Self-monitoring the Breath for the Prevention of Cardio-metabolic Risk

Danila Germanese
Massimo Magrini,
Marco Righi and
Ovidio Salvetti

Institute of
Information Science and Technology,
ISTI-CNR, Pisa, Italy
mail: danila.germanese@isti.cnr.it
mail: massimo.magrini@isti.cnr.it
mail: marco.righi@isti.cnr.it
mail: ovidio.salvetti@isti.cnr.it

Mario D'Acunto

Biophysics Institute
IBF-CNR, Pisa, Italy
mail: mario.dacunto@cnr.it

Abstract—Breath analysis techniques offer a potential revolution in health care diagnostics because of their un-obtrusiveness and their inherent safety. However, while standard instrumentation such as mass spectrometers use laboratory settings to provide a correlation between exhaled substances and physical conditions, to fully realize the potential of breath analysis as a self-monitoring tool, its application must take place also in the clinics and at home and not only in a laboratory. This basic requirement has stimulated the necessity to develop cheap, portable, real time, easy-to-use devices for reliable breath tests and analysis. In this paper, we present the design of a portable breath analyzer, able to sense a set of breath volatile organic compounds (VOCs), to perform a processing of the data collected and to generate an output easily interpreted both by physicians and patients.

Keywords–Breath analysis; Gas sensors; Self-monitoring; Home-care; Portable device; Signal processing.

I. INTRODUCTION

Breath analysis may play a key role in health care diagnostics [1]. This is because each breath contains fundamental information about the health status of an individual: breath molecules are the product of the composition of inspiratory air and the volatile substances in blood. In addition, also cells in the mouth, in the upper airways, in the gastrointestinal tract contribute volatile molecules to the exhaled breath. Consequently, the challenge is to extract from the breath meaningful data which can be correlated to subject's health.

On one hand, despite its great potential, breath analysis has not yet been employed in the ordinary diagnostic clinical trials. First of all, the main bottleneck is the lack of standardized protocol to collect breath sampling and to avoid all the confounding factors (such as inspired ambient air [2], breath flow rate, heart rate [3]). Moreover, the standard instrumentation (gas chromatography-mass spectrometry, for instance) is very expensive, time consuming and its use often requires highly qualified personnel. On the other hand, the main advantage of breath analysis is its un-obtrusiveness and safety. As a consequence, it may be a very suitable diagnostic tool, especially for those people who have to control a set of parameters daily.

These requirements have solicited the necessity to develop cheap, portable, real time, easy-to-use devices for breath analysis, in order to promote not only its purchase, but also its use in every type of setting (in home environment, for instance). In human breath, more than 200 volatile molecules have been identified and assessed. Some of such molecules were correlated to various diseases such as diabetes, oxidative stress, lung cancer, gastrointestinal diseases, etc. [4][5][6][7]. Due to such complex pattern of breath compounds, a design of a portable device for breath analysis should be based on selected chemical sensors able to sense specific VOCs. Generally, a long-term vision for breath analysis performed with a portable device should follow some basic requirements:

- the breath analysis device should be compact, easy to use, and able to follow, in real time, the breath molecules trend;
- the device should be able to transmit breath analysis results also to a remote personal computer (the family doctor's one, for instance);
- since it should be based on array of gas sensors, cross-correlation between sensors should be carefully taken in account to improve the sensitivity and the reliability of the overall device.

Recently, e-noses have gone in this direction. Formerly designed for broader applications (environmental gases monitoring, for instance), in recent years the idea of exploiting e-noses also for clinical applications has gained increased interest [8]. E-noses allow for performing breath analysis in a very short time, being quicker than a gas chromatograph. Since they are able to perform breath tests in real time, in many studies they have been employed to monitor volatile biomarkers related to cancer [9], for instance, in infectiology [10], and also to evaluate VOCs related to asthma [11]. Nevertheless, the majority of such e-noses exploit very expensive technology [12][13] or require complex circuitry [14][15].

In this paper, we describe how self-monitoring of someone's own well-being state could be done by means of a low cost device (called Wize Sniffer, WS) whose basic features

have been described in [16][17]. In particular, the WS was designed to detect a set of breath molecules related to cardio-metabolic risk. Nevertheless, its modular configuration allows for detecting a broader set of molecules, simply changing the gas sensors placed in gas samplig box. The WS is entirely based on low-cost technology: the semiconductor-based gas sensors are commercial, and breath signals are analyzed by a widely employed open source controller: Arduino Mega2560. In addition, it is programmed to also send breath analysis results also to a remote care center.

