Transcutaneous Measurement of Glucose Using Raman spectroscopy

by

Jeankun Oh

B.S., Mechanical Engineering
Pusan National University
(1995)

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

Master of Science in Mechanical Engineering

at the
Massachusetts Institute of Technology
May, 2004
[June 2004]

Author ....
Department of Mechanical Engineering
May 7, 2004

Certified by ....
Michael S. Feld
Professor of Physics, Thesis Supervisor

Read by ....
Peter T. C. So
Associate Professor of Mechanical Engineering, Thesis Reader

Accepted by ....
Ain A. Sonin
Chairman, Department Committee on Graduate Students

Transcutaneous Measurement of Glucose Using Raman Spectroscopy

By

Jeankun Oh

Submitted to the Department of Mechanical Engineering on May 18, 2004 in partial fulfillment of the requirements for the Degree of Master of Science in Mechanical Engineering

Abstract

Transcutaneous glucose measurement would provide the ability to obtain frequent measurements without the pain and risk of infection associated with obtaining a blood sample and eliminates the need for reagents. Because of these potential benefits, a large effort is being made within the academic and industrial research communities to develop alternative rapid and easily manageable analysis methods, including transcutaneous methods. Several optical techniques have been proposed and investigated widely to replace the traditional method. Transcutaneous measurement would be of particular benefit to the millions of people with diabetes who should monitor glucose levels multiple times per day and today depend on the finger pricking devices.

We have used the advantages provided by NIR Raman spectroscopy, i.e. sharp and distinct spectral features combined with a large probe volume, in order to simultaneously quantify multiple analytes (glucose, urea, cholesterol, triglycerides, total protein, albumin, hemoglobin, hematocrit (hct)) in serum and whole blood samples. Based on the successful measurements, we have moved toward transcutaneous measurements of glucose from 16 healthy human subjects as the first target of blood analytes. The average RMSEP of 13.17 mg/dL and r^2 values of 0.79 show promise that the Raman spectroscopy can be developed to achieve clinical accuracy requirements. Furthermore, the fact that glucose features could be recognized in the calibration spectra from measurements of volunteers whose glucose levels changed in a relatively large range, provides evidence that the calibration is based upon glucose.

Thesis Supervisor: Michael S. Feld
Title: Professor of Physics

Acknowledgements

Many people made this research possible. I want to thank my advisor, Michael Feld, for providing guidance and support, all while allowing me to do this study in my own time. Both he and Ramanchandra Dasari have helped me to learn about science and research.

Taewoong Koo introduced me to this project. He gave me good intuition for this research and lessons about how to handle problems on this study. A team of people helped me advance this research. Annika M. K. Enejder helped me in every aspect of this study. Martin Hunter showed the guidance for optical system design. Thomas Scecina helped me with his abundant experience. Wei-Chuan Shih worked with me on this research as he will advance this research further. He will for sure work on this study successfully.

And I thank mechanical engineering department for allowing me to work on this interesting project. I especially thank my academic advisor, Peter So.

I would like to thank NIH and Bayer Diagnostics Corporation for funding this research. Finally, I want to thank my parents and my wife for their love and support.

Contents

List of Figures

List of Tables

1. Introduction

1.1 Background

"Man may be the captain of his fate, but he is also the victim of his blood sugar"

Wilfrid Oakley [Trans. Med. Soc. Lond. 78, 16 (1962)]

Over the last century, human behavior and lifestyle have changed, resulting in a dramatic increase of diabetes over the world. Especially the past two decades have seen an explosive increase in the number of people diagnosed with diabetes as well as obesity. Diabetes itself may not be a serious disease but the real risks lie in its complications such as heart disease and stroke, blindness, kidney disease, nerve disease, amputations, and so on. It is clear that tight control of blood glucose level will lower the risk of complications.

> "The global figure of people with diabetes is set to rise from the current estimate of 150 million to 220 million in 2010, and 300 million in 2025 (Figure 1). Most cases will be of type 2 diabetes, which is strongly associated with a sedentary lifestyle and obesity. This trend of increasing prevalence of diabetes and obesity has already imposed a huge burden on health-care systems and this will continue to increase in the future." [Zimmet *et al.* 2001]

This article contains a huge number of examples which warn of the rapid increase of diabetes over the world, illustrating Figure 1.1 with statistics.

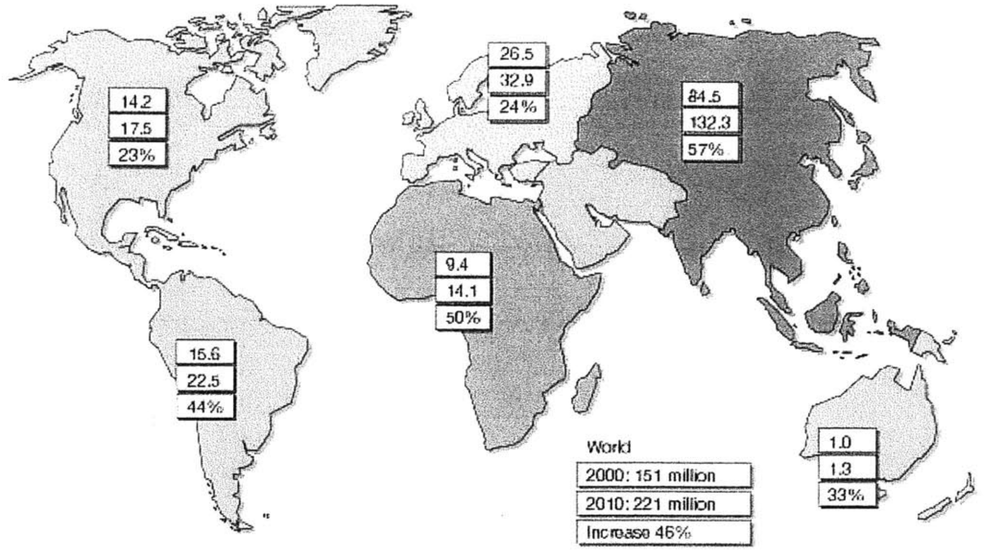


Figure 1.1. Numbers of people with diabetes (in millions) for 2000 and 2010 (top and middle values, respectively), and the percentage increase.

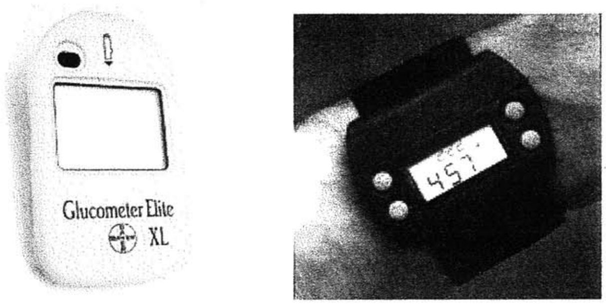
There are two main types of diabetes: Type 1 diabetes (insulin dependent diabetes) and type 2 diabetes (non-insulin dependent). Type 1 diabetes is due primarily to autoimmune-mediated destruction of pancreatic beta-cell islets, resulting in absolute insulin deficiency. People with type 1 diabetes must take exogenous insulin for survival to prevent the development of ketoacidosis. People with type 2 diabetes, which accounts for over 90% of diabetes, are able to produce insulin but are unable to use it properly. At present here is no cure for diabetes so that monitoring and closer control of blood glucose level is a key to prevent development of complications.

People with type 2 diabetes are recommended to check daily blood glucose level by themselves at least once in a day, using a conventional blood glucose checking methods based on a drop of whole blood withdrawn by means of a finger prick device, called a glucose meter (Figure 1.2 (a)). Type 1 may be required 3 or 4 times (before meals and at bedtime) for tight monitoring, since people with type 1 do not have ability to produce insulin at all.

However, to acquire blood glucose values by currently popular glucose meters, people must prick themselves with a lancet. Unfortunately, the discomfort and inconvenience of finger prinking may cause many diabetics to not monitor themselves as regularly as they should, putting themselves at risk for the serious complications. Also the cost of invasive blood glucose tests can be substantial. The glucose meters are not themselves expensive, most under $100. But the real expense lies in the test strips, which are used once and average between $0.5 and $1 apiece. A diabetic can conceivably spend $150 a month on the strips.

Consequently, a new painless (bloodless) method without consuming any disposable parts has attracted the attention of many researchers in both industry and academia. Many techniques are currently under research, including infrared absorption spectroscopy, electrochemical measurement, fluorescence spectroscopy, pulsed laser photoacoustic spectroscopy, NMR spectroscopy, spectral analysis of the chemical components in vitreous humor, and polarmetric techniques [Berkowitz BA *et al.* (1995); Feng GM *et al.* (1997); Quan KM *et al.* (1993); Sharma A *et al.* (1994); Tarr RV *et al.* (1990); Taylor MA *et al.* (1995)]. Most of these techniques, however, have practical difficulties due to the many variables which occur in actual applications, including the presence of other chemical components, change of temperature, and change of optical parameters.

The first noninvasive device to receive FDA (the Food and Drug Administration) approval has come in the market, called GlucoWatch (Cygnus in Redwood, CA) (Figure 1.2 (b)).



(a) (b)

Figure 1.2. (a) one of currently popular glucose monitoring devices, Glucometer Elite (Bayer) (b) a new bloodless technique, GlucoWatch (Cygnus)

GlucoWatch worn on the wrist, like a watch, sends a low-level electrical current through the skin, which attracts glucose and brings it to the surface where it is measured by a membrane on the GlucoWatch. However, it requires, at least twice a day, calibration using the conventional finger-lancing method and replacement with a single use disposable component, the membrane. FDA approved it not as an alternative to traditional self monitoring of blood glucose, but rather to serve as an adjunct, supplying additional information on glucose trends that are not available from self monitoring. In terms of extracting a liquid sample (interstitial fluid) instead of blood from body, this new device may be called 'minimally invasive method'.

Scientists have recognized that one of the truly noninvasive techniques may be based upon an optical device using light instead of lancet or any other small needle, since only light contacts a sample surface in optical measurements. As addressed above, various optical techniques on in vivo measurement of glucose have been studied for many years. Although Raman spectroscopy had been widely used for chemical and molecular analysis for many years, its application to in vivo measurements through human skin is relatively recent. The reason is that Raman clinical applications have been limited not only by the difficulty in collecting the inherently weak Raman signals, which are normally overwhelmed by intense tissue fluorescence, but also by the relatively slow speed of spectral acquisition to increase signal to noise (SNR) ratio. Recently, new advancing technology particularly in the areas of diode lasers and CCD detectors made it possible to enhance SNR dramatically by utilizing better collection optics design. Because Raman spectra can provide a 'fingerprint' of a substance in relatively narrow bands from which the molecular composition can be determined, good quality Raman signals will promise better results in analysis than any other spectroscopic techniques.

1.2 Raman spectroscopy

Raman spectroscopy delivers spectral information about molecular vibrations by measuring the wavelength shift of a photon due to inelastic scattering. When a photon of a monochromatic laser beam of frequency υ_i is directed on a sample, in most cases a photon of the same wavelength υ_i is emitted, which is called elastic Rayleigh scattering. During this interaction there is a low probability of the photon energy causing vibrational excitation or even rarely vibrational relaxation of the molecule and the photon is thus scattered with respectively lower or higher energy. This process is called the inelastic Raman scattering effect and these vibrational transitions are characteristic for each molecule. In inelastic scattering, the emitted photon has a frequency υ_s, which is different from the frequency υ_i of the incident photon shown in Figure 1.3. In Raman scattering, the frequency shift $\Delta\upsilon$ ($=\upsilon_i$ - υ_s) is equal to the vibrational frequency of the molecule.

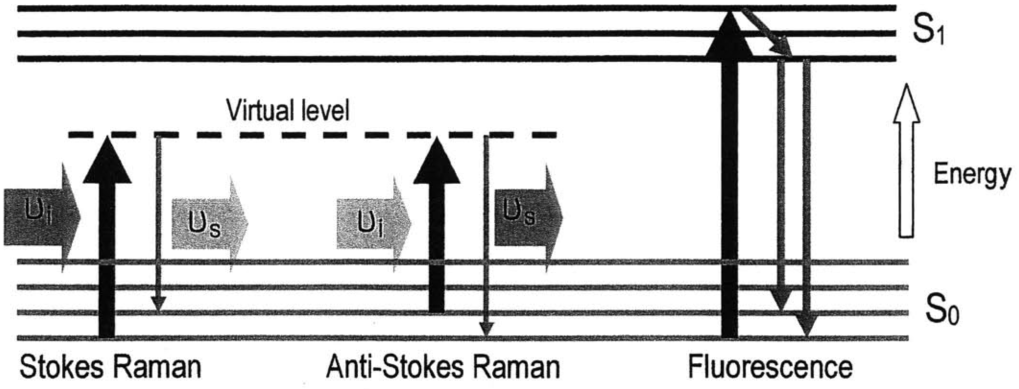


Figure 1.3. Energy level diagram of electronic states (S_0, ground electronic state; S_1, excited electronic state)

Raman spectra can be classified according to whether the emitted photon has less or more energy than the incident photon; they are known as Stokes and anti-Stokes Raman spectra respectively (Figure 1.3). Stokes Raman scattering involves a transition of a photon from the ground vibrational state to an excited vibrational state. On the contrary, in anti-Stokes Raman scattering (Raman shifts to shorter wavelengths), energy is transferred from the vibrating molecule to the Raman scattered light, which means Raman shifts to shorter wavelengths. There two advantages in suppressing fluorescence background signal as well as in utilizing the wavelength region of higher quantum efficiency of CCD (if one uses near infrared as excitation light). But the signal intensity of anti-Stokes Raman scattering is much weaker than that of Stokes Raman scattering, since the population of excited vibrational states is very much lower than the population of the ground state. To compensate the weak Raman signal, an integrating sphere was employed. However, the fact that this can hold a small volume of sample results in being not suitable for in vivo measurement.

The spontaneous Raman spectrum is a linear combination of the spectrum of each chemical component. This is important in the quantitative analysis because the linearity allows us to use linear multivariate calibration techniques. This will be discussed in detail in Chapter 4.

Since Raman shifts are independent of excitation wavelength, one has the freedom to choose an optimal excitation wavelength in the near IR for deep sampling (Section 2.2.1). Raman spectra are sharp and distinct for each molecule, so one chemical component can be distinguished easily from another, and from a broad fluorescence background, even if the Raman signal intensity is smaller than the background.

However, one of the disadvantages of Raman scattering is its weak intensity. This weakness of signal strength has been regarded as a challenge in taking high quality Raman signal of biological tissue. But the introduction of high efficiency CCD detectors, lasers and other new optical technology enables Raman spectroscopy used for a promising clinical tool.

1.3 Prior research using Raman spectroscopy

The measurements of blood analytes including glucose has been investigated for several years in George R. Harrison Spectroscopy Laboratory at MIT. The study was initiated with measuring biological analytes in aqueous samples, mixed solutions of glucose, lactic acid, and creatinine in saline solution, which was performed by lab-made Raman system [Berger *et al.* 1995, 1996]. A year later, higher level of glucose in whole blood were measured to assess the feasibility using Raman spectroscopy. These were measured in the same Raman system [Berger *et al.* 1997]. Based on these two experimental results, serum and blood samples from 69 patients were tested for quantitative measurements of blood analytes including glucose [Berger *et al.* 1999]. Correlation coefficient values (r^2) of 6 blood analytes were over 0.74 or higher; glucose was 0.83 with RMSEP 26mg/dL. However, the predictions in whole blood were not as good, since the effect of absorption mainly by hemoglobin existing in whole blood decreases signal significantly compared with clear liquid sample such as serum. Subsequent redesigning and upgrading the Raman system employing a paraboloidal mirror enabled the successful measurement in whole blood [Enerjder AMK *et al.* 2002].

Table 1.1. Accurate measurement in whole blood samples from multiple patients

Analyte	RMSEP	r^2
Glucose	19(mg/dL)	0.95
Urea	4.9(mg/dL)	0.94
Triglyceride	39(mg/dL)	0.93
Total protein	0.26(g/dL)	0.95
Albumin	0.10(g/dL)	0.99
Hematocrit	1.3(%)	0.96
Hemoglobin	0.41(g/dL)	0.94

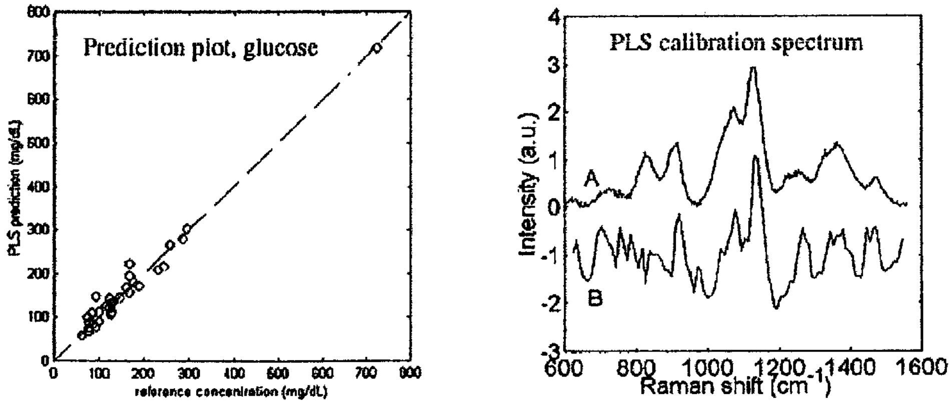


Figure 1.4. Prediction plot (left) for glucose and the corresponding PLS calibration spectrum (right). Good agreement was obtained between the latter (curve B, offset) and the Raman spectrum of pure glucose in water (curve A) [Enerjder AMK *et al.* 2002].

Glucose measurement with the high r2 value, 0.95 (Table 1.1), and good agreement in PLS calibration (Figure 1.4), indicated that the Raman system we developed is sensitive enough to measure glucose in an optically complex medium with strong and varying light-scattering and absorption properties such as human skin.

2. Experimental System

2.1 System scheme

The experimental system in the MIT spectroscopy laboratory has been upgraded from the first generation Raman system measuring blood analytes from serum [Berger AJ *et al.* 1999] to the "blood system" measuring blood analytes from whole blood [Koo T-W 2001]. The blood system was designed to collect light up to 90° by employing a paraboloidal mirror (PerkinElmer) as shown in Figure 2.1. 90° collection angle is a great benefit, comparing typical collection angles up to 30° for single lenses and up to 45° for microscope objectives. The mirror was cut 0.5" inside the optical axis to facilitate placing a sample at the focal point of the mirror. The inside surface of the mirror was coated with gold to maintain high reflectivity for the target wavelength region between 830nm and 1000nm, and 85% reflectivity was obtained.

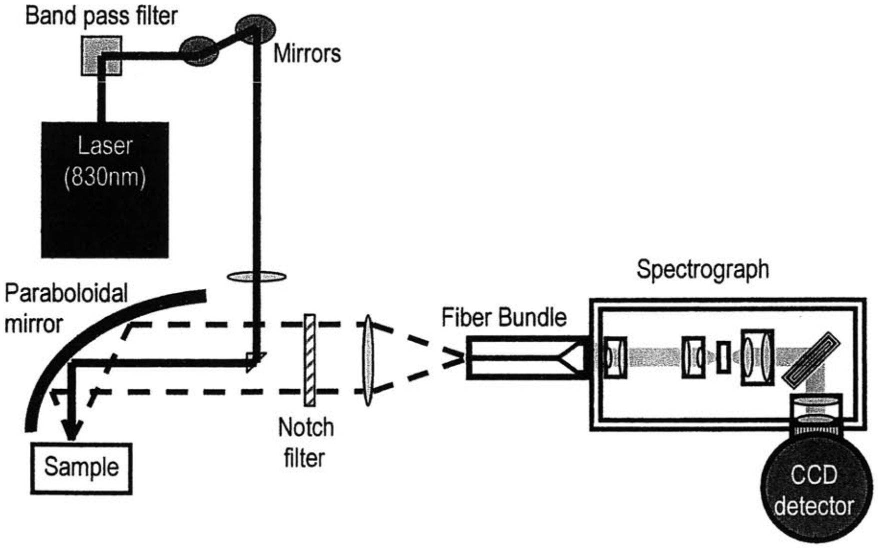


Figure 2.1. Schematic diagram of blood system

Figure 2.1 shows the schematic diagram of blood system by which successful measurements of 7 different blood analytes including glucose from whole blood samples were achieved [Enejder AMK *et al.* 2002].

