Lasers in Surgery and Medicine 36:356–364 (2005)

Autofluorescence and Diffuse Reflectance Spectroscopy for Oral Oncology

Diana C.G. de Veld, MSc,[1,2] Marina Skurichina, PhD,[3] Max J.H. Witjes, DDS, MD, PhD,[1] Robert P.W. Duin, PhD,[3] Henricus J.C.M. Sterenborg, PhD,[2]* and Jan L.N. Roodenburg, DDS, PhD[1]

[1]*Department of Oral and Maxillofacial Surgery, Division of Oncology, University Hospital Groningen, The Netherlands*

[2]*Photodynamic Therapy and Optical Spectroscopy Programme, Department of Radiation Oncology, Erasmus MC, Rotterdam, The Netherlands*

[3]*Information and Communication Theory Group, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands*

Background and Objectives: Autofluorescence and diffuse reflectance spectroscopy have been used separately and combined for tissue diagnostics. Previously, we assessed the value of autofluorescence spectroscopy for the classification of oral (pre-)malignancies. In the present study, we want to determine the contributions of diffuse reflectance and autofluorescence spectroscopy to diagnostic performance.

Study Design/Materials and Methods: Autofluorescence and diffuse reflectance spectra were recorded from 172 oral lesions and 70 healthy volunteers. Autofluorescence spectra were corrected in first order for blood absorption effects using diffuse reflectance spectra. Principal Components Analysis (PCA) with various classifiers was applied to distinguish (1) cancer and (2) all lesions from healthy oral mucosa, and (3) dysplastic and malignant lesions from benign lesions. Autofluorescence and diffuse reflectance spectra were evaluated separately and combined.

Results: The classification of cancer versus healthy mucosa gave excellent results for diffuse reflectance as well as corrected autofluorescence (Receiver Operator Characteristic (ROC) areas up to 0.98). For both autofluorescence and diffuse reflectance spectra, the classification of lesions versus healthy mucosa was successful (ROC areas up to 0.90). However, the classification of benign and (pre-)malignant lesions was not successful for raw or corrected autofluorescence spectra (ROC areas <0.70). For diffuse reflectance spectra, the results were slightly better (ROC areas up to 0.77).

Conclusions: The results for plain and corrected autofluorescence as well as diffuse reflectance spectra were similar. The relevant information for distinguishing lesions from healthy oral mucosa is probably sufficiently contained in blood absorption and scattering information, as well as in corrected autofluorescence. However, neither type of information is capable of distinguishing benign from dysplastic and malignant lesions. Combining autofluorescence and reflectance only slightly improved the results. Lasers Surg. Med. 36:356–364, 2005. © 2005 Wiley-Liss, Inc.

Key words: autofluorescence spectroscopy; cancer detection; combined classifiers; oral cancer; reflectance spectroscopy

INTRODUCTION

Early detection of pre-malignant lesions and malignant tumors may reduce patient morbidity and mortality because treatment at a less invasive stage is more successful, and therefore is of great clinical importance [1,2]. Unfortunately, (pre-)malignant lesions of the oral mucosa frequently go by unnoticed. In high-risk groups, pre-malignant and malignant lesions are therefore often diagnosed in an advanced stage. Once the patient or dentist observes a lesion, it is generally unclear whether the lesion is benign or (pre-)malignant. Current clinical diagnosis procedure therefore includes a biopsy. However, determining the optimal, that is, most dysplastic, location for biopsy is difficult. This leads to repeated biopsies and to the risk of underdiagnosis, which delays the necessary treatment.

Autofluorescence and diffuse reflectance spectroscopy have been studied as non-invasive in vivo tools for the detection of (pre-)malignant tissue alterations [3–7]. Autofluorescence of tissues under excitation with light is produced by several endogenous fluorophores. These include fluorophores from tissue matrix molecules and intracellular molecules like collagen, elastin, keratin, and NADH. The presence of disease changes the concentration of these fluorophores, which makes autofluorescence spectroscopy sensitive to tissue alterations. Diffuse reflectance is the result of single and multiple backscattering of the white excitation light. Both autofluorescence and diffuse reflectance signals are affected by the light scattering and absorption properties of the tissue. Light scattering is affected by tissue morphology, such as nuclear size distribution, epithelial thickness, and collagen content, all of which can change with the presence of disease. For the

Contract grant sponsor: Dutch Cancer Society; Contract grant number: RUG-99-1869; Contract grant sponsor: Dutch Technology Foundation (STW); Contract grant number: RNN 5316.

*Correspondence to: Henricus J.C.M. Sterenborg, PhD, Photodynamic Therapy and Optical Spectroscopy Programme, Department of Radiation Oncology, Erasmus Medical Center, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands.
E-mail: h.j.c.m.sterenborg@erasmusmc.nl

Accepted 1 November 2004
Published online 26 April 2005 in Wiley InterScience (www.interscience.wiley.com).
DOI 10.1002/lsm.20122

relevant visible wavelength range, absorption is mainly due to oxy- and deoxyhemoglobin. Therefore, absorption is affected by changes in blood content and oxygenation, which are known to accompany the presence of disease due to altered tissue metabolism and, in some cases, neovascularization. In summary, diffuse reflectance and autofluorescence spectroscopy can be used separately for detecting tissue alterations based on their sensitivities to tissue scattering and absorption properties and—in the case of autofluorescence—fluorophore concentrations.

It has been suggested that the most relevant information is contained in the fluorophore concentrations, and that tissue scattering and absorption effects merely have a negative effect on the performance of tissue diagnostic algorithms, since they are assumed to be less specific for malignant transformation [8–10]. For this reason, attempts have been made to recover the so-called intrinsic autofluorescence from the recorded autofluorescence spectra by combining autofluorescence and diffuse reflectance signals. This approach is based upon the assumption that fluorescent and reflected photons experience similar scattering and absorption events, while travelling through the tissue.

