have shown that using a self-calibration channel can significantly reduce bending losses where the bending diameter is greater than 3 cm. The distal end in contact with the cervix consisted of fibers epoxied within a stainless steel ferrule with an outer diameter of 3 mm. All the 200-µm fibers were made of identical high-OH silica/doped silica (core/cladding) with a numerical aperture (NA) of 0.22 to ensure the same bending response. Except for probe ends housed in stainless steel ferrules and the system end, the entire length of the fiber optic probe was covered with a stainless steel jacket for protection and durability. The stainless steel tube was sterilized in Cidex® OPA (ASP, Irving, CA) for 20 minutes prior to each study patient for disinfection.

The distal end that is in contact with the cervix consisted of a central collection fiber encircled by a ring of 6 illumination fibers with a center-to-center separation of 622 µm. The separation distance was chosen to match the geometry, and consequently the sensing depth, of the probe used in a previous study of the cervix in vivo [34]. Defining the sensing depth as the maximum depth that 50% of the detected photons ever penetrated, Monte Carlo simulations [34,42], showed that the mean sensing depth for wavelengths between 450 – 600 nm is 500 – 600 µm, respectively. Since the average cervical epithelial thickness is 200 – 500 µm [43,44], the probe appears to be preferentially sensitive to changes in absorption due to hemoglobin and alterations in scattering arising from collagen in the cervical stroma. Epithelial thickness does not appear to correlate with pathology, though it is dependent on age and decreases in post-menopausal women [43].

A HeNe laser was used for wavelength calibration once during instrument characterization prior to collection of clinical data in Haiti and re-affirmed at the conclusion of the study to ensure proper wavelength calibration of both spectrometers. The linearity of detector and the signal-to-noise (SNR) at select wavelengths (λ = 450, 550, 600 nm) of the system were characterized in a similar manner as in [39]. Briefly, the linearity of the detectors was ascertained by sequentially measuring the diffuse reflectance (from a Spectralon® reflectance standard) as a function of detector integration time. The SNR, as defined by Eq. (1), was quantified by the mean (MeanR(λ)) and standard deviation (StdR(λ)) of three repeated measurements from tissue mimicking phantoms (Exp 2) with the optical properties given in Section 2.3. Drift of the light source and spectrometers were monitored at 15-second intervals over 140 minutes with the sampling channel secured on a Spectralon® standard.

\[
SNR(\lambda) = 20\log_{10}\frac{MeanR(\lambda)}{StdR(\lambda)}
\]  

(1)

2.2 Monte Carlo model

A flexible and fast Monte-Carlo-based inverse model [32] developed by our group was used to extract the absorption and scattering properties of tissue mimicking phantoms and cervical tissues using diffuse reflectance spectra from 450 – 600 nm. The model has been validated...
extensively in tissue-mimicking phantoms [33], murine and hamster tumor models [45–47], and in the breast [48–50] and the uterine cervix [34] in vivo. The model is valid for a wide range of optical properties and can be used with any probe geometry and system setup provided that a one-time calibration is performed on a synthetic phantom with known absorption and scattering properties. The accuracy of the inverse Monte Carlo model [32] used is comparable for spectral bandwidth of 5 nm or less; hence, any wavelength drift less than 5 nm (e.g., due to temperature dependent expansion) can be neglected [33]. The fixed parameters of the inverse model are the wavelength-dependent extinction coefficients of the absorbers and refractive indices of the scatterers and the surrounding medium. The extinction profiles of oxyHb and deoxyHb reported by Prahl [51] are used. The free parameters that are iteratively searched during a fitting include oxyHb and deoxyHb concentrations, scatter size, and volume density of scatters. A Gauss-Newton nonlinear least-squares optimization algorithm (MATLAB, Mathworks, Natick, MA) is used to minimize the difference between the measured and the Monte Carlo-simulated diffuse reflectance. A ratio of the measured reference phantom reflectance to the modeled reference phantom reflectance gives a scaling factor that enables a direct comparison between measured and predicted reflectance spectra during the inversion process.