A multi-mode diode laser emitted the excitation beam at 830nm. A holographic band pass filter removed wavelengths other than 830nm. The filtered laser light through a 100mm focal length lens was focused onto a 1mm x 1mm x 1mm prism which was selected to minimize the shadowing of the collected Raman signals. The beam was again focused by the paraboloidal mirror at 30mm after the focal plane of the mirror. It formed a spot approximately 1mm in diameter with 280mW power at the focal plane of the mirror, where the sample was placed. The sample was placed perpendicular to the beam, and the Raman photons were collimated by the mirror and sent onto the optical fiber bundle through a 2" diameter holographic notch filter (Kaiser optical). On the entrance side, the fibers were grouped in a hexagonal array to maximize the packing fraction (0.6224), and arranged in a line shape on the other side to match the shape of the slit of the spectrograph. The Raman signals were delivered to the spectrograph, and the grating dispersed light as a function of wavelength. A liquid nitrogen cooled CCD detector collected the dispersed photons.

Based on blood system design, we upgraded the CCD detector and modified the spectrograph to a more sensitive system for transcutaneous measurements of glucose from human skin. The following Next sections will describe detailed specifications of important optical parts and system improvements for the "transcutaneous system".

2.2 System specification

Table 2.1 is a summary of system specifications of the main optical components. The blood system was used for measuring blood analytes from whole blood. The transcutaneous system was upgraded from the blood system for the goal of this study, transcutaneous measurement of glucose through human skin in this study.

Table 2.1. Summary of system specifications

<table>
<tr><th>Component</th><th>Specification</th><th>blood system</th><th>transcutaneous system</th></tr>
<tr><td rowspan="3">Laser</td><td>Wavelength</td><td colspan="2">830 nm</td></tr>
<tr><td>power output</td><td colspan="2">480 mW</td></tr>
<tr><td>power on sample</td><td>280 mW</td><td>330 mW</td></tr>
<tr><td rowspan="4">Fiber bundle</td><td>diameter</td><td colspan="2">330 μm (including cladding, core=300 μm)</td></tr>
<tr><td>Size</td><td colspan="2">Circular end diameter = 3.0 mm
Linear end height = 20.1 mm</td></tr>
<tr><td>f/#</td><td colspan="2">1.8</td></tr>
<tr><td>number of fibers</td><td colspan="2">61</td></tr>
<tr><td rowspan="2">Spectrograph</td><td>f/#</td><td>1.8</td><td>1.4</td></tr>
<tr><td>Dispersion</td><td colspan="2">16.5 m/cm^{-1}</td></tr>
<tr><td rowspan="3">CCD detector</td><td>CCD area</td><td>25 x 17 = 425 mm^2</td><td>26 x 26.8 = 697 mm^2</td></tr>
<tr><td>pixel size</td><td>22 μm x 22 μm</td><td>20 μm x 20 μm</td></tr>
<tr><td>quantum efficiency</td><td>30% @ 950 nm</td><td>60% @ 950 nm</td></tr>
</table>

2.2.1 Excitation wavelength selection

One of most important considerations for Raman spectroscopy experiments using biological samples is using the optimal excitation wavelength. We determined 830 nm NIR light as an excitation source based on 3 reasons.

One, the shot noise from a high fluorescence background signal is the primary determinant of signal to noise ratio (SNR) in this system. Raman signals are significantly weaker than background fluorescence. Thus, a high level of noise from fluorescence signal will produce a SNR in the Raman signal that can prohibit accurate detection of blood analytes of interest. Near infrared (NIR) reduces fluorescence because the lowest lying excited electronic states of most biological molecules responsible for fluorescence correspond to visible wavelengths, and therefore cannot be excited by lower-energy NIR light.

Two, NIR light penetrates relatively deeply in human tissue (1~2 mm). The penetration depth of a particular wavelength is mainly determined by absorption properties of the human body. NIR includes a specific wavelength range called the diagnostic "window", in which absorption of melanin, water and hemoglobin is smaller than in other regions as shown in Figure 2.2.

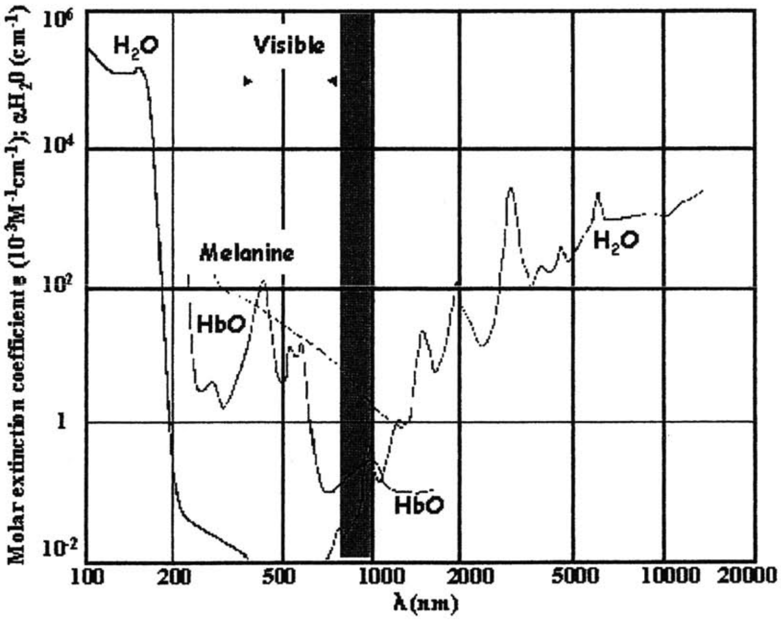


Figure 2.2. Absorption spectrum of major absorbers in human body; the grey box is the window where absorption of melanin, water and hemoglobin is relatively small.

Three, the quantum efficiency of a silicon-based detector drops at longer wavelengths. The CCD detector in our system cut off at about 1100 nm. This will be discussed at the next section.

For above reasons, we selected 830nm as our excitation wavelength. This NIR wavelength does not generate as much fluorescence background as any shorter excitation wavelength does. And also this takes advantage of the diagnostic window, so that the light can penetrate relatively deeply.

2.2.2 CCD

The blood system had the CCD detector with an active area of 25 mm x 17 mm height. The horizontal dimension of the CCD using 830 nm excitation light allows us to observe Raman signals ranging from 830 nm to 1000 nm, in which almost all unique Raman peaks of blood analytes, including glucose, appear. For this reason, the new CCD detector for transcutaneous system has almost the same width as that of the CCD in the blood system. However the new CCD is larger in its vertical dimension in order to accept the full height of fiber bundle. Figure 2.3 shows the 20mm-height fiber bundle image illuminated by a neon lamp on the new CCD detector, comparing the smaller size of previous one. This image was taken after switching the CCD, maintaining the f/1.8 spectrograph attached to the CCD. The slit frame and the f/1.8 lens frame still blocked the light from both end-sides fibers (discussed in the next section).

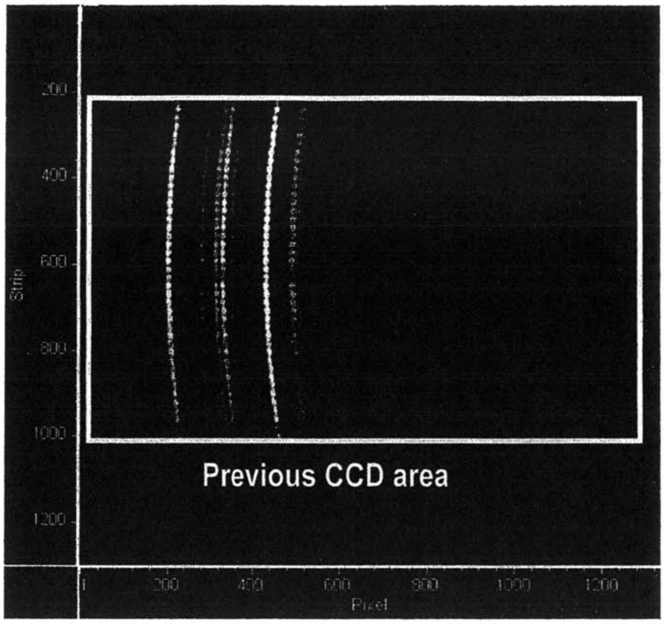


Figure 2.3 Comparison of CCD areas from both systems

Instead of the front-illuminated type of the previous CCD, the new CCD is a back-illuminated type, which has a higher quantum efficiency. As shown in Figure 2.4, the back-illuminated type increases the quantum efficiency by a factor of 2 in the wavelength range of 830 nm to 1000 nm. This doubled quantum efficiency as well as larger CCD area is the main gain in upgrading to the transcutaneous system.

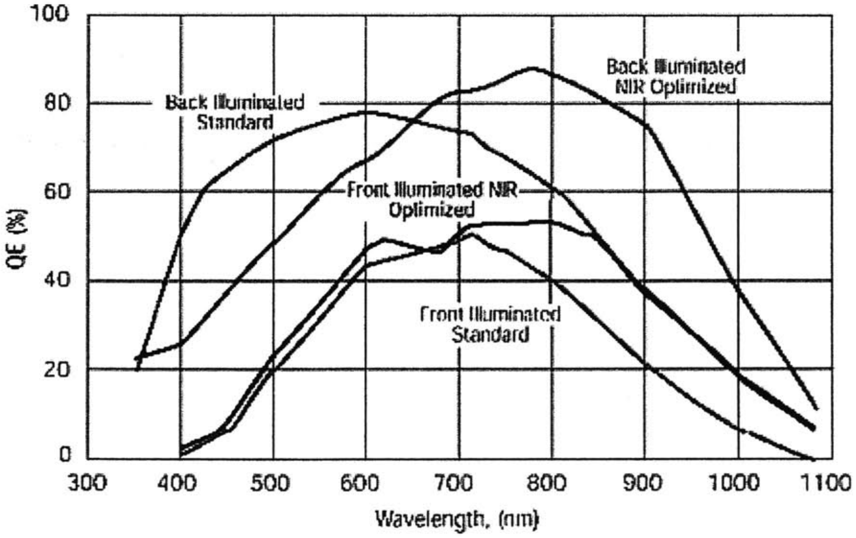


Figure 2.4. Comparison of quantum efficiency by various types of CCD [Roper Scientific].

2.2.3 Spectrograph

As described earlier in Table 2.1, the blood system initially included the f/1.8 spectrograph (Kaiser Optical) in which 4 compound lenses, two f/1.8 lenses at the entrance side and two f/1.4 lenses at the output side, were placed. The original role of the f/1.8 lenses was to collimate the light so that the holographic notch filter can work properly. We figured out those were not necessary so that the fiber bundle could sit inside of the spectrograph (Figure 2.5). Since the paraboloidal mirror produces collimated light, we moved the notch filter outside of the spectrograph and thus could remove those two lenses, further increasing throughput.

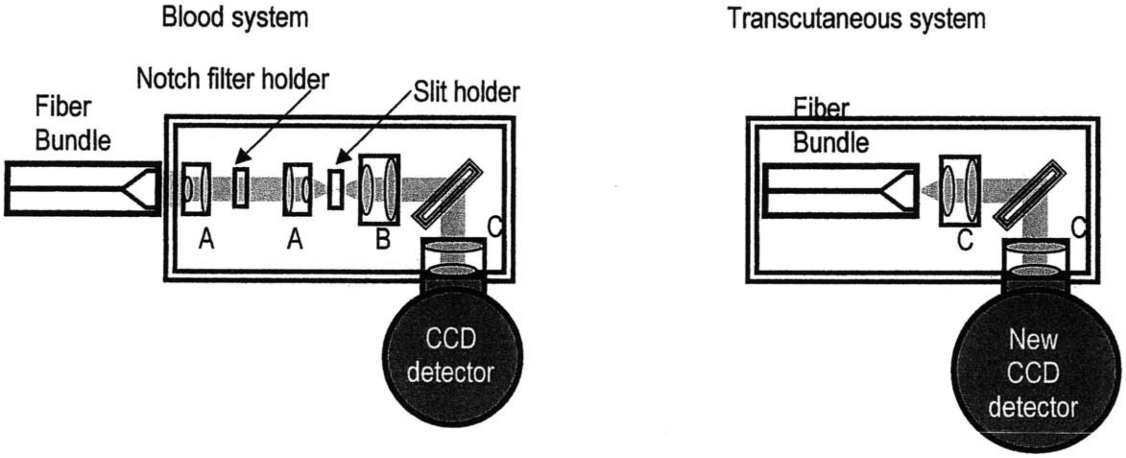


Figure 2.5. f/1.8 and f/1.4 spectrograph. (A; f/1.8, f=50 mm, B; f/1.4, f=75 mm, C; f/1.4, f=85 mm)

As discussed in the previous section 2.2.2, after switching the CCD detector to the new one with taller height, we could see fiber arrays imaged on CCD along the wavenumber axis. However we could not still see all 61 fibers due to vignetting by two f/1.8 lenses and the slit holder. We then changed lens B to another lens C and removed the slit holder. This change allowed us to observe all fibers imaged on the new CCD detector (26mm x 26.8mm), as shown in Figure 2.6.

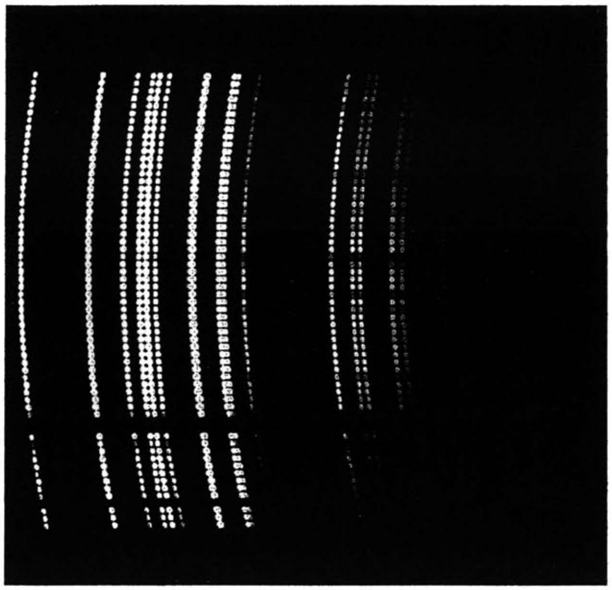


Figure 2.6. Fiber array imaged by a neon lamp on 26mm x 26.8mm CCD detector.

In Figure 2.6, the horizontal axis is pixel or wavenumber and the vertical axis is the height of CCD. We picked one (curved) vertical array at the center wavenumber region and added all of the pixel intensity over each fiber. Figure 2.7 is a normalized plot of intensity values of all 58 fibers (3 of the 61 fibers were broken) along the fiber array that we selected.

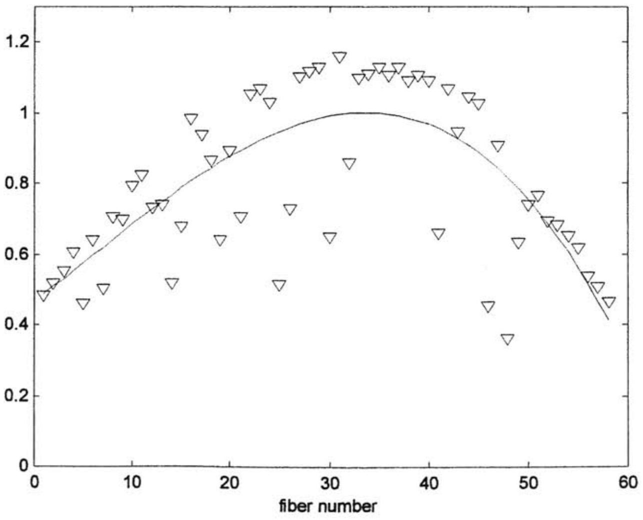


Figure 2.7. Intensity distribution of all 58 fibers.

As shown in Figure 2.7, light captured by the CCD from the end fibers was approximately 50% of the light from the center fibers because of vignetting. In order to understand the causes of vignetting in the spectrograph, we experimentally studied the light propagation path from the center and end fibers through the spectrograph using images on cardboard screens, and the effects of masks over the lenses and grating. A HeNe laser was used as a strong visible light source. A digital camera was used to take pictures of light propagation at various points in the spectrograph.

Several examples of the testing we have done are presented here. Figure 2.8 shows a top view of the spectrograph with two f/1.4 lens assemblies. Also shown are two cardboard screens mounted right after lens A and after the grating. These were used to visually observe the shapes and sizes of light beams.

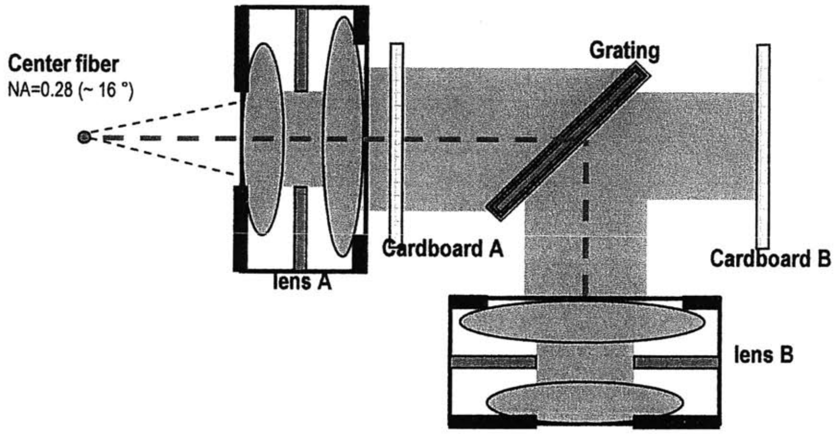


Figure 2.8. Top view of spectrograph showing two masks from the single center fiber on them

With the spectrograph illuminated with light from only the center fiber, an image with a circular shape and a diameter of 50mm was observed on the round screen placed right after lens A. It represents a full size image without any vignetting by lens A, because the f/1.8 fiber underfilled the f/1.4 lens. Also shown in Figure 2.9 is the light that propagates through the grating and onto a rectangular screen after the grating. Here the image has been reduced to 90% of the original area. The grating frame cuts off 5mm each on the right and left side, resulting in 10% loss. This loss affects the light from all fibers equally.

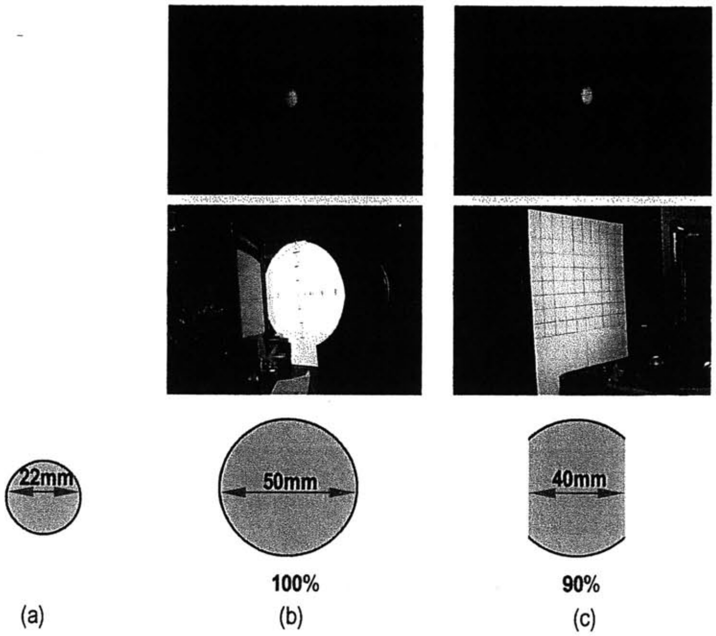


Figure 2.9. Images from the single center fiber
(a) right before, (b) right after lens A and (c) right after the grating

As compared with the same amount of vignetting loss in the light from all fibers by the grating frame, Figure 2.10 indicates different levels of light losses depending on the vertical positions of each fiber. Light from the bottom-most fiber, which is positioned 10mm from the center, experiences loss at a number of interfaces. It is clipped by the entrance aperture of lens A due to the low position of the fiber and the fiber NA, 0.28. It also is reduced somewhat by an internal aperture in lens A. Calculations based upon tracing the rays out of the fiber and the size of the image on the cardboard mask placed at the lens A output side both indicate that less than 79% of the light passes this lens. The light is dispersed by the grating and propagates through lens B (which is identical to lens A), where an identical loss occurs in the top portion of the light beam.

