Effect of hemoglobin concentration variation on the accuracy and precision of glucose analysis using tissue modulated, noninvasive, *in vivo* Raman spectroscopy of human blood: a small clinical study

Joseph Chaiken
William Finney*
Syracuse University
Department of Chemistry
Syracuse, New York 13244-4100

Paul E. Knudson†
Ruth S. Weinstock
Muhktar Khan
Joslin Center for Diabetes at
Upstate Medical University of New York
Syracuse, New York 13210

Rebecca J. Bussjager
Douglas Hagrman
Pamela Hagrman
Yiwei Zhao
LighTouch Medical, Inc.
600 E. Genesee Street
Syracuse, New York 13210

Charles M. Peterson
National Institutes of Health
National Institute of Heart Lung and Blood Diseases
Bethesda, Maryland 20892

Karen Peterson
National Institutes of Health
National Institute on Alcoholism and Alcohol Abuse
Bethesda, Maryland 20892

Abstract. Tissue modulated Raman spectroscopy was used noninvasively to measure blood glucose concentration in people with type I and type II diabetes with HemoCue fingerstick measurements being used as reference. Including all of the 49 measurements, a Clarke error grid analysis of the noninvasive measurements showed that 72% were A range, i.e., clinically accurate, 20% were B range, i.e., clinically benign, with the remaining 8% of measurements being essentially erroneous, i.e., C, D, or E range. Rejection of 11 outliers gave a correlation coefficient of 0.80, a standard deviation of 22 mg/dL with $p<0.0001$ for $N=38$ and places all but one of the measurements in the A and B ranges. The distribution of deviations of the noninvasive glucose measurements from the fingerstick glucose measurements is consistent with the suggestion that there are at least two systematic components in addition to the random noise associated with shot noise, charge coupled device spiking, and human factors. One component is consistent with the known variation of fingerstick glucose concentration measurements from laboratory reference measurements made using plasma or whole blood. A weak but significant correlation between the deviations of noninvasive measurements from fingerstick glucose measurements and the test subject's hemoglobin concentration was also observed.
[DOI: 10.1117/1.1922147]

Keywords: Raman spectroscopy; noninvasive glucose; metabolic monitoring.

Paper SS04149RR received Aug. 2, 2004; revised manuscript received Jan. 31, 2005; accepted for publication Feb. 7, 2005; published online Jun. 7, 2005.

Introduction

Studies[1] show that intensive self-monitoring of blood glucose by people with diabetes can allow them to maintain blood glucose concentrations at near normal values. Improved glucose levels delay the onset and progression of long term consequences of poorly managed diabetes including, but not limited to, peripheral neuropathy, circulatory damage, retinopathy, and early death. The advent of commercial fingerstick devices allowing self-monitoring in the late 1970s represented a landmark in diabetes care. These observations stimulated an effort[2] to discover a totally noninvasive method. In this paper we describe the results of a pilot clinical study to evaluate the accuracy and precision of a noninvasive technique based on tissue modulated Raman spectroscopy.

We reported[3] in 2000 and in a succession of subsequent articles[4–6] the first noninvasive Raman spectra of human blood *in vivo*. Those earlier studies were designed to establish that the source of the spectra was indeed blood and that the spectra were volume normalized so that quantitative noninvasive blood analysis should be possible. Since the spectra contain features that are easily seen to be associated with many well-known biological materials,[7] we found[3] that glucose is an important and feasible target analyte. Raman spectroscopy has long been an important technique for analysis of mixtures[7,8] at the millimolar concentration levels with or without the use of common chemometric[9] methods and thus could be used for other analytes in blood.

In this paper we describe improvements in instrumentation and methodology as well as in blending human performance factors into the overall method. Although a technique might be quite useful as a research tool for trained scientists and lab technicians, the same technique must meet many requirements if it is to be applicable for the large scale self-monitoring of blood glucose by untrained persons. While there have been studies[2,10] (e.g., near infrared transmission and diffuse reflectance) showing how other noninvasive or minimally invasive

*Present address: University of Michigan, Department of Chemistry, Ann Arbor, Michigan 48109.
†Present address: Medical College of Wisconsin, 8701 Watertown Plank Rd. Milwaukee, Wisconsin, 53226.
Address all correspondence to Joseph Chaiken. Tel: 315-443-4285; Fax 315-443-4070; E-mail: jchaiken@syr.edu

1083-3668/2005/$22.00 © 2005 SPIE

techniques can be used to monitor blood glucose in the hyperglycemic range, any technique applicable for the large scale self-monitoring of blood glucose must provide accurate and precise measurements in the normal glucose concentration range (79–129 mg/dL). Although there is utility in having a nonportable device for clinical settings, an inexpensive portable device is necessary if it is to be used by patients throughout the day. The algorithm employed in this study was developed because it resembles the use of single channel detectors and filters as we suspect would be required for miniaturization. Blood and circulatory abnormalities such as abnormal hemoglobin concentrations and peripheral circulatory and nervous pathologies need to be addressed before a noninvasive blood glucose monitor can be marketed to patients with diabetes.

Experimental Apparatus and Procedures

In accordance with our Institutional Review Board approved protocol, all subjects provided informed consent. Subjects included 23 males and 2 females with diabetes mellitus ranging in age from 21 to 70 years. All the subjects had one or more of the following conditions to varying degrees, neuropathy, nephropathy, and retinopathy. Statements regarding the general applicability of the spectroscopic approach can only be made regarding body mass index and age since a wide range was sampled in this study. On the other hand, there was only one Afro-American, some subjects of Hispanic ethnicity, and a large majority of Caucasian subjects so, although no obvious bias was observed, the applicability of the spectroscopic approach to subjects of differing skin tones is weakly sampled at best.

The experimental apparatus (LighTouch device) produced a glucose determination that was compared to HemoCue (HemoCue, Lake Forest, CA) glucose measurement, a validated device for capillary glucose.[11–13] The hemoglobin was measured using the HemoCue Hb device. The LighTouch device was operated by a technician not having an advanced physical science or engineering degree. Data were archived by an independent party who also performed all fingerstick blood glucose and hemoglobin measurements. Different lots of HemoCue[14] test cuvettes were mixed randomly throughout the study and whenever possible, i.e., with the discretion of the test subjects, fingerstick measurements were repeated and the results of the two fingersticks were averaged. On a single occasion two fingerstick measurements were made within minutes of each other and they differed by a large amount, i.e., $\approx$100 mg/dL. Although spiking is a known occurrence, a third measurement was attempted to be sure. The three measurements were compared and one measurement was eliminated on the basis of a Q test at 90% confidence. The fingerstick average was always paired with a single LighTouch measurement that was performed within 3–4 min of the set of HemoCue measurements.

Previously described tissue modulated spectroscopy[3–5] was used for the present study with certain modifications. The results obtained in this study were obtained using 31 mW at the sample, continuous wave, 785 nm wavelength excitation from an external cavity diode laser (Sacher Laser, Marburg, Germany) that is filtered by two laser line excitation clean-up filters purchased from Omega Optical (Brattleboro, VT). This laser power produced no sensation as demonstrated by the fact that no test subject was able to discern whether the laser was on or off. This power and focusing corresponds to less than the maximum permitted exposure,[15] which is about 33 mW. Although going to longer wavelengths will permit an increase in the maximum permitted exposure as well as other benefits (e.g., reduced fluorescence), we note that we have obtained quite good results on the same overall timescale using 785 nm excitation with powers as low as 19 mW at the sample. The filtered beam was focused using a single fused silica lens to a nominal spot size of 100 μm diameter when the laser spot is observed using a flat surface inserted into the beam at normal incidence. This is a nominal spot diameter because in practice the laser impinges on the stratum corneum of the volar side of the distal segment of the middle finger of the test subject at an angle of 53 deg. Therefore the actual spot shape on a nonscattering target is approximately elliptical with a major axis of at least 167 μm and a minor axis of about 100 μm.