2.3 Phantom validation

The accuracy with which the portable spectroscopic system and inverse Monte Carlo model could extract optical properties was evaluated using liquid phantoms with cervix-mimicking optical properties [34,52]. The phantoms consisted of lyophilized human hemoglobin (H0267 Sigma-Aldrich, St. Louis, MO) as absorbers and 1-µm monodisperse polystyrene spheres (07310 Polysciences, Warrington, PA) as scatterers. Two sets of experiments were performed – one with increasing levels of absorber (Exp 1) and another with increasing levels of scatterer (Exp 2). Exp 1 and 2 were performed on different days to assess the influence of the calibration method used. Total hemoglobin concentration ([total Hb]), and the range and wavelength-averaged values of absorption (\(\langle \mu_a \rangle \)) and reduced scattering (\(\langle \mu_s' \rangle \)) over 450 – 600 nm for Exp 1 and 2 are enumerated in Table 2 and Table 3. The expected values for \(\mu_a(\lambda)\) were determined using a spectrophotometer (Cary 300, Agilent, Lexington, MA) and Beer’s law, whereas \(\mu_s'(\lambda)\) of the phantoms were calculated using Mie theory. Phantoms 4 & H (bolded in Tables 3 and 4) were used as the reference phantom, respectively, on the 1st and 2nd days. Phantom H was also used as the reference phantom for the analysis of clinical data since it has similar optical properties as the reference phantom used previously by our group in prior clinical studies [34]. Both the self-calibrated and puck-calibrated tissue measurements were calibrated with the same reference phantom. The self-calibrated measurements were further divided by a correction factor [41] to account for the throughput differences between the sampling and self-calibration channels.

<table>
<thead>
<tr>
<th>Titration</th>
<th>[total Hb] (µM)</th>
<th>(\langle \mu_a \rangle) (cm(^{-1}))</th>
<th>(\langle \mu_s' \rangle) (cm(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.86</td>
<td>0.04</td>
<td>10.44</td>
</tr>
<tr>
<td>2</td>
<td>4.51</td>
<td>0.07</td>
<td>10.29</td>
</tr>
<tr>
<td>3</td>
<td>7.16</td>
<td>0.11</td>
<td>10.05</td>
</tr>
<tr>
<td>4</td>
<td>9.69</td>
<td>0.15</td>
<td>9.82</td>
</tr>
<tr>
<td>5</td>
<td>12.10</td>
<td>0.19</td>
<td>9.60</td>
</tr>
<tr>
<td>6</td>
<td>14.97</td>
<td>0.23</td>
<td>9.34</td>
</tr>
<tr>
<td>7</td>
<td>17.68</td>
<td>0.27</td>
<td>9.09</td>
</tr>
<tr>
<td>8</td>
<td>20.26</td>
<td>0.31</td>
<td>8.85</td>
</tr>
<tr>
<td>9</td>
<td>22.70</td>
<td>0.35</td>
<td>8.63</td>
</tr>
<tr>
<td>10</td>
<td>25.03</td>
<td>0.39</td>
<td>8.42</td>
</tr>
</tbody>
</table>
Table 3. Optical Properties (450 – 600 nm) for Titrate Scatterer Phantom Experiment (Exp/Day 2)

