Breath air measurement using wide-band frequency tuning IR laser photo-acoustic spectroscopy

Yury V. Kistenev[1,2], Alexey V. Borisov[1,2], Dmitry A. Kuzmin[2], Anna A. Bulanova[2],
Andrey A. Boyko[3], Nadezhda Y. Kostyukova[3], Alexey A. Karapuzikov[3]
[1]National Research Tomsk State University, 36 Lenin av., Tomsk, Russia;
[2]Siberian State Medical University, 2 Moscowski Trakt St., Tomsk, Russia;
[3]Special Technologies, Ltd., 1/3 Zelenaya Gorka, Novosibirsk, Russia

ABSTRACT

The results of measuring of biomarkers in breath air of patients with broncho-pulmonary diseases using wide-band frequency tuning IR laser photo-acoustic spectroscopy and the methods of data mining are presented. We will discuss experimental equipment and various methods of intellectual analysis of the experimental spectra in context of above task.
The work was carried out with partial financial support of the FCPIR contract No 14.578.21.0082 (ID RFMEFI57814X0082).

Keywords: exhaled air, molecular biomarkers, lung cancer, chronic obstructive pulmonary disease, wide-band frequency tuning IR laser photo-acoustic spectroscopy, principal component analysis, support vector machine analysis

1. INTRODUCTION

Cardiovascular, oncologic, neurodegenerative and chronic respiratory diseases are the major causes of mortality in the world, which requires development of low-cost diagnostic tools for early detection of the above diseases, preferably using non-invasive approaches[1].

Taking into account that pathological changes at cell and molecular levels appear much earlier in a form of variations of metabolism products that clinical manifestation, evidently, that control of molecule metabolite profiles due to specific cellular processes is a promising tool for early diagnostics. The principal problem is connected with not high specificity of small molecular metabolites.

The *in vitro* investigation of molecular metabolites produced by cancerous cells in the microenvironment as the source of biomarkers should help with the dilemma of their origin, due to easier control of experimental conditions, the absence of factors such as gender, age, and individual variation. However, this method does not guarantee that the collected molecular metabolites are of endogenous origin.[2] Comparison of the metabolite profiles of breath analysis and lung cancer cell cultures (LCCC)[3] had shown that, of the 68 metabolites s detected in either breath or cell culture, only 16 ones were detected in both cell culture and breath, 17 metabolites detected only in breath, 22 found only in LCCC and 13 metabolites found only in controls (non-transformed cell lines). This poor relationship indicates that *in vitro* LCCC is not a good model for the prediction of molecular metabolites in breath of lung cancer patients. A more detailed examination show that possible reason of discrepancy connected with cellular oxidative stress in living tissue.

Breath analysis in comparison with blood and urine tests is noninvasive, easy to produce and to repeat.[2] For pulmonary diseases the content of a of metabolites in exhaled air is varied under the pathological processes in the lungs, including:

— inorganic compounds, e.g., carbon dioxide, oxygen, and nitric oxide;

— non-volatile compounds usually measured in exhaled breath condensate, e.g., isoprostanes, cytokines, leukotrienes;

— volatile organic compounds VOCs which can be divided into different chemical groups, e.g., saturated hydrocarbons (ethane, pentane, aldehydes), unsaturated hydrocarbons (for example, isoprene), oxygen containing (for example acetone), sulphur containing (including ethyl mercaptane, dimethylsulfide) and nitrogen containing (including dimethylamine, ammonia). The most commonly identified VOCs are isoprene, acetone, ethanol, methanol, other alcohols and alkanes[4]. There are a number of reviews connected with analysis of potential molecular metabolites in a breath air mostly represented volatile organic compounds (VOCs).[1-2, 4-7]

In the field of screening of pulmonary diseases analysis of literature shows that acceptable specificity and sensitivity of the tests is achieved under the control of 10-20 VOCs in the breath air (see Table 1). It should be noted that for screening methods is more important specificity, not sensitivity.