In the present study, we will compare the diagnostic potential of autofluorescence spectroscopy, as has been described in our previous study [11], with the diagnostic potential of diffuse reflectance spectroscopy and a simplified form of intrinsic autofluorescence, respectively. For the extraction of our simplified form of intrinsic autofluorescence, which we will call "corrected autofluorescence," we apply a simple model of dividing autofluorescence spectra by diffuse reflectance to a variable power. We will apply various statistical methods including Principal Components Analysis (PCA) and calculate areas under the Receiver Operator Characteristic (ROC) curve for comparison of the methods. Using these methods, we will try to distinguish (pre-)malignant lesions from benign lesions, cancer from healthy mucosa, and lesions of any type from healthy mucosa. Furthermore, we will examine whether raw autofluorescence, corrected autofluorescence, and diffuse reflectance supply complementary information by combining them in the classification.

MATERIALS AND METHODS

Volunteer and Patient Population

Autofluorescence and diffuse reflectance spectra were collected from 70 healthy volunteers with no clinically observable lesions of the oral mucosa and from 155 patients with oral lesions after they had given their informed consent. The population included volunteers from the Department of Oral and Maxillofacial Surgery of the University Hospital of Groningen, as well as patients who had been referred to the same department because of an oral lesion. This study was approved by the Institutional Review Board of the University Hospital of Groningen.

Experiments

A visual inspection of the oral cavity was performed by an experienced dental hygienist. In the volunteer group, this was done to ensure that no oral lesions were present at the time of measurement. In the patient group, the dental hygienist located and described the lesions to be measured. If present, the volunteers and patients were asked to remove their dentures. All patients and volunteers rinsed their mouth during 1 minute with a 0.9% saline solution in order to minimize the influence of consumed food and beverages.

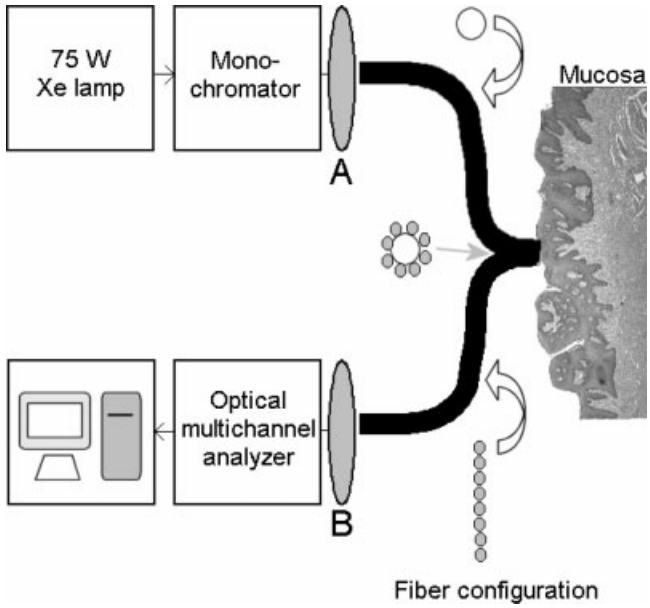

Fig. 1. Scheme of the experimental setup. For autofluorescence measurement, the monochromator picks an excitation wavelength, a 460-nm shortpass filter is placed at *A* and a 460-nm longpass filter at *B*. For diffuse reflectance measurements, the transmitting beam (zeroth order) of the monochromator is used, a 0.1% neutral density filter is placed at *A* and no filter is placed at *B*.

The measurement set-up (Fig. 1) consisted of a Xe-lamp with monochromator for illumination, a spectrograph, a custom-made set of 460-nm longpass and shortpass filters for autofluorescence measurements, and a neutral density filter for decreasing the white light excitation intensity during the diffuse reflectance measurements to prevent saturation of the CCD. Autofluorescence excitation wavelengths were 365, 385, 405, 420, 435 and 450 nm (bandwidth $\leq$15-nm full width half maximum). Using different filter sets for different excitation wavelengths would have extended the emission range, but this would have been very impractical because of the required manual filter replacements. However, since the emission spectra of the important tissue fluorophores are very broad, we expected to collect at least part of the relevant information [12,13]. For diffuse reflectance measurements, the transmitting beam (zeroth order) of the monochromator was used for excitation. The measurement probe was disinfected using 2% chlorhexidine digluconate in ethanol and covered with plastic film. The probe was placed in contact with the oral mucosa. The measurements were performed in a completely darkened room to prevent stray light from

entering the spectrograph. In our patient group, we measured four positions for each lesion: the center of the lesion, the border, the surrounding tissue, and the supposedly healthy tissue at the contralateral position. In the volunteer group, we measured 13 representative anatomical locations. The dental hygienist performed the measurements.

For each measured location and excitation wavelength, autofluorescence was recorded in three sequential measurements with a 1-second integration time. This allowed us to remove occasional spectra containing extremely high values for discrete pixels due to electronic noise. Diffuse reflectance was recorded in three sequential measurements of 0.25-second integration time for the same reason. On each measurement day, a set of calibration measurements was performed, including three sequential 0.25-second integration time diffuse reflectance measurements of spectralon. The diffuse reflectance spectra recorded in vivo were divided by the spectralon diffuse reflectance spectra to ratio out the Xenon arc lamp spectrum. A total of 48 diffuse reflectance spectra was missing in the dataset because of problems with the experimental equipment or because the anatomical location could not be measured in the patient.

In a previous study, in which we investigated the autofluorescence properties of 13 anatomical locations in the oral cavity, we concluded that oral mucosa can be divided into three categories with different spectroscopic characteristics [14]. These comprise (1) the dorsal side of the tongue, (2) the vermilion border of the lip, and (3) all other anatomical locations. In this study, we performed our data analysis within location group 3 only, since the other groups contained too few spectra for reliable analysis.