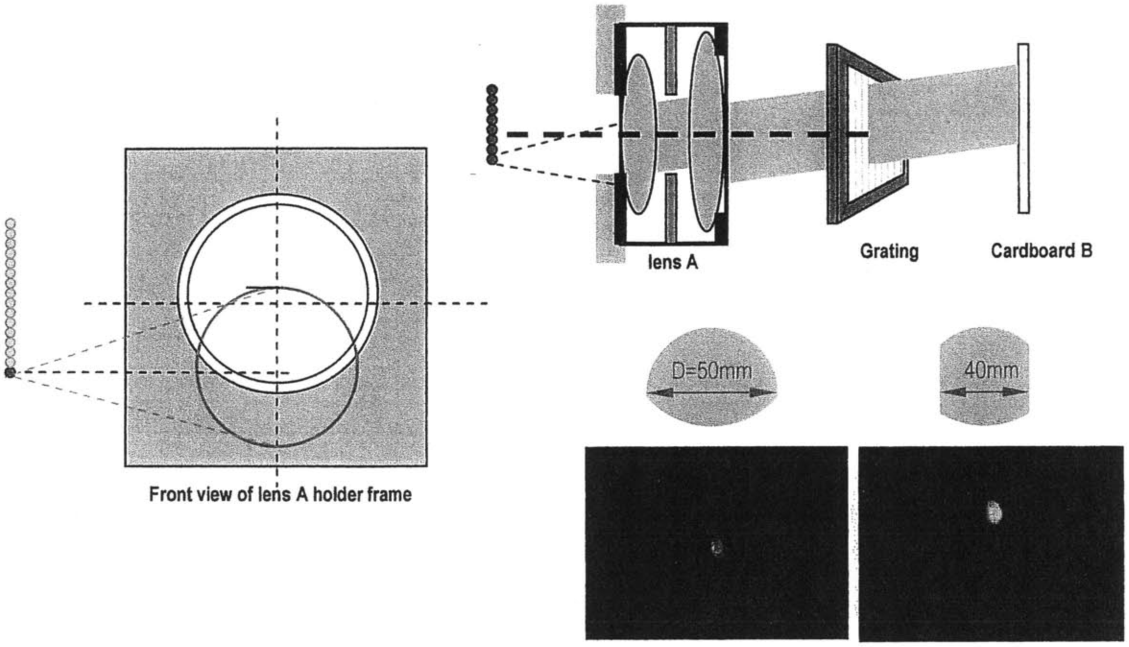


Figure 2.10. Side view and front view of spectrograph showing images from the bottom fiber on two masks after lens A and grating

As shown by these examples, both end fibers experience a considerable amount of vignetting by the two f/1.4 lenses. In the same way we used screens and masks at various parts of the spectrograph to determine causes of vignetting and the characteristics of the lenses. From this study we have learned that these two lenses operate as f/1.4 lenses only for light emanating from the center fiber. As the source of light changes to the fibers away from the center, the lenses operate at a higher f number, reaching an effective f/2 at the end fibers.

Numerical simulation modeling of vignetting was performed in collaboration with Dr. Martin Hunter using an optical ray tracing software, ZEMAX (ZEMAX Development Corporation, CA). The model by ZEMAX simplifies the two f/1.4 camera lenses in the spectrograph by using a pair of f=79mm commercially-available aspheric lenses (Melles Griot #LAG025), apertured to f/1.4. This is because the ZEMAX software database does not have this camera lens. Notice that a double set of apertures was utilized to better model the spectrograph camera lenses; this detail is important in order to better account for vignetting effects experienced by light emanating from the various vertical fiber positions at the linear fiber bundle end. A central rectangular aperture was also included to account for vignetting at the transmission grating.

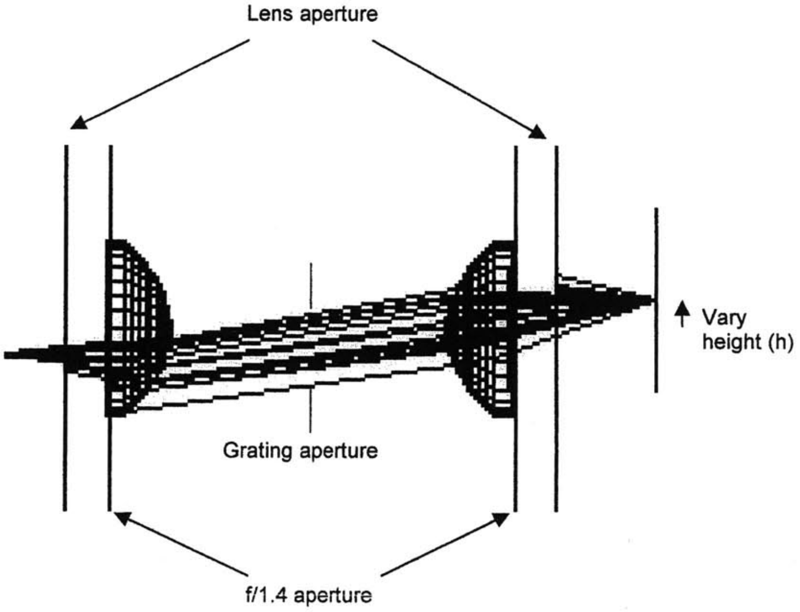


Figure 2.11. ZEMAX model for f/1.4 Kaiser Optical Systems, Inc. transmission grating and variable height fiber optic source

The model was used to measure the vignetting losses in our spectrograph as a function of fiber optic vertical distance from optical axis (h) (Figure 2.11). The results obtained are in good agreement with the experimental vignetting measurements performed with our f/1.8 fiber bundle. And vignetting losses on the order of about 50%

are expected for the outermost fibers in the 20mm-length linear fiber bundle as shown in Figure 2.12.

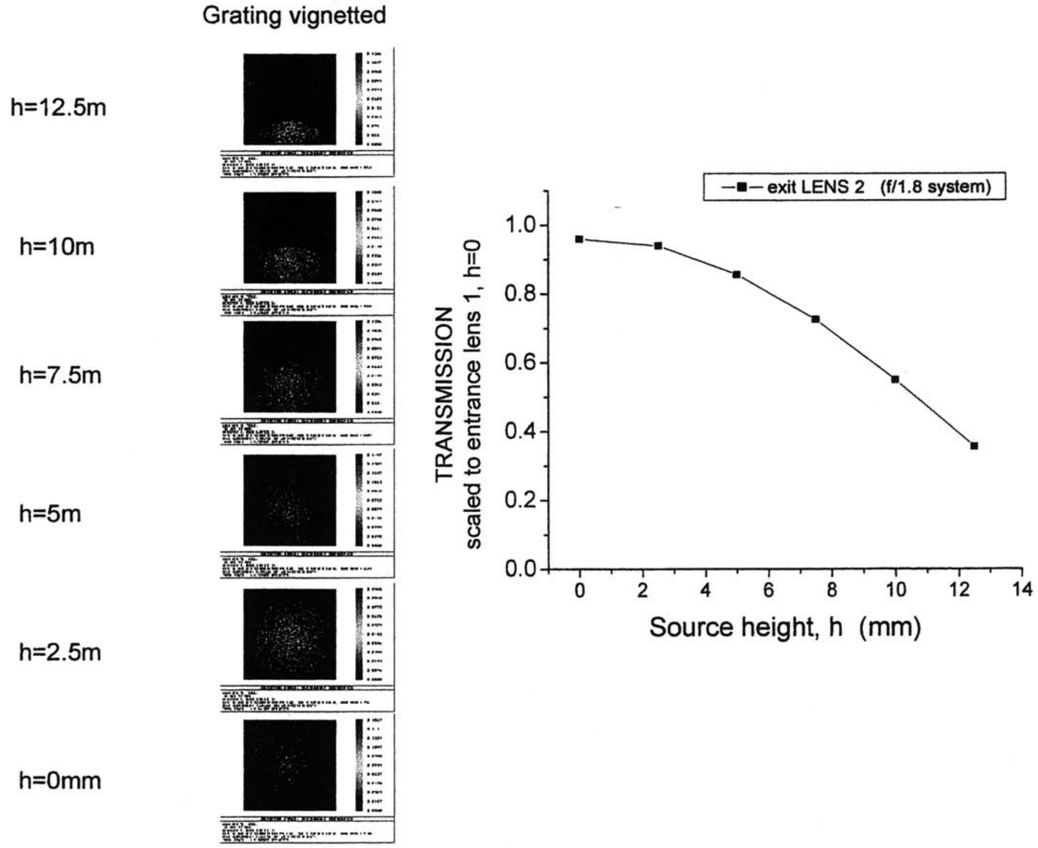


Figure 2.12. ZEMAX prediction of spectrograph vignetting losses for f/1.8 fiber bundle system Left panel: Light intensity profiles exiting 2nd lens; Right panel: Light transmittance as a function of fiber optic off-axis height, h.

2.3 System improvements

As described earlier in Table 2.1, the spectrograph and the CCD detector were changed, which accounted for the majority of system light collection efficiency. Our Raman spectroscopy systems (blood and transcutaneous systems) have one fiber bundle with 61 fibers (3 broken) as shown in Table 1. In the blood system, the f/1.8 spectrometer and 17mm height CCD was not able to make use of the full array of 61 fibers (Figure 2.3).

The upgraded spectrometer (from f/1.8 to f/1.4) and larger size of new CCD showed all fibers imaged by neon lamp as shown in Figure 2.6.

Assuming uniform illumination of the sample by the excitation laser, the scattered signal emerging from the sample is proportional to its collection solid angle Ω, and the sample area A. The product of Ω and A is often referred to as throughput or etendue. The throughput of an optical system is determined by the optical element with the lowest throughput. In the blood system, the f/1.8 spectrometer and 17 mm height CCD limited the throughput, vignetting some fibers in both sides of fiber array (Figure 2.3).

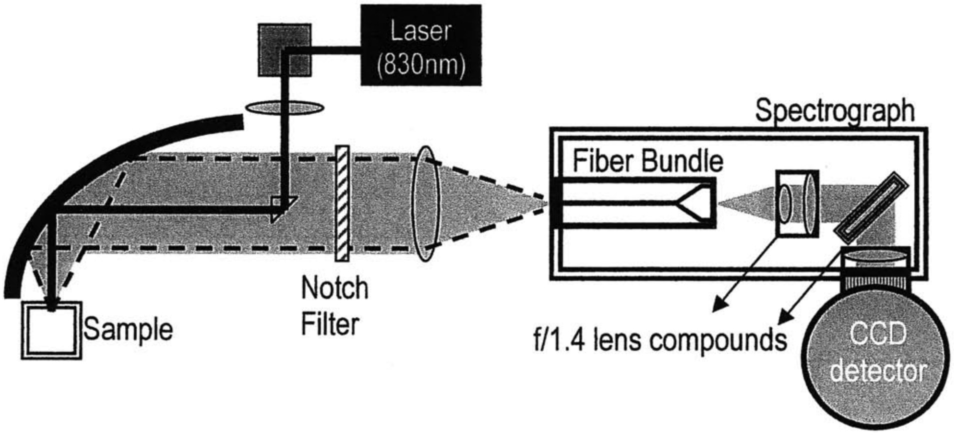


Figure 2.13. Schematic diagram of the transcutaneous system

As discussed in the previous section 2.2.3, the f/1.8 spectrograph in the blood system was upgraded to an f/1.4 spectrograph by removing two f/1.8 lens compounds (Figure 2.13). Because each compound lens had a 10% transmission loss, about 20% of total signal was increased as a result.

A new CCD detector with 1.64 times larger area and doubled quantum efficiency was attached, increasing the maximum signal level by 3 times that of the blood system. Two mirrors mounted after the holographic band pass filter were removed in order to shorten the laser light path, resulting in 18% more power on the sample location. Table 2.2 shows that the transcutaneous system is expected to collect 4.7 times more signal than the blood system with the majority of the improvement resulting from adopting the new CCD detector.

Table 2.2. Summary of system improvements

	Blood System	Transcutaneous System
New CCD detector (area x QE) CCD type Area of CCD	x 1 Front illuminated NIR optimized 25 x 17 = 425 mm2 (0.022mm pixel size)	x 3.3 (=1.64 x 2) Back illuminated NIR optimized 26 x 26.8 = 697 mm2 (0.02mm pixel size)
Reduce the number of optical elements from spectrograph	x 1	x 1.2
Improved alignment and shorten path-length (power increase)	x 1 (280mW)	x 1.2 (330mW)
Total	x 1	x 4.7

Two different types of samples, translucent liquid (methanol) and turbid liquid (whole blood), were tested to evaluate the improvements quantitatively. Under the same experimental conditions, the measurements in both systems were performed and compared in terms of signal intensity changes. Methanol has a strong Raman peak at 1034cm^{-1} and negligibly low background signal. But the whole blood signal consists of the Raman signal and the background fluorescence signal that is typically larger than the Raman signal. We could observe increases in the Raman signal from methanol by a factor of 4. Using whole blood, the Raman signal as well as the background fluorescence increased by a factor of 4; Figure 2.14 compares the total signal changes in whole blood collected from both systems.

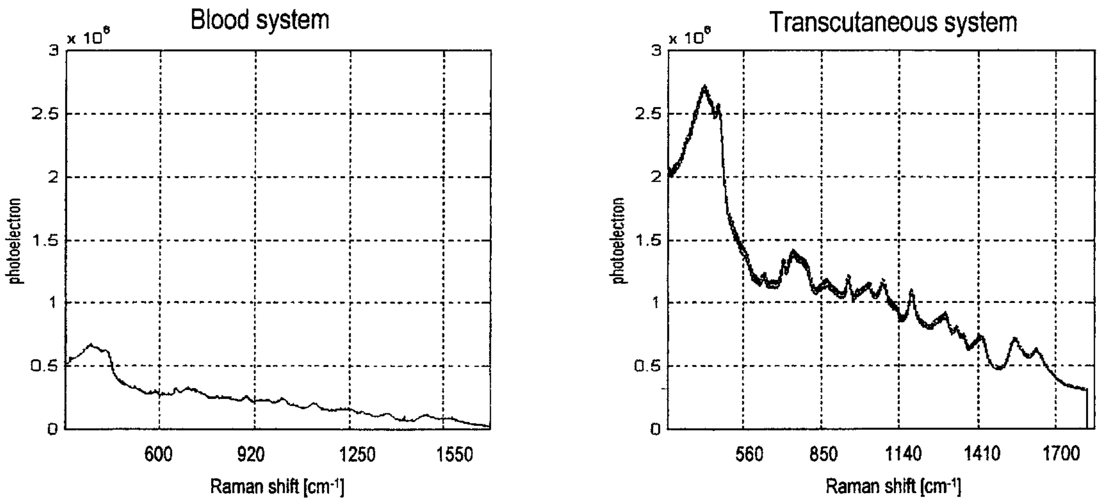


Figure 2.14. Comparison of whole blood spectra measured from blood and transcutaneous systems

The typical fluorescence background, which generated the dominant source of noise, was much stronger than any Raman signal of skin components. We found the SNR of our system is almost equivalent to that resulting from shot noise, which will be discussed in the next section. Accordingly, we could expect the SNR to be improved approximately by a factor of 2.

2.4 Signal to noise ratio (SNR)

The current setting of our transcutaneous raman system on the CCD is high gain, approximately 4 [e^-/count], which means 4 photoelectrons are generated for each photon incident on the CCD corresponding to 1 CCD count in a measurement. We verified the gain of our CCD experimentally by the fact that in a shot noise limited measurement the standard deviation of a series of measurements in photoelectrons, the rms value will equal the square root of the average signal level in the number of photoelectrons. All signal and noise intensities shown in this section are based on photoelectron counts, by multiplying the spectra we collected in CCD counts by the factor of 4.

Out of several different ways to define signal, we will consider the definition of signal (S) which is the average number of photoelectrons at a given wavelength collected over many measurements. And then we can define the total associated noise at the same wavelength as the standard deviation over the series of measurements used to determine the signal. There are 3 different types of noise in our spectral measurements, read noise, dark current noise and shot noise.

Dark noise (σ_{dark}) arises from the statistical variation of thermally generated electrons within the silicon layers comprising the CCD. Dark current describes the rate of generation of thermal electrons at a given CCD temperature. Dark noise, which also follows a Poisson relationship, is the square root of the number of thermal electrons generated within a given exposure. The dark current of our transcutaneous system CCD is <0.01 e-/pixel/sec @-100°C.

Read noise (σ_{read}) refers to the uncertainty introduced during the process of quantifying the electronic signal on the CCD. The major component of read noise arises from the on-chip preamplifier. This depends on the detector temperature but not on the collection time. The specification for read noise in our CCD is 3 e- rms @ 20 kHz.

Shot noise (σ_{shot}) or photoelectron noise, refers to the inherent natural variation of the incident photon flux. Photoelectrons collected by a CCD exhibit a Poisson distribution and have a square root relationship between signal and noise. Therefore $\sigma_{shot}=\sqrt{S}$, where S is the total number of photoelectrons collected. This is the fundamental limiting noise factor in designing an optical system such ours.

When calculating the overall signal-to-noise ratio, all noise sources, the total noise (σ_{total}) and all spectral background need to be taken into consideration. Therefore the final expression for SNR is

$$SNR = \frac{S}{\sigma_{total}} = \frac{S}{\sqrt{\sigma_{shot}^2 + \sigma_{read}^2 + \sigma_{dark}^2}} \tag{2.1}$$

In general, a biological sample such as human skin exhibits a large amount of fluorescent background, compared with Raman signal. So the major source of noise comes from the shot noise. Thus the SNR equation can be simplified as;

$$SNR \approx \frac{S}{\sqrt{\sigma_{shot}^2}} = \sqrt{S} \text{ [no unit]} \tag{2.2}$$

Based on this equation, SNR can be improved under given experimental conditions simply by increasing collection time. Let S be the signal measured in a unit time interval, for instance, 1 second. If the total collection time increase to m seconds, the signal S over m seconds will be m*S. It is assumed that this is a shot noise limited case which means that due to a strong signal shot noise is the main source of noise, ignoring the two other minor sources of noise, dark and read noise. Since shot noise from each measurement is independent, shot noise from m second measurement adds as the square root of m. Then the total noise over m seconds (σ_{total_m}) is

$$\sigma_{total\_m} = \sqrt{m} \cdot \sigma_{shot} \tag{2.3}$$

where σ_{shot} is the shot noise over m seconds. Combining both equations, the SNR over the total accumulation time, m seconds (SNR_m) is

$$S / N_m = \frac{m \cdot S}{\sqrt{m} \cdot \sigma_{shot}} = \sqrt{m} \cdot SNR \tag{2.4}$$

So the SNR increases with the square root of the total collection time m, assuming that our system is shot noise limited, which will be discussed in the end of this section.

To calculate the actual SNR of our system, data from an experiment on the 9th volunteer, one of 20 human volunteers in a transcutaneous study (discussed later in chapter 5) was selected. Because all spectra taken by the system and read from the computer were in units of CCD counts, they needed to be converted to photoelectron number. Each spectrum was taken for 3 min exposure time with 90 2-second shots and averaged by 90. We need to clarify the terms which will be used for the following discussion. 20 spectra of human skin will be called 20 samples and each 2 second measurement will be called a frame. Then we can describe our spectral data as 20 samples averaged by 90 frames. The photoelectron counts from all 20 samples are shown in Figure 2.15.