In the paper, Section II summarizes the detected VOCs; in Section III, the hardware/software architecture is described; Section IV reports the WS functionality tests and the different data analysis approaches. We conclude the paper in Section V.

II. Cardio-metabolic risk prevention

The WS was conceived in the framework of SEMEOTI-CONS (SEMEiotic Oriented Technology for Individuals CardiOmetabolic risk self-assessmeNt and Self-monitoring, grant N. 611516) European Project [18], which aimed to develop a multi-sensory platform able to assess individuals well-being state by detecting in the human face all those signs related to cardio-metabolic risk [19]. Such multisensory, interactive platform included a sensorized Mirror (the *Wize Mirror*) and the WS. In particular, the WS was designed to help the user monitor his/her noxious habits for cardio-metabolic risk, by detecting the following VOCs:

- **Carbon monoxide** (CO). More than 5000 compounds in cigarette smoke are dangerous. CO, in particular, decreases the amount of oxygen that is carried in the red blood cells. It also increases the amount of cholesterol that is deposited into the arteries;

- **Ethanol** (C_2H_6O). Moderate ethanol consumption, in healthy subjects, reduces stress and increases feelings of happiness and well-being, and may reduce the risk of coronary heart disease. Heavy consumption of alcohol, instead, causes addiction and leads to an accumulation of free radicals into the cells, causing oxidative stress.

In addition, the device can also provide useful information about metabolism, user's carbohydrates adsorption and vascular status by detecting these other molecules:

- **Oxygen and carbon dioxide** (O_2 and CO_2): the amount of O_2, which is retained in the body, and the one of CO_2, which is produced as a by-product, can be considered as a measure of the metabolism;

- **Hydrogen** (H_2): it is related to the carbohydrates breakdown in the intestine and in the oral cavity by anaerobic bacteria;

- **Hydrogen sulfide** (H_2S): it is a vascular relax agent; for instance, it has a therapeutic effect in hypertension.

III. Hardware and Software architecture

A. Wize Sniffer's sensor platform

The core of a portable device designed to detect volatile molecules (wheather they derive from ambient air, for example, or human exhaled breath) is the gas sensors array. For this purpose, different technologies and sensors' transduction principles are exploited to assess the type and the concentration of the gases under investigation [8]. Regarding the WS, our aim was to find a trade off between good sensitivity, low cost and small dimension. As we mentioned in the previous section, the WS was developed to detect a set of molecules related to those noxious habits for cardio-metabolic risk; nevertheless, our aim was to design a modular sensor platform in order to detect a broader set of molecules, by simply changing the sensors according to the VOCs to be identified. As a consequence, the sensors' ease of integration in the circuitry was another requirement we needed.

On one hand, optical gas sensors, as well as quartz crystal microbalance (QCM)-based gas sensors and surface acoustic wave (SAW)-based gas sensors are very sensitive; on the other hand, they are expensive (especially in the case of optical gas sensors) and need complex circuitry (in the case of QCM and SAW gas sensors). Also, carbon nano-fiber (CNF) based gas sensors are very expensive, especially for their manufacturing. We chose metal oxide semiconductor (MOS)- based gas sensors: they are low cost and easy to integrate in the circuitry; they have very small dimension, long life and rapid recovery. In Table I, all the employed MOS-based gas sensors are listed.

TABLE I. Sensors integrated in the WS sensor platform

Detected molecule	Sensor	Best detection range
Carbon monoxide	MQ7	20-200 ppm
	TGS2620	50-5000 ppm
Ethanol	TGS2602	1-10 ppm
	TGS2620	50-5000 ppm
Carbon dioxide	TGS4161	0-40000 ppm
Oxygen	MOX20	0-16%
Hydrogen sulfide	TGS2602	1-10 ppm
Hydrogen	TGS821	10-5000 ppm
	TGS2602	1-10 ppm
	TGS2620	50-5000 ppm
	MQ7	20-200 ppm

