POLITECNICO DI MILANO

Master of Science in Biomedical Engineering

Department of Electronics, Information and Bioengineering

POLITECNICO
MILANO 1863

DEVELOPMENT OF A DEVICE FOR THE AUTOMATIC SAMPLING AND SEPARATION OF EXHALED BREATH IN ELECTRONIC NOSE SYSTEMS

Supervisor: Prof. Pietro CERVERI

Co-Supervisor: Eng. Davide MARZORATI

Candidate:

Francesco DE GRAZIA

ID 875238

Academic Year 2018-2019

POLITECNICO DI MILANO

Master of Science in Biomedical Engineering

Department of Electronics, Information and Bioengineering

DEVELOPMENT OF A DEVICE FOR THE AUTOMATIC SAMPLING AND SEPARATION OF EXHALED BREATH IN ELECTRONIC NOSE SYSTEMS

Francesco De Grazia

Acknowledgments

First of all, I would like to thank my supervisor Prof. Pietro CERVERI for giving me the opportunity to work on this project and for his suggestions and support during these months.

Also, I must express my very profound gratitude to Eng. Davide MARZORATI for providing me with unfailing hints and continuous encouragement throughout the duration of this thesis and further. I will always be grateful for everything he taught me and for the kindness and patience he had in doing so. I wish him every success in life and academic career.

I would like to thank Eng. Matteo ROSSI for having been a good labmate and for all the stimulating discussions we shared. Best wishes for a good career to him, too.

A special thanks to Dr. Giulia SEDDA from the European Institute of Oncology for her clarifications and help about a complicate topic as lung cancer diagnosis.

Finally, I'm very thankful to my family, all my friends and my girlfriend for providing me with continuous support and encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Francesco DE GRAZIA

Abstract

Introduction Lung cancer is the worldwide leading cause of cancer death. Lung cancer survival is directly correlated with disease stage, and one the main causes of the high mortality rate is the late-stage tumor detection. Early diagnosis is therefore the key factor in significantly improving the overall survival rate.

Different screening tools, mainly based on imaging techniques, have been used for lung cancer diagnosis during the last 50 years. Low Dose Computed Tomography (LDCT) is the current pivotal approach because of its high accuracy, anyway its invasiveness and high false positive rate limit its application to high risk population screening. An alternative test appears to be necessary and, in order to make it widespread, it should be non-invasive, cost-effective and easy to use.

One of the most promising approaches being recently investigated in this field is the quantitative analysis of volatile organic compounds (VOCs) contained in the exhaled breath. Changes in the VOCs mixture may be directly related to the presence of a disease, since they reflect the metabolic activity of an individual.

Several technological solutions have been proposed as tools to analyze exhaled breath for lung cancer diagnosis, among which there are Electronic Nose (eNose) systems. Even if many of them achieved interesting results, there is still no validated clinical application of this technique, mainly because of the lack of standard guidelines and protocols.

Aim of the work Politecnico di Milano university, in collaboration with the Department of Thoracic Surgery of the European Institute of Oncology (IEO), is working on a prototype of eNose with the aim of overcoming the main limitations highlighted in the literature. In the context of this project, we present in this work the development of a device for the automatic sampling and separation of the exhaled breath.

The basic idea behind this device is that the separation of the two components of

exhaled breath, consisting of Anatomical Dead Space and Alveolar air, could increase the concentration of endogenous VOCs (those contained in alveolar air) with respect to a mixed air sample, thus improving the accuracy of the whole analysis. This separation is based on the real time monitoring of the CO_2 concentration in the airway gas and is automatic, thanks to the application of an algorithm able to switch an electrovalve with the proper timing.

Materials and methods The hardware of the system is composed by an electronic and a hydraulic part. A PSoC 5LP platform is used to control the sensing part, consisting of a fast-responding CO_2 sensor (based on NDIR principle), the actuating part, i.e., a three-way electrovalve, and a Bluetooth module used for wireless communication with the PC. Other components, mainly conditioning circuitry, LEDs and power supply, have been integrated on a circuit board. The hydraulic components are placed in a simple circuit: the user can exhale through a tee-mouthpiece assembly, then the air is directed through polyamide tubes to the CO_2 sensor first, and to the electrovalve later; the breath is finally collected in two Tedlar bags. In addition, a plastic case was 3D printed as an external protection and to fit all components in it.

The firmware was implemented in PSoC Creator, while the software was written in Python. After a first attempt of implementing an algorithm able to sample multiple breaths in succession, the device was simplified to be a single-breath sampler due to some practical limitations. Firmware and software together operate a standard routine. First of all, the system is powered up and it has to be connected via Bluetooth with the PC on which the software (a simple GUI) is running. Then the user can decide to change the plot settings, to save data in a file and when to start the acquisition. Once the acquisition is started, the subject has to maximally expire through the mouthpiece and the CO_2 concentration values are read from the sensor via UART, processed by a simple algorithm controlling the valve opening and plotted in the dedicated area of the GUI. In particular, the valve is switched when the reconstructed signal reaches a plateau. At the end of the exhalation, the operator has to stop the session and to lock the sampled bags outlet valve. Finally, to prepare the instrument for a new acquisition, the device has to be cleaned of the exhaled air remained in it until the CO_2 sensor reaches its baseline (this operation can be performed by inspiring the air out of the device while monitoring the CO_2 concentration on the GUI, or using an external pump). All operations are guided by the software, and two LEDs indicate the status (ON/OFF) of the device and the current phase of the acquisition (Waiting, Measuring or

Cleaning).

Results To test the effectiveness of the separation, a trial involving ten nonsmoking healthy volunteers (aged 24 to 26) was held. The subjects followed a hygienic procedure from the day before the test, intended to reduce the interferences in breath composition. Each of them carried out three acquisitions: 2 with the breath sampler and 1 of mixed-expiratory air (i.e., without performing any kind of separation), sampling a total of 5 bags per person. The sampled bags were subsequently analyzed with the electronic nose prototype under current development as part of the collaboration between IEO and POLIMI, containing five electrochemical gas sensors sensible to different combinations of VOCs.

The recorded data were then processed in MATLAB to check if the patterns relative to dead space (DS), alveolar air (AV) and mixed-expiratory breath (ME) could be classified as three different contributions. Each signal was low-pass filtered and five features were extracted from every single waveform. Those features, taken alone, were sufficient to consider as statistically different the DS and AV classes, but not the AV and ME ones. To reduce data dimensionality (which is high because of the presence of 5 sensors), Principal Component Analysis (PCA) was applied and each feature was projected in the space defined by the first two principal components. Again, the three classes were not fully separable, thus the next step was to use PCA on all five features taken together. Thanks to this last operation, the samples formed distinct regions in this new bidimensional space. At last, the samples were classified by making use of Discriminant Analysis. Two models were fitted, one linear (LDA) and one quadratic (QDA), and used to compute the separation boundaries between DS, AV and ME acquisitions. Linear boundaries were sufficient to discriminate with success the different breath contributions, hence a quadratic model was not necessary in this study.

Conclusions The developed device operates as a single-breath sampler able to separate the dead space and the alveolar contributions of a maximally forced expiration. The obtained results demonstrate that the two portions of air, if analyzed using an electronic nose, produce two signals which are different between them, and also from an exhaled breath sampled without any kind of separation. Even if this conclusion cannot be considered as a validation of the instrument, it is encouraging and may be taken as a starting point to investigate the question more in depth. In future works, the system could be improved to sample multiple breaths and to be more reliable, while the cleaning phase should be made easier and faster. Furthermore, new studies should include validation tools and involve a larger population.

Sommario

Introduzione Il cancro al polmone è la principale causa di morte per cancro in tutto il mondo. La sopravvivenza a questo tumore è direttamente correlata con lo stadio della malattia al momento della diagnosi, e una delle cause a cui è dovuto l'alto tasso di mortalità è la diagnosi tardiva. La diagnosi precoce è pertanto il fattore chiave per aumentare in maniera significativa il tasso di sopravvivenza dei pazienti.

Nel corso degli ultimi 50 anni sono stati utilizzati diversi strumenti di screening per il tumore al polmone, basati principalmente su tecniche di imaging; la tomografia computerizzata a basso dosaggio (LDCT) è l'approccio cardine attuale grazie alla sua elevata accuratezza, tuttavia la sua invasività e l'alto tasso di falsi positivi limitano la sua applicazione allo screening della popolazione ad alto rischio. Un test alternativo sembra necessario e, al fine di renderlo diffuso, esso dovrebbe essere non invasivo, economicamente conveniente e di facile utilizzo.

Uno degli approcci più promettenti, tra quelli recentemente in fase di studio, è l'analisi quantitativa dei composti organici volatili (VOCs) contenuti nell'esalato. Variazioni nella miscela di VOCs possono essere direttamente correlate alla presenza di una patologia, in quanto riflettono l'attività metabolica del paziente analizzato.

Diverse soluzioni tecnologiche sono state proposte come strumenti di analisi dell'esalato per la diagnosi precoce del cancro al polmone; esse sono conosciute col nome di Nasi Elettronici. Anche se molti dei sistemi presenti in letteratura hanno raggiunto buoni risultati, non esiste ancora un'applicazione clinica convalidata di questa tecnologia, principalmente a causa della mancanza di standard e linee guida.

Scopo della tesi Il Politecnico di Milano, in collaborazione con il Dipartimento di Chirurgia Toracica dell'Istituto Europeo di Oncologia (IEO), sta lavorando a un prototipo di naso elettronico con l'obiettivo di superare i principali limiti evidenziati in letteratura. Nel contesto di questo progetto, presentiamo in questo lavoro di tesi lo sviluppo di un dispositivo per il campionamento e la separazione automatica dell'esalato.

L'idea alla base di questo dispositivo è che la separazione delle due porzioni dell'espirato, ossia lo spazio morto anatomico e l'aria alveolare, possa aumentare la concentrazione dei VOCs endogeni (quelli contenuti nell'aria alveolare) rispetto ad un campione di aria mista, e quindi migliorare la sensibilità dell'intera analisi. Tale separazione si basa sul monitoraggio in tempo reale della concentrazione di CO_2 nel respiro ed è automatica, grazie all'applicazione di un algoritmo in grado di commutare un'elettrovalvola con la giusta tempistica.

Materiali e Metodi L'hardware del sistema è composto da una parte elettronica ed una idraulica. Un microcontrollore PSoC 5LP è impiegato per gestire sia la parte di rilevamento, costituita da un sensore di anidride carbonica a risposta rapida (basato sul principio NDIR), che la parte di attuazione, consistente in un'elettrovalvola a tre vie, più un modulo Bluetooth, utilizzato per la comunicazione wireless con il PC. Altri componenti, principalmente circuiteria di condizionamento, LED e alimentazione, sono stati integrati su un circuito stampato. La parte idraulica è meno articolata: l'utente espira attraverso un boccaglio, quindi l'aria viene diretta, attraverso dei piccoli tubi in poliammide, prima al sensore di CO_2 e poi all'elettrovalvola; il respiro viene infine raccolto in due sacche in Tedlar. Inoltre, un case di plastica è stato stampato in 3D per contenere tutti i componenti.

Il firmware è stato implementato in PSoC Creator e, insieme ad un software scritto in Python, gestisce una routine standard. Dopo un primo tentativo di algoritmo in grado di operare su più respiri in successione, a causa di alcune limitazioni pratiche il dispositivo è stato semplificato per essere un campionatore di singolo respiro. Prima di tutto, una volta acceso il sistema, esso deve essere collegato via Bluetooth con il PC su cui è in esecuzione il software (che consiste in una semplice interfaccia grafica). Quindi l'utente può decidere se modificare le impostazioni del grafico, se salvare i dati in un file e quando avviare l'acquisizione. Una volta avviata l'acquisizione, il soggetto deve eseguire un'espirazione forzata massimale tramite il boccaglio; i valori di concentrazione di anidride carbonica vengono letti dal sensore tramite UART, elaborati da un semplice algoritmo che controlla l'apertura della valvola e tracciati sull'apposito grafico presente nella GUI. In particolare, la valvola viene commutata quando il segnale ricostruito raggiunge un plateau. Terminata l'espirazione, l'operatore deve interrompere la sessione e sigillare la valvola d'apertura delle sacche appena riempite. Infine, per preparare lo strumento ad una nuova acquisizione, è necessario rimuovere l'aria satura di CO_2 rimasta nel dispositivo al fine di riportare il sensore al suo valore di base (questa operazione può essere eseguita inspirando l'aria dal dispositivo mentre si monitora la situazione dalla

GUI, oppure utilizzando una pompa esterna). Tutte le operazioni sono guidate passo passo dal software, mentre due LED indicano lo stato (ON/OFF) del dispositivo e la fase corrente dell'acquisizione (attesa, misurazione o pulizia).

Risultati Per testare l'efficacia della separazione effettuata dal dispositivo, è stato condotto uno studio su dieci volontari sani non fumatori (di età compresa tra 24 e 26 anni). I soggetti hanno seguito una procedura igienica dal giorno prima del test, allo scopo di ridurre le interferenze nella composizione del respiro. Ciascuno di essi ha effettuato tre acquisizioni: 2 con il separatore di respiro ed una senza effettuare la separazione, ottenendo un totale di 5 sacche per persona. Ogni sacca è stata successivamente analizzata con il prototipo di naso elettronico in sviluppo per la collaborazione tra IEO e Politecnico di Milano, contenente cinque sensori di gas elettrochimici sensibili a diverse combinazioni di VOCs.

I dati registrati sono stati quindi elaborati in MATLAB per verificare se i pattern relativi a spazio morto (DS), aria alveolare (AV) e respiro misto-espiratorio (ME) possano essere classificati come tre contributi distinti. Ciascun segnale è stato filtrato con un filtro passa-basso e cinque feature sono state estratte da ogni singola forma d'onda. Queste feature, prese da sole, sono sufficienti per considerare come statisticamente differenti le classi DS e AV, ma non quelle AV e ME. Per ridurre la dimensionalità dei dati (che è elevata a causa della presenza di 5 sensori), è stata applicata l'Analisi delle Componenti Principali (PCA) e ciascuna feature è stata proiettata nello spazio definito dalle prime due componenti. Ancora una volta, le tre classi non sono risultate completamente separabili, quindi il passo successivo è stato quello di utilizzare la PCA sulle cinque feature prese insieme. Grazie a quest'ultima operazione, i dati si sono distribuiti in regioni distinte nel nuovo spazio bidimensionale. Infine i campioni sono stati classificati tramite l'applicazione dell'Analisi Discriminante. Due modelli, uno lineare (LDA) e uno quadratico (QDA), sono stati addestrati e utilizzati per calcolare le linee di separazione tra le classi DS, AV e ME. I margini lineari sono risultati sufficienti per discriminare con successo i diversi contributi del respiro quindi, in questo studio, un modello quadratico è superfluo.

Conclusioni Il dispositivo sviluppato funziona come un campionatore di respiro singolo ed è in grado di separare lo spazio morto dall'aria alveolare durante l'acquisizione di un'espirazione forzata massimale. I risultati ottenuti dimostrano che i due contributi dell'esalato, se analizzati da un naso elettronico, producono due segnali diversi tra loro, e anche da un campione misto-espiratorio (cioè acquisito senza alcun tipo di separazione). Anche se questa deduzione

non può essere assunta come una validazione dello strumento, essa è incoraggiante e può essere considerata come un punto di partenza da cui approfondire la questione. Nei lavori futuri, il sistema potrebbe essere migliorato al fine di campionare più respiri in serie o, perlomeno, per avere maggiore affidabilità, mentre la fase di pulizia dovrebbe essere resa più semplice e immediata. Inoltre, possibili nuovi studi dovrebbero includere l'utilizzo di strumenti di convalida e coinvolgere un campione di popolazione più ampio.

Contents

List of Figures

List of Tables

List of abbreviations

ANN	Artificial Neural Network
AUC	Area Under the Curve
AV	Alveolar
CDA	Canonical Discriminant Analysis
CSA	Colorimetric Sensor Array
COPD	Chronic Obstructive Pulmonary Disease
DA	Discriminant Analysis
DS	Dead Space
EBC	Exhaled Breath Condensate
FENO	Fraction of Nitric Oxide
GC-MS	Gas Chromatography-Mass Spectrometry
ICA	Independent Component Analysis
IDE	Integrated Development Environment
IMS	Ion Mobility Spectrometry
LC	Lung Cancer
LDA	Linear Discriminant Analysis
LDCT	Low Dose Computed Tomography
LOOCV	Leave One Out Cross-Validation

ME	Mixed-Expiratory
MOS	Metal Oxide Semiconductors
NDIR	Non-Dispersive Infrared
NLST	National Lung Screening Trial
NSCLC	Non-Small-Cell Lung Cancer
PCA	Principal Component Analysis
PCB	Printed Circuit Board
PLA	Polylactic Acid
PN	Pulmonary Nodule
ppb	Parts per billion
ppmv	Parts per million volume
PLS-DA	Partial Least Squares Discriminant Analysis
QDA	Quadratic Discriminant Analysis
QMB	Quartz Microbalance
ROC	Receiver Operating Characteristic
SCLC	Small-Cell Lung Cancer
SPME	Solid Phase MicroExtraction
SVM	Support Vector Machines
VOCs	Volatile Organic Compounds

Chapter 1

Introduction

1.1 Background

Cancer is a generic term for a large group of diseases that can affect any part of the body. According to the World Health Organization, cancer is the second leading cause of death globally, and it is responsible for an estimated 9.6 million deaths in 2018 [1].

Between 30-50% [1] of cancers can be prevented by avoiding risk factors and adopting existing evidence-based prevention strategies, while others have a high chance of survival if diagnosed early and treated adequately. Anyway, in many cases the mortality is still high, as shown in Table 1.1. Lung cancer (LC) hence represents both the most spread and the deadliest cancer worldwide, and the main reason for this is its late diagnosis. If LC is diagnosed at an early stage, when small and before spreading, people have a better chance of living longer. Patients are usually diagnosed at late stages when curative treatment may no longer be effective (the five-year survival rate for early detection lays between 70 and 90%, while it falls down to 12% in case of late diagnosis [2]). Consequently, early diagnosis is a key factor to significantly improve the overall survival in high risk populations of asymptomatic patients.

Lung cancers can be divided into benign and malignant. Malignant tumors are classified into two main groups, small cell carcinomas (small-cell lung cancer, SCLC), which represent about 15-20% of total lung cancers, and non-small cell carcinomas (non-small-cell lung cancer, NSCLC) that are about 70%. NSCLC are further divided into three histological types: adenocarcinomas (50%), squamous cell or epidermoid (30%) and large cell cancers (10%). Hence, in presence of a suspected lung cancer, it is necessary to go through an appropriate process that provides a careful diagnosis based on cytology and/or histology (typing) as well as a careful assessment of the disease's extent (staging) [3].

Table 1.1: *Incidence (million cases) and mortality (million deaths) of the major cancers in 2018 according to WHO [1].*

	Cancer types						
	Lung	Breast	Colorectal	Prostate	Skin	Stomach	Liver
Incidence	2.09	2.09	1.80	1.28	1.04	1.03	N.A.
Mortality	1.76	0.627	0.862	0.782	N.A.	0.783	0.782

The symptoms of lung cancer strictly depend on the anatomical location of the disease, the level of aggressiveness and the type of growth. The most common symptoms consist of: persistent cough, dyspnea, chest pain, haemoptysis (coughing producing blood) and dysphonia (voice alteration). Usually these symptoms do not appear until the disease is already at an advanced stage. Even when LC shows symptoms, it is not easy for the physician or the patient to attribute them to lung cancer, since they may be incorrectly linked to other diseases, such as an infection or long-term effects from smoking. The only true primary prevention (risk factor reduction) is to quit smoking and reduce environmental exposure to known carcinogens. Secondary prevention (early detection of diseases already in progress) instead includes population screening, early diagnosis and treatment of pre-neoplastic lesions, but in the case of lung cancer has unfortunately not shown great efficacy [4].

The LC screening process has evolved over the last 50 years. Various studies for population screening were carried out using chest X-ray, CT and sputum cytology, with the chest X-ray being the most common diagnostic tool until 2011 [5]. In 2011 the results of a consistent study (named National Lung Screening Trial, NLST) were published: 53454 subjects underwent three rounds of low-dose computed tomography (LDCT) annual screening combined with chest radiographs. This randomized clinical screening trial showed how LDCT, compared to chest radiography, was able to reduce the LC mortality of 20% [6]. The LDCT technique appears to be relatively simple, but some of its settings are important and should be accurately defined in order to achieve good diagnostic quality and minimize the delivered dose. In addition, LDCT examinations are not as easy to read as they may initially appear and the management of positive results can be a complex process [7]. Screening with LDCT is also known to have some downsides that need to be considered. One drawback of this test is that it also finds a lot of abnormalities that have to be checked out with more tests, and that could also turn out not to be cancer. This may lead to additional tests, for instance other CT

scans or more invasive tests such as needle biopsies or even surgery to remove a portion of lung in some people. These tests can sometimes lead to complications (like a collapsed lung), or rarely death, even in people who do not have cancer (or who have very early stage cancer). LDCT scans also expose people to a small amount of radiation with each test. It is less than the dose from a standard CT, but it is more than the dose from a chest x-ray. Some people who are screened may end up needing further CT scans, which means more radiation. When done in tens of thousands of people, this radiation may cause a few people to develop breast, lung, or thyroid cancers later on [4]. In NLST, after three rounds of screening, approximately 39.0% of subjects were classified as positive with 96.4% of these being a false positive. This high false positive rate, togheter with the cost-benefit profile of LDCT screening, is the main limitation in the applicability of this technique to large scale populations.

For all the reasons cited up to now, there is nowadays an increasing need for an alternative method to allow earlier detection of lung cancer and screening of a larger population. In accordance with these purposes, the ideal test should be non-invasive, cost-efficient and easy to use. Within this scenario, different solutions have been proposed and one of the most promising is exhaled breath analysis. The potential of exhaled breath analysis has long attracted interest in the areas of medical diagnosis and disease monitoring because of its non-invasive nature, the availability of an unlimited sample supply (i.e. breath), and the potential to facilitate a rapid diagnosis. Breath odours were used for disease recognition even as early as Roman times, when the smell of a person's breath was smelt by physicians to associate uncontrolled diabetes with a sweet, acetone odour, liver failure with a fish-like smell and renal failure with a urine-like smell [8]. McCulloch et *al.*, more recently, demonstrated that dogs could be trained to detect lung and breast cancer in subjects with various stages of disease with almost 100% accuracy merely by smelling the subject's breath [9]. All these observations suggest that there are biomarkers in exhaled breath that are potentially useful for disease diagnosis. Over the last 40 years there have been many studies aiming at characterizing these biomarkers, and many other reports detailing differences in the way tumor tissue handles energy and metabolizes specific classes of compounds when compared with healthy tissues. More specifically, the composition of the exhaled breath reflects metabolic activity within the body; biological processes within the cells lead to the consumption and production of metabolic byproducts, which can circulate within the blood and transfer to the lungs where they are exhaled from the body. Thus, when cancer develops, various inflammatory processes as well as gene and protein changes take place creating a unique biomarkers profile, altered

with respect to the healthy condition, that is potentially reflected in the body fluids and finally in the breath [10]. Evidence that different volatile patterns occur in affected individuals has been claimed and that the presence of some of them may be correlated specifically to the lung cancer was found. Moreover, numerous volatile substances were identified, such as acetone, isoprene, benzene, xylene, pentane, ethanol and methanol [11]. Nonetheless, these studies still did not lead to a diagnostic method due to the overwhelming complexity of the analysis carried out by Gas Chromatography-Mass Spectrometry (GC-MS).

In the last decades the introduction of chemical sensors allowed to reconsider these studies in order to check whether novel diagnostic tools based on the chemical information may be set out. Owing to the variety of sampling methods and analytics, the field developed in a largely unrelated manner in three main domains: exhaled breath condensate (EBC), volatile organic compounds (VOCs) and FENO (fraction of nitric oxide in expired gas). A fourth area, that of exhaled particles, came later [12]. Within these solutions, the most promising (and the one adopted in this work) is the analysis of VOCs. Breath VOCs can be endogenous and exogenous, with the former produced by internal processes of the body and the latter introduced in the body from external sources, such as food and environmental pollution. VOCs are found at trace levels, typically parts per million volume (ppmv) and lower, thus their reliable detection poses a challenge. GC-MS studies have shown that several VOCs, which normally appear at levels of 1-20 ppb in healthy human breath, are elevated to levels between 10 and 100 ppb in lung cancer patients. To date no compound has been found as being present only in lung cancer patients exhaled breath [13], hence it is a combination of VOCs that could be diagnostic for lung cancer, rather than a unique VOC. Due to this, the new instruments are composed of arrays of non-selective chemical sensors: the sensors response is not univocally correlated with the concentration of a single compound, but rather it is a sort of combination of all the chemical information contained in each sample, somewhat resembling to the functioning of the human olfaction with odorants. Since the introduction of these arrays, now widely known as electronic noses (or eNoses), they have been applied to many different fields including medical diagnostics [14].

Several eNose systems have been proposed during the last years, with some of them reaching good results in terms of sensitivity and accuracy. Anyway, several patient-specific factors, such as tobacco smoking and comorbidities, are complicating an apparently easy analysis and could alter the results [15]. Furthermore, there are no standard guidelines and protocols that allow to compare the results and to combine the positive findings of each

approach. Because of these reasons, up to now none of these devices has been validated yet for the clinical application.

1.2 Thesis Proposal

In 2018 the Department of Thoracic Surgery of the European Institute of Oncology (IEO) of Milan, which gained experience in this field during the last decade, started a collaboration with the Politecnico di Milano university with the goal to develop a prototype of eNose that could be converted into clinical practice in the next few years. As part of this project, the work presented in this thesis consists in the development of a device for the automatic sampling and separation of the exhaled breath in two parts, namely the Anatomical Dead Space and the Alveolar (or End-Tidal) air. This separation is performed in such a way to preserve as analytical sample only the air contribution that takes part in the gas exchange at the level of the alveoli, representing the metabolic activity in the body. The idea is that, by applying this separation, only interesting VOCs would be analyzed and their concentration will be higher with respect to samples of mixed air, thus leading to a more selective test.

A device with this purpose has to be based on an electronic unit, some gas sensors and a few hydraulic components. The key part of the system should be a fast response CO_2 sensor to employ for the monitoring of the carbon dioxide concentration in the exhaled air; the expected waveform, known as Time Capnogram, is needed to identify the different phases of the respiratory cycle. Thanks to this information, a three way valve has to be controlled in such a way to perform the claimed separation. In addition, the device should be wireless (exploiting a Bluetooth connection) and equipped with a proper software for user interfacing.

1.3 Expected Results

According to the thesis proposal, what we could expect from this work is the realization of a functioning prototype and an easy-to-use software. As we will see in detail in Chapter 2, in the state of the art there are not many studies performing breath separation based on CO_2 monitoring, and only few of them tried to assess the effectiveness of the procedure. This lack of a golden standard, added to the impossibility to employ expensive instrumentation like GC-MS or professional CO_2 monitors for an exploratory work like this, limit the possibility to really have a validation of the device. Furthermore, at least at this stage of the work, it is worth to involve in the analysis only healthy volunteers, postponing the inclusion of

cancerous patients to a later step of the research. Given all these premises, is reasonable to expect mainly these behaviors:

- the device can correctly separate the exhaled air in two portions for a single breath and, if possible, for multiple breathings;
- the collected dead space and alveolar samples, when analyzed using the Electronic Nose prototype, show a significant difference in the sensor response both between them and compared to a mixed expiratory sample;
- the collected samples show a certain repeatability in the gas sensors response.

If these preliminary results will be satisfactory, it could be taken into consideration to integrate such a separator in a complete eNose prototype and in future to include it in a wider study involving lung cancer patients.

1.4 Thesis Outline

The present document describes in detail the whole thesis work and consists of 5 chapters. The current chapter served as an introduction to the concept of Exhaled Breath analysis, the reasons behind its necessity and to define the goal of the project. In Chapter 2 a review of the most recent systems for exhaled breath analysis in the field of lung cancer diagnosis is presented, with a focus on the technological solutions that have been proposed and in particular on the lack of valid automatic breath samplers. In Chapter 3 a systematic description of the Hardware, Firmware and Software solutions is provided. In Chapter 4 the results of the performed tests are exposed. Finally, in Chapter 5, the results are discussed and future developments are proposed.

Chapter 2

Technologies for Exhaled Breath Analysis

2.1 General Pipeline of Exhaled Breath Analysis

This first section is intended as an overview of the whole procedure to conduce exhaled breath analysis for lung cancer diagnosis and to explain some of the main choices to face in each step of the experimental study. This is a necessary step to prepare the reader to the more detailed descriptions reported in the next pages.

The general pipeline of exhaled breath analysis can be broadly broken down into three steps: breath sample collection, sample analysis, and data analysis [16]. There are several ways of achieving the desired goal in each section. For example, for breath sample collection, factors such as the type of breath to be collected (i.e., mixed expiratory or end-tidal), single or multiple exhalation, and choice of breath capture technology are just some of the options to be considered. A general schema is reported in Figure 2.1.

2.1.1 Breath Sample Collection

Breath sample collection is a central topic in the current work. Many issues could influence the composition of VOCs in the acquired sample, such as the type of sampling (total versus alveolar breath), the sampling duration (single-breath versus fixed-time or fixed-volume breathing), effect of expiratory flow and breath hold, food/medications, smoking and comorbidities.

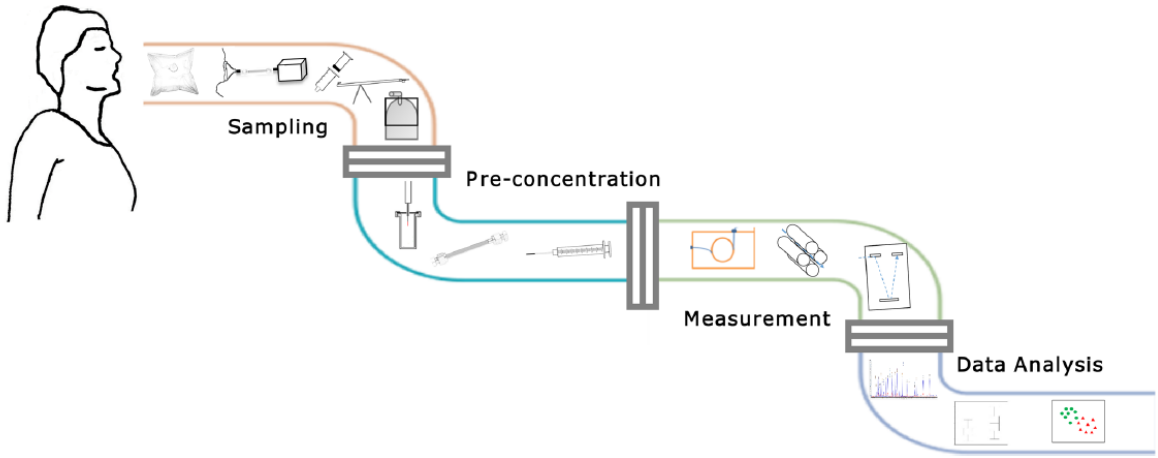


Figure 2.1: *General breath sampling pipeline. The first section shows the breath sampling containers, the second the pre-concentration methods (when used), the third the most common measurements instruments and the fourth the results of data processing and analysis. Image taken from [17]*

As introduced in Chapter 1, the broad topic of this work is the sampling of alveolar breath. In general, during breath sampling, there is a choice made as to the portion of the breath that can be collected, and this can be broadly divided into late expiratory, end-tidal (or alveolar), and mixed expiratory (total breath including anatomical dead-space air). Breath phases can be identified by capnography, i.e. by monitoring the CO_2 concentration during time, as represented in Figure 2.2.

Before proceding to the explanation of the three breath phases, a brief description of dead space and alveolar air volume is necessary. **Dead space** represents the volume of ventilated air that does not participate in gas exchange. It is the volume of air filling the conducting zone of respiration, made up by the nose, trachea, and bronchi (about 150 mL). **Alveolar air**, instead, is the volume of air in the respiratory zone (i.e., bronchioles, alveolar duct, alveolar sac and alveoli) that does take part in gas exchange [18]; its amount is dependent on the alveolar ventilation. A schematic drawing is reported in Figure 2.3.