Tissue modulation was accomplished using a 2.1 mm diameter hole in a 1-mm-thick aluminum plate. The hole also serves as the aperture for the light and is beveled outward on the side opposite the surface that contacts the fingertip. Three small nubs arranged in a triangle around the hole on the side facing the fingertip have dimensions taken from standard Braille. These nubs allow test subjects to use their sense of touch to orient the location of their fingertip relative to the aperture. This hole/aperture size is smaller than earlier prototypes, allowing substantially less protrusion of stratum corneum into the aperture thereby providing both a more mechanically stable focal point for the optical system and a more uniform stress field to affect the tissue modulation process itself. After a measurement cycle, the nubs and the hole produce temporary but observable indentations on the skin allowing the LighTouch operator to assess the position of the measurement, to suggest to what degree the test subject was motionless during the measurement cycle, and to indicate how hard the test subject pushed against the tissue modulator during the pressed period. It is important that the aperture be motionless throughout the tissue modulation cycle.

The light emanating from the irradiated zone is collected and collimated by a fused silica single lens before it is filtered by a holographic notch filter, (Kaiser Optical Systems, Ann Arbor, MI), and subsequently refocused by another lens onto the input side of a hexagonal packed, nearly circular profile, 60 fiber$\times$100 μm, fiber bundle. The fiber bundle, (Process Instruments, Salt Lake, UT), is configured to form a line image on the output side where it brings the light to a 1200 grooves/mm spectrograph (also Process Instruments, Salt Lake, UT). The entire collection and dispersal system is approximately $f=2.1$. The spectrograph disperses the collected light onto a charge coupled device (CCD) camera (Andor Technologies, South Windsor, CT) having 256 vertical and 1024 horizontal pixels and is operated at $-85\,^{\circ}$C.

In order to explore quantitatively the tissue modulation process, some experiments were performed with a TA.XT texture analyzer (Stable Micro Systems, Surrey, England). There are many ways to configure a TA.XT. In our study the TA.XT precisely moves a mechanical probe while simultaneously recording the force, displacement, and time of the probe with respect to a fixed reference position and time. The force and

displacement are both recorded with a bandwidth of 0.2 kHz, with the force accurate and precise to ± 100 mg and the displacement resolution accurate and precise to ± 10 μm. We used a probe of our own design that contacts the volar side of a fingertip with a flat aluminum surface having a 0.21 cm hole such that the overall interaction of the probe with the fingertip is essentially identical to that of a fingertip making contact and pressing on the tissue modulator aperture. The finger itself is held at rest in a bed also of our own design that is machined out of aluminum. The bed has a half-cylindrical cross section and allows the test subject's hand and fingertip to rest comfortably while the TA.XT brings the probe down onto the fingertip. In this way the fingertip remains motionless while the TA.XT probe moves.

A LighTouch device data collection sequence for an uninitiated test subject begins with a short training session. Sitting in a dark room with only the laser and computer monitor turned on, the test subject is allowed to observe the wavelength-dispersed output of the Andor camera in real time, i.e., a continuously updated sequence of 20 ms frames. Initially the test subject is requested to place only sufficient pressure against the aperture as is needed to insure that the skin is flush against the metal forming the aperture. While there are independent means to assure that this is the case, in this study the test subjects used their tactile powers. Typically this ''unpressed stage'' of a tissue modulation cycle corresponds to about 1 N total force. On a single frame basis[3] a spectrally broad fluorescence is always observed in addition to small but unmistakable Raman features corresponding to amide I and CH_2 deformation modes at about 1670 and 1450 cm^{-1} Raman shift, respectively. The test subject is then requested to push gently against the aperture while watching the real time response of the CCD camera. As long as the physical contact is not broken, any additional pressure, typically 2–7 N depending on the relative size of the person's finger and the person's blood pressure, results in a relative emptying of the irradiated capillary bed. In this ''pressed'' state, all subjects were able to observe the fluorescence signal and Raman signal decrease in concert. Instructing the test subject to release some pressure, *while maintaining continuous contact with the aperture*, makes an equally obvious increase in the CCD response. The test subject is invited to press and release a few times while maintaining constant contact and observing the real-time response to gain experience with the ''feel'' of the process.

Observing these changes allows the test subject to calibrate their own hand as a ''servo'' unit with regard to producing either a pressed or unpressed state. In some cases we allowed the test subject to experience a short practice run in which he/she was instructed to produce an unpressed state for 10 s and then a pressed state for 10 s during which time he/she was not permitted to see the real time CCD response because the computer monitor was intentionally turned off. The transition between states was initiated by audio cues from the software to the test subject. Afterwards the software produced a graphical representation of the blood volume versus time profile by plotting the integral of the fluorescence (see later) as a function of frame number, i.e., time. This representation allows the test subject and LighTouch operator to measure how steady the finger was held during the entire test period and to check how well the two stages of the tissue modulation process were executed. This training nearly always produced a test subject who was confident in his/her ability to execute an unpressed to pressed tissue modulation sequence. The entire training period never exceeded 5 min.

For this study, a measurement sequence consisted of 100 s of unpressed and 100 s of pressed states. All testing occurred in the dark without benefit of any real-time feedback of any kind to either the LighTouch operator or the test subject other than an audio cue to transition from the unpressed state to the pressed state. To test the difficulty in teaching/learning to execute a tissue modulation cycle, no test subject was ever given two chances to obtain a glucose reading. It is clear, e.g., by use of pressure sensors, observing the deformation pattern on the fingertip skin and other measurements, that the applied pressure in the two states and the position of the measurement varies from one test subject to another. It is also possible to observe some kinds of unintended motion during the tissue modulation cycle by observing the blood volume versus time plot immediately after completion of the cycle. Using this plot the LighTouch operator can select a set of unpressed frames and a set of pressed frames in equal numbers that are then coadded, respectively, before being subtracted, accumulated pressed from accumulated unpressed, to yield a tissue modulated fluorescence/Raman spectrum. In this way, at least some frames that were corrupted by unintentional motion or other sources of artifacts, e.g., obvious CCD spikes, could be excluded from subsequent processing.

The tissue modulation process as implemented in this study is not invulnerable to artifacts associated with surface imperfections on the size scale comparable to the physical extent of the laser spot and larger. Thus surface imperfections like cracking caused by excessive skin dryness, or trauma caused by physical injury, e.g. scarring due to long-term fingerstick blood glucose measurements, can be expected to lead to spurious raw data and therefore spurious blood analyte concentrations Although it is possible to use other fingers for measurements in some cases without recalibration, in this study only the middle finger was used. The data presented in the results section correspond to the results of every tissue modulation cycle regardless of the conditions of the test subjects' skin or peripheral circulatory or nervous systems. Patients were accepted for study based on their willingness to participate and their fingers were not examined before being invited to participate. The reproducibility of measurements on single individuals or simply people with similar skin condition makes us suspect that prescreening of subjects would lead to better results.