Unfortunately, humidity and cross-sensitivity strongly affect the behavior of MOS-based gas sensors [20]. In our case, humidity is a strong influencing parameter, as we deal with human breath. For this reason, we i)integrated a temperature and humidity sensor into the gas sampling box (Sensirion SHT11); ii)put a heat and moisture exchanger (HME) filter at the mouthpiece to reduce the contribution due to the water vapor from 90%RH to 65-70%RH; iii)investigated the behavior of Wize Sniffer's sensors both in response to a humidity variation and under precise measurement conditions: 30C+/-7% and 70%RH+/-5%, that are the ones that occur in the gases store chamber when a breath analysis is performed, as shown in Figure 1. Calculating the sensors' humidity drift is useful to potentially compensate it during the data processing. Figure 2 shows how the humidity strongly affects sensors' output (in this case, the one of MQ7 gas sensor). The relationship between humidity and sensors' output generally can be modeled by means of a power law:

$$V_{out} = f(hum) = a * (hum^b) + c \qquad (1)$$

where a and c are constant. In addition, we considered the entire range of humidity variation (for instance, 50%-55% Relative Humidity (RH) in the case of MQ7, as shown in Figure 2) and then, we calculated the slope of the curves. Based on the slope, drift coefficients were assessed (see Table II) as the decrease in sensors' output (Volt) per unit decrease in humidity, as given in (eq. 2):

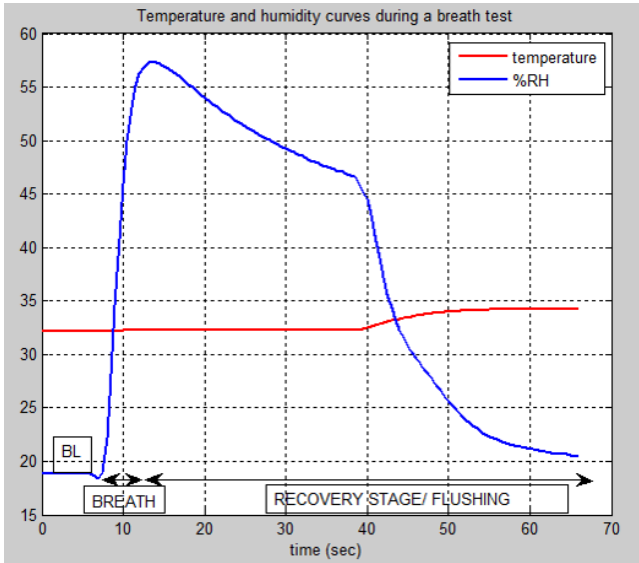$$S_d = \Delta V / \Delta hum \qquad (2)$$

Figure 1. Temperature and relative humidity in the gas sampling box when a breath analysis is performed.
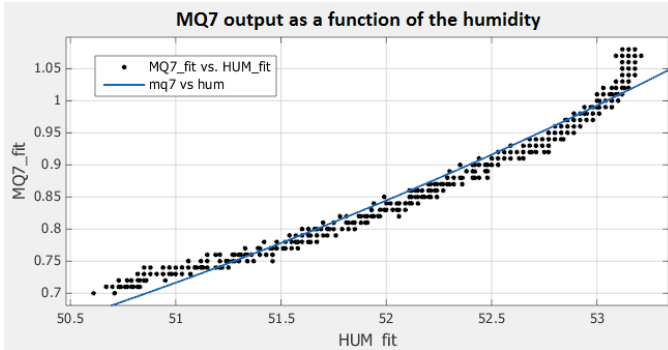

Figure 2. MQ7 output when a rise in humidity occurs.

TABLE II. Sensors' drift due to humidity

Sensor	$\Delta V/\Delta hum$ (mV)
MQ7	296
TGS2620	60
TGS2602	82
TGS821	120
TGS2444	84

By keeping the humidity constant, sensors' output will depend on the gas concentration only. For this purpose, we investigated the sensors' output in response to a well-known gases concentration. The sensors were put into a vial. The humidity into the vial was kept at 70%RH+/-5% by means of a saturated solution of NaCl placed on the bottom; then, we injected well-known gases concentration and registered sensors' output. The raw sensors'output were read by an Arduino Mega2560 connected via serial port to a personal computer. The experimental data were displayed in real time on the computer screen and stored as text files for later processing. For example, in Figure 3, we can see TGS2620 output when well-known concentration of carbon monoxide, ethanol and hydrogen were separately injected into the vial. Also in this case, the relationship between sensors' output and gases con-