2.1.1.1 Late expiratory breath

Late expiratory breath sampling involves discarding the initial portion of exhaled breath (estimated dead space) and the subsequent capture of air at the end of the breath cycle. Minimization of dead space (Phase I in Fig. 2.2) sampling allows a greater relative contri-

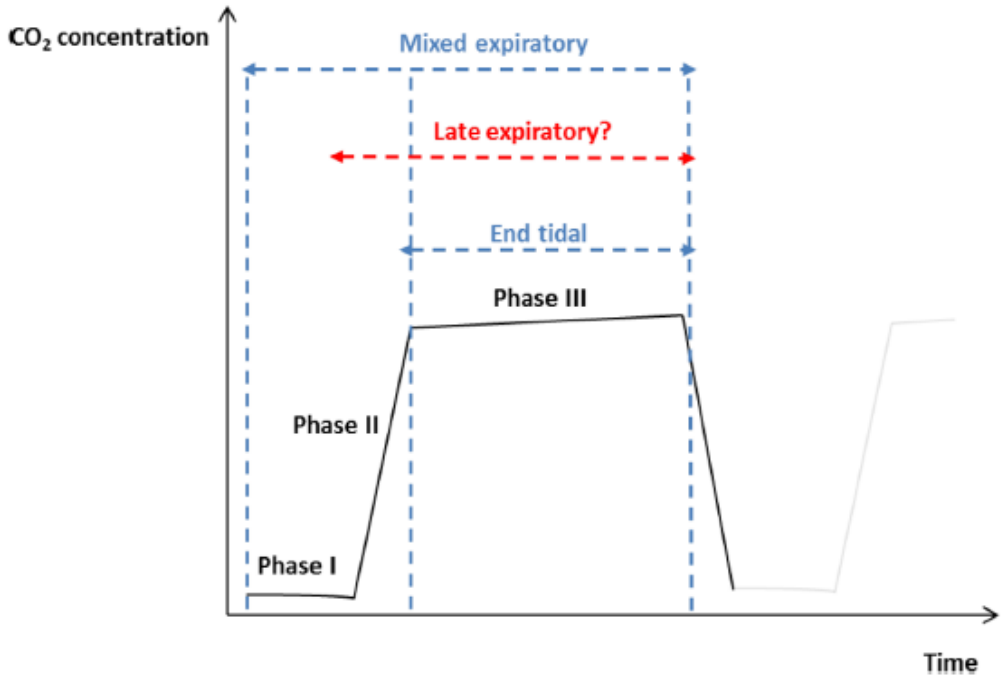


Figure 2.2: *Schematic visual representation depicting a single exhaled breath phases by capnography. Phase I is dead space, Phase II transition, Phase III alveolar air. Image taken from [17].*

bution of endogenous VOCs in the resultant sample, as well as a reduction of the levels of exogenous VOCs. In some methods this simply mandates excluding the first few seconds of exhalation from an individual before the breath sample is collected [19, 20]. Time-controlled breath samples have been shown to be unreliable [21], and there is no known optimal exclusion time duration, also with various timings used in different studies. Concerns regarding reproducibility also arise due to distinct physiological properties of individuals such as cardiac output and pulmonary ventilation which may also introduce unwanted variability even within individuals sampled repeatedly in different physiological states [22]. Thus, with several concerns associated with this type of breath, more effort is still required before late expiratory breath sampling would be suitable for use in the clinic. The ideal system would need to adapt to the current physiological state of each individual to collect a representative sample and minimise dead space contamination, but this would be at the cost of the simplicity.

2.1.1.2 Alveolar (or End-tidal) breath

End-tidal or 'Alveolar' breath indicates the air collected from the start to the end of phase III of the breath cycle. This type of air is stated to contain high concentrations of endogenous

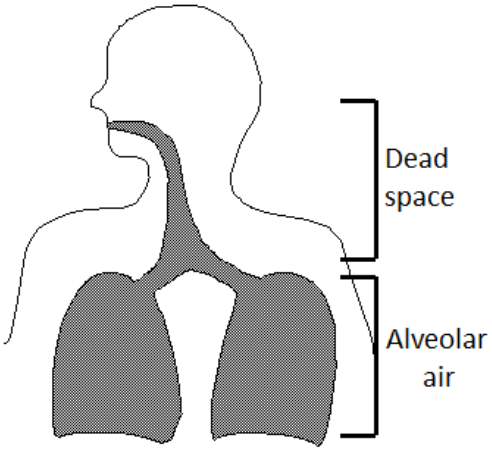


Figure 2.3: *Schematic anatomical representation of dead space and alveolar volume in the airways. Image adapted from http://www.raosyth.com*

VOCs and minimal contaminants. This type of breath differs from late expiratory in terms of the confidence of obtaining a representative (and personalized) end-tidal sample, using e.g. a visual cue (such as the real-time capnogram) to collect air only from phase III. CO_2 visual control accounts for the most common method used to collect end-tidal breath and it involves monitoring CO_2 concentrations during exhalation. During phase I of exhalation, CO_2 levels are generally low but rise during transition (phase II) and subsequently approaches a plateau signaling the start of alveolar phase III. Breath CO_2 levels can be monitored via a device known as a capnometer that enables visualization of the various phases for guidance on when to begin breath capture. There are devices which allow manual removal of air when alveolar phase has been reached [21, 23] and also trying automatic capture [24]. Selective sampling of the alveolar compartment can reduce oral contaminant concentrations (i.e. exogenous VOCs), but is technically more demanding [12]. The great benefit is that, since breath collection is adapted to each individual as samples are collected at phase III, variability regarding collection of samples is minimized and thus it would be more suitable for use in a clinic due to the availability of a benchmark. There are still many steps before clinic adoption, but at least it might enable comparison of data between laboratories and studies.

2.1.1.3 Mixed expiratory breath

Mixed expiratory breath, finally, can be considered as the simplest type of breath that can be obtained since it involves acquiring all phases of exhaled air. It may be an attractive option

due to its simplicity, however it may not provide the best quality of breath sample due to a greater abundance of environmental, mouth, and nose contaminants. Although there are feature selection models to aid in candidate biomarker selection, unless subject numbers are very large, there may still be a considerable chance of a false positive result in this situation; i.e., identifying an exogenous VOC as a candidate marker.

In summary, the ideal breath sampling method would be simple, tailored to personal physiology, allowing targeted selection of airway and/or alveolar air and eliminating sampling from the dead space and environment. This combination is not yet possible, and all current methods necessitate compromise in one or more areas. Possible improvements could be obtained by using the controls that are available such as capnography or a well defined late expiratory breath protocol.

2.1.2 Breath Sample Containers

The solution adopted in the great majority of studies consists in the temporary storage of breath samples in polymer bags prior to the analysis. Polymer bags thus encompass the majority of breath collection containers, of which Tedlar® bags (E.I. du Pont de Nemours and Company, Wilmington, DE, USA) are the most commonly used [17]. These bags are equipped with a push lock valve to avoid air leakage, as shown in Figure 2.4.

Figure 2.4: *Example of commercial Tedlar bags. Image taken from: www.indiamart.com*

Other types of bags have been used in many studies, such as Mylar bag (made by VOCs chemically inert materials) [25, 26], Bio-VOC [27, 28], ultra-clean balloons (equipped with

a charcoal reservoir for the removal of contaminant VOCs) [29], ALTEF polypropylene bag [30], Gas bulbs [13, 31] or Aluminium gas bag [32]. None of them has anyway emerged to be more reliable than the others, since no standard procedure was established for their use and cleaning, making difficult their comparison. Several cleaning techniques are reported in the literature, with some authors using nitrogen and others employing argon. Furthermore, there are no guidelines about the number of times the bags should be cleaned neither referring how long they could be stored before the analysis without affecting their content [33].

Direct breath sampling onto pre-concentration materials is also possible [34], but less used. VOC concentrations are generally in the parts per billion to parts per trillion (nM to pM) range, for this reason pre-concentrations of breath samples by adsorption onto sorbent traps or coated fibres may be required to provide a sufficient signal on the used analyzer [12]. In addiction, there is the isolated case of the Aenose device which does not require a temporary storage for the breath sample since the exhaled air is analyzed in real time.

As conclusion, the ideal collection container for clinical practice should be cost-effective, user-friendly, durable, inert and importantly allow neither ingress of environmental nor egress of breath VOCs.

2.1.3 Sample Analysis

Exhaled breath analysis can be roughly split into two main streams:

1) an analytical molecule identification-based stream;

2) a sensor technology, pattern recognition-based stream.

The analytical mass spectrometry (MS) track, often coupled to a separation technique like gas chromatography (GC), is focused on identifying biomarker compounds related to particular disease conditions and the accompanying pathophysiology. The second is a cross-reactive sensor technology that is purely based on pattern recognition of complex mixtures. This is represented by eNose technologies that are used for probabilistic predictive values in relation to health and disease. The difference between the GC-MS and eNose are several, starting from the financial cost to the VOC analysis, but one of the most important factors is that eNoses are able to produce a characteristic fingerprint from the pattern recognition of VOCs (the so called 'breathprint'). This can differentiate healthy controls from individuals affected by lung cancer, while GC-MS aims to identify a specific compound [35]. The description of the

solutions adopted for both the streams needs a detailed dissertation, therefore it is postponed to Section 2.2.

A third alternative which is worth to be mentioned is canine scent detection. Dogs have a highly developed sense of smell with a detection threshold at several parts per trillion. Although canine scent detection by trained dogs seems relatively simple and inexpensive, relatively few data have been published. McCulloch in 2006 was the first to use dogs to detect lung cancer [9]. He trained five dogs to identify exhaled breath samples of subjects with lung cancer and breast cancer. The sensitivity of the canine detection technique for biopsy-confirmed lung cancer (n=55) was 99%, with 99% specificity, while in breast cancer (n=31) the sensitivity was 88% and specificity 98%, with equal accuracy scored by all dogs. Ehmann *et al.* instead showed that dogs were able to identify lung cancer with sensitivity 71% and specificity 93% [36].

2.1.4 Data Analysis

Exhaled breath data are mainly time series and their analysis presents several alternatives, due primarily to the great availability of signal processing and pattern recognition techniques and secondly to the variability of data acquisition methods. The process is usually divided in two phases:

- data pre-processing

- real data analysis and classification

Pre-processing is a preparatory phase which is not mandatory but always recommended to enhance the quality of data and compensate for data-gathering non-idealities. The operations to be performed strongly depend on the adopted acquisition technique, but the most of the times consist in data cleaning, transformation and filtering, dimensionality reduction and other kinds of compensation. Data cleaning, transformation and filtering are performed to correct for noisy signals, eliminate outliers or false data and turn them in the form suitable for post-processing. Dimensionality reduction instead is necessary in presence of multivariate data (such as recordings from multiple sensors) to reduce their complexity and the number of variable to process. The methods used in this context belong to the family of Blind Source Separation techniques, of which the most commonly used in the field of breath analysis are Principal Component Analysis (PCA) [26] and Independent Component Analysis (ICA), as we will see in Section 2.2.

The second step is the central part of the process, in which pattern recognition techniques are used to classify the patient as positive (cancerous) or negative (healthy). The possibilities are several, including Artificial Neural Network (ANN) [37, 38], Logistic Regression, Linear Discriminant Analysis (LDA), Decision Trees [39] and Support Vector Machines (SVM) [26]. Each of these techniques has been used in literature to analyze breath samples collected with electronic noses, but none of them has still been elected as a reference. Their applicability depends on the amount of acquired data and the complexity of the model to be developped, and the process of validation of the fitted model is not always possible because of the reduced dimension of the datasets. In many cases this issue is overcome by making use of advanced machine learning techniques such as cross-validation and bootstrap. The most common method in electronic nose studies is leave one out cross-validation (LOOCV), the extreme case of k-fold cross-validation. The idea of k-fold cross-validation is to randomly divide the data into k equal-sized parts, leave out part k, fit the model to the other k-1 parts (combined), and then obtain predictions for the left-out k^{th} part. This is done in turn for each part k = 1,2,. . . ,K, and then the results are combined; setting K equal to the total number of available data yields n-fold or leave-one out cross-validation. Only a couple of studies attempted to validate the model on a large scale dataset (more than 1000 subjects), but even in these cases LC patients represented only a small portion of the total number of subjects (3% and 10% [40, 41]).

The lack of a model validated on a large scale dataset represents not only a deficiency for the comparison of results between the various research studies, but most of all the final limitation for the potential application of the technology in clinical practice.

2.2 Major Systems and Studies in the State of the Art

In Section 2.1 the reader had a general description of how exhaled breath analysis can be carried on. Now we are going to focus our attention on the devices used to sample and analyze exhaled breath in the most relevant studies of last decades, in particular on electronic noses. Such a review is important to understand the basic working principle of every system and the encountered limitations that research aims to solve in the next future.

Breath sample analysis, as mentioned in paragraph 2.1.3, is mainly divided in an analytical current and a sensor technology stream (based on pattern recognition of VOCs breathprints).

In this section the solutions adopted in literature will be analyzed more in detail.

2.2.1 Gas Chromatography and Mass Spectrometry (GC-MS)

Gas Chromatographs and Mass Spectrometers are sensitive and highly accurate instruments for gaseous molecules separation and recognition, for this reason they have been the first to be utilized in literature to analyze the low concentrations of VOCs (parts per billion, ppb) in breath samples. Furthermore, they are often used in combination with pre-concentration methods such as solid phase microextraction (SPME). The aim of the early studies, in fact, was the identification of those compounds in the airway gas (if any) able to differentiate a cancerous from an healthy patient.

In the GC-MS method, in short, exhaled breath is collected and temporarily stored in designated containers (e.g., in inert bags or sorption tubes). Then, a helium stream is used to carry the sample through a long, heated capillary column where the VOCs are separated based on their chemical properties (GC). Molecules are consecutively ionized, separated by their mass/charge (m/z) ratio (MS) and finally identified by a spectral library in the software [10]. The entire path is schematized in Figure 2.5.

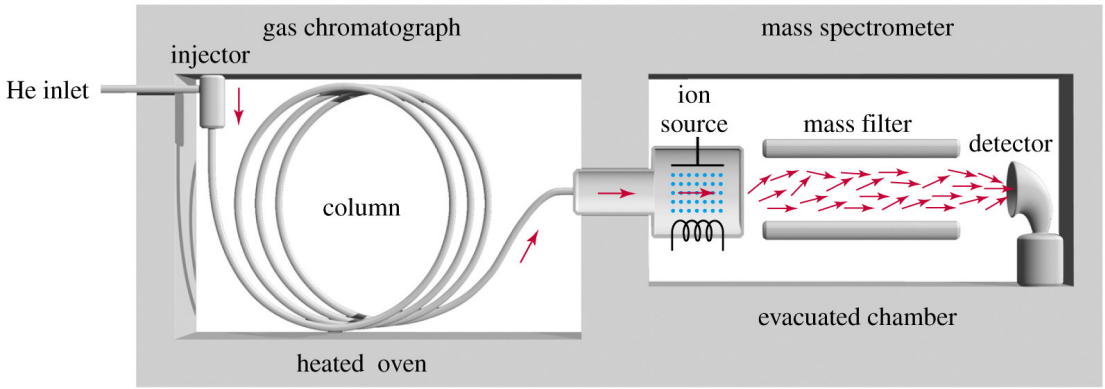


Figure 2.5: *Block diagram of a gas chromatogram-mass spectrometer. Image taken from: Organic Spectroscopy International, 2014*

The first study including lung cancer patients was in 1985, when Gordon et al identified a total of 22 VOCs showing a large difference between examined breath samples of 12 LC patients and 17 healthy volunteers. The authors, using three of these VOCs chosen according to their peak and occurrence in the subjects (acetone, methyl ethyl ketone and n-propanol), were able to accurately classify 93% of the samples [42]. In 1999 Phillips *et al.* collected breath samples from 108 patients with an abnormal chest radiograph and identified a combination

of 22 discriminating VOCs, demonstrating that for stage 1 LC the 22 VOCs had 100% sensitivity and 81.3% specificity [43]. Later on, the same authors identified primary LC with a sensitivity of 89.6% and a specificity of 82.9% using nine VOCs, noticing that patients with primary LC had breath test findings that were consistent with the accelerated catabolism of alkanes and monomethylated alkanes [44]. During the next years many other studies examined different combinations of VOCs [45, 46, 47, 48] or investigated the differences between patients with NSCLC and chronic obstructive pulmonary disease(COPD) [49, 50], LC and non-pathological smokers [51], LC and colon, breast and prostate cancers [25]. On the other side, some researches found contrasting results. Kischkel *et al.* in 2010 noticed that differences in exhalation profiles in cancer and noncancer patients did not persist if physiology and confounding variables such as smoking history, inspired substance concentrations, age and sex were taken into account [52]; more recently, in 2016, Schallschmidt *et al.* reached a similar conclusion when examining 24 VOCs that were suggested as potential cancer markers in previous studies [13].

Even though GC-MS has proved to be a useful analytical tool for identifying specific biomarkers in breath, its potential use as an early diagnostic tool is limited due to its complexity, long analysis time, the need for qualified operators and high cost. Moreover, several studies indicate that there is not a single molecule that can be correlated with lung cancer, but relative concentrations of several compounds are required for cancer detection [53]. For these reasons, researchers have focused their attention on new methods that are easily accessible, economically viable and potentially real-time, and raised the possibility of using electronic noses composed of non-selective sensors that respond not to the unambiguous concentration of a single compound as for GC-MS, but to a combination of all VOCs [5].

2.2.2 Ion Mobility Spectrometry (IMS)

A second approach to examine exhaled breath VOCs is Ion mobility spectrometry (IMS). The principle of IMS systems is breaking down analytes in the gas phase into ions using a 550 MBq 63Ni β-radiation ionising source (Ni) [54]. The ions travel down a chamber at a speed related to their size, mass and geometry. Then, they hit a Faraday plate at the end of the chamber. The collision of each ion with the plate generates an electrical signal; the various signals, when combined, produce an ion spectrum which is a fingerprint for the exhaled breath, as schematized in Figure 2.6.

Studies using IMS for VOCs examination are not so common as those emplying GC-

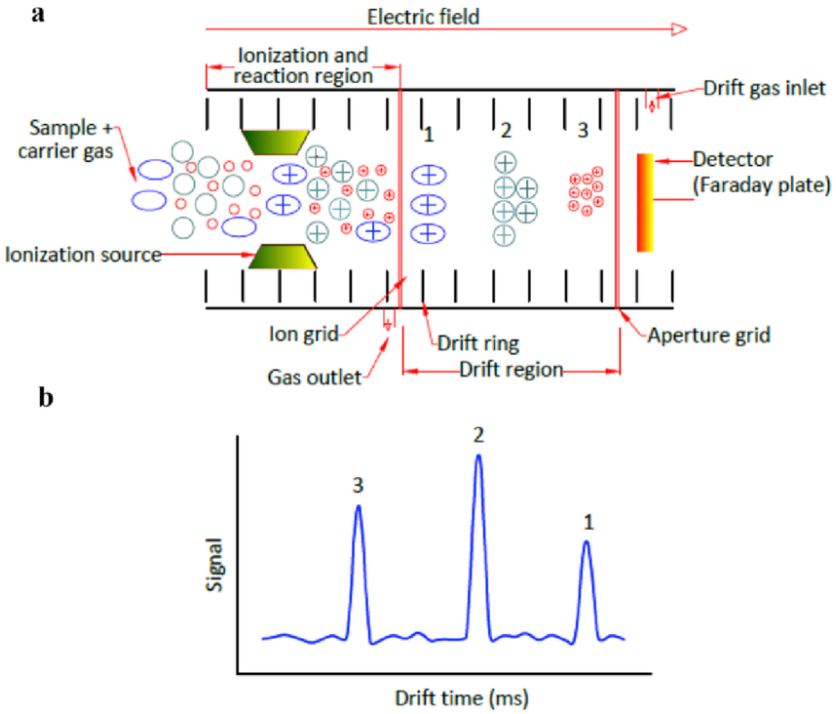


Figure 2.6: *Schematic representation of the (a) ion mobility spectrometry working principle and (b) resulting ion mobility spectrum. Image taken from [55].*

MS, in particular in the case of exhaled breath analysis. Westhoff *et al.* in 2009 used a combination of 23 peak regions within the IMS chromatogram to discriminate between 32 patients with lung cancer and 54 healthy subjects with 100% accuracy [56]. Later on, in 2014, Handa *et al.* used IMS to discriminate between the exhaled breath of 50 patients with lung cancer histologically proven by bronchoscopic biopsy samples and 39 healthy volunteers. They employed a decision tree algorithm to separate patients with lung cancer including adenocarcinoma, squamous cell carcinoma and small cell carcinoma. One hundred-fifteen separated VOC peaks were analyzed and a decision tree algorithm achieved a sensitivity of 76% and specificity of 100% [39].

2.2.3 Electronic Noses

Whenever biomarkers are not uniquely known, as in the case of VOCs for LC diagnosis, the analytical approach does not represent the ideal solution and many non-selective sensors are needed to obtain a fingerprint of the measured VOC mixture. In opposition to the instruments of Section 2.2.1 and 2.2.2, they are designed to respond to the mix of compounds in the sample rather than identifying individual compounds, producing as output a pattern representing the

mix of VOCs. The registered "breathprints" are treated with pattern-recognition techniques to be identified as pertinent to certain diseases [12]. Several technologies of this type, with the indisputable advantage of being small, portable and relatively inexpensive, have been developed to analyze exhaled breath samples. Because of their similarities to the working principle of the olfaction, these devices assumed in literature the name of Electronic Nose (or eNose) systems. In this section the most relevant implementations are exposed and the major findings are presented.

2.2.3.1 Quartz Microbalance

The quartz microbalance (QMB) is a sensor array of oscillating quartz crystals coated with different metalloporphyrins to which VOCs adsorb or desorb, changing the mass of the sensors (Δm) and consequently the fundamental oscillation frequency (Δf) of the electrical signal of the oscillator circuit at which each sensor is connected. Metalloporphyrins are versatile molecules that can host several interaction mechanisms, from weak and non-selective dispersion forces to more specific coordination. The phenomenon is described, at a first approximation, by a relation known as Sauerbrey law [57]:

$$\Delta f = -\frac{C_f f_0^2}{A}\Delta m \tag{2.1}$$

where A is the coated area, C_f the mass sensitivity constant and f_0 the fundamental frequency.

The first in employing this principle were Di Natale *et al.*, that developed an electronic nose (*LibraNose*, University of Rome Tor Vergata and Technobiochip) composed by eight QMB gas sensors coated with different metalloporphyrins. In a study of 2003 [58] they sampled the breath of 42 subjects with lung cancer, 18 healthy volunteers and 9 post-surgery LC patients in sealed bag with volume of about 4 liters through multiple and repeated breaths. Then, after maximum 5 min from the sampling time, sampled breaths were analyzed in the electronic nose sensor chamber. The typical signal of one of the sensors is reported in Figure 2.7 as an example.
By making use of partial least squares discriminant analysis (PLS-DA), they achieved 90.3% accuracy in the discrimination of LC patients from healthy people. Anyway the class of post-surgery patients was correctly individuated only in 44% of the cases, with the remaining samples classified as healthy references.

In 2010 the same group employed again the *LibraNose* to discriminate between lung cancer, diverse lung diseases (e.g. COPD) and reference controls. In addition, they also tested

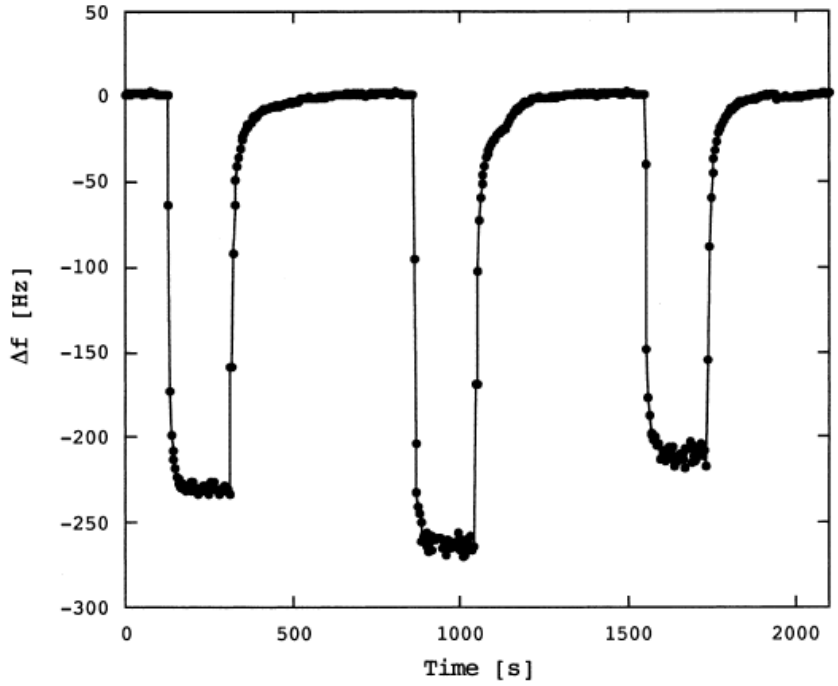
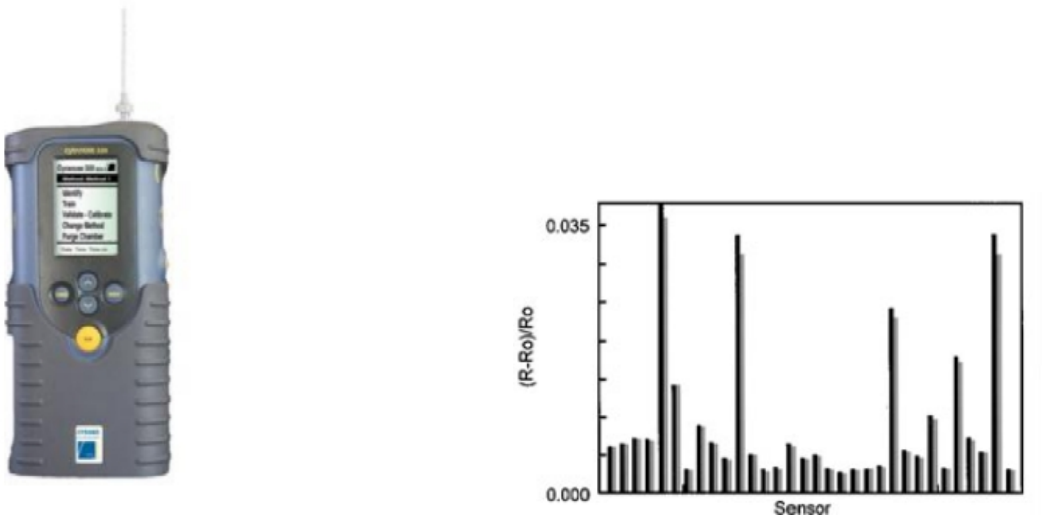


Figure 2.7: *Typical response of the Ru-Tpp sensor to 3 successive samples related to post-surgery, cancer affected, and reference breaths, respectively. Image taken from [58].*

the effect of some compounds whose concentration was demonstrated to be significatively different with respect to normal populations with GC-MS studies. 28 LC subjects and 36 controls were tested, and results showed a satisfactory identification rate of LC subjects but also a non-negligible sensitivity to breath modification induced by other affections. Anyway, they demonstrated that breath samples of control individuals drift towards the LC group when added with either single or mixtures of these alleged cancer-related compounds [59]. In this study a first form of separation between dead space and alveolar air was performed, its procedure will be explained in Section 2.3.

More recently, Gasparri *et al.* conducted a similar work using QMB sensors and applied the breath separation. The study involved 70 LC patients and 76 controls and used PLS-DA to build classification models to discriminate cancer versus negative (both in metabolic diseased and in non-metabolic diseased populations) and for cancer stage assessment. The results showed a sensitivity of 81% and a specificity of 91% in the classification of LC patients from healthy controls, demonstrated a good ability in the separation of the subgroup of stage I from the rest of the group corresponding to stages II/III/IV (sensitivity of 92%) and also highlighted that the difference between the breath composition of LC patients and a control



(a) Cyranose 320 **(b)** Typical Smellprint

Figure 2.8: *(a) Cyranose 320 electronic nose. Image taken from: www.sensigent.com; (b) example of a typical smellprint derived from the 32 sensor responses from a healthy control subject (black bars) and a patient with lung cancer (gray bars). Image taken from [26]*

population is not affected by comorbidity conditions (sensitivity and specificity values equal to 85 and 88% in the metabolic group and 76 and 94% in the non-metabolic group) [35].

All those results are encouraging, in particular the findings about stage assessment suggest that breath analysis could be useful particularly at an early stage, opening good perspectives for patients with lung cancer when they still are asymptomatic.

2.2.3.2 Conductive Polymer Gas Sensors

Electronic Noses employing conductive polymer gas sensors are made by arrays of chemiresistors whose resistance changes depending on the VOCs absorption or desorption onto the surface. The Cyranose 320, shown in Figure 2.8a, is the most popular example. It is an handheld analyzer, used in many fields, containing an array of 32 carbonblack polymer composite chemiresistors; each sensor in composed by an insulating polymer containing a mixture of conductive particles and two electrodes used to apply the voltage. The measured resistances are transmitted to a processor able to convert them into a response pattern, defined *'smellprint'*, similar to the one reported in Figure 2.8b, i.e. a bar graph of the responses of each of the sensors in the array [54]. Finally the processor identify the vapor by matching, through an advanced algorithm, the pattern of the unknown vapor to a known vapors library.

There are several applications of Cyranose 320 in exhaled breath analysis for LC diag-

nosis. Machado *et al.*, in 2005, used it to discriminate 14 LC patients and 20 controls based on their VOCs profile. They applied principal components analysis (PCA) to reduce the data from 32 individual responses to vectors called principal components and used Canonic Discriminant Analysis (CDA) to create a classification model of the type cancer versus noncancer, obtaining correct cross-validation results in 71.6% of cases. Then, applying a SVM classifier to a group of unknown samples from a different group taken 5 times for each subject, outcomes of the five samplings were concordant in 92% of cases. The model was finally validated on 14 patients, achieving an overall accuracy of 85%, sensitivity of 71.4% and specificity of 91.9% [26].

Dragonieri *et al.* in 2009 investigated the use of Cyranose 320 to distinguish 10 patients with NSCLC from 10 with COPD and 10 healthy subjects. Similarly to Machado *et al.*, they applied PCA and CDA to data achieving accuracies of 85% in LC-COPD discrimination and of 90% in healthy controls recognition [60]. These results, even if obtained on a small population, demonstrate how exhaled air VOCs differ between 2 different smoking-related pathologies.

More recently, in 2017, Tirzite *et al.* extended the analysis to other lung diseases (asthma, pulmonary embolism, pneumonia and benign lung tumors) and to a larger number of subjects (165 in the cancer group, 91 in the non-cancer group and 79 healthy volunteers). Using SVM, they correctly classified LC patients and healthy volunteers in 98.8% of cases, cancer and non-cancer group patients in 87.3% of cases and predicted patients in the cancer-COPD group in all 79 cases, with lower prognosis rate in other mixed subgroups diagnosis [61].

2.2.3.3 Gold Particle Nanosensor Array

This class of electronic nose is constituted by gold particle nanosensor arrays covered with mixture of compounds responsive to a variety of odorants. The working principle is similar to that of the other categories: the organic film component is the site for the sorption of VOCs, while the electrodes resistance changes its electrical conductivity with different behaviors depending on the VOCs adsorbed.

In 2009 Peng *et al.*, after having identified a number of VOCs as possible LC biomarkers using GC-MS and SPME, designed an array of chemiresistors based on 5nm gold nanoparticles as represented in Figure 2.9.

The response each sensor undergoes to when exposed to a breath sample is unique because of the different chemical composition of the materials and has a profile of the type represented

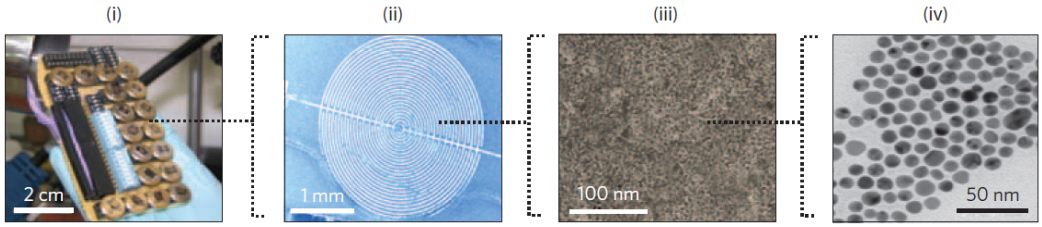


Figure 2.9: *A photograph of the array of chemiresistors (i), a scanning electron microscopy image for the chemiresistor (ii), a scanning electron microscopy image of a gold nanoparticles film located between two adjacent electrodes (iii), and a transmission electron micrograph of the monolayer-capped gold nanoparticles (iv). Image taken from [62]*

in Figure 2.10.

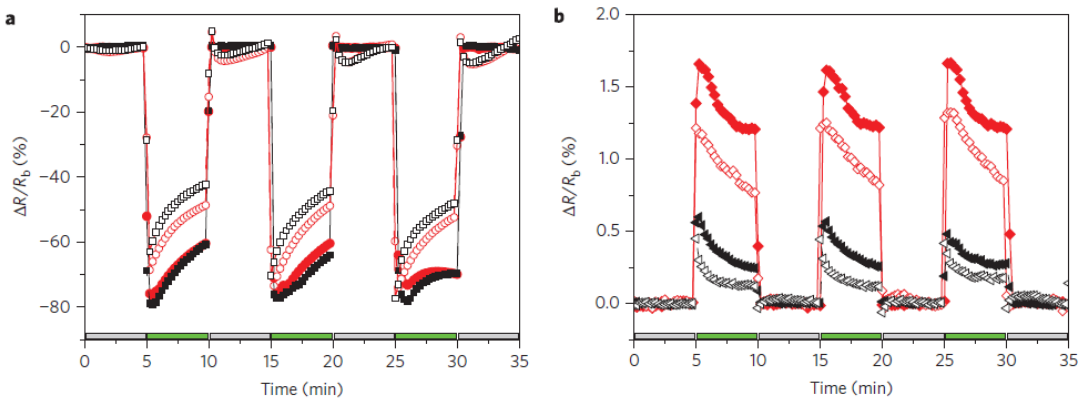


Figure 2.10: *a) Typical responses of the chemiresistors to real breath samples DR/Rb (where Rb is the baseline resistance in the absence of analyte and DR is the resistance change in presence of the analyte) upon exposure to healthy breath (filled symbols) and lung cancer breath (open symbols), as representative examples for sensors having positive responses. b) Typical responses of gold nanoparticles as representative examples for sensors having negative responses. Image taken from [62]*

PCA was used to analyze the nine-sensor array responses having clear discrimination, with no overlap, between LC and healthy breath clusters [62]. In a second study of 2010 they also demonstrated a good separation between patterns of healthy subjects and patients with lung, colon and breast cancers; only in the case of prostate cancer a minimal overlap was found with the cluster of healthy subjects [25].

