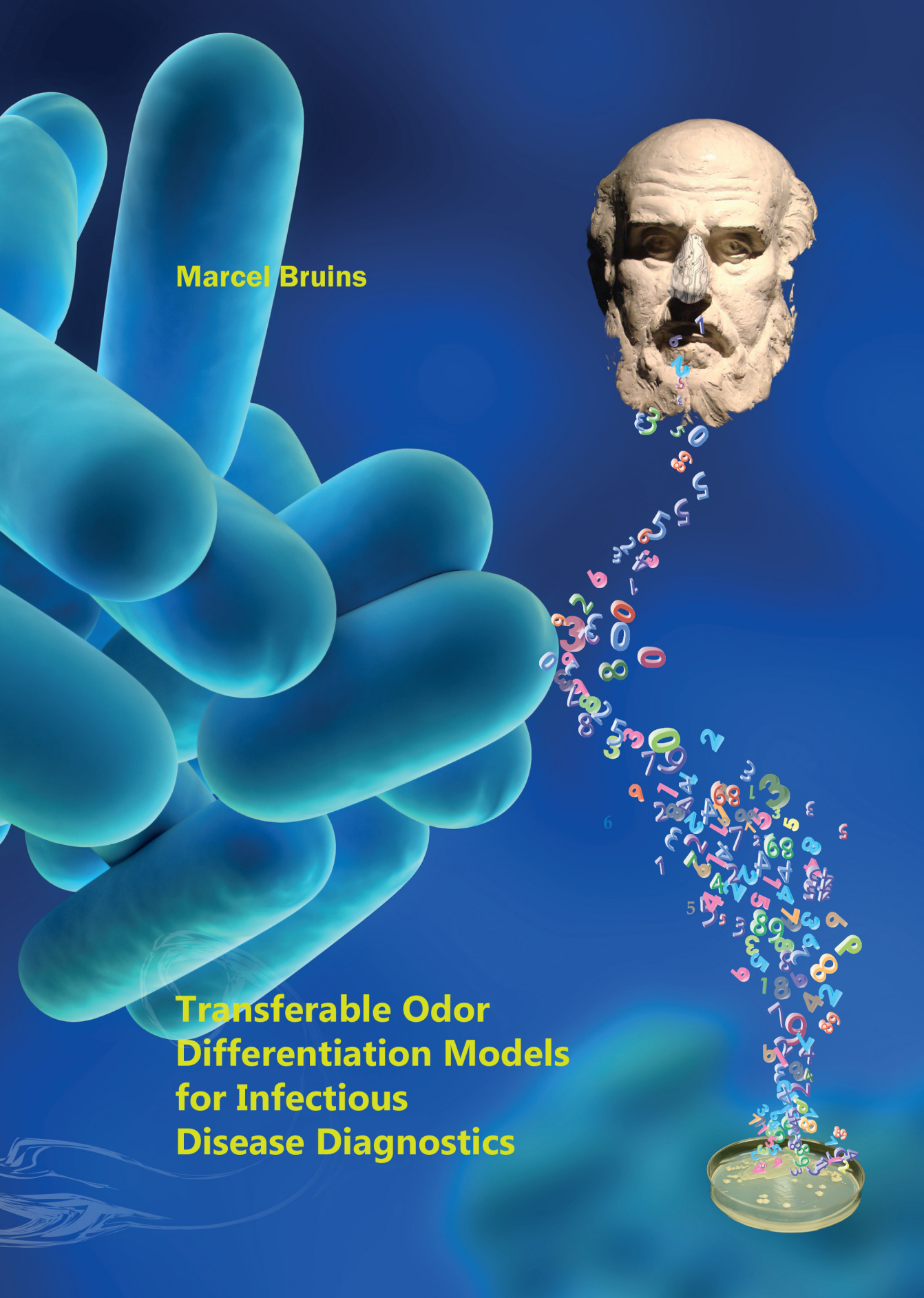


Marcel Bruins

Transferable Odor Differentiation Models for Infectious Disease Diagnostics

Transferable Odor Differentiation Models for Infectious Disease Diagnostics

Marcel Bruins

The printing of this thesis was financially supported by:

ISBN/EAN: 978-94-6108-600-6

Print: Gildeprint, Enschede, The Netherlands

Layout and Design: DoorMies.nl

Transferable Odor Differentiation Models for Infectious Disease Diagnostics

Overdraagbare geur-differentiatie modellen voor infectie ziekten diagnostiek

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam

op gezag van de rector magnificus
prof.dr. H.A.P. Pols
en volgens besluit van het College van Promoties.

