Implementation of Machine Learning for Breath Collection

Paulo Santos[1,2], Valentina Vassilenko[1,2], Fábio Vasconcelos[1] and Flávio Gil[2]
[1]*Laboratory for Instrumentation, Biomedical Engineering and Radiation Physiscs (LibPhys-UNL), Faculdade de Ciência e Tecnologias da Universidade NOVA de Lisboa, Campus FCT UNL, 2896-516 Caparica, Portugal*
[2]*NMT, S.A., Edíficio Madan Parque, Rua dos Inventores, 2825-282 Caparica, Portugal*

Keywords: Exhaled Air, Selective Air Acquisition, Air Sampling/Monitoring, Breath Rhythm Imposition, Modelled Breath Algorithm, Average Time of Expiration, Machine Learning Algorithm.

Abstract: Economic and technologic progresses states the analysis of human's exhaled air as a promising tool for medical diagnosis and therapy monitoring. Challenges of most pulmonary breath acquisition devices are related to the substances' concentrations that are source (oral cavity, esophageal and alveolar) dependent and their low values (in ppb_v - ppt_v range). We introduce a prototype that is capable of collecting samples of exhaled air according to the respiratory source and independent of the metabolic production of carbon dioxide. It also allows to access the breathing cycle in real-time, detects the optimized sampling instants and selects the collection pathway through the implementation of an algorithm containing a machine learning process. A graphical interface allows the interaction between the operator/user and the process of acquisition making it easy, quick and reliable. The imposition of breath rhythm led to improvements in accuracy of obtaining samples from specific parts of the respiratory tract and it should be adapted according to their age and physiological/health condition. The technology implemented in the proposed system should be taken into consideration for further studies, since the prototype is suitable for selectively sampling exhaled air from persons according to its age, genre and physiological condition.

1 INTRODUCTION

The detection and measurement of exhaled substances is advantageous as a reliable, reproducible and non-invasive diagnostic and prognostic tool in a wide variety of medical conditions to assess different vital organ functions (Miekisch *et al.*, 2004; Baumbach *et al.*, 2009; Di Francesco *et al.*, 2005; Manolis *et al.*, 1983; Miekisch *et al.*, 2006; Amman *et al.*, 2007; Dweik *et al.*, 2008). The human exhaled air accommodates a complex mixture of molecules which are expelled in every breath (~75% of nitrogen, ~15% of oxygen, ~5 of carbon dioxide (CO_2) and ~6% of water vapour, inorganic compounds, volatile organic compounds (VOCs) and aerosols). By measuring the concentration of those molecules, it is possible to quantify each person's individual score reflecting the state of health (Lourenço *et. al*, 2014).

There are different main targets in the analysis of exhaled air capable to identify potential diseases, but VOCs are the most studied and interesting to look for as biomarkers of pathological conditions (Miekisch *et al.*, 2004; Baumbach *et al.*, 2009; Di Francesco *et al.*, 2005; Manolis *et al.*, 1983; Miekisch *et al.*, 2006; Amman *et al.*, 2007; Dweik *et al.*, 2008; Lourenço *et al.*, 2014; Mazzatenta *et al.*, 2013). The concentration of these compounds in the exhaled air varies depending on the respiratory origin of exhaled air to be analyzed, including oral cavity, esophageal and alveolar air (Phillips *et al.*, 1999; Di Natale *et al.*, 2014; Ruzsanyi *et al.*, 2013).

Furthermore, the concentration of most of the VOCs present in the exhaled air is very low (ppb_v – ppt_v or μgl^{-1} – ngl^{-1} range). Thus, the detection of such small amounts in fractions of exhaled air from different respiratory origins has revealed itself one of new challenges to overcome in the most recent pulmonary breath sampling devices.

Even though there are several studies in this field, the clinical importance of these compounds is yet to be discovered. This work does not aim to evaluate any group of compounds or specific VOCs. Instead, focus will be given to the process of exhaled air sampling according to user's characteristics by evaluating the influence of imposing a controlled breath rhythm. These aspects obviously can influence the studies involving the analyses of samples containing these compounds.