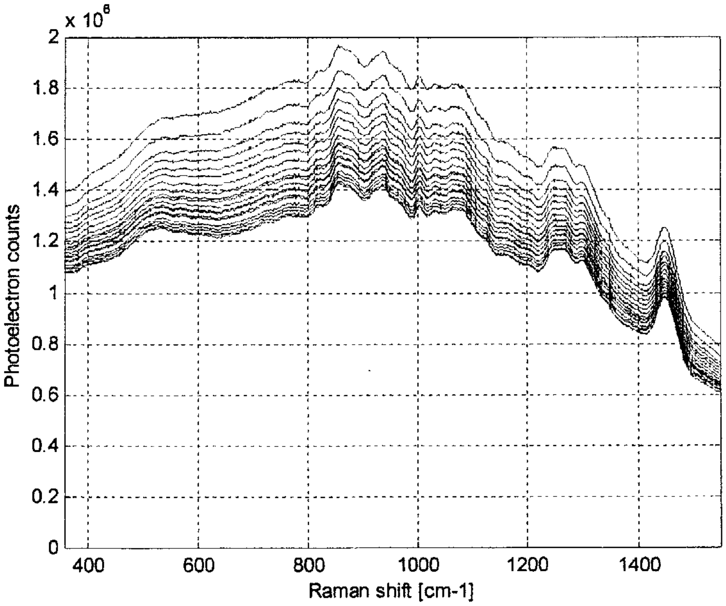


Figure 2.15. All 20 samples of the 9th human volunteer.

The noise for computing SNR is defined as the standard deviation of these 20 samples along all pixel points corresponding to wavenumbers. Figure 2.16 displays the standard deviation of 20 samples along various pixels.

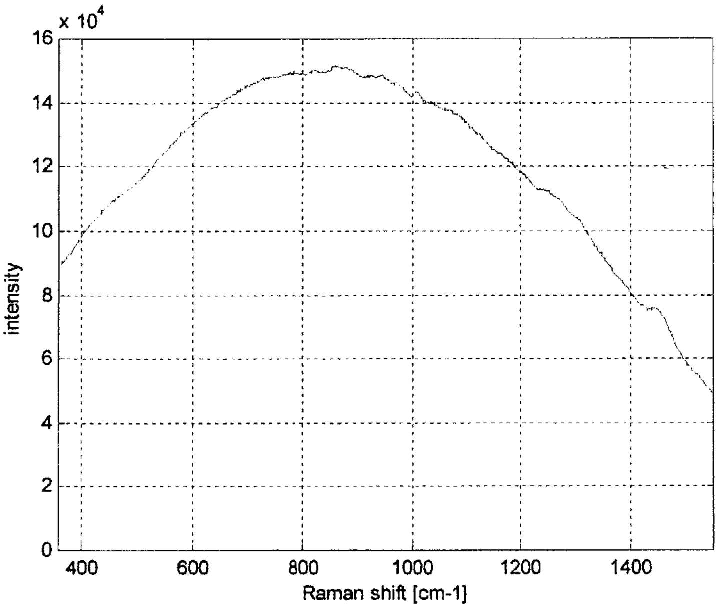


Figure 2.16. The standard deviation of 20 samples.

The spectral shape in the standard deviation plot and the large level of the standard deviation are caused by the fact that the spectra changed not only because of noise sources but also of significant effect of fluorescence photobleaching. Biological samples usually have the dominant signal of fluorescence even with excitation wavelengths as long as 830 nm. Total measurement time for 20 samples of about 2 hours leads to a considerable amount of signal decrease, resulting in the high level of standard deviation values. To eliminate the effect of photobleaching and evaluate SNR due to random noise, the 10th sample was selected as a sample which has approximately an average signal level, as shown in Figure 2.17.

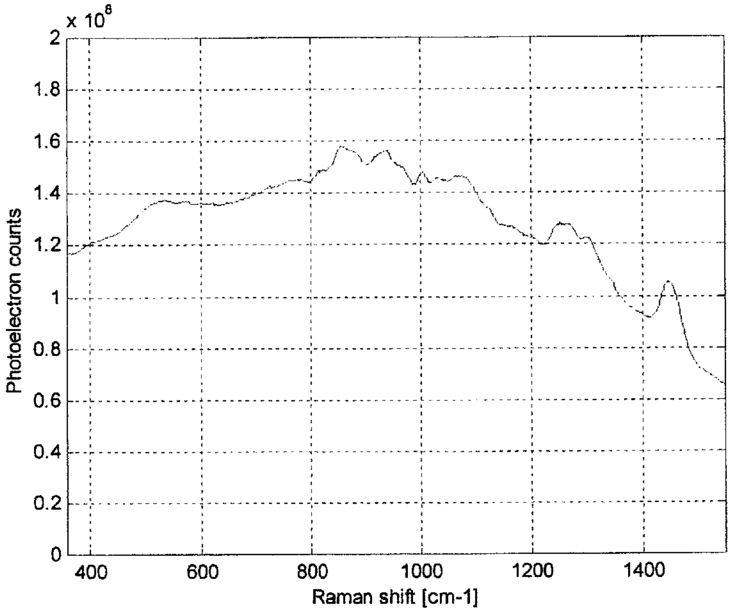


Figure 2.17. The 10th spectra of the 9th human volunteer.

Because of the much higher noise level calculated by the standard deviation of 20 samples (3-minute-spectra) as shown in Figure 2.15, we estimated the noise in the sample measurement by calculating the noise over 90 frames (90 2-second-spectra of the 10th sample) based on Eq. (2.3). It is true that there is still a certain amount of fluorescence bleaching with 90 frames measured for 3 minutes, but it is relatively small. And also, in order to reduce the effect of fluorescence background change by photobleaching, the noise is calculated after taking the overall signal to a (slowly varying) 5th order polynomial and then subtracting this from the overall signal to obtain the Raman contribution. Because the 5th order polynomial is smoothly subtraction from the original spectrum does not quantitatively affect the noise. Figure 2.18 displays the standard deviation plot along all pixel points. The standard deviation at the 597th pixel of 90 2-second-spectra, $\sigma_{2\sec}$, is taken as the noise for 2-second spectra' set.

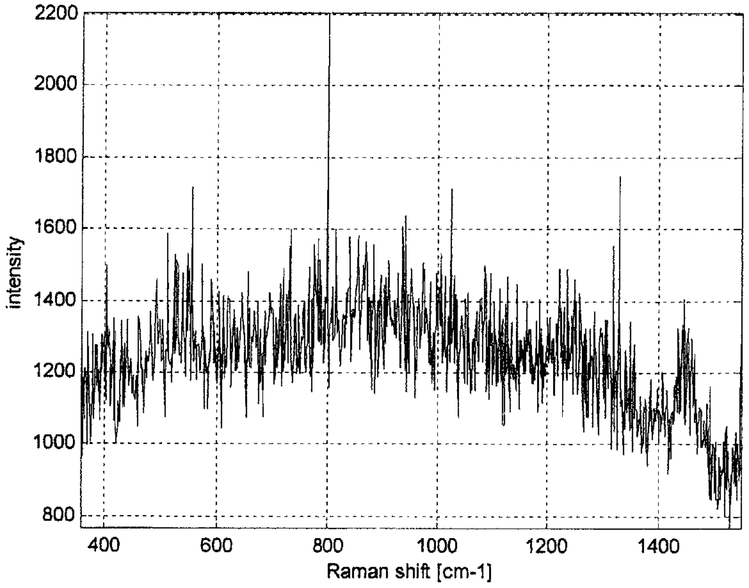


Figure 2.18. The standard deviation of 90 2-second-spectra of the 10th sample after the 5th order polynomial subtraction.

As shown in Figure 2.17, the average signal of 90 frames, $S_{2\text{sec}}$, is read from the intensity at pixel 597, one of highest intensity pixel positions near pixel 600, which is 1,507,850 in photoelectron number. Then the signal and the noise of the 2-second measurements ($S_{2\text{sec}}$ and $\sigma_{2\text{sec}}$) can be related to the signal and noise of the samples ($S_{3\text{min}}$ and $\sigma_{3\text{min}}$) by the following equations;

$$S_{3\min} = m \cdot S_{2\sec} \tag{2.5}$$

$$\sigma_{3\min} = \sqrt{m} \cdot \sigma_{2\sec} \tag{2.6}$$

Combining Eq. (2.5) and (2.6), we can calculate the $SNR_{3\text{min}}$.

$$SNR_{3\min} = \frac{S_{3\min}}{\sigma_{3\min}} = \frac{m \cdot S_{2\sec}}{\sqrt{m} \cdot \sigma_{2\sec}} = \frac{S_{2\sec}}{(\sigma_{2\sec} / \sqrt{m})} \tag{2.7}$$

Applying the calculated values of $S_{2\text{sec}} = 1507850$, $\sigma_{2\sec} = 1281$ and $m = 90$, the $SNR_{3\text{min}}$ by the Eq. (2.7) yields,

$$SNR_{3\min} = \frac{S_{2\sec}}{(\sigma_{2\sec}/\sqrt{m})} = \frac{1.5078\times10^{9}}{(1281/\sqrt{90})} = 11169 \tag{2.8}$$

Since it is assumed that the SNR of our system is limited by shot noise, we can verify the result of Eq. (2.8) using Eq. (2.2) with original 3 min spectra. Let the signal of 3 minutes, $S_{3\min}$, be the total signal of the 10th sample at the same pixel 597.

$$S_{3\min} = \sum_{i=1}^{m} S_{2\sec\_i} = 1.35\times10^{8}[e^{-}] \tag{2.9}$$

Then the $SNR_{3\min}$ can be calculated by Eq. (2.2).

$$S/N_{3\min} \approx \frac{S_{3\min}}{\sqrt{\sigma^{2}_{shot\_3\min}}} = \sqrt{S_{3\min}} = \sqrt{1.35\times10^{8}} = 11619 \tag{2.10}$$

As a result, this verifies that the method to estimate the $SNR_{3\min}$ by Eq. (2.8) is almost equivalent to the calculation by Eq. (2.10), which is the case for a shot noise limited system.

2.5 System Stability

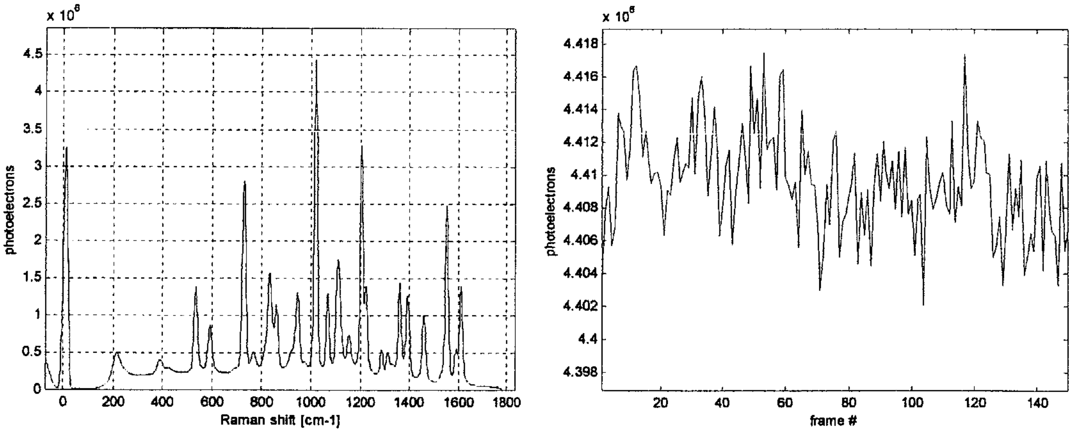
System stability is a critical characteristic when performing many experiments using an optical system. We have used measurements of Indene over 6 days to evaluate the stability of our system. We then developed new ways to take system drift into account and correct for it.

2.5.1 Experiments

Knowledge of system stability is necessary in order to understand the quality of experimental data. We used Indene in a quartz cuvette for this experiment because it is stable enough to be used for wavenumber calibration, its Raman peaks are sharp enough to determine each peak position easily, and its peaks are spread over the wavelength range that we are using. Although Indene is subject to photobleaching, illumination time of less than one minute per one set of measurement is short enough to avoid photobleaching.

We established an experimental plan to determine how stable the system is in terms of intensity and wavelength drifts. Each frame was measured for 0.4 seconds because any longer exposure time saturated the CCD. Each sample had a one minute measurement with 150 0.4-second frames to track short-term drift and to get better SNR sample. We collected 96 one minute samples over 6 consecutive days to monitor long-term drift. For these 6 days, the sample position was left unchanged, with only the laser turned on and off each day.

2.5.2 Intensity and wavelength drifts



(a) (b)

Figure 2.19. (a) Indene spectra and (b) intensity drift of the 49th sample.

Figure 2.19 presents two plots of (a) 150 0.4-second Indene spectra of the 49th sample and (b) their highest peak intensity drift. Based on the SNR calculation in section

2.4, a 1 minute measurement demonstrated that the system is stable enough to show that the intensity drift was mainly caused by photon shot noise. In addition, there was no observable wavelength drift in this short time period. But over a long term period, we measure one pixel drift (approximately 1.4 wavenumber) in wavelength over 6 day.

The stability test was studied in two parts, intensity drifts and wavelength drifts, using 96 samples measured over 6 days, which are the same samples used in section 2.4. In order to track both drifts, 9 peaks of the indene Raman spectrum including the highest peak were selected, which are spread widely from the first to the last peak over the whole wavenumber range of CCD.

Figure 2.20 presents an overall view of intensity drift over 6 days by measuring the area under the highest peak. Because of the system resolution, we found that the area under a peak is a better way to track intensity drift rather than the point value at the peak position. Each day's measurement had various laser warm-up times and gap-times between each sample. For instance, day 4 had 16 samples over 4 hours with 30 minute warm-up time and day 3 had 16 samples over 8 hours with 1 hour warm-up time.

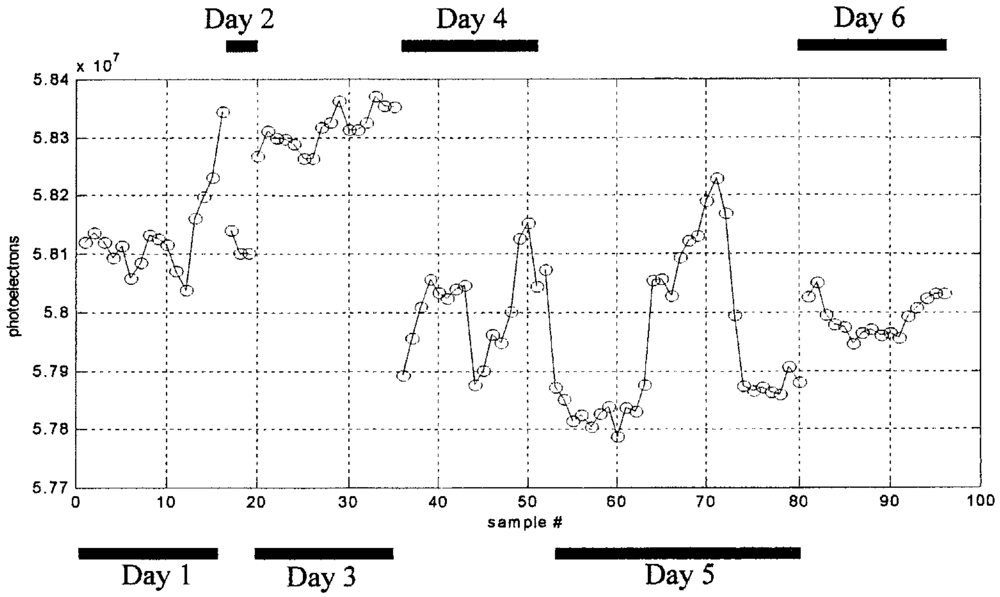


Figure 2.20. Intensity drifts under the highest peak curve over 6 days.

A total intensity drift of approximately 1% of amplitude at the highest peak intensity was observed over 6 days. The drift pattern was irregular, but the drift patterns of all 8 other peaks were similar and fluctuated in the range of 1%. Tracking of the position of the highest peak indicated a one pixel drift in wavelength between day 1 and 2, as shown in Figure 2.21. The drifts of other peaks were within one pixel, but the drift levels were not the same. One of the reasons for this drift is the relatively low spectral resolution of our system, so that variations in the Raman peak at the 674 pixel point (Figure 2.21) are relatively coarse. This peak may be regarded as varying between 2 pixels of similar intensity. There are several possible reasons for this, such as system alignment change and room temperature and humidity changes.

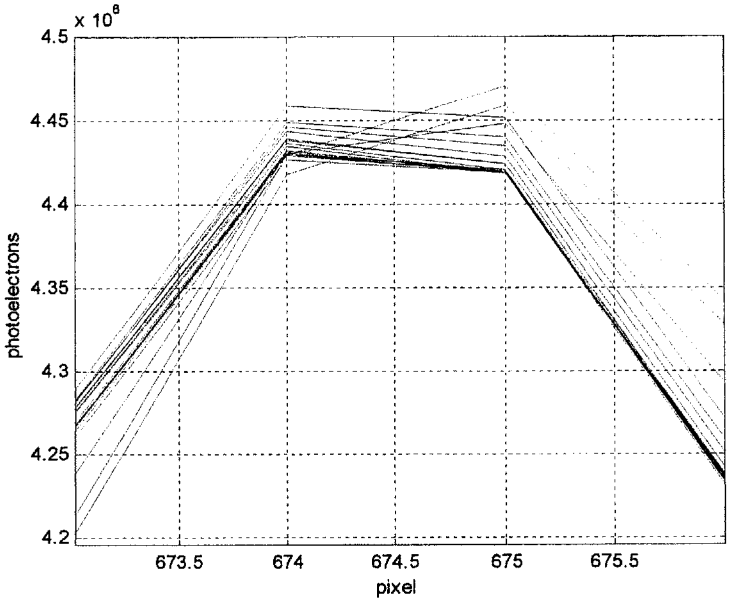


Figure 2.21. Drifts on wavelength of the highest peak between the day 1 and 2.

2.5.3 Enhanced method to correct drifts

It was found that our system stability was close to the shot-noise limit for a short-term experiment. But there were 1% intensity drifts and 1 pixel (1.4 wavenumbers) drifts in wavelength over 6 days. Theses long-term drifts will be discussed here and compared with the effects of shot noise of individual samples to determine whether those drifts were caused by the uncertainty of photons that we cannot control or by system variations. We averaged all 96 samples to generate a spectrum with minimum noise, and then added 96 random noise sets to it to generate 96 numerical samples whose only source of fluctuation is shot noise. These noise sets were created numerically based on the same level of noise as was measured in the system and described in section 2.4. Using the same methods to track the drifts, Figure 2.22 compares the intensity drifts from the original and the numerical samples.

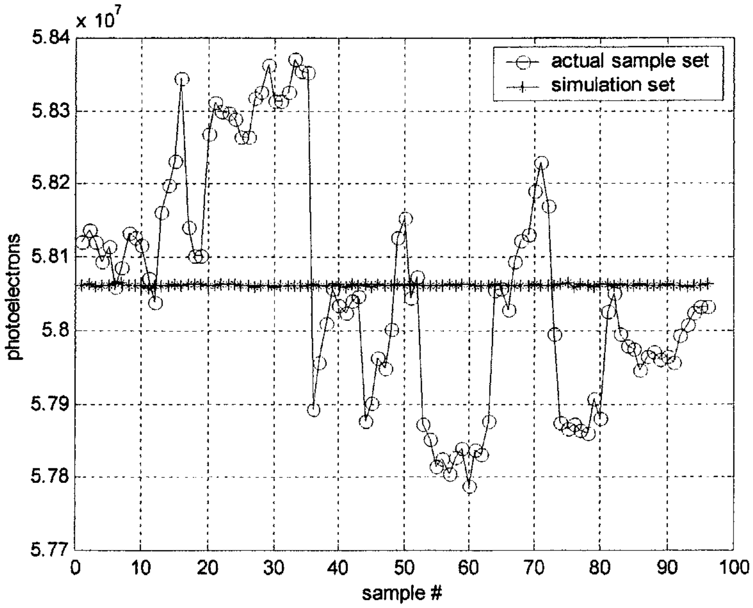


Figure 2.22. Intensity drifts from the original and the numerical samples.

