The Wize Sniffer Knows What You Did:

Prevent Cardio-Metabolic Risk by Analyzing Your Breath

Danila Germanese

Institute of Information Science and Technology
ISTI-CNR
Via G. Moruzzi 1, Pisa, Italy
Email: danila.germanese@isti.cnr.it

Mario D'Acunto

Institute of Biophysics
IBF-CNR
Via G. Moruzzi 1, Pisa, Italy
Email: mario.dacunto@ibf.cnr.it

Marco Righi
and Massimo Magrini
and Ovidio Salvetti

Institute of Information Science and Technology
ISTI-CNR
Via G. Moruzzi 1, Pisa, Italy
Email: marco.righi@isti.cnr.it
Email: massimo.magrini@isti.cnr.it
Email: ovidio.salvetti@isti.cnr.it

***Abstract*—Its un-obtrusiveness and its inherent safety make breath analysis a very promising technique in healthcare diagnostics. On one hand, it enables the monitoring of biochemical processes: the volatile organic compounds (VOCs) from the metabolic processes are generated within the body, travel via the blood, participate to the alveolar exchanges and appear in exhaled breath; on the other hand, breath is easily and non-invasively accessible. Nevertheless, despite its great potential, breath analysis is not widely used in clinical practice: the high costs for standard analytical instrumentation (i.e., gas chromatograph-mass spectrometer), the need for specialized personnel able to read the results and the lack of standardized protocols to collect breath samples, set limits to its exploitation. Here, we describe the Wize Sniffer (WS), a portable device based on low cost technology, able to collect and analyze in real time the composition of the breath. In particular, by means of the WS, the user can evaluate his/her own cardio-metabolic risk score by self-monitoring the composition of the breath. Indeed, the presented device is able to detect, in real time, all those VOCs related to the noxious habits for cardio-metabolic risk. Nonetheless, the modular configuration of the WS, makes it usable also for other applications by changing the type of the gas sensors according to the molecules to be detected.**

Keywords–Breath analysis; E-noses; Gas sensors; Self-monitoring; Signal processing; Bio-signals; Medical device; Cardio-metabolic risk prevention.

I. INTRODUCTION

Since the time of Hippocrates, classical medicine has used the sense of smell as an indicator of human diseases [1], [2]: the fruity-smelling breath underlined the presence of diabetes; the stale beer-like odor of the skin was typical of the persons with tuberculosis; the butcher's-like smell of the skin suggested yellow fever, etc. Therefore, early medical practitioners recognized that the presence of human diseases changed the odors released from the body and breath.

It was in 1784 when, for the first time, Lavoisier and Laplace identified the presence of carbon dioxide in human exhaled breath. However, it is commonly recognized that the modern breath analysis started in 1971, when Linus Pauling demonstrated that breath is a mixture of more than 200 volatile molecules at the levels of part per million (ppm), part per billion (ppb), or lower [3].

Breath is the product of the composition of inspiratory air, molecules deriving from ingested food and beverages or from dermal adsorption (exogenous molecules), and all the volatile substances in the blood, which are produced endogenously as part of our normal (or disease-related) metabolism and participate to alveolar exchanges according to their types, concentrations, volatilities and rates of diffusion. In addition, also cells in the mouth, upper airways, and gastro-intestinal tract contribute volatile molecules to the breath. Human exhaled breath is composed of nitrogen (75%), oxygen (13%), water vapor (6%), carbon dioxide (5%). The remaining 1% is composed of a series of volatile organic compounds (VOCs) that are peculiar for each individual. As a consequence, it is correct to think that every one of us has his/her own *breath-print*, which can tell a lot about the state of health.

Breath analysis, for its un-obtrusiveness and its inherent safety, may play a very important role in health care diagnostics. It may be used to detect disease, monitor disease progression, or monitor a therapy. Indeed, many studies aimed to correlate breath VOCs to various diseases such as diabetes, lung cancer, gastrointestinal diseases, etc. [4], [5], [6]. Exhaled pentane and ethane were investigated as lipid peroxygenation product in case of oxidative stress [7]; isoprene (the major hydrocarbon present in human breath) was suggested to be linked with cholesterol synthesis [8] and cardiac output [9]; breath ammonia may be a useful biomarkers both for the evaluation of clinical treatments in case of renal diseases [10], [11] and for monitoring the level of severity in case of liver diseases [12]; increased levels of breath carbon monoxide may be due to airway inflammation in asthma and in chronic obstructive pulmonary disease (COPD)[13].

Comparing with other traditional methods such as blood test, breath analysis is non-invasive, real-time, and harmless to not only the subjects but also the personnel who collects the samples. Nonetheless, despite its great potential, only few breath tests (among which: carbon monoxide test for neonatal jaundice, ethanol test for drunk drivers, hydrogen test for the

evaluation of gastrointestinal transit time, for the monitoring of intestinal bacterial overgrowth and for the assessing of h.pylori infection, nitric oxide test for the evaluation of asthmatic disease) are commonly used in clinical practice nowadays. In [14] T.H. Risby and S.F. Solga give a fair view of current status of breath analysis and try to explain the reasons why it has not gained a wider use yet. One of such reasons is related to the high costs of the specific, standard instrumentation for gas analysis (i.e., gas chromatograph, mass spectrometer) and the need of expert personnel to perform the analysis, which also are very time consuming [10].

In recent years, the idea of exploiting e-noses also for clinical applications has gained the attention of the scientific community [15]. Being quicker than a gas chromatograph, as they are able to follow the trend in time of breath molecules, in many studies they have been employed in different fields of medicine: in oncology, for instance, to identify lung cancer-related breathprint [16], in infectiology [17], in respiratory medicine to evaluate asthma [13] or to discriminate between healthy subjects and patients suffering from chronic obstructive pulmonary disease (COPD) [18]. Nevertheless, the majority of such e-noses exploit very expensive technology [19], [20] or requires complex circuitry [21], [22].

By developing the Wize Sniffer (WS) [1], [23], described in this paper, we aimed to overcome this limitations:

- the WS is a portable, real-time device, which might be used not only in laboratory settings, but also in doctor's office, or in home environment;
- it is very easy to use, also for non-specialized personnel, thus allowing the self-monitoring of own health state. In addition, it is programmed to send breath analysis results also to a remote care center;
- it is entirely based on low cost technology: the employed gas sensors are commercial, semiconductor-based and easily embeddable in the circuitry; breath signals are analyzed by a widely employed open source controller: Arduino Mega2560.