Santos P., Vassilenko V., Vasconcelos F. and Gil F.
Implementation of Machine Learning for Breath Collection.
DOI: 10.5220/0006168601630170
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017)*, pages 163-170
ISBN: 978-989-758-216-5

1.1 Breath Sampling Devices

Multiple apparatus and methods are used in breath studies, which in general are performed with patients either providing their breath for storage and subsequent later analysis (offline analysis) or breathing directly into the analyzer for immediate analysis (online analysis). Different approaches are used, depending on the breath constituents to analyze, being gas-phase the more indicated for VOC analysis (Beauchamp *et al*., 2015). Nevertheless, controlling breath sampling is crucial to enable the identification of part of the respiratory tract from which the sample derived and to ensure that comparative data is generated between studies. The control is actually made by using CO_2, flow, pressure temperature or humidity sensors (Beauchamp *et al.*, 2015).

However, there are several constraints while sampling the exhaled air for analysis (Alonso *et al*., 2013). Despite different breath sampling methods/devices used and several types of analysis developed in the last decade, most of them are lacking in accuracy and precision in the collection of an exhaled air sample (Alonso *et al*., 2013). The technology used in exhaled air collection, introduce high variability in breath samples due to: the way the air is expelled, the breath frequency, the length and depth of the breath cycle and the mental and physical condition of the patient (Basanta *et al*., 2007; Droz *et al*., 1986; Risby, 2008).

1.2 Respiratory Cycle Monitoring

The real time monitoring of the patient's respiratory cycle allows the identification of respiratory phases and the definition of the instants for breath collection. The identification of the alveolar air portion assumes critical importance due to the presence of diverse constituents from endogenous origin in equilibrium with the alveolar capillary blood vessels.

Presently the most used method for identification of the different phases of the breathing cycle is the capnography which allows to monitor the concentration or partial pressure of carbon dioxide (PCO_2) in the respiratory gases by providing information about the production of carbon dioxide, pulmonary perfusion, alveolar ventilation and respiratory patterns (Bhavanishankar *et al.*, 1995; Mimoz *et al*., 2012; Bhavanishankar *et al*., 1992).

Yet, the use of capnography has some limitations regarding the collection of selective samples of exhaled air, because it varies with (a) the inherent variation of breath composition and concentration of each constituent throughout the breathing cycle; (b) the speed of the breath, which affects the composition of the mixture between alveolar air and dead space air; (c) the depth and frequency of breathing, which control changes from autonomous to conscious breathing, when a person is asked to provide a sample of breath.

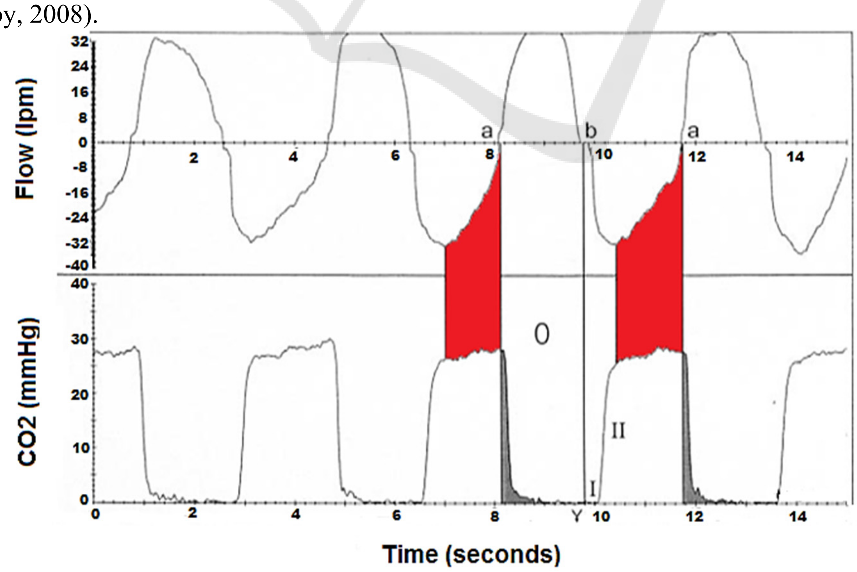

Figure 1: Graphical comparison between respiratory flow rate (at the top) and time capnogram (at the bottom). The ab segment represents inspiration and the ba segment represents expiration on the respiratory waveforms. The red area represents the alveolar air region, while the shaded area under the CO_2 curve represents the inspiratory phase of the respiratory cycle, thus constituting rebreathing. (Adapted from Bhavani-Shankar and Philip, 2000).