<table>
<thead>
<tr>
<th>Titration</th>
<th>[total Hb]</th>
<th>(&lt;\mu_s&gt;)</th>
<th>(&lt;\mu_a&gt;)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>7.33</td>
<td>0.07</td>
<td>7.49</td>
</tr>
<tr>
<td>B</td>
<td>6.93</td>
<td>0.07</td>
<td>11.07</td>
</tr>
<tr>
<td>C</td>
<td>6.51</td>
<td>0.06</td>
<td>14.86</td>
</tr>
<tr>
<td>D</td>
<td>6.11</td>
<td>0.06</td>
<td>18.57</td>
</tr>
<tr>
<td>E</td>
<td>5.70</td>
<td>0.05</td>
<td>22.36</td>
</tr>
<tr>
<td>F</td>
<td>13.03</td>
<td>0.12</td>
<td>7.49</td>
</tr>
<tr>
<td>G</td>
<td>12.30</td>
<td>0.12</td>
<td>11.19</td>
</tr>
<tr>
<td><strong>H</strong></td>
<td><strong>11.57</strong></td>
<td><strong>0.11</strong></td>
<td><strong>14.97</strong></td>
</tr>
<tr>
<td>I</td>
<td>10.85</td>
<td>0.10</td>
<td>18.68</td>
</tr>
<tr>
<td>J</td>
<td>10.12</td>
<td>0.10</td>
<td>22.46</td>
</tr>
</tbody>
</table>

2.4 Clinical protocol

Haiti was chosen as the study site since Latin American countries are among those with the highest incidence and mortality of cervical cancer in the world [6]. The study protocol was reviewed and approved by the Institutional Review Boards at Duke University Medical Center (DUMC) in Durham, NC, USA and Misyon Sante Fanmi Ayisyen (Family Health Ministries, FHM) in Leogane, Haiti. Informed written (or oral if the patient is illiterate) consent was obtained from patients admitted to FHM Cervical Screening Clinic or Dr. Merisier’s private clinic for cervical cancer screening based on either condition: (1) positive Papinicolau (Pap) smear or (2) seropositive for highly virulent human papilloma virus (HPV) strains (i.e., 9, 16, and 18). Diffuse reflectance spectra were collected from 49 sites in the cervical transformation zone of 21 female patients aged 30 – 62 years (mean ± SD: 40.3 ± 8.5 years). Pregnant women were excluded from the study and all but two recruited patients were pre-menopausal. Of the 49 sites examined, 16 were colposcopically abnormal after the application of acetic acid and 33 sites were colposcopically normal. Biopsies results were not available due to limited lab access and financial hardship of patients so only colposcopically normal sites were included in subsequent analysis of the data for this study.

Diffuse reflectance from 450 - 600 nm was collected from all (up to three) colposcopically abnormal sites immediately following visual examination at low magnification of the cervix with the application of 5% acetic acid. This was followed by an optical measurement on a colposcopically normal site from the same patient. All data were acquired (within one minute) following the application of dilute acidic acid since in reality tissues would have residual acetic acid before optical interrogation with our probe if VIA or VIAM was performed. Optical interrogation of colposcopically normal and abnormal sites was conducted prior to biopsy to avoid confounding absorption due to superficial bleeding. Identification of abnormal site, placement of the probe on the cervix, and biopsies were made by the same gynecologist (DM). A probe holder, constructed of Delrin® and attached to the speculum, was used to stabilize the fiber optic probe and prevent motion-induced artifacts. The probe was placed in contact with a specific site on the cervix and subsequently locked in place with the probe holder for the duration of the measurement.

Two calibration schemes were employed for the study. Traditionally, a diffuse reflectance standard was used to correct for drifts in source or system throughput and calibration spectra were obtained either before or after clinical spectra acquisition. In this study the diffuse reflectance standard consists of a puck (SRS-99 LabSphere, North Sutton, NH) coated with spectrally flat Spectralon® (LabSphere, North Sutton, NH) in the UV-visible-NIR wavelengths. The second calibration method consisted of a self-calibration channel incorporated in the fiber-optic probe to account for real-time fluctuations in system throughput (e.g., bending, changes in LED output, etc), as well as to streamline operation of
the spectroscopic device by eliminating pre- or post-study calibration measurements by an operator in the field [41].