Dynamics and Fluctuations in Biomedical Photonics XIII, edited by Valery V. Tuchin, Kirill V. Larin, Martin J. Leahy, Ruikang K. Wang, Proc. of SPIE Vol. 9707, 97070M
© 2016 SPIE · CCC code: 1605-7422/16/$18 · doi: 10.1117/12.2214645

As to lung cancer, 112 potential markers of in the exhaled air were registered during the last ten years.[8] Among them are: 36 hydrocarbons, 7 alcohols, 8 aldehydes, 2 acids, 12 ketones, 12 aromatic compounds, 2 heterocycles, 2 nitriles, 5 terpenes, 9 ethers, 1 sulfide, 2 halogenated compounds, and 15 compounds from other chemical classes (the examples are presented in the Table 2).

Table 1. The dependence of specificity and sensitivity of the pulmonary diseases tests based on breath air component analysis on the quantity of measured VOCs

№	Quantity of VOCs	Specificity, %	Sensitivity, %	Objects of classification	Ref.
1.	15	71	100	Lung cancer / Control	[9]
2.	21	80	100	Lung cancer / Control	[10]
3.	22	81	100	Lung cancer / Control	[11]
4.	13	80		Lung cancer / Control	[12]
5.	6	81	100	Chronic obstructive pulmonary disease / Control	[13]

Table 2. Examples of classes of VOCs, typical for lung cancer[8]

Hydrocarbons	Alcohols	Aldehydes	Ketones	Aromatic compounds	Terpenes
2 - methyl-propane 5-methyl-tridecan	1-octene-3-ol	Pentanal Hexanal Octanal Nounal	6-methyl-5-hepten-2-one	Mixture of benzophenone	trans-caryophyllene

The technical background of the method consists of various analytical methods. The Selected Ion Flow Tube Mass Spectrometry (SIFT-MS), Proton Transfer Reaction Mass Spectrometry (PTR-MS), Gas Chromatography-Mass Spectrometry (GC-MS) with thermal desorption or solid-phase micro extraction (SPME), laser absorption spectroscopy (LAS), ion mobility spectrometry (IMS), and electronic noses containing a variety of gas sensors and semiconductor-based sensor arrays have been used.[12] Basic characteristics of these methods are presented in the Table 3.

Table 3. A comparison of the characteristics of the available breath research techniques[14]

Analytical method	Mode of operation	Limit of detection	Sensitivity	Specificity
SIFT-MS	Direct / Real time	ppbv	High	High
PTR-MS	Direct / Real time	pptv	High	Medium-High
IMS	Real-time	ppbv	Medium	Medium
GC-MS	Pre-concentration	pptv-ppbv	Very-high	Very-high
LAS	Real-time	ppbv	High	High
Sensor arrays	Reference to a database	ppbv	Medium	Medium

The analytical methods in the Table 3 are placed according to their complexity, which is important for routine applications. LAS and sensor arrays are most easy to use, but gas sensors are often much less sensitive, usually lack specificity, and are prone to drift.

For elimination of disadvantages of sensor arrays the new types of sensors and coatings are being developed, in particular, the chemical sensors which change their color when a certain VOC's appear (colorimetric sensors), or change frequency of quartz resonator etc.[15,16]. Different types of selective pre-sampling are applied.

Photoacoustic spectroscopy (PAS) is the variant of LAS, which is one of the most sensitive methods of gas analysis, especially with use of coherent radiation sources and the intracavity photoacoustic detection[17]. PAS has very low detection limit (in optimal conditions ~ 1 ppt), sufficient selectivity. There is no need of sample pre-concentration, small volume of gas sample (several ml) is enough; there is ability to monitoring of a component in real time in the regime of constant flow of gas mixture through the cell.