centration can be modeled by means of an equation similar to eq.1. Nevertheless, when a breath analysis is performed, a mix-
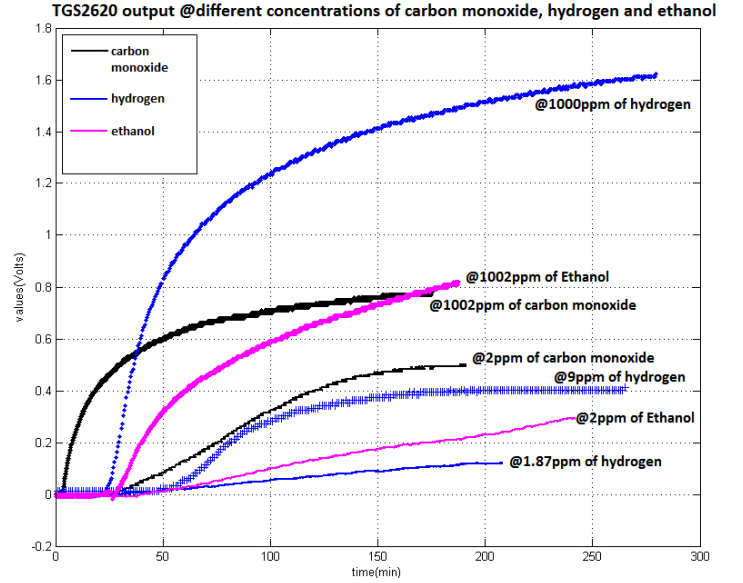

Figure 3. TGS2620 output when a well-known concentration of carbon monoxide, hydrogen and ethanol were injected into the vial.

ture of gases spreads into the gas sampling box and chemically interacts with the sensors. In this case, the phenomenon known as *cross sensitivity* makes these sensors non selective. Such behavior affects the method for data analysis, as described in Section IV.

Finally, the final set-up of the device is shown in Figure 4 and Figure 5. In the first one, the internal configuration of the device is reported: the gas sampling box, on the right, has a capacity of 600 ml according to the tidal volume [21]. The MOS-based gas sensors are placed within such box. Sensors' output is read and pre-processed by a widely used open source controller: an Arduino Mega2560. In Figure 5, the two configurations of the WS are shown: the WS can work both as a Wize Mirror tool and as a stand-alone device. In both cases, the user blows into a disposable mouthpiece, where a HME filter is placed. A flowmeter monitors the exhaled breath volume. Breath gases reach the gas sampling box by means of a corrugated tube. A fan is switched on between two breath tests to purge the gas sampling box with ambient air and to recovery sensors' steady state.

B. WS Software

Given its unobtrusiveness and its safety, breath analysis may be used as a daily monitoring analysis tool. To fully exploit its potential, its application must take place not only in laboratory settings, but also in the clinics.

In addition, our aim was to develop a device which could be used also in home environment and which could be able to send breath analysis results to a remote personal computer (for instance, to the one of the own family doctor) thus promoting independent living in community-based, home, and long-term care settings [22]. Arduino samples sensors' signals every 250 ms, saves raw vector data and extracts the maximum value of raw breath curve. Several parameters and features can be derived from breath curves [23] to fully characterize
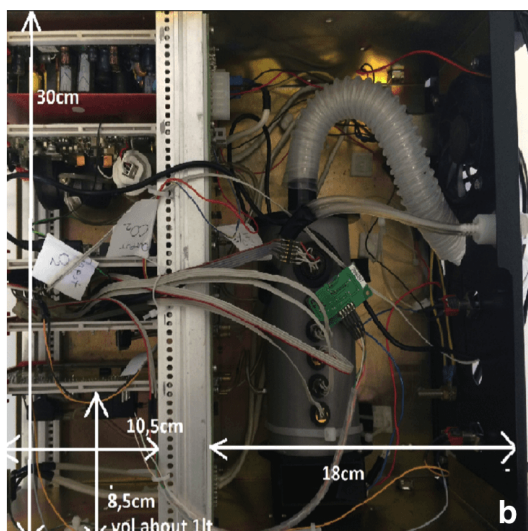
Figure 4. Final set-up of the Wize Sniffer. Internal configuration.

Figure 5. Final configuration of the Wize Sniffer. On the left, the WS is used a Wize Mirror tool. On the right, it is used in a stand-alone configuration.

them. We chose to calculate the value at curve plateau as it better describes the chemical balance between sensors'sensing element and target gases. Such data are then processed and analyzed, as described in Section IV.

In order to send breath analysis results also to a remote personal computer, we implemented a client-server architecture. It means that, after performing a test and processing the results, the device, by means of an internet connection and a TCP/IP communication protocol, can send the results to the physician, for instance. Arduino is programmed to execute a daemon on port 23. By implementing a Telnet server, it waits for a command line from the remote personal computer and then can provide the data. A measure is valid if the user's exhaled volume equals at least the one of gas sampling box.

IV. FUNCTIONALITY TESTS

The aim was to assess if the WS was able to monitor and evaluate the individuals' noxious habits for cardio-metabolic risk (smoke and alcohol intake in particular).