Peled *et al.* in 2012 examined 72 patients with a chemical nanoarray of 18 cross-reactive sensors (of which 16 based on spherical gold nanoparticles) to evaluate the classification ability of benign and malignant Pulmonary Nodules (PNs). The eNose was able to distinguish the

19 benign versus the 53 malignant PNs with accuracy of 88% and AUC of 0.986 [63].

2.2.3.4 Metal Oxide Sensors

Electronic Nose technology using Metal Oxide Sensors exploits the conductivity changes of these aspecific sensors to detect VOCs particles. The Aeonose, which is the most popular device of this group, is composed by three different micro hotplate metal oxide sensors and a Tenax tube. Air particles cause redox reactions at the sensors surface creating a VOCs profile similarly to the technologies described in this section. The peculiarity of Aeonose is that it is portable and does not require any extra container for breath sampling, hence it allows real time analysis.

Figure 2.11: *The Aeonose device. Image taken from www.enose.nl*

R. van de Goor *et al.* in 2018 employed five versions of this eNose in a study enrolling 60 LC patients and 107 controls. They developed an ANN model and achieved sensitivity of 83%, specificity of 84% and an overall accuracy of 83%,with the model of the training set having an AUC of 0.84 [37].

Blatt *et al.* in 2007 used an array of six MOS sensors (developed by *SACMI s.c.*) whose heating element operates in a range of temperatures going from 200°C to 400°C; the response consists in a change in resistance as reported in Figure 2.12.

After data normalization, some features were extracted and PCA was employed to reduce dimensionality. Three classifiers were developed to analyze the exhaled breath of 101 volunteers (58 healthy and 43 suffering from different LC types) reaching maximum accuracy of 92.6%, sensitivity of 96.5% and specificity of 91.4% [38].

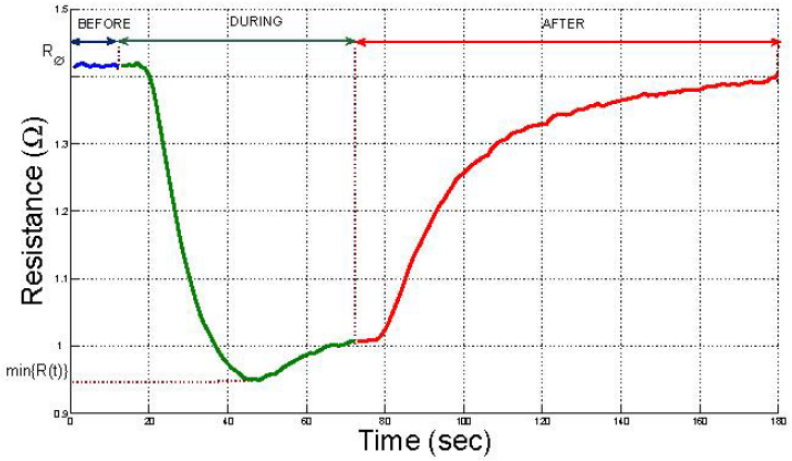


Figure 2.12: *Example of typical sensor response before, during and after the measuring phase. Image taken from [38]*

Devices of this type, even if very sensitive, have high power consumption due to the working temperature limiting their applicability as portable devices.

2.2.3.5 Colorimetric Sensor Array

The Colorimetric Sensor Array (CSA) is constituted by disposable cartridges of dots filled with chemically sensitive compounds such as metalloporphyrins. Each of these dots has different sensitivity to a subgroup of VOCs and the interaction with a gas causes them to change color, as depicted in Figure 2.13. The colors of the spots on the cartridge are thus scanned before and after exposure to the sample and their variation is measured.

Mazzone *et al.* used a CSA analyzer in three studies to compare the exhaled breath of LC subjects and control groups. In 2007 they used an array with 36 chemically sensitive spots generally responsive (i.e, not sensitive to one or two specific groups of volatiles) and enrolled 49 subjects with lung cancer and 94 controls, both healthy (n=21) and affected by other types of pathologies (n=73). Their random forest method had an error rate of 14.1%, sensitivity of 73.3% and a specificity of 72.4% for the diagnosis of lung cancer [64]. In 2012 instead, with a platform improved using selected chemically responsive dyes, they examined the breath of 92 LC patients and 137 controls and developed four logistic prediction models to compare the various subgroups. The major finding of this study was that clinical risk factors (including age, sex, smoking status and COPD) raise the model accuracy in discriminating between the two groups obtaining an Area Under the Curve (AUC) of the Receiver Operating

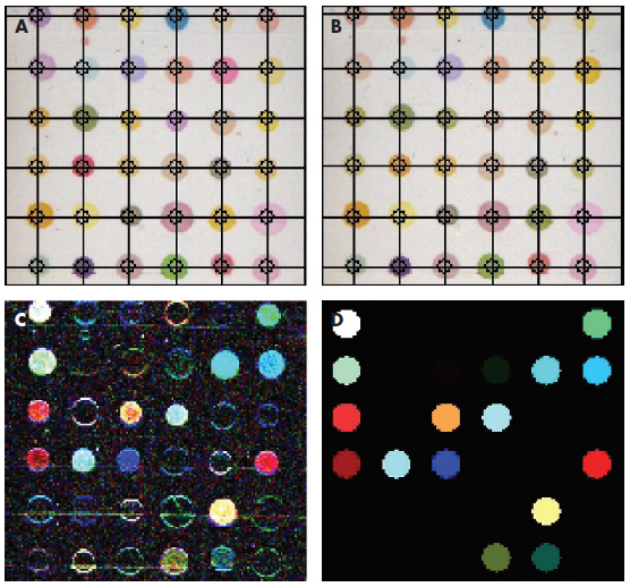


Figure 2.13: *(A) Colorimetric sensor array at baseline and (B) after the exposure to exhaled breath. (C) The difference in the colors from time A to B. (D) Cleaned-up version of panel C. Image taken from [64]*

Characteristic curve (ROC) equal to 0.811 for the global cancer versus non-cancer signature and equal to 0.825 and 0.849 including the individual histologies [65]. In their third research work of 2015 the CSA was furtherly improved with new pigments, more sensitive and also in greater quantity, a more sophisticated imaging system and end-tidal CO_2 control of alveolar breath sampling. 97 LC patients and 182 control subjects were analyzed achieving AUCs for models of cancer and subgroups versus control ranging from 0.794 to 0.861, again higher when the histology subgroups were compared with control subjects [66].

CSA, compared to other electronic noses, have the great advantage that they are not temperature and humidity dependent, thus they correctly operate with no need for water vapor removal from the breath samples.

2.2.3.6 Type-Different Sensor Array

In a reduced number of works, some research groups embedded arrays of cross response sensors, hence classes of sensors which are not gas specific, in their devices. All these sensors respond differently to the same gas sample and each of them has different behaviors when exposed to different chemicals.

Li *et al.* in 2017 selected 14 gas sensors of 4 different types (MOS, hot wire gas, catalytic combustion gas and electrochemical gas sensors) based on their reactivity to those which

are supposed to be the major VOCs in human breath of LC patients, according to previous studies. These sensors were embedded in a gas reaction chamber and controlled by a main control chip (STM32F10). The system is depicted in Figure 2.14.

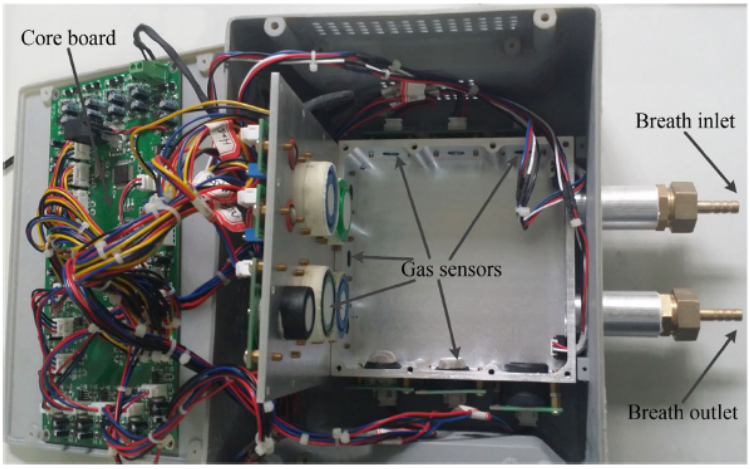


Figure 2.14: *Core board and gas reaction chamber of the type-different sensor array platform developed by Li* et al.*. Image taken from [33]*

A total of 52 breath samples were analyzed (from 24 LC patients, 5 patients with other respiratory diseases, 10 healthy smokers and 10 healthy non-smokers), and two classifiers were developed to distinguish the subgroups. An example of the sensors response is reported in Figure 2.15. The model that showed the best performance had sensitivity, specificity and accuracy respectively of 91.58%, 91.72% and 91.59% [33].

These findings, together with the possibility to integrate the advantages of the previously described sensing solutions, suggest that this kind of approach could potentially increase the diagnostic ability of eNose systems.

2.3 Exhaled Breath Separation: First Solutions and Findings

Up to now we described the major systems for exhaled breath analysis in the field of lung cancer early diagnosis, with a particular focus on electronic nose devices, and we discussed the most important findings of each of them. Anyway the purpose of this work is not the realization of a complete eNose platform, but the development of an automatic separator of the breath sample in its dead space and alveolar contributions. In this section we will thus see the first solutions proposed in research with this objective and briefly discuss their advantages and limitations.

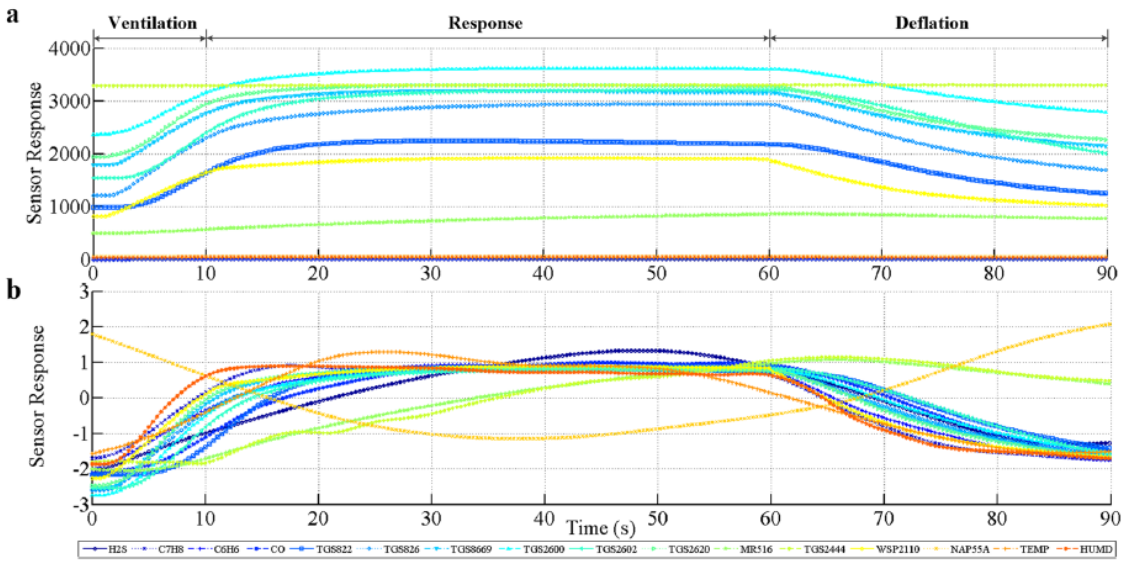


Figure 2.15: *Typical response curves of the sensor arrays (a) before preprocess and (b) after preprocess. Image taken from [33]*

The rationale behind the separation of exhaled breath in dead space and alveolar portions arises from the influence of dead space air on exhaled VOCs. In fact, concentrations of substances in blood relate to alveolar concentrations by their blood-gas partition coefficient or solubility, and compounds of interest for this kind of examination have extremely low concentrations falling in the range of 10-12 to 10^{-9} mol/l [31]. Because of this, exhaled air chemical analysis may be affected considerably by dilution and contamination with dead space gas. More in detail, the concentrations of molecules produced in the lower airways but not in the upper part (alveolar portion) may be diluted by the dead space air; also, some molecules are produced only in the upper airways (dead space) and have not metabolic origin, being impurities. Two main approaches have been introduced in literature to avoid this kind of problem:

- the partitioning of the first proportion of exhaled air via a T-valve (time or volume-controlled separation)
- the monitoring of CO_2 concentration in the exhaled breath and selection of the alveolar plateau period (CO_2 controlled separation)

CO_2-controlled methods provide more precise results, while the time or volume-controlled separation is usually simpler and more feasible [11], as described in the following paragraphs.

2.3.1 Time or volume-controlled separation

D'Amico *et al.* in 2010 designed a simple sampling procedure to remove the upper part of the respiratory tract (i.e., from mouth to lungs) prior to the analysis with the LibraNose descripted in Section 2.2.3.1. The employed sampler is composed by a mouthpiece connected to two sterile Tedlar bags (see 2.1.2 for further details); the smaller bag, intended to the dead space portion, is always kept open while the other, having larger volume in that dedicated to the alveolar contribution, has access regulated by a three-way valve. In particular, in this experiment the first bag had volume of 0.5L (about 4 times the average dead space volume of 0.125L) and the second bag 3L. Thus, during an exhalation, the breath fills the smaller bag until the increased resistance of the air opens the valve allowing the inflation of the second bag. Finally, only the second bag is kept to be analyzed. Authors observed that perturbation sources, such as food, tend to decay faster in the second bag compared to the first, making the analysis less sensitive to them [59].

Gasparri *et al.* in 2016 applied the same procedure, again in a study using a QMB sensor system (descripted in Section 2.2.3.1). Differently from the previous case, they stated that the upper airways contribution is not completely segregated to the first bag but the second bag contains mixed-exhaled breath, still with a prevalence of alveolar contribution. [35].

Blatt *et al.* instead decided to use a spirometer to evaluate each volunteer exhalation capacity and, at the end of the exhalation, diverted the flow into the bag. Then they took two measures from each bag using their MOS sensors array [38].

Bikov *et al.* in 2014 conducted a study to assess the influence of expiratory flow rate, breath hold and anatomic dead space on the detection of lung cancer with electronic nose. In particular, to investigate breath sampling, they performed two procedures: in one they discarded the first 500 mL of exhaled air using a small-resistance T-valve and collected the remaining air, in the other the dead space air was not discarded and mixed breath was collected. They showed that expiratory flow rate, breath hold and dead space influence significantly the "breathprints" in healthy individuals but not in LC patients.

2.3.2 CO_2 controlled separation

Schubert *et al.* in 2001 performed probably the first form of end-tidal CO_2 control of alveolar breath sampling. They employed a fast-responding infrared absorption mainstream CO_2 sensor (*930, Siemens-Elema, Solna, Sweden*), inserted between the Y piece of the respiratory circuit and the patient, to monitor the carbon dioxide level and control an electrically oper-

ated two-way valve. Adsorption traps containing 80 mg of activated charcoal were mounted onto the two outlets of the valve and connected to a roller pump; volatile substances were concentrated by adsorption onto the activated charcoal. A schematic of the system is reported in Figure 2.16. The control of the valve is regulated as follows: the valve is switched into "ex-

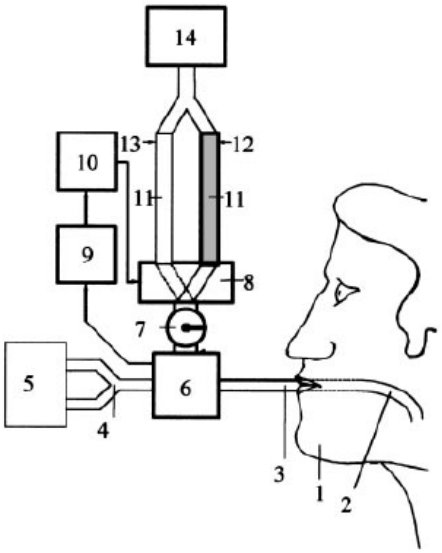


Figure 2.16: *Schematic drawing of the CO2-controlled sampling device designed by Schubert et al. 1, patient; 2, endotracheal tube; 3 and 4, respiratory tubing; 5, ventilator; 6, CO2 mainstream sensor; 7, stopcock; 8, electrical 2-way valve; 9, CO2 analyzer; 10, electronic processing unit; 11, trap; 12, alveolar sampling (expiratory) position; 13, mixed inspiratory and dead space sampling (inspiratory) position; 14, roller pump. Image taken from [31]*

piratory" position when the increasing ramp of the capnogram exceeds a static threshold, and into "inspiratory" position when the CO_2 concentration fell below 90% of the dynamic maximum. An example of the signal with the switching time is reported in Figure 2.17.
The authors declared that this kind of alveolar gas sampling was reliable and precise and, as a demonstration of the effectiveness of the separation, showed that median expired isoflurane and isoprene concentrations were respectively 1.75 and 2 times higher in the CO_2-controlled samples than in the mixed expired samples, while expired acetone and pentane concentrations were not different [31]. The system, anyway, has a big limitation in the static expiratory threshold, which needs to be manually lowered in case of patients showing a slow rise of expired CO_2 concentration (e.g., in obstructive lung disease) or having low expired CO_2 concentrations (e.g., during hyperventilation).

Mazzone *et al.* in 2015 tried to solve this issue using a 75% fall in the upward slope of

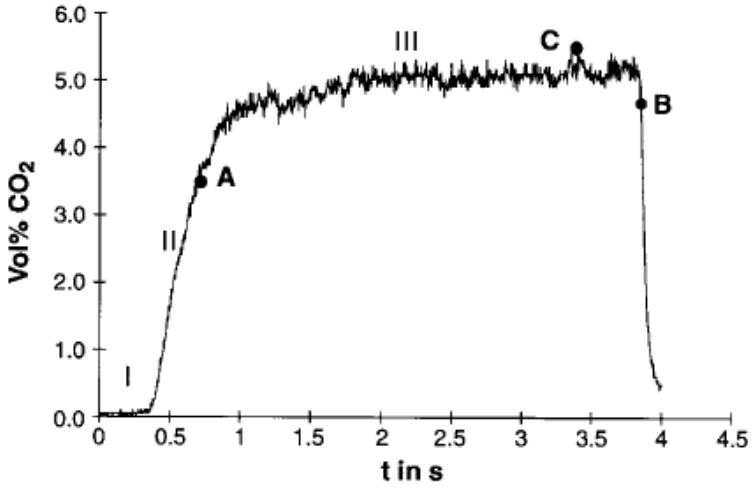


Figure 2.17: *Capnogram (CO_2 volume vs time) and switching points of the valve. A) start of alveolar sampling at a static threshold; B) end of alveolar sampling at 90% of the preceding maximum CO2 concentration; C) maximum CO2 concentration during alveolar phase; I) CO2 free inspiratory phase; II) mixing phase; III) alveolar phase. Image taken from [31]*

the end-tidal CO_2 curve to trigger the opening of the valve; once the valve has been switched, the alveolar portion of the breath is diverted on their CSA (see Section 2.2.3.5). Anyway they reported that the normalization of sensor changes to end tidal CO_2 values did not influence the accuracy of the model [66].

P. Salvo *et al.* in 2015 proposed a breath sampler for both single and multiple breaths. Their proposal to overcome the issue of inter-subject variability was to use an adaptive threshold to switch the valve when necessary. The procedure starts with a training phase in which the subject breathes for 30s and the initial value of maximum CO_2 pressure (PCO_{2max}) is extracted; the threshold is then computed by subtracting a constant bias B (equal to 0.4kPa) from this value. While the device is being used, an adaptive filter (more specifically an exponential smoothing filter) updates breath by breath the PCO_{2max} as follows:

$$pres_s = \alpha * mag_{s-1} + (1 - \alpha) * pres_{s-1} \tag{2.2}$$

where *pres* is the adapted value of PCO_{2max}, α is the smoothing factor ($0< \alpha <1$), s is the breath number and *mag* is the average of the local maxima of PCO_2 at breath *s-1*. The valve is opened when the pressure is higher than the threshold and the derivative is positive, it is closed in the opposite situation. An example of the adaptive threshold is reported in Figure 2.18. To evaluate the breath sampler, the CO_2 partial pressure calculated during the

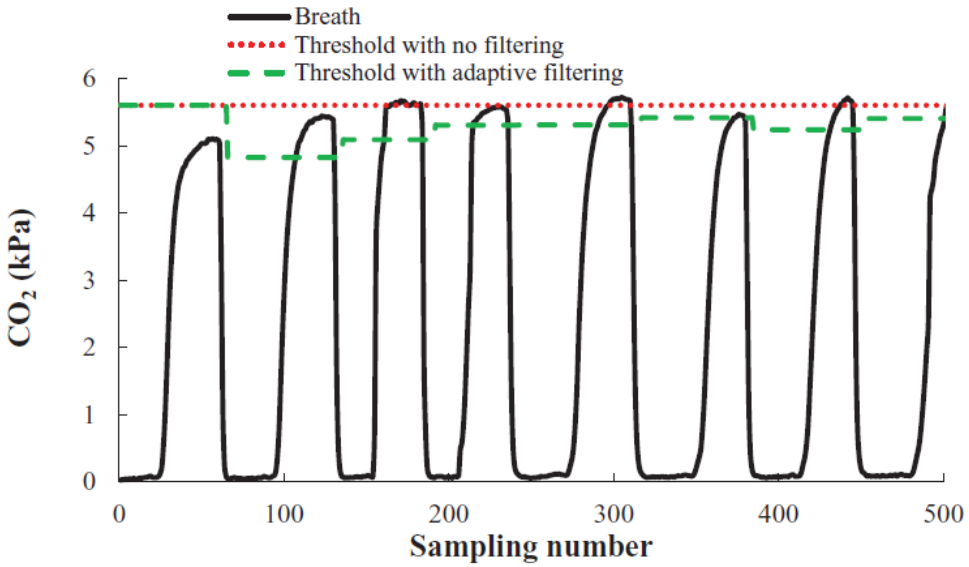


Figure 2.18: *Comparison of thresholds calculated with and without the exponential smoothing filter in the study of P. Salvo et al. Image taken from [24]*

sampling was compared with the one measured offline within the bag, showing negligible deviations between the two values and validating the possibility of an efficient sampling of selective fractions of exhaled air [24].

Miekisch *et al.* in 2008 investigated instead the impact of sampling procedures on the results of breath analysis comparing CO_2-controlled, time-controlled and mixed expiratory breath sampling. Their device for CO_2-controlled sampling is based on single-use plastic T-pieces and a fast-responding infrared absorption mainstream CO_2 monitor (*Capnogard, Novametrix, USA*). The real-time capnogram was used as visual control to manually close the far end of the sampling device during the alveolar phase and fill the bag with alveolar air. The measure used to investigate the difference between the different sampling methods is the ratio of PCO_2 in the different samples to end tidal CO_2 in ten healthy volunteers. The most important result is that significant differences were found between partial pressure ratios in alveolar and mixed expiratory samples and between mixed expired and samples taken 1 s after the start of exhalation [23].

All the observations reported in this section encourage research in this field to go on with CO_2-controlled alveolar sampling to improve the results already obtained in analysis of mixed expiratory samples. Furthermore, the need for standardization is a valid reason to focus on CO_2-plateau controlled separation, rather than on time or volume-controlled separation.

Chapter 3

Materials and Methods

In this chapter we describe the instruments employed to realize the device and the related software. For the clarity of the reader, the chapter starts with a characterization of the single parts of the hardware, continues with a brief explanation of the firmware implemented to control these components and ends with a description of the software that allows the user to interface with the prototype. The three streams have been conduced in parallel and are fully interconnected between them. An overview of the system is represented in Figure 3.1.

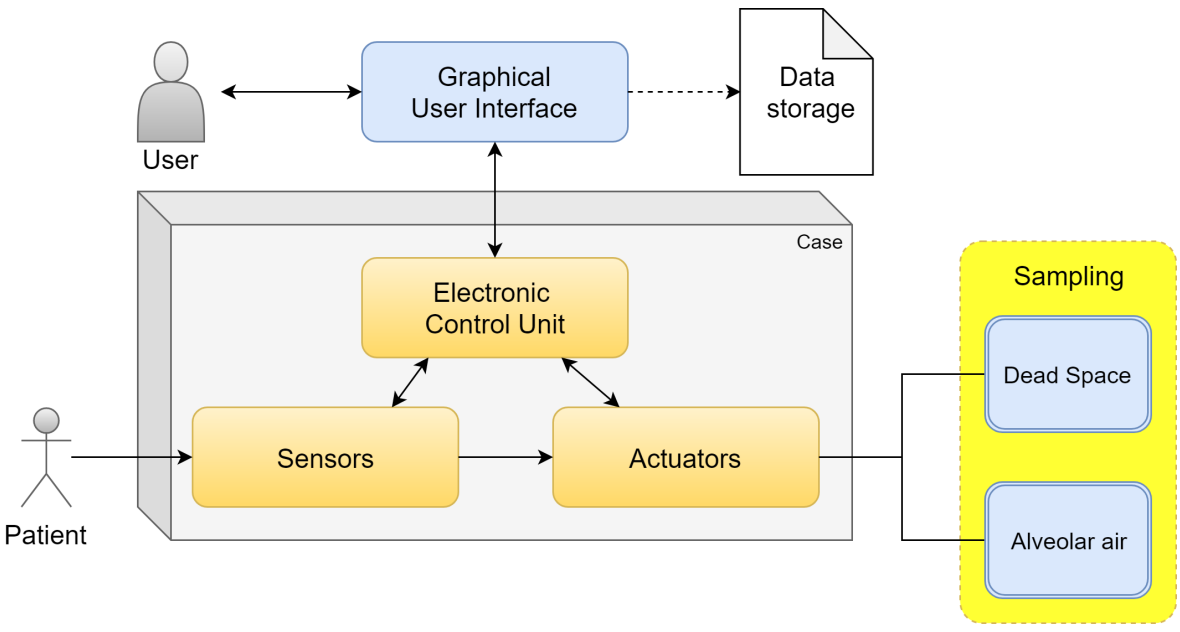


Figure 3.1: *Schematic overview of the system, showing the relation between hardware, firmware and software.*

3.1 Hardware

The hardware is mainly divided in electronic and hydraulic components. In addition, an electronic board has been designed and printed to integrate the microcontroller, the sensing and the actuating parts in a single circuit. A small plastic case was 3D printed to contain all pieces inside it and reduce the overall dimension of the device. A schematic overview of the system is represented in Figure 3.1. In the following sections everything will be described in detail and all choices will be motivated.

3.1.1 Electronic components

The electronic components employed in the project are mainly a microcontroller, the sensing part and a module for Bluetooth communication, all of them with the necessary conditioning circuitry. It's important to underline that the gas sensors, even if initially used for preliminary tests, have not been included in the final device for breath separation but only in the prototype of Electronic Nose under current development as part of the collaboration between Polimi and IEO and used for the final analysis; anyway their description is necessary to better understand the results reported in Chapter 4. The overall schematic is reported in Figure 3.2.

3.1.1.1 Microcontroller

The platform employed in this project is PSoC 5LP CY8CKIT-059 (where PSoC stands for Programmable System on Chip). It is one of the most recent low cost prototyping platforms produced by Cypress and it mounts on a single chip a 32 bit microcontroller based on an ARM Cortex-M3 architecture able to operate up to 80 MHz. It embeds a 24 channels DMA controller and several analogic and digital components with programmable logic, thus resulting very functional and versatile for embedded systems design. It is provided of a micro-USB port with Full Speed USB 2.0 connectivity and is made of two divisible parts, allowing the user to separate the platform with the USB connector (the programmer) from the remaining part of the board once the firmware is fully programmed. Thanks to its characteristics, it results ideal for the prototyping of the device in this project. It also features a very versatile IDE, as we will see later in Section 3.2.

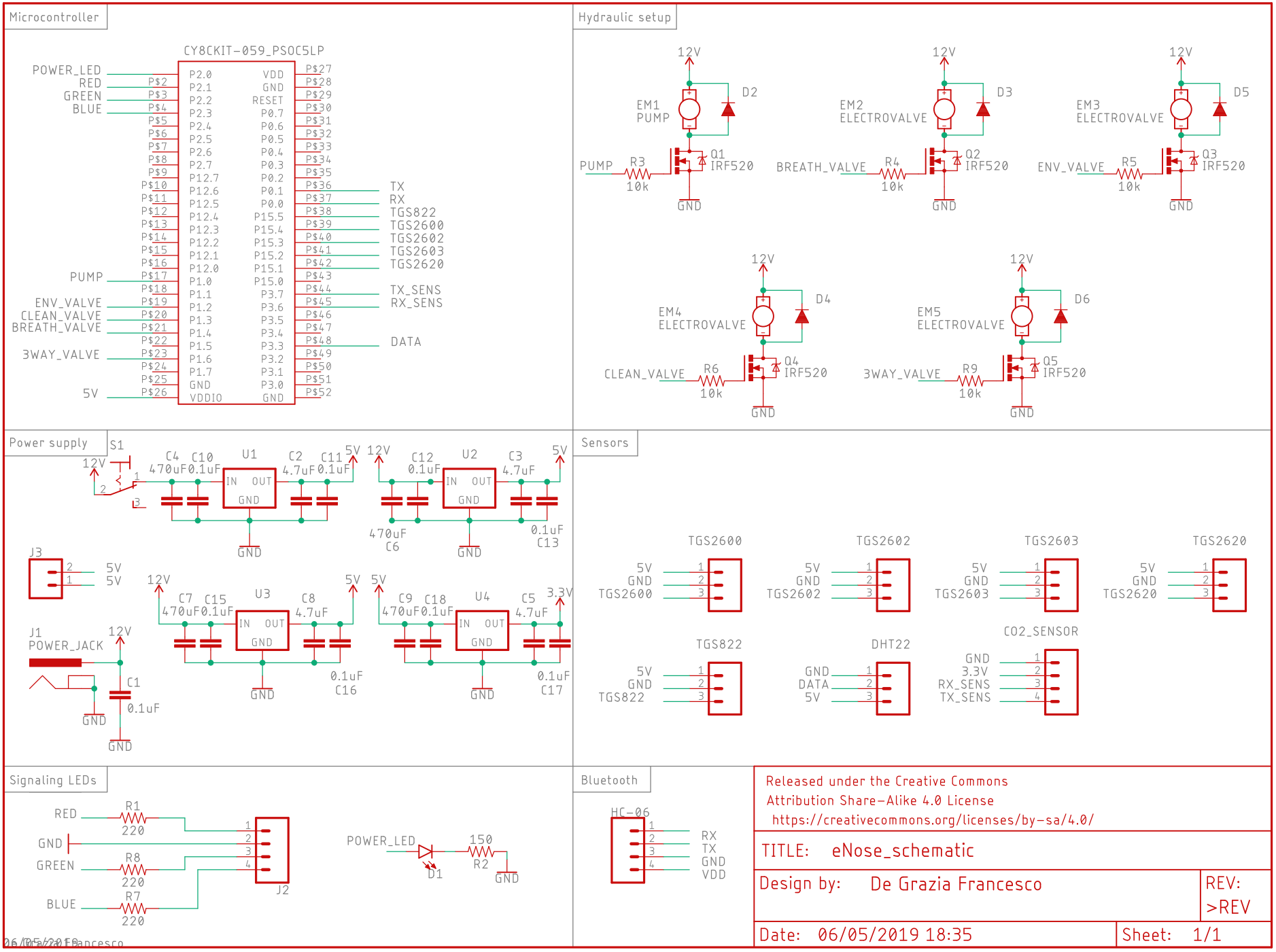


Figure 3.2: *Schematic of the whole circuit, designed in Eagle.*

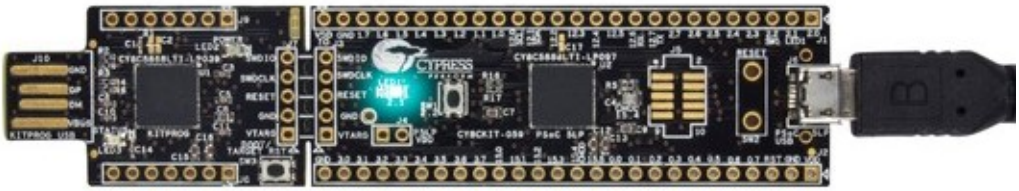


Figure 3.3: *Microcontroller PSoC 5LP. Image taken from www.cypress.com*

3.1.1.2 CO_2 sensor

As stated in Section 1.2, the core of the system is the sensor to be used for the CO_2 monitoring. For this purpose, it has to:

- be fast responding (or, at least, have a response time suitable for breath to breath analysis);
- measure the CO_2 concentration in the correct range and with a sufficient resolution;
- be cost effective, such that it could be a valid alternative to the expensive CO_2 monitors used in clinic;
- have low power consumption, being suitable for an embedded application like this.

An extensive research has been conducted to find a sensor with such characteristics. The most commonly available CO_2 sensors are based either on electrochemical or on Non-Dispersive Infrared (NDIR) sensing principle (the latter being an optical absorption method). The selection between the two technologies strongly depends on the application; a brief comparison of the principal parameter of interest for this project is reported in Table 3.1.