Sample Description

We measured diffuse reflectance and autofluorescence spectra in 70 healthy volunteers (37 men and 33 women) with a mean age of 50 years (range 18–85, standard deviation 15 years). Of the 70 volunteers, 4 were measured again on another occasion. Not in all volunteers could all 13 locations be measured, due to, for example, retching reflexes or problems with opening the mouth. Only the first measurement sessions were taken into account in the data analysis, and measurements at the dorsal side of the tongue or the vermilion border of the lip were omitted. We removed the measurement sessions for which either autofluorescence or diffuse reflectance spectra were missing due to problems with the experimental equipment. As a result, a total of 581 healthy oral mucosa spectra, distributed almost equally over the 11 anatomical locations, could be included in the data analysis.

Our patient population consisted of 155 persons (mean age 57, range 20–91, standard deviation 13 years). Some patients suffered from multiple lesions, so that a total of 172 unique lesions could be measured. Several lesions were measured for two or three times at different occasions for comparison, leading to a total of 199 lesion measurements sessions. Only first measurements on a specific lesion were included in the data analysis, resulting in 172 remaining measurement sessions. We removed 28 sessions because they were recorded from the dorsal side of the tongue or the vermilion border of the lip. A further 11 measurement sessions were left out of the analysis for different reasons: (1) because an accurate diagnosis was not available for the lesion at the time of measurement, (2) we were not certain enough that the probe had been located at the correct position because the lesion was hardly visible or very small, or (3) because the patient had already been receiving therapy [11]. We removed measurement sessions recorded from benign lesions of which the diagnosis was overly clear, such as aphtous lesions or lingua geographica [11]. Finally, we removed seven lesion measurement sessions for which either autofluorescence or reflectance spectra were missing due to problems with the experimental setup at the time of measurement. This resulted in a total of 115 lesion measurement sessions, of which 88 were benign, 11 dysplastic, and 16 cancerous. Because of the relatively small sample of dysplastic lesions, we decided not to separate these into mild, moderate, or severe dysplasias. We did not have permission to perform additional biopsies, therefore only those lesions for which the dental surgeon decided that a biopsy was necessary could be diagnosed by means of histopathology. However, all dysplastic and malignant lesions were histologically proven. Biopsies were always performed after the spectra had been acquired so as not to influence the spectra. If no histopathology was available, diagnosis was obtained by visual inspection by the dental surgeon. Details of lesions included in the data analysis are summarized in Table 1.

TABLE 1. Description of Lesions Included in the Data Analysis

Lesion type	Benign lesions	Dysplastic lesions	Malignant lesions
Oral leukoplakia (48)	40	8	—
Erosive leukoplakia (2)	1	1	—
Oral lichen planus (30)	30	—	—
(Non-specific) ulcus (5)	5	—	—
Candidiasis (10)	10	—	—
Erythroplakia (2)	—	2	—
Actinomycosis (1)	1	—	—
Scar (1)	1	—	—
Cancerous tumours (16)	—	—	16
Total	88	11	16

Data Processing

Autofluorescence spectra preprocessing. After manual removal of occasional spectra disturbed by electronic noise, the three or remaining two sequentially recorded autofluorescence spectra from each location and excitation wavelength were averaged. Detection wavelength calibration was performed for all spectra using a mercury–argon calibration lamp. Background spectra recorded from the same tissue site at the same excitation wavelength were subtracted. The spectral regions below 455 nm and above 867 nm were omitted since they contained no fluorescence. Spectra were averaged per four CCD pixels and eventually consisted of 199 sampling points.

Diffuse reflectance spectra preprocessing. After preprocessing of the recorded spectra in a similar way as for autofluorescence spectra, all diffuse reflectance spectra recorded in vivo were divided by the spectralon spectrum to ratio out the lamp spectrum. However, a Xenon lamp shows large peaks in the long wavelength part of the spectrum, which are unstable in time. Dividing the tissue spectrum by the spectralon spectrum therefore results in extreme noise in the near infrared spectral region, even though the maxima of the Xenon peaks maintain at the same wavelength. For this reason, we cut the diffuse reflectance spectra above 700 nm. Also, because the Xenon lamp has a very low intensity around 350 nm, hardly any diffuse reflectance is collected in this area. This also results in extreme noise when the tissue diffuse reflectance is divided by spectralon diffuse reflectance. We therefore skipped the part below 400 nm as well. After preprocessing, the spectra consisted of 556 points covering the 400–700 nm range.

Corrected autofluorescence preprocessing. A first order approximation of intrinsic autofluorescence spectra was obtained by dividing the autofluorescence spectra by diffuse reflectance spectra recorded at the same anatomical location to some variable power. We have applied the following model. For diffuse light, the total diffuse reflectance $R_d(\lambda)$, which is expressed as a fraction of the total incident flux, is related to the tissue absorption coefficient $\mu_a(\lambda)$ by

$$R_d(\lambda) = e^{-\mu_a(\lambda)<l>_{r,\lambda}}$$

where $<l>_{r,\lambda}$ is the expected mean pathlength traveled by the diffuse reflectance photons in the tissue [15]. All these variables are wavelength-dependent. The recorded autofluorescence F_r is affected in the same way:

$$F_r(\lambda) = e^{-\mu_a(\lambda)<l>_{f,\lambda}} \cdot F_i(\lambda)$$

where $F_i(\lambda)$ is the intrinsic fluorescence as evoked by the fluorophores, and $<l>_{f,\lambda}$ is the expected mean pathlength traveled by the fluorescence photons in the tissue [15]. In a simple model with a homogeneous distribution of scatterers and fluorophores, the absorption coefficient is constant at all locations and only depends on the wavelength of the photons.