As described in [24], the WS underwent a clinical validation in three research centers: CNR in Pisa and Milan, CRNH (Centre de Recherche en Nutrition Humaine) in Lyon. The validation campaign involved 77 volunteers, with different habits and lifestyle. The subjects had to answer Audit test and Fagerstrom test, which respectively assessed the alcohol and smoke dependence, and other questionnaires about their lifestyle.

Taking into account the methodological issues about breath sampling [25], we outlined a measuring protocol, which considered mixed expiratory air sampling, since our interest was focused on both endogenous and exogenous biomarkers. The subjects took a deep breath in, held the breath for 10s, and then exhaled once into the corrugated tube trying to keep the expiratory flow constant and to completely empty their lungs. The study was approved by the Ethical Committee of the Azienda Ospedaliera Universitaria Pisana, protocol n.213/2014 approved on September 25th, 2014; all patients provided a signed informed consent before enrollment.

As mentioned before, MOS-based gas sensors are strongly affected by cross-sensitivity. It means that such type of sensors is not selective for one substance only, but they are sensitive for a set of VOCs. Such characteristic makes the quantitative analysis of the detected VOCs very difficult. As a consequence, we exploited another approach for data analysis, more classical, based on multivariate methods of pattern recognition. Pattern recognition exploits sensors' cross-correlation and helps to extract qualitative information contained in sensors' outputs ensemble. Then, first Principal Component Analysis (PCA) was performed, in order to provide a representation of the data in a space of dimensions lower than the original sensors' space. After assessing, by the PCA, the presence of clusters (see Figure 6), the data were processed with a K-nearest neighbor (KNN) classification algorithm, previously trained with the data coming from another acquisition campaign. The aim was to classify the subjects according to their habits: Healthy (that means, no cardio-metabolic risk), Light Smoker, Heavy Smoker, Social Drinker, Heavy Drinker, LsSd (Light smokers, Social drinker), LsHd (Light smokers, Heavy drinkers), HsSd (Heavy smokers, Social drinker), HsHd (Heavy smokers, Heavy drinker). The Audit and Fagerstrom questionnaires were our ground truth. The KNN classifier was able to correctly classify in 89,61% of cases. Errors were due to TGS2602 and TGS2620 cross-sensitivity for hydrogen. In fact, for instance, three no-risk subjects were classified as social drinker probably because of high hydrogen contribution, which caused a rise in these sensors voltage output.

Then, the population under study was increased, up to 169 subjects. Figure 7 provides a summary of subjects' habits and lifestyle in general. In particular, the subjects were divided into "low risk population", "medium risk population", "high risk population" basing on the sum of scores relative to Audit (AS), Fagerstom (FS) and lifestyle questionnaires, which were our ground truth also in this case. Also, the measuring protocol for breath sampling was the same, as well as data pre-processing: the parameter extracted from raw breath curve by Arduino was the value at the curve plateau, again. Given the significant
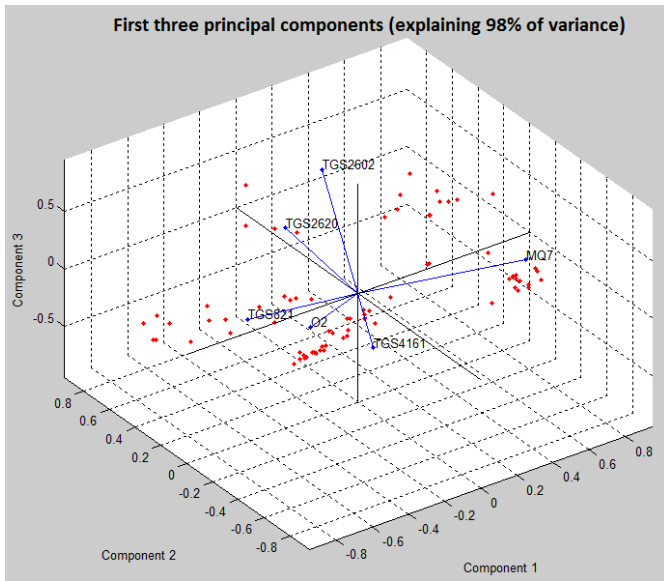
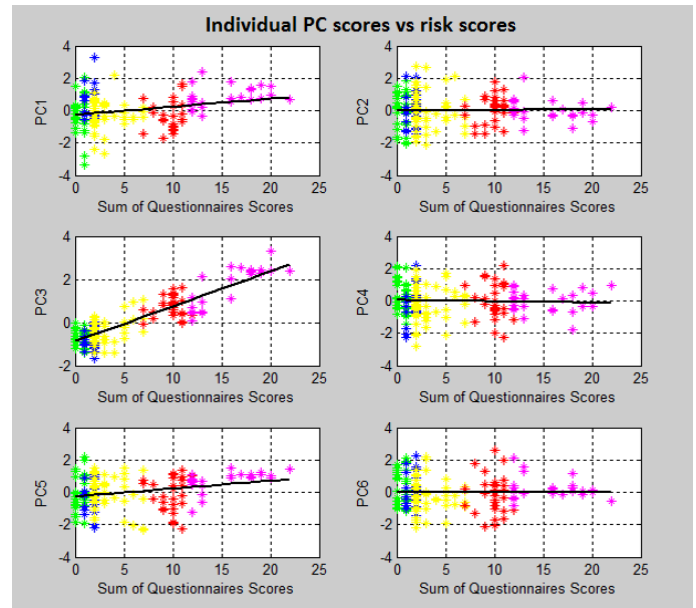Figure 6. First three principal components.