Despite the possibility of diagnose several diseases, the modifications in the shape of capnograms (related with such diseases) also difficult clear identification of the expiratory segments (Kodali et al., 2013).

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, the actual research in breath analysis tries to pursue a suitable device and a precise protocol for sampling exhaled air independently of the subject's metabolic production of CO_2, the smoking habits, the type of food eaten, the stomach, esophagus and mouth condition of the patients.

Bear that in mind, this particularly work aims to describe the influence of the implementation of machine learning for selective breath sampling by using a novel technology where a respiratory cycle model is adapted to individual breathing characteristics of the user.

2 BREATH MODELLING & ITS IMPLEMENTATION

Due to the above mentioned limitations of the capnography and, since the measurement of the respiratory flow rate may yield the same effect of determining the respiratory phases in a cheaper and easier way, both were compared to identify the region corresponding to the alveolar air and to modulate it in a mathematical function. By overlapping a time capnogram with a fluxogram (Bhavani-Shankar and Philip, 2000), the area related to the end-tidal breath was clearly identified (figure 1). Considering that, only a flowmeter can be used for selective assessment to the last segment of the expiration in a fluxogram which corresponds to the alveolar region with higher CO_2 concentration.

2.1 Respiratory Cycles' Modelling

The collection of selective portions of exhaled air by using respiratory flow measurements, implies the use of reference respiratory rhythms for the users to follow. By measuring the exhaled air flow of multiple subjects and by determining the total time of each respiratory cycle and the transition between respiratory phases (inspirations and exhalations), Vassilenko *et. al.* (2013) were able to calculate average signals that best describe three breathing rhythms (slow, normal and fast). The average signals for the three respiratory rhythms are shown in figure 2 and were used as references for development of mathematical models which precisely characterises breath rhythms of every patient (Vassilenko *et al.,* 2013).

Since it overcomes the disadvantages regarding the imprecise identification of selective portions of exhaled air (associated with variable depth and frequency of breathing), the proposed method for monitoring and selective sampling of exhaled air through respiratory flow sensing represents a reliable alternative method to capnography approaches.

2.2 Implementation of Respiratory Cycles' Models

The prototype developed by the authors performs real time flow measurements and captures alveolar air by synchronizing modelled respiratory cycles with the user's breathing cycle. The prototype comprises hardware cells controlled by an intelligent control software loaded on several computing devices

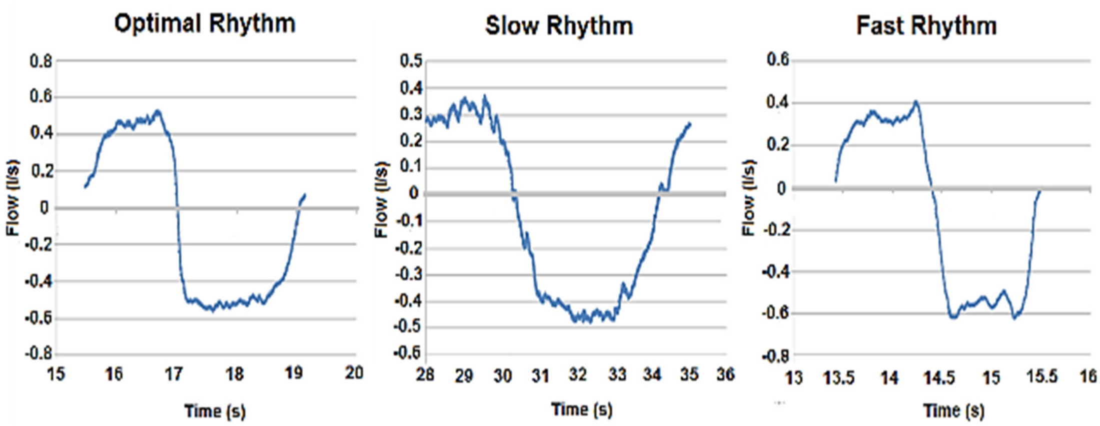

Figure 2: Representation of the average signals obtained for each pace imposed to tested individuals (Vassilenko *et al.,* 2013).

