A shifted-excitation Raman difference spectroscopy (SERDS) evaluation strategy for the efficient isolation of Raman spectra from extreme fluorescence interference

Medhanie Tesfay Gebrekidan, a,b Christian Knipfer, b,c Florian Stelzle, b,c Juergen Popp, d,e Stefan Will a,b and Andreas Braeuer a,b*

A biochemical characterization of pathologies in biological tissue can be provided by Raman spectroscopy. Often, the raw spectrum is severely affected by fluorescence interference. We report and compare various spectra-processing approaches required for the purification of Raman spectra from heavily fluorescence-interfered raw spectra according to the shifted-excitation Raman difference spectroscopy method. These approaches cover the entire spectra-processing chain from the raw spectra to the purified Raman spectra. In detail, we compared (1) area normalization versus z-score normalization, (2) direct reconstruction of the difference spectra versus reconstruction of zero-centered difference spectra and (3) collective baseline correction of the reconstructed spectra versus piecewise baseline correction of the reconstructed spectra and, finally, (4) analyzed the influence of the shift of the excitation wavelength on the quality of the reconstructed spectra. Statistical analysis of the spectra showed that – in our experiments – the best results were obtained for the z-score normalization before subtraction of the normalized spectra, followed by zero-centering of the difference spectra before reconstruction and a piecewise baseline correction of the pure Raman spectra. With our equipment, a wavelength shift from 784 to 785 nm provided reconstructed spectra of best quality. The analyzed specimens were different tissue types of pigs, tissue from the oral cavity of humans and a model solution of dye dissolved in ethanol.

**Keywords:** shifted-excitation Raman difference spectroscopy; fluorescence; fluorescence rejection; reconstruction of Raman spectra

### Introduction

Raman signals provide spectroscopic information that is molecule-specific and unique to the nature of the specimen under investigation. [1–6] Therefore, in recent years, the application of Raman spectroscopy has been markedly increasing in various fields. [7–10] As biological materials like proteins, carbohydrates, lipids, nucleic acids and deoxyribonucleic acid (DNA) feature different molecular structures, [11] their Raman spectra are also different. Thus, various biological tissues can be identified and differentiated from each other based on their Raman spectrum. [12] When any physiological change or pathological process changes the native chemical composition of a specimen can be identified or quantified experimentally if the Raman process is excited in the (near) infrared spectral region.

The acquisition of high-quality Raman spectra from which the chemical composition of a specimen can be identified or quantified remains one of the main challenges in applied Raman spectroscopy, especially when rather short measurement times (in the order of seconds or shorter) have to be realized. This is due to the low Raman scattering cross section: The small probability that an incident photon is scattered according to the Raman effect from a molecule is orders of magnitude smaller than the cross sections of potentially interfering absorption-emission processes, such as laser-induced fluorescence. As the strong broadband fluorescence appears in the same spectral region as the weak Raman signals, fluorescence interference may cause a background in the spectrum, which in the worst case, biological tissues inherit a rather high fluorescence. This may swallow up the weak Raman signals in the shot noise of the fluorescence signal.

Usually, fluorescence interferences can be avoided experimentally if the Raman process is excited in the (near) infrared spectral region. However, extreme fluorescence interference makes it necessary to develop strategies to efficiently isolate Raman spectra from heavy fluorescence interference.

The current work reports and compares various spectra-processing approaches required for the purification of Raman spectra from heavily fluorescence-interfered raw spectra according to the shifted-excitation Raman difference spectroscopy method. These approaches cover the entire spectra-processing chain from the raw spectra to the purified Raman spectra. In detail, we compared (1) area normalization versus z-score normalization, (2) direct reconstruction of the difference spectra versus reconstruction of zero-centered difference spectra and (3) collective baseline correction of the reconstructed spectra versus piecewise baseline correction of the reconstructed spectra and, finally, (4) analyzed the influence of the shift of the excitation wavelength on the quality of the reconstructed spectra. Statistical analysis of the spectra showed that – in our experiments – the best results were obtained for the z-score normalization before subtraction of the normalized spectra, followed by zero-centering of the difference spectra before reconstruction and a piecewise baseline correction of the pure Raman spectra. With our equipment, a wavelength shift from 784 to 785 nm provided reconstructed spectra of best quality.
region, for example, with a typical Nd-doped solid-state laser-emission wavelength around 1064 nm. Unfortunately, this is not applicable if short measurement times have to be realized, because of the poor quantum efficiencies of detectors in the (near) infrared spectral region and their high thermal noise. Consequently, the Raman process has to be excited in the visible or the ultraviolet (UV) spectral range. As UV radiation is harmful to most of the biological materials, one has to live with the fact that the same biological materials tend to fluoresce when excited in the visible spectral region. Excitation in the red spectral region, such as at approximately 785 nm, has been found optimal for characterization of biological tissues because of the relatively low fluorescence background and a still acceptable quantum efficiency of silicon based CCD detectors.[13] Nevertheless, the undesired fluorescence background still interferes with the desired Raman signals. Consequently, the purification of the desired Raman signals from the acquired spectra containing Raman and fluorescence signals is one essential spectrum-processing step before the purified Raman spectrum can be interpreted reliably.

This work is mainly motivated by our observation that in the many papers on fluorescence rejection reported in literature, the mathematical processing strategies were not described in a sufficiently comprehensive way to make a straightforward transfer to our spectra possible. This is why we here present a detailed description, comparison and assessment of the pre-processing and post-processing steps we applied for the purification of Raman signals from heavily fluorescence-interfered spectra. In this study, we combined – as it is illustrated in Fig. 1 – the shifted-excitation Raman difference spectroscopy (SERDS) for physical fluorescence suppression with mathematical fluorescence suppression approaches and found that the combination of both approaches assured the most efficient fluorescence suppression and best reconstruction quality of the Raman spectra. A more detailed description of Fig. 1 will follow in the results section.

Figure 1. Overview of the processing steps required for the purification of the pure Raman spectrum from the interfering fluorescence.

