



**HAL**  
open science

## Pilot Study for Validation and Differentiation of Alveolar and Esophageal Air

Paulo Santos, Valentina Vassilenko, Carolina Conduto, Jorge M. Fernandes,  
Pedro C. Moura, Paulo Bonifácio

► **To cite this version:**

Paulo Santos, Valentina Vassilenko, Carolina Conduto, Jorge M. Fernandes, Pedro C. Moura, et al.. Pilot Study for Validation and Differentiation of Alveolar and Esophageal Air. 12th Doctoral Conference on Computing, Electrical and Industrial Systems (DoCEIS), Jul 2021, Costa de Caparica, Portugal. pp.331-338, 10.1007/978-3-030-78288-7\_32 . hal-03685932

**HAL Id: hal-03685932**

**<https://inria.hal.science/hal-03685932v1>**

Submitted on 2 Jun 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Pilot Study for Validation and Differentiation of Alveolar and Esophageal Air

Paulo Santos<sup>1,2</sup>, Valentina Vassilenko<sup>1,2</sup>, Carolina Conduto<sup>1</sup>, Jorge M. Fernandes<sup>1,2</sup>,  
Pedro C. Moura<sup>1</sup> and Paulo Bonifácio<sup>1,2</sup>

<sup>1</sup>Laboratory of Instrumentation, Biomedical Engineering and Radiation Physics (LibPhys-UNL), NOVA School of Science and Technology, Campus FCT UNL, Caparica, Portugal

<sup>2</sup>NMT, S.A., Edifício Madan Parque, Rua dos Inventores, Caparica, Portugal  
ph.santos@campus.fct.unl.pt , vv@fct.unl.pt

**Abstract.** Breath analysis is an expanding scientific field with great potential for creating personalized and non-invasive health screening and diagnostics techniques. However, the wide range of contradictory results in breath analysis is explained by the lack of an optimal standard procedure for selective breath sampling. Recently we developed novel instrumentation for selective breath sampling, enabling the precise collection of a pre-determined portion of exhaled air using AI (Machine Learning) algorithm. This work presents pilot study results for validation of developed technology by differentiation of alveolar and oesophageal air obtained from the healthy population (n=31). The samples were analyzed in-situ by Gas Chromatography-Ion Mobility Spectrometry (GC-IMS) apparatus, and obtained spectra were processed with proper multivariate classification tools. The results show a promising performance of proposed AI-based technology for breath sampling adapted to users' age, genre, and physiological conditions.

**Keywords:** Alveolar Air, Selective Air Acquisition, Breath Sampling, Machine Learning, Medical Instrumentation, Principal Component Analysis

## 1 Introduction

Molecular analysis of exhaled volatile compounds (breathomics) has been innately attractive, allowing non-invasive, quick, and straightforward observations of multiple biochemical processes of the human body. Indeed, modern technology breakthroughs have placed the analysis of exhaled air metabolites as a reliable non-invasive diagnostic and a prognostic tool for a wide variety of medical conditions to assess different vital organ functions [1]. Recently the non-invasive detection of biomarkers in breath associated with tuberculosis (TB), chronic obstructive pulmonary disease (COPD), asthma and an array of cancers have been reported [2,3].

In every single human breath, more than one thousand different molecules are exhaled. Volatile Organic Compounds (VOCs) arise in exhaled breath, urine, blood, saliva, and faeces, and are also emitted by skin [4]. VOCs can have an endogenous metabolic origin either related or not to previous exogenous exposures (environment,

smoking, medication, food, etc.) [1] or can have a microbial origin (*i.e.*, in the gut or airways) [5,6]. The concentration of exhaled VOCs is also influenced by the compound-specific blood/gas partition coefficient (*i.e.*, respiratory gas exchange), cardiac output and alveolar minute volume. Therefore, the exhaled air composition also shifts according to its respiratory origin or source. The first exhaled portion of a tidal breath corresponds to oesophageal air, approximately, one third (around 150ml) of the tidal breath volume (500ml in an average adult). It is known as physiological dead space and contains molecules from mouth, larynx, trachea, and bronchi. This air type has increased concentrations of exogenous VOCs (mainly linked to room air, environmental exposure, odour, food, *etc.*) The remaining two-thirds of a complete exhalation is composed of end-tidal or alveolar air (*i.e.*, 350mL).[7] Alveolar air has been deeply studied due to the direct exchange of breath constituents with systemic blood. Such phenomenon provides an analysis of a higher concentration of endogenous VOCs, offering a metabolic window of the biochemical processes of the body [7].