Extracting the glucose concentration from the tissue modulated spectra was accomplished in a similar but not identical manner as previously published.[3] As can be seen in Figs. 1 and 2 the modulated spectrum contains both fluorescence and Raman features. From those spectra and previously published results based on *in vivo* and *in vitro* spectra of authentic glucose, we have ascertained the wavenumber range, $\approx$375–$\approx$686 cm^{-1}, contains the most glucose information balanced against the least tendency to contain off-axis Rayleigh scattered light. To illustrate the starting point for that procedure the spectrum of a large glucose concentration-spiked gel spectrum is included in Fig. 2. The glucose spiked calibration material contains glucose at 1 M concentration and as such the glucose Raman signal overwhelms all other contributions

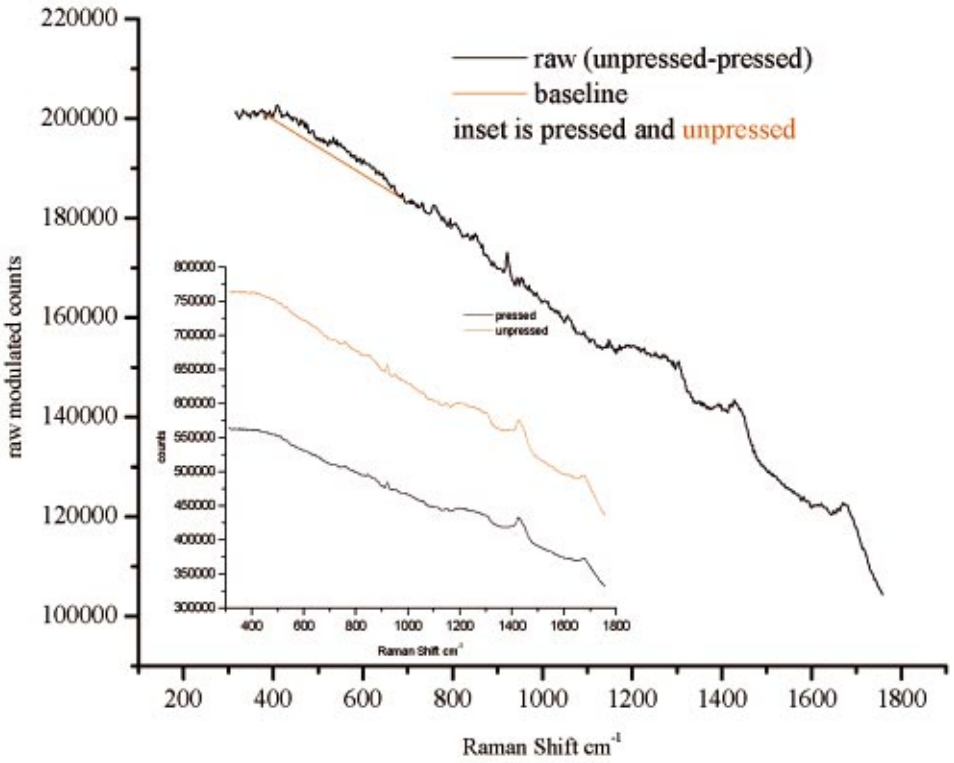

Fig. 1 Representative raw pressed (black inset) and unpressed (red inset) spectrum and difference spectra, not corrected for instrument response.

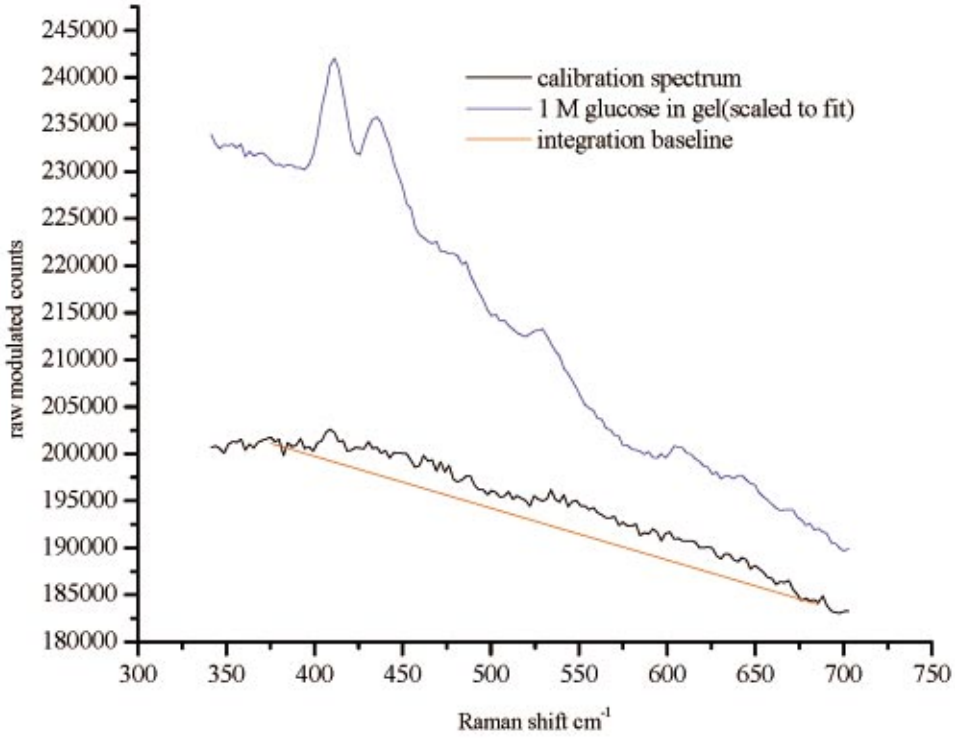

Fig. 2 Raw difference spectrum (black) showing base line (red) used for integration. Also shown in blue is an appropriately scaled spectrum of ≈1800 mg/dL glucose in gelatin that allows observation of relevant glucose features.

at virtually all Raman shifts. To obtain an integral over the same Raman feature(s), the same set of integration wavenumber limits found in our earlier studies were used as follows. First, a set of ten pixels were averaged at each of the end points of the same spectral region to define the endpoints of a base line. A straight line between these two points was used as the base line. The raw tissue modulated spectrum was integrated down to the base line between the end points as indicated in Fig. 2.

To obtain blood volume normalization for the spectra, each raw tissue modulated, i.e., difference, spectrum was integrated to zero from about 1000 cm^{-1} of Raman shift to the highest shift accessible with the CCD detector which was usually about 1800 cm^{-1}. Because the fluorescence constitutes the majority of the emission at all Stokes shifted wavelengths relative to the exciting wavelength, and because whatever amount Raman contributes to that emission is also a measure of the volume of material that responds to tissue modulation, simply summing the raw counts in the difference spectrum, over the range least affected by unfiltered Rayleigh and other stray light, is sufficient to obtain a measure of the blood volume. This integral is then divided into the Raman feature integral shown in Fig. 2. We suggest that at least in principle, it should be possible to implement this essentially digital procedure in the analog domain using optical filters and single channel detectors. In this case we might use one narrow filter for each of the ten-point end point averages, a third filter for the $\approx$350–$\approx$600 cm^{-1} glucose integration and a fourth filter for the $\approx$1000–$\approx$1800 cm^{-1} blood volume measurement. A portable device designed to monitor only glucose might be possible using this scheme although this is a subject of further work. Generally speaking, for multiple analytes, we would expect that dispersive optics, multichannel detection, and digital processing would be more appropriate.

Previously[3] we referred to the ratio of these integrals as glucose concentrations in ''integrated normalized units,'' (INUs). To calibrate the LighTouch device INUs to mg/dL for human testing, a series of measurements, i.e., integrals of Raman features in blood volume normalized spectra, were paired with contemporaneous fingerstick glucose measurements in mg/dL. The resulting data pairs were plotted, fit to a linear regression and then inverted to yield blood glucose values in terms of subsequent ratios of integrals from identically processed data. We have found that such a regression based on data from one or more individuals yields a calibration that can be applied to anyone else.

Results

The size of fingers varied as did the amount of pressure exerted by each subject, in either the pressed or unpressed stages. The position of the irradiated zone also varied because the tissue modulation aperture was oriented in the same manner for all the subjects but each subject fit in differently. Representative raw data from $\approx$350 to $\approx$1800 cm^{-1} corresponding to the pressed and unpressed states and their difference are shown in Fig. 1. Figure 2 shows the $\approx$350–$\approx$600 cm^{-1} range as well as the integration range and the baseline that was used to obtain a glucose measurement. For comparison, $\approx$1800 mg/dL glucose in gelatin, arbitrarily scaled to allow easy visual comparison, is also shown. The representative raw data corresponded to a HemoCue blood glucose of $\approx$100 mg/dL.