The WS was developed in the framework of SEMEOTICONS European Project [24], which aimed to develop the Wize Mirror, (WM) a multi-sensory platform having the appearance of a mirror. The WM, by means of a series of depth sensors and multispectral cameras, is able to assess individual's well-being state by detecting in his/her face all those signs related to cardio-metabolic risk [25], [26]. The WS was designed to be a WM's tool, in order to detect in human breath the molecules related to the noxious habits for cardio-metabolic risk: alcohol intake, wrong diet, smoke. Not only: we aimed to develop a device whith a modular core, and which could be also used for broader applications [27], [1].

In the paper, Section II lists the molecules detected by the WS; Section III describes the devices general hardware/software architecture; Section IV explains the WS functionality tests and the experimental results, later discussed in Section V.

II. The detected VOCs

Our aim was to develop a device, which could help the user to monitor his/her noxious habits for cardio-metabolic risk simply by analysing the breath composition. Therefore, within the WS, an array of semiconductor-based gas sensors is able to detect the following molecules:

- Carbon monoxide (CO): in human body, it is naturally produced by the action of heme oxygenase on heme when the macrophages of the spleen remove old and damaged erythrocytes from the circulation. Also, it is the major compound in cigarette smoke. An increase of CO in blood is very dangerous, as it leads hemoglobin to carry less oxygen through the vessels, because CO usurps the space in hemoglobin that normally carries oxygen, forming carboxyhemoglobin [28]. It also increases the amount of cholesterol that is deposited into the arteries. CO normal levels in exhaled breath are 2-3.5ppm; increasing levels can be detected in smokers (13.8 - 29ppm);
- Ethanol (C_2H_6O): endogenous ethanol levels are 0-3.9ppm (mean 0.62ppm), normally lower than the ones found in subjects' breath after alcoholic drinks ingestion. However, moderate ethanol consumption, in healthy subjects, reduces stress and increases feelings of happiness and wellbeing, 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 [29].

The WS can also provide useful information about metabolism, carbohydrates adsorption and vascular status by detecting:

- Oxygen and carbon dioxide (O_2 and CO_2): exhaled air has a decreased amount of oxygen and an increased amount of carbon dioxide. These amounts show how much O_2 is retained within the body for use by the cells and how much CO_2 is produced as a by-product of cellular metabolism. Exhaled O_2 amount is about 13.6%-16%. Mean CO_2 concentration in exhaled breath is about 4% (= 40000ppm) [30]. Individual's breathing rate influences the level of CO_2 in blood and, as a consequence, in exhaled gas. Breathing that is too slow causes respiratory acidosis (that results in an increase of CO_2 partial pressure in blood, which may cause hypertension), while breathing that is too rapid causes a decrease in CO_2 in blood that leads to hyperventilation and respiratory alkalosis;
- Hydrogen (H_2): it is related to the carbohydrates breakdown in the intestine and in the oral cavity by anaerobic bacteria [31]. Breath hydrogen levels vary within a day and from day to day; fasting levels range between 0.3 and 34.1ppm (mean 9.1ppm). However, it may vary also among individuals, especially in case of lactose intolerance and celiac diseases;
- Hydrogen sulfide (H_2S): in healthy subjects, concentrations of volatile sulfur-containing compounds in blood are very low. The body uses sulfur compounds in order to neutralize the action of free radicals [32]. Among the sulfur-containing volatile molecules, hydrogen sulfide is considered as a vascular relaxant agent, as it has a therapeutic effect in various cardio-vascular diseases.

III. WHITHIN THE WS: HARDWARE AND SOFTWARE

Here, we describe the hardware and software platforms of the WS.

A. Hardware

In designing our device, we took into account an important issue: the greater demands on improvements in effectiveness, smartness and lower costs of biomedical instruments used in daily healthcare applications [33], resulted from increasing limitations of healthcare financial resources as a consequence of budgetary cuts or constraints.

For this reason, we aimed to design a device entirely based on low-cost technology. In Figures 1, 2 and 3, WS' hardware is shown.

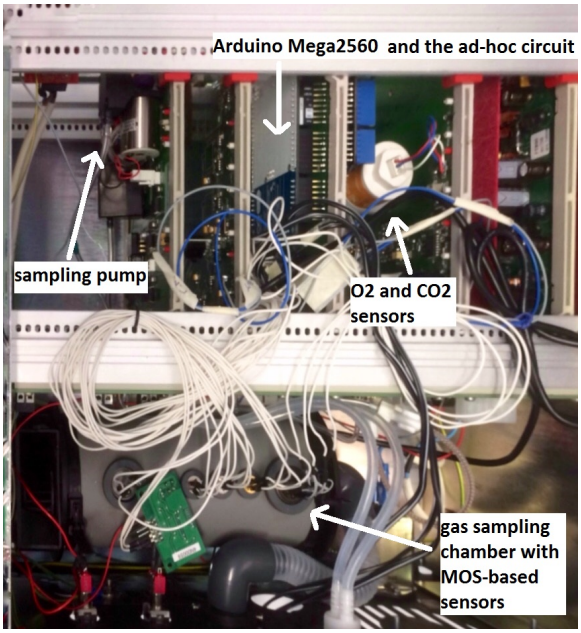

Figure 1. Wize Sniffer's hardware, internal configuration

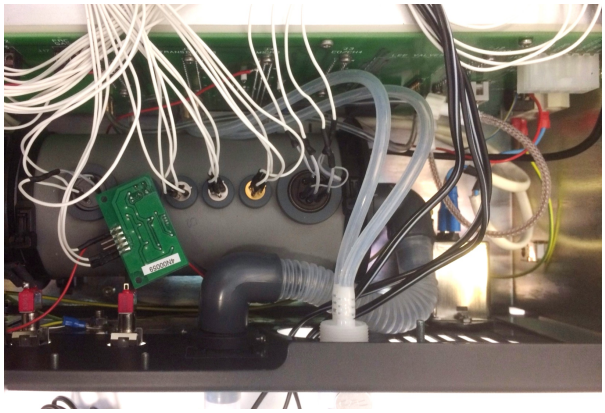

Figure 2. Wize Sniffer's gas sampling chamber detail

The exhaled gases flow into a corrugated tube, made of polyvinyl chloride (PVC), and reach the gas sampling chamber. A heat and moisture (HME) filter, made of hygroscopic material, absorbs the water vapor present in the exhaled breath: as it