As alveolar and oesophageal (*i.e.*, dead space air excluding mouth) carry specific clinical information, devices used to collect exhaled VOCs should be intelligent by adapting to patients' breathing characteristics and controlling the sampling procedure when isolating an air portion for further analysis.

Notwithstanding the recent and quick development of breathomics, no robust and repeatable breath-related VOCs' profiling technologies have been clinically validated. The major challenges for the validation of breath sampling technologies are linked to lack of standardized procedures to accurately collect extremely low concentrations of source-dependent exhaled air constituents that can be originated from oral cavity, oesophagus, and alveoli. Unsolved issues related to exhaled breath sampling include correction for ambient inspired volatiles, type of sampling (total versus alveolar breath), sampling duration (single-breath versus fixed-time or fixed-volume breathing), the effect of expiratory flow, breath hold, type of materials, sample pretreatment, the effect of humidity, food/medications, exercise, smoking and co-morbidities [9].

A full control of the breath sampling procedure by assessing the patients' respiratory cycle enables the identification of part of the respiratory tract from which the sample derived from. Presently the most used devices that control breath sampling identify the different phases of the breathing cycle by measuring CO<sub>2</sub> concentration, breathing flow, pressure, temperature, or humidity. However, most of these methods introduce high variability in breath samples due to: (a) the way the air is expelled, (b) the breath frequency, (c) the length and depth of the breath cycle and (d) the mental and physical condition of the patient. The problems remaining to be solved comprise the question of how to achieve accurate, selective, and repeatable sampling, how to ensure easy and safe handling, and maybe most importantly, the issue of sample stability to allow a proper chemical analysis. Therefore, up-to-date research in breath analysis tries to pursue a suitable device and a precise protocol for sampling exhaled air independently of patients: age, gender, metabolic production of CO<sub>2</sub>, smoking habits, nutrition, and health conditions; thus, harmonizing the breath collection methods already approached by the scientific community. The research herein presented aims to disclose the benefits of implementing Artificial Intelligence in an advanced breath sampling prototype capable of selectively collecting exhaled air samples according to the user's respiratory source, independently of their metabolic production of carbon dioxide [10].

## 2 Relationship to Applied Artificial Intelligence Systems

This paper reports the results of the performance of a novel ML-based system for a selective and adaptive sampling of breath, as well as its differentiation, through the experimental measurements by fast, reagent-free, and portable analytical instrumentation of Ion Mobility Spectrometry with Gas-Chromatography pre-separation (GC-IMS) and high sensitivity and selectivity for VOC in ppb<sub>v</sub> (parts per billion by volume) or ppt<sub>v</sub> (parts per trillion) concentrations range [10,11].

Most breath samplers comprise some algorithms to analyze the user respiration cycles to detect inspiration and expiration phases and to, consequently, determine the time window for breath sampling. However, and erratic respiratory rhythms that determined the time window for a sample is shorter and brought up multiple issues when obtaining small portions of exhaled air. Only the system patented by Capnia, Inc. (WO2015143384 A1, 2015) is configured to impose a breath frequency to users (young children and non-cognizant patients) to avoid those erratic respiratory episodes. Even so, that imposed frequency does not adapt itself with the user's breathing pace by combining a set of users' characteristics as the proposed system does.