For any given wavelength, there will be a difference in mean expected pathlengths for diffuse reflectance and fluorescence light. We can write $<l>_{f,\lambda} = k(\lambda) \cdot <l>_{r,\lambda}$, with $k(\lambda)$ depending on the tissue under investigation. Therefore,

$$F_i(\lambda) = \frac{F_r(\lambda)}{e^{-\mu_a(\lambda)<l>_{f,\lambda}}} = \frac{F_r(\lambda)}{e^{-\mu_a(\lambda)k(\lambda)<l>_{r,\lambda}}} = \frac{F_r(\lambda)}{R_d(\lambda)^{k(\lambda)}}$$

We now assume that $k(\lambda)$, the ratio between the pathlengths for fluorescence and diffuse reflectance light, is independent of wavelength. In reality, this is not exactly the case since the scattering properties of the tissue, which for a large part define the pathlength, are wavelength-dependent, too. However, based on the results that show little blood absorption dips and for our simple purpose, we are satisfied with this simplification. Our intrinsic autofluorescence spectra to a first approximation were therefore recovered by dividing the recorded fluorescence by the diffuse reflectance to a variable power k. The variable power was fitted in such a way that the blood absorption dips that appear around 545 and 575 nm in the corrected autofluorescence spectra were as small as possible. This was achieved by minimizing an approximation of the area of the 575 nm (second) blood absorption dip, which is represented by a triangle spanned by the spectral data points at wavelengths 563, 582, and 604 nm. These data points were selected empirically by visual inspection of the resulting spectra. For each wavelength point, three pixels were averaged. Empirically, we found that including the 545-nm blood absorption dip in the algorithm made the results worse, that is, the resulting spectra contained more blood absorption dips, probably because this 545-nm dip is less pronounced, so that noise in the spectrum starts playing a role. The same was true for trying to approximate the area with a polynomial. Probably, the spectral details of the blood absorption dip are more subject to noise than its more robust triangular representation.

Various refined models have been developed to recover intrinsic (auto)fluorescence spectra from recorded fluorescence spectra by means of a diffuse reflectance spectra [9,16–18]. These models are much more complete in taking into account scattering effects, probe geometries, etc. However, the models are also very complex and require tissue phantom measurements, which introduce other assumptions into the model. We therefore chose to use a simple approximation in this study. To our opinion, this was justified by the question that we wanted to answer: whether the removal of blood absorption artifacts would improve the classification. If this were the case, then removing most of the blood absorption artifacts up to a level at which they were not visible in the spectra would already imply an improvement in classification.

Statistical Analysis

Diffuse reflectance spectra. Diffuse reflectance spectra were normalized using three different normalization methods: normalization by the unit area (UA), Savitzky-Golay smoothing+derivative (SGS+D), and by the standard normal variate transformation (SNV). Afterwards, we performed PCA and applied Artificial Neural Networks.

For classification based on the PCs, we used the Karhunen–Loeve Linear Classifier (KLLC), also known as the regularized linear classifier assuming normal distributions, the Quadratic Classifier assuming normal distributions (QNC), the 1-nearest neighbor classifier (1-NN), and the Pseudo-Fisher Linear Discriminant (PFLD) based on normal distributions of data classes. These classifiers were applied to the retrieved first 5, 10, 20, or 30 PCs. We calculated ROC curve areas using Leave-One-Out (LOO) classification for all methods. In these curves, sensitivities for detection of lesions are plotted against corresponding values of (1-specificity). The more accurately a method separates the data classes, the closer the corresponding ROC-AUC (Area Under the ROC-Curve) approximates 1. We compared the areas under different ROC curves. This allowed us to make a fair judgment of the effectiveness of different methods without being constricted to single values of sensitivity and specificity, which largely depend on the threshold value chosen [19]. The calculations were repeated for the combined set of center and border measurements to investigate whether this affected the results.

Autofluorescence spectra. Autofluorescence spectra that had been corrected for blood absorption as well as raw autofluorescence spectra were classified in the same way as described above.

Complementarity of diffuse reflectance and (corrected) autofluorescence. Possibly, autofluorescence and diffuse reflectance spectra can supply supplementary information about the tissue properties. To test this hypothesis, we compared the scores for all individual lesions using different spectral information. We compared autofluorescence and corrected autofluorescence, both at the six excitation wavelengths using 10 PCs, and diffuse reflectance spectra using 10 or 30 PCs. For all these methods, we classified lesions versus healthy tissue using the KLLC and investigated the correlations between the predictions of the different classifiers.

Next, we applied six different combining rules to test whether the combining of autofluorescence, corrected autofluorescence and reflectance spectra could improve the results and thus whether they contain complementary information. The majority rule voting classifies the sample into the class that is chosen in majority by the three classifiers (for three types of spectra). A second combiner is obtained by training the nearest mean classifier on the three labels given by the separate classifiers. The remaining combiners were based on posterior class probabilities. Such probabilities are given by each classifier and represent its confidence that a sample belongs to a certain class. For example, a tissue measurement can yield a confidence of 0.9 of representing a lesion ($P_{lesion} = 0.9$), and a confidence of 0.1 of being healthy ($P_{healthy} = 0.1$). If only one classifier is used, then this measurement will, of course, be classified as a lesion ($P_{lesion} > P_{healthy}$). When three classifiers are used, different combining rules can be applied to the separate confidences. The product rule defines the total confidence of representing a lesion as the product of the confidences of representing a lesion for the three different types of spectra, and similarly for healthy tissue. The maximum rule defines the total confidence as the maximum of the confidences of the three classifiers, and the mean rule defines it as their mean value.

The fixed combining rules (like the product, the mean, and the maximum) ignore the distribution of the obtained posterior probabilities (confidences) that might be very informative. Therefore, besides the fixed combining rules, we have applied the trained combiner (in our case, the NMC) to posterior probabilities obtained by three classifiers constructed on three types of spectra.