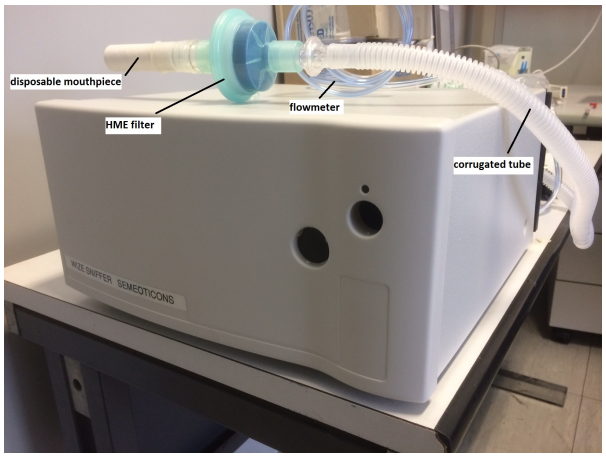

Figure 3. Wize Sniffer's hardware, external configuration. Its dimensions are: 30x30x14cm

will be described later, the employed gas sensors'conductivity response is strongly affected by humidity. Integrating such a filter allows for reducing the humidity of -30% and also for holding users' oral bacteria. A PNT Flow-Ree flowmeter allows for monitoring the user's flow rate and for calculating the exhaled gas volume. The core of the WS is the signal measurement module, that is the sensor array, composed of six semiconductor-based gas sensors, placed within the gas sampling box. The latter was made up of acrylonitrile-butadiene-styrene (ABS) and Delrin, which are two materials that do not interfere with sensors's sensitivity, and its capacity is 600ml according to the tidal volume [30]. Within the gas sampling box, also a sensor for temperature and humidity (Sensirion SHT11) is placed. In addition, a sampling pump injects, at a fixed rate (120ml/sec), the sampled exhaled gas to other two sensors, which have faster response time and work in *flowing-regime*. They detect oxygen and carbon dioxide and are respectively based on an electrochemical cell and an infrared source. Sensors' raw output are pre-processed and stabilised by a signal conditioning module: a series of voltage buffer amplifiers (LM124-N, Texas Instrument) transfers sensors' signals from the measurement module to a widely employed open source controller: an Arduino Mega2560 with Ethernet module. Finally, in order to facilitate sensors' recovery time, a flushing pump was integrated on one side of sampling chamber. After each breath test, it can be switched on in order to "purge" the chamber with ambient air and recovery sensors' baseline.

In Table I, the employed gas sensors are listed. 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, simply by 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.

Optical, carbon nano-fiber (CNF), quartz crystal microbalance (QCM), metal oxide semiconductors (MOS), conducting polymers (CP), and surface acoustic wave (SAW), are the most common gas sensor types employed in e-noses [15], [34].

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-ber (CNF) based gas sensors are very expensive, especially for their manufacturing. We chose metal oxide semiconductor (MOS)- based gas sensors: they show long term stability and reproducibility of gas response [35], great metallurgical and chemical stability of the sensing material [35], high sensitivity towards target gases, short reaction and recovery time, easy calibration. In comparison to other types of gas sensors, MOS-based gas sensors' availability, small dimensions, compactness and low cost make them the most widely used gas sensors [36].

Whithin the gas sampling chamber, six Taguchi semiconductor-based gas sensors, manufactured by Figaro Engineering [37] (costs: 25-40 Euro), were integrated.

TABLE I. SENSORS INTEGRATED IN THE WS SENSOR PLATFORM

Detected molecule	*Sensor*	*Best detection range*
Carbon monoxide	TGS2442	50-1000ppm
	MQ7	20-200ppm
	TGS2620	50-5000ppm
Ethanol	TGS2602	1-10ppm
	TGS2620	50-5000ppm
Carbon dioxide	TGS4161	0-40000ppm
Oxygen	MOX20	0-16%
Hydrogen sulfide	TGS2602	1-10ppm
Hydrogen	TGS821	10-5000ppm
	TGS2602	1-10ppm
	TGS2620	50-5000ppm
	MQ7	20-200ppm

Unfortunately, humidity and cross-sensitivity strongly affect the behavior of MOS-based gas sensors [35], as shown in Figure 4. The water vapour undergoes dissociative adsorption on metal oxide surfaces and the resultant ions are adsorbed on the metal oxide surface, impeding, in many cases, the response of sensor by lowering the sensitivity of the sensing element [38], [39]. However, a distinction should be done between n-type and p-type-based sensing materials [40]. As shown in Figure 4, when humidity icreases, resistance of the n-type-based film decreases, resulting in a rise in voltage output.

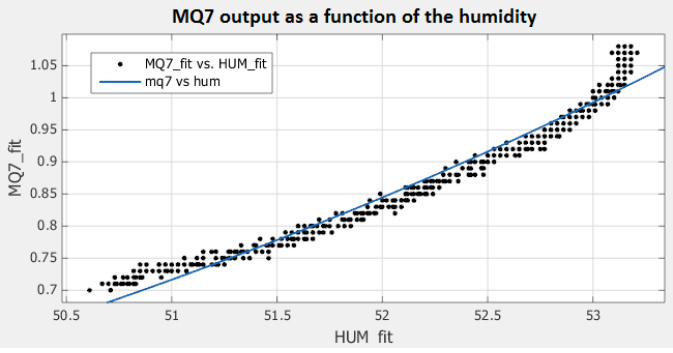

Figure 4. MQ7 output when a rise in humidity occurs.

In our case, humidity plays a very crucial role, as we deal with human breath. Therefore, we deemed it necessary to take steps to manage this factor and optimize its effects. First, as previously described, we put a humidity filter, made of hygroscopic material, in order to absorb the majority of the water vapor present in exhaled breath and reduce the humidity in the gas sampling chamber from a 90% to a 70-60%. In addition, we integrated, into the gas sampling chamber, also a temperature and humidity sensor (Sensirion SHT11) in order to monitor these two parameters. Indeed, we noted that the temperature inside the chamber remains almost constant before, during and after each breath test. On the contrary, relative humidity shows a variation of about 35% while the subject is performing a breath test (Figure 5). For this purpose, we i) calculated sensors' drift due to variations in humidity; ii) investigated sensors' sensitivity in precise measurement conditions [1] (30 °C±7%, 70%RH±5%, that are the ones that occur in the sampling box during a breath test, as shown in Figure 5.