To study the influence of applied tissue-probe contact pressure, a feature that has yet to be accounted for in the current generation of our probe, diffuse reflectance spectra were acquired from 19 sites in 19 patients with low, medium, and high contact pressures. After identifying a colposcopically normal site, the gynecologist gradually increased the pressure, pausing at each pressure level for one to two seconds for data acquisition. Low pressure was defined as having a gentle touch but ensuring a close contact between the distal end of the probe and the tissue. Medium pressure was defined as ensuring a closed contact with minimal visible compression of the tissue. High pressure was defined as exerting the maximum pressure without causing significant pain to the patient. At each of the exerted probe pressures, the probe was held in place by the probe holder. Three repeated scans were acquired at each applied pressure level on all 19 sites. Means of three repeated scans per site (per pressure) were used in subsequent analysis.

2.5 Statistical analysis

MATLAB (MathWorks, Natick, MA) was used to perform the Student t-tests and Wilcoxon rank sum tests. The Student t-test was used when data could be assumed to be normally distributed, as determined visually and by using the Lillifors test for normality.

3. Results

3.1 System characterization

The linearity of both the tissue-sensing and self-calibration detectors was excellent with correlation coefficients between intensity and integration time greater than 0.99 at 450, 500, 550, and 600 nm over a 16-bit dynamic range. For Phantom H, the SNR at 450, 500, 550, and 600 nm were 59, 69, 64, and 54 dB, respectively. For other phantoms, the SNR was at least 30 dB at any wavelength. The self-calibration channel also had a higher SNR (> 40 dB) compared to the sensing channel (> 30 dB). Since the SNR plateaued at higher intensities, both detectors are shot-noise dominated (data not shown) [53]. A system drift test over 140 minutes using a reflectance standard as the sample showed that the sample and self-calibration intensities at 574 (peak wavelength) converged to within 1% of steady-state intensity within four minutes (data not shown).

3.2 Phantom results

Extracted versus expected [total Hb] and <\mu'> in phantom experiments 1&2 are shown in Fig. 2. As shown in Fig. 2(b), both puck calibration and self-calibration techniques result in similar extraction accuracies for [total Hb]. However, as shown in Fig. 2(d), the errors for scattering are significantly reduced (2 – 3 times) with self-calibration compared to those calibrated with the puck. Yu et al. [41] attributed the poor extraction of scattering to drifts in the overall intensity of the collected diffuse reflectance, which is crucial to the accurate extraction of scattering; whereas measurement of absorption is more sensitive to the spectral shape [40].

Table 4 also summarizes the percent errors in optical property extraction from tissue-mimicking phantoms from the above phantom studies as well as for previously published phantom studies where different pixels on the same detector (CCD) of an imaging spectrometer were used to resolve the sensing and calibration channels [41]. Overall errors are comparable between the phantom studies reported here and that reported by Yu et al. [41]. However, the \mu' range reported in this study is significantly larger than that tested by Yu et al. [41] and hence, the errors in <\mu'> for different-day phantoms using puck calibration are also concomitantly larger [41].
Fig. 2. Extracted vs. expected [total Hb] when (a) target and reference phantoms are from the same experiment (or day) and (b) target and reference phantoms are from different experiments (or days). Extracted vs. expected $\mu_s'$ when (c) target and reference phantoms are from the same experiment (or day) and (d) target and reference phantoms are from different experiments (or days). Either calibration method yielded similar extraction errors for [total Hb] and $\mu_s'$ when same-day reference phantom is used. Extraction error for [total Hb] was similar even when a different-day reference phantom was used. However, extraction error was substantially higher for $\mu_s'$ when puck-calibration was used in lieu of self-calibration.