Table 3.1: *Brief comparison of Electrochemical and NDIR CO2 sensors.*

	CO2 sensor	
Parameter	**NDIR**	**Electrochemical**
Measurement Range	50ppm to 100%	400ppm to 90%
Response Time	<5 s typical	<2 min typical

In our case the choice falls on NDIR technology, mainly because of the response time,

and the chosen sensor is SprintIR-W20, produced by *Gas Sensing Solutions* and represented in Figure 3.4.

Figure 3.4: *Photo of the SprintIR-W20 CO_2 sensor with flow through adaptor. Image taken from www.gassensing.co.uk*

The reason why NDIR sensors are faster than electrochemical ones is due to their working principle: they have no heated filaments, hence there are no chemical reactions to detect a gas or compound, but instead measurements are taken based on the physical absorption of IR, an almost instantaneous process. More specifically, the Beer-Lambert law relates the level of IR radiation transmitted through an absorbing medium such as gas (CO_2 in this case), as shown in Equation 3.1.1.2:

$$I(c, \lambda) = I_0(\lambda)e^{-k_g cl} \tag{3.1}$$

thus it's possible to compute the reduction in intensity of the IR radiation (transmitted by an emitter) from an initial value I_0 (measured in absence of the target gas) to the intensity $I(c,\lambda)$ received by an IR detector due to the concentration c of a particular gas having wavelength λ. The absorption index of the target gas (e.g. CO_2) at a given wavelength is denoted by k_g and the IR optical path length given by l. The principle of operation is resumed in the stages shown in Figure 3.5.

The emitters included in IR systems are usually broadband and emit radiation across a spectrum of 2.5 to 12.5 µm. Sensors can be made specific to a particular gas or compound by detecting only IR absorption over a small wavelength range thanks to a bandpass filter covering the detector. The absorption spectrum shown in Figure 3.6 demonstrates how CO_2 absorbs IR at approximately 4.26 µm.

SprintIR has the following features:

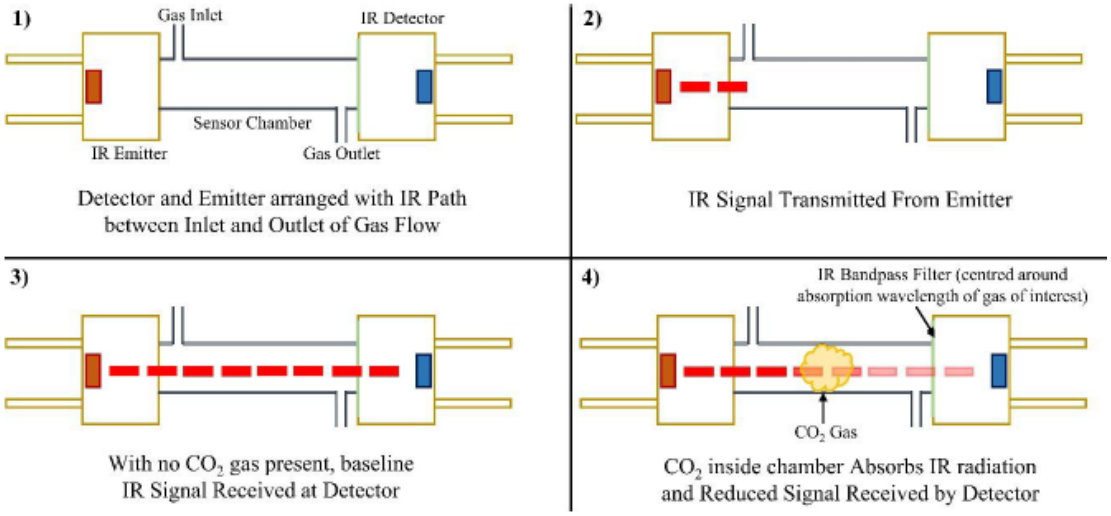


Figure 3.5: *Basic working principle of NDIR sensing*

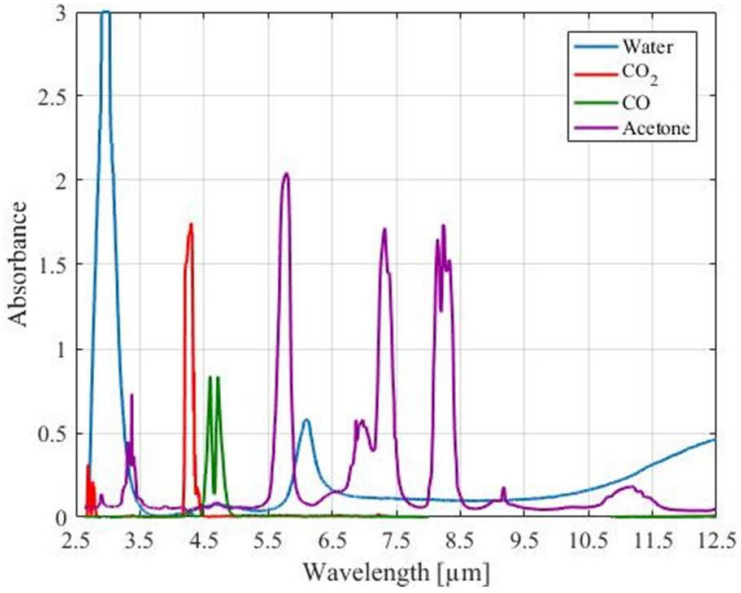


Figure 3.6: *Absorbance of IR radiation by water, CO_2, CO and acetone.*

- high speed sensing, up to 20 measurements per second (20 Hz)
- measures up to 20% CO_2 concentration
- relatively low price (160 €, still lower than professional CO_2 monitors)
- low power/energy consumption (35 mW on average)

hence it satisfies the requirements of this application. It has digital (UART) 8 bit output with RS232 interface and standard setup baud rate equal to 9600 (factory default outputs

are average and raw CO_2 values), being easily interfaced with a microcontroller as PSoC 5LP. The recommended supply voltage is 3.3 V, and only GND, 3.3 V, Rx and Tx are required for bidirectional serial connection (all other pins should be left unconnected), as represented in Figure 3.7.

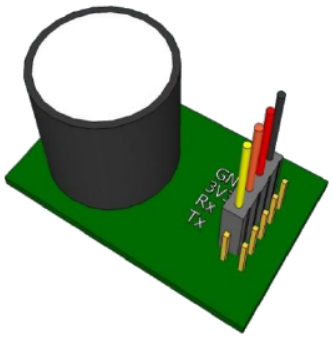


Figure 3.7: *Wiring schematic of SprintIR W20*

The initial calibration was performed by leaving the sensor for 5 minutes in fresh air, as described in the datasheet, since no better calibration tools are available in our laboratory (e.g. nitrogen or other calibration gases). The calibration routine aims at finding the lowest CO_2 value experienced by the sensor during this short period, and adjusts its zero point so that it would have read 400 ppm when it reads the lowest CO_2 value (400 ppm is assumed to be the global background level in fresh air). Over time, the zero point needs to be calibrated to maintain the long-term stability of the sensor.

3.1.1.3 Bluetooth Module

HC-06 is a very simple, low cost module for Bluetooth communication, having reduced dimensions ($27mm \times 13mm \times 2mm$) and low power consumption (average current of about 8 mA); nevertheless it ensures a good wireless communication. It allows to transform a UART/USART port, most commonly known as serial port, in a Bluetooth port; it mounts a 2.4 GHz antenna, has operative voltage of 3.1 V to 4.2 V and implements the Full Speed USB 1.1 protocol. Its wiring is based on 4 pins, two for the power supply (VCC, GND) and two for the serial communication (TX, RX), as represented in Figure 3.8.

Its usage in this project is necessary to make the device independent, more versatile and easy to use.

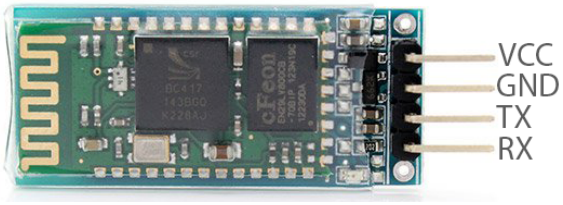


Figure 3.8: *Bluetooth module HC-06 with relative pinout*

3.1.1.4 Gas Sensors

As stated at the beginning of this section, the electrochemical gas sensors we are going to describe in this paragraph are not part of the device developed for this thesis project, but they belong to the prototype of eNose in development for the PhD project founded by the IEO. This device has been used for the breath analysis described in Chapter 4. Even if it should be a type-different sensor array (see Section 2.2.3.6 for details) similar to the one of Li *et al.* [33], only MOS sensors have been included up to now. Each one of them, all manufactured by FIGARO and powered at 5 V, is sensible to different gases at low concentrations as described in Table 3.2.

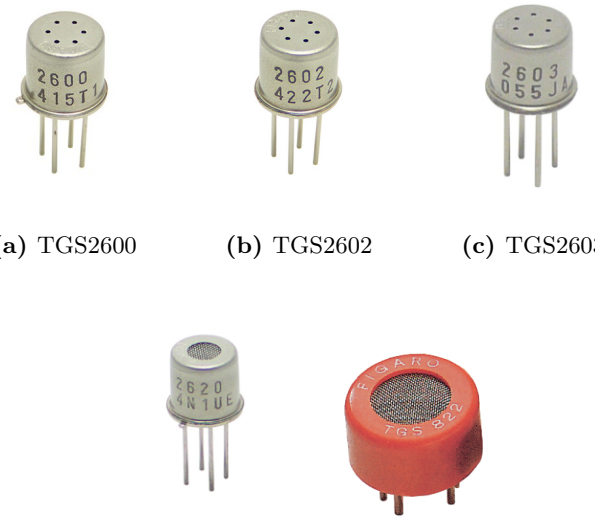
Table 3.2: *Key characteristics of the gas sensors used in the electronic nose prototype.*

Model	Range (ppm)	Detectable gases
TGS2600	1-30	Hydrogen, ethanol, butane, etc.
TGS2602	1-30	VOCs (Toluene, hydrogen sulfide, ethanol, etc.)
TGS2603	1-10	Trimethylamine, methyl mercaptan, etc.
TGS2620	50-5000	Ethanol, hydrogen, butane, etc.
TGS822	50-5000	Acetone, ethanol, benzene, etc.

The sensors are represented in Figure 3.9.

Even if usually used in applications such as air cleaners, ventilation control and air quality monitors, they have been chosen for this application because of the high sensitivity to low concentrations of gaseous air contaminants, low price (about 17€), small size, long life and low power consumption.

Sensors of the type TGS26xx (i.e., TGS2600, TGS2602, TGS2603, TGS2620) dispose of a sensing element comprised of a MOS layer formed on an alumina substrate of a sensing chip together with an integrated heater. They require two voltage inputs, heater voltage (V_H) and



(a) TGS2600 **(b)** TGS2602 **(c)** TGS2603

(d) TGS2620 **(e)** TGS822

Figure 3.9: *Pictures of the electrochemical MOS sensors composing the array.*

circuit voltage (V_C), and features 4 pins, respectively:

1 Heater

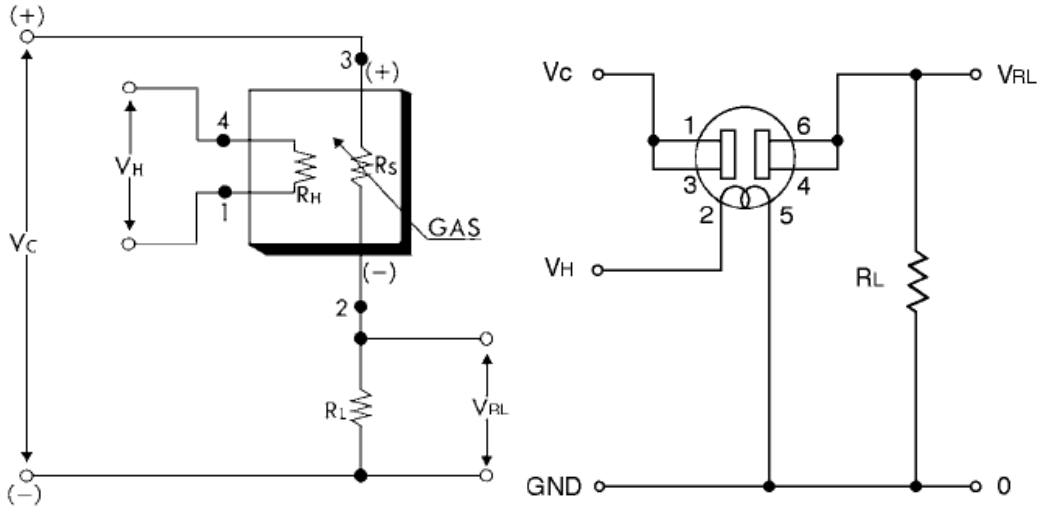
2 Sensor Electrode (-)

3 Sensor Electrode (+)

4 Heater

The integrated heater, to which the voltage V_H is applied, is required to maintain the sensing element at a specific temperature, optimal for sensing. The circuit voltage, instead, is applied to measure the output voltage (V_{OUT}) across the load resistor (R_L), which is connected in series with the sensor. The recorded signal derives from a conductivity change depending on the gas concentration in the air. The circuit schematic is reported in Figure 3.10a.
In this case, a common power supply line delivering 5 V is used for both V_C and V_H, a load resistor (R_L) value of 10 kΩ has been chosen to keep power consumption of the semiconductor below a limit of 15 mW, while R_H approximate value at room temperature is 83 Ω and R_s measures 10-90 kΩ in air (these values change sensor by sensor).

Sensor TGS822 is slightly different. First of all, it is composed by a sensing element in SnO_2, sintered to form a thick film on the surface of an alumina ceramic tube; also, it has 6 pins, whose standard conditions are:



(a) TGS26xx **(b)** TGS822

Figure 3.10: *Schematics and pinout of the gas sensors.*

- heater voltage (V_H) of 5 V;
- circuit voltage (V_C) of maximum 24 V, regulated to have power consumption P_S not higher than 15 mW (in our circuit is equal to 5 V);
- load resistor (R_L) whose value can be chosen to be higher than 0.45 kΩ (in our case 10 kΩ).

The schematic is reported in Figure 3.10b.

The sensitivity characteristics of each sensor in completely different, and the one of TGS2600 is shown as an example in Figure 3.11; it's possible to notice how the resistance ratio R_S/R_0 varies according to the gas concentration and depending on the gas type, where R_S is the sensor resistance in gaseous environment and R_0 the one in fresh air. The temperature and humidity dependency characteristics can be found in the datasheet, too.

3.1.1.5 Additional Components

Other electronic components are included in the current prototype. The power supply has been realized with the following components, represented in Figure 3.12:

a) a AC-DC wall adapter, supplying as output voltage 12V and currents up to 3A;

b) a small power jack ($2.1mm \times 2.5mm$) placed on the circuit board (Figure 3.13b);

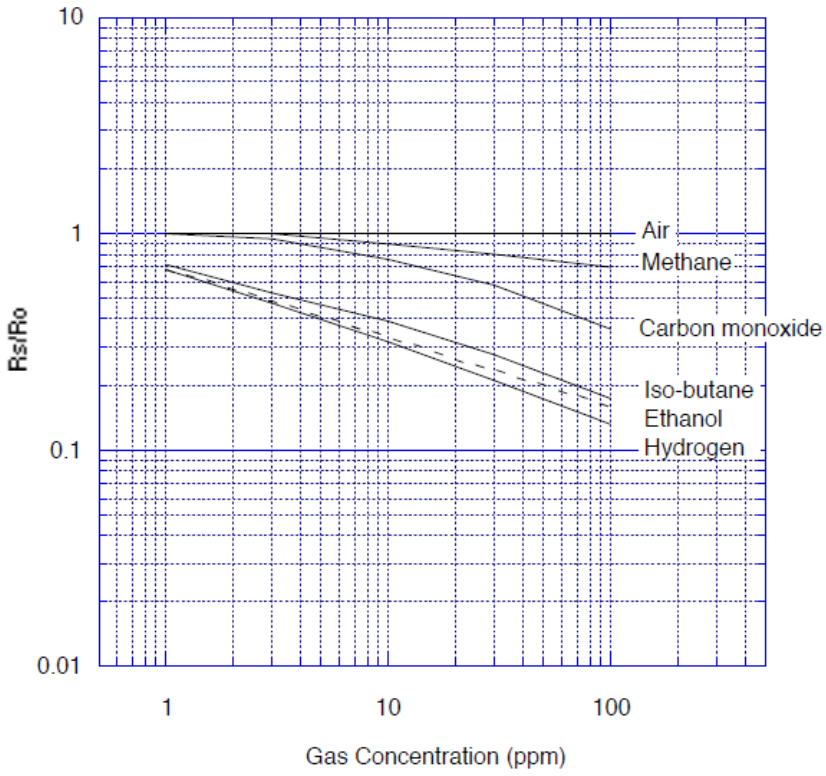


Figure 3.11: *Sensitivity characteristics of sensor TGS2600*

c) voltage regulators with output voltage of 5V or 3.3V, depending on the sensor to interface;

d) couples of electrolytic and ceramic capacitors to keep the voltage lines stable;

e) a single pole, double throw toggle switch.

Two LEDs, shown in Figure 3.14, are used to visualize the status of the device:

1) a status indicator (a simple green LED) to report when the device is switched on;

2) a phase indicator (an RGB led) to show the phase of the procedure; in particular, blue color indicates the streaming phase, while yellow the cleaning phase.

Finally, some JST connectors have been soldered on the board to provide the sensors' connections.

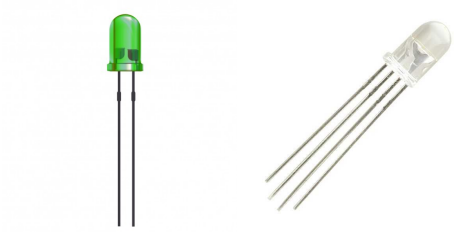
3.1.2 Hydraulic components

Being this a respiratory device, there are necessarily some hydraulic components in it. As already stated, a core part consists in a three-way valve used to switch between the dead

(a) 12V wall adapter **(b)** Power jack **(c)** Voltage regulator

(d) Electrolytic and ceramic capacitors **(e)** Switch

Figure 3.12: *Additional components used for the power supply of the device.*



(a) Status LED **(b)** Phase LED

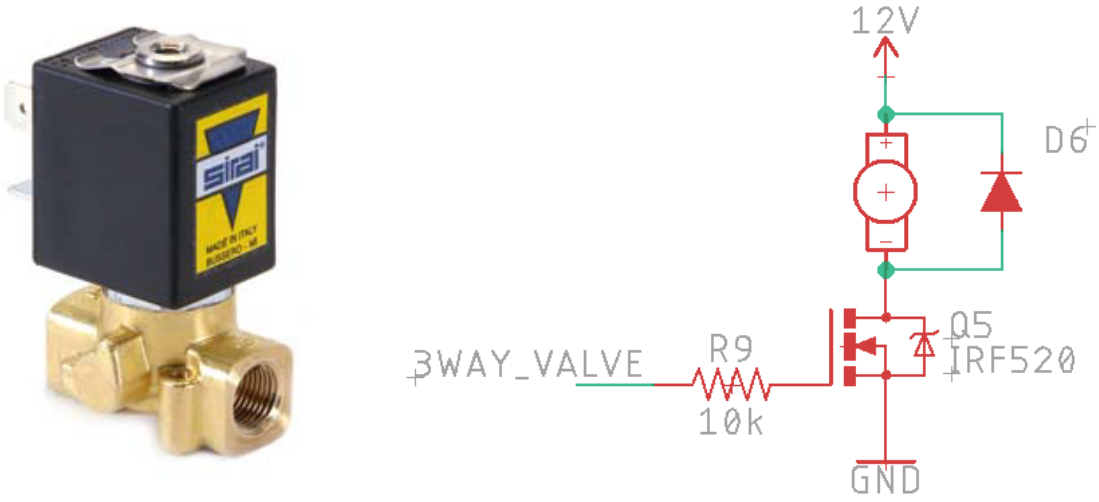
Figure 3.13: *Device LEDs. They are externally mounted on the device case and used as indicators.*

space and the alveolar sample bag during the acquisition. In addition, a mouthpiece and some tubes for breath directing are also included. In next paragraphs they will be described more in detail.

3.1.2.1 Three-Way Valve

The valve chosen for this project is a direct operated 3 port solenoid valve for water and air produced by *Sirai (Italy)*, model L372V03C. It is shown in Figure 3.14a.

It has been chosen because of its small dimensions (internal diameter of 1.6 mm, overall



(a) Valve external appearance

(b) Conditioning circuit schematic

Figure 3.14: *Three-way valve used in this project and its conditioning circuit schematic.*

external dimensions $53mm \times 33mm \times 25.5mm$) and its fast opening time (about 10 ms). It embeds a 12 V electromagnet to control its opening and closure, and is normally closed when de-energized. As many other inductive components, the electrovalve is conditioned as reported in the schematic of Figure 3.14b:

- a flyback diode is connected across the electromagnet itself to eliminate possible sudden voltage spikes seen when the supply current is suddenly reduced or interrupted;
- a power MOSFET (IRF520N) is used as a fast switch to open/close the valve with a signal from the microcontroller;
- a simple 10 kΩ resistor limits the current flowing to the MOSFET.

3.1.2.2 Other components

Other hydraulic components used in this project are the Tee-Mouthpiece Assembly (model QT00854-P, produced by *QuinTron*) and the tubes in PA 12 PHL (Polyamide 12) used to direct the breathing gas from the mouthpiece to the CO_2 sensor, the three-way valve and finally to the two Tedlar sampling bags. The Tee-Mouthpiece Assembly includes a plastic T-piece, a one-way diaphragm valve and a mouthpiece, as represented in Figure 3.15.

Furthermore, it must be named the employment of a pump and a few two-way valves (same type of the three-way valve of Section 3.1.2.1) in the electronic nose prototype used

Figure 3.15: *Composition of the Tee-Mouthpiece Assembly: at left, the mouthpiece; on the top, the T-piece; on the bottom, the diaphragm valve.*

to obtain the results in next section. Their schematics are reported in Figure 3.2, since there are connectors also on the breath sampler board that allow to connect pumps or other valves in future versions of the prototype.

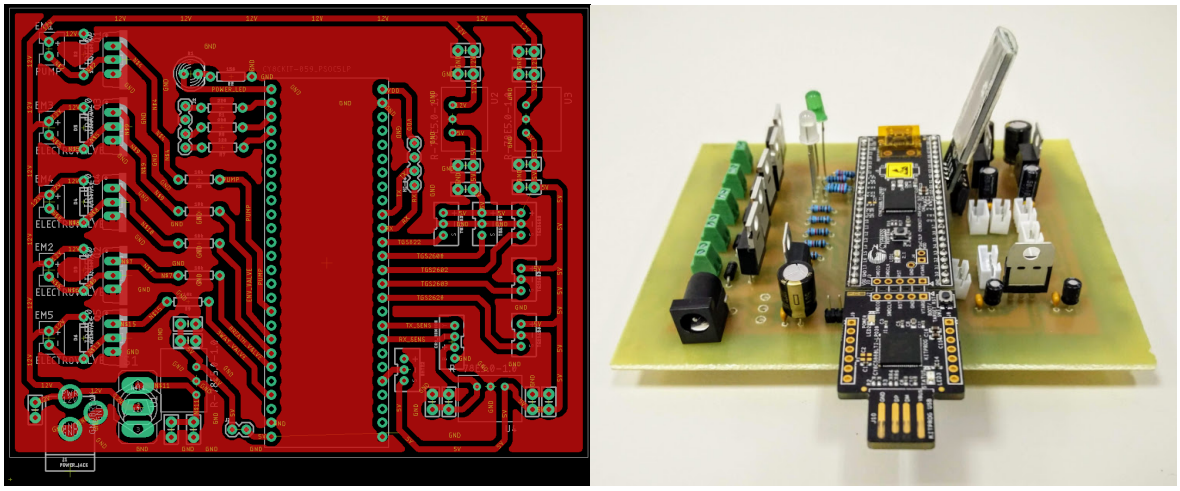
3.1.3 Circuit Board

All the components cited up to now are integrated in a single circuit and controlled by the PSoC microcontroller. The prototype underwent different stages of evolution starting from a breadboard version up to the final one, in which there is a Printed Circuit Board (PCB) with all conditioning circuitry soldered on it and all sensors and actuators that can be connected through wires and JST connectors. The PCB was designed in EAGLE (*Autodesk*) starting from the schematic of Figure 3.2, and is reported in Figure 3.16a.
It was printed with the toner transfer method and realized manually. The full production process involved the following steps:

1) the circuit was designed with the software EAGLE (Figure 3.16a);

2) it was printed in mirror format on photographic paper with a laser printer, as shown in Figure 3.17a (laser toner contains particles of plastic, which are "welded" together onto the paper);

3) the image (printed side down) was then placed onto a copper clad board and ironed onto the surface using a hot iron for about 10 minutes. When heated sufficiently, the toner became sticky and adhered to the board as in Figure 3.17b;



(a) Circuit board design **(b)** Printed Circuit Board

Figure 3.16: *PCB design (at left) and its physical realization (at right) with all components and connectors already soldered.*

4) once the paper had released all toner traces onto the board, it has been removed (Figure 3.17c);

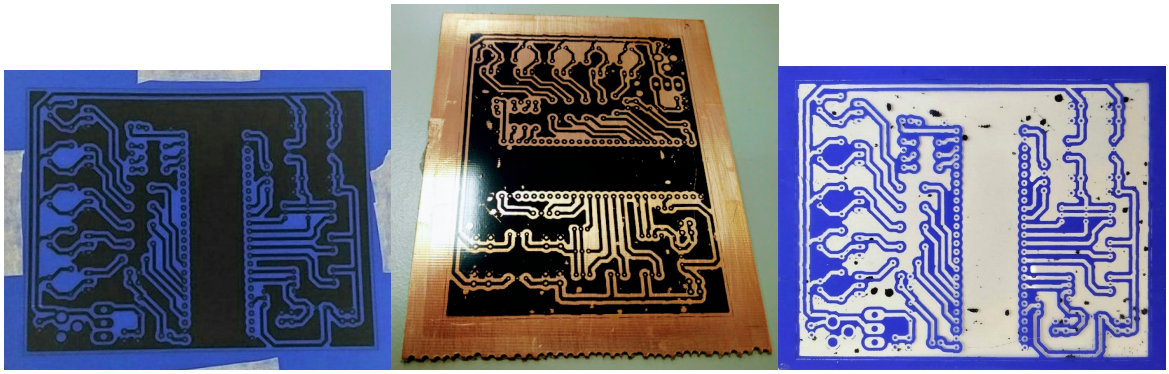
5) since it's very difficult to transfer all toner on the board, there were some gaps or holes in the transfer. To recover from this situation, the gaps were filled with a pen that uses an etch resistant ink;

6) the board was then immersed for a few minutes into an acid bath composed by one part of muriatic acid and two of hydrogen peroxide (Figure 3.17d). This solution etches the copper, with the toner trace working as a protective layer to preserve the printed circuit;

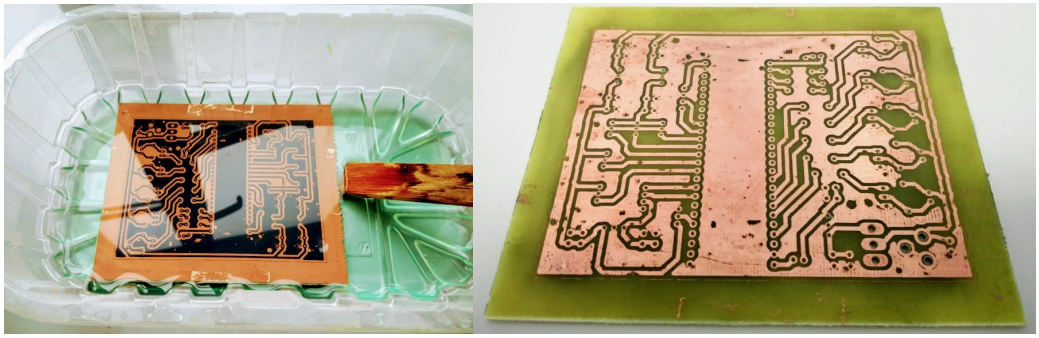
7) once the board was thoroughly etched and rinsed, the etch resistant toner was cleaned off with acetone.

8) then, a continuity test from one end of each trace to the other was performed with a multimeter to make sure that all traces were continuous. The remaining gaps were fixed by making a solder bridge over the breaks;

9) finally, the PCB was drilled (Figure 3.17e) and all components were soldered on it. The final result is visible in Figure 3.16b.



(a) Print on photographic paper **(b)** Ink impression after ironing **(c)** Ink mask after ironing



(d) Board etching in acid bath **(e)** Rinsed and drilled board

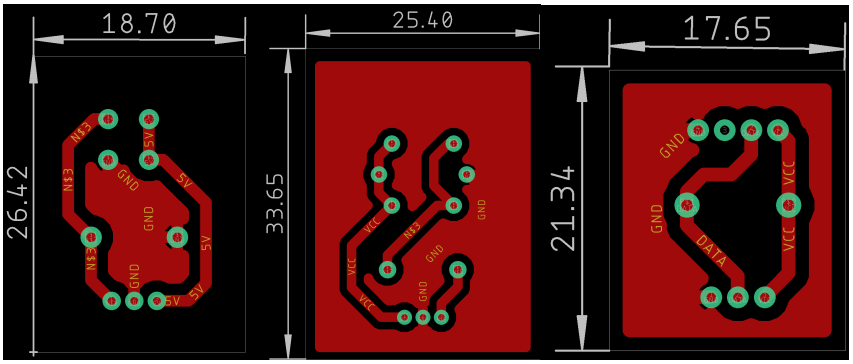
Figure 3.17: *Stages of board printing: a) the circuit is printed in mirror format on photographic paper with a laser printer and b) it's ironed onto the board so that all toner traces are released; c) the ink mask is removed after ironing, then d) the board is etched in an acid bath, e) rinsed and finally drilled to house components.*

3.1.3.1 Sensors boards

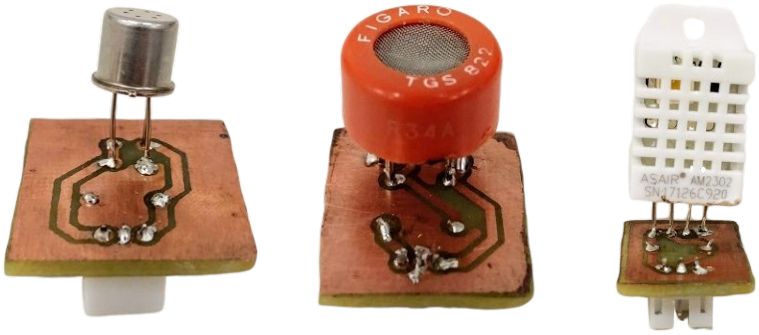
Apart from the main circuit, some other mini-boards have been realized to test the gas sensors described in Section 3.1.1.4 and a temperature sensor which was later excluded from the project (DHT11). These boards are connected to the main circuit by means of JST connectors and allow to change the sensors when necessary: the connectors have 3 lines (5V, GND and signal) and the single sensor conditioning is implemented directly on the mini board, giving flexibility to the main circuit.

3.1.4 Case Design

A simple case was designed in SolidWorks and realized with the *RoboxDual* 3D printer; the material used is polylactic acid (PLA), a common materials for 3D printing. The case is



(a) TGS26xx CAD **(b)** TGS822 CAD **(c)** DHT11 CAD



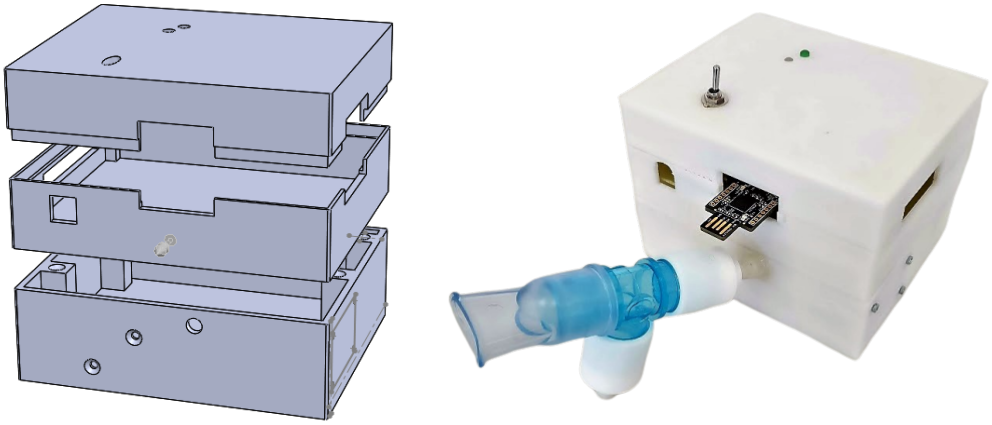
(d) TGS26xx PCB **(e)** TGS822 PCB **(f)** DHT11 PCB

Figure 3.18: *a) TGS26xx type b) TGS822 and c) DHT11 sensors boards design (on top) and d) e) f) their physical realization (on bottom).*

composed by 3 parts creating two separate layers: the lower is designed to house the valve and the CO_2 sensor; the upper contains the PCB with all connectors and circuitry.
The CAD with the assembly of the three parts is reported in Figure 3.19a.
On the lowest level it is possible to see the holes to fix the sensor and the valve with some screws (Figure 3.20a). The medium and the upper pieces, instead, form two lateral windows to have an outlet for the PSoC programmer and the sensors connecting wires (Figure 3.20b). On the top there are three holes for the LEDs and the power switch (Figure 3.19b).