784 nm excitation wavelength

785 nm excitation wavelength

Intensity

Wavelength/nm

Normalization
(area normalization vs z-score normalization) and Subtraction

Reconstruction
(zero centered reconstruction vs direct reconstruction)

Baseline correction
(Collective baseline correction vs piecewise baseline correction)

Purified Raman Spectrum

J. Raman Spectrosc. 2016, 47, 198–209 © 2015 The Authors. Journal of Raman Spectroscopy published by John Wiley & Sons Ltd. wileyonlinelibrary.com/journal/jrs
Therefore, we now continue by providing some information about methods for fluorescence background elimination in general and describing the working principle of the SERDS technique.

Purification of the Raman signals from fluorescence-interfered spectra

Various techniques based on Raman scattering such as near-infrared dispersive Raman,[16–19] the Fourier transform Raman,[17] surface-enhanced Raman[18,19] and UV resonance Raman spectroscopies have been applied to analyze the structure of various biological molecules. To combat the annoying fluorescence interference, many approaches based on instrumental, experimental and computational methods have also been followed.[20] These approaches may be coarsely classified as time-gating methods, shifted-excitation techniques, improved sample preparation and algorithm-based baseline correction methods.[21]

Algorithm-based baseline correction methods include frequency filtering using fast Fourier transformation,[22] wavelet transformation,[23,24] polynomial fitting[15,25–27] and baseline correction using asymmetric least squares (ALS).[28–30] The frequency filtering method involves frequency content separation based on the fact that the fluorescence background has slow frequency components, while the Raman signals have high frequency components.[29] This method can cause artifacts to the Raman spectra because of the similarity of the frequencies of the noise and the Raman signals. Wavelet transformation is a powerful tool in signal processing.[31] However, fluorescence rejection based on wavelet transformation highly depends on the decomposition method used and the shape of the background.[28] Thus, useful spectral information might become lost, and spectral distortions might occur. The fast and simple polynomial fitting is one of the most frequently used methods for fluorescence rejection.[32] In this method, a polynomial function of usually fourth to sixth order is fitted to the fluorescence background. Then the fitted function is subtracted from the raw spectrum to receive the purified Raman spectrum. As the fluorescence background varies from sample to sample, polynomial fitting depends on the spectral fitting range and the chosen polynomial order.[32] In order to improve the performance of polynomial fitting and to reduce its limitations, Lieber and Mahadevan-Jansen[33] proposed an iterative polynomial fitting (ModPoly) method. The raw spectrum is fitted with a polynomial function. Then the raw spectrum and the polynomial-fit spectrum are compared at each wavenumber. Those values of the polynomial-fit spectrum exceeding the respective values of the original spectrum are replaced with the values of the original spectrum, which results in a new spectrum. This new spectrum is again fitted with a polynomial function, and the now-derived polynomial-fit spectrum is again compared with the original spectrum. Again, those values of the polynomial-fit spectrum exceeding the respective values of the original spectrum are replaced with the values of the original spectrum and so on. This method works better than the single polynomial fitting in situations where the level of the Raman signals is in the order of or even exceeds the fluorescence background level.[32] Cao et al.[27] developed a method that automatically chooses a polynomial order to avoid user intervention, which is called adaptive minmax. It involves multiple fits with different polynomial orders and chooses the one that gives the minimum error. Another technique based on curve fitting (ALS) has been presented by Eiler and Boelens[30] which uses the Whittaker smoother[34] to obtain a slowly varying estimate of the fluorescence background. Cadusch et al.[20] developed an algorithm based on least squares and compared it with the automated polynomial fitting proposed by Lieber and Mahadevan-Jansen (ModPoly),[33] and an around ten times lower normalized error ratio was obtained for a different number of peaks. He et al.[28] improved the ALS[30] and obtained a 16 times improvement of the root mean square error. The polynomial methods mentioned earlier have in common that the polynomial-fit result is not influenced by any initial knowledge or guess of the evaluating person. Only the order of the polynomial has to be chosen ahead. Therefore, the polynomial methods provide general fit results, which are not influenced by the experimentalist’s assumptions or guesses.

Shifted-excitation Raman difference spectroscopy

The shifted-excitation wavelength technique for fluorescence background elimination is based on Kasha’s rule,[35] which states that the fluorescence emission is unaltered for a small change in the excitation photon energy, while the Raman spectrum shifts according to the excitation photon energy change. Therefore, the subtraction of two spectra from each other acquired with slightly different excitation wavelengths makes the elimination of the fluorescence background possible, while a Raman difference spectrum remains. This SERDS has been established as a useful tool for applying Raman spectroscopy to samples with strong fluorescence interference.[36–38] Shreve et al.[39] illustrated the simplicity and effectiveness of SERDS by analyzing Raman scattering from CHCl₃ in the presence of added fluorescent laser dye. They were able to observe Raman peaks that are three orders of magnitude weaker than the fluorescence background. Using 784.5- and 785.5-nm excitation wavelengths, Appiah et al.[40] demonstrated the effectiveness of SERDS in situations where intense fluorescence interference prevents the use of common Raman spectroscopy. They were able to show the effectiveness of SERDS as compared with polynomial methods by extracting the ethanol spectrum from dark Jamaican rum. However, the processed spectrum of ethanol was not perfectly centered on the baseline, which might be caused by incomplete fluorescence elimination. Volodin et al.[41] were able to detect methanol in red wine using SERDS and confirmed from the results that it more accurately eliminated the fluorescence than baseline fitting methods. Matousek et al.[42] also reported the capability of SERDS to eliminate both the fluorescence background and the systematic noise from spectra. Sowoidnich and Knoeffel[43] applied the SERDS method for the in situ classification of meat species and were able to use the fingerprint Raman spectra for differentiating among beef, pork, chicken and turkey meat. Martins et al.[16] also made in vitro and in vivo experiments on human tooth and skin tissues, respectively, using the SERDS technique. They described the SERDS method as a very systematic and reproducible way to eliminate undesired luminescence background. Moreover, Noack et al.[36] applied SERDS in combination with support vector regression for online analysis of algae production and showed its potential as a non-intrusive tool for online monitoring of biotechnological processes. It should be mentioned here that the fluorescence signal intensity relative to the Raman signal intensity can change during a SERDS experiment because of photo bleaching of the fluorescent compound. Photo bleaching effects also can alter the signature of the broadband fluorescence background, meaning that the fluorescence background in some spectral regions is altered differently than in others. In these cases, the complete mitigation of the fluorescence interference is not straightforward, and the SERDS technology alone might not be able to reveal pure Raman signals.
Material and methods