2. WIDE-BAND FREQUENCY TUNING IR LASER PHOTO-ACOUSTIC SPECTROSCOPY

The potential of LAS/PAS method is defined by the available for registration spectrum range. In the Table 4 the spectral regions, which are suitable for registration of typical pulmonary diseases markers, are presented. Evidently, that for simultaneous detection of 1-20 VOCs one should use tunable laser sourced with tuning range from 3 μm to 10-11 μm.

Frequency conversion using optical parametric oscillator (OPO) is one of the effective ways to generate widely tunable coherent light in spectral range from visible to mid-IR range. We developed gas analyzer LaserBreeze[18] based on OPO method with tuning range from 2.5 to 10.7 μm. In OPO we used two types of nonlinear elements: periodically poled lithium niobate structure (PPLN) and mercury thiogallate crystal $HgGa_2S_4$ (HGS). Nd:YLF laser (10 ns, 0.5 – 1.5 kHz, 1.5 mJ) was used as a pump source. The linewidth of developed OPOs was 3-4 cm^{-1}. The average power of OPO based on PPLN structure was 20mW (1700 Hz). The average power of OPO based on HGS crystal was 9 mW (900 Hz). The double channel resonant photo-acoustic cell was used for recording absorption spectra of gaseous samples. The LaserBreeze gas analyzer block-scheme is presented in Figure 1, its appearance is presented in the Figure 2. Main technical characteristics of LaserBreeze gas analyzer are shown in Table 5.