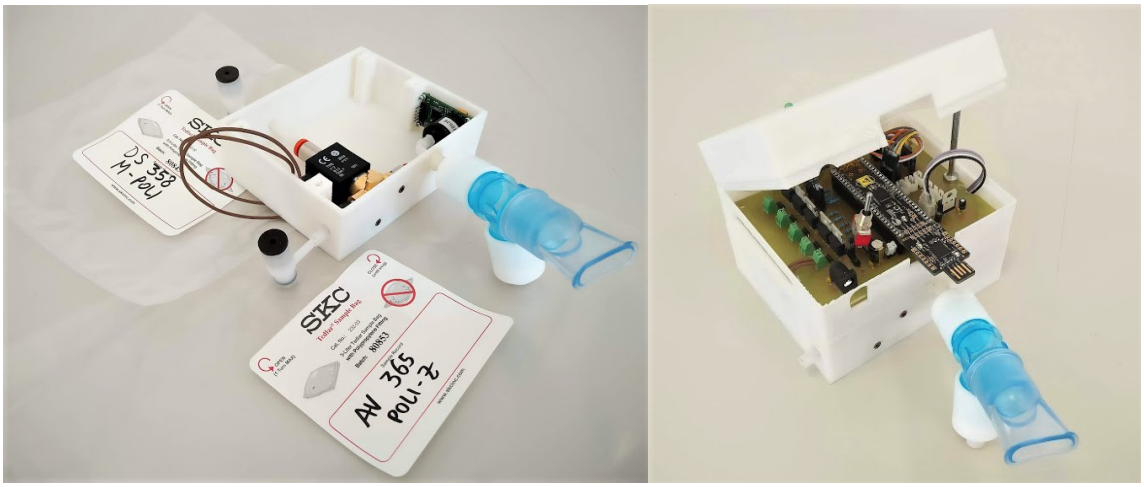
3.2 Firmware

In this section, the implemented firmware will be described. It was programmed exploiting the Integrated Development Environment (IDE) *PSoC Creator 4.2*, which enables concurrent hardware and firmware editing, compiling and debugging of PSoC 5LP. It includes:



(a) (b)

Figure 3.19: *a) CAD project of the case assembly and b) its physical realization.*



(a) (b)

Figure 3.20: *a) Lower part of the case, containing the sensor and the valve, and b) upper level, containing the board.*

- hardware design and easy-to-use wiring tool;
- over 150 production-ready components;
- full communications libraries including I_2C, USB, UART, SPI, CAN, LIN, and Bluetooth Low Energy;
- digital peripherals with powerful graphical configuration tools;

- analog signal chain tools with amplifiers, filters, ADC and DAC;
- dynamically generated API libraries;
- free C source code compiler;
- integrated source editor with inline diagnostics and built-in debugger.

It is a very flexible IDE and results suitable for our application. At first, the components used in the Top Design will be described, later the implemented algorithm will be schematically explained from the PSoC point of view.

3.2.1 Top Design

The Top Design is the schematic editor of PSoC Creator, i.e. the page in which the top-level schematic file can be customized. In this page it's possible to add to the workspace a list of virtual components available in the catalog that allow to configure very easily the real components present on the microcontroller. Once a component is added, it can be simply configured by double-clicking on it and opening its configure dialog menu; to complete its configuration, in many cases it is necessary to connect it to other components or to one or more pins, both digital or analog, that represent the physical pins of the device. Then the virtual pin elements need to be assigned to the real pin (or to the corresponding port) on the Pins page. The pinout page of the current project is reported in Figure 3.21.

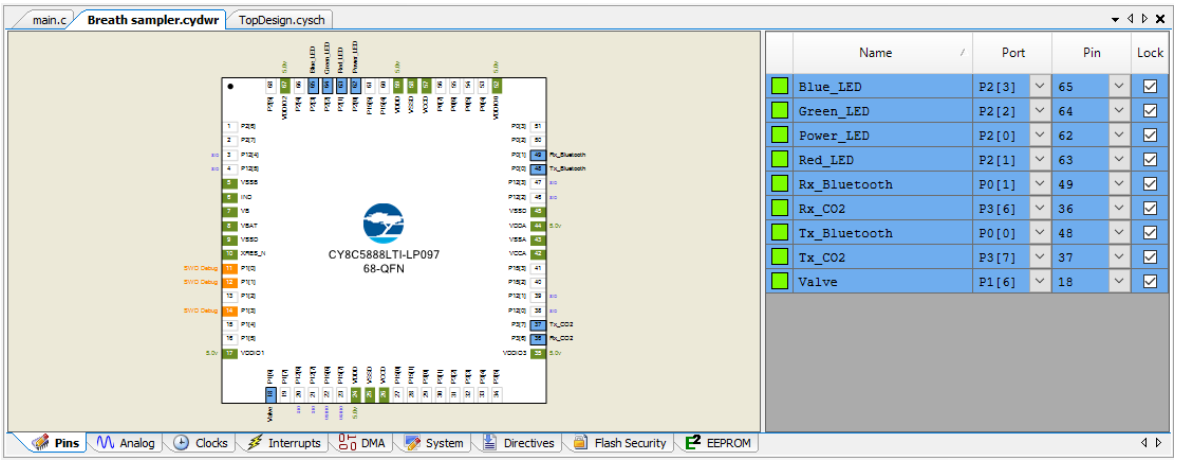


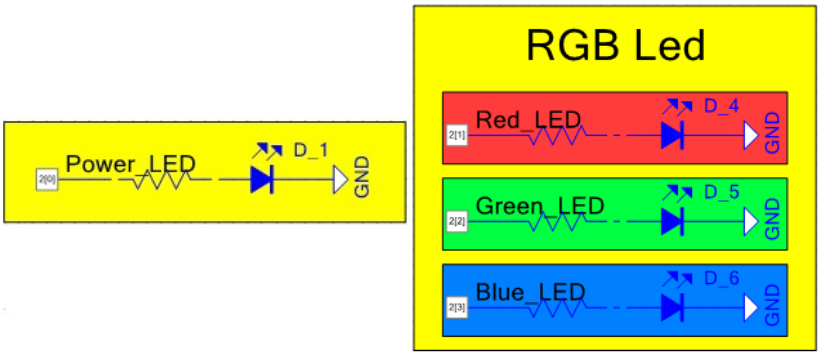
Figure 3.21: *Pins page of the PSoC creator project. On the left, it's possible to see the active pins as highlighted in blue on the microprocessor; on the right, there is a table with the correspondance between the virtual pins names and the physical ports to which they are assigned.*

Finally, by means of the Build button, the project C code is generated and the device can be programmed by compiling the *main.c* file.

In this project, we employed two UART blocks (one for Bluetooth communication and the other to read data from the CO_2 sensor), 4 digital pins for the two LEDs and a fifth pin to control the three-way valve. We will now see each of them more in detail.

3.2.1.1 Power and Status LEDs

As seen in Section 3.1.1.5, a simple green LED and a RGB LED have been chosen respectively as power and status indicators, i.e. to signal when the device is switched on and which phase of the routine it is executing.



(a) Power LED **(b)** Status LED

Figure 3.22: *a) Power and b) Status LED virtual components in the Top Design.*

Since they do not need particular conditioning, at least for this application, their virtual components are just digital pins, represented togheter with off-chip resistors, LEDs and GND symbols as in Figure 3.22.

The power LED is switched on by forcing the high level output to its pin with the function *Power_LED_Write(HIGH)*, while the RGB LED can assume different colors depending on the voltage applied to each pin. In this project we need only two colors, hence no PWM component is employed but each of the three lines is used independently.

3.2.1.2 Three-way valve

The three-way valve, as explained in Section 3.1.2.1, is commuted between the open/close state by means of a power MOSFET used as a switch. For this reason the only signal needed is HIGH/LOW voltage, thus, as in the case of LEDs, a digital pin is sufficient; the remaining

Table 3.3: *Main parameters chosen for the UART Bluetooth component.*

Parameter	Value/Choice
Mode	Full UART (Tx+Rx)
Bits per second	9600
Data bits	8
Clock	Internal
Interrupt Source (Rx)	on byte received

off-chip hardware components visible in Figure 3.23 have been added only for clarity reasons.

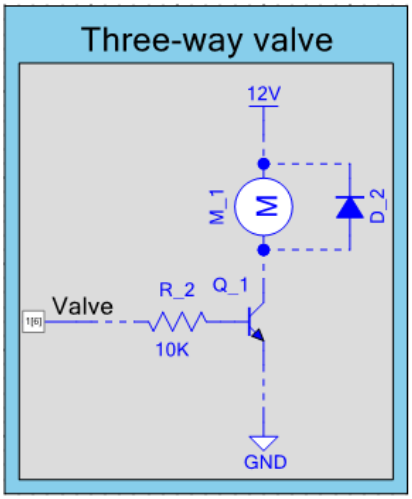


Figure 3.23: *Three-way valve virtual component in the Top Design.*

3.2.1.3 UART Bluetooth

The Bluetooth communication is based on the Universal Asynchronous Receiver Transmitter (UART) protocol, commonly referred to as RS232 or RS485. PSoC Creator has a proper component to handle it with simplicity for the user, which provides support to configure the number of data bits, stop bits, parity and so on. It is reported in Figure 3.24. In this specific case, the main parameters have been chosen as reported in Table 3.3.

The two pins (Rx to receive and Tx to transmit data) have been chosen in such a way to have correspondence between the Rx channel of the PSoC and the Tx of the Bluetooth module and vice versa.

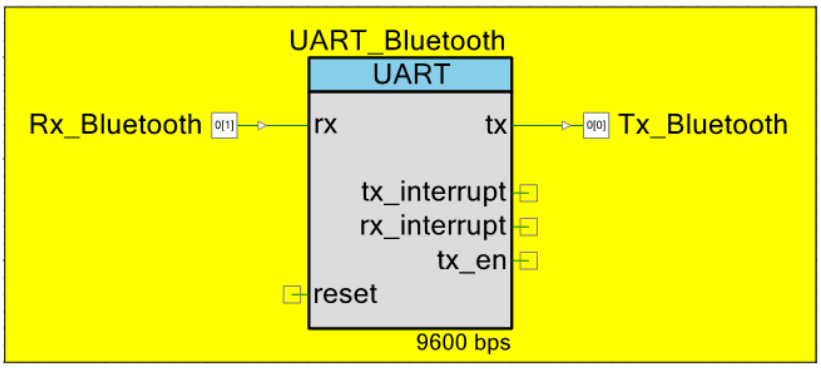


Figure 3.24: *Bluetooth virtual component in the Top Design.*

3.2.1.4 UART CO_2

The CO_2 sensor, as described in 3.1.1.2, has digital 8 bit output with RS232 interface, thus the necessary virtual component is again a UART block as shown in Figure 3.25.

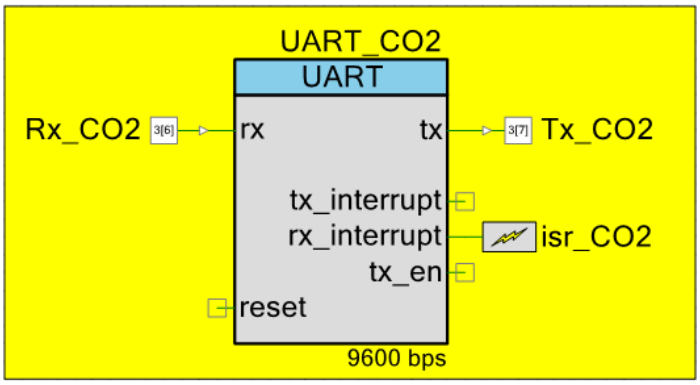


Figure 3.25: *CO_2 sensor virtual component in the Top Design.*

The only difference with the Bluetooth element of the previous paragraph is the presence of an Interrupt block (called "isr_CO2") that defines an hardware-triggered interrupt every time a new character is received. The aim of this interrupt will be explained in Section 3.2.2.

3.2.2 Algorithm

The current section presents a summary of the overall algorithm functioning. The operations are schematically resumed in Figure 3.26 and are divided in two main blocks, respectively the main and the UART CO_2 Interrupt Service Routine (ISR). The routine is very simple and is complementary to the part executed by the Software (described in Section 3.3); here we will explain it from the firmware point of view.

When the device is switched on, all components described in Section 3.2.1 are initialized making use of the function *startSystem()*: the power LED is switched on, the two UARTs are started and the ISR is initialized. Starting from this point, the ISR executes the carbon dioxide level reading via the CO_2 sensor UART. The sensor digital output is reported as:

$$Z\ \#\#\#\#\#\ z\ \#\#\#\#\#\ \backslash r\ \backslash n$$

where $Z\ \#\#\#\#\#$ shows the CO_2 concentration after digitally filtering and $z\ \#\#\#\#\#$ shows the instantaneous CO_2 concentration without any digital filtering (i.e., the raw value). The $\backslash n$ character is used to detect when a line is complete, then the raw value is reconstructed and a new reading starts. In the meanwhile, in the main, the device enters the Waiting phase, in which it expects to receive a character from the software via the Bluetooth UART. Depending on the received character, a new phase is initialized:

- a "c" starts the CO_2 sampling phase, in which the CO_2 concentration data reconstructed in the ISR are sent to the software, the valve opening is controlled and the Status LED is switched to blue color;
- a "g" starts the Cleaning phase, signaled by yellow color, in which data are still sent to the software, but there is no valve control;
- a "s" stops the streaming (or the cleaning) and returns to the Waiting phase.

The sequence of the three phases, which seems to be random in the firmware, is regulated in the software, as we will see later.
The control of the valve is accomplished with the function *ValveControl(val_prev, val_curr)*, where *val_prev* and *val_curr* are respectively the previous and the current values of CO_2 concentration.

The initial idea was to implement a **multi-breath sampler** similar to the one developed by Salvo *et al.* [24], described in Section 2.3.2. An algorithm was implemented to accomplish the following tasks:

- identify each respiratory act and its inspiratory and expiratory phases;
- compute and update the $CO_{2,max}$ value of each expiration based on Equation (2.2);
- open the valve whenever the signal derivative (approximated as the difference between *val_prev* and *val_curr*) is higher than a fixed value and *val_curr* overcomes a dynamic threshold depending on the current value of $CO_{2,max}$;

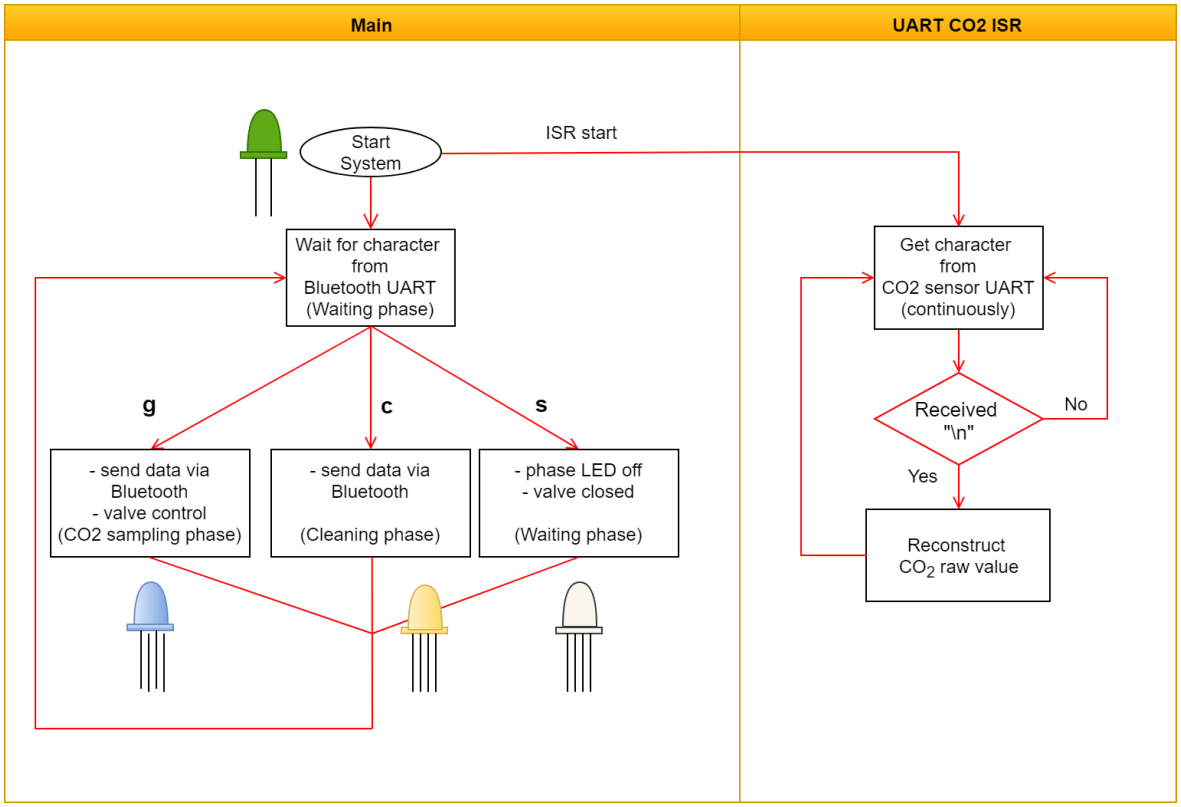


Figure 3.26: *Schematic diagram of the firmware algorithm.*

- close the valve whenever the signal derivative is lower than a fixed negative value and *val_curr* decreases below a dynamic threshold;
- identify spurious respiratory act (e.g. having too high frequencies or not enough expiratory volume) and exclude them from the updating of $CO_{2,max}$, to make the algorithm more robust.

The algorithm was developed at first in MATLAB (as post-processing on saved data) and later in PSoC Creator (as real-time approach). Some preliminary tests have been conducted with a first version of the prototype not including the three-way valve (which was still in delivering); their results are reported in Chapter 4. As will be discussed in Chapter 5, the multi-breath approach resulted unfeasible once tried with the final prototype, thus the system was simplified and modified to sample a single breath.

The **single-breath sampler**, based on the current version of the function *ValveControl*, works instead on two separate phases:

1) the CO_2 sampling phase, in which a single deep breath is sampled and the valve is switched only once;

2) the Cleaning phase, necessary after each sampling to remove the instrumental dead space air remained in the system and to keep the CO_2 sensor to its baseline value (about 450 ppm).

During the sampling, the time instant when there is the passage from dead space to alveolar fraction is identified as the point in which the signal reaches its plateau (in this case, being in presence of a deep breath, the plateau is the instant in which the sensor saturates to 20 000 ppm); since there is a certain air volume that fills the sensor cap, the tubes and the valve before reaching the sampling bag, named *instrumental dead space*, the valve switching is postponed of a time interval computed empirically and equivalent to 30 samples (i.e., about 3 seconds). The cleaning, instead, is managed by the software, hence the only role of the firmware in this phase is to send the CO_2 sensor readings to the Bluetooth UART and to wait for a "s" to be stopped.

3.3 Software

In the current section, the Software will be characterized in both its graphical appearance (the User Interface) and the underlying implementation (the functioning algorithm). The employed programming language is Python 3, with Tkinter being the fundamental package for GUI development. Python is a multi-paradigm language that supports object-oriented and structured programming, it is open-source and contains over 130000 packages with a wide range of functionalities, being ideal to develop a desktop application as the one we are going to illustrate. In the following paragraphs, all details will be given and every choice will be motivated.

3.3.1 Graphical User Interface

The Graphical User Interface is the part of the software that allows the user to interact with the electronic device through icons, buttons, labels and other graphical objects. In the biomedical field, the user is generally a physician (or any kind of healthcare professional) who has not deep knowledge of the engineering behind a system, but instead needs to manage the device efficiently and in the simplest way as possible and to have availability of all necessary data and parameters.

In this project the GUI is a fundamental instrument since, in combination with the firmware, it allows to:

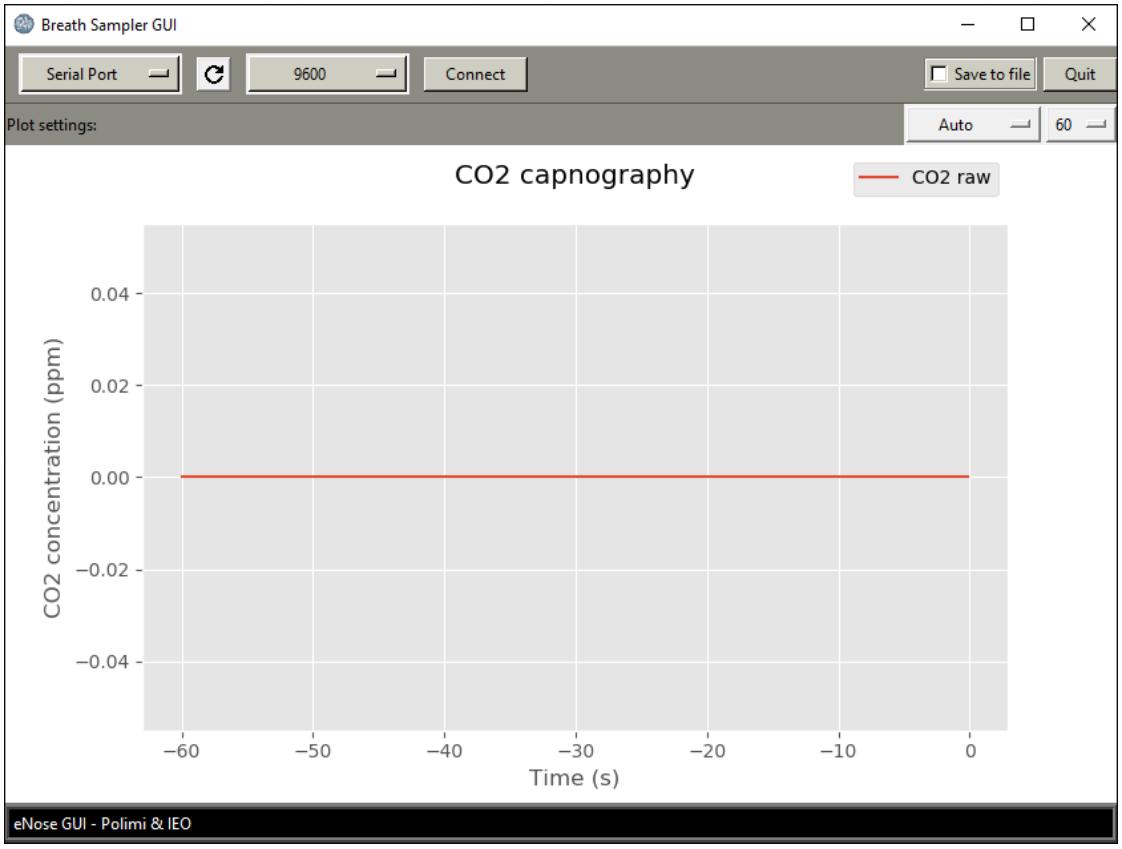


Figure 3.27: *Starting window of the Graphical User Interface.*

- establish the Bluetooth connection with the device;
- start and stop both the Sampling and the Cleaning phase;
- save data in a file when necessary;
- visualize in real-time the recorded CO_2 concentration.

The starting window is reported in Figure 3.27; it is composed by three main blocks:

– ***Top Navigator***, the grey bar on the top, which is the control point of the software. It contains, from left to right:

1. the *Serial Port list*, that shows all available serial ports and lets the user choice the one corresponding to the device Bluetooth UART;
2. the *Refresh Button* (the one containing the circular arrow) which, if pressed, gets the updated list of ports in the *Serial Port list*;
3. the *Baudrate list*, a second option menu containing the possible choices of baudrate for the serial communication. The default baudrate is equal to 9600 bps;

4. the *Connect* button, used to establish a connection once the serial port and the baudrate have been chosen. When the device is connected, its label changes to *Disconnect* and it accomplishes the opposite task;

5. the *Save to file* checkbutton. If set to on, it opens a window to insert patient's generalities (Figure 3.28) and creates a *.csv* file in which the streaming data are saved.

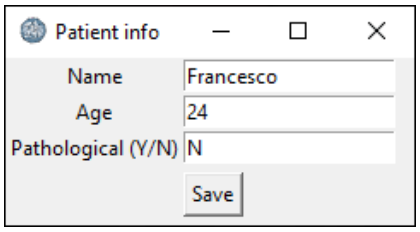


Figure 3.28: *Window to get patient's info. It appears when the user decides to save data in a file.*

6. the *Quit button*, which closes the application.

– ***Plot Area***, the central part of the window. It contains the area in which the CO_2 concentration data are plotted against time and a bar with the plot settings. The bar itself contains two option menus used to set the plot time interval (from 5 to 60 s) and the displayed amplitude (full scale or automatic).The library used to create the plot is *Matplotlib*, a Python 2D plotting library which produces quality figures with full control of line styles, font properties, axes properties, etc.

– ***Bottom Label***, the black bar on the bottom, in which a feedback message is printed every time the software changes operation or an error occurs.

3.3.2 Standard Functioning

The current section completes the description of the device working principle started in Section 3.2.2, but from the software point of view.

When the application has started and the system has been switched on, the window of Figure 3.27 appears. The user can now select the serial port (usually COM7) and the baudrate (9600 bps for this program) from the relative option menus and start the Bluetooth connection with the device by pressing the *Connect* button. In case the serial port or the baudrate are not selected correctly, a reminder message will appear in the *Bottom label*, otherwise a first

attempt of connection will start. Example of the printed messages are reported in Figure 3.29.

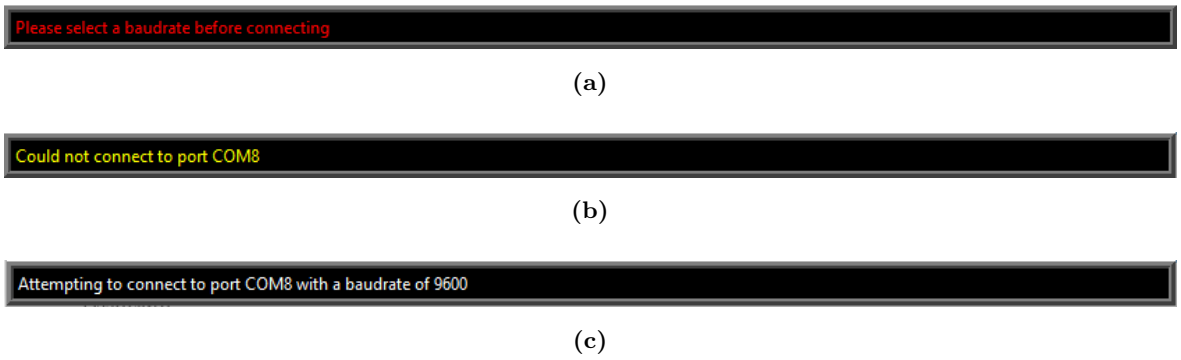


Figure 3.29: *Error or reminder messages printed in the Bottom label in case of a) not selected baudrate, b) wrong serial port or c) attempt to connect with the device.*

Once the device is correctly connected, the *Bottom label* will signal it with a message and two new widgets will appear in the *Top Navigator*, as represented in Figure 3.30.

Figure 3.30: *Top Navigator updated after the connection.*

The first is the *Start Session* button, whose function is to give start to the CO_2 sampling phase. The second one, the *Phase label*, will signal the current phase during each step of the procedure. Before starting the acquisition, the user can decide to save data in a *.csv* file by selecting the *Save to file* checkbutton and entering the data requested in the *Patient info* window of Figure 3.28. Then, when the *Start Session* button is pressed, the CO_2 sampling phase is started, the *Phase label* and the *Bottom label* are updated and the data are plotted in the *Plot Area* almost in real-time, as represented in Figure 3.31.

The *Connect* button is now disabled, while the *Start Session* button becomes the *Stop session* button and has to be pressed once the breath sampling is finished. In case the session is not stopped by the user, it will be terminated automatically after a standard time. At the end of the CO_2 sampling phase, a message box appears to signals the end of the session and ask the user if he/she wants to clean the sensor (Figure 3.32).

In case of negative answer, the application returns to the Waiting phase, otherwise a second message box appears to give information on how to set the device for the last part of the procedure. The former *Stop session* button changes its label to *Start cleaning*, and it can be

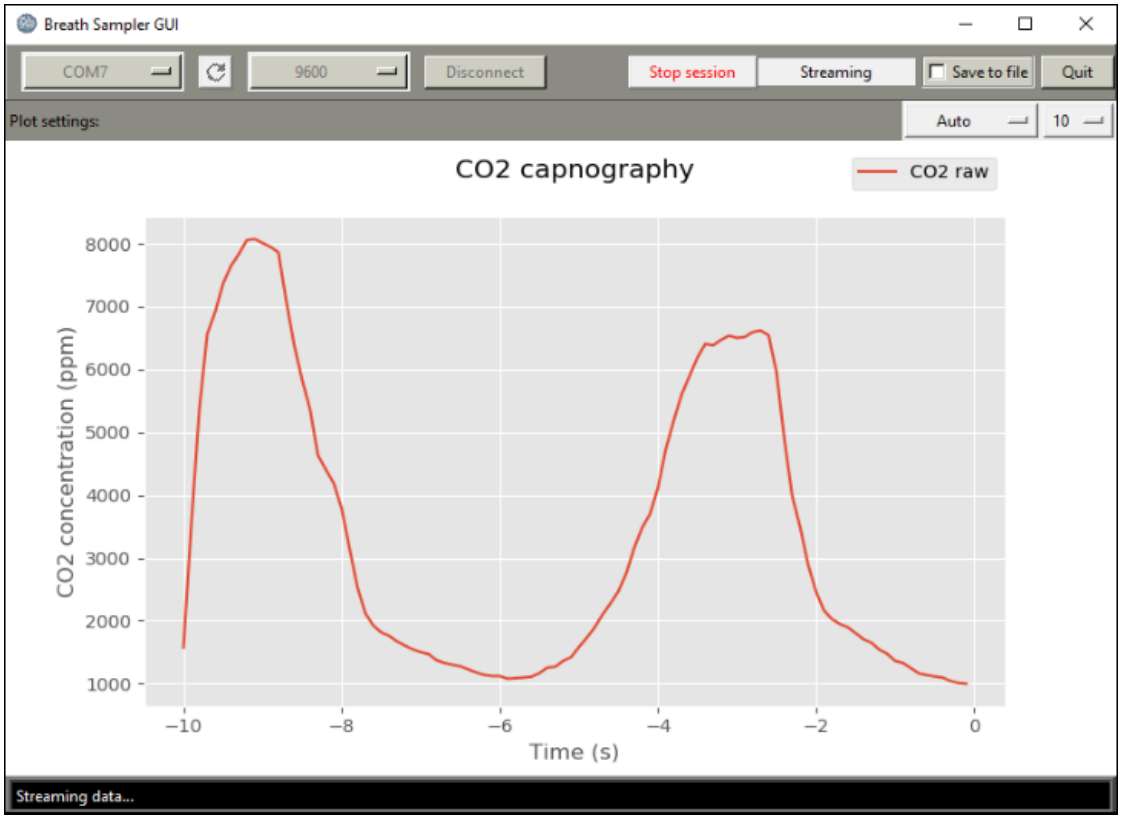
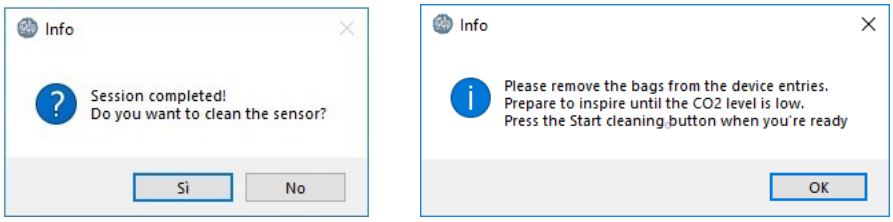


Figure 3.31: *GUI during the CO_2 sampling phase. Concentration data are plotted against time in real-time.*



(a) **(b)**

Figure 3.32: *Message boxes appearing at the end of the acquisition. a) The user can decide to clean the sensor or to go back to the Waiting phase; b) instructions to prepare the Cleaning phase.*

pressed to start the Cleaning phase. The user has two possibilities to clean the sensor:

1. take the air full of CO_2 out of the device by inspiring through the mouthpiece;

2. use an external pump (as the one of the electronic nose) to extract the air.

The cleaning is considered to be complete when the CO_2 concentration level reaches its baseline, more specifically the algorithm looks for 5 consecutive samples with values lower than

the threshold of 500 ppm. If the cleaning takes more than 60 s, it is stopped automatically. The procedure is now finished and the application is ready to start a new sampling.

3.3.3 Software Implementation

The described application is a multi-threaded program, i.e., it contains two or more parts that can run concurrently, share information and communicate between each other. Threads have their own life cycle, it means their execution is not necessarily continuous, but they can also be run only in certain phases of the program. The current implementation is schematized in Figure 3.33.

The program starts by launching the **main thread**, which sets up the starting GUI and initiates the Waiting phase. This thread handles the selection of the serial port and of the baudrate, the eventual refresh of the *Serial Port list* and, in case the user decides to save the data in a file, also the opening of the *Patient's info window* and the creation of the file (the light blue color and the dashed lines indicate that this path is not mandatory).

Then, when the *Connect* button is pressed, the **Connection thread** is started and it tries to establish the Bluetooth connection with the device. In case there is one of the errors described in the previous section, the program goes back to the main thread, otherwise the connection is established and the Connection thread is interrupted. At this point the application is ready to start the acquisition.