Samples

In a first set of experiments, *ex vivo* Raman measurements were conducted on resected sample tissues from four domestic pigs that averaged 6 months of age. These experiments were carried out to find out which spectral processing steps result in the best differentiability of various tissue types. The sagittal split heads from which the tissue was derived were provided from a local slaughterhouse. Spectroscopic measurements were conducted within a time frame of 6 h after preparation of the specimens to minimize protein degradation due to exsiccation. The absence of systemic or local diseases was ensured in the specimens. Five different tissue types were obtained from each pig head: fat, cortical bone, gland, mucosal and cancellous bone. Squares of approximately 3 × 3 cm were cut out from each tissue using surgical scissors and a water-cooled band saw. After preparation and rinsing with saline solution, all specimens were stored in non-transparent plastic boxes and cooled at 4 °C.

In a second set of experiments, *ex vivo* experiments were also conducted on resected formalin-fixed tissue from the oral cavity of humans that has a size of 1 × 2 cm approximately. These experiments were carried out to find out which excitation wavelength shift required for the realization of the SERDS technique results in the smallest loss of spectral information and in the best signal-to-noise ratio. Tissue samples were prepared by clinical professionals of the department of maxillofacial surgery from the University Hospital of Erlangen (Universitätsklinikum Erlangen). The study protocol on human oral tissue was approved by the Ethics Committee of the University Clinic Erlangen (Ref. no. Az. 243_12).

In a third set of experiments, measurements were made in pure ethanol and in a solution of cryptocyanine (dye) dissolved in ethanol (0.0245 mg cryptocyanine in 1 ml of ethanol). The cryptocyanine absorbs the excitation wavelengths in the near-infrared regime and as a consequence fluoresces. Therefore, the spectra acquired from the solution are highly interfered with fluorescence, while the spectra taken from pure ethanol yield the pure Raman spectrum of ethanol. The comparison of the Raman spectra taken directly from pure ethanol and the one reconstructed from the fluorescence-interfered spectra taken from the solution shows how the reconstruction processes manipulate the purified spectrum.

Experimental setup

Figure 2 shows the setup of the self-developed Raman sensor. A diode laser (Toptica DLpro) with a variable laser wavelength between 770 and 810 nm and linewidth of less than 500 kHz was used as the excitation light source.

The excitation beam is launched into a glass fiber, which guides the laser radiation to a Raman probe. Inside the Raman probe, a short pass filter (785 nm) suppresses wavelengths longer than 785 nm originating from fiber-light interactions when the excitation light passes through the glass fiber. The excitation laser beam is then reflected via a dichroic mirror, which is highly reflective for the excitation wavelength (we never set the excitation wavelength longer than 785 nm) but transparent for wavelengths longer than 785 nm. It is then focused through a lens onto the sample with a focal spot diameter of approximately 200 μm. A portion of the excited signals (this is mainly elastic light scattering signals, fluorescence and the desired Raman signals) is detected in back-scattering direction through the same lens. The red-shifted fluorescence and Raman signals pass the dichroic mirror towards another lens focusing them onto a detection glass fiber bundle guiding the signals...
from the Raman probe to the spectrometer (Ventana from Ocean Optics). The elastic light scattering signals are filtered out, first, by the dichroic mirror reflecting them towards the excitation glass fiber and, second, by a long pass filter mounted between the dichroic mirror and the signal focusing lens. The combination of a half-wave plate and a Glan-laser prism allowed for the adjustment of the laser excitation power, which we measured behind the Raman probe, to approximately 100 mW. The Glan-laser prism was installed to let the vertical polarization of the excitation laser beam pass.

The Ventana spectrometer analyzes the spectra between 800 and 940 nm, which corresponds to Raman shifts from 200 to 2000 cm\(^{-1}\). The spectral resolution is specified at 810 nm to be 10 cm\(^{-1}\). At 810 nm, a wavenumber difference of 10 cm\(^{-1}\) corresponds to approximately 0.6 nm. Therefore, signals with a wavelength difference of minimum 0.6 nm can be spectrally resolved as two different peaks. With 1024 pixels along the spectral axis of the detector, one pixel corresponds to approximately 0.137 nm (~2.44 cm\(^{-1}\)), which is below the spectral resolution.

Experimental procedure

From each piece of tissue, spectra were taken from 40 different locations with minimum 1-mm distance between them. This distance is more than four times the focal spot size. For the pig tissue, at each location, measurements were made with the Raman sensor using the excitation wavelength 784 and 785 nm. Based on these measurements, we compared different methods of processing of the raw data according to the SERDS evaluation method including normalization, zero-centering, reconstruction and baseline correction. For the tissues of the oral cavity of humans, we applied five different excitation wavelengths of 784, 784.1, 784.4, 784.5 and 785 nm in order to analyze the impact of excitation wavelength difference onto the quality of the reconstructed spectra. These combinations of excitation wavelengths result in excitation wavelength (wavenumber) shifts of 0.1 nm (~1.63 cm\(^{-1}\)), 0.3 nm (~4.88 cm\(^{-1}\)), 0.4 nm (~6.50 cm\(^{-1}\)), 0.5 nm (~8.13 cm\(^{-1}\)), 0.6 nm (~9.74 cm\(^{-1}\)), 0.9 nm (~14.62 cm\(^{-1}\)) and 1 nm (~16.25 cm\(^{-1}\)). From them, some are above and others are below the spectrometer’s specified spectral resolution of 10 cm\(^{-1}\). For each excitation wavelength, 50 single-spectra were recorded and saved, from which one mean spectrum was computed per excitation wavelength. The acquisition of one single-spectrum took 100 ms. As the time required for read-out of one single-spectrum is negligibly short, the acquisition of 50 single-spectra took approximately 5 s.