Table 4. Errors in Extraction of Phantom [total Hb] and $\mu_s'$ in the Wavelength Range of 450 – 600 nm

<table>
<thead>
<tr>
<th>Two different spectrometers</th>
<th>Puck-Calibration</th>
<th>Self-Calibration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target/Reference</td>
<td>Same Day</td>
<td>Different Day</td>
</tr>
<tr>
<td>% Error in [total Hb]</td>
<td>6 ± 5</td>
<td>8 ± 5</td>
</tr>
<tr>
<td>% Error in $\mu_s'$</td>
<td>5 ± 4</td>
<td>5 ± 4</td>
</tr>
<tr>
<td>Same spectrometer</td>
<td>Same Day</td>
<td>Different Day</td>
</tr>
<tr>
<td>% Error in [total Hb]</td>
<td>8 ± 5</td>
<td>8 ± 5</td>
</tr>
<tr>
<td>% Error in $\mu_s'$</td>
<td>13 ± 6</td>
<td>13 ± 6</td>
</tr>
</tbody>
</table>

3.3 Self-calibrated vs. puck-calibrated clinical spectra from the cervix in vivo

Representative clinical spectra (450 – 600 nm) and extracted optical parameters from two colposcopically normal cervical sites calibrated using both puck and self-calibration are shown in Fig. 3. Solid lines and broken lines indicate bests fits (N = 100) to the puck calibrated spectra and self-calibrated spectra, respectively. Error bars in Fig. 3 represent standard deviations between three repeated scans, which are small compared to difference attributed to different calibration standards. All spectra calibrated post-measurement using a Spectralon puck should have calibrated intensities of less than unity since the puck has a diffuse reflectance of nearly 100%. Figures 3(a)–3(c) are from a site in which the puck-calibrated reflectance ratio was less than one and both calibration methods led to an excellent fit. For this site, diffuse reflectance divided by the puck is actually higher than that divided by the self-calibration channel. Although the extracted $\mu_s(\lambda)$ are identical, puck-calibration led to extraction of higher $\mu_s'(\lambda)$ than those using self-calibration. In some instances, puck-calibrated diffuse reflectance spectra had values that exceeded unity which maybe a result of significant...
system drift between the time of measurement and calibration with the puck. Figures 3(d)–3(f) are from a site in which the puck-calibrated reflectance is greater than one resulting in a poor fit and extraction of absorption and scattering that reached the floor and ceiling, respectively, set in the least squares search algorithm. Only the self-calibrated spectrum in this case resulted in an excellent fit. Thus, all puck-calibrated spectra that exceeded a ratio of greater than one, led to poor fitting and, consequently, reached the pre-set boundary constraints in the least squares optimization algorithm. These spectra were eliminated from subsequent analysis.

Fig. 3. Representative (a) diffuse reflectance, (b) extracted $\mu_a(\lambda)$, and (c) $\mu'_a(\lambda)$ from a colposcopically normal site in which the puck-calibrated reflectance ratio was less than one and both calibration methods led to a good fit. Although the extracted $\mu_a(\lambda)$ are identical, puck-calibration led to extraction of higher $\mu'_a(\lambda)$ than those using self-calibration. Representative (d) diffuse reflectance, (e) extracted $\mu_a(\lambda)$, and (f) $\mu'_a(\lambda)$ from a colposcopically normal site in which the puck-calibrated spectra exceeded one and led to poor fitting and the extraction of optical properties that reached boundary constraints in the least squares search algorithm. PC: puck-calibrated and SC: self-calibrated. Red asterisks and blue diamonds represent measured diffuse reflectance that has been calibrated using the puck and self-calibration measurement, respectively. Solid lines indicate best fits using puck calibrated spectra and broken lines indicate self-calibrated spectra. Error bars indicate standard deviations from three repeated scans.