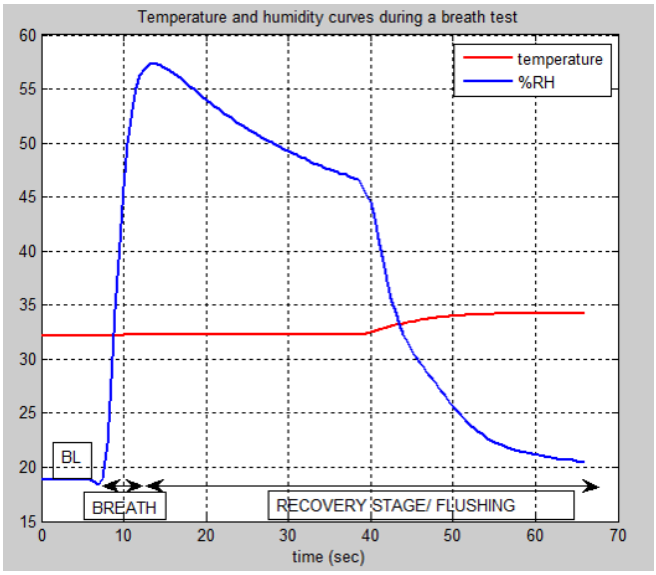

Figure 5. Temperature and relative humidity in the gas sampling box when a breath analysis is performed

The relationship between humidity and sensors' output V_{out} (as can be observed in Figure 4) generally can be modeled by means of a power law (eq. (1)), as reported also by Ho Sohn and coworkers [41]:

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

where a, b and c are constant and specific for each TGS sensor. Calculating sensors' humidity drift is useful to potentially compensate it during the data processing. We considered the entire range of humidity variation (for instance, 50%-55%RH in the case of MQ7) and then we calculated the slope of the curves. Based on the slope, drift coefficient S_d was assessed for each sensor (see Table II) as the decrease in sensors output ΔV (Volt) per unit decrease in humidity, Δhum as given in eq. 2:

$$S_d = \frac{\Delta V}{\Delta hum} \quad (2)$$

By keeping the humidity constant, sensors' output depends on the gas concentration only. For this purpose, we investigated sensors' behavior in response to a well-known gas concentrations at a fixed humidity and temperature conditions. The used experimental set-up was the one reported in Figure 6.

TABLE II. SENSORS DRIFT DUE TO HUMIDITY

Sensor	ΔV / Δhum *(mV)*
MQ7	296
TGS2620	60
TGS2602	82
TGS821	120
TGS2444	84

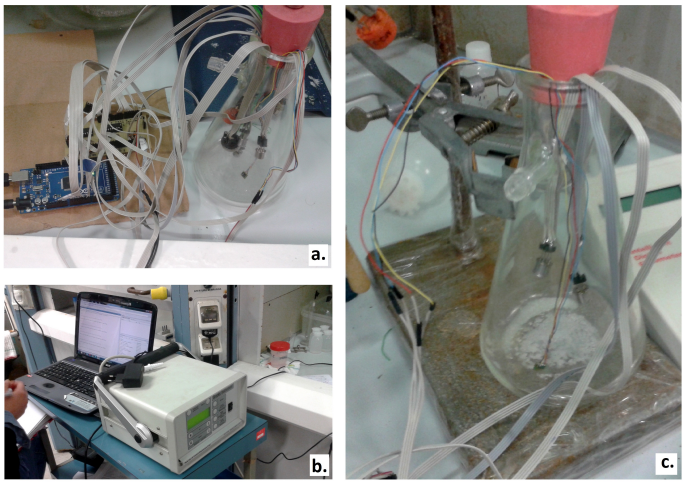

Figure 6. Experimental set-up. a) The gas sensors' raw output are read by an Arduino Mega2560 connected via USB to a personal computer; b) the signals are displayed in real time on the computer screen; c) the gas sensors into the vial, where is placed a saturated solution of $NaCl$ on the bottom.

The gas sensors were put into a vial, where the humidity was kept at 70%RH±5% by means of a saturated solution of $NaCl$ placed on the bottom. Measurements were performed only after the sensors were operated at a fixed temperature for several hours (at least 2 hours for warm-up).

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 7, we can see TGS2620 output when well-known concentrations of carbon monoxide, ethanol and hydrogen were separately injected into the vial.

In general, the relationship between sensors' output V_{out} and each gas contribution can be modeled by means of eq. (3).

$$V_{out} = f([gas]) = a * ([gas]^b) + c \quad (3)$$

Where c is a constant, b is the constant power-law exponent and a can be considered as the sensor's sensitivity coefficient. These parameters are specific for each TGS sensor. We also found that the used gas sensors were sensitive to concentration lower than their best detection range (reported in Table I).

Nevertheless, when a breath analysis is performed, a mixture of gases spreads into the gas sampling box and chemically interacts with the sensors. In this case, the phenomenon known as cross sensitivity makes the semiconductor sensors non-selective. In addition, the detection threshold (that is, the minimum concentration of gas necessary to a meaningful change in sensors' conductivity) depends not only on absolute

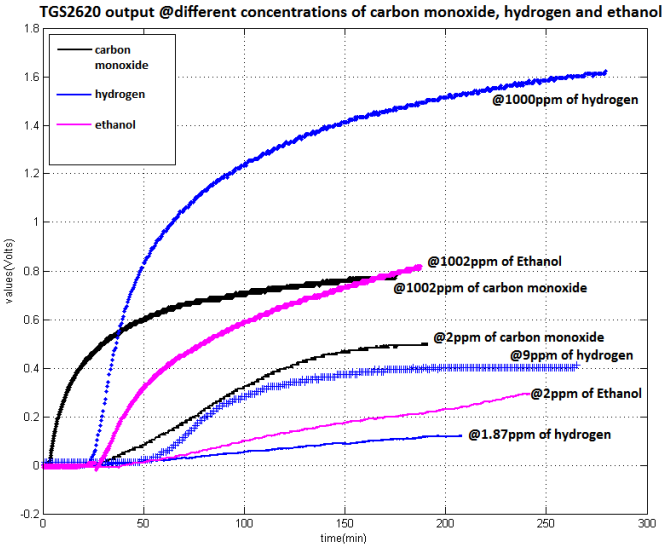

Figure 7. TGS2620 output when well-known concentrations of CO, H_2 and C_2H_6O were separately injected into the vial.

sensitivity to that particular gas but also on concentrations of the other gases, which partially mask the response to the gas of interest, as shown in eq. (4) reported by P. Clifford and coworkers [35] (the equation refers to the case of only two mixed gases, as an example).

$$(\frac{R_j}{R_{0j}})^{-\frac{1}{\beta_j}} = \frac{(1 + \sum K_j * [G_{1j}]^{n_{1j}} * K_j * [G_{2j}]^{n_{2j}})}{[O_2]} \quad (4)$$

Where $\frac{R}{R_0}$ is the j-th sensor's variation in resistance, β is the power law exponent, specific for each j-th sensor, $[G_1]$ and $[G_2]$ represent the concentrations of the two mixed gases, n is an integer or fractional integer power, specific for each j-th sensor, K can be considered as the j-th sensor's sensitivity coefficient and $[O_2]$ is the oxygen partial pressure. In some cases, for some terms of the summation, there is only one term per gas, for oters there is a product of several.