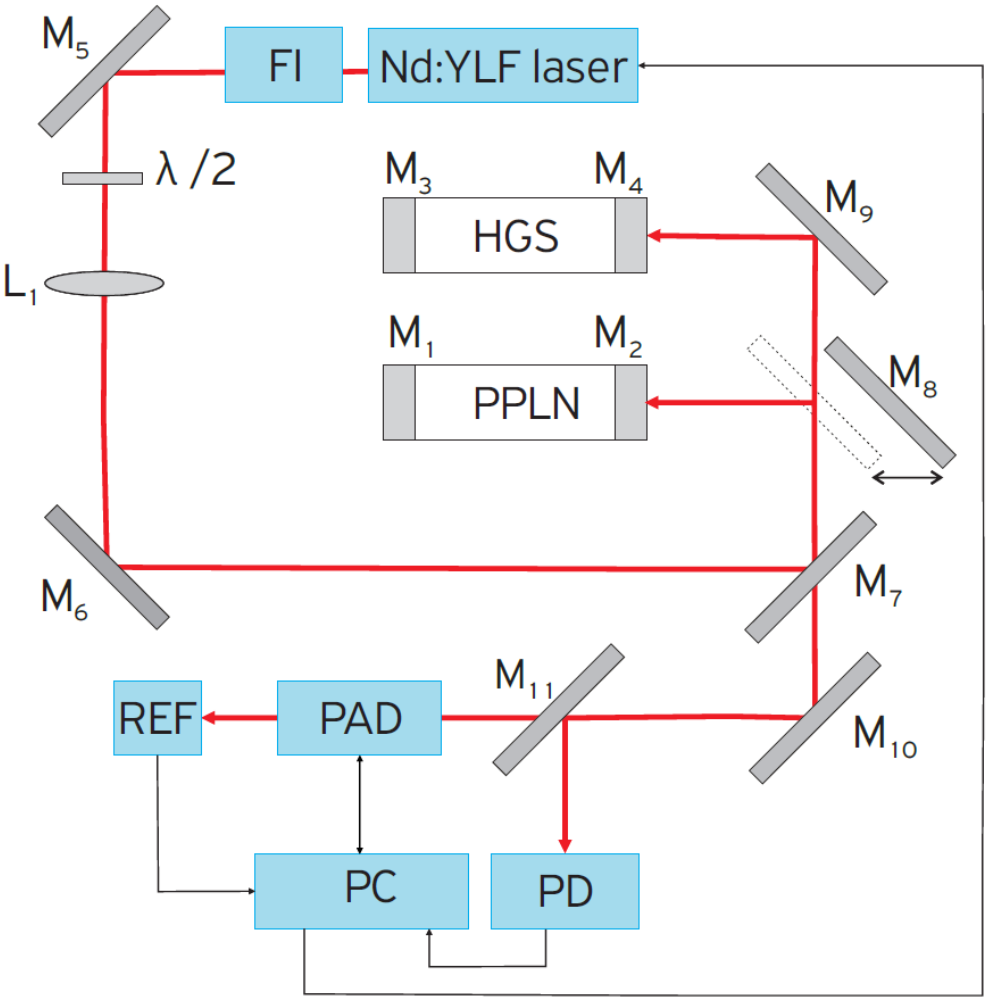

Figure 1. LaserBreeze gas analyzer block-scheme. Here, FI is the Faraday isolator, PAD is the photoacoustic detector, REF is the reference gas-cell with special gas-mixture, PD is the pyroelectric photodetector, PC is the personal computer, M_8 is the reflecting mirror placed on motorized translation stage.

Figure 2. The LaserBreeze gas analyzer appearance

Table 4. Main volatile organic compounds[19]

VOCs	Suitable for detecting absorption bands, μm
acetone, C_3H_6O	7.35
acetylene, C_2H_2	3.05
ammonia, NH_3	10.35
butane, C_4H_{10}	3.387, 10.45
carbon dioxide, CO_2	4.24, 10.6
carbon dioxide (13 isotope), $^{13}CO_2$	4.408
carbon monoxide, CO	4.62
ethane, C_2H_6	3.348
ethanol, C_2H_5OH	9.38
ethyl acetate, $C_4H_8O_2$	8.03, 9.47
ethylene, C_2H_4	10.53
methane, CH_4	7.7
nitrogen dioxide, NO_2	6.25
nitrogen oxide, NO	5.25
nitrous oxide, N_2O	3.89
pentane, C_5H_{12}	3.372
propane, C_3H_8	3.375
sulfur dioxide,SO_2	7.28

Table 5. Main technical characteristics of LaserBreeze gas analyzer[18]

Parameter	Value
Concentration sensitivity	No worse than 1×10^{-3}ppm
Number of detected molecular biomarkers	No less than 20
Relative error in determining VOCs concentration	No more than 10-30%
Reliability and selectivity of VOCs identification	No less than 95%
Scanning range of OPO radiation	2.5 to 10.7 μm
Sample volume	No more than 50 cm^3
Detection time for one VOCs in a sample	No more than 3 s
Detection time for 10 VOCs in a sample	No more than 2 min

3. DATA ANALYSIS

Statistical methods of estimation significant variations between the target and control groups have limited use at the small sample sets typical for breath air analysis. The most common method used is the Principal Component Analysis (PCA), which essentially does a reduction in the multidimensional data.[20] The basic idea of PCA is to find the minimum number of new features which are enough for recovery of the basic features by linear transformation, possibly with insignificant errors. PCA projects correlated variables into a lower number of uncorrelated variables called principal components. A specific feature of PCA is that the hidden connections and patterns that are typical for the investigated data set can be revealed.

Mathematical background of PCA consists in decomposition of initial experimental data 2D-dimension matrix X ($I \times J$) in a form of matrix product[20]:

$$X = T \cdot P^t + E = \sum_{a=1}^{A} t_a \cdot p_a^t + E \quad (1)$$

where I is the quantity of samples of experimental data, J is quantity of features of investigated objects, T ($I \times A$) is the scores matrix, P ($A \times J$) is the loadings matrix, E is the residuals matrix, A is the quantity of principal components. In our case these features of the state under investigation are the set of absorption coefficients of exhaled air sample in laser source frequency detuning branch of the used gas analyzer. PCA is useful, if $A << J$. In this case the method allows, first of all, to separate most informative features of the state, by other words to reduce the dimension of the feature space, and to decrease noise, and secondly, to estimate the relative position of the studied objects in the reduced space of principal components (PC).