RESULTS

General Description of the Data

Examples of autofluorescence, corrected autofluorescence, and diffuse reflectance spectra of a single lesion have been depicted in Figure 2. The shape of autofluorescence spectra in general has been described before [11]. Corrected autofluorescence appeared similar, but the prominent blood absorption dips were missing. Frequently, the intensity ratios for different locations changed after the correction, as can be seen in Figure 2. The average power k as fitted for optimal reduction of the blood absorption dips was 0.84 with a standard deviation of 0.32 for the healthy tissue dataset, and 0.84 ± 0.33 for the lesion set. This means that in our model, the expected mean pathlength of the fluorescence photons is approximately 84% of that of the reflectance photons. That it is shorter is consistent with reality, because fluorescence is evoked at a certain average depth in the tissue, while reflectance photons travel from probe through tissue to probe. Diffuse reflectance spectra showed clear blood absorption dips around 400–450, 540–550, and 570–580 nm. In general, the slope of the > 620 nm part of the diffuse reflectance spectra appeared flat and no further relevant spectral features could be distinguished. Much variance was observed between different spectra, especially in spectra recorded from the center or border of lesions. There were no significant differences in diffuse reflectance signal intensity between center, border and surroundings of lesions. However, center, border as well as surroundings of lesions showed significantly less diffuse reflectance intensity than contralaterally measured diffuse reflectance ($P < 0.001$). In about 89% of all cases, the contralateral position yielded higher fluorescence intensity than border, center, and surroundings of the lesion. We believe that this decrease in intensity is caused by blood absorption. Between the different types of lesions (benign, dysplastic, and malignant), no significant differences in diffuse reflectance intensity were observed for spectra recorded at the center or border of the lesion. Other spectral features than those related to blood absorption, in particular spectral slopes, did not show any trends in relationship to lesion type or probe location at the lesion.

Statistical Results

Autofluorescence spectra corrected by diffuse reflectance spectra. All statistical results are summarized in Table 2. Distinguishing cancerous lesions from

healthy tissue was very successful, resulting in ROC areas of up to 0.98, in this case corresponding with a sensitivity of 94% and a specificity of 94% (minimum Euclidian distance to the maximum value of 100% for sensitivity and specificity). For all methods, there seemed to be no clear dependence of the results on the excitation wavelength.

For the classification of lesions of any type versus healthy mucosa, the results were slightly less positive than for the classification of cancer versus healthy tissue. However, the best result still gave an ROC area of 0.90, in this case corresponding with a sensitivity of 83% and a specificity of 86%. The results did not seem to depend much on the excitation wavelength, however, 365 nm excitation was slightly more successful.

The results for the relevant clinical question, distinguishing benign from dysplastic and cancerous lesions, were all unsatisfying (ROC areas <0.70 for all excitation wavelengths and classification methods).

Diffuse reflectance spectra. Distinguishing malignant lesions from healthy mucosa gave good results. For all three normalization methods, the KLLC gave the best results with ROC areas of 0.88–0.93 for 20 or 30 PCs (best results: sensitivity 82%, specificity 88%). The classification of all lesions combined against healthy mucosa gave comparable results (best results: sensitivity 89%, specificity 80%). The results for distinguishing benign from dysplastic and cancerous lesions were slightly better than for autofluorescence corrected by diffuse reflectance. Most classifiers gave ROC areas <0.70, but some combinations of normalization method and classifier gave values that were higher (best results: ROC-AUC 0.77, sensitivity 69%, specificity 77%). This can be considered as reasonable but not good enough for clinical application.

Autofluorescence spectra. Our analysis of autofluorescence spectra that were not corrected for the influence of blood absorption has been described more extensively before [11]. The classification of malignant lesions versus healthy oral mucosa was very successful, resulting in ROC areas of 0.90–0.98 for different normalizations and classifiers. Separating all lesion types combined from healthy oral mucosa was best achieved by applying the KLLC, resulting in a maximum ROC area of 0.88 (sensitivity 89%, specificity 71%). The results for the relevant clinical question, that is to distinguish malignant and dysplastic lesions from benign lesions, were all unsatisfying (ROC areas <0.65).

Complementarity of Diffuse Reflectance and (Corrected) Autofluorescence

The results based on raw and corrected autofluorescence spectra all were highly correlated.We compared 12 auto-