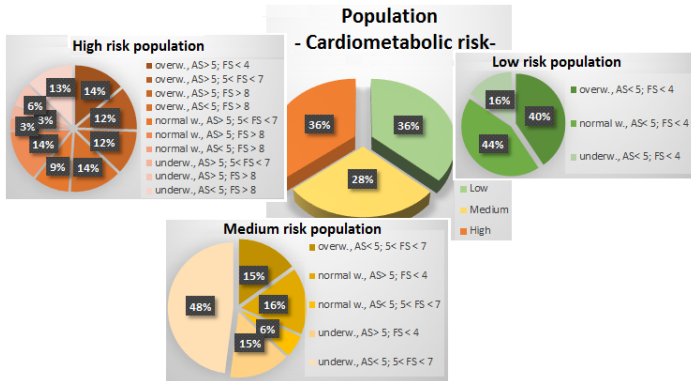Figure 8. PC scores against subjects' risk scores.

Figure 7. Population under study.

data set. If, on one hand, PCA exploits the real sensors' cross-correlation, ICA originates from the assumption that the data has a non-Gaussian distribution, which often is a property of the gas sensors' array measurement data [27]. In our case, breath signals and the environmental ones (noise) get mixed with each other before the chemical interaction with the sensor array. As a consequence, each sensor's output is the result of a combination of different gaseous contributions.

We applied FastICA algorithm to our data set, and plotted individual independent components (IC) against subjects' RS. As shown in Figure 9, in this case, sharper linear trends emerge. Then, the data set was split into two data-sets (train

number of subjects, in this case we tried to implement a method of data analysis which was able to predict subjects' risk score (RS, it means, sum of Audit, Fagerstom and lifestyle questionnaires scores), based on the breath data.

Sensors' raw data first were zero-centered and normalized, thus putting in evidence the qualitative aspects of the data. Then, also in this case, the principal components were extracted and the PC scores were plotted against the subjects' RS (Figure 8). As can be deduced from the colours (green points derive from no-risk subjects, the blue ones from low-risk subjects, the yellow ones from medium risk subjects, the red ones from high risk subjects, the magenta ones from very high risk subjects), subjects' RS are arranged in ascending order.

Except for PC3, from a visual, exploratory analysis, we saw that the PC scores did not have a sharp increasing or decreasing linear trend with respect to RS, thus not having enough information to contribute to any prediction model. Such result matches the one reported in [26]. Being inspired by this study, we also implemented an Independent Component Analysis (ICA) on our data.

ICA is a high-order transformation method for data representation which extracts independent components from the
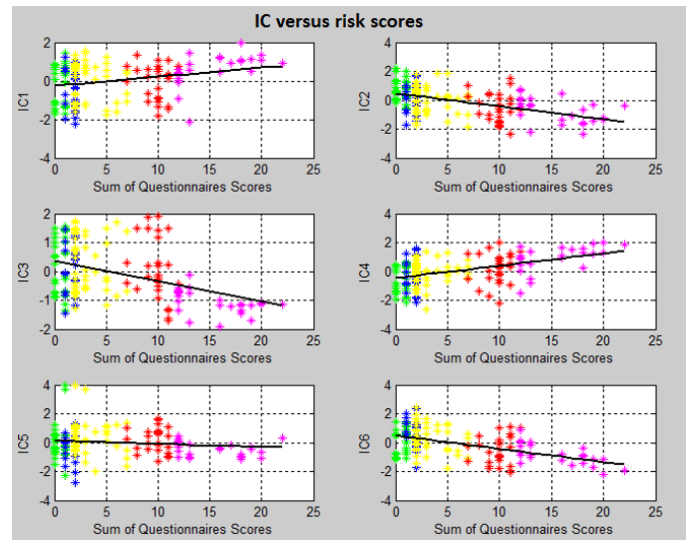

