The classification of the patients with pulmonary diseases using breath air samples spectral analysis

Yury V. Kistenev' , Alexey V. Borisov, Dmitry A. Kuzmin' , and Anna A. Bulanova

Citation: **1760**, 020028 (2016); doi: 10.1063/1.4960247
View online: http://dx.doi.org/10.1063/1.4960247
View Table of Contents: http://aip.scitation.org/toc/apc/1760/1
Published by the American Institute of Physics

The Classification of the Patients with Pulmonary Diseases Using Breath Air Samples Spectral Analysis

Yury V. Kistenev[1,2,a)], Alexey V. Borisov[1,2], Dmitry A. Kuzmin[1,2,b)], and Anna A. Bulanova[2]

[1] *National Research Tomsk State University, Lenin Avenue 36, Tomsk, 634050 Russia*
[2] *Siberian State Medical University, Moscovsky Trakt 2, Tomsk, 634050 Russia*

[a)] Corresponding author: yuk@iao.ru
[b)] kuzmin.da@ssmu.ru

Abstract. Technique of exhaled breath sampling is discussed. The procedure of wavelength auto-calibration is proposed and tested. Comparison of the experimental data with the model absorption spectra of 5% CO_2 is conducted. The classification results of three study groups obtained by using support vector machine and principal component analysis methods are presented.

INTRODUCTION

The study of qualitative and quantitative composition of the exhaled breath samples (EBS) is an urgent task for design of non-invasive express diagnostics tools of a diseases [1–6]. This field is often referred to as exhaled metabolomics or “breathomics”. Laser absorption spectroscopy (LAS) methods are very attractive among the suitable for practical medicine technical approaches for routine measurements [3]. The ability of LAS equipment depends on tuning range of used laser source. It should be take into account that absorption vibration bands of a large number of volatile biomarkers of various diseases are in the IR region approximately between 2 and 10 μm [5]. So the realization of the potential of LAS in breathomics is associated with the creation of the laser radiation sources with ultra-wide tuning range.

We use the laser optical-acoustic gas analyzer “LaserBreeze”, developed by “Special Technologies” (Novosibirsk, Russia), which combines the method of optical-acoustic detection (OAD) of gas components and laser source, created on the basis of combination of two optical parametric oscillators (OPO) [7–9]. The threshold sensitivity of the registered volatile gaseous components concentration is up to 1 ppb, the necessary amount of EBS is about 50 ml. Usage of two OPO, working on different nonlinear crystals, provides wide tuning range of laser radiation from 2.5 to 10.5 μm.

Only OPO provides a wide wavelength range tuning of laser radiation at relatively low cost. This offers great opportunities for breathomics. However, to realize these opportunities, methodological aspects of similar equipment using should be considered.

TECHNIQUE OF EXHALED BREATH SAMPLING

Since the OAD cell has a sufficiently small volume, we used a plastic container (disposable syringe of 160 ml—below sampler) with an airtight plastic cap for the sampling and transportation of EBS. The following conditions of sampling were established to reduce the influence of external factors:

– sampling was carried out in the morning (07.30–09.30) before meals;
– for smokers sampling was 40 min after smoking;
– multiple oral water rinsing should be executed before sampling.

Physics of Cancer: Interdisciplinary Problems and Clinical Applications (PC’16)
AIP Conf. Proc. 1760, 020028-1–020028-5; doi: 10.1063/1.4960247
Published by AIP Publishing. 978-0-7354-1418-1/$30.00