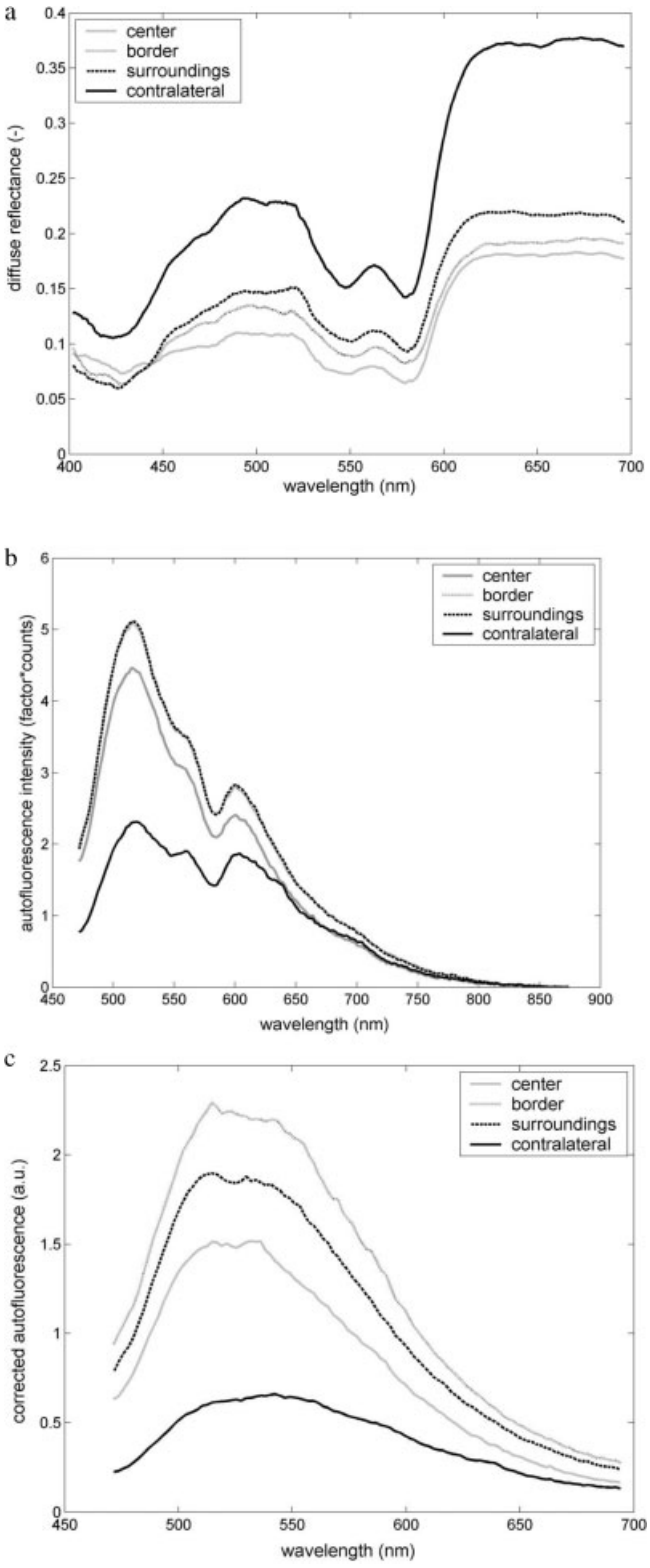

Fig. 2. Example spectra of one single lesion (**a**) diffuse reflectance, (**b**) autofluorescence, and (**c**) autofluorescence corrected for blood absorption. Autofluorescence excitation wavelength was 405 nm, the lesion was a benign lichen of the ventral side of the tongue. Please note the disappearing of blood absorption dips in the corrected autofluorescence spectra, and the differences in intensity in comparison with uncorrected autofluorescence spectra that were proved by this correction.

TABLE 2. Best Results in Terms of Areas Under the ROC Curve, Applied Classifier Noted Between Brackets

Type of spectra (normalization)	Corrected autofluorescence spectra	Diffuse reflectance spectra
Distinguishing malignant lesions from healthy oral mucosa		
Unit area normalization	0.83–0.96 (KLLC with 5, 10, 20, or 30 PCs)	0.90–0.93 (KLLC with 20 or 30 PCs)
Savitzky–Golay smoothing + derivative	0.89–0.97 (NN-1 with 5, 10, 20, or 30 PCs)	0.88–0.89 (KLLC with 20 or 30 PCs)
Standard normal variate transformation	0.89–0.98 (NN-1 with 10, 20, or 30 PCs)	0.89–0.92 (KLLC with 20 or 30 PCs)
Distinguishing all types of lesions from healthy oral mucosa		
Unit area normalization	0.83–0.88 (KLLC with 10, 20, or 30 PCs)	0.89–0.90 (KLLC with 20 or 30 PCs)
Savitzky–Golay smoothing + derivative	0.77–0.83 (NN-1 with 10, 20, or 30 PCs)	0.87–0.90 (KLLC with 20 or 30 PCs)
Standard normal variate transformation	0.82–0.90 (QNC with 10, 20, or 30 PCs)	0.88–0.89 (KLLC with 20 or 30 PCs)

fluorescence methods (6 excitation wavelengths, 2 types of autofluorescence spectra) pairwise, and found that, on the average, the pairs predicted the same classification (lesion or healthy) in 86% of the cases. The two types of diffuse reflectance spectra (using 10 or 30 PCs) agreed on lesion classification in 80% of the cases. However, the agreement in classification between the methods based on diffuse reflectance on one hand and those based on autofluorescence on the other, showed less agreement in lesion classification. On the average, autofluorescence and reflectance classifications agreed in 66% of the cases.

Next, we examined the possibilities of combining the three different methods to see if any classification improvement could be obtained, and thus to test whether the methods contain complementary information. The best result was obtained for the product rule (ROC area 0.90, sensitivity 88%, specificity 74%). This was slightly higher than the best result for uncorrected autofluorescence spectra at 365-nm excitation (ROC area 0.88, sensitivity 89%, specificity 71%). Combining only reflectance and corrected autofluorescence gave similar results (ROC area up to 0.89). Combining was also tested for the classification of malignant and dysplastic versus benign lesions. The overall results for the KLLC were slightly better (best result: ROC-AUC 0.73 for the product rule and the maximum rule, sensitivity 68%, specificity 70%) than for corrected autofluorescence and reflectance separately, but still too low for clinical application. We can conclude that combining autofluorescence and reflectance spectra has a positive influence on the results, but the improvement is small.

DISCUSSION

In agreement with our results for autofluorescence spectra that had not been corrected for blood absorption [11], the classification of cancerous lesions versus healthy oral mucosa was excellent for corrected autofluorescence as well as for diffuse reflectance spectra. The results for corrected and uncorrected autofluorescence spectra were almost equal (best results: ROC area 0.98 and 0.97, respectively). The results for diffuse reflectance spectra seemed to be slightly less positive (maximum ROC area 0.93). This can probably be explained by the loss of information that was contained in the porphyrin-like peak, which appears in fluorescence but not in diffuse reflectance spectra and is especially useful for classifying ulcerating tumors. However, these findings also mean that blood absorption and scattering effects are very efficient for distinguishing cancer from healthy mucosa.

While the influence of blood absorption has at least for a large part been reduced, the corrected autofluorescpectra give almost equal results as the raw autofluorescence spectra. This means that the fluorescence in itself is apparently also sufficient for distinguishing cancer from healthy mucosa. We can conclude that absorption and scattering effects as contained in reflectance spectra, as well as corrected autofluorescence characteristics, independently contain sufficient information for distinguishing cancer from healthy oral mucosa.

The separation of all lesion types combined from healthy oral mucosa was less successful but still gave good results, in agreement with our previous study on autofluorescence spectra that had not been corrected for blood absorption [11]. The results for diffuse reflectance, plain autofluorescence, and corrected autofluorescence were similar (best results: ROC area of 0.90, 0.88, and 0.90, respectively). Although the differences are very small, the uncorrected autofluorescence spectra gave a slightly worse result than the other two, which could perhaps implicate that separating autofluorescence from blood absorption effects simplifies the classification for our algorithms.

Three lesions were misclassified by means of (corrected) fluorescence and diffuse reflectance spectra for almost all classification methods. These lesions were studied in more detail. They turned out to be a (benign) lichen planus of the cheek, a benign leukoplakia of the cheek and a benign hyperplastic, hyperkeratotic lesion of the lateral border of the tongue. All these lesions showed higher fluorescence intensity than their surrounding and contralateral healthy tissue. We believe that this higher intensity was produced

by hyperkeratosis and that this produces the misclassifications. Probably, the algorithm relies for a large part on autofluorescence and diffuse reflectance intensity in combination with the relative depth of blood absorption dips. Similar effects have been reported in the lungs [20].