(laptops, desktops, smartphones, tablets, *etc*.). The hardware module is responsible for data acquisition, processing and its transmission to the software, and for channeling the portion of exhaled air through the sampling or elimination outlets.

The software comprises a graphical interface that imposes a breathing pace to the user according to its age, gender and physiological state and, by using an algorithm identifies the instants of sampling and communicates with the hardware to trigger the sample to be stored either in a bag or go directly to an analytical analyser. Both of that software components were updated with the implementation of a machine learning process in the algorithm and the imposition of breath rhythm to the user according to its age.

The algorithm implemented in the intelligent control software was configured to: measure the user's respiratory flow; detect user's breathing frequency; distinguish inspiratory and expiratory breath phases; synchronize the user's respiratory cycle with the representative and modeled respiratory cycle; and to calculate the average time of expiration of the user. The average time of expiration allows the selection the fraction of exhaled air to sample.

The machine learning process implemented in the algorithm of the software is based on the continuous calculation and saving of the average exhalation time values, allowing the prediction of the time of occurrence of a new expiration and, consequently, the prediction of the precise time-frame for the acquisition of the fraction of exhaled air to sample. By this way, the machine learning process learns the respiratory cycle of the user, and test it on the modelled respiratory intrinsically contained in the algorithm of the software.

In addition to the definition of global variables of the operation and from the subject (genre, age and physiological/health condition), the graphical interface also provides a feedback mechanism for communicating with the user/operator. This feedback mechanism presents a central part of the system because it provides multiple indicators for showing the breathing rhythm to be followed by the user or if the moments of breath air acquisition are occurring.

According to the information defined in the graphical interface, the user is asked to breathe according to a specific respiratory rhythm (figure 3). When the breathing pace of the user is matched the representative and modelled respiratory cycle, the initial and final instant of the exhaled air's fraction of interest is identified in the respiratory cycle. This information is communicated with the remaining system of device in order to sample the portion of interest of exhaled air between these instants. This process ensures that only a fraction of exhaled air is diverted to a collection reservoir or directly analysed by an analytical analyser.

3 TESTS OF PERFORMANCE

To evaluate the effectiveness of the machine learning process implemented on the software of the prototype and the influence of a breath rhythm imposition, two groups of individuals with different age groups (15 patients between 2 and 5 years old – children – and 30 within 18 and 27 years old – university students) were asked to make breathing test in the prototype. The patients had to achieve the beginning of breath collection and, simultaneously, the minimum number

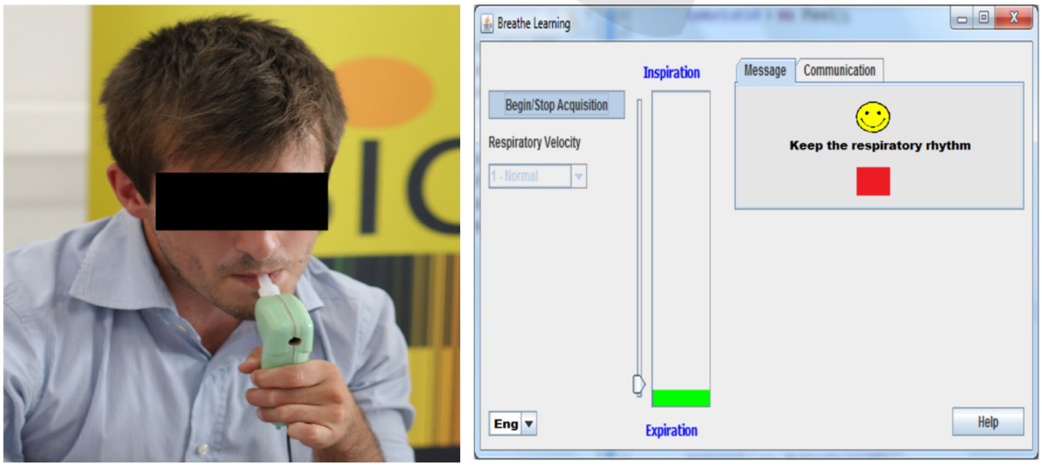

Figure 3: Prototype for alveolar air collection used on experimental tests (on the left) and the graphical interface related with breath rhythm imposition to the user (on the right).