Intensity drift of numerical samples showed relatively small variations, resulting in 0.0013% in photoelectron counts. It verifies that 1% intensity drift from the original samples is much higher than that from the simulated ones. This indicates that the drifts we measured are system drifts, and were introduced by sources of drifts such as laser power and system optics variations. It is possible that temperature and humidity fluctuation may affect the optical properties of the sample and system alignment. Laser power drift is commonly regarded as a considerable cause of system drifts which can be monitored and corrected.

We have been working with two possible solutions to monitor real-time laser intensity drift. First, direct acquisition of laser power change from an external data port of the laser may be available. But there is a question of whether the measurement from the data port would be identical with the laser output power. Two, monitoring a small portion of the laser light, diverted by a beam splitter before it illuminates the sample, may be possible by means of a photodiode. Study of these two methods is still in progress.

We assumed that the laser power drift is one of the primary causes of system drift. If one of the methods described above were to be achieved successfully, the intensity drift caused by the laser power drift could be corrected by the fact of the linearity between the laser power and Raman signal. But this does not provide a solution to correct the wavelength drift.

Here we suggest a new method for correcting wavelength drift, revising a current correction algorithm that will be described in section 3.4. The current code was originally devised to convert a spectrum based on a pixel of a CCD detector into wavenumber that does not depend on any system properties. It uses the pixel for each peak of a reference spectrum that has many sharp Raman peaks spread over the whole wavelength range of the CCD detector. The selected pixels of the Raman peaks are then matched to the wavenumbers of the reference peaks. But as shown in Figure 2.21, the peak is not sharp, so there may be considerable error in finding its frequency with acceptable accuracy. In order to reduce this error, we developed a new method to determine peak position with a precision of 0.1 pixels, by dividing one pixel into 10 sub-pixels by means of spline-curve fitting. This allows us to track more precisely the peak positions required for wavenumber calibration.

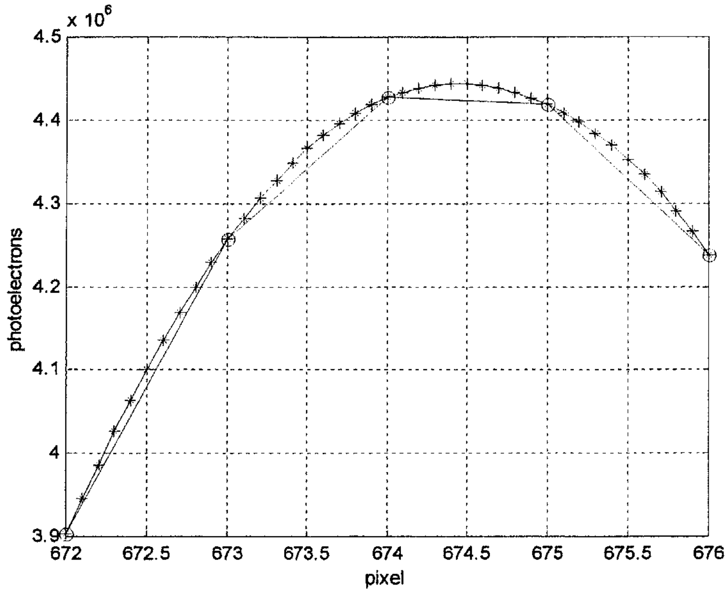


Figure 2.23. Comparisons of a normal peak based on 1 pixel and spline-curve fitting based on 1/10 sub-pixels.

Figure 2.23 plots the highest peak of the 1st Indene sample using both methods. The line with circles presents the normal one-pixel-based peak, and the line with ten star marks in between pixel-based circles is drawn using the new method. Instead of only one pixel drift by the current method, the new method enables us to track the peak drifts more precisely, as shown in Figure 2.24.

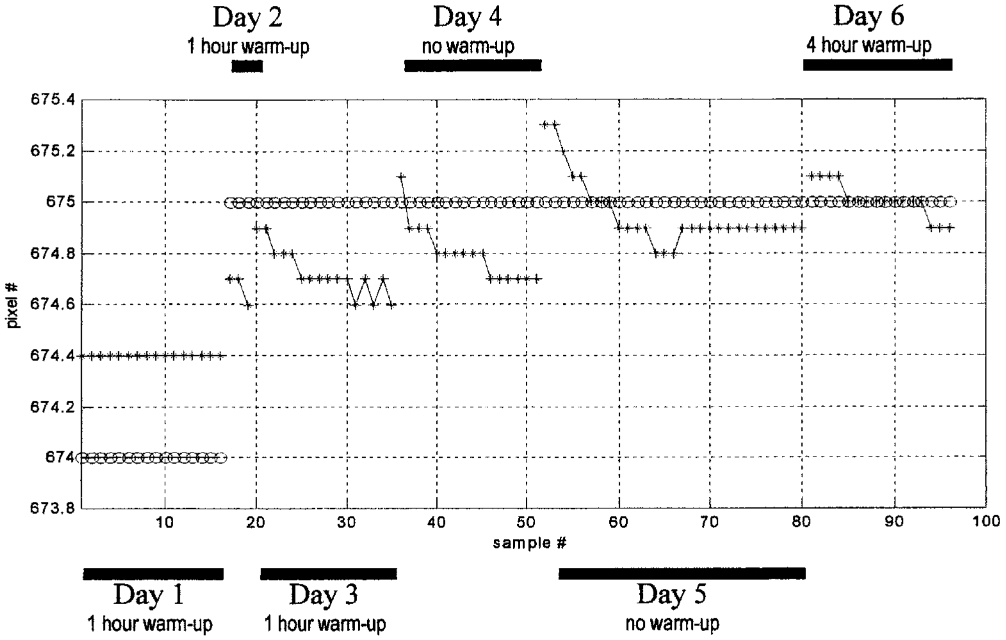


Figure 2.24. The highest peak drifts measured by the two different methods over 6 days.

There were two findings: the wavelength drift pattern revealed by the new method has a regular shape each day, and this appears very similarly in all other Indene peaks. Each day's measurements used various warm-up times, as indicated in Figure 2.24, and exhibited a similar drift pattern, except during day 1. In this typical pattern, the position of each peak starts at a higher pixel number, then moves to a lower one. We could also ascertain that the drift range depends on the warm-up time.

From the above findings, 4 hours or more warm-up time are recommended for better system stability. And the new method will be able to more precisely track and correct system wavelength drift.

3. Data processing

In our studies, the system collects the raw spectrum, which contains all of the spectral information such as the Raman and fluorescence signals from the sample, cosmic rays, and distortions due to system response. It is necessary to apply properly pre-process the raw spectra to extract useful Raman spectral information. Prior to multivariate analysis, we used five pre-processing methods to obtain Raman spectra. We built and ran those pre-processing codes in MATLAB.

3.1 Curvature correction of the fiber bundle image

One of the key parameters for the efficiency of a Raman system is the usable detection area, which is determined by the CCD detector area and the product of the usable slit width and the height. This rule lets us employ a higher fiber bundle to collect more signal, as long as the CCD detection area is not overfilled. Our custom-made CCD detector is large enough to cover our entire fiber array. The fiber bundle consists of 61 fibers. The proximal end of the bundle is arranged in a linear array, matching the entrance slit of the spectrograph, and the distal end is circular in shape to match the focused light emerging from the sample. We designed the line width of the proximal fiber array to serve as the slit by itself, without the use of any external slit. For this reason, our modified spectrograph does not have the slit and the slit holder which are typically mounted in the spectrograph. However, as the height of the linear array at the proximal end of fiber bundle increases, the image focused on the CCD detector becomes more curved. This parabolically curved image occurs when using a spectrograph with a plane grating. This phenomenon is due to the fact that rays from different positions along the length of the slit are incident on the grating with varying amounts of obliqueness. If left uncorrected, the curved slit image will degrade the peak shape and spectral resolution.

Our Raman spectroscopy systems (blood and transcutaneous systems) used the same fiber bundle with 61 fibers (3 broken) as described in Table 2.1. In the blood system, the f/1.8 spectrometer and 17mm high CCD were not able to make use of the full 61 fiber array. As described in section 2.2, the new modified spectrograph and CCD detector allowed the full height of the fiber bundle to be imaged on the CCD (Figure 3.1).

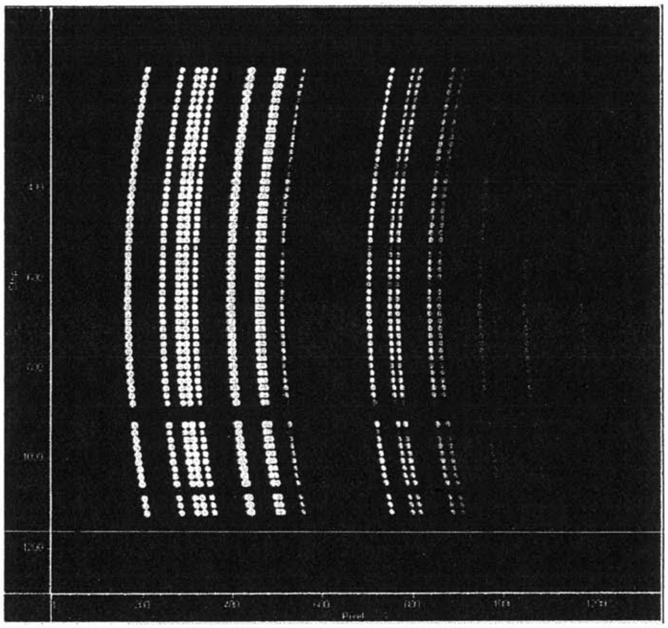


Figure 3.1. Fiber array imaged by neon lamp on 26mm x 26.8mm CCD detector.

The curved image of the fiber bundle at each wavenumber must be straightened before all the pixels in a vertical column are binned. The pixels in each column are binned into a number of separate regions. These regions are then shifted by a number of pixels, based on the distance from the region in the center, in order to prevent loss of resolution. This hardware binning for each designated section (or strip) of the CCD vertical pixels is performed before the signal is read out. We devised a mathematical algorithm to group automatically a few pixels into to a strip within which pixels are binned vertically [Berger 1998]. Figure 3.2 illustrates an expanded view of one part of Figure 3.1 with 3 of the 61 fiber images located in the central wavelength region. The curve passing through the center of each fiber can be divided into a series of steps, as shown by the white lines. Each vertically straight section will be in the form of a strip. The gray band in the figure represents one such strip.

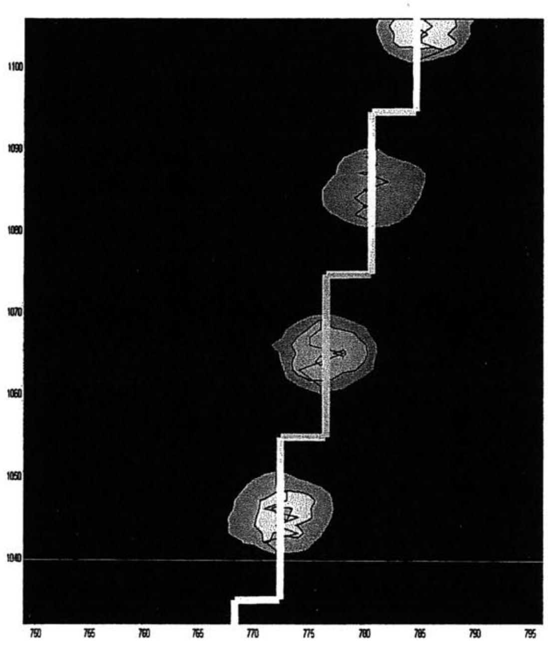


Figure 3.2. Graphical representation of the curve-aligned fiber bundle, forming a series of steps. The gray band region is binned into one strip.

3.2 Cosmic ray filtering

When spectra were collected on a CCD detector, unexpected peaks with very high intensity and narrow widths often appeared on the top of the spectra. They are called cosmic rays or cosmic spikes, caused by cosmic rays hitting the sensor are generating pockets of electrons. These events are localized spatially to a few pixels and are randomly distributed temporally, resulting in a source of data contamination. Since cosmic rays contaminate the Raman bands, they must be removed, lest quantitative analysis of the spectra may provide inaccurate results.

Early methods developed to remove them automatically sought to take advantage of their sharp features compared with relatively broader Raman peaks, but these methods have had limited success [Hill *et al.* 1992; Baraga *et al.* 1992]. More robust methods were developed later that utilize the temporal randomness to identify and eliminate them in successively acquired spectra [Takeuchi *et al.* 1993]. In our experiments the Raman signal from blood and human skin was still not strong enough to obtain high SNR, and so multiple consecutive frames were used to increase SNR. We developed a MATLAB algorithm using multiple frames to find and remove cosmic rays. One sample for in vivo measurement was collected for 3 minutes with 90 consecutive spectra of 2 second integration time each. A statistical algorithm was used to identify cosmic rays under the assumption that a sudden change of intensity over a threshold level in a spectrum beyond typical noise is due to a cosmic ray.

Here we used 90 frames statistically to calculate the median and the standard deviation that was typically generated by noise given in the system. Based on the median value, if the intensity of a pixel point was above a threshold, for instance 3 standard deviations above the background noise, then it would be recognized as a cosmic ray and filtered out automatically. Once the value filtered out, it was replaced by the median.

3.3 White light correction

Spectral line shape calibration, called white light correction, is an essential process to compare spectra from different Raman systems, or from the same system on different days, which means that spectra must be corrected for the spectral response of the system. This is a necessary method for accurate, quantitative analysis so that the spectral information is independent of the features of a given experimental system.

Wavelength-calibrated tungsten white light source is used as a white light source. It then illuminates a reflectance standard (Labsphere, Inc.). The reflectance standard is placed in the same position as the Raman signals collected from the samples. Dividing the white light spectrum by the raw spectra then corrects the system response induced by response of the optical components, the grating efficiency, spectral variations in the CCD, the quantum efficiency, and the fixed pattern noise of the CCD detector, and possibly vignetting of the optics. Figure 3.3 shows two white light spectra taken in the transcutaneous system, one with a 2.0 inch holographic notch filter (Kaiser Optical) and one without a filter. In most spectral regions, except at low wavenumber where the light is cut-off due to the notch filter, its features were reflected mainly by vignetting through the spectrograph, the grating efficiency, and the CCD detector efficiency.

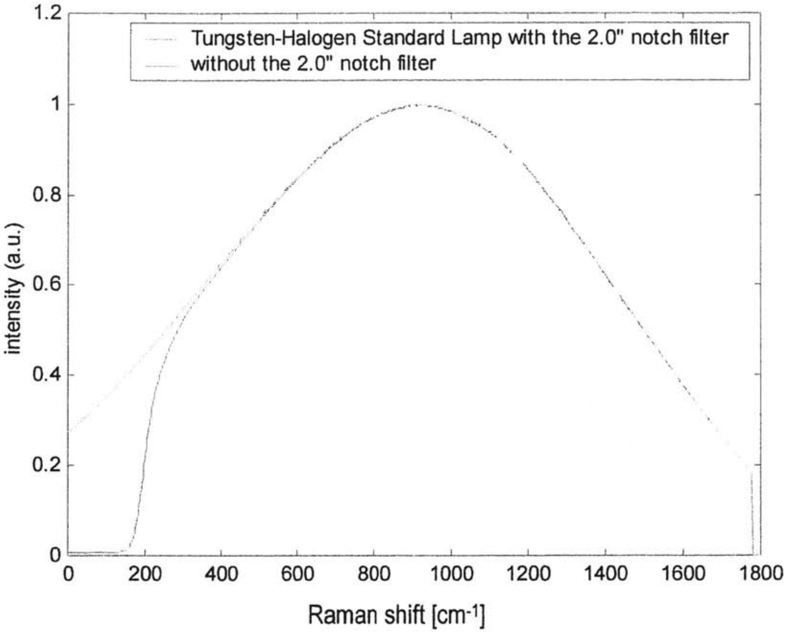


Figure 3.3. White light spectra with notch filter and without notch filter.

Our experimental study of the first transcutaneous measurement of glucose monitored the glucose Raman spectrum from a single volunteer over a short time period (about 2 hours), using the Raman system we designed. Our Raman system was stable enough that we could not see any significant spectral changes during the 2 hour data collection period. Therefore, we did not have to do the white light correction for this study.

3.4 Wavenumber calibration

Since the wavelength of the raw spectrum in wavelength is based on CCD pixel numbers, which depend not on the Raman band frequencies but on CCD structure, a conversion (called wavenumber calibration) is required to obtain an accurate frequency scale, so as to identify chemical constituents in the sample. This is also an important requirement for composing spectra collected in different systems, and also for composing spectra measured in a same system but taken on different days, after significant changes in alignment.

In principle, any known material can be used as a wavenumber calibration standard as long as it exhibits enough well defined sharp peaks. In this study, the spectral wavenumber calibration was performed using the known band positions of indene. We measured the indene spectrum before starting an experiment each day. But as the white light correction was not used for this study, the wavenumber correction was skipped since there was no significant change in alignment. In addition, as described in section 2.4, we determined that our system was stable enough for us to assume that the pixel position of a particular wavenumber shift did not change during the course of a day's measurements.

3.5 Fluorescence background subtraction

In the case of measuring biological samples such as human skin, typically strong fluorescence signals overwhelm an inherently weak Raman signal. Using 830nm excitation is one way to reduce the fluorescence background (see section 2.2.1). Nonetheless, a certain amount of fluorescence background was collected, which was still much higher than the Raman signal. We found that over 90% of the fluorescence background signal came from the biological sample itself by a test using an additional notch filter and a finger as a sample. The test was performed to check the fluorescence background from each optical component in the system induced by intense Rayleigh light surviving through the original notch filter. Consequently, this fluorescence background was present in the raw spectra along with the Raman signal. This fluorescence background varied from sample to sample and decreased by the photobleaching effect when a sample was exposed for a period of time. This varying fluorescence background interfered with our ability to accurately predict glucose concentration changes based on well overlaid Raman spectra, using a multivariate technique such as PLS. Thus, it is desirable to remove this background prior to multivariate analysis, since this background does not provide relevant information.

As the florescence background varies slowly in frequency, as compared to the Raman signal, it is possible to apply a high-pass filter to remove the broad background. It has been demonstrated that this background can be satisfactorily removed by least-square fitting a fifth order polynomial spectral curve to the data [Baraga 1992; Brennan 1995]. An example of applying this fitting procedure to our human skin spectra is shown in Figure 3.4. After subtracting low frequency fluorescence background, only sharp Raman signal is left. The corrected spectra will always have negative regions, when using this method. However, there is no effect on our data analysis, as it still indicates the intensity relative to the baseline.