The breath sampling ML-based innovative prototype was developed in our group by researchers from NMT, S.A. and NOVA School of Science and Technology Nova University of Lisbon, Portugal (patent WO/2018/047058, 12 09 2016). It performs continuous real-time breathing flow measurements to collect a pre-determined portion of exhaled air by synchronizing a previously modelled respiratory cycle with the breathing cycle of the user. Through real-time synchronization of breathing cycles, the device detects optimized sampling instants through a machine learning-based algorithm described elsewhere [10,11]. The system includes two hardware units (communication module and end-user) controlled by intelligent software loaded on a computing device (*e.g.*, desktop, or tablet). The software contains a human-machine interface (HMI) and a sampling algorithm to control and personalize breaths' sampling procedure. Such HMI enables: (a) visual feedback of operations (*i.e.* sensor signal, breathing pacer and sampling status), (b) personalization of breath collection procedure (by defining sampling features and users' characteristics such as age, gender, and physiological state) and (c) the recording real-time flow measurement data.

In turn, the sampling algorithm implemented in the intelligent control software was configured to (i) measure the user's respiratory flow, (ii) distinguish inspiratory and expiratory breath phases, (iii) synchronize the respiratory cycle measured with representative and modelled respiratory cycle and, (iv) to calculate the user's average time of expiration (ATE). The machine learning (ML) algorithm implemented in the software algorithm allows predicting the time of a new exhalation and, consequently, precise prediction of sampling timeframes. This approach enables the ML process to assess the user's respiratory cycle and test it on the modelled respiratory cycle intrinsically contained in the sampling algorithm. [11]. Afterwards, and within timeframes predicted, the software communicates with the end-user to channel the sample, either into a collection reservoir or to direct online analysis [10].

On the other side, to validate the performance of this patented technology, a PhD work plan of one of the authors consists of a demonstration of how this AI-based technology works in field measurements and if it is possible to separate the very small portions of breath, *i.e.*, the alveolar and esophageal exhaled air. For this purpose, a

series of experimental measurements described in this work was performed. The obtained samples of exhaled air were analyzed on-line by GC-IMS instrumentation. Since each spectrum contains extensive data of analyte information, they were processed with proper for spectral analysis multivariate classification tools: Principal Component Analysis, PCA (Unsupervised Learning Procedure) and Partial Least Square Discriminant Analysis (Supervised Learning Procedure).

### 3 Methodology for Breath Differentiation

The methodology herein presented was applied during *in-situ* implementation tests that focused mainly on breath volatome differentiation using the abovementioned our breath sampler prototype. Therefore, the prototype's sampling selectivity (*i.e.*, the ability for sampling and differentiation) was tested through the collection and on-line analysis of esophageal and alveolar air samples obtained from an independent cohort of university students [6]. The 31 healthy non-smokers volunteers, without lung diseases, heart conditions and gastric disorders and not pregnant, aged 20 to 35, participate in the study. Additional pre-conditions were also required: (a) 3-5 hours of fasting; (b) no prior consumption of alcohol or leguminous plants; (c) 3 hours from the last mouth wash; (d) absence of hygiene products or perfume prior to the measurements and (e) light meal before the test (avoiding products as meat or fish). Within these tests, each volunteer attended two sessions (alveolar and esophageal exhaled air acquisition) occurring on separate days. Although the laboratory environment presents a stable VOC content, as previously reported [1], all participants were exposed to environmental air for the same timeframe (*i.e.*, 30 minutes) before any breath collection.

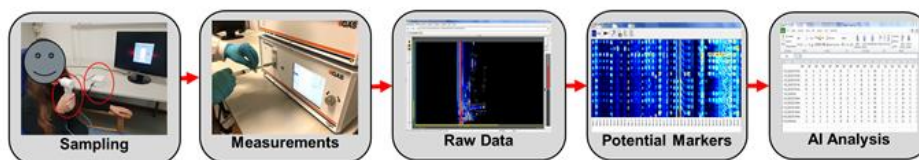
#### 3.1 Exhaled Air Sampling & Analysis