When the user presses the *Start session* button, the program passes to the CO_2 sampling phase and tries to create the Streaming thread; in case of failure, it goes back to the Waiting phase. If the thread is created with success, the software sends a message to the device to receive the data through the Bluetooth UART (see Section 3.2.2 for further details). Data are plotted one after the other and, when the *Save to file* checkbutton is set to on, they are saved ten at a time. This thread is ended when the user presses the *Stop session* button or, as an alternative, when the session time is over.

The user can now decide to clean, or not, the sensor. In case of negative answer, the procedure is finished and can be restarted from the Waiting phase, otherwise the Cleaning phase is introduced. The **Cleaning thread** is launched when the relative button is clicked, and data are plotted just as in the Streaming thread. The software therefore checks if the CO_2 concentration level lowers down to the baseline as explained in Section 3.3.2 (of if the cleaning time ends) and interrupts this last thread. The procedure is now complete and the system is ready to start a new acquisition or to be quitted.

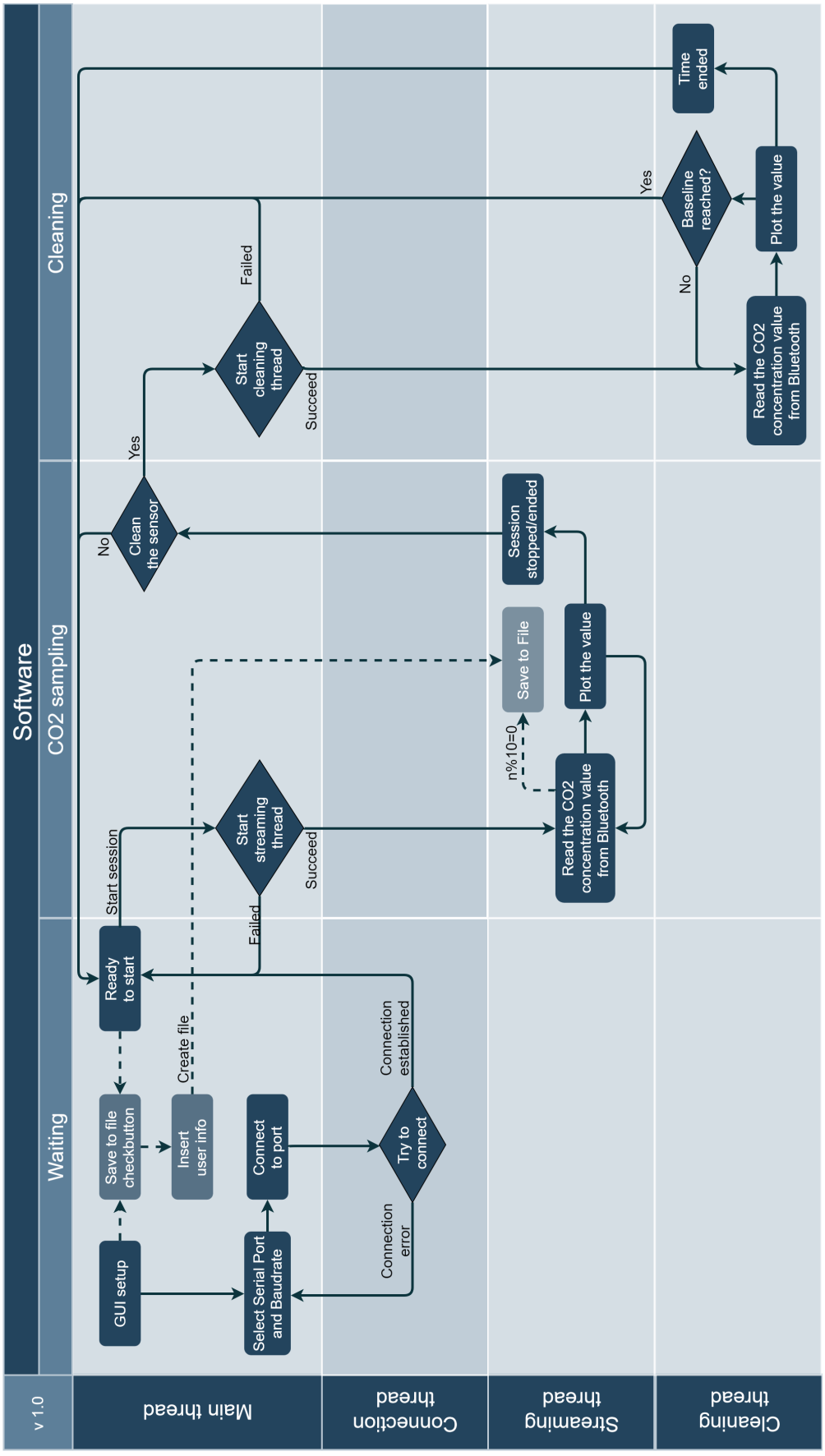


Figure 3.33: *Schematic diagram of the software implementation.*

Chapter 4

Results

In the previous chapters we have discussed the rationale of breath separation in the field of electronic noses, and we had an overview of the materials and the methodologies employed for the development of a device intended to perform this operation. We will now expose to the reader the main results obtained during the various phases of this work, while their discussion is postponed to the next chapter.

The device designed in this project is a single-breath sampler aiming at separating the two contributions contained in exhaled breath, i.e., dead space and alveolar air.
In Section 3.2.2 we introduced that the first approach aimed at developing a multi-breath sampler, but this resulted unfeasible and the system had to be simplified. In the following sections we are going to describe how the data were acquired, the results obtained with both types of acquisition and the way they were analyzed. The software used to process and plot the data in this part of the project is MATLAB.

4.1 Acquisition Protocol

4.1.1 Type of acquisition

A **multi-breath sampler** is a system able to sequentially sample the content of multiple breathing acts and to separate, in each of them, the two portions of exhaled air. The subject who undergoes the analysis only has to inhale and exhale through the mouthpiece maintaining a relaxed respiratory pattern until the operator stops the acquisition. During this time the device should fill, breath after breath, both the dead space and the alveolar sampling bags.

A **single-breath sampler** aims at performing the same operation, but on a single breathing act. The acquisition is a little more challenging: the subject has to inhale ambient air to total lung capacity, and then to exhale as hard and as completely as possible through the mouthpiece until he/she is out of breath. The exhaled air automatically fills at first the smaller bag (dedicated to dead space) and later, when the valve is switched, the other bag (with alveolar air).

Finally, a **mixed-expiratory breath** acquisition consists in sampling only exhaled air without performing any kind of separation. As we have seen in Chapter 2, many studies don't perform breath separation and still analyze mixed-expiratory air.

4.1.2 Population and hygienic procedure

The population involved in this study consists of 10 healthy volunteers with the characteristics listed in Table 4.1. No cancerous subjects were included because the study is still in an exploratory phase; furthermore, the device should be certified in order to be employed in clinic with patients.

Table 4.1: *Population involved in the study. All subjects are healthy volunteers.*

Subjects' characteristics	
Number	10
Age (years)	25±1
Sex (females/males)	4/6
Current smokers/ex-smokers/never smokers	0/2/8
Declared respiratory pathologies	None

All subjects were asked to follow a hygienic procedure from the day before the acquisition. The main guidelines were the following:

- to abstain from food and drink (except water) within eight hours before the test;
- to avoid garlic, onion and other aromatic food from the day before the test;
- not to smoke from the night before the test;
- to brush the teeth after the last meal taken the evening before the test;
- to avoid perfumes or scented soap in the twenty-four hours before the breath test.

Such a procedure is intended to reduce interferences in the breath composition, and it is adapted from the protocol described by R. Gasparri *et al.* [35]. Three acquisitions were taken for each subject: 2 with the single-breath sampler and one of mixed-expiratory air, sampling a total of 5 bags per person as depicted in Figure 4.1.

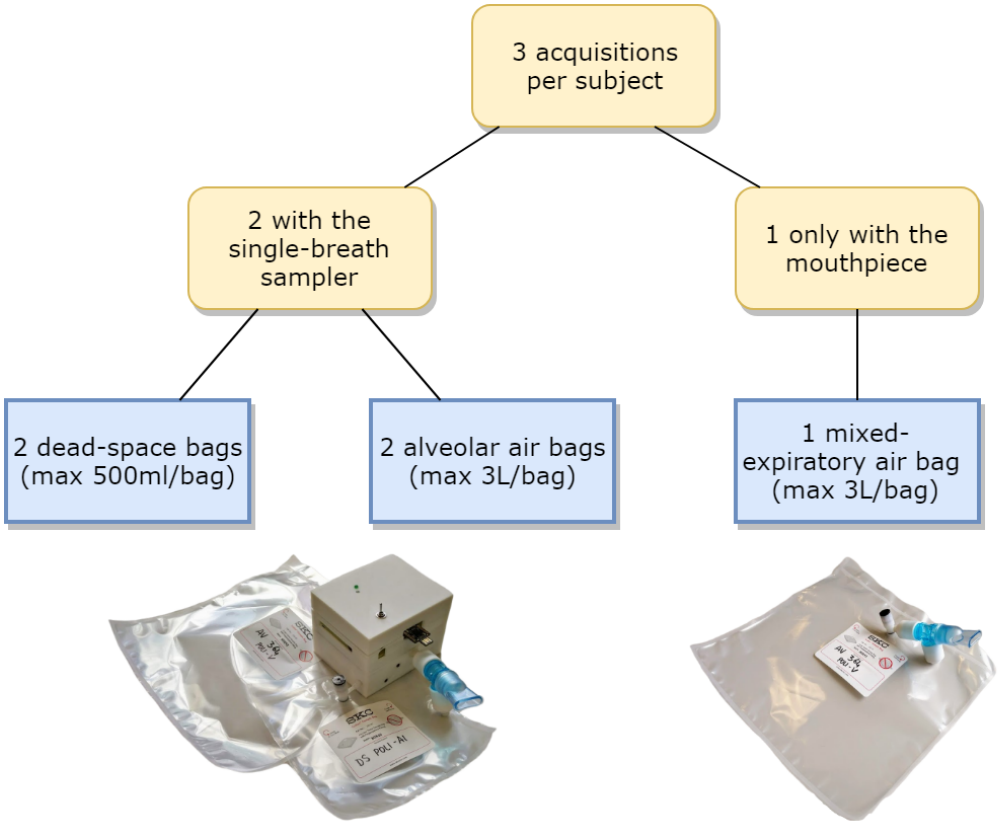


Figure 4.1: *Block diagram of the acquisition taken for each subject.*

4.2 Standard Sensors Response

The sampling bags were subsequently analyzed with the electronic nose prototype in development as part of the collaboration between IEO and Politecnico di Milano. An image of the instrument during the analysis is shown in Figure 4.2.

The eNose contains five gas sensors of the type described in Section 3.1.1.4 in a closed gas chamber. The procedure consists of three successive phases:

1. the **Cleaning phase**, in which ambient air is pumped in the gas chamber to remove remaining gas particles until the sensors response is stable;

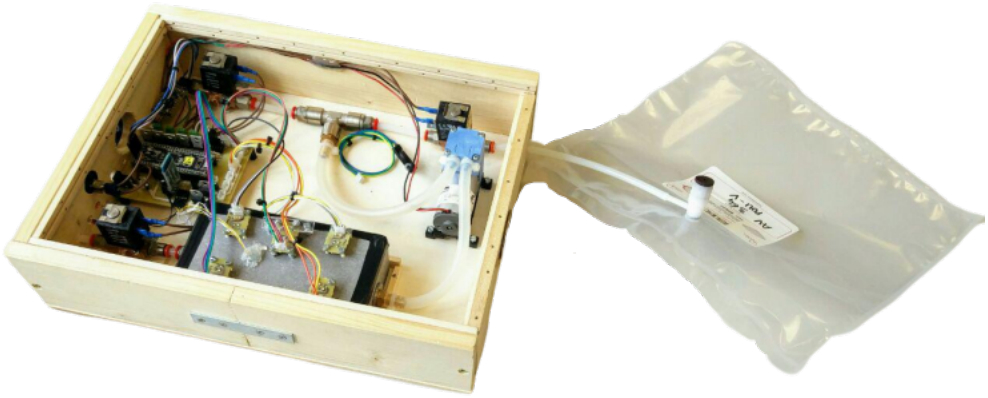


Figure 4.2: *Image of the electronic nose prototype during the analysis of a sampling bag.*

2. the **Measuring phase**, i.e., the key moment of the analysis, in which the sampled air is pumped into the chamber producing a change of the sensors' conductivity;

3. the **Recovery phase**, similar to the Cleaning phase, necessary for the instrument to return to the initial baseline.

An example of the typical sensor response is reported in Figure 4.3.

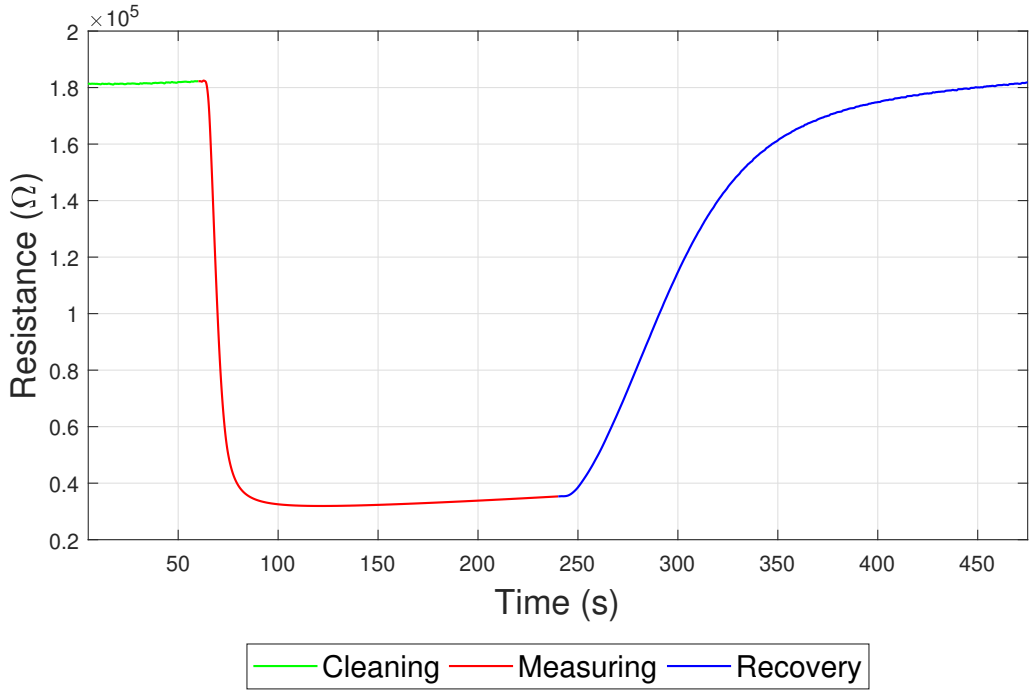


Figure 4.3: *Example of a typical sensor response.*

The cleaning and the recovery phases have a dynamics depending on the system itself.

The stability of the sensors is checked by comparing the standard deviation of the signal, computed in time intervals of 20 s, with an empirical threshold of 100 Ω: the gas chamber is considered to be clean when the standard deviation is lower than this threshold for three times in a row, thus the cleaning phase lasts at least 1 minute. Furthermore, the pump employed during both phases is activated at constant power, thus the ambient air flow is fixed and the recovery happens with the same velocity.

The response during the measuring phase, instead, is strongly dependent on the air sample and is different for each sensor, being each of them more sensible to a certain combination of volatile particles (see Table 3.2 for further details). For this reason, the signals relative to the three air contributions (dead space, alveolar and mixed-expiratory) have visible differences in their profiles, in particular in the minimum value reached at the end of the steep descent. A comparative example is reported in Figure 4.4.

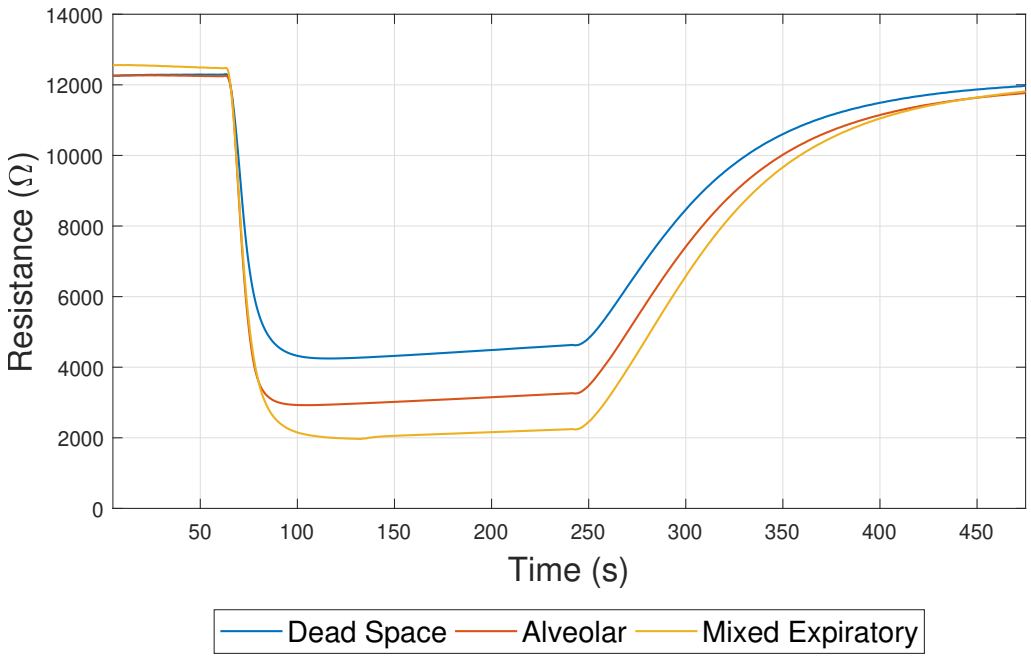


Figure 4.4: *Comparison of the signals relative to dead space, alveolar and mixed-expiratory air of one of the volunteers, recorded with sensor TGS2600.*

Even if the separation of the three waveforms appears to be clearly visible in this example, it is not so marked for all acquisitions and, furthermore, it is qualitative. For this reason, as we will see in Section 4.3, some features will be extracted and combined between them to quantitatively assess if the signals can be significantly differentiated.

4.3 Data Analysis

Data analysis is a very broad scientific field in continuous evolution and we explored some of the solutions adopted in the literature. Anyway, it is not the main topic of this thesis project and, due to time and resources limits, acquired data are not so numerous. Furthermore, there are no golden standard in this research topic and there was no possibility to employ expensive instrumentation like GC-MS or professional CO_2 monitors for an exploratory work like this. Thus, we have no claim to perform an extensive data elaboration and the aim of this part of the work is just to validate in a reasonable way the breath separation performed by the device.

4.3.1 Pre-processing

First of all, data have to be pre-processed to be ready for the successive elaboration. In fact, just by looking at the waveforms reported in Figure 4.4, two problems emerge: the initial baseline reached by the sensor during the Cleaning phase is not equal for all of them; all signals are corrupted by some noise which needs to be reduced, although not obvious to the naked eye. The solutions adopted for both issues will be now exposed.

4.3.1.1 Baseline Correction

The baseline, i.e. the level of the signal reached during the Cleaning and at the end of the Recovery phase, should ideally have the same value in all acquisitions. Anyway, the electronic nose is a prototype in continuous evolution and it is reasonable for it to have some non-idealities such as a slightly variable baseline level. This value, that will be indicated as R_0 from now on, depends on the heating of the gas sensors, thus, starting from the time when the instrument is switched on, it may increase measure after measurement. To assess the robustness of the analysis, the variability of R_0 should be negligible with respect to R_0 itself. For this reason, the distribution of R_0 for the available samples has been investigated and the result is shown in Figure 4.5.

As a second step, the baseline has been subtracted from each signal. This measure has been taken just to align the starting point of the signals and to have a visual feedback of the resistance reductions relative to the three air contributions (Figure 4.6). Anyway, in the successive steps, this baseline correction has not been taken into account for the computation of the features.

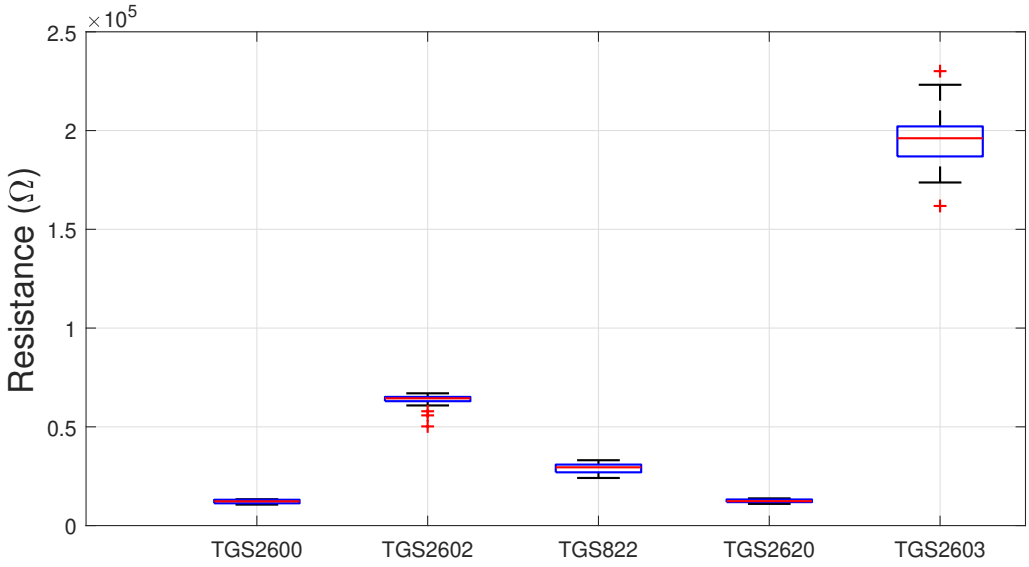


Figure 4.5: *Boxplot reporting the distribution of R_0 in the acquired signals.*

4.3.1.2 Noise Reduction

All recorded signals are corrupted by some noise which is probably related to the activation of the pump. The pump, indeed, is in operation during the entire Cleaning and Recovery phases, and during the first part of the Measuring phase. Since it is difficult to notice such noise in the images shown up to now because of their scale, two zoomed regions are reported in Figure 4.7.

Two types of filters have been implemented, taking into account that the noise is affecting in particular the low frequencies:

- a **moving average** filter with window size equal to 30 samples;
- a **low-pass** FIR filter with the parameters reported in Table 4.2 and the Kaiser windowing as design method.

Table 4.2: *Design parameters of the low-pass FIR filter.*

Filter parameters	
Passband frequency (Hz)	0.05
Stopband frequency (Hz)	0.25
Passband ripple (dB)	0.5
Stopband attenuation (dB)	65

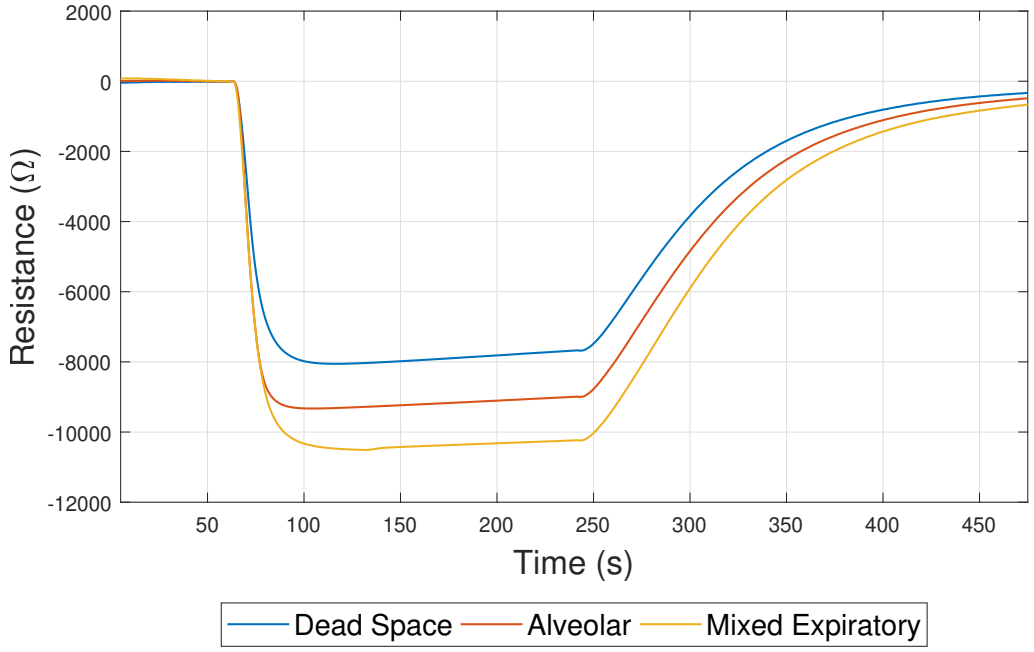


Figure 4.6: *Example of the signals of all five sensors after baseline correction.*

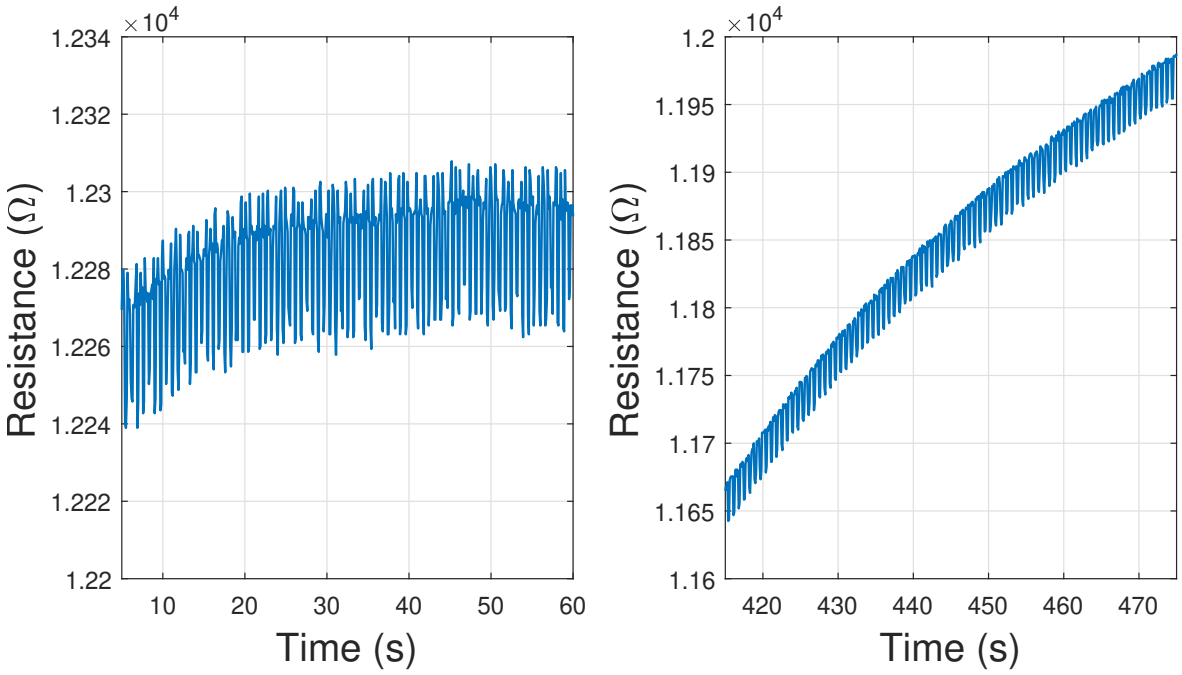


Figure 4.7: *Zoomed intervals of the Cleaning (at left) and Recovery phases (at right). The noise due to the pump activation is now visible.*

The parameters employed in both methods have been tuned empirically, since the noise is not uniform in the three signal intervals. The results of the filtering operations are shown in Figure 4.8, again on two zoomed regions.

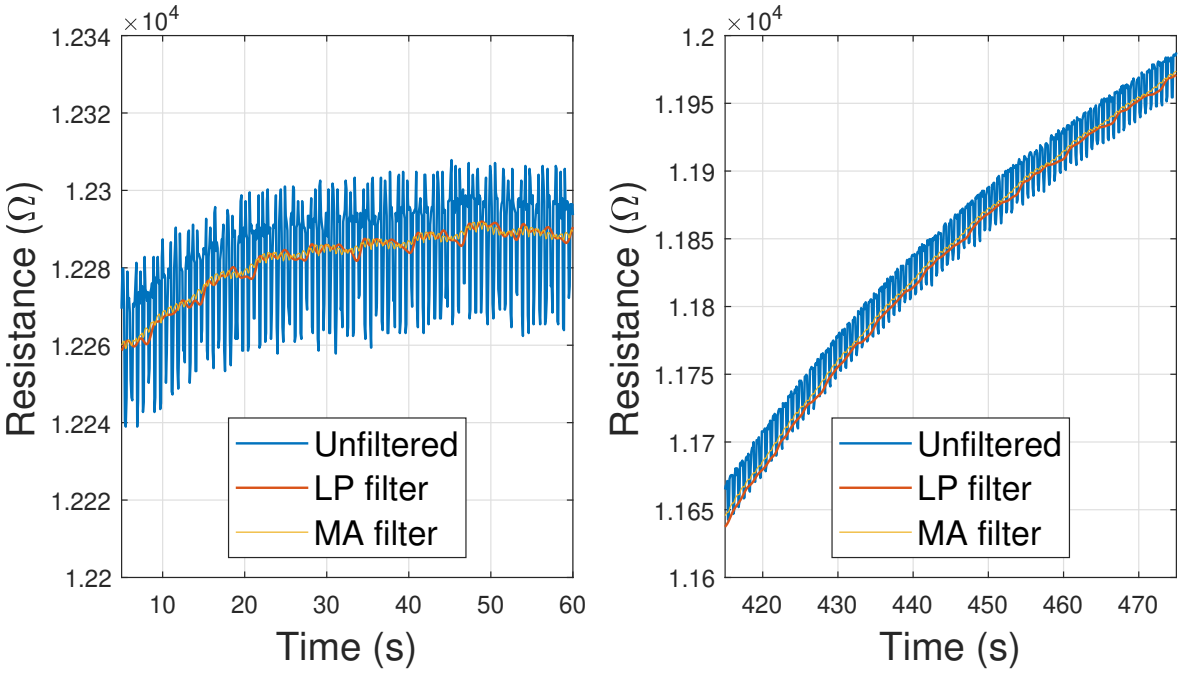


Figure 4.8: *Zoomed intervals of the Cleaning (at left) and Recovery phases (at right), showing the effect of the two filtering methods.*

Both filters seems to work well, at least for the requirements of this work. We decided to employ the low-pass filter since its output envelope appears to be a bit smoother in all regions of the signal.

4.3.2 Feature Extraction

Feature extraction is a generic expression to indicate those operations that, starting from the initial set of measured data, compute derived values (the features) intended to be informative and non-redundant, easing the subsequent learning and classification steps. In this project we extracted five features from each signal, thus, considering that data are recorded by five different sensors, there are 25 features in total. The choice of those parameters is based on considerations done in the literature, in particular 3 out of 5 are adapted from the study of Blatt *et al.* [38], one from the work of Martinelli *et al.* [67] and the last one has been added as our contribution.

Let define $R(t)$ the resistance curve, R_0 the value of resistance reached during the Cleaning phase and R_{min} the minimum reached during the Measurement phase, we have chosen the following features:

1. *Classic*, the ratio beetween the baseline and the minimum resistance:

$$C = \frac{R_0}{R_{min}} \tag{4.1}$$

2. *Delta*, the resistance change during the Measuring phase:

$$\delta = R_0 - R_{min} \tag{4.2}$$

3. *Slope*, the maximum inclination of the resistance curve during the Measuring phase:

$$S = max\left(\frac{dR(t)}{dt}\right) \tag{4.3}$$

4. *Relative Integral*, i.e., the area contained in the curve during the Measuring phase, relative to the baseline:

$$I_{rel} = \int_{meas} \frac{R(t)}{t \cdot R_0} \tag{4.4}$$

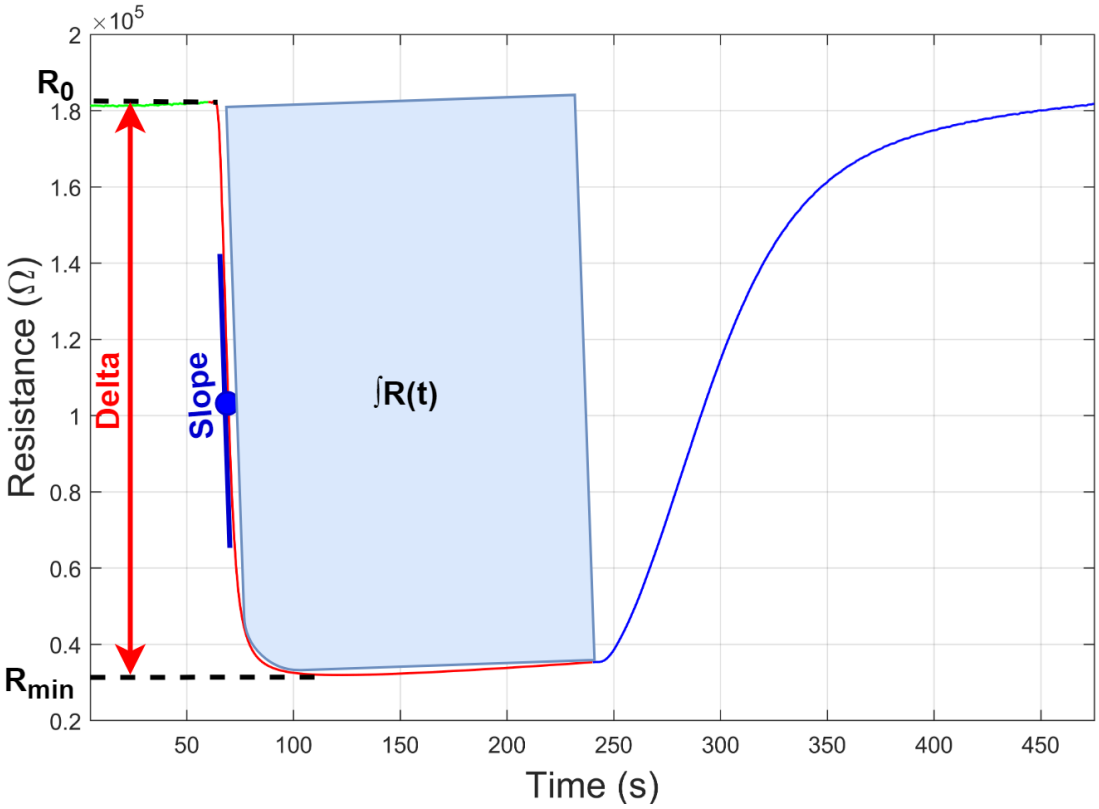


Figure 4.9: *Visual representation of some of the parameters employed for feature computation.*

5. *Phase Integral*, which is the area of the phase space of sensors'response:

$$I_{phase} = \int_0^{R(t)} \frac{dR(t)}{dt} \cdot dR(t) \tag{4.5}$$

The phase space is a dimension having as canonical variables the resistance change itself *R(t)* and its derivative *R'(t)*. The trajectory resulting in this space represents the dynamic evolution of the system in time, and offers a further information to be added to the static parameters computed up to now. In particular, the lower part of the curve represents the Cleaning and Measuring phases, while the upper one is relative to the Recovery. In our case the resulting shape is not perfectly closed since the baseline reached by the sensors at the end of the Recovery phase is slightly different from that of the Cleaning.