To calibrate the device, one individual (male, aged 68, 165 lbs, Caucasian, type II diabetic) performed 18 fingersticks over 2 days to obtain the paired LighTouch-HemoCue data shown in Fig. 3. Although included in the data set plotted in Fig. 3, for calibration purposes one outlier was rejected based on a Q test[16] and another on the basis of an instrumental inconsistency detected after the measurements were completed (wrong time per frame was inadvertently selected). This was on a day when less than ten points were collected and a small sample Q test with 90% confidence limits was appropriate. The remaining points were fit with a linear regression that was subsequently inverted to yield a linear transform for the raw LighTouch INU data to glucose in mg/dL. Over the next 14 weeks the same person contributed seven additional measurements that were combined with single measurements on 24 different people. Figure 3 contains the entire data set obtained from the 25 volunteers including the initial 18 measurements used to provide the concentration calibration in a Clarke error grid.

The error grid was developed[17] to allow comparison of the performance of different kinds of fingerstick based devices since simple correlation coefficient and other statistical measures do not fully relate to clinical utility. Furthermore the zones allow for easier discussion of individual points. The zones in the grid are labeled with letters having the following meanings. Zone A denotes ''clinically accurate.'' Zone B denotes ''clinically benign'' or ''acceptable'' because such values lead to no treatment of the patient. Zone C is ''unacceptable'' because such values lead to over correction in the blood glucose level by the patient. Zone D denotes a ''dangerous failure to detect and treat'' and zone E leads to ''erroneous treatment.'' The LighTouch device produced 92% of measurements in the A and B zones. The remaining 8% of measurements were essentially wrong since they fell in the other zones. Six of the erroneous measurements are color coded to permit tracking them through some but not all of the analyses that follow.

In all but three cases the test subjects consented to an additional fingerstick for the purpose of measuring their blood hemoglobin concentration. The average of the hemoglobin range observed in this study was 14.3 g/dL$\pm$1.6 (1σ) with a range of 10.5–17.2 g/dL. It is known that either reference device (i.e., hemoglobin or glucose) used in this study has systematic bias and less precision at the extremes of hemoglobin concentration.[12–14] As can be seen in Fig. 4, there was no correlation between the HemoCue glucose level with the HemoCue hemoglobin level with $r=0.016$. We observe in Fig. 5 that the LighTouch glucose level plotted against the HemoCue hemoglobin measurement had a correlation coefficient of $r=-0.42$ when all points were included as in the inset. Note that there is a single point corresponding to LighTouch glucose of 364 mg/dL that exerts a large effect on the linear fit. This point also corresponded to the lowest hemoglobin observed and corresponds to the single point in the E zone in Fig. 3. Without that point in Fig. 5 we obtain linear correlation of $r=-0.27$.

As must also be true based on Figs. 4 and 5, the deviations between the LighTouch glucose measurements and the HemoCue glucose measurements in Fig. 6 are also weakly

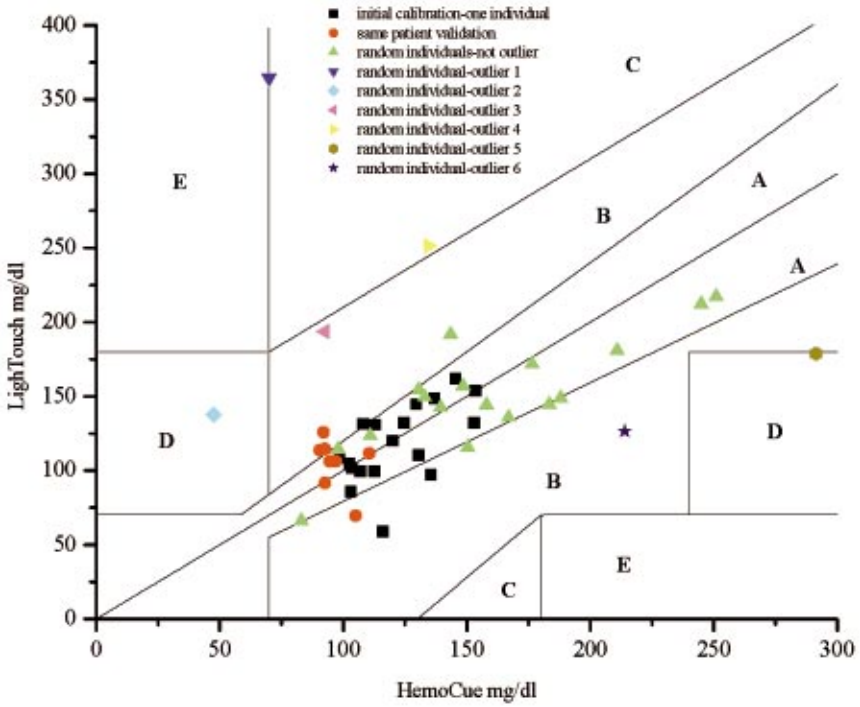

Fig. 3 Clarke error grid analysis of 49 data points. Zone labels are as described in text. The 18 one individual calibration points are shown in black. The eight same patient validation points are shown in red. The data points from random patients are shown in green triangles except for the ones we shall label and number as "random individual outliers." The various colors and shapes are maintained consistently in subsequent relevant figures to allow tracking of these points throughout the paper.

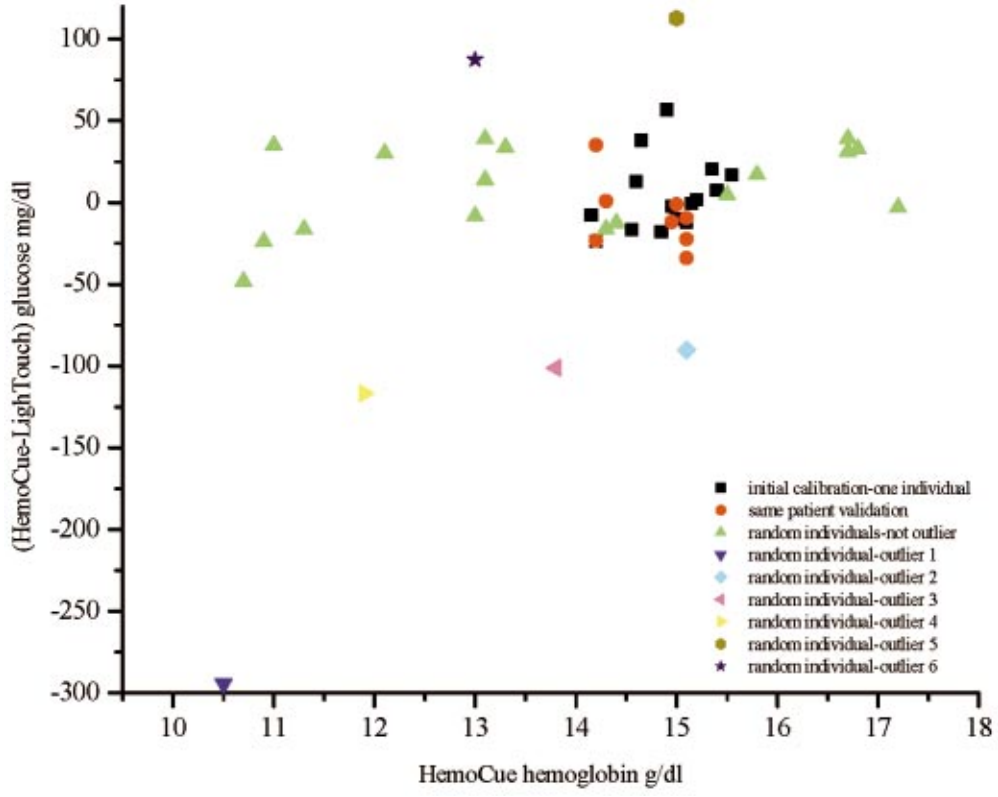

Fig. 6 The difference between HemoCue and LighTouch glucose measurements plotted as a function of HemoCue hemoglobin. With all points included as shown linear regression gives $r=0.38$, SD=56.0, $N=46$, $p=0.009$ and with one (E zone) outlier removed $r=0.19$, SD=41.2, $N=45$, $p=0.212$.