In previously reported work [11,12], have demonstrated that ML algorithm led to improvements in the sampling procedure, reducing standard deviation for average time of exhalation (ATE), the number of cycles and time required to start sampling. Therefore, a similar protocol for breath sampling was applied during breath differentiation tests. Thus, after powering and plugging the advanced prototype into a computing device, the prototype's HMI suggested a breathing rhythm to each user according to its age, gender, and physiological state. Afterwards, each volunteer was asked to breathe until the fraction of exhaled air was sampled by the device. Such process ensured that only a specific fraction of breath was diverted to 5 ml syringes. Breath samples were analyzed by BreathSpec® device (from GAS Dortmund). Such device couples an Ion Mobility Spectrometer (IMS) with Gas Chromatograph (GC), where gas analytes are pre-separated in the chromatographic column [14]. A description of IMS principle and details of its work can be found in a previous publication [15]. However, it is worth mentioning the ability of this instrumentation to deliver results without the need for any chemical reagent or sample pre-processing. Exhaled breath analytes were analyzed for 15 minutes with a carrier gas flow of 10 mL/min for the first 5 minutes and 50 mL/min for the remaining time. The constant flux of 150 mL/min for drift gas was used. As a result, the 3D GC-IMS spectra were obtained. The spectra

features include the drift time (Dt), retention time (Rt) and intensity (I), as it shown in Fig.2. Dt (in milliseconds) is relative to the Reactive Ion Peak (RIP), representing the number of ions available to ionize analytes. Rt indicates the time VOCs to travel the GC column, and is expressed in seconds [15].

### 3.2 Data Analysis

Laboratory Analysis Viewer (LAV) software (version 2.2.1) was used to visually compare spectra and create area-sets for each relative intensity peak. Relative intensities within each area-set were exported for multivariate data analysis (MVA) analysis: Principal Component Analysis (PCA) using IBM SPSS Statistics software (version 23) and Partial Least Square Discriminant Analysis (PLSDA).

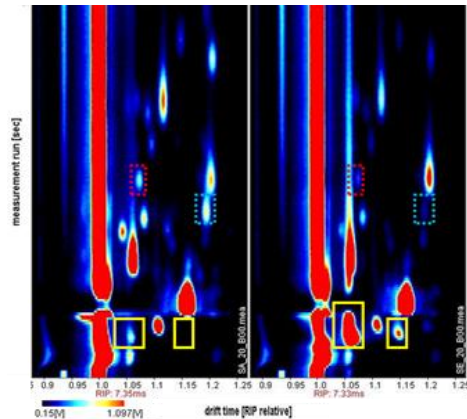


**Fig. 1.** Workflow of the processes applied throughout the study for breath differentiation. From left to right: breath sampling, measurements, spectra analysis, area-set definition, and AI analysis.

## 4 Results & Discussion

The analysis of esophageal and alveolar air samples (Fig. 2) revealed 31 and 33 GC-IMS spectral intensity peaks, respectively. A small group of intensity peaks were found to be subject specific (individual variability) while others showed relevant differentiation (regularly or not) for both types of exhaled air. The intensity of GC-IMS spectral peaks correspondent to ethanol (monomer and dimer) was significantly higher ( $p < 0.05$ ) for esophageal air than for alveolar air.

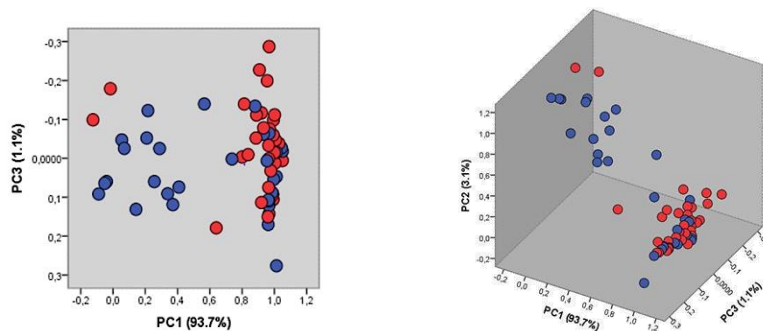
Alveolar and esophageal air specific fingerprints spectra bring forward differences between the two types of air. Besides the fact that most VOCs are simultaneously present in both alveolar and esophageal airs, the pivotal divergences can be seen in compounds' intensity. In Fig. 2., four intensity peaks, are marked in both alveolar and esophageal spectra. Ethanol's monomer (left) and dimer (right) are marked with a solid yellow line, whereas both the red and blue dashed lines point to two still not identified compounds. Although being a very common compound in indoor and outdoor airs [1], ethanol shows a source-dependent behavior for breath samples, exhibited by significant variations in the intensities of the spectra for both types of air. For esophageal air, ethanol's characteristic peaks (solid yellow line) show higher intensity values than for alveolar air samples. This may be related to the influence of the (a) food ingestion path across the esophagus or (b) environmental air nearness. However, the two unidentified compounds exhibit an opposite behavior (higher in alveolar air) and, therefore, can be correlated with metabolic VOCs participating in gas lung exchanges also reflecting concentrations in blood. [16]