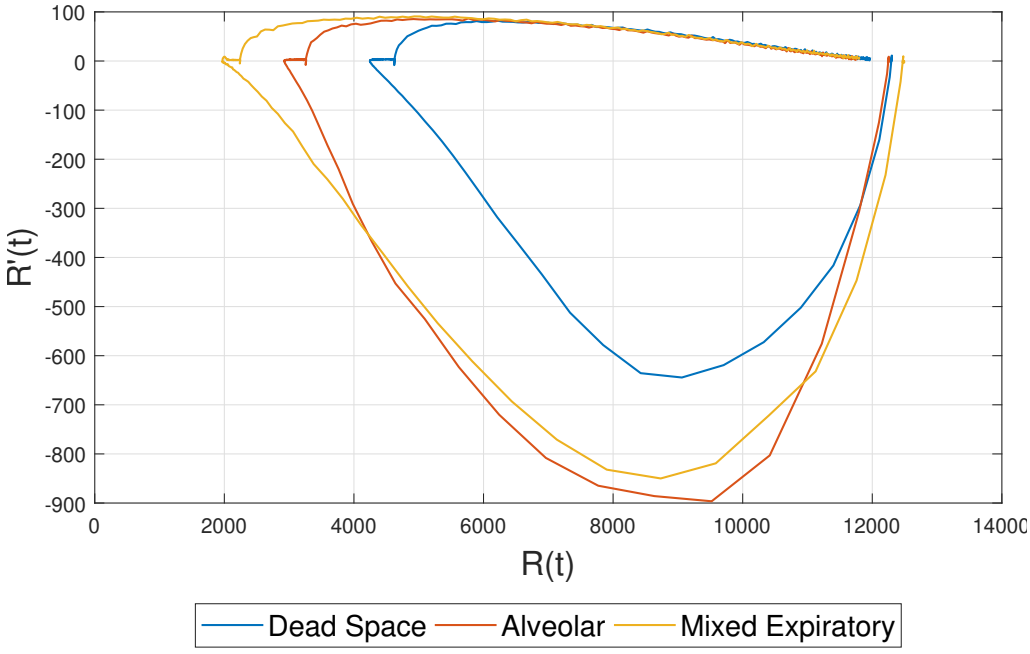


Figure 4.10: *Example of the reconstructed trajectory of a signal in the phase space.*

At this point, the features of each breath sample have been labeled as DS, AV and ME (respectively Dead Space, Alveolar and Mixed-Expiratory) to be recognized in a second moment. After that, they have been merged into a single vector and z-score standardization has been applied (this operation scales data to have mean 0 and standard deviation 1). Once standardized, all features have been re-divided in the 3 populations‘ vectors and displayed in boxplots. The boxplot of the standardized distribution of Classic for all five sensors is reported in Figure 4.11 as an example.

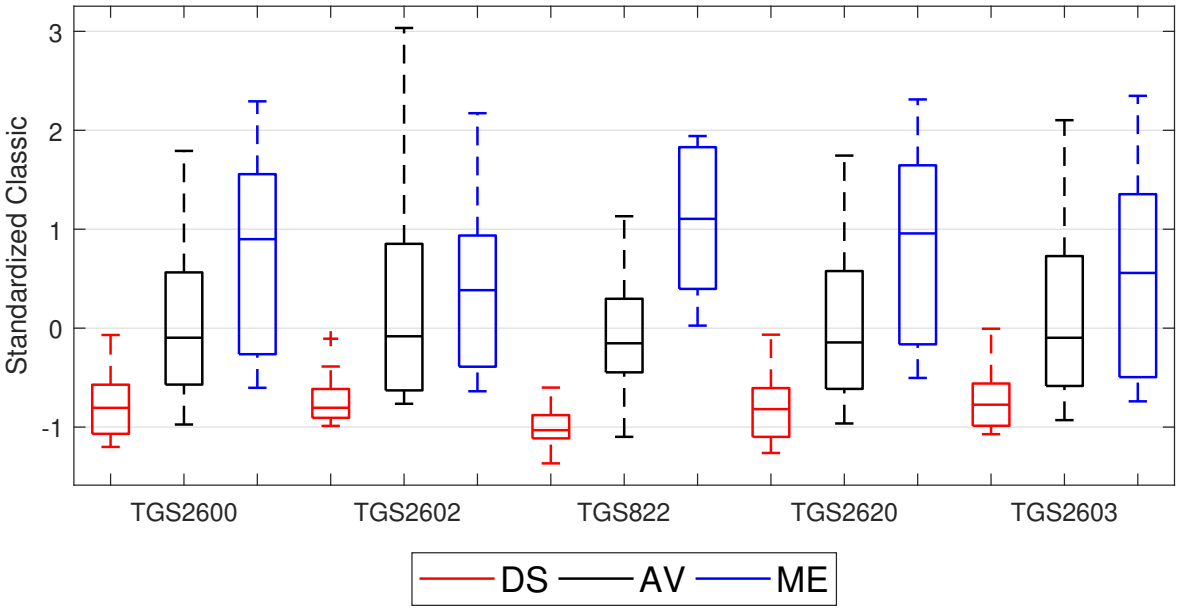


Figure 4.11: *Boxplot of the standardized distribution of Classic.*

Finally, to test the separability of breath contributions, some statistical tests have been applied. First of all, the *Lilliefors test* was used to check if the extracted features satisfy the normality condition. The result of this test is 1 if the null hypothesis that the data come from a distribution in the normal family is rejected at the 5% significance level, and 0 otherwise. The output is reported in Table 4.3.

Table 4.3: *Output of the Lilliefors test applied to the extracted features. The output is 0 if the distribution can be assumed to be normal, 1 otherwise.*

	DS					**AV**					**ME**				
Sensor nr.	*1*	*2*	*3*	*4*	*5*	*1*	*2*	*3*	*4*	*5*	*1*	*2*	*3*	*4*	*5*
Classic	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0
Delta	0	0	0	1	1	0	0	0	0	0	1	0	0	1	0
Phase Integral	0	1	1	0	0	0	1	1	0	0	1	0	1	1	0
Relative Integral	1	1	0	0	0	0	1	1	0	0	0	1	1	1	1
Slope	0	1	0	1	1	0	1	0	1	0	1	1	1	1	1

Then, on the basis of this output, two alternative tests have been applied to check the separability of DS and AV samples, and AV and ME samples:

- the *Two-sample t-test*, comparing the means when populations are normal (more specif-

Table 4.4: *Output of the Two-sample t-test and of the Two-sided Wilcoxon rank sum test, applied to compare the distributions of the extracted features. The output is 1 if the compared distributions can be assumed to have different mean or median, 0 otherwise.*

	DS vs AV					**AV vs ME**				
Sensor nr.	*1*	*2*	*3*	*4*	*5*	*1*	*2*	*3*	*4*	*5*
Classic	1	1	1	1	1	1	0	1	1	0
Delta	1	1	1	1	1	0	0	1	1	0
Phase Integral	1	1	1	1	1	0	0	1	0	0
Relative Integral	1	1	1	1	1	1	1	1	1	1
Slope	1	1	0	1	1	0	0	0	0	0

ically, it tests the null hypothesis that the data comes from independent random samples from normal distributions with equal means and equal but unknown variances);

- the *Two-sided Wilcoxon rank sum test*, comparing the medians when populations are not normal (the null hypothesis, in this case, is that data are samples from continuous distributions with equal medians).

Again, the matrix containing the test decisions has output 1 if the test rejects the null hypothesis at the 5% significance level, and 0 otherwise. It is reported in Table 4.4.

4.3.3 Dimensionality Reduction

The successive step of the analysis has been the features projection in a lower dimensional space. There are two main reasons behind this operation:

1. none of the extracted features, taken alone, is sufficient to discriminate the three breath contributions exhaustively;

2. multivariate data are complex not only to be managed, but also to be visualized.

The technique adopted for dimensionality reduction in this project is Principal Component Analysis. PCA is a procedure using an orthogonal transformation to convert a set of observations of possibly correlated variables (in our case the extracted features) into a set of values of linearly uncorrelated variables called principal components. These principal components are nothing more than a linear combination of some of the original variables, and each of them accounts for as much of the variability in the data as possible. Thus, by selecting the

first few principal components that explain the most of the total variance, it is possible to look at data as if they are in a new simpler coordinate system.

First of all, PCA was applied on each single feature. The variances of the first principal components and their cumulative values (up to the 95%) are displayed in Figure 4.12.

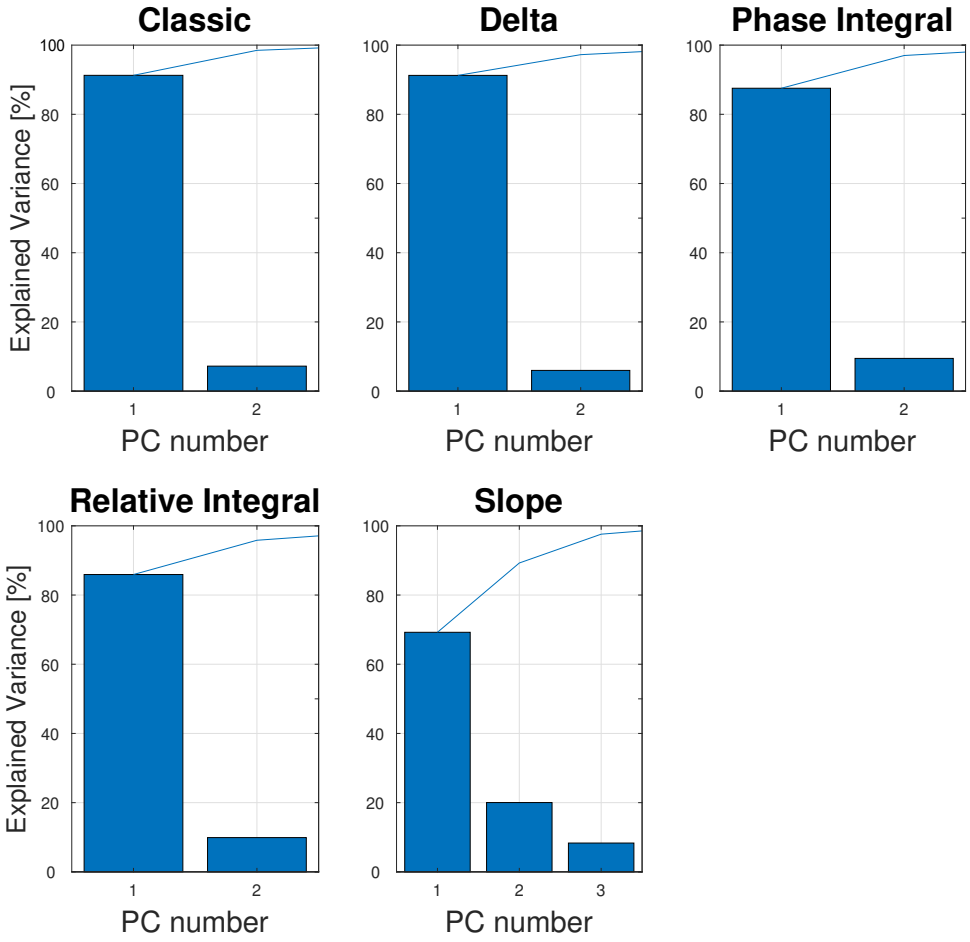


Figure 4.12: *Barplots indicating the variance explained by each of the principal components for every single feature. On the background, a blue line indicates the cumulative value of the variance.*

Since the first two components reach a total explained variance equal or higher than the 90% for all features, data have been projected in a two-dimensional space, achieving the distributions reported in Figure 4.13.

In the successive phase, PCA was applied on the whole set of features to further improve the analysis. In this case, the first two principal components together explain the 86% of the total variance, while the first three the 90% (Figure 4.14). Anyway, two components were sufficient to plot the three families of samples as separate clusters (Figure 4.15) and are preferable for visualization purposes.

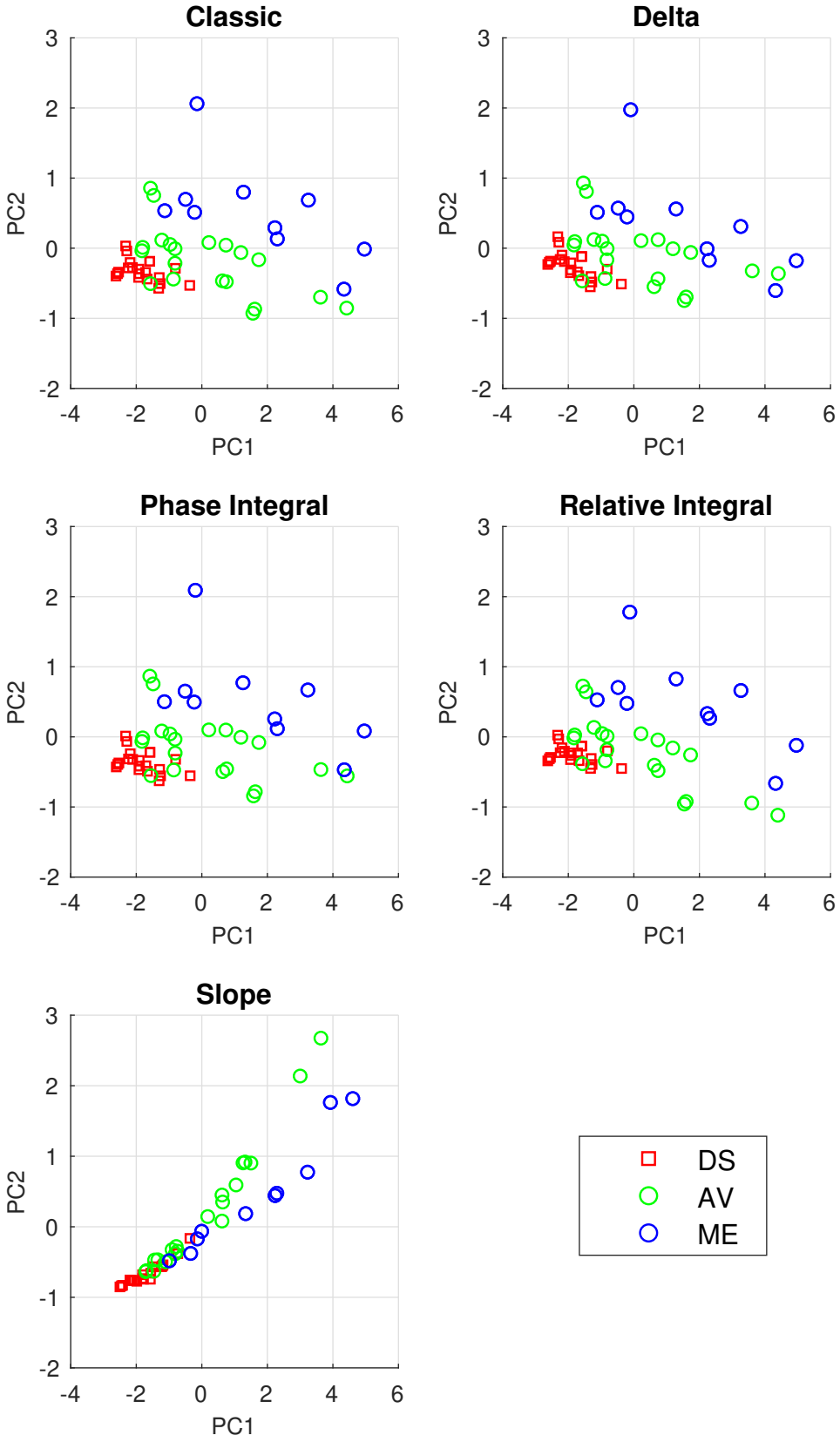


Figure 4.13: *Features projected singularly in the two-dimensional space defined by the first two principal components.*

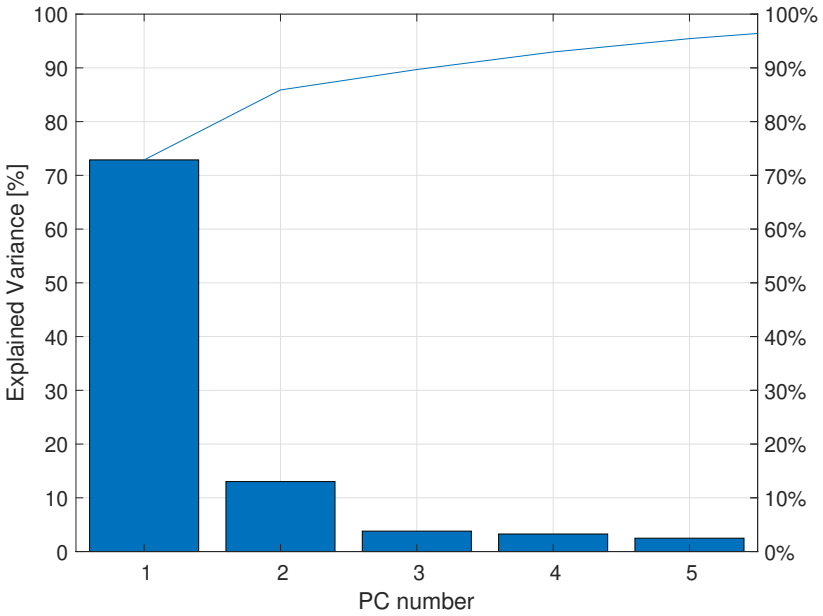


Figure 4.14: *Variance explained by the first principal components obtained using PCA on all features.*

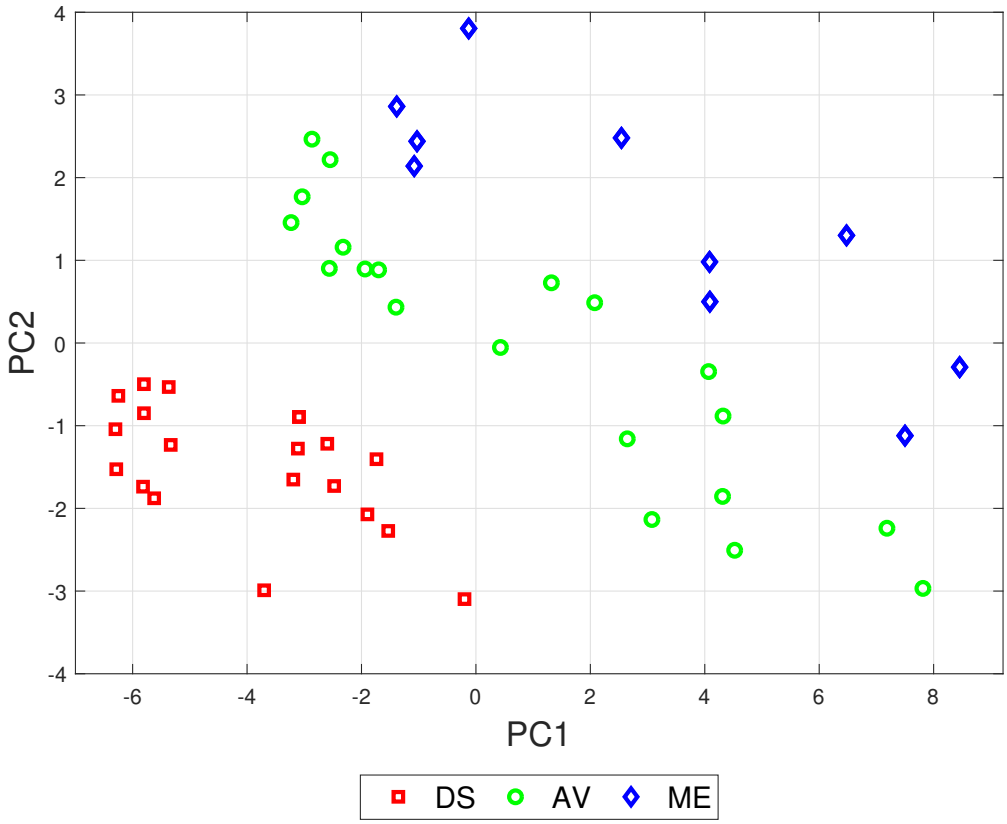


Figure 4.15: *Orthogonal projection of the breath samples in the two-dimensional space defined by the first two principal components, obtained by applying PCA to the whole set of extracted features.*

4.3.4 Discriminant Analysis

The last step of the analysis consists in the discrimination of the three breath contributions with a statistical tool, as a validation of the separation performed with the instrument. The choice fell on Discriminant Analysis, a supervised learning method used to find a linear combination of features that separates two or more classes of objects. DA is closely related to analysis of variance (ANOVA) and regression analysis, but it is based on continuous independent variables and a categorical dependent variable (i.e. the class label). For this reason, it appears ideal to complete this analysis without excessively increasing complexity. Depending on the complexity of the classifier to be developed, the boundary between classes can be linear, quadratic and so on. In this project, we attempted to separate the classes with Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA).

4.3.4.1 Linear Discriminant Analysis

LDA fits the linear model that best separates two or more classes based on the input variables (called predictors, in our case are the projected features) and their labels (the breath classes). The linear boundary between two classes is then reconstructed as:

$$K + \begin{bmatrix} x_1 & x_2 \end{bmatrix} L = 0.$$

where K and L are respectively the coefficient *Const* (the intercept) and the 2×1 vector *Linear* (the slope) of the fitted boundary and x_1, x_2 are the predictors coordinate relative to the axes of the orthogonal system defined with PCA. The boundaries to separate the three classes (i.e., first between DS and AV, then between AV and ME) have been reconstructed and plotted as in Figure 4.16. None of the samples has been misclassified.

4.3.4.2 Quadratic Discriminant Analysis

QDA attempts to classify data with quadratic boundaries. It works better than LDA in presence of non-linear distributions of predictors, at the cost of a higher complexity of the model. The boundaries, in this case, are reconstructed as:

$$K + \begin{bmatrix} x_1 & x_2 \end{bmatrix} L + \begin{bmatrix} x_1 & x_2 \end{bmatrix} Q \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0.$$

where Q is a 2×2 coefficients matrix weighting the quadratic combination of predictors. The result is shown in Figure 4.17. The two quadratic boundaries separate again the three

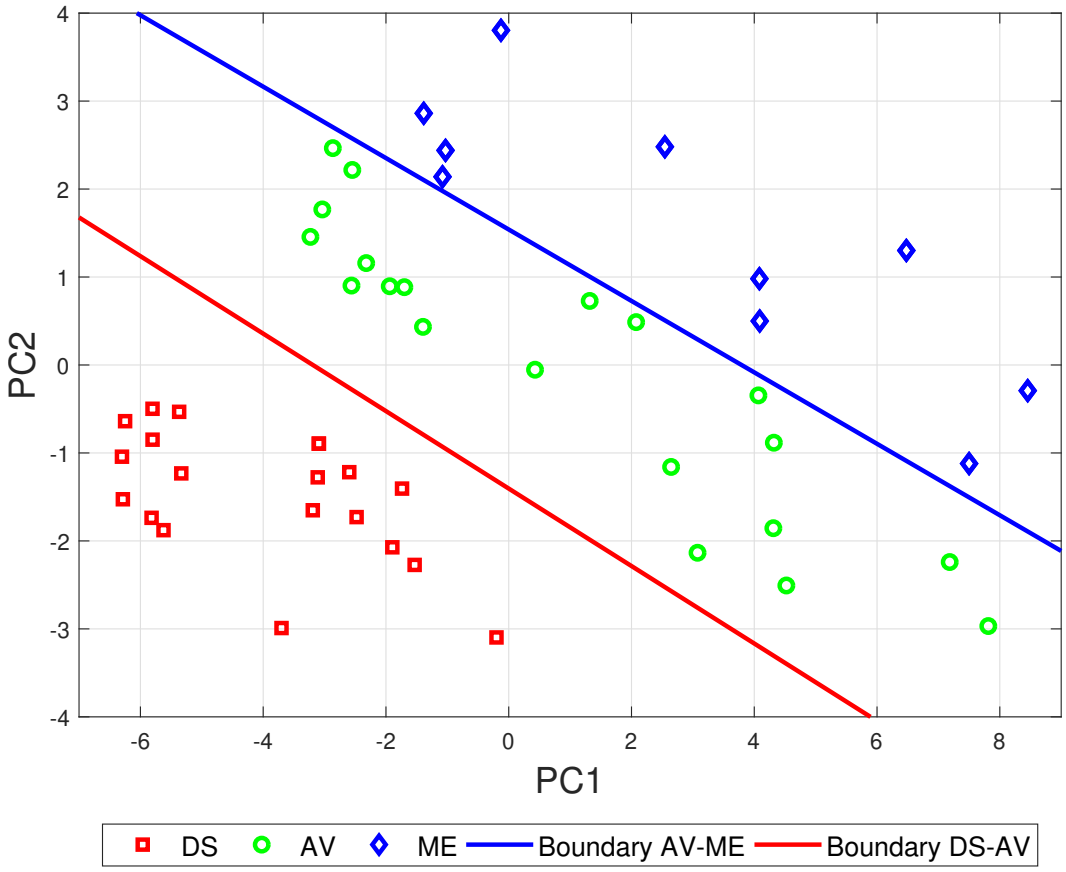


Figure 4.16: *Application of LDA to the data projected in the new orthogonal space determined with PCA. The two linear boundaries separate the three classes without any misclassification.*

classes without any misclassification. In this case, being data linearly separable, a quadratic classifier has no real advantage since it only increases the complexity of the model without any improvement in the quality of the analysis.

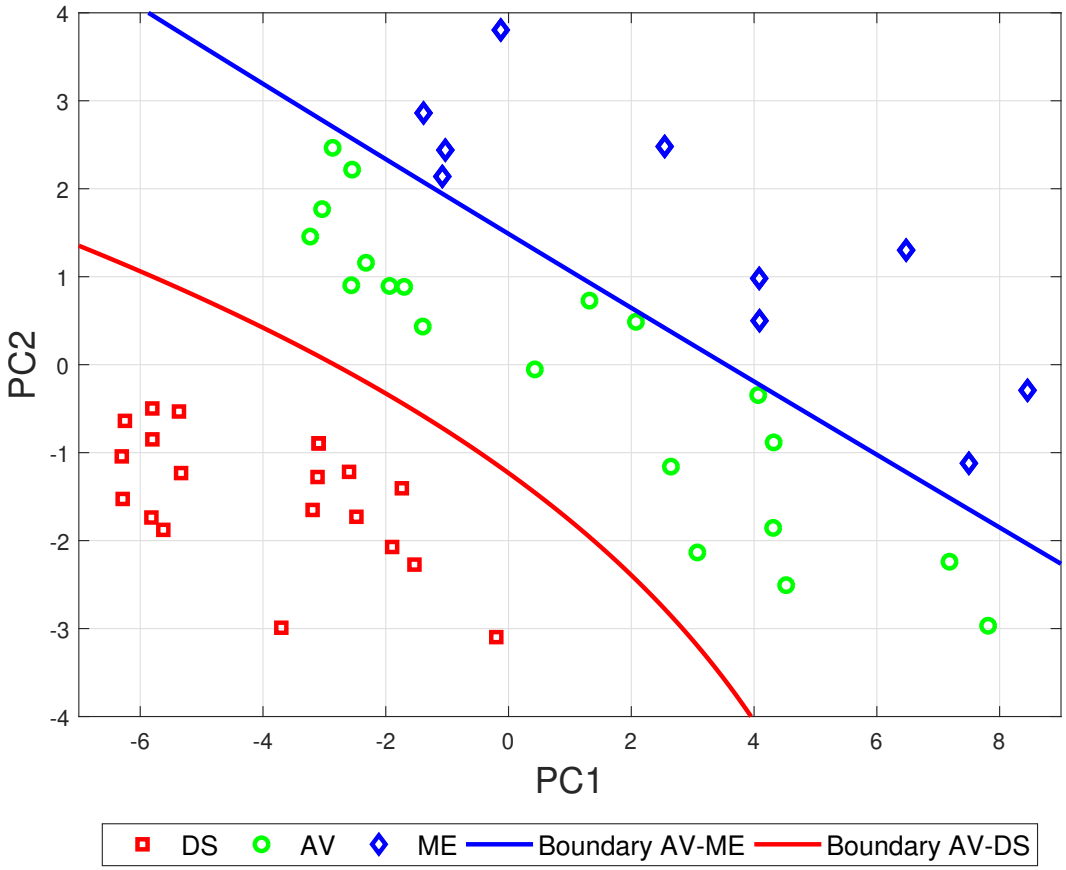


Figure 4.17: *Application of QDA to the data projected in the new orthogonal space determined with PCA. The two quadratic boundaries separate the three classes without any misclassification.*

Chapter 5

Conclusions

In this work, according to the thesis proposal (Section 1.2), the development of an automatic breath sampling device has been presented. It aims at the separation of the two components of exhaled breath, namely dead space and alveolar air, with the purpose to improve the analysis performed with electronic nose systems, usually based on the whole exhalation (i.e., without any separation).

The two main objectives explained in Section 1.3, i.e., the realization of a functioning prototype and of an easy-to-use software, have been reached. More specifically, the proposed solution is based on the acquisition of a single breath through maximal-forced expiration. It exploits a fast-responding CO_2 sensor (based on NDIR principle) to record the carbon dioxide concentration signal and a three-way solenoid valve to switch from one sampling bag (DS) to the other (AV), as schematized in Figure 5.1. A user-friendly software (a GUI written in Python) guides the user throughout the whole acquisition process and, if needed, allows to save data.

To validate the instrument, ten healthy volunteers have been enrolled for an experimental trial. Each of them had to perform two tests making use of the device, and one acquisition with no separation (mixed-expiratory breath). The sampled bags were then analyzed with an electronic nose prototype based on electrochemical gas sensors. The recorded signals have finally been processed in MATLAB to assess if the three classes of breath are separable. All results are reported in Chapter 4 and will be now discussed and compared with the outcomes we expected at the beginning of the project.

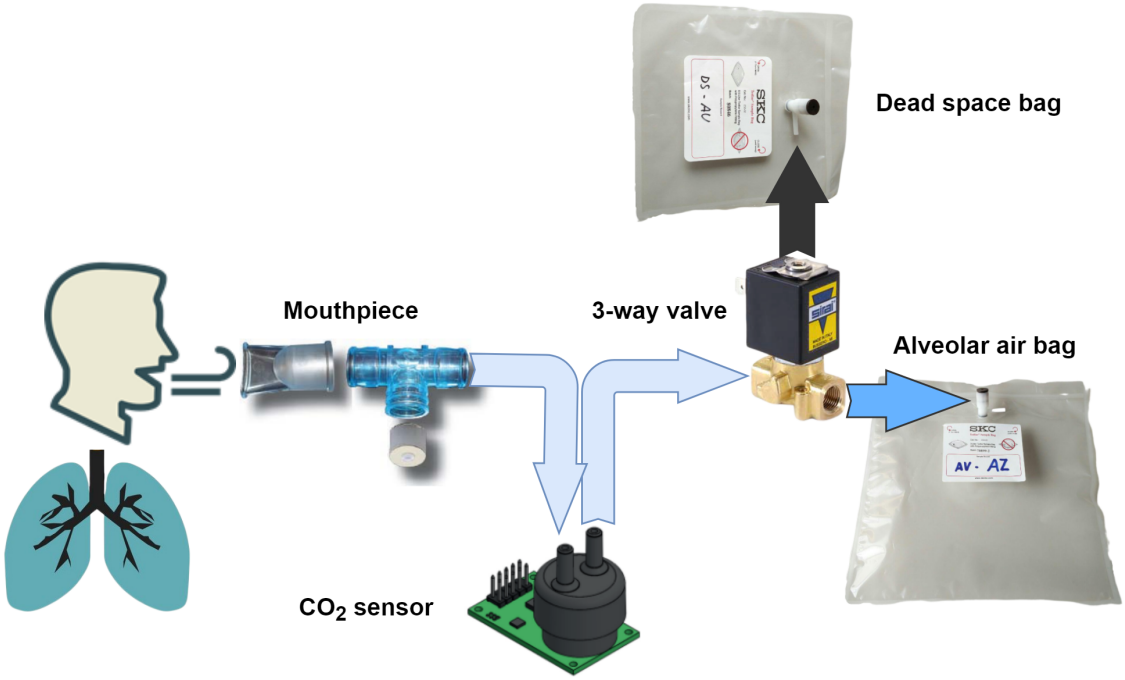


Figure 5.1: *Schematic representation of the device working principle.*

5.1 Discussion

As reported in the previous chapters, one of the expected results of the project was the development of a sampler of single breath and, if possible, of multiple breathings. In its initial stages, the prototype was designed with the aim to perform the second operation (a description of the implemented multi-breath algorithm is reported in Section 3.2.2). This approach was theoretically functioning, at least when the prototype was not completed yet, with the behavior shown in Figure 5.2.