We also tried to investigate the cross sensitivity of our sensors. In Figure 8, we can see TGS2620 response when well-known mixed concentrations of the three gases (carbon monoxide, ethanol and hydrogen) were injected into the vial at the same time.

In this way, how the different VOCs add together and influence gas sensors output can be understood. The single gas contribution can be modeled by a power law similar to eq. (3), but each of them has its "weight" on the overall output, as shown in eq. (4). However, because of the multitude of factors involved, understanding the interaction mechanism behind the MOS-based gas sensors' sensing property in general remains an open issue [40].

Finally, in Figure 9, the WS performing a breath test is shown.

B. The Software

Given its un-obtrusiveness and its safety, breath analysis may be used as a daily monitoring analysis tool. To fully

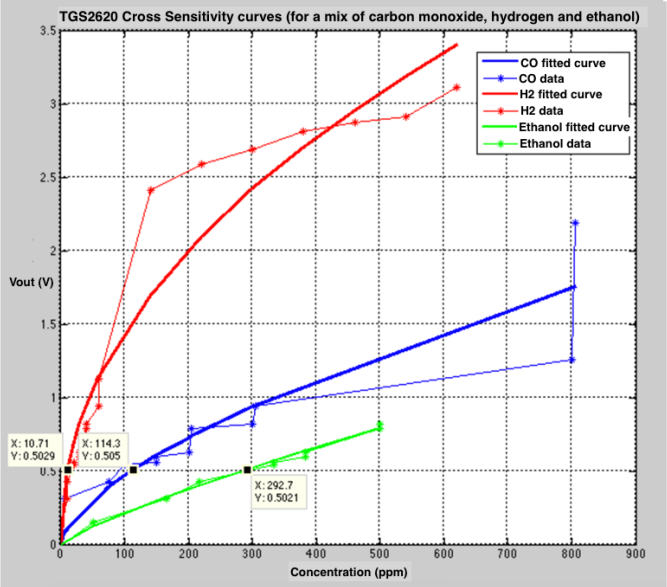

Figure 8. TGS2620 output when well-known mixed concentrations of CO, H_2 and C_2H_6O were contemporarly injected into the vial.

Figure 9. The WS while performing a breath test.

exploited its potential, its application must take place not only in laboratory settings, but also in the clinics, in doctors' offices, and at home. Our aim was to develop a device, which could be used in home environment and which could be able to send breath analysis results also to a remote personal computer (for instance, to the one of the own family doctor).

Arduino Mega2560 with Ethernet board samples sensors signals every 250msec, saves raw vector data and extracts the maximum value of raw breath curve. Several parameter and features can be derived from breath curves [42] to fully characterized them. We chose to calculate the value at curve plateau as it better describes the chemical balance between sensorssensing 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 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 users exhaled volume equals at least the one of gas sampling box (600ml).

IV. WS VALIDATION

The aim of the validation 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). For this purpose, as described in [1], 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 campaign involved 77 volunteers, male and female, between 30-65 years of age, with different habits and lifestyles. People answered Audit and Fagerstrom tests, which respectively assessed their alcohol and smoke dependence, and other questionnaires about lifestyle in general.

Exhaled breath composition is strongly inuenced by breath sampling method [43], as well as by breath flow rate [44], posture [45], ambient air [46], lung volume [47]. In our case, also factors such as BMI, sex, age, subjects lifestyle may inuence the breath composition: for example, alcohol disposal in men is different than the one in women, and, in addition, it may depend on body mass index (BMI)[48], as well. Therefore, breath composition may exhibit not only a strong inter-variability (among different subjects), but also a marked intra-variability (relative to the same subject). Moreover, standard protocols for breath sampling do not exist. The definition of precise guidelines to collect breath sample would be useful also to avoid such factors that influence the breath composition.

As a consequence, we considered, on one hand, all these issues about human breath variability and influencing factors, and, on the other hand, the methodological issues about breath sampling [43]. Since our interest was focused on both endogenous and exogenous biomarkers, we drafted a protocol which exploited the mixed expiratory air sampling. Such method of sampling entails collecting total breath, including the air contained in the upper airways, which involves volatile compounds that do not participate to alveolar exchanges (dead space air) [46]. 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. However, mixed expired air consists of dead space, transition phase and alveolar phase. The dead space and the transition phase contain breath compounds from the upper airways, whereas the alveolar phase contains the VOCs resulting from alveolar exchanges, which better represent the individuals metabolic conditions. As a consequence, a controlled identification of the respiratory phases was performed, by monitoring the curve of exhaled carbon dioxide (i.e., capnogram) and discriminating between dead space air fraction and end-tidal exhaled air. In addition, the applied manoeuvres such as breath holding [49], high exhaled volume, lower exhalation flow rate [50], [51], [52] and single exhalation [53] lead to an incrase in alveolar VOCs concentrations in breath samples, thus reducing the impact of the previously listed influencing factors (especially ambient air).

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 not selective, thus impeding to calculate the exact molecules' concentrations. Multivariate methods of pattern recognition techniques usually allow for overcoming this problem. Pattern recognition exploits sensors' cross-correlation and helps to extract qualitative information contained in sensors' outputs ensemble. Therefore, 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' multidimensional space. From a visual, exploratory analysis of the data, the presence of clusters (see Figure 10) was observed. Then, a Knearest neighbor (KNN) classification algorithm, previously trained with the data collected during another acquisition campaign, was adopted to classify the subjects according to their habits: Healthy (that means, not in danger of cardio-metabolic diseases), Light Smoker, Heavy Smoker, Social Drinker, Heavy Drinker, LsSd (Light smoker and Social drinker), LsHd (Light smoker and Heavy drinker), HsSd (Heavy smoker and Social drinker), HsHd (Heavy smoker and Heavy drinker).

The outcomes of the Audit and Fagerstrom questionnaires were our ground truth. The KNN classifier was able to correctly classify in 89,61% of cases. Errors were probably 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 the high presence of hydrogen in their breath, which caused a rise in these sensors' voltage output.

Then, the number of volunteers increased up to 169 subjects. They were classified by clinicians into "low risk population", "medium risk population", "high risk population" and "very high risk population", on the base of their Risk Score (RS), that is, the sum of the scores relative to Audit (AS), Fagerstrom (FS) and lifestyle questionnaires, our ground truth also in this case. Given the significant number of subjects, we tried to implement a simple model, which was able to predict subjects RS on the base of breath data. First, sensors' raw data were zero-centered and normalized, thus putting in evidence their qualitative aspects. Then, also in this case, the principal components were extracted and the PC scores were plotted against the subjects RS, as shown in Figure 11.