Box and whisker plots of extracted [total Hb], Hb saturation (HbSat), and $<\mu'_a(\lambda)>$ using puck-calibrated and self-calibrated data are shown in Fig. 4. The number of sites shown in Fig. 4 excluded data which resulted in zero extracted absorption using puck-calibrated data (18 out of total 33 colposcopically normal sites). Although extracted [total Hb] can vary by up to 17%, the extracted absorption parameter, [total Hb] ($P = 0.86$) and HbSat ($P = 0.15$) are not significantly affected by the calibration method used when considering all colposcopically normal sites. However, $<\mu'_a(\lambda)>$ is significantly associated with the calibration method used (*$P < 0.03$) when considering all sites and could vary by over 20% depending on the calibration method used. Disparities between the results from the two calibration methods in the phantom studies (puck calibration resulted in higher $<\mu'_a(\lambda)>$ values) were recapitulated in the clinical measurements, i.e., the extracted $<\mu'_a(\lambda)>$ showed higher median values and greater variance when calibrated with a puck post-measurement in the cervical tissue studies. The results above suggest that there are system drifts that are corrected by the self-calibration, but not by the puck calibration performed at a later time point. Since the same coating (i.e., Spectralon®, LabSphere, North Sutton, New Hampshire) is used in the self-calibration channel and in the puck, both calibrations have the same spectral response although the absolute intensities may
differ due to different measurement geometry. However, the difference in reflectivity between the two methods alone could not have caused the calibrated reflectance to differ since the the puck- and self-calibrated tissue spectra are each divided by a specific reference phantom spectrum, which itself is also calibrated in an identical manner as the tissue spectral data. This ratio should cancel out any systematic effects attributed to reflectivity of the material in the calibration standard. Calibration of the tissue spectra by reference phantom spectra is a necessary step prior to inversion with the inverse Monte Carlo model [41].

Another confounding variable commonly encountered in contact probe spectroscopy is the applied probe pressure. Representative diffuse reflectance spectra corrected through self-calibration at low, medium, and high applied probe pressures from the same colposcopically normal site are shown in Fig. 5. Reflectance seems to increase with the applied pressure, leading to decreased absorption and increased scattering. Variation within each applied pressure is small (error bars in Fig. 5) compared to differences between different pressure levels.
Fig. 5. (a) Representative diffuse reflectance (450 – 600 nm) from a colposcopically normal site calibrated using self-calibration channel at low (red asterisks, color online), medium (blue diamonds), and high (black triangles) contact pressures. Error bars indicate standard deviation between three repeated scans at each pressure. Dashed lines are best least squares fits (100 fits) to the mean of the measured diffuse reflectance using the Monte Carlo-based inverse model using self-calibration. Diffuse reflectance increases as the applied contact pressure increases. (b) Extracted absorption spectrum ($\mu_a(\lambda)$) from the same colposcopically normal site at low (red broken line), medium (blue dashed line), and high (black solid line) pressures. [Total Hb] at different pressures varied over 11.5 µM (103% of value at medium pressure). (c) Extracted reduced scattering spectra ($\mu_s'(\lambda)$) from the same colposcopically normal site at low (red broken line), medium (blue dashed line), and high (black solid line) pressures. Extracted $\mu_s'$ at different pressures varied over 3 cm$^{-1}$, or 30% of the value extracted at medium pressure.

To account for inter-patient and inter-site variations and isolate the effect of applied probe pressure, extracted absorption and scattering at low pressure were subtracted from those at medium and high pressures, respectively, as shown in Fig. 6. [Total Hb] decreased significantly with applied pressure ($P < 0.01$ and 0.05 for high vs. low and medium vs. low pressures, respectively). $\mu_s'$ also significantly increased with applied pressure (high vs. low pressure with $P < 0.005$). Hb Saturation was not significantly associated with pressure as it is a normalized quantity of oxygenated Hb and [total Hb], which are similarly impacted with increasing pressure.
4. Discussion

We have presented the effect of two common confounding factors – calibration and contact pressure – on the extraction of absorption and scattering contrast, namely [total Hb] and $\langle\mu_s'\rangle$, respectively. Scattering contrast was especially sensitive to shifts in system throughput and hence was significantly affected by the calibration technique used. Hence, a real-time self-calibration channel should be used to accurately decouple tissue diffuse reflectance spectra from instrument dependent response. Absorption contrast, specifically [total Hb] and HbSat, were more sensitive to spectral shape as opposed to absolute intensity; and hence were not significantly associated with the calibration technique used. The applied pressure significantly affected the extraction of [total Hb] and $\langle\mu_s'\rangle$.