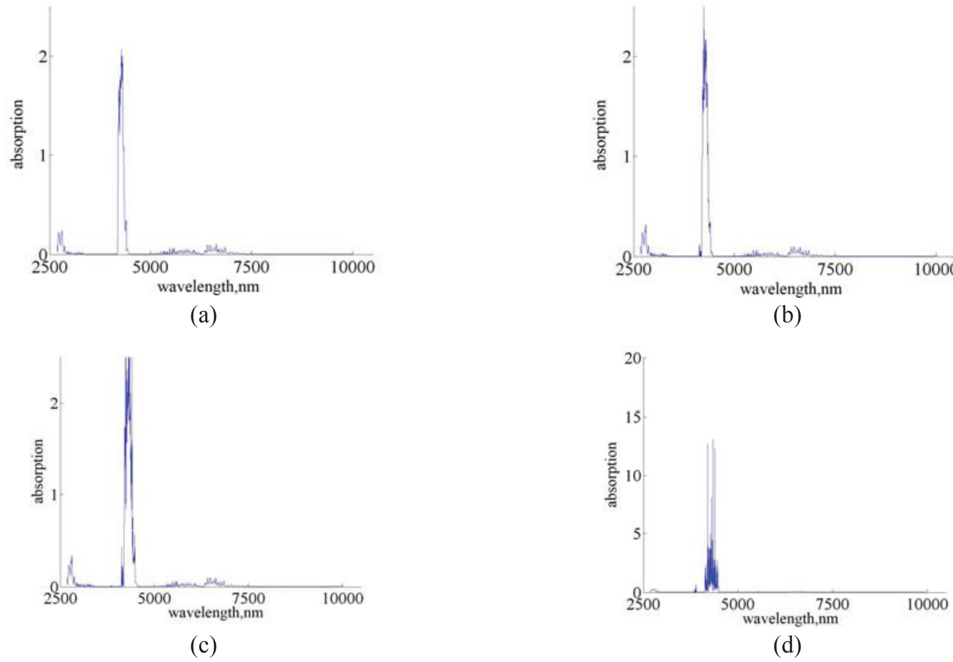

FIGURE 1. The absorption spectra of EBS at 60 (a), 120 (b), 180 (c) and 240 min (d) after sampling

Sampling proceduree was as follows: the participant made a calm exhalation, the initial portion of the air volume of 100–250 ml (so-called “dead volume”) in the surrounding area and rest part of exhalation blew in sampler. The container that supports the room temperature was used for transportation of samplers with EBS.

To evaluate the restriction of EBS time transportation, absorption spectra profiles of the same EBS were measured at 60 (Fig. 1a), 120 (Fig. 1b), 180 (Fig. 1c) and 240 min (Fig. 1d) after sampling.

As can be seen from the plots, the values of the absorption coefficients in the spectral region near 4000 nm significantly increased in comparison with the initial profile of absorption spectrum after 180 min (Fig. 1c). So, the “degradation” of the sample was irrelevant during 2 hours after sampling, which ensures the accuracy of the measurements of EBS absorption spectra.

WAVELENGTH AUTO-CALIBRATION

The laser source based on the OPO does not have the internal rapper of wavelength, as in the case, for example, of laser sources with discrete set of generation lines (e.g., gaseous laser). In the “LaserBreeze”, for these purposes hermetically soldered reference cell, which is filled with a gas mixture of known composition, is used.

This cell can be used to evaluate stability and reliability of the EBS absorption spectra measurements. Figure 2 shows the absorption spectra of the reference cell’s content, recorded at different times, and the template spectrum. All these spectra are identical, which indicates the reliability of the results.

Figure 3 shows a part of the two EBS absorption spectra. As can be seen, a wavelength shift appears in some areas of the spectra. The correction of the experimental spectra of the wavelength was implemented using the following algorithm. Let there be experimentally obtained by “LaserBreeze” absorption spectra $f(\lambda)$. Let us define the range of wavelengths $\{\lambda\}$ from λ_{min} to λ_{max} with step h. Let us associate absorption spectra in the range $\{\lambda\}$ with the spectra $f(\lambda)$ by the following rule:

$$f_c(\lambda_i) = \frac{1}{n} \sum_{\lambda_i < \lambda_j < \lambda_{i+1}} f(\lambda_j). \quad (1)$$

Here, n is the quantity of absorption coefficients within the interval $[\lambda_i, \lambda_{i+1}]$.

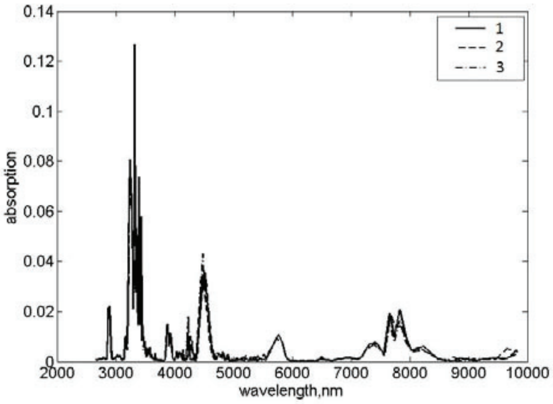

FIGURE 2. The absorption spectra of reference cell's content, recorded in different times (*1*, *2*) and theoretical profile of the spectrum (*3*)