De openbare verdediging zal plaatsvinden op
vrijdag 28 februari 2014 om 09:30 uur

door

Marinus Gerhardus Bruins

geboren te Almelo

PROMOTIECOMMISSIE:

Promotor:
Prof.dr.dr. A. van Belkum

Overige leden:
Prof.dr. H.P. Endtz
Prof.dr. D. van Soolingen
Prof.dr.ir. H.J.C.M. Sterenborg

Copromotoren:
Dr.ing. W.W.J. van de Sande
Dr.ir. A. Bos

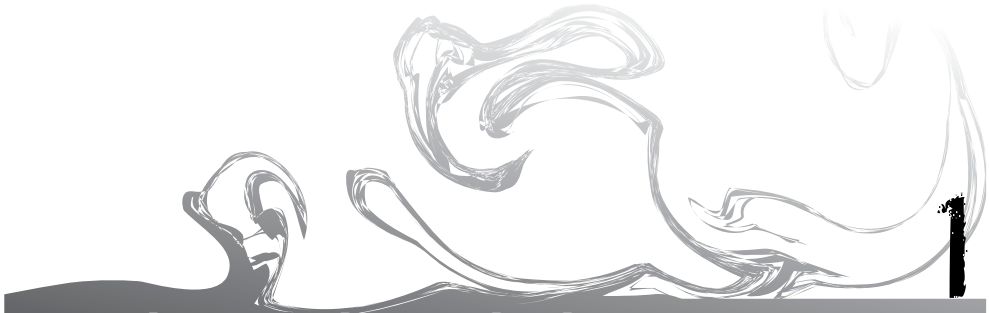
The whole of science is nothing more than a refinement of everyday thinking.

Albert Einstein

contents

Part I

General Introduction



Smelling as a diagnostic aid

chapter 1

M. Bruins
A. van Belkum
A. Bos

Chapter 5 - "The Use of Electronic Nose Devices in Clinical Microbiology"
The Role of New Technologies in Medical Microbiological Research and Diagnosis
Bentham Publishers **2011** 90-101

part of Chapter 10 - Biofysische Methoden
Moleculaire Diagnostiek
Heron-reeks, Syntax media **2011** 421-426

Introduction

The use of smell as a diagnostic aid in the detection of diseases has been known since ancient times. For example, Hippocrates mentioned the diagnostic value of smell in his work called "Aphorisms" [1] written in 400 BC and he wrote in Section IV: *81; " If a patient passes blood, pus, and scales, in the urine, and if it has a heavy* ***smell****, ulceration of the bladder is indicated";* and in Section V: *11. "In persons affected with phthisis, if the sputa which they cough up have a heavy* ***smell*** *when poured upon coals, and if the hairs of the head fall off, the case will prove fatal."* These quotes show that the olfactory diagnosis of (infectious) diseases has a longstanding history, though the possibility of using odor to identify a particular bacterial species was only re-discovered in 1922, when Omelianski described the aroma-producing capabilities of several species of microorganisms [2]. These aromas are a mixture of so-called volatile organic compounds (VOCs) produced by the microorganism. In the current age of modern immunological and molecular diagnostics, the art of olfactory diagnosis is slowly disappearing. However, given the simplicity and ease of olfactory diagnostics, this field may still provide an attractive means for laboratory-based identification or bedside diagnostics. Ultimately, this will require the development of specific instrumentation that is not susceptible to bias or to inter-individual quality variation. This is where modern electronics and software will continue to play an important role.

This thesis focuses on the applicability of (electronic) olfaction in the diagnosis of microbial infections. Nowadays the olfactory diagnosis is used in a much wider scope as it is also being used to diagnose cancer and asthma, for instance. The generic information presented here for the use of microbial diagnosis can be applied to the entire field of medical diagnostics

Everyone who has ever cultured microorganisms knows that certain bacterial species produce odors during growth, some of which are species-specific. Some of these species may be clinically important ones. This odor may be a characteristic sign of the presence of a particular bacterial species, a fact that can be important in the diagnosis of an infectious disease. An experienced person can be able to determine which microorganism is present within a specimen, just by carefully sniffing a specimen. Sometimes this can even be done before it has been cultured. Since this yields a certain risk for the person and does not meet good laboratory practice, an automated approach is needed to retain this knowledge. A few examples of bacterial species and the type of odor they produce are shown in Table 1.1.