We have seen that combining classifiers based on autofluorescence and on reflectance spectra slightly improves the result. However, for the most relevant clinical question, the resulting sensitivities and specificities are still too low for clinical application.

In the literature, in vivo diffuse reflectance spectroscopy and autofluorescence spectroscopy have been applied for distinguishing between benign and (pre-)malignant tissue types in various organs. Georgakoudi et al. and others have proposed a method called trimodal spectroscopy, in which diffuse reflectance, autofluorescence and light scattering spectroscopy are used together for tissue diagnosis [8,13,21,22]. For this method, diffuse reflectance and autofluorescence spectra are measured, and two models are applied to extract intrinsic autofluorescence (IFS) and light scattering spectra (LSS). DRS, IFS, and LSS are then used in a majority rule to establish the diagnosis. This method has proven successful in various organs. Studies performed for Barrett's esophagus (16 patients) and the cervix (44 patients) gave higher sensitivities and specificities for the combined methods than for any of the methods separately. For a study using the same techniques in 8 volunteers and 15 patients with oral lesions, similar results were obtained when distinguishing lesions from healthy mucosa (sensitivity 96%, specificity 96%) [22]. These results were again higher than for the separate spectroscopic techniques. The distinction between cancerous and dysplastic tissue was made with a sensitivity of 64% and a specificity of 90%. In comparison with our results for the oral cavity, the results found by Muller et al. for distinguishing lesions from healthy mucosa are better. We believe that this is caused by the differences in patient population. Our lesion set was very diverse, which may complicate the classification. Unfortunately, the generally more difficult benign versus dysplastic/malignant classification was not performed by Muller et al. because of the small amount of benign lesions included.

Lin et al. compared in vivo autofluorescence and diffuse reflectance spectra from brain tumours and normal brain tissue [23]. A two-step algorithm based on fluorescence and diffuse reflectance intensity at 460 nm produced a sensitivity of 89% and specificity of 76% for distuinguishing tumor-bearing from normal brain tissue. Nordstrom et al. measured autofluorescence and diffuse reflectance spectra from the cervix and applied multivariate analysis [24]. Autofluorescence obtained high sensitivities (86–91%) and specificities (87–93%) for distinguishing cervical intraepithelial lesions from normal squamous tissue. However, metaplasia (benign) and cervical intraepithelial lesions could not be separated. For diffuse reflectance, the results for distinguishing CIN from normal tissue were lower, but the classification of metaplastic versus dysplastic lesions was more successful (sensitivity 77%, specificity 76%). This is a striking similarity with our own results, in which diffuse reflectance spectra were more successful for distinguishing different lesions types, too. No attempts were made to develop an algorithm that combined the two types of spectra.

Although other organs cannot truly be compared to the oral cavity, which has a complex anatomy and in which many diverse lesions occur, we can conclude that classification of lesions by means of diffuse reflectance and autofluorescence spectroscopy is fairly good. However, the classification of (pre-) malignant versus benign lesions is more difficult. These results are in agreement with our own.

Reflectance spectroscopy has also been studied on its own as a method for distinguishing different tissue types. Koenig et al. used diffuse reflectance spectroscopy to detect bladder carcinoma [25]. They found a sensitivity of 91% and a specificity of 60% for distinguishing nine malignant and two dysplastic lesions from six normal sites and nine benign lesions, using an algorithm that was based on the total amount of blood in the tissue. Ge et al. studied colonic dysplasia and neoplasia by means of diffuse reflectance spectroscopy in a large patient population [26]. They found predictive accuracies of 81–85% for distinguishing adenomatous (dysplastic) from hyperplastic (benign) polyps using different pattern recognition tools (sensitivity 89%, specificity 75%). Zonios et al. studied colonic polyps with reflectance spectroscopy as well [7]. The spectra were analyzed using an analytical light-diffusion model. Differences in hemoglobin concentration and effective scatterer size were observed between normal and adenomatous tissue sites, but no classification was attempted. Mirabal et al. have applied reflectance spectroscopy using variable source-detector separation distances for the detection of cervical neoplasia [27]. They found a sensitivity of 72% and a specificity of 81% for distinguishing squamous intraepithelial lesions from normal squamous mucosa, and similar for squamous intraepithelial lesions versus normal columnar mucosa.

From these studies, we can conclude that classification of (pre-)malignant versus benign lesions is performed with higher sensitivities and specificities than in the oral cavity, although the results are still insufficient for clinical application. From our studies and from the literature we can conclude that although lesions of the oral mucosa can be reliably distinguished from healthy mucosa, correct classification of lesion types is not possible. Therefore, the current setting in which fluorescence spectroscopy, reflectance spectroscopy, or combination of these techniques are used is not applicable in clinical use for the oral cavity.

ACKNOWLEDGMENTS

We are grateful to all volunteers and patients for their valuable contribution and to Mirjam Wouda, Irènke de Jong-Orosz, and Ada Schokkenbroek for performing the many measurements.