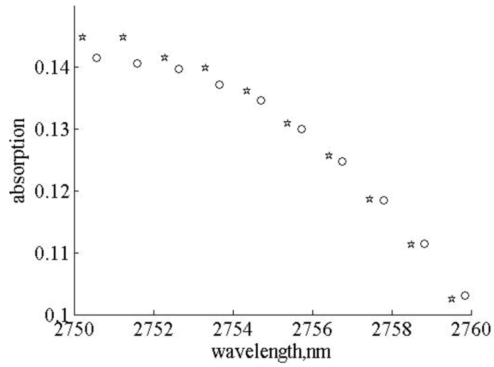

FIGURE 3. Part of EBS absorption spectrum of the same sample in primary and repeated measurements

The result is a set of spectra $\{f_c\}$ in the same wavelength range $\{\lambda\}$, each element of which corresponds to the experimental absorption spectra. Figure 4 shows the EBS absorption spectra undergone correction procedure.

COMPARISON OF THE EXPERIMENTAL DATA WITH THE SIMULATED SPECTRA

The mixture of 5% CO_2, dissolved in N_2 (production of "PGS", Zarechny, Russia), has been selected as the calibration gas mixture. The comparison of the results with the calculated spectra using database HITRAN are shown in Fig. 5. One can see that measured absorption spectra were very close to calculated spectra.

CLASSIFICATION

To estimate ability of classification we used below the data set, consisting of 109 absorption spectra of EBS from lung cancer patients (N = 16), patients with pneumonia (N = 16), patients with chronical obstructive pulmonary disease (COPD) (N = 40) and healthy volunteers (N = 37) measured by "LaserBreeze".

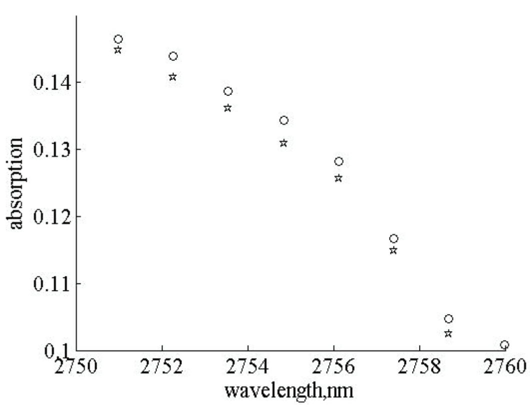

FIGURE 4. Part of EBS absorption spectrum of the same sample in primary and repeated measurements after averaging procedure

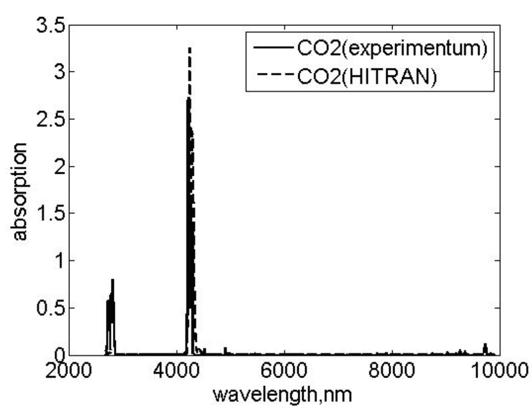

FIGURE 5. The experimental absorption spectra of 5% CO_2 in comparison with the calculated model absorption spectra of 5% CO_2 from HITRAN database

TABLE 1. Application of SVM with the respective kernels for the group of patients with pneumonia and join group of patients with lung cancer, COPD and healthy volunteers.

Classification variants	SVM kernel	Sensitivity		Specificity	
		Mean	Dispersion	Mean	Dispersion
Pneumonia (6/10) – COPD (6/34)	MLP	0.9834	0.0004	0.7920	0.0006
Pneumonia (6/10) – Lung cancer (6/10)	Gaussian Radial Basis Function	0.9789	0.0020	0.7041	0.0097
Pneumonia (6/10) – Healthy (6/31)	MLP	0.9333	0.0444	0.7950	0.0005

The two-stage analysis was applied. At the first stage, the initial feature space was transformed into the space of principal components using the method of principal component analysis (PCA) [10, 11]. At the second stage, the set of the experimental EBS spectra for four groups was classified using the support vector machine method (SVM). According to SVM approach we used random division of experimental data repeated over 50 times on training and testing sets. The training set was varied from three to ten spectra from each subgroup. The results of classification were transformed in terms of the specificity and sensitivity [12]. The mean and variance of these parameters is presented in the Table 1. The designations in brackets for the corresponding nosology denotes the number of elements in the training set / number of elements in testing set.