Results and discussion

Here, we provide a detailed description of how we pretreat raw spectra in order to obtain the best fluorescence rejection for our applications with the following SERDS technique. Additionally, we provide a detailed description of how we post-process the obtained reconstructed SERDS spectra in order to obtain even purer Raman spectra from the originally extremely fluorescence-interfered raw spectra. In this study, we introduce an improved spectral processing approach compared with the evaluation strategy that we applied in our previous study.\(^{[2]}\) Different processing possibilities are compared with each other. The obtained pure Raman spectra are finally analyzed for the differentiation of the different tissue types from the pigs based on a principal component analysis (PCA) method. It should be emphasized here that our findings are especially true for the spectra we obtained from the three measurement objects, which are the pig tissues, the tissues from the oral cavity of humans and the dye solutions. For other original spectra, maybe better fluorescence mitigation results could be obtained following other approaches. Nevertheless, the here presented description of the different approaches might be beneficial for attempts of fluorescence rejection of others.

Then, we discuss (with respect to the equipment used in this study) the impact of the shift of the excitation photon energy (shift of the excitation wavelength) onto the quality of the reconstructed spectra. As the long-term aim of our research activities is the provision of a Raman measurement technique for the indication of early stages of cancer in the oral cavity of humans, the corresponding investigations were made with resected tissue from the oral cavity of humans.

Finally, we compare the purest Raman spectra taken directly from pure ethanol with Raman spectra reconstructed from the fluorescence-interfered spectra taken from the cryptocyanine–ethanol solution. This comparison makes accessible how, especially, the mathematical reconstruction processes manipulate reconstructed spectra with respect to the spectral axis.

Figure 1 provides a first view over the different processing steps between the raw spectra and the pure Raman spectra. In the following, the single-processing steps are described in detail, considering two averaged raw Raman spectra excited with 784 and 785 nm, respectively. The influence of the excitation wavelength difference will be discussed afterwards.

Normalization and subtraction

Following Kasha’s rule, the fluorescence contribution contained in both raw spectra shown at the top of Fig. 1 should ideally be the same. Only the Raman contributions are supposed to be shifted according to the excitation photon energy shift. In reality, the fluorescence contributions are not identical but only similar in both spectra,\(^{[3]}\) eventually because of photo bleaching of a fluorescent compound. Therefore, the subtraction of the two spectra will not provide pure Raman difference information but will still contain fluorescence-difference artifacts (‘left-over fluorescence’). Thus, to bring the similarity of the two fluorescence contributions closer to the ideal situation, the raw spectra are normalized first. Two normalization approaches, area normalization and z-score normalization (standardization), are applied and compared with each other.

The basic idea of area normalization is to make the area or integral under the curve of the two spectra in a certain spectral range the same. The spectral intensity values at each spectral position of either of the two spectra are scaled by the area of the two spectra:

\[
S_{n1}(\lambda) = \frac{S_1(\lambda)}{A_1},
\]

\[
S_{n2}(\lambda) = \frac{S_2(\lambda)}{A_2}
\]

where \(S_{n1}(\lambda)\) and \(S_{n2}(\lambda)\) are the area-normalized (index n) spectra, \(S_1(\lambda)\) and \(S_2(\lambda)\) are the spectra excited at 784 nm and 785 nm, respectively, and \(A_1\) and \(A_2\) are the areas/integrals under their curves, respectively. After area normalization, the areas/integrals of the two area-normalized spectra are the same.
In the z-score normalization, the two spectra are converted into a common scale with an average of zero and standard deviation of one:

\[
S_{zn1}(\lambda) = \frac{S_1(\lambda) - \text{mean}(S_1(\lambda))}{\text{std}(S_1(\lambda))}
\]

and

\[
S_{zn2}(\lambda) = \frac{S_2(\lambda) - \text{mean}(S_2(\lambda))}{\text{std}(S_2(\lambda))}
\]

Here, the z-score normalized intensity value \(S_{zn1}(\lambda)\) (or \(S_{zn2}(\lambda)\)) at each spectral position is expressed with respect to both the mean intensity value of all spectral locations \(\text{mean}(S_1(\lambda))\) (or \(\text{mean}(S_2(\lambda))\)) and the standard deviation of the intensity values of all spectral locations \(\text{std}(S_1(\lambda))\) (or \(\text{std}(S_2(\lambda))\)). The mean and the standard deviation are computed from the signal intensities along the whole spectral range of each spectrum and are consequently single-valued. After z-score normalization, the areas/integrals of the two z-score normalized spectra are zero.

Figure 3(a) and (b) shows the two spectra once area-normalized and once z-score normalized, respectively. Depending on which normalization approach is used, the difference spectra (Fig. 3(c)) are also different. The difference spectrum obtained from the z-score normalized spectra is closer to the zero baseline than the difference spectrum computed from the area-normalized spectra. It will be shown later that the quality of a reconstructed Raman spectrum is the better the closer the difference spectrum is to the zero baseline in regions in which no Raman signals exist. Therefore, we preferentially apply the z-score normalization method. In cases in which the signatures (not the intensities) of the spectra of the fluorescence background are identical in both spectra, the area normalization method would provide difference spectra sitting on the baseline.

Furthermore, the reproducibility of a number of the difference spectra is significantly improved if the z-score normalization is applied. From 40 different sample points on one piece of pork's gland tissue, we collected averaged Raman spectra excited again at 784 and 785 nm. The 40 difference spectra obtained after the z-score normalization were reproduced with an average standard deviation of 0.053 along the spectral range from 600 to 2000 cm\(^{-1}\) (corresponds to the wavelength range from 823 to 930 nm in Fig. 3). However, the average standard deviation of the difference spectra obtained from the area-normalized spectra is 0.12, which is more than twice as much.