Self-calibration offers many advantages over one-time puck calibration measurements. Since system throughput such as fiber bending or source fluctuations may depend on the actual physical configuration of the system or vary over time, it is important to capture the variation in real-time through a self-calibration channel, as opposed to a one-time puck calibration measurement. Drifts in system response can result in significant differences in extracted scattering contrast, which is heavily dependent on the intensity of the diffuse reflectance measured. Extracted absorption contrast, however, may be more dependent on the shape of the spectrum as opposed to the calibrated intensity. Previously, several groups including ours [31,35,37] have observed a decreasing, yet not statistically significant, trend in scattering with dysplastic transformation of the cervix using quantitative diffuse reflectance spectroscopy. A real-time calibration channel could potentially minimize errors due to variations in system throughput and enable the observation of such scattering contrast. An integrated self-calibration channel also obviates the need for separate calibration measurements (typically acquired before or after the tissue measurements) and thereby significantly reduces the training requirement and operation burden for the user.

Ruderman et al. [54], using polarization gated reflectance spectroscopy on oral mucosa in vivo, observed a significant difference between [total Hb] and scattering intensity, but not in oxygenation, between gentle and firm pressures (0.009 – 0.012 N/mm² and 0.15 – 0.20 N/mm², respectively). When applied pressure increased, [total Hb] and total scattering intensity from their cross-polarization channel (i.e., deeper oral mucosa) decreased and increased, respectively, which were similar to what we observed in cervical mucosa.
Compression and displacement of local vasculature and scatterers likely led to the observed changes in [total Hb] and <µs> when pressure was varied. When applied pressure was high, elastic deformation of the tissue caused blood vessels to be pushed away and a decrease in overall blood content in the area under the probe tip. No significant change in HbSat was observed with varying pressures until six seconds after the application of firm pressure, relative to gentle pressure. Reif et al. [55] observed decreasing trend in oxygen saturation with increasing pressure (0.04 – 0.2 N/mm²) on thigh muscles in mice. Their integration times (i.e., up to five seconds) were approximately an order of magnitude longer than those used in this study and muscles may demonstrate greater oxygen saturation changes than epithelial tissue due to their greater metabolic rate.

The technology described in this paper is not intended to be the solution to low cost cervical screening but a tool that will allow for quality control in the measurement of cervical tissue spectra with minimal operator bias for testing in resource-limited setting or places where quality control needs to be automated rather than expert or environment dependent. If the sources of contrast obtained with this technology prove to be clinically useful, advances in optical technologies will be leveraged to reduce the cost of this technology. Future spectroscopic systems intended for clinical use, particularly where operator training is not viable, should incorporate a real-time self-calibration channel and collect diffuse reflectance spectra at a consistent pressure to maximize data integrity. By monitoring the ratio of the reflectance collected by the sampling channel to the self-calibration channel with the fiber optic probe secured to a repeatable standard daily, a non-technical user can triage if fiber damage (e.g., crack in fiber) has occurred and replace the probe when necessary. The addition of these important functionalities, as well as low operator training and field compatible power and package requirements, will enable the collection of reliable clinical data in resource-limited settings which can then provide the basis for redesigning future systems that can be implemented for low cost cervical cancer screening where colposcopy followed by diagnostic biopsy is not available.

Acknowledgments

The authors wish to acknowledge Julia Chapman for her assistance with clinical data acquisition in Haiti and Kevin Chang for his help in creating the custom LabVIEW software. This work was supported by a biomedical engineering for global health grant from the Duke Provost Office and John T. Chamber’s Fellowship.