of cycles, the time required to start breath sampling and the average time of exhalation (ATE) were registered. This method was applied two times for each individual, in which firstly patient's autonomous breath rhythm was suggested and then with a respiratory pace imposition to the subject through the feedback present on the graphical interface.

3.1 Number of Respiratory Cycles

The results show that the number of cycles needed to begin breath sampling are lower when a breath rhythm was imposed to both groups of patients. More specifically, when the university students breathed autonomously, the number of cycles registered till the acquisition started are higher (9.70 ± 2.22; mean ± standard deviation) comparably with the same number for an imposed breathing (8.56 ± 2.12). The results are more evidently with children when comparing the number of respiratory cycles needed to initiate the sampling by an autonomous breathing (17.61 ± 3.31) and an imposed one (13.93 ± 2.49).

3.2 Average Time of Exhalation

The results presented in figure 4 illustrate the comparison of the average time of exhalation (ATE) between an imposed/controlled breathing rhythm and an autonomous rhythm of the patients, and the relationship of such feature for two aging groups.

For both aging groups (children and university students), when respiratory rhythm was autonomous, the distribution of ATEs is significantly uneven when compared with the distribution of ATEs for an imposed breathing rhythm. The values of the standard deviation for an autonomous breathing (424 and 966 ms, for children and university students, respectively) are almost 4 times higher when compared with the corresponding values of standard deviation for the imposed rhythm (120 and 250 ms, for children and university students, respectively). For both acquisition methods, the average values of ATEs are also significantly lower for children (974 and 1032 ms, for autonomous and controlled breathing, respectively) comparably to the older university students (2031 and 1752 ms, for autonomous and controlled breathing, respectively).

3.3 Time Required to Start Breath Sampling

Figure 5 displays box and whiskers plots related to the time necessary to begin sampling the portion of interest of exhaled air in order to distinguish both procedures of breath sampling (with and without an imposed rhythm) for both aging groups.

The time required to start breath sampling presents several differences regarding the type of breathing applied to the patients. For children (2-5 years old), the time needed to begin sampling the portion of interest of exhaled air with a controlled rhythm of breathing (32.89/31.88-34.60/s; median/interquartile range/) is lower when compared with an autonomous breath rhythm of the patient (35.51/33.65-41.06/s). However, this decrease in the