**Reconstruction**

The normalization and subtraction phases were performed in the wavelength regime. Then the subtraction of the spectra cancels out the fluorescence, and the Raman signals remain shifted. Thus, after subtraction, the remaining processing steps were accomplished in the wavenumber regime.

The difference spectrum is then considered for the reconstruction of a pure Raman spectrum, for which a straightforward linear mathematical method based on recurrence relation\(^{[42]}\) is implemented. Figure 4 illustrates the working principle of the method. At first, the intensity values of the to-be-reconstructed Raman spectrum (\(R(\nu)\)) in Fig. 4) are all initialized to zero.

Afterwards, the intensity of the reconstructed Raman spectrum \(R(\nu)\) at a specific wavenumber \(\nu\) is computed by adding the intensity value of the difference spectrum \(D(\nu)\) and the intensity value of the updated reconstructed Raman spectrum \(R(\nu)\) shifted by the excitation wavenumber shift \(\Delta\nu\), to the left (lower wavenumbers) from the specified wavenumber (Eqn (4)). This procedure is repeated.

Figure 3. Comparison of difference spectra (c) computed from z-score normalized (b) and area-normalized (a) raw spectra.
for all wavenumber positions on the spectral axis from the left to the right side of the spectrum:

\[ R(\nu_i) = D(\nu_i) + R(\nu_i - \Delta \nu) \]  

(4)

Figure 5 shows a difference spectrum \( D(\nu_i) \) and the corresponding reconstructed spectrum \( R(\nu_i) \).

Apparently, the fact that the difference spectrum is not centered at the zero baseline causes a broadband background in the reconstructed spectrum. Therefore, we recommend the application of ‘zero-centering approach’, which we show in Fig. 6, for the further minimization of the left-over fluorescence, before the reconstruction is carried out. In this ‘zero-centering approach’, the difference spectrum is center-fitted using an ALS fit.\(^{30}\) Here, the weights are updated iteratively in such a way that higher weights are assigned to small residuals and lower weights are assigned to large negative and positive residuals in the difference spectrum.\(^{34}\) This measure assures that Raman signatures, which would feature large negative or positive residuals to the center-fitted spectrum, are not eliminated. The resulting fit function (gray line in Fig. 6(a)) then passes through the center of the difference spectrum.

In a next step the zero-centered fit function (gray line in Fig. 6(a)) is subtracted from the difference spectrum (blue spectrum in Fig. 6(a)), which results in the zero-baseline centered difference spectrum (Fig. 6(b)). Finally, the spectrum, which is shown in Fig. 6(c), can be reconstructed from the zero-baseline centered difference spectrum.

When comparing the reconstructed spectra with (Fig. 6(c)) and without (Fig. 5(b)) zero-baseline centering it becomes obvious that neither of them perfectly sits on the zero baseline, but that the peak signatures, especially in the wavenumber region from 600 to 2000 cm\(^{-1}\), are much better resolved for the reconstruction from a zero-baseline centered difference spectrum. Therefore, we recommend the application of the zero-centering approach before the reconstruction is applied.

**Baseline correction**

Through the z-score normalization, the zero-centering and the reconstruction, the interfering fluorescence background has been significantly reduced. This improvement becomes clear when comparing the raw spectrum in Fig. 3(a) and the processed spectrum in Fig. 6(c). While in Fig. 3(a), one cannot identify any signatures of a Raman spectrum, in Fig. 6(c) (thanks to the SERDS method with previous z-score normalization and subsequent zero-centering), the signatures of the Raman signals are clearly visible. Now, after the mainly SERDS-based elimination of most of the interfering fluorescence, even a single polynomial fit can, in principle, eliminate the still remaining fluorescence background. In order to improve the performance of the baseline correction and in order to increase the reproducibility of Raman spectra, we applied a piecewise fitting approach based on ALS (piecewise ALS).

The reconstructed spectrum is divided automatically into sub-spectra. Neither the number of sub-spectra nor their actual spectral ranges are predefined as initial guesses. Only the maximum spectral range of one sub-spectrum is defined, which in our case was set to maximum 200 data points. As one can see in Fig. 7, the spectral regions are separated by local minima. Thus, the global minimum within the first 200 data points (this is a local minimum with respect to the entire spectral range) represents the border between the first and the second sub-spectra. Starting from this minimum, the next global minimum within the following 200 data points is identified as the border between the second and the third
sub-spectra. This strategy is repeated as long as the right margin of the spectrum is reached. In our case, four sub-spectra labeled as regions 1–4 in Fig. 7 were identified. As Raman signals never would appear as local minima, this measure assures that no Raman signals are eliminated by this method. Then ALS curve fitting is applied to fit the background of each sub-spectrum. The fit functions are finally concatenated to cover the whole spectrum and subtracted from the reconstructed spectrum. As it can be seen in Fig. 7, the subtraction gives a pure Raman spectrum of tissue.

![Figure 6](image-url)

Figure 6. (a) Difference spectrum (blue spectrum) and asymmetric least square fit (gray curve), (b) zero-centered difference spectrum (blue spectrum minus gray curve) and (c) corresponding reconstructed Raman spectrum.

![Figure 7](image-url)

Figure 7. Piecwis asymmetric least squares (piecewise ALS) baseline correction. Blue line is the reconstructed spectrum, the black dashed line is the piecewise ALS fit and the red spectrum is the purified Raman spectrum obtained by subtracting the piecewise ALS fit from the reconstructed spectrum.