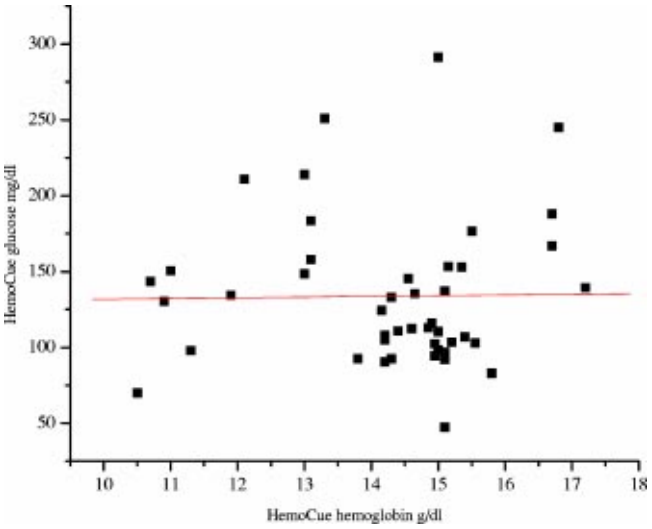

Fig. 4 Hemocue glucose plotted as a function of HemoCue hemoglobin. Linear regression gives $r=0.016$, SD=50. $N=49$, $p=0.92$.

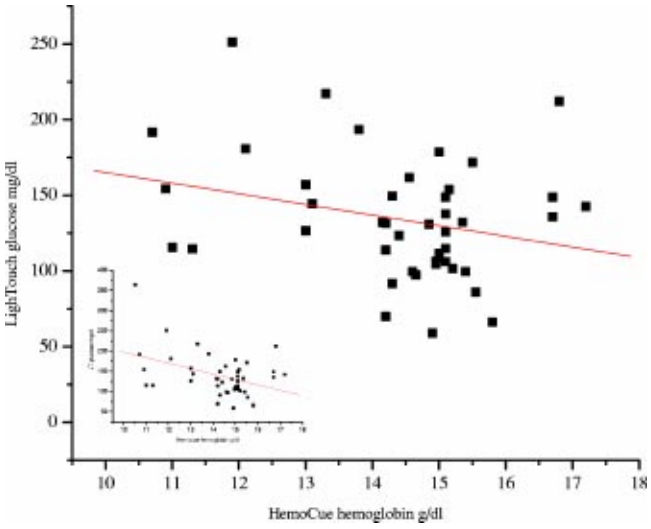

Fig. 5 LighTouch glucose plotted as a function of HemoCue hemoglobin. With all points included (inset) linear regression gives $r=-0.42$, SD=48, $N=46$, $p=0.004$ and with one (E zone) outlier removed $r=-0.27$, SD=39.1, $N=45$, $p=0.073$.

correlated with the hemoglobin concentration. The same E zone outlier point in all the earlier figures is evident and linear regression applied to whole data set produces $r=0.38$. Without that point there is a weaker linear correlation of 0.19 between the deviations and the blood hemoglobin concentration.

To further probe the role of the tissue modulation process itself in glucose concentration measurement, in Fig. 7 we plotted the deviations between LighTouch and HemoCue glucose measurements as a function of the ''total modulated fluorescence.'' The total modulated fluorescence is simply the integral of the tissue modulated light emitted from the capillary bed that is used for the blood volume normalization. Including all the data we find that the one E zone point again appears to be an outlier and including all the points linear regression produces a correlation coefficient of zero. Excluding this point yields a correlation of -0.19. The color coded erroneous points from the Clarke grid can be seen to form a ring around a central clump of points.

To probe the effect of varying the hemoglobin concentration on the tissue modulation process, we plotted the observed total modulated fluorescence as a function of the observed HemoCue hemoglobin measurement in Fig. 8. Using all the data in linear regression, $r=-0.07$ and dropping the data points corresponding to the four highest modulations, which appear to be outliers, we obtain $r=0.25$. The data suggest that the maximum amount of modulated fluorescence increases with the hemoglobin concentration. We did not measure nor attempt to control the amount of pressure applied by each of the test subjects and that would certainly be expected to have a large role in determining the amount of blood modulated. For a given hemoglobin concentration, pressing harder would be expected to move more blood and produce a larger fluorescence modulation. For all but one, which falls near the average hemoglobin concentration and the average modulated fluorescence, the color coded outliers from the Clarke grid form a ''boundary,'' indicated by the line connecting those points near the top of the cluster of data points, regardless of the hemoglobin concentration.

Figure 9 shows the results for two different size fingers, each from a different person, of a measurement of total force as a function of displacement of the TA.XT probe from the position of first stratum corneum contact, as the probe is pressed towards the bone (distal phalanx). The pressure is seen to increase as the probe is brought towards the bone while the finger is being essentially squeezed between the probe and the resting surface. The probe is then backed off rather quickly, 2 mm/s. This produces a type of hysteresis because the finger does not re-expand as fast as the probe is backed off from the position of highest compression. Based on the timescales involved, we assume that the blood and other fluids do not refill the capillary bed fast enough to maintain the total force experienced at the equivalent position during the squeezing cycle. The logarithm of the total force is plotted because it accentuates an inflection point that is always observed during the squeezing cycle. The position of the inflection point in both sets of data is marked by the horizontal line. The corresponding displacement for each of the inflection points is different because the gross size of the fingers is different. The petite young female displays the inflection point at a smaller displacement than does the larger male finger.

Since it is known[12] that the HemoCue itself has a bias at larger glucose values, we plotted the deviation between HemoCue and LighTouch as a function of the average between the HemoCue and LighTouch measurements (Fig. 10). A significant correlation was not found.

Discussion

The Clarke error grid analysis in Fig. 3 shows that 92% of the LighTouch measurements occur within zones A and B indicating that the LighTouch device could be used to measure blood glucose by various individuals with a single calibration. Therefore, in general it is possible to obtain high quality glucose concentration measurements across the range of hemoglobin values and blood volume modulations sampled. To im-

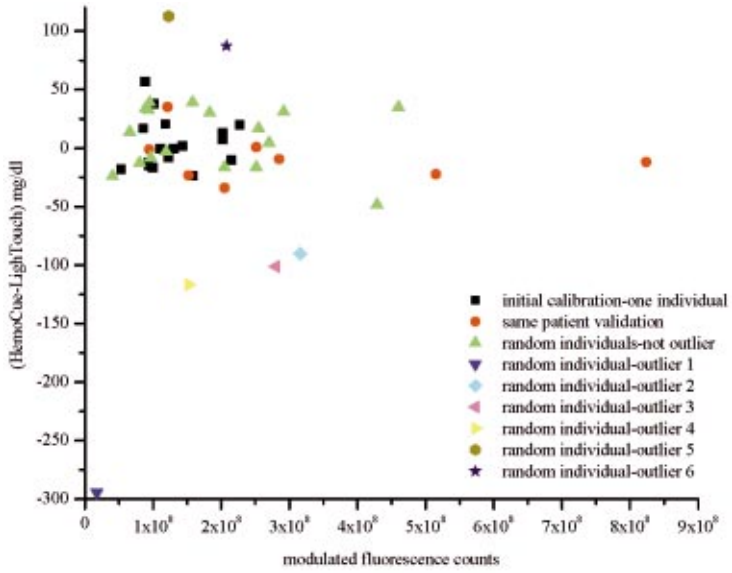

Fig. 7 The difference between HemoCue and LighTouch glucose measurements plotted as a function of total modulated fluorescence counts. With all points included linear regression gives $r=0.000$, SD=58.7, $N=49$, $p=0.97$ and with one outlier (E zone) removed $r=-0.19$, SD=41.2, $N=48$, $p=0.21$.