Species of Microorganism	Odor
Clostridium difficile	Horse-like
Haemophilus spp	Burnt Caramel
Streptococcus milleri	Butterscotch
Pseudomonas aeruginosa	Almond
Streptomyces spp	Earth
Eikenella corrodens	Bleach
Streptococcus constellatus	Caramel

Table 1.1. Bacterial species and their characteristic odors as perceived by the human nose. The relationship between disease and aroma producing organisms were already reported by Omlanski in 1923[2].

The biological nose

For all animals, including humans, smell is one of the major senses. The actual sensing part of smell is located in the roof of the nasal cavities called the olfactory epithelium, and has a surface area of approximately 3 cm^2 on either side. When air is inhaled, it passes the olfactory epithelium and some of the molecules present will interact with sensory receptor cells. Not all molecules can interact with sensory receptors since they are embedded in a lipid-rich mucous layer produced by the Bowman's glands. This mucous layer limits the kind of molecules that can reach the receptor cells; only molecules that are somewhat water and fat soluble, have a low polarity and sufficient high vapor pressure, are soluble in the mucus and can reach the receptors. On average, a human nose contains 400 different olfactory receptors [4]. Each receptor only responds to a limited and unique set of odorants. The effect is shown in figure 1.1 where a response matrix is shown for 7 odorants and just 4 different receptors. The unique combination of activated receptors allows for a very large set of (mixed) odorants to be identified.

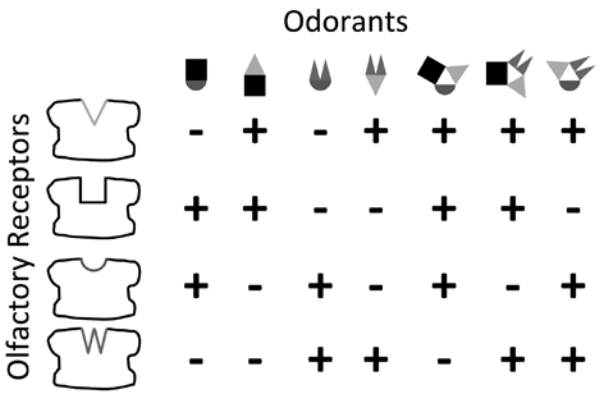


Figure 1.1. Example receptors (1–4) and their potential activation (+) by seven different odorants.

Figure 1.2 shows the anatomy of the human olfactory bulb. The olfactory epithelium contains about 50 million receptors called cilia. These cilia are whip-like extensions of the actual olfactory receptor cell. Each olfactory receptor cell has between 8 and 20 cilia. The olfactory receptors are connected to synaptic structures called glomeruli; each glomerulus receives the information of 10-100 olfactory receptors. In turn the glomeruli converge in mitral cells which finally send the signal to the brains through the olfactory nerve.

The entire biological track contains three different converging steps (cilia → receptor; receptor → Glomerulus and Glomerulus → Mitral cells) condensing the information about a thousand times [4]. The brain combines the signals received and determines what combination of molecules is detected and which odor "belongs" to this combination of molecules.

Since not every olfactory receptor reacts with every molecule, a large number of receptor types is needed to discriminate between as many varieties of odors as possible.