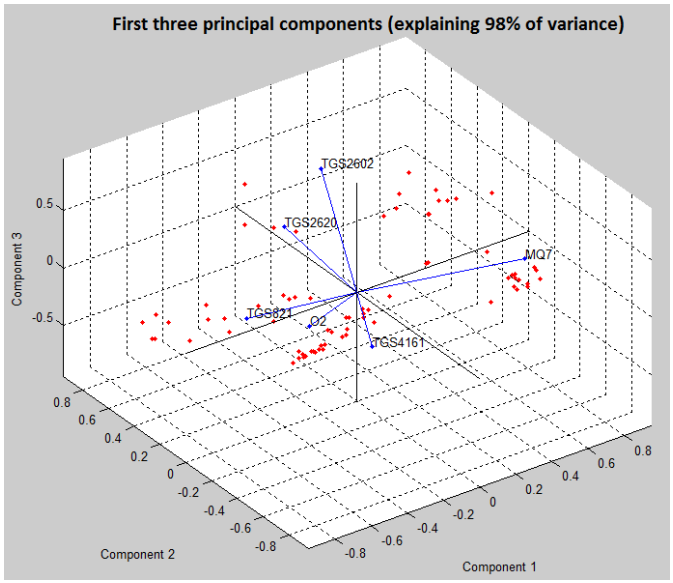

Figure 10. First three principal components. The presence of several clusters can be observed.

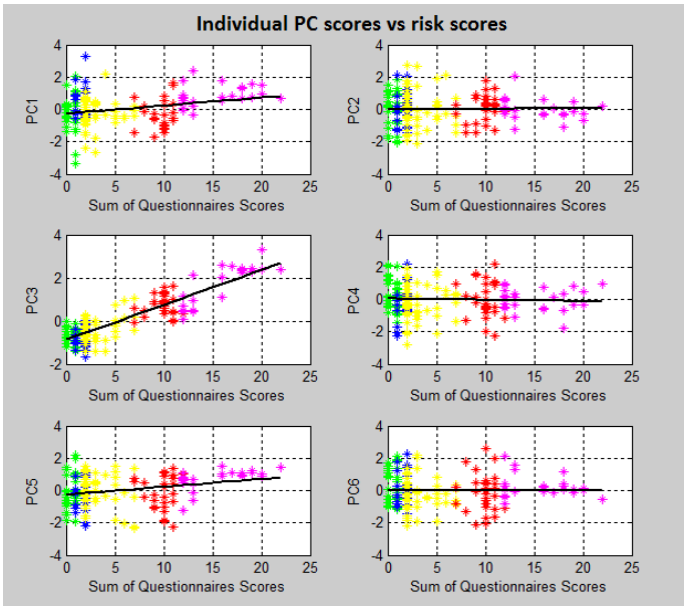

Figure 11. PC scores against subjects risk scores arranged in ascending order.

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 were arranged in ascending order. Except for PC3, from an 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 matched the one reported in [54]. 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 component from

the 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 [55]. In our case, breath signals and the environmental ones (noise) get mixed with each other before the chemical interaction with the sensor array. In addition, due to sensors' cross-sensitivity effect, the conductivity response depends on a linear combination of individual gas terms, where the effects of one gas can be masked by the combined effects of others. In addition to this "competition" among gases, there is an associative interaction by which the effects of one gas are enhanced by the presence of another. As a consequence, each sensors 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 12, in this case sharper linear trends emerged.

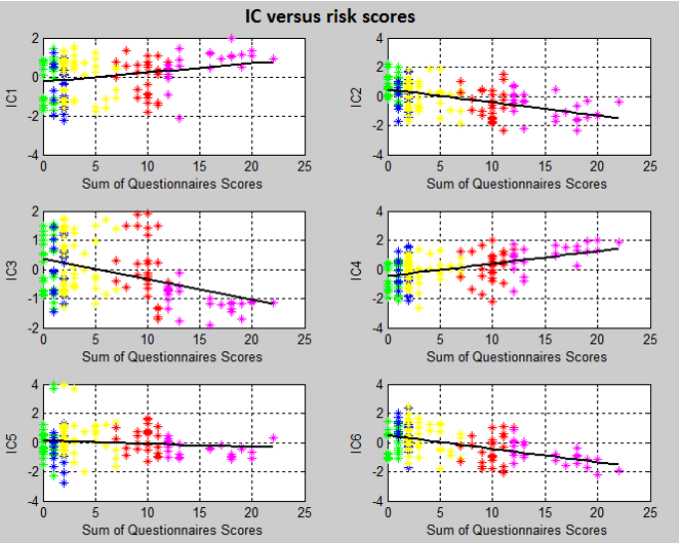

Figure 12. IC scores against subjects risk scores arranged in ascending order.

Then, the data set was split into two data-sets (train 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 volunteers' RS and the breath data pre-processed by ICA. Then, such model was validated by using the validation data set. In Figure 13 we can see that the correlation coefficient (r) between actual and estimated risk scores was 0.8976.

V. Conclusion

The field of breath analysis is as old as the one of medicine. Since the time of Hippocrates, classical medicine has exploited the sense of smell to identify human diseases.

For its un-obtrusiveness, and its inherent safety, breath analysis may be a very useful tool in health care diagnostics: to detect disease, or to monitor disease progression, or to evaluate the success of a therapy. With respect to other traditional methods such as blood test, breath analysis is non-invasive, real-time, and harmless to not only the subjects but also the personnel who collect the samples. Nevertheless, to fully exploit such potential, breath analysis should be performed

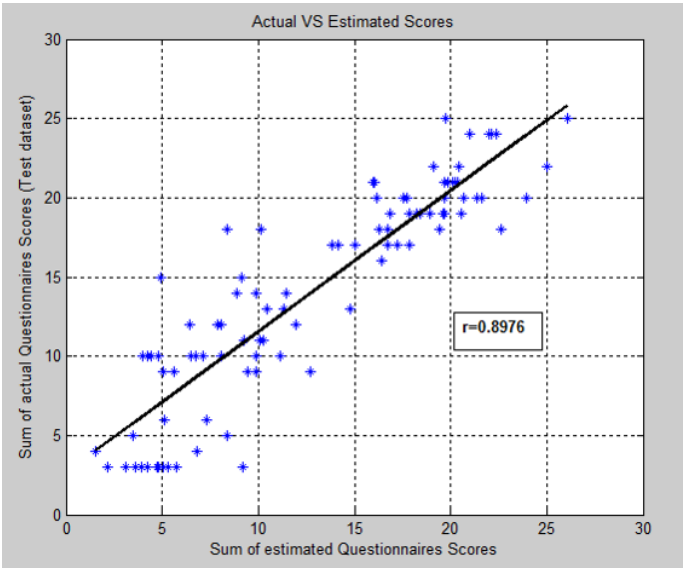

Figure 13. Actual risk scores versus predicted ones

by using a simple, portable, easy to use instrumentation. In addition, an important issue should be considered, that is the greater demands on improvements in lower costs of biomedical instruments used in daily healthcare applications, resulted from increasing limitations of healthcare financial resources as a consequence of budgetary cuts or constraints.