Figure 9. PC scores against subjects' risk scores arranged in ascending order.

data set and validation data set) to build the prediction model, which was developed by means of the Matlab LinearModel Tool. Indeed, by using the independent components, a linear

regression model was built to establish a relationship between the risk score and the breath data pre-processed by ICA. Then, such model was validated by using the validation data set. In Figure 10, we can see that the correlation coefficient (r) between actual and estimated risk scores is 0.8976.
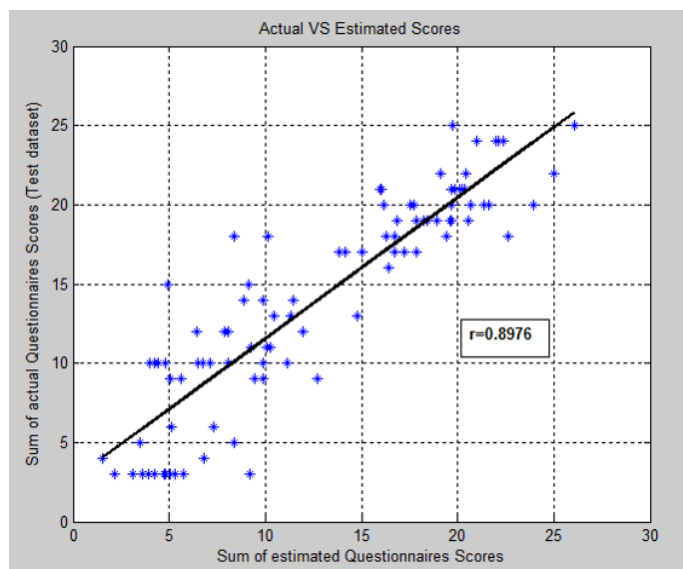

Figure 10. Actual risk scores against predicted ones.

V. CONCLUSION

In this paper, we describe how breath analysis could be exploited for daily self-monitoring by using a portable, low cost, very easy to use device that we developed and called Wize Sniffer. In the presented use case, the WS could help the user to monitor his/her habits and prevent the cardio-metabolic risk. Nevertheless, the WS modular configuration allows for changing the gas sensors according to the molecules (and then, to the related diseases) to be monitored. The core of such type of devices is the gas sensor array. We chose to use MOS-based gas sensors, because of their low cost, their ease to be integrated in the circuitry, their long life and rapid recovery time. Nonetheless, they are affected by humidity, which is, in our case, a parameter to be taken into account, as human breath is composed of 90% water vapor. As reported in Section 4, we faced this issue first by integrating a HME filter and then by calculating sensors' drift due to humidity in order to possibly compensate it. Another peculiarity of MOS-based sensors is the cross-sensitivity, which makes difficult any quantitative analysis approach. In this case, we faced this problem by using multivariate methods of analysis. In addition, our data analysis directly provides the user with him/her cardio-metabolic risk score. The safety and the unobtrusiveness of the device allow for a daily monitoring which, even if without a real diagnostic meaning yet, could represent a pre-screening, useful for an optimal selection of more standard medical analysis.

REFERENCES

[1] W. Miekisch, J. Schubert, and G. Noeldge-Schomburg, "Diagnostic potential of breath analysis - focus on volatile organic compounds," Clinica Chimica Acta, Elsevier, vol. 347, 2004, pp. 25–39.

[2] F. Di Francesco and al., "Implementation of fowlers method for endtidal air sampling," J. of Breath Research, 2008, p. 037009(10pp.).

[3] J. Jones and S. Clarke, "Effect of expiratory flow rate on regional lung emptying," Clin. Sci., 1969, pp. 343–56.

[4] F. Di Francesco, R. Fuoco, M. Trivella, and A. Ceccarini, "Breath analysis: trends in techniques and clinical applications," Microchemical J., vol. 79, 2005, pp. 405–10.

[5] D. Guo, D. Zhang, N. Li, L. Zhang, and J. Yang, "A novel breath analysis system based on electronic olfaction," IEEE Transaction on Biomedical Engineering, vol. 57, 2010, pp. 2753 – 2763.