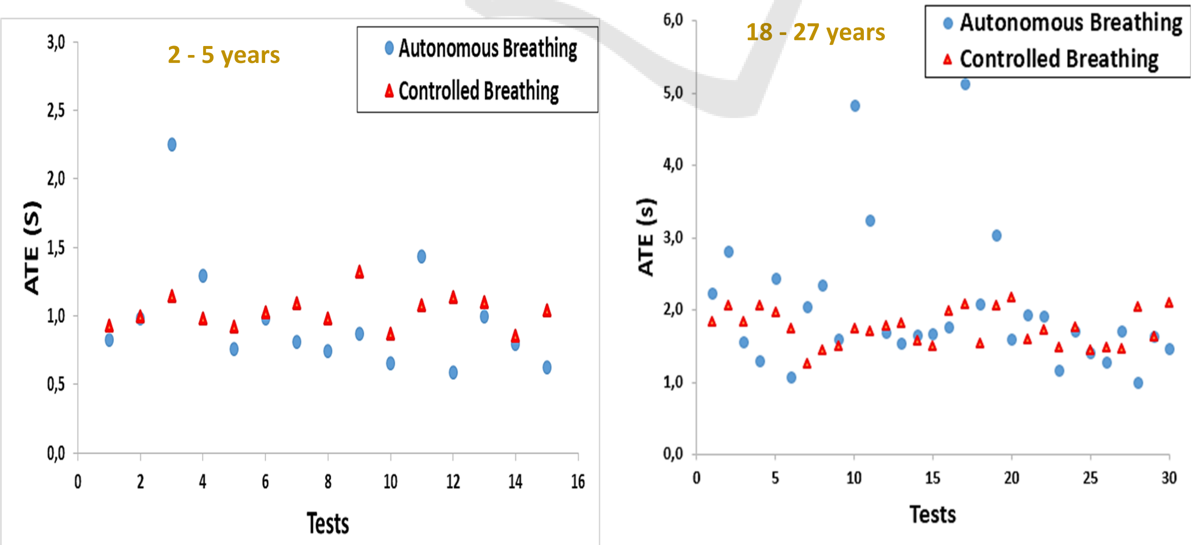

Figure 4: Distribution of results of average time of exhalation (ATE)) for free breath cycles (autonomous breathing) and for an optimized imposed rhythm (controlled breathing) with children (on the left) and university students (on the right).

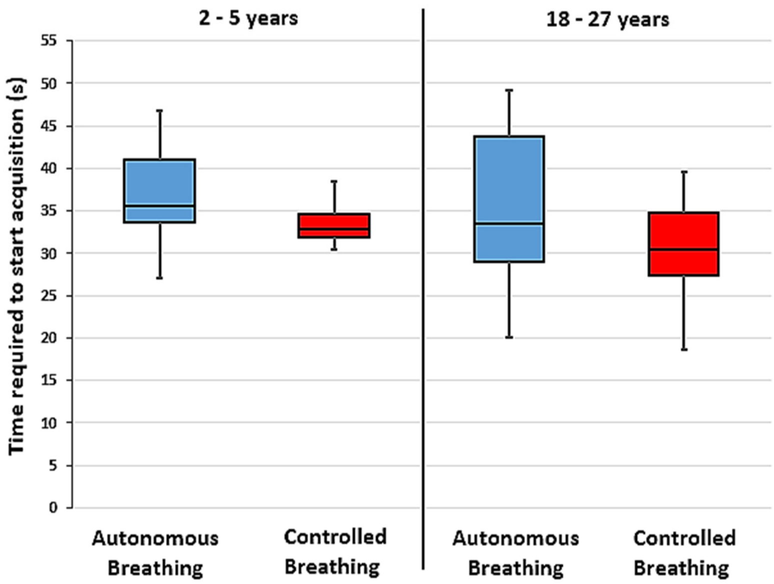

Figure 5: Time required to start breath sampling for free breath cycles (autonomous breathing) and for an optimized imposed rhythm (controlled breathing) with children (on the left) and university students (on the right). Data are displayed in box and whiskers plot (box depicts median with first and third quartiles, whiskers show first quartile – 1.5 interquartile range and third quartile+1.5 interquartile range).

time required to start the sampling is not so evidently for university students (18-27 years old) where, for an autonomous breathing (33.59/29.04-43.73/s), this required time is similar compared with time obtained for an imposed breathing rhythm (30.41/27.48-34.79). Of note, the decreased interquartile range of the time need for breath sampling with the controlled rhythm for both groups of patients compared with the time obtained for an autonomous rhythm.

4 DISCUSSION

The results of performance tests applied to the prototype show that, when concerning the number of respiratory cycles and time needed begin the breath sampling, the imposition of breath rhythm to the patients (children and university) is more efficient since less time is spent and the user is not required to make unnecessary long breaths, which can lead to fatigue (Roussos *et al.*, 1996). This increased efficiency in start of selective exhaled air sampling is related with a quicker prediction of the time of occurrence of a new expiration and, consequently, the prediction the precise time-frame for its collection which are ultimately related with the machine learning process implemented in the prototype. These results of the prototype's performance tests are more evident for children, where the suggestion of an appropriated breath rhythm have crucial importance due to their inability of autonomously maintain a breath rhythm.

The results obtained for average time of exhalation (ATE) show that the breath rhythm imposed to patients should be adapted according to their aging group and physiological/health condition. Moreover, the stabilized values of ATE and the lower interquartile range of the time required to begin sampling the portion of interest of exhaled air, for both aging groups, indicates the imposition of an aging-suitable breath rhythm as the reliable way of using the prototype for collection of exhaled air.