The performance of piecewise ALS is compared with an iterative polynomial fitting (ModPoly) method, which was proposed by Lieber and Mahadevan-Jansen. Figure 8(a) and (b) shows 40 Raman spectra of gland tissue of a pig head – baseline corrected using piecewise ALS and ModPoly, respectively. The standard deviation of the Raman spectra is also plotted alongside as a black dashed line. With the piecewise ALS fitting, a better reproducibility of the spectra was obtained. The standard deviation of the Raman peak at around 1200 cm⁻¹ is 0.15 for the piecewise ALS fitting. It is 0.44 for the ModPoly. The similarity of the Raman spectra is improved by almost three times. Moreover, the standard deviation of the Raman peak between 800 and 1000 cm⁻¹ is decreased from 0.26 to 0.14, which is around a twofold improvement. The overall standard deviation is also improved by two times. The two small peaks between 1000 and 1200 cm⁻¹ are also resolved better for all the spectra in the case of the piecewise ALS baseline correction approach. Similarly, the three peaks at the rear left wavenumbers of

![Figure 8](image-url)

Figure 8. Comparison of polynomial fitting versus piecewise asymmetric least squares (ALS) fitting – applied to extract Raman spectra from 40 gland tissue sample points. Single-spectra are given in color; standard deviations from the evaluation of spectra by either method are shown as black dashed lines.
the spectral range are easier to identify as compared with the result of the ModPoly.

**Purified (processed) Raman spectra**

Combining the principle of the shifted-excitation Raman difference technique with baseline correction based on curve fitting and applying simple data processing regimes, pure Raman spectra from a highly fluorescence-interfered sample are obtained. Following, the Raman spectra are analyzed for the differentiation of tissues from pigs.

Principal component analysis (PCA) was performed on the Raman spectra in the spectral range between 600 and 2000 cm\(^{-1}\). PCA is a well-known statistical tool applied to reduce dimensionality of data sets consisting of interrelated variables while keeping the meaningful dimensions that account for the majority of their differences.\[^{[44]}\] It involves projection of the original data set onto new orthogonal axes (principal components), which contains large percentage of variance of the data to produce new data sets with reduced dimensionality.

The output of the analysis showed that the overall variance in the datasets is contained within a few principal components. Almost all the variance in the pigs’ spectra (98.2\%) is contained in the first ten principal components. The first principal component accounts for 68.5\% of the total variance in the Raman spectra. The rest of the 98.2\% variance is shared among the other nine principal components in a decreasing manner proportional to the eigenvalue to which they correspond. Figure 9(a) shows the result of PCA to help visualize the scattering of the Raman spectra of the different tissues and their reproducibility from one measurement to another. The PCA scores show that the scattering of the tissues is small and well separated from each other.

According to Fig. 9(a), the spectra of cortical bone and cancellous bone are well separated from the other tissues. However, the two tissues themselves overlap in Fig. 9(a). Their purified Raman spectra are rather similar (Fig. 9(b)). The Raman spectra of fat are more reproducible, compared with the other tissue types. This is most likely due to the low fluorescence of the raw spectrum of fat. Gland tissue showed the highest scattering as compared with the other tissues. This is due to the fact that the corresponding raw spectra suffered from the most intense fluorescence background. The PCA scores of gland tissue and mucosal tissue are also close to each other. The Raman spectrum of mucosal tissue has a double peak around 900 cm\(^{-1}\), while the gland tissue spectrum has only one strong peak (Fig. 9(b)). These spectra also show considerable differences at around 1200 cm\(^{-1}\).

Finally, we have to prove that the here-proposed pre-processing and post-processing mechanisms in combination with SERDS result in better results than state-of-the-art pure polynomial or pure SERDS techniques. On this purpose, we compare in Fig. 10 Raman spectra reconstructed from the raw spectrum taken from gland (Fig. 1 (top)) using three different approaches. Surprisingly, the

![Figure 9](https://example.com/figure9.png)

**Figure 9.** (a) Principal component analysis scores of the Raman spectra of five different pig tissues (fat, cortical bone, gland, cancellous bone and mucosal) taking from four different pigs along principal component 1 (PC1) and principal component 3 (PC3). (b) Purified single Raman spectra of five different pig tissues (fat, cortical bone, gland, cancellous bone and mucosal).
ModPoly approach can reveal few Raman signatures, although in the raw spectrum of the gland, no Raman peaks are recognizable at all. Nevertheless, the signal-to-noise ratio of the ModPoly reconstructed Raman spectrum is worse compared with the others. The spectrum reconstructed using the SERDS method after z-score normalization (this is the spectrum shown in Fig. 5(b)) also shows some of the Raman signatures, but most of them are still hidden under broad bands resulting from the reconstruction of a non-zero-centered difference spectrum. This reconstructed spectrum features a better signal-to-noise ratio than the ModPoly reconstructed one. The clearest Raman signatures can be found in the spectrum reconstructed according to our suggestions. This means that the SERDS subtraction is applied after z-score normalization, that the reconstruction is applied after zero-centering and, lastly, that a baseline correction using a piecewise ALS approach is applied.

**Influence of the shift of the excitation photon energy (shift of the excitation wavelength) onto the reconstructed Raman spectra**

As mentioned before, the long-term aim of our research activities is the development of a Raman technique for the indication of early stages of cancer in the oral cavity of humans. Therefore, the experiments relevant for this subchapter were performed using tissue from the oral cavity of humans instead of pig tissue. Our experimental results showed that setting the excitation wavelength shift to less than 0.5 nm resulted in a loss or distortion of Raman peaks. At this point, we want to recall that the spectral resolution of the spectrometer is specified to be 10 cm⁻¹ at 810 nm, which would correspond to a wavelength difference of approximately 0.6 nm. We reconstruct the Raman spectrum of oral tissue (mucosal) of humans for the aforementioned excitation wavelength shifts of 0.1, 0.3, 0.4, 0.5, 0.6, 0.9 and 1.0 nm. In order to keep Fig. 11 clear, normalized and reconstructed Raman spectra (‘normalized’ because of the normalization before subtraction) are shown, for example, wavelength shifts, only. The Raman peaks become more visible and stronger, if the wavelength shift increases. The reconstructed Raman signatures are assigned to molecular vibrations based on two references.[2,45]