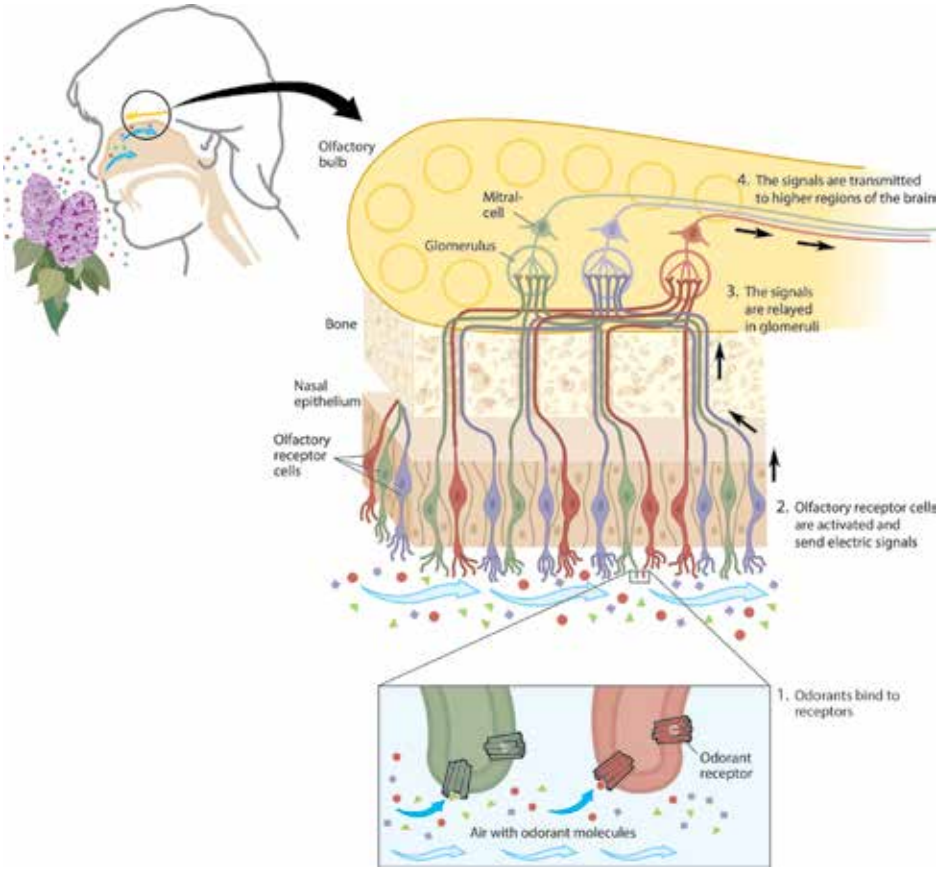


Figure 1.2. Schematic representation of the human olfactory bulb. ©The Nobel Assembly at Karolinska Institutet.

The electronic nose

The concept of an electronic nose (eNose) originated in the 1970s. Up to then, analytical chemistry had been pre-occupied with developing highly specific sensors and methods aimed at identifying unique substances. However, the availability of personal computing made it possible to apply pattern recognition techniques to complex data measurements, including analytical chemistry measurements. The general idea was to develop a broadly responsive sensor system and use pattern recognition systems to match "unknown response" patterns to previously "observed response" patterns, thereby identifying the specific odors present within (complex) mixtures. In fact, this is analogous to the way humans smell complex mixtures of unknown substances containing a characteristic odor; hence the name "eNose". The principles underlying this approach are illustrated in figure 1.3 and compared to the routing of human smelling.

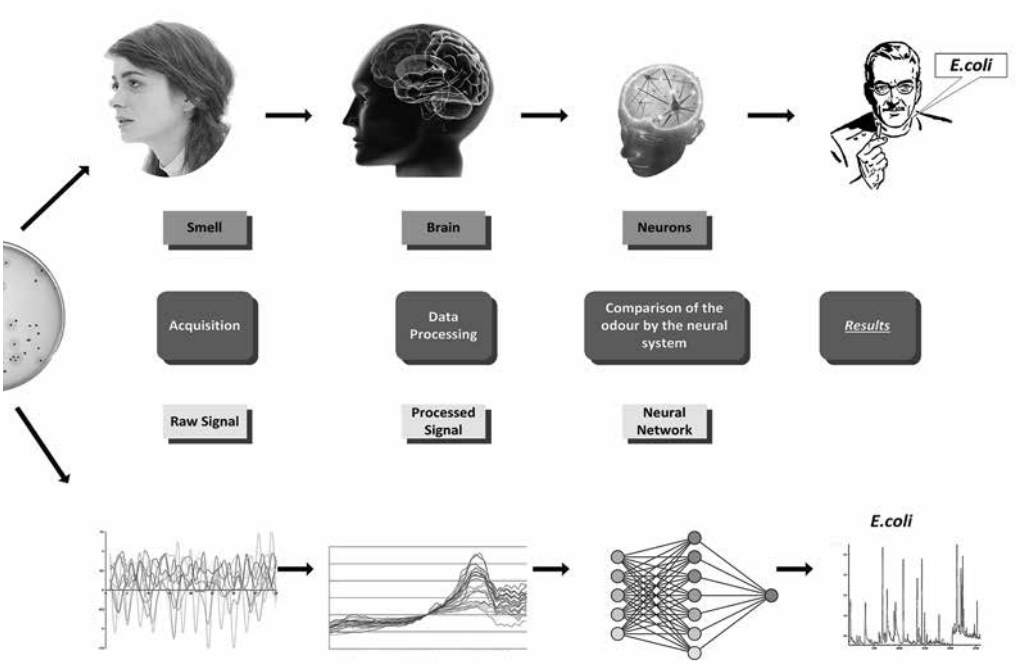


Figure 1.3. Comparison of the steps required to detect an odor using either a biological or an eNose.