The majority of the existent breath samplers and ventilators comprise several algorithms to analyse respiration cycles of the user in order to detect inspiration and expiration phases and to, consequently, determine the time-window for breath sampling. However, and for cases of dyspnea with erratic respiratory rhythms, that determined time-window for sampling can be too short and can brought up multiple difficulties when obtaining such small portions of exhaled air. Only the system patented by Capnia, Inc. (patent number WO2015143384 A1, 2015) is configured to impose a breath frequency to users (young children and non-cognizant patients) in order to avoid those erratic respiratory episodes. Even so, that imposed frequency does not adapts and "learns" with the user's breathing pace such as the proposed system does.

The protocol necessary to use the prototype, in which the patient has to follow the directions given by the system and try to maintain the breathing with the same rhythm that appears on the graphical interface, also suggests the introduction of improvements in the accuracy and precision on obtaining samples of a specific part of the respiratory tract, which consequently led to the increase in the repeatability of the analysis applied to these samples. Furthermore, it is excluded the introduction of variability during breath sampling related with breathing frequency, amplitude of the respiratory cycle, the mental and physical condition of the patient, as well as, the method applied by the person who asks for the patient to breathe.

5 CONCLUSIONS

The research work demonstrated herein presents a suitable and novel technology and related protocol of using it for selectively sampling exhaled air regarding the subject's: metabolic production of CO_2, smoking habits, type of consumed food, stomach, esophagus and oral cavity conditions. Moreover, the implementation of a user-dependent's respiratory cycle model on the prototype used in this work could allow a more accurate way to collect portions of exhaled air according to the exhaled air's respiratory origin. This collection is done from single or multiple exhalations, for online or posterior analytical analysis for medical diagnosis and/or therapy monitoring, in a quick, reliable, non-invasive way, applied at any stage of life.

The imposition of a respiratory rhythm according to the characteristics of the user (age, gender and physiological/health condition) and the machine learning process implemented on the prototype led to improvements in the accuracy in sampling breath from specific parts of the respiratory tract and decreases the variability of the samples related with breath frequency, amplitude of the breath cycle, mental and physical condition of the patient.

However, the implemented algorithm have to be optimized for better performance in real healthcare environments and the respiratory rhythm appearing in graphical interface should be interactively adapted according to all age groups, especially to the elderly and children who have more difficulty to follow this method. We also believe that future and similar applications for mobile devices should be developed to help the patients to learn and train the respiratory rhythm while the respective portable sampling equipment for analysis is not commercially available. The final application should be suitable to different group stages simplifying the breath sampling process.

ACKNOWLEDGEMENTS

The authors would like to thank all volunteers that offered their time to perform tests for the acquisition of their respiratory cycles and for the tests of performance of the prototype. We thank to parents/guardians of the children who executed the same tests, for authorize their participation and to the daycare of FCT (center of pre-school education) for providing space and conditions to its implementation. The authors would also thank the Fundação para a Ciência e Tecnologia (FCT, Portugal) for co-financing the PhD grant (PD/BDE/114550/2016) of the Doctoral NOVA I4H Program.

REFERENCES

Miekisch, W, Schubert, JK & Noeldge-Schomburg, GFE 2004, 'Diagnostic potential of breath analysis - focus on volatile organic compounds', *Clinica Chimica Acta*, vol. 347, pp. 25-39.

Baumbach, JI 2009, 'Ion mobility spectrometry coupled with multi-capillary columns for metabolic profiling of human breath', *J. Breath Research*, vol. 3, pp. 16.

Di Francesco, F, Fuoco, R, Trivella, MG & Ceccarini, A 2005, 'Breath analysis: trends in techniques and clinical applications', *Microchemical Journal,* vol. 79, pp. 405-410.

Manolis, A 1983, 'The diagnostic potencial of breath analysis', *Clinical Chemistry*, vol. 29, pp. 5-15.