REFERENCES

1. Hyde N, Hopper C. Oral cancer: The importance of early referral. Practitioner 1999;243(1603):753, 756–761.

2. Silverman S. Early diagnosis of oral cancer. Cancer 1988;62(8 Suppl):1796–1799.
3. Majumder SK, Ghosh N, Kataria S, Gupta PK. Nonlinear pattern recognition for laser-induced fluorescence diagnosis of cancer. Lasers Surg Med 2003;33(1):48–56.
4. Mayinger B, Jordan M, Horner P, Gerlach C, Muehldorfer S, Bittorf BR, Matzel KE, Hohenberger W, Hahn EG, Guenther K. Endoscopic light-induced autofluorescence spectroscopy for the diagnosis of colorectal cancer and adenoma. J Photochem Photobiol B 2003;70(1):13–20.
5. Wang CY, Tsai T, Chen HM, Chen CT, Chiang CP. PLS-ANN based classification model for oral submucous fibrosis and oral carcinogenesis. Lasers Surg Med 2003;32(4):318–326.
6. Zellweger M, Grosjean P, Goujon D, Monnier P, van den BH, Wagnieres G. In vivo autofluorescence spectroscopy of human bronchial tissue to optimize the detection and imaging of early cancers. J Biomed Opt 2001;6(1):41–51.
7. Zonios G, Perelman L, Backman V, Manoharan R, Fitzmaurice M, Van Dam J, Feld MS. Diffuse reflectance spectroscopy of human adenomatous colon polyps in vivo. Appl Opt-OT 1999;38(31):6628–6637.
8. Badizadegan K, Backman V, Boone CW, Crum CP, Dasari RR, Georgakoudi I, Keefe K, Munger K, Shapshay SM, Sheets EE, Feld MS. Spectroscopic diagnosis and imaging of invisible pre-cancer. Faraday Discuss 2004;126:265–279.
9. Diamond KR, Farrell TJ, Patterson MS. Measurement of fluorophore concentrations and fluorescence quantum yield in tissue-simulating phantoms using three diffusion models of steady-state spatially resolved fluorescence. Phys Med Biol 2003;48(24):4135–4149.
10. Georgakoudi I, Jacobson BC, Muller MG, Sheets EE, Badizadegan K, Carr-Locke DL, Crum CP, Boone CW, Dasari RR, Van Dam J, Feld MS. NAD(P)H and collagen as in vivo quantitative fluorescent biomarkers of epithelial precancerous changes. Cancer Res 2002;62(3):682–687.
11. De Veld DC, Skurichina M, Witjes MJ, Duin RP, Sterenborg DJ, Star WM, Roodenburg JL. A clinical study for classification of benign, dysplastic and malignant oral lesions using autofluorescence spectroscopy. 2004;9(5):940–950.
12. Drezek R, Sokolov K, Utzinger U, Boiko I, Malpica A, Follen M, Richards-Kortum R. Understanding the contributions of NADH and collagen to cervical tissue fluorescence spectra: Modeling, measurements, and implications. J Biomed Opt 2001;6(4):385–396.
13. Georgakoudi I, Sheets EE, Muller MG, Backman V, Crum CP, Badizadegan K, Dasari RR, Feld MS. Trimodal spectroscopy for the detection and characterization of cervical precancers in vivo. Am J Obstet Gynecol 2002;186(3):374–382.
14. De Veld DC, Skurichina M, Witjes MJ, Duin RP, Sterenborg DJ, Star WM, Roodenburg JL. Autofluorescence characteristics of healthy oral mucosa at different anatomical sites. Lasers Surg Med 2003;32(5):367–376.
15. Wilson BC. Measurement of tissue optical properties: Methods and theories. In: Welch AJ, van Gemert MJC, editors. Optical-thermal response of laser irradiated tissue. New York: Plenum Press; 1995. pp 233–303.
16. Muller MG, Georgakoudi I, Zhang Q, Wu J, Feld MS. Intrinsic fluorescence spectroscopy in turbid media: Disentangling effects of scattering and absorption. Appl Opt-OT 2001;40(25):4633–4646.
17. Wu J, Feld MS, Rava RP. Analytical model for extracting intrinsic fluorescence in turbid media. Appl Optics 1993; 32(19):3585–3595.
18. Zhang Q, Iler MG, Wu J, Feld MS. Turbidity-free fluorescence spectroscopy of biological tissue. Opt Lett 2000;25(19): 1451–1453.
19. Metz CE. Basic principles of ROC analysis. Semin Nucl Med 1978;8(4):283–298.
20. Bard MLP, Amelink A, Skurichina M, den Bakker M, Burgers SA, Van Meerbeeck JP, Duin RP, Aerts JGJV, Hoogsteden HC, Sterenborg HJCM. Improving the specificity of fluorescence bronchoscopy for the analysis of neoplastic lesions of the bronchial tree by combination with optical spectroscopy: Preliminary communication. Lung Cancer 2005;47(1):41–47.
21. Georgakoudi I, Jacobson BC, Van Dam J, Backman V, Wallace MB, Muller MG, Zhang Q, Badizadegan K, Sun D, Thomas GA, Perelman LT, Feld MS. Fluorescence, reflectance, and light-scattering spectroscopy for evaluating dysplasia in patients with Barrett's esophagus. Gastroenterology 2001;120(7):1620–1629.
22. Muller MG, Valdez TA, Georgakoudi I, Backman V, Fuentes C, Kabani S, Laver N, Wang Z, Boone CW, Dasari RR, Shapshay SM, Feld MS. Spectroscopic detection and evaluation of morphologic and biochemical changes in early human oral carcinoma. Cancer 2003;97(7):1681–1692.
23. Lin WC, Toms SA, Johnson M, Jansen ED, Mahadevan-Jansen A. In vivo brain tumor demarcation using optical spectroscopy. Photochem Photobiol 2001;73(4):396–402.
24. Nordstrom RJ, Burke L, Niloff JM, Myrtle JF. Identification of cervical intraepithelial neoplasia (CIN) using UV-excited fluorescence and diffuse-reflectance tissue spectroscopy. Lasers Surg Med 2001;29(2):118–127.
25. Koenig F, Larne R, Enquist H, McGovern FJ, Schomacker KT, Kollias N, Deutsch TF. Spectroscopic measurement of diffuse reflectance for enhanced detection of bladder carcinoma. Urology 1998;51(2):342–345.
26. Ge ZF, Schomacker KT, Nishioka NS. Identification of colonic dysplasia and neoplasia by diffuse reflectance spectroscopy and pattern recognition techniques. Appl Spectrosc 1998; 52(6):833–839.
27. Mirabal YN, Chang SK, Atkinson EN, Malpica A, Follen M, Richards-Kortum R. Reflectance spectroscopy for in vivo detection of cervical precancer. J Biomed Opt 2002;7(4):587–594.