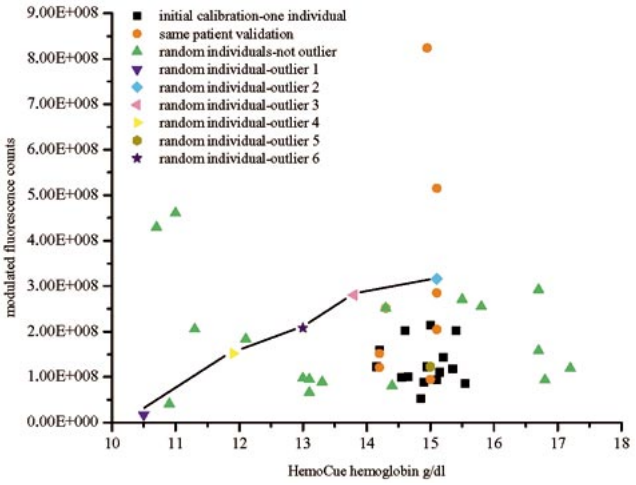

Fig. 8 The modulated fluorescence counts plotted as a function of hemoglobin concentration. With all points included linear regression gives $r=-0.07$, SD=1.45E8, $N=45$, $p=0.64$ and with four highest modulation outliers removed $r=0.25$, SD=7.47E7, $N=41$, $p=0.106$. Black line connects five of six "outliers" based on Clarke error grid analysis (Fig. 3) forming boundary for rejecting data as described in text.

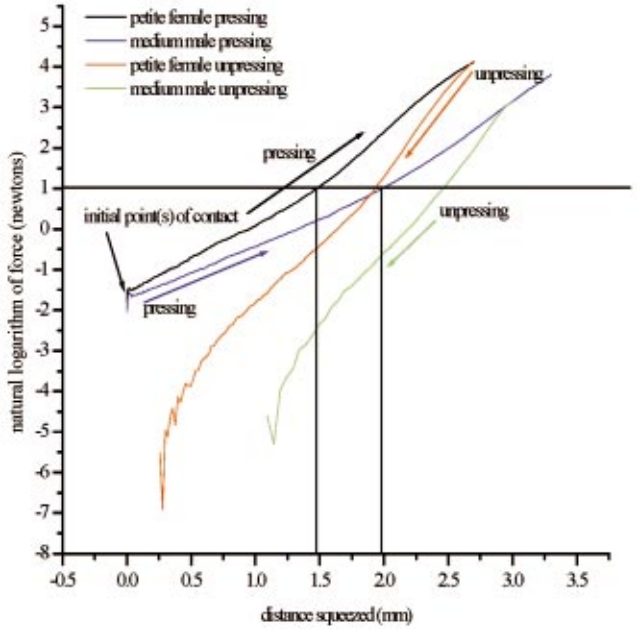

Fig. 9 TA.XT texture analyzer experiment simulating tissue modulation of fingertip showing natural logarithm of force variation with amount of pressing, i.e., displacement or distance squeezed from resting state and then during release of pressing.

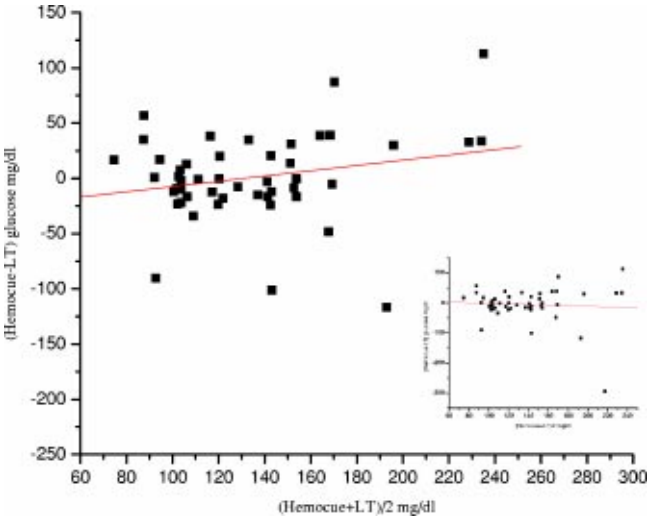

Fig. 10 The difference between HemoCue and LighTouch glucose measurements plotted as a function of the average of the HemoCue and LighTouch glucose measurements. With all points included (inset) linear regression gives $r=-0.07$, SD=58.5, $N=49$, $p=0.644$ and with one outlier removed $r=0.23$, SD=39.7, $N=48$, $p=0.12$.

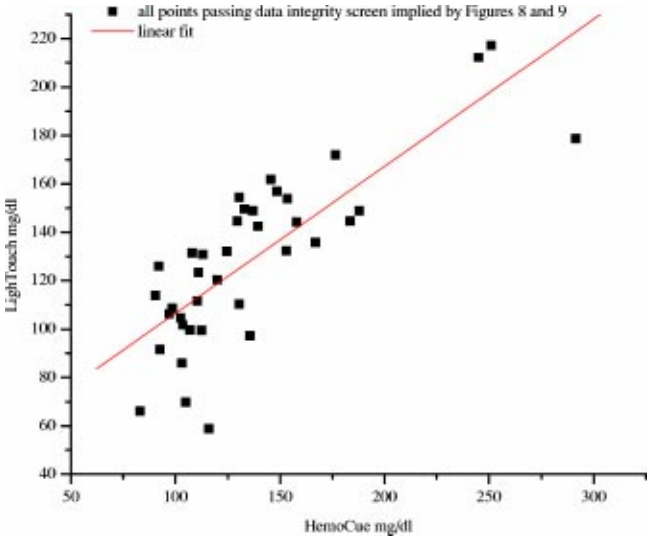

Fig. 11 LighTouch glucose plotted as a function of HemoCue glucose. With data selected as described in text $r=0.80$, SD=22, $N=38$, $p<0.0001$ and with three highest points excluded $r=0.70$, SD=21, $N=36$, $p<0.0001$.

prove the performance of the LighTouch device, the sources of the 8% of measurements that were erroneous need to be identified and if possible eliminated. We note that each concentration measurement is comprised of a blood volume measurement and a glucose measurement that to a large extent are independent of each other. The glucose based Raman signal is intrinsically much weaker than the fluorescence based blood volume measurement and to a large extent is much more spatially localized. Thus the two types of measurements are susceptible to only partially overlapping sources of errors. For just one example, the Raman signal is more susceptible to errors due to optical misalignment and other effects that can be traced to various sources. In what follows we shall attempt to account for the several outliers that can be seen in Fig. 3 on the basis of random and systematic errors.

As reported earlier[5] very small tissue modulations lead to low signal to noise. We suggest that very small modulations can be caused by either only a very small amount of blood actually being moved during the modulation process or because a reasonable amount of blood is moved but the hemoglobin concentration is much smaller than average. In the former case the Raman signal for glucose could have low signal to noise if only due to shot noise and so the concentration measurement will be compromised. The datum in the E zone likely corresponds to the latter situation since it also corresponds to an abnormally low hemoglobin concentration, the lowest observed in the study. As will be discussed later, in the latter case the Raman signal can have good signal to noise and the precision of the glucose concentration measurement would be expected to be good although biased in a systematic manner due to the variation in hemoglobin concentration. We do not think the E zone datum corresponds to some type of systematic error.

It is always possible to have an erroneous measurement at any particular hemoglobin concentration or blood volume modulation due to purely instrumental factors. No attempt was made to remove spikes from the CCD response during this study. True single pixel spikes can have only a very small effect on the blood volume measurements because hundreds of pixels are employed and the fluorescence is strong. Nevertheless, even a small spike in the region being used for glucose Raman measurement will usually be important. A later study should incorporate spike detection and removal on a per frame basis to minimize the effect of this source of error. Although we cannot be certain, a CCD spike in the spectral region used for the glucose Raman signal could easily be the source of the E zone datum.