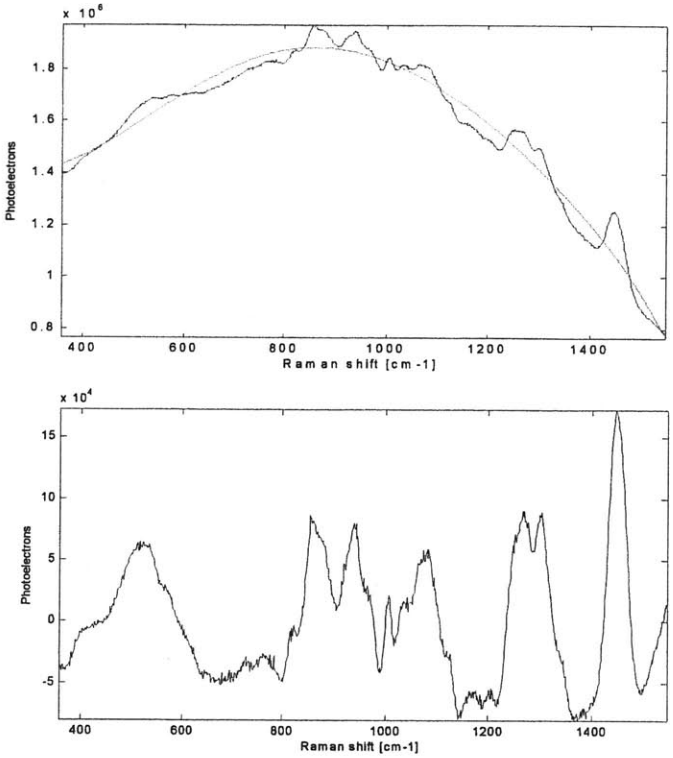


Figure 3.4. The first sample of the 9th human volunteer's skin measurement; Upper plot, raw spectrum in black before fluorescence background subtraction with fifth-order polynomial fit in gray; Lower plot, Raman spectrum after polynomial fit subtraction.

4. Data analysis

Raman spectra consist of numerous spectral peaks, each one due to a specific vibration of a particular molecule. The intensities of their spectral peaks (or bands) at different wavelengths can provide quantitative information about the concentration of each chemical component, providing unique information. The complexity of biological samples often leads to difficulty in interpreting their spectral information, since the presence of multiple components results in many overlapping Raman bands. In addition, the high level of noise mainly caused by fluorescence background makes it difficult to analyze relatively weak Raman peak intensities into accurate quantitative information.

Extracting meaningful, chemical information from such a complex data set can be accomplished by using mathematical techniques known as multivariate analysis, or chemometrics. Such sophisticated analysis techniques enable us to take full advantage of information contained in the spectra. In order to apply multivariate techniques to successfully analyze Raman spectra, two assumptions are required. One is that the Raman spectrum of a mixture of chemicals is simply a linear superposition of each component spectrum existing in the mixture. The second is that there is a linear relationship between signal intensity and chemical component concentration. Both assumptions in biological tissue were experimentally demonstrated [Manoharan 1992]. Manoharan conducted a set of experiments using Raman spectroscopy on mixtures that were similar in composition to that of biological tissue, and found that both of conditions hold within experimental accuracy, allowing the use of Raman spectroscopy to extract precise chemical concentrations from the mixture.

Based on this linearity assumption, Raman spectra measured from samples can be formulated in a numerical form:

$$S=C\cdot P \tag{4.1}$$

where S is the matrix of Raman spectra taken from the sample in experiments. C includes concentration information of all components, and P is the matrix of all component spectra that exist in the sample. In concentration measurements, the purpose is to find a projection vector that predicts concentrations from spectra:

$$C=S\cdot B \tag{4.2}$$

where B is the matrix of each component of the projection vector. For each analyte in every data set, there exists only one ideal projection vector that models the data set most accurately.

Various multivariate analysis methods have been developed to estimate the concentrations efficiently and accurately. Different cases depending on conditions given for analysis require different techniques; for instance, the ordinary least squares method (OLS) is useful in the case that a complete set of the component spectra is known. When the concentrations of all the calibration samples are known, the classical least squares method (CLS) can be employed. Those are often classified as explicit techniques, while other techniques, called implicit techniques, do not require that all of the component spectra are known. As implicit techniques, principal component regression (PCR) and partial least squares method (PLS) can analyze Raman spectra without requiring a complete set of concentrations or component spectra [Haaland *et al.* 1988; Geladi *et al.* 1986]. Explicit methods are more robust, but they require the composition of the mixture to be completely characterized. In this dissertation, OLS and PLS are discussed in the next sections.

4.1 Ordinary least squares Method (OLS)

In Raman spectroscopy, OLS method is a simple but powerful tool if the Raman spectra of all components are known. For example, OLS can be used to predict the concentrations of a simple sample; glucose, urea, and ethanol in water. In this case, the Raman spectra of each component can be measured and put to the matrix P in Eq. (4.1).

The goal of OLS is to obtain the concentrations, C, from the measured Raman spectra, S. From the Eq. (4.1), we can directly obtain concentration from $C=S\cdot P^{-1}$, if the matrix P is invertible. But the number of spectra and the number of variables in the spectra are rarely equal. By an alternate method, if $(P\cdot P^T)$ is invertible, then multiplying P^T and taking the inverse of $(P\cdot P^T)$ enable us to find the projection (or calibration) matrix B;

$$S\cdot P^T=C\cdot P\cdot P^T$$

$$C=S\cdot P^T\cdot (P\cdot P^T)^{-1}$$

$$B=P^T\cdot (P\cdot P^T)^{-1} \tag{4.3}$$

Once the pure components spectra are known, one can construct the calibration matrix B, from which all of the concentrations of an unknown sample can be predicted.

4.2 Partial least squares method (PLS)

As mentioned above, one needs to know what type of information is available before determining the most appropriate analysis technique in a given situation. OLS is the best technique to use if the spectral components of the system are completely characterized, while implicit methods, PCR and PLS are required when information about only one or a few components is available. PCR and PLS are often called data compression techniques, because they minimize the number of parameters needed to characterize the measured spectra. PCR generates a set consisting of a small number of specially chosen basis spectra, called principal components, which can accurately characterize all of the spectral changes in a set of spectral data. These principal components can be calculated using singular-value decomposition. If a sample set has one chemical component, only one principal component is necessary to explain the spectral variance, and the other principal component is probably due to noise or measurement error, which is not generally useful.

PLS is a similar technique from the point of view that both technique decompose the spectral data set and find the relationship between the decomposed spectra and the concentrations. The difference comes from the fact that PLS utilizes the concentration information in the decomposition process. The decomposition is performed with the covariance matrix of spectra and concentrations, the benefit of which includes reducing the effect of noise and obtaining components that are better correlated to the concentrations of the component of interest. The use of the covariance matrix of spectra and concentrations allows the averaging of noise in spectral measurements by concentration measurements, or the averaging of error in concentration measurements by spectral measurements. PLS and PCR have similar performance if noise in the spectral data and errors in the reference concentration measurements are negligible. Otherwise, PLS generally provides better analysis than PCR [Thomas *et al.* 1990].

4.3 Preprocessing for PLS

Curvature correction, cosmic ray removal and fluorescence background subtraction were explained in Chapter 3. In addition, several additional preprocessing steps to help PLS work efficiently, spectral range selection, mean centering, and binning.

The basic idea of spectral range selection is to include a spectral range that contains the spectrum of the specific chemical component of interest. Multivariate calibration methods mathematically try to find the spectral components that are correlated with the species concentration information. With the presence of a strong signal which varies distinctively from sample to sample, the algorithm may try to fit this strong signal, neglecting the variance of other regions. In addition, it is not logical to use a spectral range which does not contain the spectrum of the specific chemical component. Also we determined that the prediction errors were not significantly affected by various spectral ranges as long as the ranges contained the main spectral features of the component of interest. With no special algorithm devised for selecting the optimal spectral range, we chose the wavelength range, from 340 to 1550, which includes all of main features of glucose. Figure 4.1 illustrates this range from a spectrum of high-content glucose solution in clear water, and a spectrum of water alone. We employed this range to analyze the human skin spectral data.

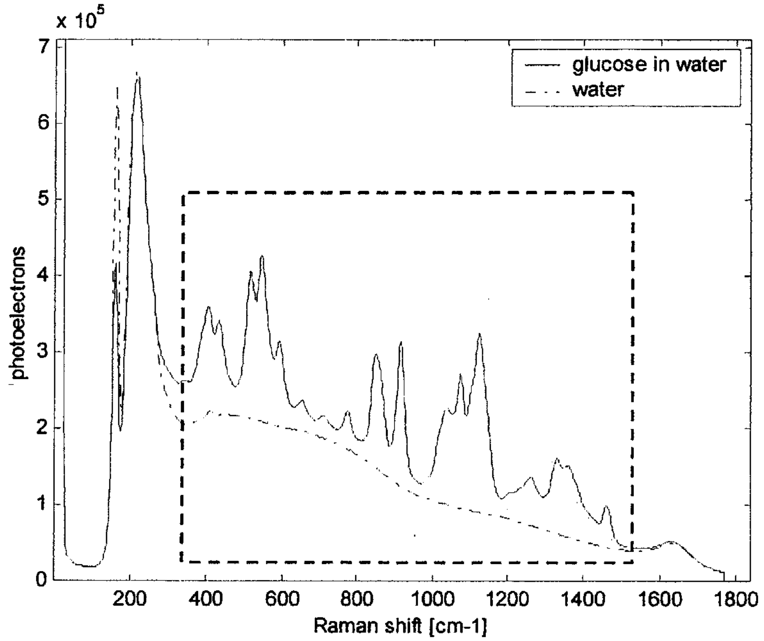


Figure 4.1. Spectral range selection from 340 to 1550 (cm^{-1}) in dash-line box; 2 second measurement of glucose in water plotted by a solid line. The dashed line shows the spectrum of pure water obtained under the same experimental conditions. Those were contained in a cuvette which has magnesium fluoride (MgF_2) exit windows.

In our study, PLS was always conducted after mean centering, since it emphasizes the differences in the sample set, reducing the complexity of the data by one degree of freedom, and prevents certain spectra in the data set from being weighed more than other spectra. Without mean centering in PLS, the first spectral component is the statistical mean of the data set, which is often not the information that is needed.

Although we have improved the system to collect more Raman signal, the noise, which is mainly shot noise coming from a high level of fluorescence background, is still a problem, compared with the relatively low Raman signal of the target chemical component [Beger *et al.* 1997]. Several noise reduction methods have been devised. We used one of them, a binning method. Here binning is performed by software, unlike the hardware binning that is performed in the CCD. Binning is combining the intensities of a certain number of pixels. Binning increases SNR by the square root of the number of pixels. Additionally, it can simplify the model and increase the robustness of multivariate analysis. The primary concern in using binning is degrading Raman spectral features. In our work, binning with the size of 10 was used, which means that 10 pixels were combined into one data point. Figure 4.2 shows that key features of Raman spectrum of glucose, the target analyte, are maintained in binned spectrum.

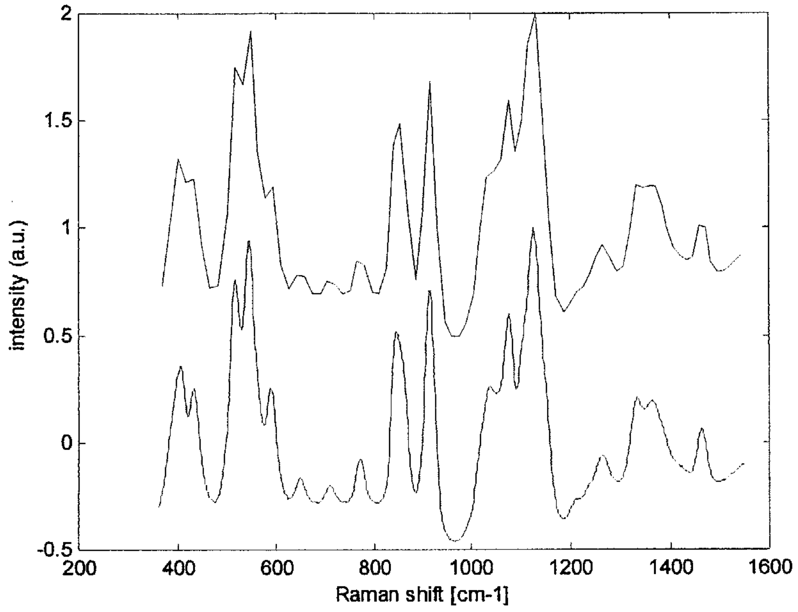


Figure 4.2. Comparisons of Raman spectrum of glucose in water before binning (lower plot) and after binning (upper plot).

4.4 Validation and prediction error

PLS requires a calibration set of reference measurements (e.g. we use the glucometer for reference measurements of glucose). Using this technique, we decompose the original Raman spectra into orthogonal B-vectors or calibration spectra by utilizing the changes in concentration measurement obtained from the reference. That is, changes in concentration are related to changes in the spectra to determine which unique parts of the spectra are correlated with changes in concentration. This process is generally called calibration. In calibration, a mathematical model (calibration spectra) is developed to describe the relationship between chemical concentrations and Raman spectra. Once the calibration spectra are generated with the reference measurement, they can be projected onto other spectra to predict the concentration of the new data. In this process, an independent data set, which does not include the data used for the model building, is often used for testing the robustness of a model, since it is desirable to use an independent data set for validation. However, this requires a large number of samples. When the number of samples is not sufficient to perform an independent validation, cross-validation can be performed. Cross-validation allows for the efficient use of a data set, because only a small number of samples are reserved, and the spectra of the remaining samples are used for calibration. Since the data set used for calibration and the data set used for validation are independent, the validation is performed without bias. In the human volunteer study, a single sample was spared from each calibration in a "leave-one-sample-out" cross-validation, and the calibration and validation process is repeated as many times as the number of samples in the data set.

In concentration predictions, prediction errors indicate the quality of the predictions. For each sample, the prediction error is the difference between the predicted concentration and the concentration measured by a reference technique. For a set of samples, the prediction error for a data set is the standard deviation of the prediction errors for all samples. It is also called the root mean standard error of prediction (RMSEP). The correlation coefficient, r, is often used to quantify the accuracy of the other method. Although the tolerable limits of correlation are somewhat arbitrary, "correlation coefficients of over 0.9 are usually deemed acceptable, and indicate that the new method agrees satisfactorily with the reference method." [Pincus 1996]

5. Experiments for noninvasive measurements of glucose

5.1 preliminary experiments

We conducted preliminary studies before we performed experiments using human subjects. First, we estimated that the amount of glucose signal in the range, which could be shown in Raman signal from a human body noninvasively, would be shown with proper noise given our system. Then we measured glucose using tissue-like samples, tissue phantom and pigskin based samples, which provided well controlled samples having similar optical properties to human skin.

5.1.1 Human skin and glucose spectrum

Glucose is the carbohydrate essential to all body cells as an energy source. The human body controls its concentration precisely by insulin secreted from the pancreas. The glucose concentration in a normal human body lies typically in between 45 and 180 mg/dL (2.5 to 10.0 mM) in plasma [Threatte *et al.* 1996]. People with diabetes cannot control glucose concentration efficiently, so that an extreme case of over 1000 mg/dL has been reported. Glucose concentrations in that concentration range are expected to be obtained spectroscopically when measuring a whole blood or plasma sample. While whole blood is a uniform mixture of blood cells and plasma, human skin is an inhomogeneous composition of the blood-tissue matrix affecting the optical properties and the glucose distribution. Almost all parts of human skin are covered by the most superficial layer, the epidermis, the thickness of which ranges from 40 μm in the lips to over 1 mm in the soles of the feet. Melanin, the major source absorbing the light around

830 nm as shown in Figure 2.2, is in the epidermis. The dermis, located beneath the epidermis, also varies in thickness depending on the location of the skin. It is 0.3 mm on the eyelid and 3.0 mm on the back. The dermis contains blood vessels as well as many specialized structures. While whole blood is present in blood vessels in the dermis, interstitial fluid (ISF) lies between cells in the epidermis and dermis. ISF is the fluid that surrounds the individual cells in the body. Nutrients diffuse from the capillaries into the interstitial fluid where they are absorbed into the cells. The importance of ISF is that blood glucose and ISF glucose levels are essentially equal. However, rapidly changing glucose levels create a lag between blood and ISF measurements. Overall, the differences do not much influence the in vivo measurement of glucose, because they are minor and negligible (lags in glucose levels last usually less than 10 minutes).

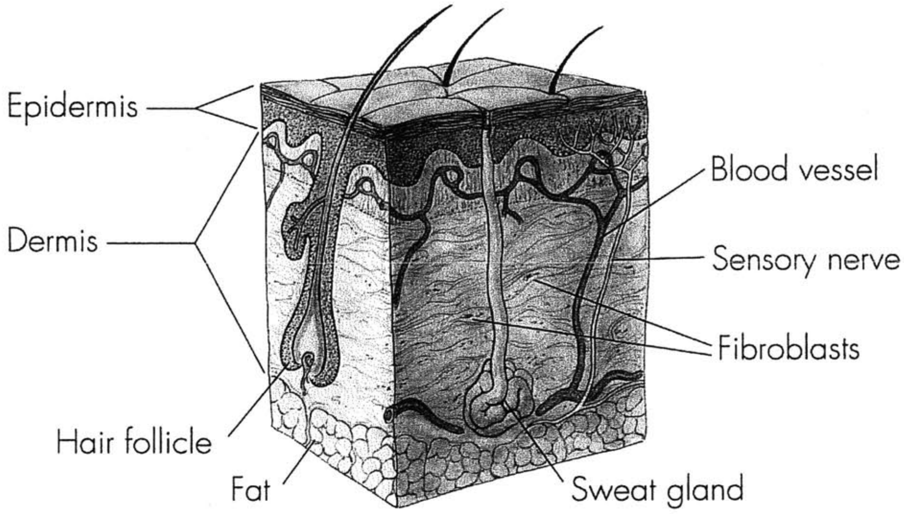


Figure 5.1. Epidermis and dermis structure.

We assumed that the sources of glucose Raman signal from human body are glucose resolved in whole blood and ISF. This assumption allows us to estimate the glucose Raman signal intensity quantitatively, investigating the weight fraction of both in total biological fluids of human skin. The total water content in skin is 69.4 weight% [Widdowson *el al.* 1960], of which 37% is extracellular [Roe *et al.* 1998] with a glucose concentration equivalent to that in whole blood [Jensen *et al.* 1995]. The remaining 63% is intracellular with a glucose concentration corresponding to approximately 10% of that in the extracellular fluid [Guyton *et al.* 1997]. As a result, 30% of glucose signal that can be obtained in vitro will be collected in vivo.

Using the glucose spectrum measured in our system (Figure 4.1), we can estimate the intensity of a glucose Raman signal in human skin, based on the above weight fraction. Figure 5.2 shows the Raman spectra of 90 mg/dL and 180 mg/dL glucose with a typical level of noise calculated in section 2.2.4. Both glucose concentrations are in the range of glucose found in healthy people. These glucose signals were taken from glucose in water and rescaled to reflect the factor, 30%. They were performed under the same experimental conditions as is used in human volunteer studies, such as a 2-second exposure time.

Both noisy spectra still exhibit discernable glucose bands. When defining the signal level for glucose as the height of the highest peak of glucose, the SNRs with glucose at 90 and 180 mg/dL are 4.7 and 9.3, respectively.

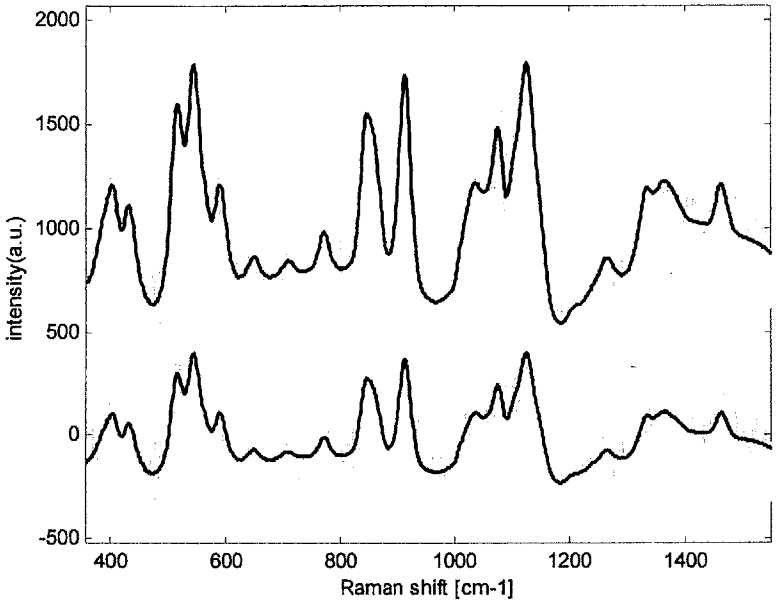


Figure 5.2. Glucose Raman spectra with give noise level; 90 mg/dL (lower plot) and 180 mg/dL (upper plot).