**Fig. 2.** GC-IMS spectra representative of alveolar (left) and esophageal (right) airs. Most visually discriminant VOCs are highlighted (ethanol (monomer and dimer) marked in yellow solid line and two unidentified VOCs marked with a red and blue dashed line).

Although visual spectra comparison provides a qualitative assessment of the most concentrated compounds, proper statistical data analysis allows to perceive hidden (or not so visible) evidence and differentiate patterns for each sample type. The application of unsupervised PCA (Fig. 3.), highlights the differences between the profile of alveolar (red points) and esophageal (blue points) air samples. Bi-dimensional representation of PC1 and PC3 displays higher visual differences between both clusters.

Principal Components (PC) exhibited 93.7% [PC1], 3.1% [PC2] and 1.1% [PC3] of the total explained variance of the data set for the observations of both types of exhaled air. Some esophageal measurements (53.3% of the esophageal cases) were clustered within the cohort of alveolar air samples that were more correctly grouped (only 6.5% of alveolar cases were wrongly clustered as esophageal). This result indicates the collection of a mixture of exhaled air types (usual phenomenon in breath sampling) that is thought to be caused either by: (a) the increased temperature of the prototype's solenoid valve and working failures posteriorly linked to an arrival to its functioning hysteresis region (hardware drawback); (b) individual variation of esophageal sphincter closure degree when breathing; (c) individual differences in respiratory physiology.



**Fig. 3.** Representation of alveolar (red) and esophageal (blue) air patterns after PCA (for 2 (left) and for the 3 (right) most relevant principal components - PC). Total variance is also included for each principal component (PC1, PC2 and PC3).

Measured VOCs spectra from exhaled air were also processed by another Multivariate Data Analysis, a Partial Least Square Discriminant Analysis (PLSDA). However, the results for alveolar and esophageal air differentiation obtained by PLSDA were less clear than by PCA ones, thus are not presented.

## 5 Conclusions & Future Work

In this paper a novel AI-based technology for selectively sampling exhaled air according to subjects' characteristics (*i.e.*, age, gender, metabolic production of CO<sub>2</sub> and physiological state) was presented. Results obtained from breath measurements showed differentiation in exhaled air profiles related to breath source/origin. Two distinguishable VOC patterns were determined, and ethanol (monomer and dimer) was identified as one of the main characteristic compounds of esophageal exhaled air.

In this work, two multivariate classification tools were applied to identify better the unperceived variations between both types of exhaled air samples: Unsupervised Learning Procedure, PCA, and Supervised Learning Procedure, PLSDA. PCA comprised the most suitable approach to differentiate both exhaled air profiles so far, thus revealing the ability of our device to collect breath samples according to the respiratory source/origin. Furthermore, a deeper analysis resorting to alternative classification tools, such as t-SNE t-distributed Stochastic Neighbor Embedding, is planned to evaluate cluster separation.

Our breath sampler and its capability for breath differentiation through AI must be highlighted as the potential of application for clinical screening is increased for several diseases whose VOC biomarkers concentration depend on the source of air to sample. Therefore, the workflow process herein presented can be adopted, and widely used as it proves obvious advantages of implementing AI in breath sampling devices. Thus, this paper paves the way to implement this methodology for clinical purposes by improving the development of standard procedures for selective breath sampling.

**Acknowledgments.** The authors thank all volunteers for participating in the study. The work benefitted from the continuous support of the combined effort of NOVA School of Science and Technology and NMT, S.A. Partial support came from Fundação para a Ciência e Tecnologia (FCT, Portugal) through the PhD grant (PD/BDE/114550/2016).