Miekisch, W & Schubert, JK 2006, 'From highly sophisticated analytical techniques to life-saving diagnostics: Technical developments in breath analysis', *Trac-Trends in Analyt. Chem.*, vol. 25, pp. 665-673.

Amann, A, Spanel, P & Smith, D 2007, 'Breath analysis: the approach towards clinical applications', *Mini reviews in medicinal chemistry*, vol. 7, pp. 115-129.

Dweik, RA & Amann, A 2008, 'Exhaled breath analysis: the new frontier in medical testing', *Journal of Breath Research*, vol. 2, no. 3, 030301.

Lourenço, C & Turner, C 2014, 'Breath analysis in disease diagnosis: methodological considerations and applications', *Metabolites*, vol. 4, pp. 465-498.

Mazzatenta, A, Di Giulio, C & Pokorski, M 2013, 'Pathologies currently identified by exhaled biomarkers', *Respiratory Physiology & Neurobiology*, vol. 187, pp. 128-134.

Phillips, M, Herrera, J, Krishnan, S, Zain, M, Greenberg, J & Cataneo, R 1999, 'Variation in volatile organic compounds in the breath of normal humans', *Journal of*

Chromatography B: Biomedical Science and Applications, vol. 729, pp. 75-88.
Di Natale, C, Paolesse, R, Martinelli, E & Capuano, R 2014, 'Solid-state gas sensors for breath analysis: A review', *Analytica Chimica Acta*, vol. 824, pp. 1-17.
Ruzsanyi, V 2013, 'Ion mobility spectrometry for pharmacokinetic studies-exemplary application', *Journal of Breath Research*, vol. 7, no. 4, 046008.
Beauchamp, J 2015, 'Current sampling and analysis techniques in breath research - results of a task force poll', *Journal of Breath Research*, vol. 9, 047107.
Alonso, M & Sanchez, JM 2013, 'Analytical challenges in breath analysis and its application to exposure monitoring', *Trends in Analytical Chemistry*, vol. 44, pp. 78-89.
Basanta, M, Koimtzis, T, Singh, D, Wilson, I & Thomas, CL 2007, 'An adaptive breath sampler for use with human subjetcs with an impaired respiratory function', *Analyst*, vol. 132, no. 2, pp. 153-163.
Droz, PO & Guillemin MP 1986, 'Occupational exposure monitoring using breath analysis', *J. Occup. Med.*, vol. 28, no. 8, pp. 593-602.
Risby, TH 2008, 'Critical issues for breath analysis', *Journal of Breath Research*, vol. 2, 030302.
Bhavanishankar, K, Kumar, AY, Moseley, HSL & Ahyeehallsworth, R 1995, 'Terminology and the current limitations of time capnography – A brief review', *Journal of Clinical Monitoring*, vol. 11, pp. 175-82.
Mimoz, O, Benard, T, Gaucher, A, Frasca, D & Debaene, B 2012, 'Accuracy of respiratory rate monitoring using a non-invasive acoustic method after general anaesthesia', *Br. J. Anaesth*, vol. 108, pp. 872-875.
Bhavanishankar, K, Moseley, H, Kumar, AY & Delph, Y 1992, 'Capnometry and Anesthesia', *Canadian Journal of Anaesthesia*, vol. 39, pp. 617-632.
Kodali, BS 2013, 'Capnography outside the operating rooms', *Anesthesiology*, vol. 118, pp. 192-201.
Bhavani-Shankar, K & Philip, JH 2000, 'Defining segments and phases of a time capnogram', *Anesth. Analg.*, vol. 91, no. 4, pp. 973-977.
Dias, F, Alves, J, Januário, F, Ferreira, JL & Vassilenko, V 2013, 'Prototype and Graphical Interface for Selective Exhaled Air Acquisition', *Proc. of the Intern. Conf. on Biomedical Electronics and Devices,* vol. 1, pp. 216-219.
Roussos, C & Zakynthinos, S 1996, 'Fatigue of the respiratory muscles', *Intensive Care Med.*, vol. 154, pp. 1099-1105.
Capnia, Inc. 2015, 'Selection, segmentation and analysis of exhaled breath for airway disorders assessment', WO2015143384 A1.