5.1.2 Dosimetry and safety

Although the significant advantage of Raman spectroscopy over other spectroscopic techniques is that its spectrum has distinct and pronounced peaks, its primary problem is the inherent weakness of the Raman signal. Because of this, there is a need for high excitation power and relatively long signal collection time in order to reduce noise. Keeping the laser power below the level of damage in measurements in vivo is critical. It was necessary to understand how much power is allowed and ensure that the laser power to be used in vivo, 330 mW, is safe.

The American National Standards Institute (ANSI) set a guideline for safe use of lasers and laser system [ANSI.Z136-1 2000, ANSI.Z136-3 1996]. For an exposure to an 830 nm excitation beam for longer than 10 seconds, the ANSI standard recommends less than 0.36 W/cm^2 irradiance on skin as a level of "comfort". However these recommendations are extremely conservative. We have looked for other investigators' experimental conditions related to this issue. There have been NIR Raman spectroscopy studies using several organs, such as cervix, skin, and retina. Their excitation power levels varied from 2.5 to 20,000 W/cm^2. None of them has reported any physical damage of samples [Shim *et al.* 1997; Caspers *et al.* 1998; Bakker Schut *et al.* 2000; Buschman *et al.* 2000; Caspers *et al.* 2000; Hata *et al.* 2000; Shim *et al.* 2000; Ermakov *et al.* 2001; Utzinger *et al.* 2001].

In our laboratory, a preliminary experiment was conducted using another Raman system designed for clinical applications. Samples of normal aorta of a cadaver were exposed to a focused bean (830 nm, 200 μm diameter at the sample surface) of 280 mW radiant power for up to 600 seconds. When the histology of the tissue samples was examined to investigate the effects of NIR laser exposure on tissue, no thermal damage was found. Similar experiments were also carried out with skin, in which case no evidence of histological changes was observed [Motz JT 2003]. Although we could not directly compare this experiment with in vivo measurements, it seemed that our laser exposure power was safe, considering additional mechanisms for heat dissipation, such as blood flow. This experiment was performed prior to our human volunteer study.

5.1.3 Tissue phantom study

To demonstrate the feasibility of in vivo glucose measurements, a preliminary experiment was performed with a blood tissue phantom, a mixture of polystyrene spheres, hemoglobin and water. Polystyrene spheres and hemoglobin played the roles of a scatterer and an absorber respectively. The optical properties of the phantom are similar to those of Caucasian skin (the scattering coefficient μ_s of 23 mm^{-1}, the absorption coefficient μ_a of 0.07 mm^{-1}, and scattering anisotropy factor g of 0.86) [Troy *et al.* 1996]. The goal of this experiment was to measure various concentrations of glucose in a reproducible skin-like sample.

We prepared 11 tissue phantom samples, spiking various glucose concentrations in a wide range from 0 to 360 mg/dL with 18 mg/dL increase per sample. Raw spectra were collected and pre-processed, as explained in Chapter 3. For each sample, 30 frames of 2-second spectra were taken, and the total collection time was 1 minute. The excitation beam power was 330 mW on the sample, and the area of the beam incident on the sample was approximately 1 mm^2. The spectra were cosmic ray filtered and fluorescence background subtracted in the manner described in Chapter 3. Multiple frames were averaged to improve the SNR of the spectral data. The raw spectra from this experiment are shown in Figure 5.3, and the preprocessed spectra, Raman spectra are shown in Figure 5.4. The overall intensity of tissue phantom raw spectra is compatible to that of human skin raw spectra in Figure 3.4.

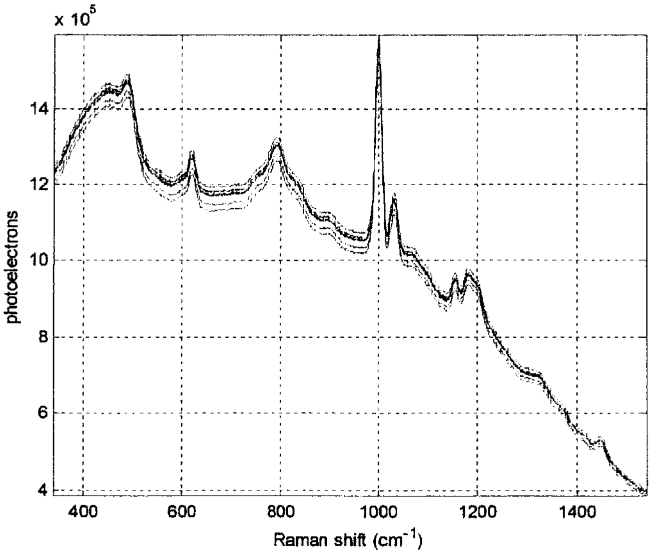


Figure 5.3 Raw spectra of 11 tissue phantom samples.

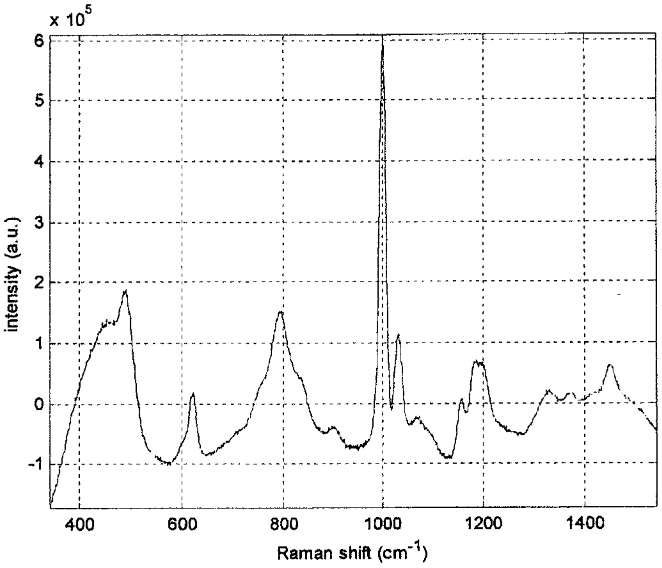


Figure 5.4 Raman spectra of 11 tissue phantom samples after processing.

PLS was performed with leave-one-sample-out calibration and with binning size 8 as described in Chapter 4. The predicted concentrations are plotted in Figure 5.5. The prediction error was 12.7 mg/dL (0.71 mM) for 1-minute exposure time. The x-axis represents reference concentrations that we already know. The y-axis represents concentrations predicted by the multivariate technique, here PLS. The dashed line in grey is for "zero-prediction-error" data line, and the magnitude of the vertical distance from each data point to the line is the prediction error for that data point.

The correlation coefficient, r, indicates how two variables are correlated. A correlation coefficient close to 1 means close correlation between the variables, and r close to 0 does little correlation; being close to -1 is inverse correlation. In this tissue phantom measurement, the r^2 value of 0.99 was calculated, which means very strong correlation between the reference concentrations and the predicted concentrations.

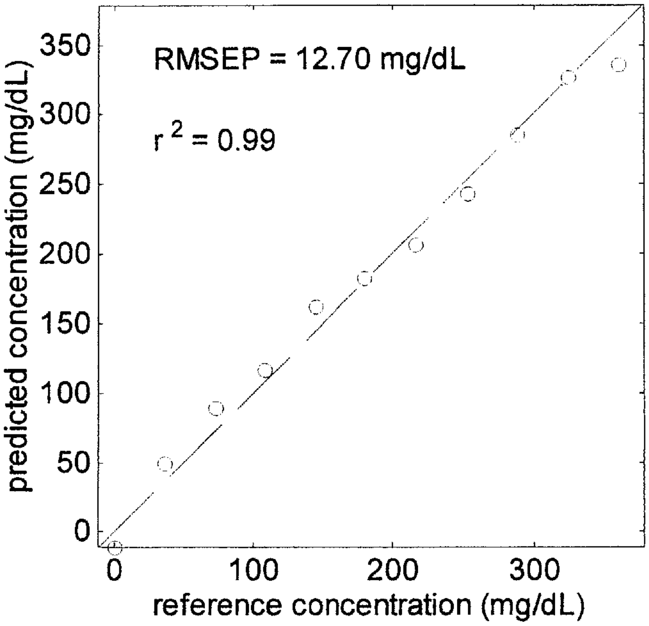


Figure 5.5 Prediction plot of glucose in tissue phantom samples.

We can verify that the predicted concentrations from PLS analysis are based on glucose by examining the calibration regression spectrum called the b-vector (described in section 4.4). The PLS calibration spectrum is compared with a glucose spectrum for comparison in Figure 5.6.

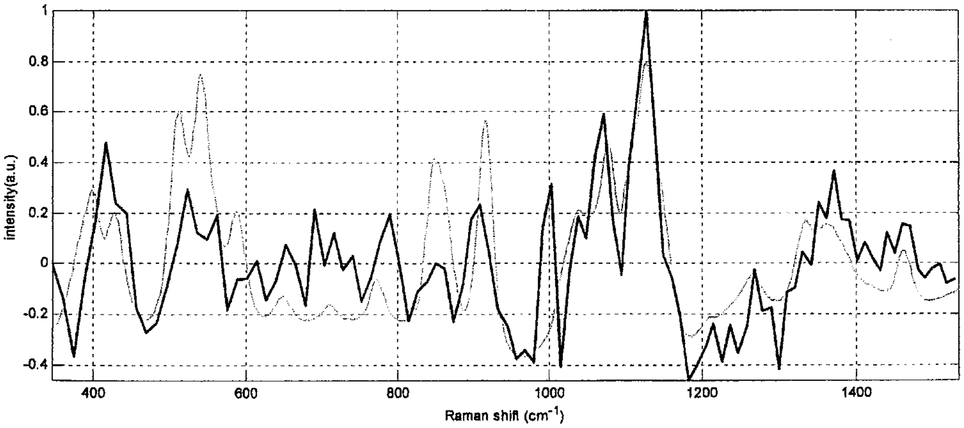


Figure 5.6 The PLS calibration spectrum (black) and the glucose Raman spectrum (grey) for comparison.

The calibration spectrum is similar to the glucose spectrum, showing most of the prominent glucose peaks, which indicates that the calibration is based on the glucose molecule. As a result, the tissue phantom measurement confirmed that glucose within a physiologic range that can be found in people with diabetes generates Raman signals of sufficient magnitude that glucose levels can be accurately predicted.

5.2 Experimental procedures

We invited 20 human volunteers to participate in the first transcutaneous study of glucose using NIR Raman spectroscopy. Our transcutaneous system described in section 2.3 was used, replacing the sample holder with a cuvette with a solid mount to hold the volunteer's arm tightly and easily. All 20 volunteers, healthy Caucasian and Asian, came to our laboratory satisfying 12-hour fasting condition, which required no food (only water allowed) for about 12 hours before the study. The spectra were collected from the right forearms held motionlessly on the sample position. After taking the first spectra, the volunteer drank 220 ml of a high glucose condensed beverage, called Sun-dex. Spectra were acquired every five minutes over a period of about 2 hours (2.3 hours, on average). Capillary blood samples were drawn every ten minutes, and a full reference data set (five-minute internals) was obtained from interpolations between measured glucose concentrations. A Hemocue glucose analyzer (Hemocue, Mission Viejo, CA) was employed for the blood analysis. This procedure, drinking Sun-dex and monitoring glucose changes for 2 to 3 hours, is often conducted in a hospital. It is called the oral glucose tolerance test (OGTT or GTT) and is a diagnosis method for diabetes and reactive hypoglycemia. A laser power of 330 mW at 830 nm was delivered to the forearm in a spot diameter of ~1 mm. Spectral data was collected in 90 frames, each 2 second (total collection time up to 3 minutes). The protocol was approved by MIT's Committee on the Use of Humans as Experimental Subjects (COUHES). Data from 4 volunteers was excluded; One volunteer was on antibiotics and showed a delayed and weak response to the Sun-dex. One volunteer developed a blister at the measurement position (apparently from a high melanin content in the skin). Two volunteers with low blood circulation in their fingers caused difficulties with blood sampling.

5.3 Analysis

A total of 431 Raman spectra from 16 volunteers were collected. Glucose levels ranged from 80 to 230 mg/dL. Figure 5.7 (a) shows the changing raw spectra from the 9th volunteer. The raw signal varied by over 20% during the 110 minute course of the test. Figure 5.7 (b) shows the Raman signals after background subtraction.

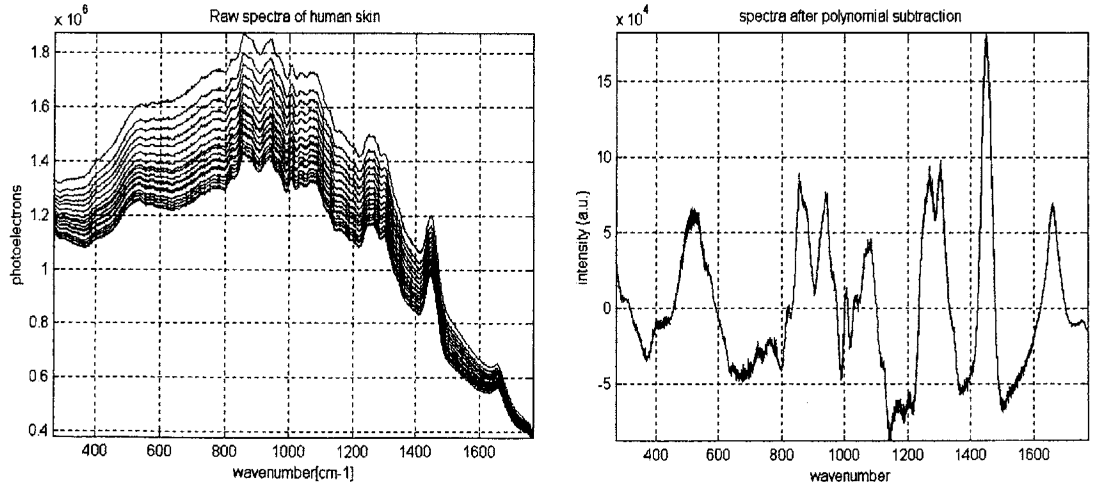


Figure 5.7. (a) Raw signal (left) and (b) Raman signal (right) from volunteer #9.

In Figure 5.7 (b), the spectra look consistent, but a closer examination reveals a drift in the peaks of as much as 7% of the Raman signal. After pre-processing such as cosmic ray filtering and fifth order polynomial subtraction, the spectra were binned to 10 pixels (corresponding to 14 cm^{-1}) and evaluated by PLS analysis with leave-one-sample-out cross validation. 5 loading vectors were used for analyzing volunteer #9. Other volunteers required 3 to 11 loading vectors to build the PLS calibration models.

5.4 Results

Figure 5.8 compares the predicted glucose concentrations from the 9th volunteer with the corresponding glucose reference concentrations. Except for one outlier (excluded), the predicted concentrations followed the reference data well. A high value of the squared correlation coefficient (r^2) of 0.95 between reference and predicted data was obtained.

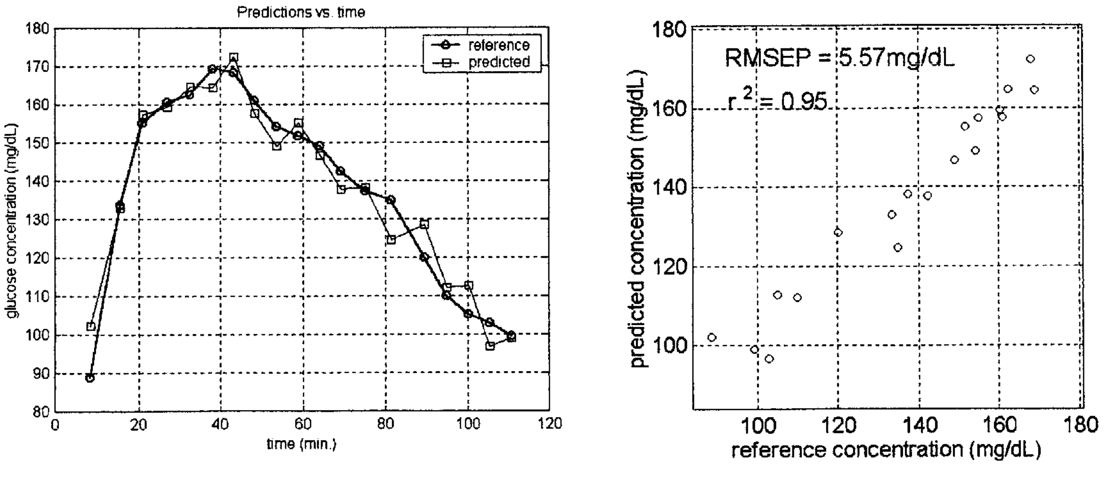


Figure 5.8. The predicted glucose concentrations and the reference concentrations plotted against time (left) and the predicted glucose concentrations vs the reference glucose concentrations (right).

An RMSEP of 5.57 mg/dL, as well as an r^2 of 0.95 is in the range of clinical accuracy. This is one of our best results. The highest error observed was from volunteer #13. The RMSEP and r^2 are 20.23 mg/dL and 0.61 respectively. Table 5.1 summarizes the results of all 16 volunteers.

Table 5.1. Prediction errors and squared correlation coefficients (Note : 4 volunteers were excluded – see section 5.2 for details).

Volunteer #	RMSEP [mg/dL]	r^2
1	18.00	0.79
4	14.01	0.87
5	8.94	0.78
6	12.90	0.82
7	10.64	0.94
9	5.57	0.95
10	12.62	0.57
11	7.01	0.96
13	20.23	0.61
14	14.22	0.86
15	17.22	0.81
16	13.70	0.81
17	8.34	0.91
18	19.99	0.37
19	12.54	0.83
20	15.46	0.79

r^2s' of 4 volunteers are over 0.9. Volunteer #18 shows very low r^2 values, probably due to smaller spread of glucose levels than the others.

When considering all 16 volunteer results, the average RMSEP and r^2 values were 13.17 mg/dL and 0.79, respectively. Figure 5.9 illustrates the Clark error grid plot of the data from all 16 volunteers.

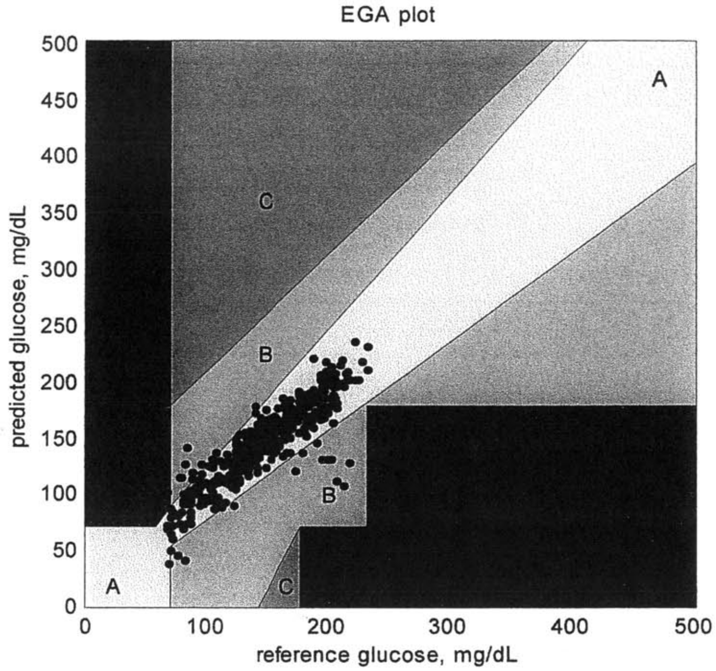


Figure 5.9. Glucose predictions plotted on Clark error grid.