Anyway, once the prototype was fully completed and the whole setup was tested, some key problems emerged: the hydraulic circuit, when the sampling bags are connected, is a closed loop, thus the inspiration causes the user to re-breath his exhaled air and the CO_2 sensor to saturate. Whilst the question of inspired air can be partially solved by transforming the dead space line into the inspiratory line (dead space collection is not essential), the issue related to sensor saturation is more difficult to handle and requires to re-design the device or to integrate new components. In fact, at the end of the exhalation, the user would still inspire the air sampled in the alveolar bag and the sensor will record high levels of CO_2,

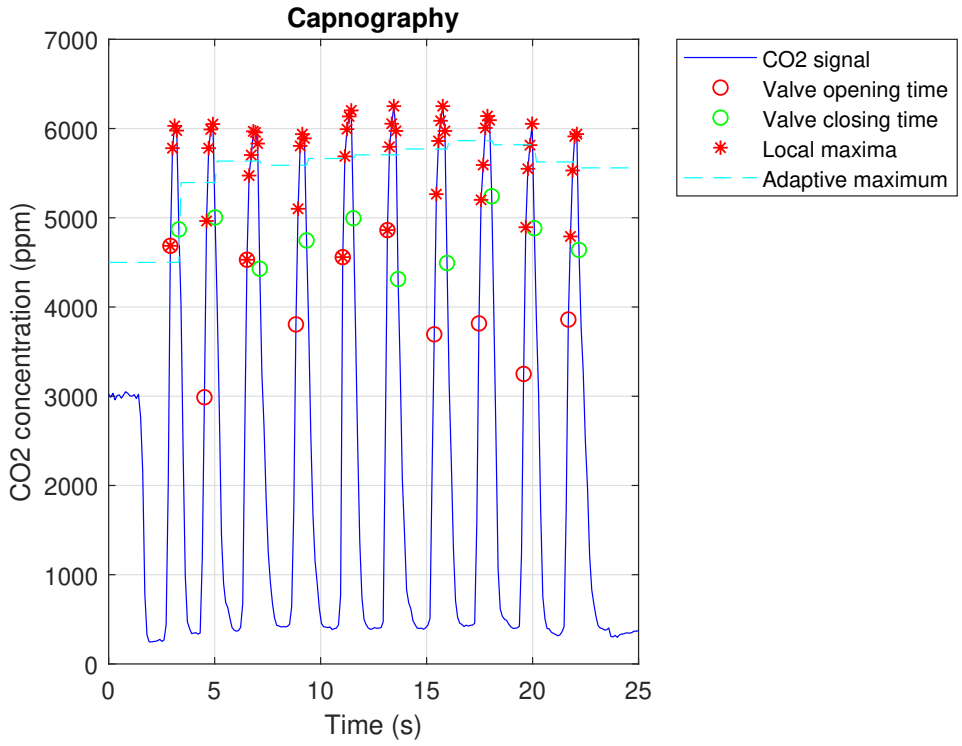


Figure 5.2: *Visual representation of the working principle of the multi-breath sampler. The algorithm, breath after breath, computes an adaptive threshold (based on local maxima) used to open and close the electrovalve.*

hence the algorithm would not cause the valve to switch. The easiest solution, at this point, was to simplify the system to a single-breath sampler and to replicate a procedure similar to the one manually performed at the IEO [35], but automatic.

The results relative to the single-breath sampler are more satisfactory and will be now commented. The acquisition protocol, which is similar to that of spirometry, is quite simple but has the disadvantage to be highly dependent on patient cooperation and effort. For this reason, it is not fully reproducible, therefore two measures have been taken for each volunteer. Furthermore, because of the impedance of the hydraulic circuit, some of the volunteers reported that the test has been quite challenging for them . This is probably due to the very narrow diameter of the valve opening and of the tubes, added to the reduced dimension of the sensor cap. It represents a limitation for this system, since the effort will result even higher for patients with lung cancer or other respiratory pathologies.

The waveform recorded with the electronic nose prototype is quite similar to those found in the literature, such as in the studies of Di Natale *et al.* [14], Blatt *et al.* [38] and

Li *et al.* [33]. We have seen it has the evident problem of baseline variability measure after measurement (Figure 4.4), which raised the necessity to perform the baseline correction at least to visualize better the difference between the signals relative to the three different classes. The distributions of the R_0 relative to the recorded signals are reported in Figure 4.5 for each of the five sensors. Even if the interquartile ranges appear to be small, this is a primary issue to be managed in the future versions of the prototype. Such a problem, in fact, also limits the possibility to assess the repeatability of the collected samples, which was one of the goals prefixed at the beginning of the work.

The next operation instead, signals filtering, is not really necessary (the noise is barely visible if we look at the full dynamic of the signal, e.g., in Figure 4.6). Because of this, such a operation was performed in the simplest way as possible, obtaining an acceptable result.

About the features, some of them are dependent on R_0 and can be partially corrupted by its variability. In particular, in Figure 4.10 we could see that the trajectory in the phase space is not a perfectly closed curve because of the difference between the initial and the final values of R_0. Their statistical analysis showed that not all distributions are normal (Figure 4.11 and Table 4.3), but this is reasonable given the number of acquisitions. Furthermore, the data acquired with the device appears to generate a greater quantity of normal distributions than those collected using the mouthpiece only, and this is an encouraging result about the reproducibility of the acquisition process. The output of the Two-sample t-test and of the Two-sided Wilcoxon rank sum test (Table 4.4), instead, demonstrates that dead space and alveolar air can be considered significantly different according to all features (except for the *Slope* relative to sensor TGS822), while only the *Relative Integral* demonstrates the separability between alveolar and mixed-expiratory samples. Considering that one of the expected results is to find a significant difference between the three types of breath samples, there are two considerations to be done:

a) the statistical analysis of the single features cannot be considered to be fully discriminant, hence further processing is requested;

b) a greater number of acquisitions is necessary to have a better validation of each feature, but the one that appears to need a reformulation is the *Slope*.

The successive step, that is dimensionality reduction, is very helpful both to enhance the quality of the analysis (it somehow eliminates the elements that explain less variability in the data) and to better visualize data in a single plot. As shown in Figure 4.12, the first two

components explain more than the 90% of variance for all features except *Slope*, confirming the poor effectiveness of this parameter previously found with the t-test and the rank sum test. Also, the projection of each sample in the two-dimensional spaces (Figure 4.13) confirms that the 3 classes are not exhaustively separable with the single features, but three regions of discrimination are beginning to be outlined (apart from *Slope*, again). The application of PCA to all five features, instead, results to be much more effective. Even if the first two principal components explain together the 86% of the total variance (Figure 4.14), which is less then the 90% achieved with the single features, the projection in the bidimensional space shows now three distinct regions of samples (Figure 4.15). The three classes appears to be separable, but a final step is necessary to prove it.

The application of discriminant analysis allowed to fit both a linear (LDA) and a quadratic classification model (QDA) based on the predictors resulting from PCA. The classification boundaries obtained with the two models are represented respectively in Figure 4.16 and 4.17. Both of them successfully separate the dead space from the alveolar air samples, and the latter from the mixed-expiratory samples too, with no misclassification. In particular, the separation between DS and AV samples is clearer than that between AV and ME samples (it is quantified by the distance of each point from the boundary). As already stated, at least with the data acquired in this project, a quadratic model is unnecessary since it only increases the complexity with no real advantage. However, in case of availability of a larger amount of data, quadratic boundaries and/or the inclusion of a third principal component could be necessary.

These results are not a real validation of the developed device, since data are not compared with other studies nor analyzed with any other professional instrument. Anyway we have demonstrated that this type of automatic sampling has generated, at least in a small group of people, two air mixtures having different compositions in term of VOCs both between them and also when compared to an air sample taken without separation (which is the standard to improve).

In judging these results and the overall work, there are several limitations to be taken into account:

- the Tedlar bags used during the acquisitions were not at their first use and have not been cleaned in between an usage and the other, since we had no availability of the necessary instruments to do it. Even if they are made by chemically inert material, it is possible that some of the VOCs remained in the bags after each sampling and

conditioned the successive analysis;

- the mouthpiece assembly was not substituted with a new one after each test (as happening during clinical trials), but it was only rinsed with water and dishwashing soap. This could have some influence on the VOCs mixture sampled in each bag, too;
- no VOCs filter was used during the inspiration, thus the inspired ambient air could partially affect the exhaled breath and the successive analysis;
- the population involved in the study is too small to induce a normality study. Furthermore, all subjects are healthy, not-smokers and are all about the same age, thus it is a too limited sample to draw general conclusions;
- the instrumental dead space (i.e., the volume of the internal tubes, the valve and the sensor cap) has not been quantified, but it was only compensated with a delay in the switching time of the valve;
- the lack of an internal pump makes the cleaning process too slow and comfortless.

5.2 Future Improvements

We have seen that some of the project's expectations have been met, while some others have not. In particular, we reached the primary goals to build a device able to sample a single breath and to demonstrate that the separate contributions have significant differences in the sensors response. On the other hand, we failed in the attempt to separate multiple breathings and we were limited in testing the repeatability of the sensors response due to some limitations of the electronic nose.

Based on these evaluations, the following possible improvements have been identified:

1. **Addition of a flow measurement system:** the problem emerged in the multi-breath approach can be probably solved by adding a flow measurement system upstream of the CO_2 sensor. This could identify the inspiratory and expiratory phases based on the flow direction and open an inspiratory channel based on this information, rather than on time capnogram. Also, a flow rate measure can be somehow used to infer the volume of air entering the device: this additional parameter could be helpful to compensate with higher precision the question of the instrumental dead space, too. The ideal solution (there are several possibilities, spanning from flow meters, differential pressure sensors,

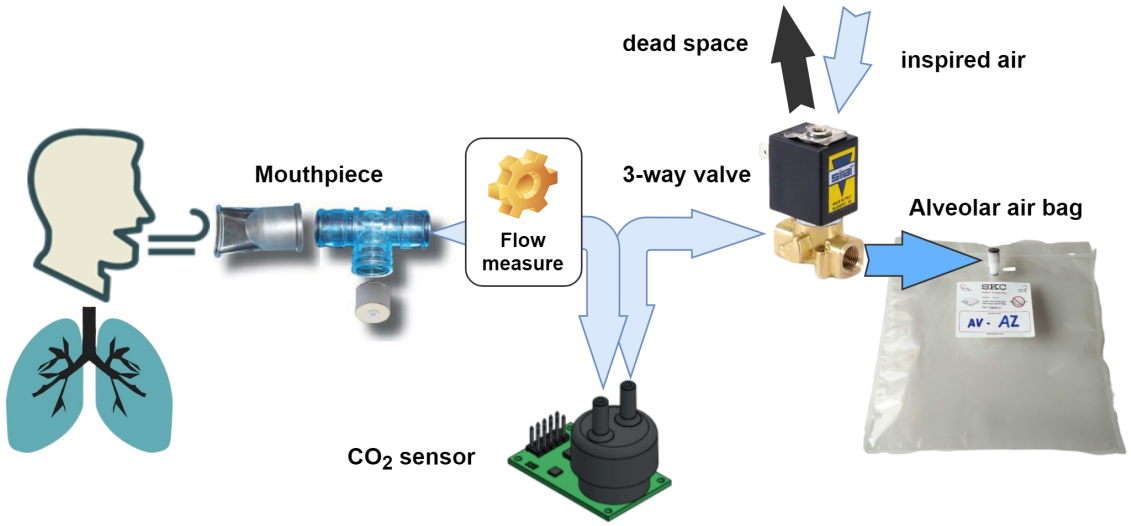


Figure 5.3: *Schematic representation of the system with the proposed modification.*

hot wire anemometers, etc.) should be compatible with the reduced dimension of the device and may not influence the acquisition procedure. A schematic representation of the modified device is reported in Figure 5.3.

2. **Insertion of a pump:** as an alternative to the solution proposed at point 1, a micro-pump can be introduced in the system to automate the Cleaning phase. The system would still be a single-breath sampler, but the procedure would become easier and faster.

3. **Validation on lung cancer patients:** a larger population study is necessary to really test the effectiveness of this kind of breath separation. Moreover, the inclusion of lung cancer patients would ascertain if the VOCs profile relative to alveolar air is different than that relative to mixed-expiratory breath also in the pathological case.

4. **Reduction of the hydraulic circuit impedance:** as previously stated, the test resulted challenging for some of the volunteers because of the high impedance of the hydraulic circuit. Since it could be possible that patients with compromised respiratory functions undergo breath tests, the system could be re-designed to have a lower impedance and to make the expiration more comfortable for the user. This can be achieved by widening the diameter of the conducting pipes, at the price of increasing the instrumental dead space.

5. **Definition of more effective features:** the features extracted in this work were sufficient to successfully classify our samples, but cannot be enough in presence of a larger dataset. The introduction of new parameters, or the improvement of those already defined, can be useful to make the whole classification process more robust.

Bibliography

[1] World Health Organization. *Cancer - Key facts*, 2018 (accessed March 2, 2019).

[2] Peter Goldstraw, Kari Chansky, John Crowley, Ramon Rami-porta, Hisao Asamura, Wilfried E E Eberhardt, Andrew G Nicholson, Patti Groome, and Alan Mitchell. The IASLC Lung Cancer Staging Project : Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classi fi cation for Lung Cancer. 11(1):39–51, 2015.

[3] European Institute of Oncology (IEO). *Lung Cancer - Prevention and Diagnosis*, 2018 (accessed March 4, 2019).

[4] American Cancer Society. *small-cell lung cancer/detection,diagnosis,staging*, 2018 (accessed March 4, 2019).

[5] Roberto Gasparri, Rosalia Romano, Giulia Sedda, Alessandro Borri, Francesco Petrella, Domenico Galetta, Monica Casiraghi, and Lorenzo Spaggiari. Diagnostic biomarkers for lung cancer prevention Diagnostic biomarkers for lung cancer prevention. *Journal of Breath Research*, 12(2):27111, 2018.

[6] Denise R. Aberle, Amanda M. Adams, Christine D. Berg, William C. Black, Jonathan D. Clapp, Richard M. Fagerstrom, Ilana F. Gareen, Constantine Gatsonin, Pamela M. Marcus, and JoRean D. Sicks. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *Yearbook of Pulmonary Disease*, 2012(5):78–80, aug 2012.

[7] Cristiano Rampinelli, Daniela Origgi, and Massimo Bellomi. Low-dose CT : technique , reading methods and image interpretation. 12:548–556, 2012.

[8] Michael Phillips. Breath tests in medicine. *Scientific American*, 267(1):74–79, 1992.

[9] Michael McCulloch, Tadeusz Jezierski, Michael Broffman, Alan Hubbard, Kirk Turner, and Teresa Janecki. Diagnostic accuracy of canine scent detection in early- and late-stage lung and breast cancers. *Integrative Cancer Therapies*, 5(1):30–39, 2006. PMID: 16484712.

[10] Inbar Nardi-Agmon and Nir Peled. Exhaled breath analysis for the early detection of lung cancer: recent developments and future prospects. *Lung Cancer (Auckland, N.Z.)*, 8:31–38, 2017.

[11] Andras Bikov, Marton Hernadi, Beata Zita Korosi, Laszlo Kunos, Gabriella Zsamboki, Zoltan Sutto, Adam Domonkos Tarnoki, David Laszlo Tarnoki, Gyorgy Losonczy, and Ildiko Horvath. Expiratory flow rate , breath hold and anatomic dead space influence electronic nose ability to detect lung cancer. pages 1–9, 2014.

[12] Ildiko Horváth, Peter J Barnes, Stelios Loukides, Peter J Sterk, Marieann Högman, Anna-Carin Olin, Anton Amann, Balazs Antus, Eugenio Baraldi, Andras Bikov, et al. A european respiratory society technical standard: exhaled biomarkers in lung disease. *European Respiratory Journal*, 49(4):1600965, 2017.

[13] Kristin Schallschmidt, Roland Becker, Christian Jung, Wolfram Bremser, Irene Nehls, G. Leschber, S. Frese, T. Walles, and J. Neudecker. Comparison of volatile organic compounds from lung cancer patients and healthy controls - challenges and limitations of an observational study. *Journal of Breath Research*, 10(4):Article 046007, 1 – 17, 2016.

[14] Corrado Di Natale, Antonella Macagnano, Eugenio Martinelli, Roberto Paolesse, Giuseppe D'Arcangelo, Claudio Roscioni, Alessandro Finazzi-Agrò, and Arnaldo D'Amico. Lung cancer identification by the analysis of breath by means of an array of non-selective gas sensors. *Biosensors and Bioelectronics*, 18(10):1209–1218, sep 2003.

[15] Davide Marzorati, Luca Mainardi, Giulia Sedda, Roberto Gasparri, Lorenzo Spaggiari, and Pietro Cerveri. A review of exhaled breath: a key role in lung cancer diagnosis. *Journal of breath research*, 13(3):034001, 2019.

[16] N. J. Rattray, Z. Hamrang, D. K. Trivedi, R. Goodacre, and S. J. Fowler. Taking your breath away: metabolomics breathes life in to personalized medicine. *Trends in Biotechnology*, 32:538–548, 2014.

[17] Oluwasola Lawal, Waqar M Ahmed, Tamara ME Nijsen, Royston Goodacre, and Stephen J Fowler. Exhaled breath analysis: a review of 'breath-taking' methods for off-line analysis. *Metabolomics*, 13(10):110, 2017.

[18] Sal Intagliata, Alessandra Rizzo, and Steve S Bhimji. Physiology, lung dead space. In *StatPearls [Internet]*. StatPearls Publishing, 2019.

[19] Mar Castellanos, Gemma Xifra, José Manuel Fernández-Real, and Juan M Sánchez. Breath gas concentrations mirror exposure to sevoflurane and isopropyl alcohol in hospital environments in non-occupational conditions. *Journal of Breath Research*, 10(1):016001, jan 2016.

[20] Juan M. Sanchez and Richard D. Sacks. Development of a multibed sorption trap, comprehensive two-dimensional gas chromatography, and time-of-flight mass spectrometry system for the analysis of volatile organic compounds in human breath. *Analytical Chemistry*, 78(9):3046–3054, 2006. PMID: 16642992.

[21] Wolfram Miekisch, Sabine Kischkel, Annika Sawacki, Tina Liebau, Maren Mieth, and Jochen K Schubert. Impact of sampling procedures on the results of breath analysis. 026007.

[22] K. A. Cope. Effects of ventilation on the collection of exhaled breath in humans. *Journal of Applied Physiology*, 96(4):1371–1379, 2004.

[23] Wolfram Miekisch, Andreas Hengstenberg, Sabine Kischkel, Udo Beckmann, and Maren Mieth. Construction and Evaluation of a Versatile CO 2 Controlled Breath Collection Device. 10(1):211–215, 2010.

[24] P. Salvo, C. Ferrari, R. Persia, S. Ghimenti, T. Lomonaco, F. Bellagambi, and F. Di Francesco. A dual mode breath sampler for the collection of the end-tidal and dead space fractions. *Medical Engineering and Physics*, 37(6):539–544, 2015.

[25] Gang Peng, M Hakim, YY Broza, S Billan, R Abdah-Bortnyak, A Kuten, U Tisch, and H Haick. Detection of lung, breast, colorectal, and prostate cancers from exhaled breath using a single array of nanosensors. *British journal of cancer*, 103(4):542, 2010.

[26] Roberto F Machado, Daniel Laskowski, Olivia Deffenderfer, Timothy Burch, Shuo Zheng, Peter J Mazzone, Tarek Mekhail, Constance Jennings, James K Stoller, Jacqueline Pyle,

et al. Detection of lung cancer by sensor array analyses of exhaled breath. *American journal of respiratory and critical care medicine*, 171(11):1286–1291, 2005.

[27] Christopher Phillips, Neil Mac Parthaláin, Yasir Syed, Davide Deganello, Timothy Claypole, and Keir Lewis. Short-term intra-subject variation in exhaled volatile organic compounds (vocs) in copd patients and healthy controls and its effect on disease classification. *Metabolites*, 4(2):300–318, 2014.

[28] Mrinal Kumar Das, Subasa Chandra Bishwal, Aleena Das, Deepti Dabral, Ankur Varshney, Vinod Kumar Badireddy, and Ranjan Nanda. Investigation of gender-specific exhaled breath volatome in humans by gcxgc-tof-ms. *Analytical chemistry*, 86(2):1229–1237, 2013.

[29] Michael Phillips, Renee N Cataneo, Jose Alfonso Cruz-Ramos, Jan Huston, Omar Ornelas, Nadine Pappas, and Sonali Pathak. Prediction of breast cancer risk with volatile biomarkers in breath. *Breast cancer research and treatment*, pages 1–8, 2018.

[30] Sean W Harshman, Brian A Geier, Maomian Fan, Sage Rinehardt, Brandy S Watts, Leslie A Drummond, George Preti, Jeffrey B Phillips, Darrin K Ott, and Claude C Grigsby. The identification of hypoxia biomarkers from exhaled breath under normobaric conditions. *Journal of breath research*, 9(4):047103, 2015.

[31] Jochen K Schubert, Karl-Heinz Spittler, Guenther Braun, Klaus Geiger, and Josef Guttmann. Co2-controlled sampling of alveolar gas in mechanically ventilated patients. *Journal of Applied Physiology*, 90(2):486–492, 2001.

[32] Amalia Z Berna, James S McCarthy, Rosalind X Wang, Kevin J Saliba, Florence G Bravo, Julie Cassells, Benjamin Padovan, and Stephen C Trowell. Analysis of breath specimens for biomarkers of plasmodium falciparum infection. *The Journal of infectious diseases*, 212(7):1120–1128, 2015.

[33] Wang Li, Hongying Liu, Dandan Xie, Zichun He, and Xititan Pi. Lung Cancer Screening Based on Type-different Sensor Arrays. *Scientific Reports*, 7(1):1–12, 2017.

[34] Maria Basanta, Roger M Jarvis, Yun Xu, Gavin Blackburn, Ruth Tal-Singer, Ashley Woodcock, Dave Singh, Royston Goodacre, CL Paul Thomas, and Stephen J Fowler. Non-invasive metabolomic analysis of breath using differential mobility spectrometry

in patients with chronic obstructive pulmonary disease and healthy smokers. *Analyst*, 135(2):315–320, 2010.

[35] Roberto Gasparri, Marco Santonico, Claudia Valentini, Giulia Sedda, Alessandro Borri, Francesco Petrella, Patrick Maisonneuve, Giorgio Pennazza, Arnaldo D'Amico, Corrado Di Natale, et al. Volatile signature for the early diagnosis of lung cancer. *Journal of breath research*, 10(1):016007, 2016.

[36] Rainer Ehmann, E Boedeker, U Friedrich, J Sagert, J Dippon, G Friedel, and T Walles. Canine scent detection in the diagnosis of lung cancer: revisiting a puzzling phenomenon. *European respiratory journal*, 39(3):669–676, 2012.

[37] Rens van de Goor, Michel van Hooren, Anne-Marie Dingemans, Bernd Kremer, and Kenneth Kross. Training and validating a portable electronic nose for lung cancer screening. *Journal of Thoracic Oncology*, 13(5):676–681, 2018.

[38] Rossella Blatt, Andrea Bonarini, Elisa Calabro, Matteo Della Torre, Matteo Matteucci, and Ugo Pastorino. Lung cancer identification by an electronic nose based on an array of mos sensors. In *2007 International Joint Conference on Neural Networks*, pages 1423–1428. IEEE, 2007.

[39] Hiroshi Handa, Ayano Usuba, Sasidhar Maddula, Jörg Ingo Baumbach, Masamichi Mineshita, and Teruomi Miyazawa. Exhaled breath analysis for lung cancer detection using ion mobility spectrometry. *PloS one*, 9(12):e114555, 2014.

[40] Morad K Nakhleh, Haitham Amal, Raneen Jeries, Yoav Y Broza, Manal Aboud, Alaa Gharra, Hodaya Ivgi, Salam Khatib, Shifaa Badarneh, Lior Har-Shai, et al. Diagnosis and classification of 17 diseases from 1404 subjects via pattern analysis of exhaled molecules. *ACS nano*, 11(1):112–125, 2016.

[41] Lu Kou, David Zhang, and Dongxu Liu. A novel medical e-nose signal analysis system. *Sensors*, 17(4):402, 2017.

[42] SM Gordon, JP Szidon, BK Krotoszynski, RD Gibbons, and HJ O'Neill. Volatile organic compounds in exhaled air from patients with lung cancer. *Clinical chemistry*, 31(8):1278–1282, 1985.

[43] Michael Phillips, Kevin Gleeson, J Michael B Hughes, Joel Greenberg, Renee N Cataneo, Leigh Baker, and W Patrick McVay. Volatile organic compounds in breath as markers of lung cancer: a cross-sectional study. *The Lancet*, 353(9168):1930–1933, 1999.

[44] Michael Phillips, Renee N Cataneo, Andrew RC Cummin, Anthony J Gagliardi, Kevin Gleeson, Joel Greenberg, Roger A Maxfield, and William N Rom. Detection of lung cancer with volatile markers in the breath. *Chest*, 123(6):2115–2123, 2003.

[45] Michael Phillips, Nasser Altorki, John HM Austin, Robert B Cameron, Renee N Cataneo, Joel Greenberg, Robert Kloss, Roger A Maxfield, Muhammad I Munawar, Harvey I Pass, et al. Prediction of lung cancer using volatile biomarkers in breath 1. *Cancer biomarkers*, 3(2):95–109, 2007.

[46] Michael Phillips, Nasser Altorki, John HM Austin, Robert B Cameron, Renee N Cataneo, Robert Kloss, Roger A Maxfield, Muhammad I Munawar, Harvey I Pass, Asif Rashid, et al. Detection of lung cancer using weighted digital analysis of breath biomarkers. *Clinica chimica acta*, 393(2):76–84, 2008.

[47] Amel Bajtarevic, Clemens Ager, Martin Pienz, Martin Klieber, Konrad Schwarz, Magdalena Ligor, Tomasz Ligor, Wojciech Filipiak, Hubert Denz, Michael Fiegl, et al. Noninvasive detection of lung cancer by analysis of exhaled breath. *BMC cancer*, 9(1):348, 2009.

[48] Magdalena Ligor, Tomasz Ligor, Amel Bajtarevic, Clemens Ager, Martin Pienz, Martin Klieber, Hubert Denz, Michael Fiegl, Wolfgang Hilbe, Wolfgang Weiss, et al. Determination of volatile organic compounds in exhaled breath of patients with lung cancer using solid phase microextraction and gas chromatography mass spectrometry. *Clinical chemistry and laboratory medicine*, 47(5):550–560, 2009.

[49] Diana Poli, Paolo Carbognani, Massimo Corradi, Matteo Goldoni, Olga Acampa, Bruno Balbi, Luca Bianchi, Michele Rusca, and Antonio Mutti. Exhaled volatile organic compounds in patients with non-small cell lung cancer: cross sectional and nested short-term follow-up study. *Respiratory research*, 6(1):71, 2005.

[50] Diana Poli, Matteo Goldoni, Massimo Corradi, Olga Acampa, Paolo Carbognani, Eveline Internullo, Angelo Casalini, and Antonio Mutti. Determination of aldehydes in exhaled breath of patients with lung cancer by means of on-fiber-derivatisation spme–gc/ms. *Journal of Chromatography B*, 878(27):2643–2651, 2010.

[51] Patricia Fuchs, Christian Loeseken, Jochen K Schubert, and Wolfram Miekisch. Breath gas aldehydes as biomarkers of lung cancer. *International Journal of Cancer*, 126(11):2663–2670, 2010.

[52] Sabine Kischkel, Wolfram Miekisch, Annika Sawacki, Eva M Straker, Phillip Trefz, Anton Amann, and Jochen K Schubert. Breath biomarkers for lung cancer detection and assessment of smoking related effects - confounding variables, influence of normalization and statistical algorithms. *Clinica Chimica Acta*, 411(21-22):1637–1644, 2010.

[53] Nuria Queralto, Anders N Berliner, Brett Goldsmith, Raymond Martino, Paul Rhodes, and Sung H Lim. Detecting cancer by breath volatile organic compound analysis: a review of array-based sensors. *Journal of breath research*, 8(2):027112, 2014.

[54] Annette G. Dent, Tom G. Sutedja, and Paul V. Zimmerman. Exhaled breath analysis for lung cancer. *Journal of Thoracic Disease*, 5(SUPPL.5), 2013.

[55] Gopi Krishna Kafle, Lav R Khot, Sindhuja Sankaran, Haitham Y Bahlol, Jessica A Tufariello, and Herbert H Hill Jr. State of ion mobility spectrometry and applications in agriculture: A review. *Engineering in Agriculture, Environment and Food*, 9(4):346–357, 2016.

[56] Michael Westhoff, Patric Litterst, Lutz Freitag, Wolfgang Urfer, Sabine Bader, and Joerg Ingo Baumbach. Ion mobility spectrometry for the detection of volatile organic compounds in exhaled breath of patients with lung cancer: results of a pilot study. *Thorax*, 64(9):744–748, 2009.

[57] DS Ballantine, RM White, SJ Martin, AJ Ricco, ET Zellers, GC Frye, and H Wohjlten. Acoustic wave sensors academic press. *San Diego, CA, USA*, 1997.

[58] Corrado Di Natale, Antonella Macagnano, Eugenio Martinelli, Roberto Paolesse, Giuseppe D'Arcangelo, Claudio Roscioni, Alessandro Finazzi-Agro, and Arnaldo D'Amico. Lung cancer identification by the analysis of breath by means of an array of non-selective gas sensors. *Biosensors and Bioelectronics*, 18(10):1209–1218, 2003.

[59] Arnaldo D'Amico, Giorgio Pennazza, Marco Santonico, Eugenio Martinelli, Claudio Roscioni, Giovanni Galluccio, Roberto Paolesse, and Corrado Di Natale. An investigation on electronic nose diagnosis of lung cancer. *Lung cancer*, 68(2):170–176, 2010.

[60] Silvano Dragonieri, Jouke T Annema, Robert Schot, Marc PC van der Schee, Antonio Spanevello, Pierluigi Carratú, Onofrio Resta, Klaus F Rabe, and Peter J Sterk. An electronic nose in the discrimination of patients with non-small cell lung cancer and copd. *Lung cancer*, 64(2):166–170, 2009.

[61] Madara Tirzīte, Māris Bukovskis, Gunta Strazda, Normunds Jurka, and Immanuels Taivans. Detection of lung cancer in exhaled breath with an electronic nose using support vector machine analysis. *Journal of breath research*, 11(3):036009, 2017.

[62] Gang Peng, Ulrike Tisch, Orna Adams, Meggie Hakim, Nisrean Shehada, Yoav Y Broza, Salem Billan, Roxolyana Abdah-Bortnyak, Abraham Kuten, and Hossam Haick. Diagnosing lung cancer in exhaled breath using gold nanoparticles. *Nature nanotechnology*, 4(10):669, 2009.

[63] Nir Peled, Meggie Hakim, Paul A Bunn Jr, York E Miller, Timothy C Kennedy, Jane Mattei, John D Mitchell, Fred R Hirsch, and Hossam Haick. Non-invasive breath analysis of pulmonary nodules. *Journal of Thoracic Oncology*, 7(10):1528–1533, 2012.

[64] Peter J Mazzone, Jeffrey Hammel, Raed Dweik, Jie Na, Carmen Czich, Daniel Laskowski, and Tarek Mekhail. Diagnosis of lung cancer by the analysis of exhaled breath with a colorimetric sensor array. *Thorax*, 62(7):565–568, 2007.

[65] Peter J Mazzone, Xiao-Feng Wang, Yaomin Xu, Tarek Mekhail, Mary C Beukemann, Jie Na, Jonathan W Kemling, Kenneth S Suslick, and Madhu Sasidhar. Exhaled breath analysis with a colorimetric sensor array for the identification and characterization of lung cancer. *Journal of Thoracic Oncology*, 7(1):137–142, 2012.

[66] Peter J Mazzone, Xiao-Feng Wang, Sung Lim, James Jett, Humberto Choi, Qi Zhang, Mary Beukemann, Meredith Seeley, Ray Martino, and Paul Rhodes. Progress in the development of volatile exhaled breath signatures of lung cancer. *Annals of the American Thoracic Society*, 12(5):752–757, 2015.

[67] Eugenio Martinelli, Christian Falconi, Arnaldo D'Amico, and Corrado Di Natale. Feature extraction of chemical sensors in phase space. *Sensors and Actuators B: Chemical*, 95(1-3):132–139, 2003.