An important aspect of the eNose concept is that a substance, or mixtures of substances, can only be recognized after a calibration phase. That is, in order to identify a complex pattern of odors, it is first necessary to develop a database that defines one or more of the "odor patterns". This means that all eNoses currently available rely on the use of a searchable (digital) database where characteristic odor signal patterns obtained from previous measurements are systematically stored. In this way, novel odor signal patterns can be matched with an existing odor profile via comparative pattern recognition analysis, ultimately leading to pattern (and hence odor) identification.

Currently, commercial eNoses are most frequently used in the food industry, in environmental monitoring and for research purposes. These involve relatively simple applications as their main purpose is the detection of "abnormal" situations i.e. during continuous monitoring of a process line or environment. The eNose detects a deviation and triggers an investigation into the deviating odor. These kinds of applications are very useful in situations where every deviation from a normal situation is cause for action or suspicion e.g. CO and H_2S monitoring for human safety.

At the moment, there is no commercially available application of eNose technology in medical diagnostics. However, recent and active research into possible medical diagnostic applications of eNose technology is ongoing. Showing, in general, promising results [5, 6, 7, 8].

eNose Technologies

eNoses generally measure the so-called headspace of a sample. The headspace is the gaseous part of a closed sample container which contains the volatile components of a solid or liquid sample collected in the container. Technically, a gaseous sample does not have a headspace but for uniformity the entire sample is referred to as headspace nonetheless.

In literature [9] GC-MS (gas chromatography-mass spectrometry) [10, 11] and IMS (ion mobility spectrometry) [12] are sometimes described as eNose devices as well. In our opinion, however, these techniques are being used to separate, quantify and identify volatile chemicals but do not indicate whether or not the compound is associated with an odor. In fact, an odor is mostly a "blend" of compounds that combine to create a specific olfactory signature. This blend is actually separated into its components by these techniques, thereby destroying the specific odor itself and are, therefore, "technically speaking" not genuine eNoses. A genuine eNose acquires a sample "as-is" and does not alter the sample. Of course there will always be very small changes to a sample due to detection technique but these are so small that they will generally not influence the sample. A major advantage of such a technique is the continued availability of the original sample for further analysis.

The 3 basic sensor technologies commonly incorporated into modern eNose technologies are:

1. A *quartz microbalance (QMB)* [13, 14, 15] consists of a quartz crystal with a chemically active surface, usually a polymer. The resonance frequency of the crystal shifts due to change in mass when gas molecules adsorb to the surface. These minute shifts are then measured using high-frequency electronics (which are complex and expensive). However, small temperature variations in the environment may also generate similar frequency shifts requiring the maintenance of strict environmentally-controlled temperature conditions.

2. *Conducting polymer (CP) sensors* [14, 16] are polymers, which are either intrinsically conducting or non-conducting with embedded graphite. The graphite provides an electrical resistance path that can be easily quantified. When gas molecules associate with the polymer, it swells, breaking contact points between the graphite particles and thus changing the resistance. In intrinsically conducting polymers the conductivity may change due to the association of analyte molecules. Similar to QMB sensors, very small temperature and/or humidity changes will result in expansion/contraction of the sensor material which in turn will lead to changes in resistance profiles, again requiring strictly controlled environmental conditions. Although the number of possible polymer variants is enormous (and thus the number of sensor array variants), polymers tend to be rather unstable. Strong oxidizers can fairly easily disrupt these polymers and thereby irreversibly change or even destroy the sensor.

3. *Semi-conducting metal oxide (MO) sensors* [13, 14, 17] are based on the principle that certain metal-oxides behave as semiconductors at high temperatures. Sensors based on this principle are designed to join a heating element and a sensor element (usually a sintered metal-oxide with or without catalyst). Both elements are separated by a very thin isolating membrane. Redox-reactions occurring at the sensor surface at elevated temperatures result in changes in resistance that can be measured using appropriate electronic hardware and software. These redox-reactions depend on the nature of the metal-oxide/catalyst, the reacting gas(es), and the temperature. As a “rule-of-thumb”, a minimum of at least 0.1% of ambient oxygen is required for normal operation. Depending on sensor type and temperature, a very broad range of substances will exhibit redox reactions, though notable exceptions are CO_2 (which will not oxidize further), nitrogen and the noble gases. At lower oxygen concentrations the resistance of the metal oxides decreases drastically, or alternatively the conductivity increases.