In this paper, we presented the Wize Sniffer, a portable device, able to analyze in real time the composition of human breath, and entirely based on low-cost technology: an Arduino Mega 2560 (a widely employed open source controller) and a semiconductor-based Taguchi gas sensor array.

Although such type of sensors are strongly affected by humidity (which in our case plays a crucial role) with a few, inexpensive arrangements we faced this issue, by integrating a hygroscopic filter behind the mouthpiece, a temperature and humidity sensor within the gas sampling box and by calculating sensors'drift due to humidity. In addition, a robust data processing helped us to face with sensors' cross-sensitivity and human breath inter-variability and intra-variability. When performing a breath analysis by means of non-selective gas sensors, one has to face, on one hand, with an uncertainty of measure which derives from all those factors that affect the gas sensors' behavior; on the other hand, one deals with un uncertainty due to all the physiological conditions that may inuence the breath composition. Pattern recognition algorithms turn out the best way to overcome such problem.

The Wize Sniffer is very simple to use, also by non-specialized personnel: in the presented use case, it provides the user with a very easily interpretable outcome. The WS is able to calculate, by means of a simple regression model, the user's Risk Rcore (with respect to his/her noxious habits) and potentially help him/her to prevent his/her cardiometabolic risk.

However, the user can also send his/her breath analysis results and Risk Score to a remote healthcare center by means of an internet connection: indeeed, Arduino is programmed to implement a simple Telnet Server.

Finally, a strong point of the Wize Sniffer is also its adjustability: thanks to its modular design, the gas sensors can be changed according to the molecules (and then, to the related diseases) to be monitored. Such characteristic allows for using such device in broader applications. Indeed, recently the WS has been used to discriminate and monitor patients with acute liver diseases by evaluating ammonia levels in their breath. The evaluation of WS performances also in the case of cirrhotic patients will be discussed in a future work.

In general, the safety and the un-obtrusiveness of a device for breath analysis, like the presented one, 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. Undoubtedly, further big efforts will be done in order to introduce breath analysis in clinical practice. Scientists and manufacturers should collaborate in order to standardize, on one hand, the architectural principles which e-noses have to be based on, and on the other hand, the procedures for breath sampling, in order to obtain compatible signals and outcomes that may be used and processed by different e-nose systems and shared among physicians all over the world.

ACKNOWLEDGMENT

This work was funded in the framework of the Collaborative European Project SEMEOTICONS (SEMEiotic Oriented Technology for Individuals CardiOmetabolic risk selfassessmeNt and Self- monitoring), grant N. 611516.
Paolo Paradisi, Massimo Martinelli, Massimo Guidi and COSMED s.r.l. are warmly acknowledged for useful support.

REFERENCES

[1] D. Germanese, M. D'Acunto, M. Magrini, M. Righi, and O. Salvetti, "A Low Cost Technology-based Device for Breath Analysis and Self-monitoring," in Proceedings of the Second International Conference on Advances in Signal Image and Video Processing, SIGNAL2017. IARIA, May 2017, pp. 8–13.

[2] J. Pereira et al., "Breath analysis as a potential and non-invasive frontier in disease diagnosis: an overview," Metabolites, vol. 5, 2015, pp. 3–55.

[3] R. Dweik and A. Amann, "Exhaled breath analysis: the new frontier in medical testing," Journal of Breath Research, vol. 2, 2008, p. 030301.

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

[5] D. Hill and R. Binions, "Breath analysis for medical diagnosis," International Journal on Smart Sensing and Intelligent Systems, vol. 5, 2012, pp. 465–498.

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

[7] M. Phillips, R. Cataneo, J. Greenberg, R. Grodman, and M. Salasar, "Breath markers in oxidative stress in patients with unstable angina," Heart Disease, vol. 5, 2003, pp. 95–99.

[8] R. Salerno-Kennedy and D. Cashmark, "Potential applications of breath isoprene as a biomarker in modern medicine: a concise overview," Wienklin Wochenschr, vol. 17, 2005, pp. 180–186.

[9] F. Pabst, W. Miekisch, P. Fuchs, S. Kischkel, and J. Schubert, "Monitoring of oxidative and metabolic stress during cardiac surgery by means of breath biomarkers: an observational study," Journal of Cardiothoracic Surgery, vol. 2, 2007, p. 9pp.

[10] 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.

[11] I. Essiet, "Diagnosis of kidney failure by analysis of the concentration of ammonia in exhaled breath," Journal of Emerging Trends in Engineering and Applied Sciences, vol. 6, 2013, pp. 859–862.

[12] R. Adrover et al., "Breath-Ammonia Testing of Healthy Subjects and Patients with Cirrhosis," Digestive Diseases and Sciences, vol. 57, 2012, pp. 189–195.

[13] P. Montuschi, N. Mores, A. Trov, C. Mondino, and P. Barnes, "The electronic nose in respiratory medicine," Respiration, vol. 85, 2013, pp. 72–84.

[14] T. Risby and S. Solga, "Current status of clinical breath analysis," Applied Physics B, vol. 85, 2006, pp. 421–426.

[15] A. Wilson, "Recent progress in the design and clinical development of electronic-nose technologies," Nanobiosensors in Disease Diagnosis, vol. 5, 2016, pp. 15–27.

[16] R. Machado et al., "Detection of lung cancer by sensor array analysis of exhaled breath," American J. of Respiratory and Critical Care Medicine, vol. 171, 2005, pp. 1286–1291.

[17] N. Yusuf et al., "In-vitro diagnosis of single and poly microbial species targeted for diabetic foot infection using e-nose technology," BMC Bionformatics, vol. 16, 2015, pp. 158–169.

[18] O. Sibila et al., "Identification of airway bacterial colonization by an electronic nose in chronic obstructive pulmonary disease," Respiratory Medicine, vol. 108, 2014, pp. 1608–1614.

[19] URL: http://www.bedfont.com/ [accessed: 2017-08-30].

[20] D. Walt et al., "Optical sensor array for odor recognition," Biosensors and Bioelectronics, vol. 13, 1998, pp. 697–699.