Metabolomics data often show non-linear patterns, but these problems are well handled using non-linear methods. The most often used of them are the Kernel-based models as the support vector machine (SVM) for data classification[21]. Classifying data is connected with a common task of machine learning. SVM classification includes training stage, so data from each group were randomly divided into two sets, one of which used for the training and the second for testing the classification rules. The first stage is based on construction of a hyperplane that has the largest distance to the nearest training-data point of any class, since in general the larger the margin the lower the generalization error of the classifier. At the second stage this hyperplane used for classification.

4. RESULTS

Consider the application the SVM classifier for separate of multiple absorption spectra of exhaled air (ASEA) measured by LaserBreeze gas analyzer from the lung cancer (LC) patients, patients with chronical obstructive pulmonary disease (COPD) and healthy volunteers.

Table 6 shows an example of direct application of the SVM method for binary separation of ASEA for LC and COPD patients. Training sets were included 7 ASEA for every illness. The test set was included 8 ASEA from LC patients and 24 ASEA from COPD patients. The spectra from the training sets were not included in the test set.

Table 6. The example of direct binary separation of ASEA for LC and COPD patients using the SVM method

LC patients		**COPD patients**	
Patients in the test set	Identified by the SVM method patients	Patients in the test set	Identified by the SVM method patients
8	1	24	31

It is obvious that the direct application of the SVM method for classification of spectra of ASEA is not effective, due to peculiarity of the initial data. Calculations had been shown that using of various kernels in SVM does not give suitable results. However, the consistent using of the PCA and SVM methods gives a significant improvement of the binary separation of ASEA for LC and COPD patients from above mention sets. The Table 7 shows the classification results with different kernels for several combinations of used principal components.

Table 7. The example of direct binary separation of ASEA for LC and COPD patients using the consistent using of the PCA and SVM methods

SVM Kernel	Used Principal components	LC patients in the test set	COPD patients in the test set	LC patients, correctly identified by the SVM method	LC patients, not identified by the SVM method	COPD patients, correctly identified by the SVM method	COPD patients, not identified by the SVM method
Linear	First and third	8	24	8	5	19	0
Quadratic	First and third	8	24	8	0	24	0
Polynomial	First and third	8	24	6	0	24	2
Gaussian Radial Basis Function	First and third	8	24	8	6	18	0
Linear	First and fourth	8	24	8	3	21	0
Quadratic	First and fourth	8	24	8	0	24	0
Polynomial	First and fourth	8	24	5	0	24	3
Gaussian Radial Basis Function	First and fourth	8	24	8	6	18	0

The results in the table show that the SVM classification with quadratic kernel based on the data from combination of the first and third or the first and fourth principal component of the ASEA is most effective. The calculations had been shown that the SVM classification based on the usage of larger number of principal components of the ASEA has less effectiveness.

5. CONCLUSION

The presented results show high potential of wide-band frequency tuning IR laser photo-acoustic spectroscopy method based on OPO laser source in combination with PCA and SVM methods of the experimental spectra analysis for screening diagnostics of pulmonary diseases.

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 RFMEFI57814X0082).

REFERENCES

[1] Pereira, J., Porto-Figueira, P., Cavaco, C., Taunk, K., Rapole, S., Dhakne, R., Nagarajaram, H., Câmara, J. S., "Breath Analysis as a Potential and Non-Invasive Frontier in Disease Diagnosis: An Overview," Metabolites 5, 3-55 (2014).

[2] Schmidt, K. and Podmore, I., "Current Challenges in Volatile Organic Compounds Analysis as Potential Biomarkers of Cancer," Journal of Biomarkers 2015, 1-16 (2015).

[3] Kalluri, U., Naiker, M., Myers, M. A., "Cell culture metabolomics in the diagnosis of lung cancer – the influence of cell culture conditions," J. Breath Res. 8, 1-10 (2014).