The Raman peaks due to the C–C twisting (protein) at 646 cm⁻¹, ring breathing mode in DNA bases at 678 cm⁻¹ and DNA at 748 cm⁻¹ are not visible for an excitation wavelength shift Δλ = 0.1 nm. For Δλ = 0.4 nm, the peaks at 828 cm⁻¹ (stretch DNA) and 883 cm⁻¹ (νCH₂) are recognizable as compared with the peaks for Δλ = 0.1 nm. The peaks at 1078, 1270 and 1330 cm⁻¹, which are due to C–C or C–O stretch (lipid), Amide III (protein) and CH₂ twisting (lipid), respectively, are easily recognizable for Δλ = 0.5 and 1 nm but distorted for Δλ = 0.4 nm. The other peaks (1445, 1656 and 1747 cm⁻¹) are recognizable except for Δλ = 0.1 nm. The only difference is the intensity of the Raman signal; i.e it is stronger for Δλ = 1 nm. Thus, setting the excitation wavelength shift to a value of less than 0.5 nm resulted in a loss of information misclassification of the sample due to the removal and distortion of peaks during reconstruction. The reconstructed spectra for Δλ = 0.5 and 1 nm do not show any loss or distortion of peaks. All peaks are recognizable. The intensity of the signal is the only difference between the reconstructed spectra. However, the shoulder at 1270 cm⁻¹ is more clearly seen for Δλ = 1 nm. Moreover, the peaks at wavenumbers below 800 cm⁻¹ are resolved better for this shift. The results show that it is possible to extract a Raman spectrum from a highly fluorescence-interfered raw spectrum of biological tissue by applying an excitation wavelength shift of 0.5 nm (here, corresponding to a wavenumber shift of 8.15 cm⁻¹). However, to obtain better resolved spectra with an improved signal intensity and, as a result, in order to minimize misclassification error of tissues, in this study, it is recommended to use a higher excitation wavelength shift of around 1 nm (here, corresponding to wavenumber shift of 16.3 cm⁻¹).

**Influence of the mathematical spectral processing methods onto the spectral information**

The charm of the SERDS technique is that the fluorescence is eliminated because of physical approach. Therefore, the SERDS technique in contrast to purely mathematically based polynomial fluorescence rejection approaches should not affect the Raman features of the spectrum. In other words, mathematical polynomial rejection methods bear the risk of not only eliminating the interfering fluorescence but also eliminating or influencing Raman
signatures. Therefore, we carried out Raman measurements first in pure ethanol and then in a solution of cryptocyanine dissolved in ethanol. Cryptocyanine is a dye that we excite to fluoresce with the excitation laser, seeking to excite solely the Raman process. We used this dye because the spectral signature of its fluorescence imitates the fluorescence from biological tissue that we analyzed. Therefore, by the comparison of the pure Raman spectrum taken from pure ethanol with the Raman spectrum reconstructed from the dye solution, we assess whether and to what extent our proposed technique affects the Raman signatures. Figure 12 (top) shows that the raw spectrum taken from the dye solution is heavily interfered by fluorescence, much like the raw spectra of gland tissue (Fig. 1). Figure 12 (bottom) shows the Raman spectrum reconstructed from the raw spectrum shown in Fig. 12 (top) and the pure reference Raman spectrum taken from pure ethanol. For the reconstruction, we again applied first the z-score normalization, second the subtraction, third the zero-centering, fourth the reconstruction and fifth the piecewise ALS baseline correction. We used exactly the same evaluation code and parameters used before in the context of the measurements of the pig tissues and the tissues from the oral cavity of humans.

The comparison between the reconstructed and the pure Raman spectra yields that all spectral features identifiable in the pure Raman spectrum can also be found in the reconstructed spectrum. Also the peak positions of the ethanol Raman peaks are not shifted remarkably. Of course, the reconstructed Raman spectrum shows more Raman signatures, as this spectrum also contains the Raman signal of the rather complex molecule cryptocyanine. We have also seen the increase in the strength of these exact Raman signatures as the concentration of the dye increases in the mixture with ethanol. An overlap of Raman peaks of ethanol and cryptocyanine can also contribute to small spectral shifts of the peak maxima. Therefore, we cannot clearly state here whether the little shifts of the peak maxima are due to the interference of the mathematical methods with the spectral features or peak–peak interferences. We can state that the methodology proposed here does influence the reconstructed spectrum only marginally. This is demonstrated by the good agreement between the Raman signal peaks of ethanol in both the pure and the reconstructed spectra.

**Conclusion**

We compared different methods for the processing of spectra that are relevant for the purification of the Raman spectra in the context of the SERDS. The results show that it is possible to extract ‘pure’ Raman spectra from heavily fluorescence-interfered raw spectra of biological tissue. Based on the comparison of different data processing methods, we recommend to use a z-score normalization before the subtraction of the normalized spectra from each other. Furthermore, the difference spectrum should be zero-centered before reconstruction. The reconstructed spectrum is recommended to be brought to the baseline by a piecewise polynomial correction.

A shift of 1 nm between the two excitation wavelengths of 748 and 785 nm resulted in the highest quality reconstructed Raman spectrum. Throughout all post-processing steps, care has been taken not to eliminate Raman signals from the processed spectra.