[21] F. Lai and H. Huang, "Fabrication of high frequency and low-cost surface-acustic wave lters using near eld phase shift photolitography," Microelectronic Engineering, vol. 83, 2006, pp. 1407–1409.

[22] 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, vol. 161, 2012, pp. 114–123.

[23] M. D'Acunto et al., "Wize sniffer - a new portable device designed for selective olfaction," in Proceedings of the Seventh International Conference on Health Informatics, HEALTHINF2014. SCITEPRESS Digital Library, Mar. 2014, pp. 9–14.

[24] URL: http://www.semeoticons.eu/ [accessed: 2017-08-30].

[25] S. Colantonio et al., "A smart mirror to promote a healthy lifestyle," Biosystems Engineering,Special Issue: Innovations in Medicine and Healthcare, vol. 138, 2015, pp. 33–43.

[26] P. Henriquez et al., "Mirror mirror on the wall... an unobtrusive intelligent multisensory mirror for well-being status self-assessment and visualization," IEEE Transaction on Multimedia, vol. 19, 2017, pp. 1467–1481.

[27] D. Germanese et al., "A low cost, portable device for breath analysis and self-monitoring, the Wize Sniffer," De Gloria A. (eds) Applications in Electronics Pervading Industry, Environment and Society. ApplePies 2016. Lecture Notes in Electrical Engineering, 2017, pp. 51–57.

[28] S. Zevin, S. G. Saunders, S. G. Gourlay, P. Jacob, and N. Benowitz, "Cardiovascular effects of carbon monoxide and cigarette smoking," Journal of the American College of Cardiology, vol. 38, 2001, pp. 1633–1638.

[29] T. Risby, Volatile organic compounds as markers in normal and diseased states. ed. N. Marczin and M. Yacoub, IOS Press, 2002, pp. 113–122.

[30] D. Shier, J. Butler, and R. Lewis, Eds., Holes Human Anatomy and Physiology. McGraw-Hill, 11th Ed, 2007.

[31] D. Thompson et al., "Extra intestinal influences on exhaled breath hydrogen measurements during the investigation of gastrointestinal disease," Gut, vol. 26, 1985, pp. 1349–1352.

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

[33] A. Wilson and M. Baietto, "Applications and advances in electronic-nose technologies," Sensors, vol. 9, 2009, pp. 5099–5148.

[34] K. Arshak, E. Moore, G. Lyons, F. Harris, and S. Clifford, "A review of gas sensors employed in electronic nose applications," Sensor Review, vol. 24, 2004, pp. 181–198.

[35] P. Clifford and D. Tuma, "Characteristics of semiconductor gas sensors I. Steady state gas response," Sens Actuators, vol. 3, 1982, pp. 233–254.

[36] H. Nagle, R. Gutierrez-Osuna, and S. Schiffman, "The How and Why of Electronic Noses," IEEE Spectrum, vol. 1998, 35, pp. 22–31.

[37] URL: http://www.figarosensor.com/ [accessed: 2017-08-30].

[38] C. Wang, L. Yin, L. Zhang, D. Xiang, and R. Gao, "Metal Oxide Gas Sensors: Sensitivity and Influencing Factors," Sensors, vol. 10, 2010, pp. 2088–2106.

[39] Q. Qi et al., "Electrical Response of Sm_2O_3- Doped SnO_2 to C_2H_2 and Effect of Humidity Interference," Sens. and Actuat. B, vol. 134, 2008, pp. 36–42.

[40] P. Shankar, J. Bosco, and B. Rayappan, "Gas sensing mechanism of metal oxides: The role of ambient atmosphere, type of semiconductor and gases - A review," ScienceJet, vol. 4, 2015, pp. 1–18.

[41] J. Ho Sohn, M. Atzeni, L. Zeller, and G. Pioggia, "Characterisation of humidity dependence of a metal oxide semiconductor sensor array using partial least squares," Sens. and Actuat. B, vol. 131, 2008, pp. 230–235.

[42] J. Yan et al., "Electronic nose feature extraction methods: A review," Sensors, vol. 15, 2015, pp. 27 804–27 831.

[43] W. Miekisch, S. Kischkel, A. Sawacki, T. Lieban, M. Mieth, and J. Schubert, "Impact of sampling procedures on the result of breath analysis," Journal of Breath Research, vol. 2, 2008, p. 026007 (7pp).

[44] J. Jones and S. Clarke, "Effect of expiratory flow rate on regional lung emptying," Clin. Sci., vol. 37, 1969, pp. 343–356.

[45] N. Anthonisen, P. Robertson, and W. Ross, "Gravity-dependent sequential emptying of lung regions," Journal of Appl. Physiol., vol. 28, 1970, pp. 589–595.

[46] F. Di Francesco et al., "Implementation of fowlers method for end-tidal air sampling," Journal of Breath Research, vol. 2, 2008, p. 037009 (10pp).

[47] J. Jones, "The effect of pre-inspiratory lung volumes on the result of the single breath O2 test," Respiratory Physiology, vol. 2, 1967, pp. 375–85.

[48] K. Maudens, L. Patteet, A. van Nuijs, C. Van Broekhoven, A. Covaci, and H. Neels, "The influence of the body mass index (BMI) on the volume of distribution of ethanol," Forensic Sci Int., vol. 243, 2014, pp. 74–78.

[49] M. Larstad, K. Toren, B. Bake, and A. Olin, "Determination of ethane, pentane and isoprene in exhaled air effects of breath-holding, flow rate and purified air," Acta Physiol., vol. 189, 2007, p. 8798.

[50] J. Anderson, W. Lamm, and M. Hlastala, "Measuring airway exchange of endogenous acetone using a single-exhalation breathing maneuver," J. of Applied Physiology, vol. 100, 2006, pp. 880–889.

[51] M. Hlastala and J. Anderson, "The impact of breathing pattern and lung size on the alcohol breath test," Ann. Biomed. Eng., vol. 35, 2007, p. 264272.

[52] A. Jones, "Quantitative measurement of the alcohol concentration and the temperature of breath during a prolonged exhalation," Acta Physiol. Scand., vol. 114, 1982, p. 407412.

[53] S. George, A. Babb, and M. Hlastala, "Dynamics of soluble gas exchange in the airways: III. Single exhalation breathing maneuver," J. of Applied Physiology, vol. 75, 1993, p. 243949.

[54] 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, vol. 19, 2008, pp. 236–246.

[55] M. Kermit and O. Tomic, "Independent component analysis applied on gas sensor array measurement data," IEEE Sensors Journal, vol. 3, 2003, pp. 500–511.