Taking into account the modulated blood volume and hemoglobin concentration in the manner described later, resulted in Fig. 11. In this case we reject only the five data points that form the boundary in Fig. 8 and the six data points having larger modulated blood volume. Rejecting these 11 points results in $r=0.80$, SD=22, for $N=38$. At least two of the rejected points are not outliers in the sense that since leaving them in actually improves the correlation and decreases the standard deviation relative to the two cases just described. As will be described later, we suggest that rejecting the 11 points on the basis of being above the boundary does not require independent knowledge of the subject's hemoglobin concentration. It is our hypothesis that the 5 boundary points identified from the Clarke error grid analysis and the ones above the boundary points in Fig. 8 are associated with subjects who pressed too hard or too weakly during the pressed and unpressed tissue modulation stages, respectively.

The deterministic consequences of pressing too hard or weakly are discussed a few paragraphs later, however, if this is the case then measuring the applied pressure during the tissue modulation process and giving the "human servo" feedback in real time should avoid this problem and improve the yield of good measurements per attempts. Whether these suggestions are correct or not can be checked in a subsequent clinical study using a LighTouch device that incorporates the earlier real time feature. Since the higher glucose values have

a disproportionately favorable effect on the correlation, arbitrarily rejecting the three highest values results in $r=0.70$ for $N=35$ and all remaining HemoCue glucose values between 188 and 83. There were insufficient instances of HemoCue measurements in the hypoglycemic range, i.e., <79 mg/dL, to assess the efficacy of the LighTouch device in that range.

Torjman[11] and co-workers report that for an inhomogeneous population in comparison with a laboratory plasma glucose measurement for reference ($\pm 0.1\%$), the HemoCue, in either the glucose range 79 mg/dL and below, as well as 79–140 mg/dL, has a correlation coefficient of 0.80. For 140 mg/dL and higher, $r=0.97$. Thus since about half of our measurements, 47%, corresponds to glucose values of 140 or below, and the LighTouch must obviously have finite precision itself, in no case can we expect correlations much greater than about 0.80. Thus the observed values of the correlation coefficient for linear regression suggest accuracy and precision for the LighTouch in line with current fingerstick technology. In this respect, we note the measurements of Glasmacher[12] among others in which the correlation coefficient for a number of FDA approved commercially available fingerstick devices ranges from 0.78 to 0.88 referenced to a laboratory method and all in the normal glucose range. It should also be mentioned that some of these devices have a 1%–6% probability of producing D zone measurements in the hypoglycemic range.[13] In this study the LighTouch device produced a total of about 8% of points collectively in the regions other than the A and B zones. Thus the LighTouch performance compares reasonably well with commercially available fingerstick devices but the LighTouch device requires a little more improvement before it can be substituted for fingerstick devices.

The most important source of error in all fingerstick devices is undoubtedly human error.[18] Human error undoubtedly played a role in this study since the tissue modulation process was not executed uniformly by all individuals. We have found from on-going studies that real time feedback to the test subject is very useful in alleviating this problem. One could also automate the process to attain greater uniformity and less error. Improvements that allow a decrease in the overall measurement time, such as increasing the signal collection efficiency, would be expected to significantly decrease the rate of human error regardless of whether the tissue modulation is executed by the human servo or automatically.

The results of this study also suggest that there are subtle sources of small systematic error in the LighTouch process. Although Fig. 4 reveals no correlation for this particular data set ($r=0.02$, $N=46$, $p=0.92$) between HemoCue glucose measurements and hemoglobin measurements, linear regression in Fig. 5 suggests a weak correlation ($r=-0.27$, $N=45$, $p=0.07$, or greater) between LighTouch glucose measurement and HemoCue hemoglobin measurement. Essentially the same observation is contained in Fig. 6 suggesting a weak correlation ($r=0.19$, $N=45$, $p=0.21$) between the difference between paired HemoCue and LighTouch measurements and the hemoglobin concentration. The data suggest that, as expected, the LighTouch tends to underestimate the true glucose concentration for subjects with a relatively high hemoglobin concentration.

To see why this is expected, we first point out that the LighTouch uses *modulated* fluorescence emanating predominately from hemoglobin as a measurement of blood volume. To the extent that the skin does not change position or thickness during the modulation cycle, the skin contributes nearly no modulated fluorescence. If the hemoglobin concentration is above average then a given level of observed modulated fluorescence must emanate from a smaller than average actual modulated blood volume. If the hemoglobin concentration is below average then the same given level of observed modulated fluorescence must emanate from a larger actual modulated blood volume. For a given actual glucose concentration, a smaller actual modulated blood volume will lead to a smaller glucose Raman signal thus leading to an under estimated glucose concentration. Combining this effect with the same reasoning applied to a smaller actual modulated blood volume signal accounts for the negative correlation observed in Figs. 5 and 6. Note that the variation in hemoglobin concentration is about $\pm 20\%$ so the effect, while systematic, cannot be expected to be large.

Figure 7 suggests that there is an optimal modulation, roughly between 1 and 3E8 counts for the time duration, laser power, focusing conditions, and other factors characterizing this particular LighTouch device system. In attempting to understand and improve upon the tissue modulation process, we made no allowance for the fact that the fingers of all the test subjects do not have the same capillary density. The greater the density of capillaries[19] in the irradiated zone (on average about 50 capillaries/mm^2), the more blood that can be modulated. If all the test subjects had the same capillary density, and the capillaries had roughly the same size distribution with respect to the average erythrocyte diameter, then the modulated fluorescence should scale nearly linearly with the hemoglobin concentration. Of course, there is not necessarily any relationship between hemoglobin concentration and capillary density and this is reflected in Fig. 8. Within limits, hemoglobin concentration can change on a daily or even hourly basis depending on various factors including a person's degree of hydration as affected by excess sweating due to physical exertion/stress or even eating salty foods or taking diuretics. Capillary density is an anatomical characteristic that cannot be expected to change much if at all on any daily or even weekly time scale.

The lack of any relationship between hemoglobin concentration and capillary density is reflected by the fact that at any particular hemoglobin concentration there is a vertical range of possible modulation, ''below the boundary,'' that yields accurate glucose concentrations. We expect and observe that as the hemoglobin concentration goes to zero, so does the size of the modulation. According to Fig. 8, the lower (higher) the hemoglobin concentration the lower (higher) the maximum modulated blood volume that was observed to yield an accurate glucose measurement. Both the weakness of the observed positive correlation and the observed spread of the measurements even after rejecting possible outliers are consistent with the expected variation in physical size, capillary density, and blood viscosity across test subjects.

In an earlier study[5] we observed that such outliers were often associated with test subjects who had previously been diagnosed with conditions normally associated with peripheral vascular disease. By their own statements during testing, people with diabetic neuropathy cannot feel how hard they are pressing the aperture during either part of the tissue modula-

tion cycle. The harder one presses, the greater the likelihood that the tissue modulation aperture will move during the measurement and, equivalently, the greater extrusion of skin into the aperture. Since the light collection system is designed accept light from the aperture in the plane of the tissue modulation plate, changing the position of the skin in either direction with respect to the plane of the aperture tends to decrease the amount of light collected. Therefore, a similar effect is expected during the unpressed stage if one does not press hard enough to maintain proper registration, i.e., the stratum corneum is behind the aperture. Either of these effects affects the collection efficiency of the Raman part of the concentration measurement in a manner different from the effect on the fluorescence measurement causing error and loss of precision in the glucose concentration measurement.

Taken together, Figs. 8 and 9 suggest that by measuring the applied pressure during a tissue modulation cycle, it will be possible to automatically compensate for differences in the size and capillary density of different test subjects. The slight inflection in the measured pressure with increasing displacement defines a point where the blood has been removed from the tissue modulated region and simple compression of the remaining static tissue is the only process that remains. By maintaining the applied pressure below this level, that is approximately equal for different test subjects, and corresponds to a different displacement of the tissue modulation aperture with respect to the surface of the stratum corneum, extrusion of tissue through the tissue modulation aperture, and gross movement of the aperture itself can be minimized if not avoided completely. This approach is being implemented and will be tested to see if our hypotheses are correct in the next generation LighTouch device.