## References

1. Moura, P., Vassilenko, V., Fernandes, J., Santos, P.: Indoor and Outdoor Air Profiling with GC-IMS. In : Technological Innovation for Life Improvement, DoCEIS 2020. IFIP Advances in Information and Communication Technology, 2020, Caparica, Portugal, vol. 577, pp.437-444 (2020)
2. M., P., Basa-Dalay, V., Blais, J., Brothamley, G., Chaturvedi, A., Modi, K., Pandaya, M., Natividad, M., Patel, U., Ramraje, N., Schmidt, P., Udwardia, Z.: Point-of-care breath test for biomarkers of active pulmonary tuberculosis. *Tuberc.* 92, 314-320 (2012)



3. Van Berkel, J. J. B. N., Dallinga, J. W., Möller, G. M., Godschalk, R. W. L., Moonen, E. J., Wouters, E. F. M., Van Schooten, F. J.: A profile of volatile organic compounds in breath discriminater COPD patients from controls. *Respiratory Medicine* 104(Elsevier), 557-563 (2010)
4. Lacy Costello, B., Amann, A., Al-Kateb, H., Flynn, C., Filipiak, W., Khalid, T., Osborne, D., Ratcliffe, N. M.: A review of the volatiles from the healthy human body. *Journal of breath research* 8-014001 (2014)
5. Gonçalves, M., Fernandes, J., Fetter, V., Diniz, M., Vassilenko, V.: Novel methodology for quick detection of bacterial metabolites. In : 2019 IEEE 6th Portuguese Meeting on Bioengineering (ENBENG), Lisbon, Portugal (2019)
6. Conduto, C., Vassilenko, V.: Desenvolvimento do dispositivo para teste não-invasivo da infecção por bactéria *H. pylory*. Master Thesis, NOVA University of Lisbon (2018)
7. Xu, M., Zhentao, T., Yixiang, D., Yong, L.: GC-Based Techniques for Breath Analysis: Current Status, Challenges, and Prospects. *Critical Reviews in Analytical Chemistry* 46, 291-304 (2016)
9. Phillips, M.: Method for the collection and assay of volatile organic compounds in breath. *Analytical Biochemistry* 247, 272-278 (1997)
10. Vassilenko, V., Santos, P. H. C.: System for controlled and selective sampling of exhaled air and corresponding operating procedure. WO/2018/047058, 12 09 2016
11. Santos, P. H. C., Vassilenko, V., Vasconcelos, F., Gil, F.: Implementation of Machine Learning for Breath Collection. In : BIODEVICES 2017 - 10th International Conference on Biomedical Electronics and Devices, Proceedings; Part of 10th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2017, Porto, Portugal, pp.163-170 (2017)
12. Santos, P., Roth, P., Fernandes, J., Fetter, V., Vassilenko, V.: Real Time Mental Stress Detection Through Breath Analysis. In Springer, C., ed. : *Technological Innovation for Life Improvement*. DoCEIS 2020. IFIP Advances in Information and Communication 2020, vol. 577, pp.403-410
13. Phillips, M., Cataneo, R., Chaturvedi, A., Kaplan, P. D., Libadoni, M., Mundada, M., Patel, U., Zhang, X.: Detection of an Extended Human Volatome with Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry. *PLoS One* 8(9):e75274 (2013)
14. Kanu, A., Hill Jr., H.: Ion mobility spectrometry detection for gas chromatography. *Journal of Chromatography A*, 12-27 (2008)
15. Fernandes, J. M., Santos, P. H. C., Vassilenko, V.: Algorithm for Automatic Peak Detection and Quantification for GC-IMS Spectra. In : *Technological Innovation for Life Improvement*, DoCEIS 2020. IFIP Advances in Information and Communication Technology 2020, Costa de Caparica, Portugal, vol. 577, pp.369-377 (2020)
16. Wondka, D., Bhatnagar, A.: Selection, Segmentation and Analysis of Exhaled Air for Airway Disorders Assessment. (US Patent WO/2015/143384, 20 03 2015)