[4] Dent, A.G., Sutedja, T.G., Zimmerman, P.V. "Exhaled breath analysis for lung cancer". J. Thorac. Dis. 5, s540-s550 (2013).

[5] Lechner, M., "Mass spectrometric profiling of low-molecular-weight voltatile compounds," Diagnostic Potential and Latest Applications Current Medicinal Chemistry 14, 987-995 (2007).
[6] Hakim, M., Broza, Y. Y., Barash, O., Peled, N., Phillips, M., Amann, A., Haick, H., "Volatile Organic Compounds of Lung Cancer and Possible Biochemical Pathways," Chem. Rev. 112(11), 5949-66 (2012).
[7] Ulanowska, A., Ligor, M., Amann, A., Buszewski, B., "Determination of Volatile Organic Compounds in Exhaled Breath by Ion Mobility Spectrometry", Chem. Anal. 53, 953-965 (2008).
[8] Amann, A., Mochalski, P., Ruzsanyi, V., Broza Y.Y., Haick, H. "Assessment of the exhalation kinetics of volatile cancer biomarkers based on their physicochemical properties", J Breath Res. 8 (1), 016003 (11 pages) (2014).
[9] Cao, W., "Breath Analysis: Potential for Clinical Diagnosis and Exposure Assessment," Clinical Chemistry 52, 800-811 (2006).
[10] Ager, C., Bajtarevic, A., "Noninvasive detection of lung cancer by analysis of exhaled breath," BMC Cancer 9, 1-16 (2009).
[11] Gleeson, K., Phillips, M., "Volatile organic compounds in breath as markers of lung cancer: a cross-sectional study," The Lancet 353, 1930-1933 (1999).
[12] Carbognani P., Poli, D., "Exhaled volatile organic compounds in patients with non-small cell lung cancer: cross sectional and nested short-term follow-up study," Respir Res. 6(1), 1-10 (2005).
[13] Dallinga, J.W., Möller, G.M., Godschalk, R.W.L., Moonen, E.J., Wouters, E.F.M., Van Schooten, F.J. Van Berkel, J.B.N., "A profile of volatile organic compounds in breath discriminates COPD patients from controls," Respiratory Medicine 104 (4), 557-563 (2010).
[14] Lourenço, C., Turner, C., "Breath Analysis in Disease Diagnosis: Methodological Considerations and Applications," Metabolites 4, 465-498 (2014).
[15] Wilson, A.D., Baietto, M. "Advances in Electronic-Nose Technologies Developed for Biomedical Applications". Sensors 11(1), 1105-1176 (2011).
[16] Montuschi, P., Mores, N., Trové, A., Mondino, C., Barnes, P.J. "The Electronic Nose in Respiratory Medicine". Respiration 85 (1), 72-84 (2013).
[17] Miklos, A., Hess, P., Bozoki, Z. "Application of acoustic resonators in photoacoustic trace gas analysis", Review of scientific instruments 72 (4), 1937-1955 (2001).
[18] Karapuzikov, A.A. et al. "LaserBreeze Gas Analyzer for Noninvasive Diagnostics of Air Exhaled by Patients. Phys. Wave Phenomena 22(03), 189-196 (2014).
[19] Kistenev, Yu. V., Karapuzikov, A.I., Kostyukova, N.Yu., Starikova, M.K., Boyko, A.A., Bukreeva, E. B., Bulanova, A. A., Kolker, D.B., Kuzmin, D.A., Zenov, K.G., Karapuzikov, A.A., "Screening of patients with bronchopulmonary diseases using methods of infrared laser photoacoustic spectroscopy and principal component analysis," Journal of Biomedical Optics 20 (6), 1-6 (2015).
[20] Pomerantsev, A. L., Rodionova, O. Ye., "Concept and role of extreme objects in PCA/SIMCA," Journal of Chemometrics 28(5), 429-438 (2014).
[21] Burges, C. J. C., "Geometry and invariance in kernel based methods," Advances in Kernel Methods: Support Vector Learning, 89-116 (1999).