The Error Grid Analysis (EGA, the Clarke Error Grid) is used to assess the clinical relevance of the differences between the predicted measurements and the reference ones [Clarke *et al.* 1987; Cox *et al.* 1997]. As shown above, almost all predictions are in Zones A and B in the EGA plot (except for one point). While results in zones C, D, and E are potentially dangerous and therefore clinically significant errors, results in zones A and B are considered clinically acceptable.

As a result of the narrow peaks inherent in Raman spectroscopy, inspecting the PLS calibration spectra allows us to evaluate the extent to which the concentration predictions are based on glucose spectral features. A comparison of the calibration vector for volunteer #9, one of best results, and the glucose spectrum is shown in Figure 5.10.

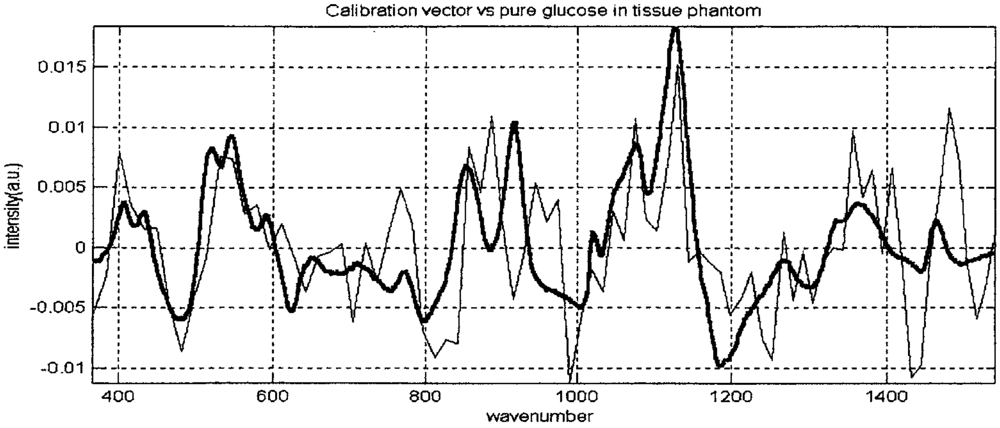


Figure 5.10. The PLS calibration spectrum (thin line) compared to the glucose spectrum (thick line).

The most prominent peak of glucose is located in the wavenumber range of 1000-1200 cm^{-1}. This peak and several glucose peaks, such as two peaks at 400-600 cm^{-1}, can be recognized in the calibration spectrum. Five of volunteers have a good correlation, visually corresponding to the glucose feature. It should be noted that the calibration spectrum is never identical to the glucose spectrum, because it is also composed of spectra from interfering components in skin.

A high degree of correlation between the calibration spectrum and the glucose spectrum is a good indication that the calibration is based on glucose. However, not all of 16 calibration spectra show such good correlations to glucose.

6. Conclusion and Future Directions

6.1 Conclusion

We have improved the blood system to make it more suitable for transcutaneous measurements, collecting larger Raman signals by a factor of 4. Using this improved system, the first transcutaneous Raman measurements of glucose were performed. We collected in vivo Raman spectra from 20 healthy human volunteers. Based on results of 16 volunteers, the average RMSEP of 13.17 mg/dL shows promise that Raman spectroscopy can be developed to achieve the requirements of clinical accuracy. Furthermore, glucose features could be recognized in the calibration spectra from measurements of those volunteers whose glucose levels changed over a relatively large range. Five of the PLS calibration spectra showed strong correlation with the glucose spectrum especially in the unique glucose Raman band region, 1000-1200 cm^{-1}. This indicates that, for these volunteers at least, the predictions were clearly based on glucose.

The volunteers who participated in our experiments were all healthy people, which means that the net glucose change is typically lower than that of people with diabetes. A calibration generated from data obtained from people with diabetes is expected to be more accurate because of the wider spread of their glucose levels. Before an expanded human volunteer experiment with people with diabetes, we need to improve our system and revise the protocol for multiple volunteer studies based on issues found while conducting the first transcutaneous study, such as more stable spectrum collection, decreasing an effect of fluorescence photobleaching, etc. We continue working on this project and investigating the causes and solutions of issues which will be discussed in following sections.

6.2 System improvement

The Raman system used for our transcutaneous study employed a fiber bundle composed of f/1.8 fibers. Subsequent to that study, we have upgraded the system with a new fiber bundle (f/1.4 and 25 mm height of the slit side) to collect more signal efficiently, matching the NA of the spectrograph, f/1.4. We evaluated the new fiber in terms of signal increase, using methanol and whole blood samples in cuvettes. About 20% increases for both sample measurements were observed, and this agreed well with a ZEMAX simulation. Figure 6.1 shows that the new fiber image fills out the full height of the CCD detector.

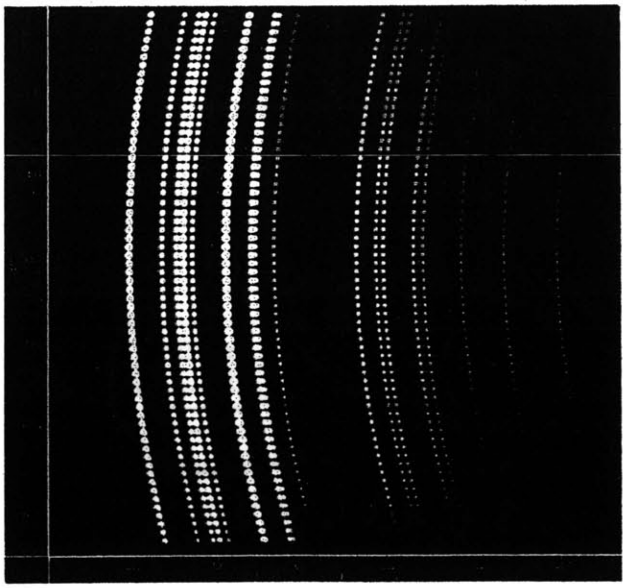


Figure 6.1. New fiber bundle imaged on the CCD detector (f/1.4, 360um core diameter 65 fibers, 25mm height).

However, this 25 mm height of the fiber array makes the distortion by the curvature more significant than that of the 20 mm height fiber. Although the curvature correction process works nicely, there is another problem concerning the wavelength dependence on the degree of the curvature [Zhao 2003]. So far, we have an on-going task to develop a new algorithm to minimize the distortion affect on the wavelength. We expect this new code to be tested soon.

Independent data taken from multiple volunteer measurements and measurements made over multiple days re necessary to build a robust model. Since we have conducted a single day measurement per each volunteer, wave number correction and white light correction processes could be skipped. However, these two corrections are critical to exchange data taken from multiple days, reflecting possible differences caused by system drift and system alignment change. To address this, we have integrated two additional features into our system. One is the white light measurement equipment (discussed in section 3.3).

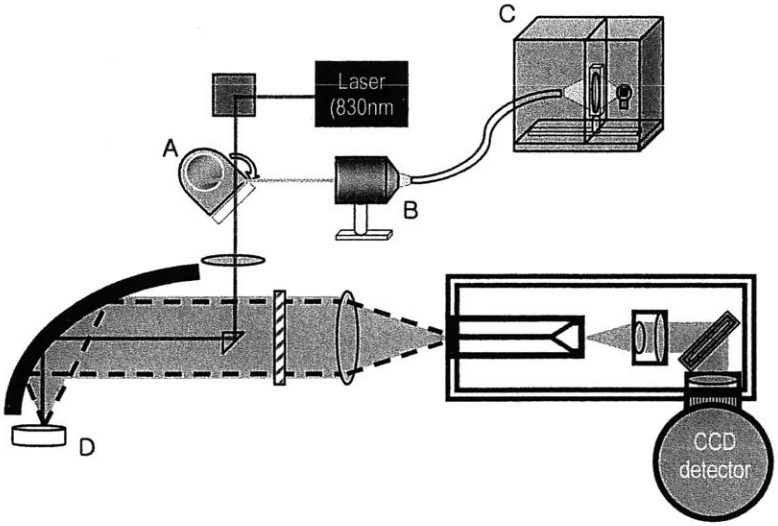


Figure 6.2. White light correction instrument, (A; mirror and flip holder, B; 10x microscope objective, C; light collection box from tungsten bulb, whose light is focused a lens to 100 um fiber, D; a diffuse reflectance standard from Labsphere, Inc.).

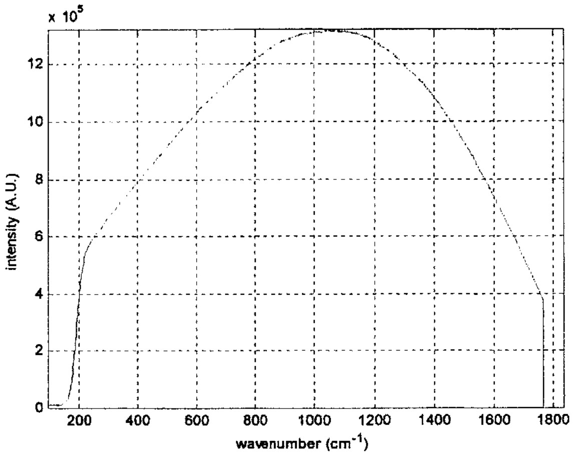


Figure 6.3. White light spectrum using the integrated white light correction instrument

Figure 6.2 is a schematic drawing of the white light correction addition. White light from a standard tungsten bulb is delivered through a single fiber. A mirror mounted on a flip holder directs the white light collimated by a 10x microscope objective, allowing the light to pass through the same path of the laser light. Before and after a volunteer measurement, the white light is recorded from a diffuse reflectance standard (Labshpere, Inc.). Figure 6.3 is an example of the white light spectrum.

As discussed in the section 2.4, we observed up to 1% intensity drift. We believe that the wavelength drifts are primarily caused by the laser. As a way to monitor the power drift of the laser, we integrated a photodiode and beam splitter optics which extracts a small amount of laser light. This device can be utilized to correct a drift of Raman spectrum intensity.

With the useful information using the newly integrated components, we expect more accurate analysis in future transcutaneous measurements.

6.3 Future directions

A more comfortable and robust holder is needed to stabilize the arm as measurements are being made. One of 4 excluded volunteers showed an unacceptable level of fluctuations in raw spectra due to movement of the arm. Although we placed bands to hold the arm tightly, (but not too tightly to interrupt normal blood flow), it was sometimes difficult to keep the arm fixed without movement due to unconscious motion. A new design for this sample mount is being discussed.

A better sense for the glucose Raman signal intensity that we can expect from human skin requires more studies on the structure of human skin. We estimated the physiological level of glucose in the human body based on the weight fraction of water, assuming the tissue matrix is a homogeneous structure. However the human skin is a multi-layer matrix with different distribution of glucose content. If we build a Monte Carlo simulation code for human skin, it will give us better understanding in theory.

One of issues that degrades the effective analysis of PLS is the monotonically decreasing Raman signal over the experiment period. This seems to be related to a 20% decrease in the background signal (primarily by fluorescence photobleaching) in the highest peak during a 2 hour measurement time. A possible cause might be that the fifth order polynomial subtraction routine did not remove the fluorescence background properly. After the background subtraction, a 7% change of Raman signal intensity in the highest peak around 1450 cm^{-1} was observed (Figure 5.7). Needless to say, this disturbs the glucose prediction. We have been investigating this issue to search for a more effective process to extract a fluorescence-background-free Raman signal.

In order to support a study on this issue, we are searching for a stable and reproducible material simulating human skin, not only in its optical properties but also relative to fluorescence photobleaching. Once we have a material satisfying those properties, it will help us determine the cause of this problem.

A tissue phantom was a candidate, but it failed in terms of the photobleaching property. We have tried to use a collagen gel sample, since we noticed that the main Raman bands came from the collagen in skin. Recently we planned to use a porcine skin sample that has been studied once as a preliminary experiment. Although we did not use the data in this thesis, we learned that the porcine skin sample did not only have the same spectral features as human skin, but also similar absolute signals. We hope that this pigskin model could be utilized to simulate the human skin, helping us to address the photobleaching and the background removal method issues.

When these issues have been addressed, we plan to conduct another study with human volunteers. In the first experiment for the non-invasive glucose measurement, we worked with 20 healthy human subjects, each monitored over only a 2.5 hour period. True demonstrations of glucose measuring performance will require independent data that is data measured on different days than the calibration data. This implies testing on the same volunteers on multiple days. As was stated in section 6.1, a better calibration can be expected by using volunteers with diabetes. In order to access diabetic volunteers over multiple days more easily, we may decide to move our system to a hospital, redesigning the current system in a smaller size to facilitate transportation and ease of placement in a hospital setting.

References

Amos A, McCarty D and Zimmet P (1997). "The rising global burden of diabetes and its complications: estimates and projections to the year 2010." Diabetic Med. **14**: S1-S85.

Andrew JB, Koo TW and Itzkan I (1999). "Multicomponent blood analysis by near-infrared Raman spectroscopy." Applied Optics **38**(13): 2916-2926.

ANSI.Z136-1 (2000). Safe Use of Lasers, American National Standards Institute.

ANSI.Z136-3 (1996). Safe Use of Lasers in Health Care Facilities, , American National Standards Institute.

Bakker Schut TC, Withes MJH, Sterenborg HJCM, Speelman OC, Roodenburg JLN, Marple ET, Bruining HA and Puppels GJ (2000). "In Vivo Detection of Dysplastic Tissue by Raman Spectorscopy." Anal Chem **72**(24): 6010-6018.

Baraga JJ, Feld MS and Rava RP (1992). "Rapid Near-infrared Raman Spectroscopy of Human Tissue with a Spectrograph and CCD Detector." Appl. Spectrosc. **46**(2): 187-190.

Berger AJ and Feld MS (1997). "Analytical method of estimating chemometric prediction error." Applied spectroscopy, **51**(5): 725-732.

Berger AJ (1998). "Measurement of analytes in human serum and whole blood samples by near-infrared Raman spectroscopy." Ph.D. Thesis. Cambridge, Massachusetts Institute of Technology.

Berkowitz BA, Garner MH, Wilson CA and Corbett RJT (1995). "Non-destructive measurement of retinal glucose transport and consumption in vivo using NMR spectroscopy." Journal of Neurochemistry **64**(5).

Brennan JF, Wang Y, Dasari RR and Feld MS (1997). "Near-Ingrared Raman Spectrometer Systems for Human Tissue Studies." Appl Spectrosc **51**(2): 201-208

Bushman HP, Marple ET, Wach ML, Bennett B, Schut TCB, Bruining HA, Bruschke AV, van der Laarse A and Puppels GJ (2000). "In vivo determination of the molecular composition of artery wall by intravascular Raman spectroscopy." Anal Chem **72**(16): 3771-3775.

Caspers PJ, Lucassen GW, Bruining HA and Puppels GJ (2000). "Automated depth-scanning confocal Raman microspectrometer for rapid in vivo determination of water concentration profiles in human skin." J Raman Spectrosc **31**:813-818.

Clarke WL, Cox D, Gonder-Grederick LA, Carter W and Pohl SL (1987). "Evaluating clinical accuracy of systems for self-monitoring of blood glucose." Diabetes Care **10**: 622-628.

Cox DJ, Gonder-Grederick LA, Kovatchec BP, Julian DM and Clarke WL (1997). "Understanding error grid analysis." Diabetes Care **20**:911-912.

Enejder AMK, Koo TW, Oh J, Hunter M, Sasic S, Horowitz GL and Feld MS (2002). "Blood analysis by Raman spectroscopy." Optics Letters **27**(22): 2004-2006.

Ermakov IV, McClane RW, Gellermann W and Bernstein PS (2001). "Resonant Raman deterction of macular pigment levels in the living human retina." Opt Lett **26**(4): 202-204.

Feng GM, Huang YC, Chang GJ, Chang M and Chan C (1997). "A true phase sensitive optical heterodyne polarimeter on glucose concentration measurement." Optics Communications **41**(5-6).

Geladi P and Kowalski BR (1986). "Partial least-squares regression: A tutorial." Analytica Chimica Acta, **185**: 1-17.

Guyton AC, Hall JE, Schmitt W (1997). Human physiology and mechanisms of disease (Philadelphia; WB Saunders company).

Haaland DM and Thomas EV (1988). "Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extracation of qualitative information." Analytical Chemistry, **60**: 1193-1202.

Hata TR, Scholz TA, Ermakov IW, McClane RW, Khachik F, Gellermann W and Pershing LK (2000). "Non-Invasive Raman Spectroscopic Detection of Carotenoids in Human Skin." J Invest Derm **115**: 441-448.

Hill W and Rogalla D (1992). "Spike-correction of weak signals from charge-coupled devices and its application to Raman spectroscopy." Analytical Chemistry **62**: 2575-2579.

Jensen BM, Bjerring P, Christiansen JS and Orskov H (1995). Scand J Clin Lab Invest **55**, 427-432.

Koo T-W (2001). "Measurement of blood analytes in turbid biological tissue using near-infrared Raman spectroscopy." Ph.D. Thesis. Cambridge, Massachusetts Institute of Technology.

Manoharan R, Baraga JJ, Feld MS and Rava RP (1992). "Quantitative Histochemical Analysis of Human Artery Using Raman-Spectroscopy." Journal of Photochemistry and Photobiology B-Biology, **16**(2): 211-233.

Motz JT (2003), "Development o In Vivo Raman Spectroscopy of Atherosclerosis." Doctoral thesis, Massachusetts Institute of Technology, 2003.

Pincus MR (1996a). "Assessment of liver function." In Clinical diagnosis and management by laboratory methods." Henry JB ed. (Philadelphia: W. B. Saunders Company.

Quan KM, Christison GB, MacKenzie HA and Hodgson D (1993). "Glucose determination by pulsed photoacoustic technique: an experimental study using a gelatin-based tissue phantom." Physics in Medicine & Biology, **38**(12).

Roe, JN and Smoller BR (1998). Critical Reviews in Therapeutic Drug Carrier Systems 15, 199-241.

Roper Scientific GmbH Germany at http://www.roperscientific.de/theory.html.

Sharma A and Quantrill N (1994). "Measurement of glucose using fluorescence quenching." Spectrochimica Acta. Part A: Molecular Spectroscopy **50A**(6).

Shim M and Wilson B (1997). "Development of an In Vivo Raman Spectorscopic System for Dia Gnostic Applications." J Raman Spectrosc **28**: 131-142.

Shim MG, Song L-MWK, Marcon NE and Wilson BC (2000). "In Vivo Near-infrared Raman Spectroscopy: Demonstration of Feasibility During Clinical Gastrointestinal Endoscopy." Photochem Photobiol **72**: 146-150.

Takeuchi H, Hashimoto S and Harada I (1993). Applied Spectroscopy **47**: 129.

Tarr RV and Steffes PG (1990). "Non-invasive blood glucose measurement system and method using stimulated Raman spectroscopy." United States Paten 5,243,983.

Taylor MA, Joones MN, Vadgama PM and Higson SPJ (1995). "The characterization of liposomal glucose oxidase electrodes for the measurement of glucose." Biosensors & Bioelectronics **10**(3/4).

Thomas EV and Haaland DM (1990). "Comparison of multivariate calibration methods for quantitative spectral analysis." Analytical Chemistry, **60**: 1091-1099.

Troy TL, Page DL and Sevick-Muraca EM (1996). "Optical properties of normal and diseased breast skin." J. Biomed. Opt. **1**, 342-55.

Utzinger U, Heintzelman DL, Mahadevan-Janson A, Malpica A, Follen M and Richards-Kortum R (2001). "Near-Infrared Raman Spectroscopy for in vivo Detection of Cervical Precancers." Appl Spectrosc **55**(8):955-959.

Widdowson, EM and Dickerson JWT (1960). Boichem J **77**, 30-43.

Zhao J (2003). "Image Curvature Correction and Cosmic Removal for High-Throughput Dispersive Raman Spectroscopy." Applied Spectroscopy 57:1368-1375.

Zimmet P, Alberti KGMM and Shaw J (2001). "Global and societal implications of the diabetes epidemic." Nature **414**: 782-787.