The calculations were performed for the following kernels: Linear, Quadratic, Polynomial up to 15 degree, Gaussian RadialBasis Function (kernel parameter changed from 0.001 to 20), Multilayer Perceptron (MLP) kernel [13, 14].

It is shown that wide-band IR absorption spectra of EBS in combination with SVM classification to the experimental data presented in the principal components space allow one to identify the representatives of target group (pneumonia) with high sensitivity, specificity and small variance.

CONCLUSIONS

LAS methods are very attractive for qualitative and quantitative analysis of the exhaled breath samples. The ability of LAS technique depends on tuning range of used laser source. The usage of optical parametric oscillators which provide wide ultra-wide tuning range are very promising, but there is a specificity to use the similar laser sources. The procedure of wavelength auto-calibration is shown to improve greatly the stability of wavelength binding. Comparison of the experimental data with the model absorption spectra of 5% CO_2 showed high agreement. The classification for four groups using SVM and PCA analysis methods has provided rather high sensitivity and specificity and small variance.

ACKNOWLEDGMENTS

This work was partially supported by the Federal Target Program for Research and Development, Contract No. 14.578.21.0082 (unique identifier of applied scientific research and experimental development RFMEFI57814X 0082).

The study reported in this article was conducted according to accepted ethical guidelines involving research in humans and/or animals and was approved by an appropriate institution or national research organization.

The study is compliant with the ethical standards as currently outlined in the Declaration of Helsinki.

All individual participants discussed in this study, or for whom any identifying information or image has been presented, have freely given their informed written consent for such information and/or image to be included in the published article.

REFERENCES

1. A. W. Boots, L. D. Bos, M. P. van der Schee, et al., Exhaled molecular fingerprinting in diagnosis and monitoring: Validating volatile promises, Trends Molecular Med. **21**(10), 633–644 (2015).
2. M. P. van der Schee, T. Paff, et al., Breathomics in lung disease, CHEST **147**(1), 224–231 (2015).

3. C. Lourenço and C. Turner, Breath analysis in disease diagnosis: Methodological considerations and applications, Metabolites **4**, 465–498 (2014).
4. Y. V. Kistenev, E. B. Bukreeva, A. A. Bulanova, et al., Laser spectroscopy and chemometric study the specific features of air exhaled by patients with lung cancer and chronic obstructive pulmonary disease, Phys. Wave Phenomena **22**(3), 210–215 (2014).
5. Y. V. Kistenev, A. I. Karapuzikov, and N. Yu. Kostyukova, Screening of patients with bronchopulmonary diseases using methods of infrared laser photoacoustic spectroscopy and principal component analysis, J. Biomed. Opt. **20**(6), 065001 (2015).
6. W. Chuji and S. Peeyush, Breath analysis using laser spectroscopic techniques: Breath biomarkers, spectral fingerprints, and detection limits, Sensors **9**(10), 8230–8262 (2009).
7. A. A. Karapuzikov, I. V. Sherstov, D. B. Kolker, et al., Laser breeze gas analyzer for noninvasive diagnostics of air exhaled by patients, Phys. Wave Phenomena (2014).
8. D. B. Kolker, M. K. Starikova, R. V. Pustovalova, et al., A nanosecond optical parametric oscillator in the mid-IR region with double-pass pump, Instruments Experimental Techniques (2012).
9. Optical parametric oscillator within 2.4–4.3 μm pumped with a nanosecond Nd:YAG Laser, Atmospheric Oceanic Optics (2012).
10. I. T. Jolliffe, *Principal Component Analysis* (Springer, 2002).
11. W. J. Krzanowski, *Principles of Multivariate Analysis* (Oxford University Press, Oxford, 1988).
12. T. Fawcett, An introduction to ROC analysis, Pattern Recognition Lett. **27**(8), 861–874 (2006).
13. V. Kecman, *Learning and Soft Computing* (MIT Press, Cambridge, MA, 2001).
14. N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods* (Cambridge University Press, Cambridge, 2000).