Torjman[11] had previously observed that when HemoCue glucose measurements were compared with reference glucose measurements using plasma, there is a slight bias towards HemoCue underestimating the glucose concentration at high glucose concentrations. This observation motivated the comparison associated with Fig. 10 analogous to the comparison made by Torjman and co-workers. Regardless of whether the average of the HemoCue and LighTouch measurements or just the HemoCue measurements are used in the abscissa, an analogous small positive correlation with respect to linear regression is observed.

The cumulative data reported show that the LighTouch measurement system can be used to monitor blood glucose in the normoglycemic range with accuracy and precision commensurate with current FDA approved fingerstick glucose meters. Using totally noninvasive measurements we have reproduced weak dependences that have already been established for blood using conventional invasive techniques. This could only be the case if the tissue modulation process was indeed providing quantitative spectroscopic access to blood and blood glucose. The algorithm employed in this study to obtain glucose concentrations was crude and a more sophisticated approach should produce better results. Given the possibility of spectral interference from other blood analytes and possible imperfections in the tissue modulation process itself, it would seem that such interferences are not very large although we suggest that more research is needed to establish this definitively. Furthermore, a single calibration based on one individual is applicable to random people and is also stable for a period of 3 months. Continued effort towards improving and evaluating the measurement process is clearly justified. Quantitative vibrational spectroscopy of human blood is feasible as a clinical research tool for experimental and practical metabolic monitoring involving various analytes and may also become useful for the self-monitoring of blood glucose by people with diabetes. To support this conclusion we cite the excellent work of Puppels and co-workers[20] who have recently obtained noninvasive blood spectra *in vivo* and noted the potential for glucose sensing using confocal Raman microscopy.

Conclusions

We have established a method for quantitative noninvasive vibrational spectroscopy of human blood *in vivo*. The method produces glucose concentration measurements with accuracy and precision comparable with FDA approved fingerstick devices and therefore can be expected to contain information bearing on various other analytes. A single calibration of the device using data from one individual is applicable to other randomly chosen individuals and is stable for a period of at least $\approx$3 months. We have established sources of random and systematic error in the process. At least one source of systematic error is associated with hemoglobin concentration variation while random error may result from inadequate registration of the fingertip with the tissue modulation aperture at either stage of the modulation process. The results suggest that measuring the applied pressure and providing real-time feedback to the test subject during the tissue modulation process will improve the success rate of the measurement process even without prior knowledge of a test subject's hemoglobin concentration. Although there may be other sources of error and loss of precision, both systematic and random, satisfactory management of the tissue modulation process could result in a device for the clinical and self-monitoring of blood glucose and other analytes.

Acknowledgments

The authors gratefully acknowledge the conscientious efforts of Jessica Grenga and Kelly Duncan in collecting the data presented herein. Dan Godici and John Fayos of Critical Link, LLC provided software development. Deb Dews, Ethan Voss, David Kidd of the Cornell Food Science Laboratory and Danny Chawan of SRIM Enterprises provided invaluable assistance in performing the TA.XT and other experiments. They particularly appreciate the helpful comments of the reviewers.

References

1. American Diabetes Associate, ''Consensus statement: self-monitoring of blood glucose,'' *Diabetes Care* **10**, 95–99 (1987).
2. Omar S. Khalil, ''Spectroscopic and clinical aspects of noninvasive glucose measurements,'' *Clin. Chem.* **45**, 165–177 (1999).
3. Joseph Chaiken, William F. Finney, Karen Peterson, Charles M. Peterson, Paul E. Knudson, Ruth. S. Weinstock, and Paul Lein, ''Noninvasive, *in-vivo*, tissue modulated near infrared vibrational spectroscopic study of mobile and static tissues: blood chemistry,'' *Proc. SPIE* **3907**, 89–97 (2000).
4. Joseph Chaiken, William F. Finney, Xiaoke Yang, Paul E. Knudson, Karen Peterson, Charles M. Peterson, Ruth. S. Weinstock, and Douglas Hagerman, ''Progress in the noninvasive, *in-vivo*, tissue modu-

lated Raman spectroscopy of human blood,'' *Proc. SPIE* **4254**, 216–227 (2001).
5. Joseph Chaiken, William F. Finney, Paul E. Knudson, Karen Peterson, Charles M. Peterson, Xiaoke Yang, and Ruth. S. Weinstock, ''Noninvasive, *in-vivo*, tissue modulated near infrared vibrational spectroscopic study of mobile and static tissues: microcirculation and viscosity effects,'' *Proc. SPIE* **4254**, 106–118 (2001).
6. Joseph Chaiken, William F. Finney, Paul E. Knudson, Karen Peterson, Charles M. Peterson, Xiaoke Yang, and Ruth. S. Weinstock, ''Noninvasive blood analysis by tissue modulated NIR Raman spectroscopy,'' *Proc. SPIE* **4368**, 134–145 (2001).
7. Anthony T. Tu, *Raman Spectroscopy in Biology: Principles and Applications*, Wiley, New York (1982).
8. Andrew J. Berger, Irving Itzkan, and Michael Feld, ''Feasibility of measuring blood glucose concentration by near-infrared Raman spectroscopy,'' *Spectrochim. Acta, Part A* **53**, 287–292 (1997).
9. David M. Haaland and Edward V. Thomas, ''Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of quantitative information,'' *Anal. Chem.* **60**, 1193–1202 (1988).
10. Jason J. Burmeister, Mark Arnold, and Gary W. Small, ''Noninvasive blood glucose measurements by near-infrared transmission spectroscopy across human tongues,'' *Diabetes Technology and Therapeutics* **2**, 5–16 (2000).
11. Marc C. Torjman, Luis Jahn, Jeffrey I. Joseph, Katherine Crothall, and Barry J. Goldstein, ''Accuracy of the HemoCue portable glucose analyzer in a large nonhomogeneous population,'' *Diabetes Technology and Therapeutics* **3**, 591–600 (2001).
12. A. G. Glasmacher, W. Brennemann, C. Hahn, Ch. Marko, S. Wehnert, U. Lohr, and R. Kutschkow, ''Evaluation of five device for self-monitoring of glucose in the normoglycemic range,'' *Exp. Clin. Endocrinol. Diabetes* **106**, 360–364 (1998).
13. Z. Trajnoski, G. A. Brunner, R. J. Gfrerer, P. Wach, and T. R. Pieber, ''Accuracy of home blood glucose meters during hypoglycemia,'' *Diabetes Care* **19**, 1412–1415 (1996).
14. See literature provided by HemoCue, Inc. Lake Forest, CA.
15. Roy Henderson and Karl Schulmeister, *Laser Safety*, Sec. 3.8, Institute of Physics Publishing, Bristol (2004).
16. David P. Shoemaker, *Experiments in Physical Chemistry*, 6th ed., p. 41, WCB McGraw-Hill, Boston (1996).
17. William L. Clarke, Daniel Cox, Linda A. Gonder-Frederick, William Carter, and Stephen L. Pohl, ''Evaluating clinical accuracy of systems for self-monitoring of blood glucose,'' *Diabetes Care* **10**, 622–628 (1987).
18. Lois Jovanovic-Peterson, Charles M. Peterson, J. Dudley, and C. Kilo, ''Identifying sources of error in self-blood glucose monitoring,'' *Diabetes Care* **11**, 791–794 (1988).
19. Rough estimate based on capillary dimension data in *Gray's Anatomy*, Peter Williams, Ed., 38th ed., p. 1465, Churchill Livingston, New York (1999), combined with loop-width data in *Geigy Tables*, C. Lentner, Ed., 8th ed., Vol. 5, p. 545, CIBA-GEIGY, Basel (1990).
20. P. J. Caspers, G. W. Lucasesen, and G. J. Puppels, ''Combined *in vivo* confocal Raman spectroscopy and confocal microscopy of human skin,'' *Biophys. J.* **85**, 572–580 (2003).