[6] D. Hill and R. Binions, "Breath analysis for medical diagnosis," Int. J. on smart sensing and intelligent system, vol. 5, 2012, pp. 401–440.

[7] J. Fenske and S. Paulson, "Human breath emission of vocs," J. of the Air and Waste Management Association, vol. 49, 1999, pp. 594–598.

[8] A. Wilson, "Recent progress in the design and clinical development of electronic-nose technologies," Nanobios.in Dis.Diagn., 2016, pp. 15–27.

[9] T. de Meij and al., "Electronic nose can discriminate colorectal carcinoma and advanced adenomas by fecal volatile biomarker analysis: proof of principle study," Int. J. Cancer, 2014, pp. 1132–1138.

[10] N. Yusuf and al., "In-vitro diagnosis of single and poly microbial species targeted for diabetic foot infection using e-nose technology," BMC Bionformatics, 2015, pp. 158–169.

[11] P. Montuschi, N. Mores, A. Trov, C. Mondino, and P. Barnes, "The electronic nose in respiratory medicine," Respiration, 2013, pp. 72–84.

[12] URL: http://www.bedfont.com/ [accessed: 2017-07-04].

[13] D. Walt and al., "Optical sensor array for odor recognition," Biosens. Bioelectron., 1998, pp. 697–699.

[14] F. Lai and H. Huang, "Fabrication of high frequency and low-cost surface-acustic wave filters using near field phase shift photolitography," Microelectronic Eng., 2006, pp. 1407–1409.

[15] L. Fan, H. Ge, S. Zhang, H. Zhang, and J. Zhu, "Optimization of sensitivity induced by surface conductivity and adsorbed mass in surface acoustic wave gas sensors," Sensor Act. B-Chem, 2012, pp. 114–123.

[16] D. Germanese and al., "A low cost, portable device for breath analysis and self-monitoring, the wize sniffer," De Gloria (eds) Applications in Electronics Pervading Industry, Environment and Society, ApplePies 2016, Lecture Notes in Electrical Engineering, vol. 349, 2017.

[17] D. Germanese, M. D'Acunto, and O. Salvetti, "Design of a breath analysis device for self-monitoring and remote health-care," in Proceedings of the 9th Intern. Joint Conf. on Biomedical Eng. Systems and Technologies (BIOSTEC 2016) February 21–23, 2016, Rome, Italy. SciTePress, 2016, pp. 9–14.

[18] URL: http://www.semeoticons.eu/ [accessed: 2017-07-04].

[19] S. Colantonio and al., "A smart mirror to promote a healthy lifestyle," Biosystems Eng., Special Issue: Innovations in Medicine and Health-care. Elsevier, vol. 138, 2015, pp. 33–43.

[20] P. Clifford and D. Tuma, "Characteristics of semiconductors gas sensors i. steady state gas response," Sens. and Act., vol. 3, 1983, pp. 233–254.

[21] D. Shier, D. Butler, and R. Lewis, Holes Human Anatomy and Physiology, 11th ed. McGraw-Hill, 2007.

[22] A. Testa, M. Cinque, A. Coronato, and G. DePietro, "A formal approach for a dependability assessment of a mhealth monitoring system," Mob.Health, Springer S. in Bio-/Neuroinf., vol. 5, 2015, pp. 171–194.

[23] J. Yan and al., "Electronic nose feature extraction methods: A review," Sensors, vol. 15, 2015, pp. 2784–27 831.

[24] D. Germanese, M. Righi, M. Magrini, M. Guidi, M. D'Acunto, and O. Salvetti, "A device for self-monitoring breath analysis," in Proceedings of the 7th Intern. Conf. on Sensor Device Technologies and Applications (SENSORDEVICES2016) July 24–28, 2016, Nice, France. IARIA, 2016, pp. 77–82.

[25] W. Miekisch, S. Kischkel, A. Sawacki, T. Liebau, M. Mieth, and J. Schubert, "Impact of sampling procedures on the results of breath analysis," J. of Breath Research, 2008, p. 026007(7pp.).

[26] S. Balasubramanian, S. Paniraghi, C. Logue, C. Doetkott, M. Marchello, and J. Sherwood, "Independent component analysis- processed electronic nose data for predicting salmonella typhimurium populations in contaminated beef," Food Control, Elsevier, vol. 19, 2008, pp. 236–246.

[27] M. Kermit and O. Tomic, "Independent component analysis applied on gas sensor array measurement data," IEEE Sensors journal, 2003, pp. 500–511.