**References**

[1] I. J. Bigio, S. G. Bown, *Cancer Biol. Ther.* **2004**, 3, 259.
[2] C. Knipfer, J. Motz, W. Adler, K. Brunner, M. T. Gebrekidan, R. Hankel, A. Agaimy, S. Will, A. Braeuer, F. W. Neukam, *Biomed. Opt. Express* 2014, 5, 3252.
[3] I. U. Rehman, Z. Movasaghi, S. Rehman, *Vibrational Spectroscopy for Tissue Analysis*, CRC Press, Boca Raton, 2012.
[4] K. Das, N. Stone, C. Kendall, C. Fowler, J. Christie-Brown, *Laser Med. Sci.* 2006, 21, 192.
[5] N. Stone, P. Matousek, *Cancer Res.* **2008**, 68, 4424.
[6] S. Olszynska-Janus, M. Gasior-Glogowska, K. Szymborska-Malek, M. Komorowksa, W. Witkiewicz, C. Peczowicz, S. Szotek, M. Kobielarz, *Acta Bioeng. Biomech.* **2012**, 14, 121.
[7] B. Schrader, B. Dippel, I. Erb, S. Keller, T. Löchte, H. Schulz, E. Tatsch, S. Wessel, *J. Mol. Struct.* 1999, 480, 21.
[8] M. P. Buric, K. P. Chen, J. Falk, S. D. Woodruff, *Appl. Opt.* **2008**, 47, 4255.
[9] F. Rossi, M. Paolantoni, C. Clementi, B. Doherty, C. Milliani, B. G. Brunetti, A. Sgamellotti, *J. Raman Spectrosc.* 2010, 41, 452.
[10] C.-C. Lin, M.-T. Kuo, H.-C. Chang, *J. Med. Biol. Eng.* **2010**, 30, 343.
[11] B. Singh, R. Gautam, S. Kumar, V. B. Kumar, U. Nongthomba, D. Nandi, G. Mukherjee, V. Santosh, K. Somasundaram, S. Umapathy, *Curr. Sci.* 2012, 102, 232.
[12] A. Lorincz, D. Haddad, R. Naik, V. Naik, A. Fung, A. Cao, P. Manda, A. Pandya, G. Auner, R. Rabah, *J. Pediatr. Surg.* 2004, 39, 953.
[13] A. S. Williams, L. Hui, H. Lui, D. I. McLean, S. Lam, H. Zeng, *Int. J. Cancer* 2003, 107, 1047.
[14] M. A. S. Martins, D. G. Ribeiro, A. E. Pereira dos Santos, A. A. Martin, A. Fontes, H. S. Martinho, *Biomed. Opt. Express* 2010, 1, 617.
[15] D. B. Chase, *J. Am. Chem. Soc.* **1986**, 108, 7485.
[16] J. Ross, X. Agrawal, *Vib. Spectrosc.* 2011, 57, 163.
[17] D. L. Jeanmaire, R. P. Van Duyne, *J. Electroanal. Chem. Interfacial Electrochem.* 1977, 84, 1.
[18] P. Cadusch, M. Hlaing, S. Wade, S. McArthur, P. Stoddart, *J. Raman Spectrosc.* 2013, 44, 1587.
[19] J. B. Cooper, M. Abdelkader, K. L. Wise, *Appl. Spectrosc.* **2013**, 67, 973.
[20] G. Schulze, A. Jaresk, M. M. Yu, A. Lin, R. F. Turner, M. W. Blades, *Appl. Spectrosc.* **2003**, 57, 545.
[21] R. M. Ramos, I. Ruisánchez, *J. Raman Spectrosc.* **2005**, 36, 848.
[22] V. Barclay, R. Bonner, I. Hamilton, *Anal. Chem.* 1997, 69, 78.
[23] M. Gao, G. Lewis, G. M. Turner, A. Soubret, V. Ntziazihristos, *Appl. Opt.* **2005**, 44, 5468.
[24] J. F. Brennan, Y. Wang, R. R. Dasari, M. S. Feld, *Appl. Spectrosc.* **1997**, 51, 201.
SERDS evaluation strategy for isolation of Raman spectra

[27] A. Cao, A. K. Pandya, G. K. Serhatkulu, R. E. Weber, H. Dai, J. S. Thakur, V. M. Naik, R. Naik, G. W. Auner, R. Rabah, J. Raman Spectrosc. 2007, 38, 1199.

[28] S. He, W. Zhang, L. Liu, Y. Huang, J. He, W. Xie, P. Wu, C. Du, Anal. Methods 2014, 6, 4402.

[29] J. Peng, S. Peng, A. Jiang, J. Wei, C. Li, J. Tan, Anal. Chim. Acta 2010, 682, 63.

[30] P. H. Eilers, H. F. Boelens, Leiden University Medical Centre report, 2005.

[31] Z. M. Zhang, S. Chen, Y. Z. Liang, Z. X. Liu, Q. M. Zhang, L. X. Ding, F. Ye, H. Zhou, J. Raman Spectrosc. 2010, 41, 659.

[32] J. Zhao, H. Lui, D. I. McLean, H. Zeng, Appl. Spectrosc. 2007, 61, 1225.

[33] C. A. Lieber, A. Mahadevan-Jansen, Appl. Spectrosc. 2003, 57, 1363.

[34] P. H. Eilers, Anal. Chem. 2003, 75, 3631.

[35] M. Kasha, Discuss. Faraday Soc. 1950, 9, 14.

[36] K. Noack, B. Eskofier, J. Kiefer, C. Dilk, G. Bilow, M. Schirmer, R. Buchholz, A. Leipertz, Analyst 2013, 138, 5639.

[37] J. Kiefer, J. Raman Spectrosc. 2014, 45, 980.

[38] J. Kiefer, K. Noack, A. Leipertz, Am. Pharm. Rev. 2013, 16, 32.

[39] A. P. Shreve, N. J. Cherepy, R. A. Mathies, Appl. Spectrosc. 1992, 46, 707.

[40] B. Appliah, S. Dolgy, V. Ban, E. Melnik, B. Volodin, Special Issues 2012, 27, 1.

[41] B. Volodin, S. Dolgy, V. S. Ban, D. Gracin, K. Jurač, L. Gracin, Proc. SPIE 2014, 89390Y.

[42] P. Matousek, M. Towrie, A. Parker, Appl. Spectrosc. 2005, 59, 848.

[43] K. Sowoidnich, H.-D. Kronfeldt, Appl. Phys. B: Lasers Opt. 2012, 108, 975.

[44] I. Jolliffe, Principal Component Analysis, Wiley Online Library, New York, 2002, pp. 2–9.

[45] Z. Movasaghi, S. Rehman, I. U. Rehman, Appl. Spectrosc. Rev. 2007, 42, 493.