Besides the three technologies mentioned above, several other eNose sensor technologies [18] exist which are used less frequently. These include for example:

I. *Electrochemical techniques*: these are based on a flow of electrons between an electrode surface and molecules in the surrounding air. This kind of sensor mostly consists of 2 different electrodes and the electron flow between these electrodes is measured and indicative for chemicals present in the air. These sensors can be made fairly specific for a range of chemical compounds. Major drawback of these sensors is the depletion of the reagent during the measurements as it is actually continuously consumed during the process.

II. *Surface acoustic wave (SAW) procedures:* these are largely based on the same principle as the QMB sensors. Instead of measuring a change in weight as with the QMB, the SAW measures a change in length of the path that a high frequency wave travels. Just like the QMB this requires highly specific and very delicate high frequency equipment and environmental stabilization.

III. *Optical methods:* Basically, two optical techniques are being used: spectroscopy and interferometry. In case of spectroscopy, a well-defined light beam (often a laser) interacts with the sample causing absorption and/or scattering. A response is detected as a function of wavelength. Examples include absorption spectroscopy and Raman spectroscopy. In case of interferometry, light from a single source follows two different paths creating an interference pattern. Usually the light beam can exit the optical path (evanescent wave) and interact with a so-called chemo-optical layer in one path, in the other path, the light remains unaffected. An example of such a set-up is the Mach-Zehnder interferometer.

Nowadays there is a definite commercial market for eNose devices. The French company Alpha M.O.S. (Fox2000 with 6 MO sensors [19]) is the market leader since decades. The American company Smiths Detection (which acquired the Cyranose from Cyrano Sciences in 2004, Cyranose32 with 32 CP sensors [20]) and the British company Scensive Ltd (which acquired the Bloodhound assets of Bloodhound Sensors Ltd in 2005, Bloodhound BH114 with 14 CP sensors[21, 22]) are also well known. In 2012 Smiths Detection changed its focus to the military market while the civil market is covered by Intelligent Optical Systems INC. The e-nose devices of Smiths Detection and Scensive Ltd are based on conducting polymer arrays, whilst those of Alpha M.O.S use combinations of both QMB and arrays of conducting polymers. All of these devices are rather bulky, power-hungry and are intended to be used as dedicated laboratory instruments, requiring regular standardization, calibration and normalization, and cost in the range of € 7,500-€ 100,000 per machine.

"The eNose technology" of The eNose Company

In this thesis the eNose module (Figure 1.4) developed by The eNose Company (Zutphen, The Netherlands) was used in all experiments performed. This module uses a single metal-oxide (MO) sensor instead of the multitude of sensors used in other eNose technologies. A device fitted with a single module is called a MonoNose while a device fitted with more modules is called a Aerekaprobe. An Aerekaprobe is mostly used for situations where the best suited MO-sensor cannot be predicted upfront. An Aerekaprobe typically contains 4 distinct types of commercially available sensors (AS-MLC (CO); AS-MLN(NO2); AS-MLK(CH4); AS-MLV(VOC) from Applied Sensors Gmbh (Reutlingen, Germany), varying in metal oxide type and in the catalyzing agent embedded in the surface of the sensor. The 4 types are employed in triplicate in an Aerekaprobe device bringing the total number of sensors to 12. The usage of 4 different types enables the selection of the best type(s) of sensor(s) for a specific situation without having to choose upfront. The usage of a triplicate of each type reduces the influence of senor specific information due to the fact that the sensor independent part of the information will be similar for the three different sensors and the average information is used. Therefore, the major part of the sensor specific information will be discarded when extracting the average information.

Figure 1.4. The sensor module utilized in the eNose technology of The eNose Company. Left: A sensor module containing the sensor in a metal cap with a white Gore-Tex membrane on top, temperature controlling electronics and microprocessor. Middle: The sensor with cap removed (the white dot is the actual sensor material seen at 100 micron diameter). Right: A close-up of the sensor material with its attached electrodes, 2 opposing electrodes for the heating and the other 2 to measure the resistance.

Although the use of a single sensor within an eNose module tends to limit the specificity of VOC detection, the addition of a thermal cycle heating program to the sensor compensates for this lack of chemical specificity, enabling a "virtual array" of sensors [23]. The thermal cycle used in this work operated with a range between 220°C and 320°C which is applied every 20 seconds. The cycle follows a sinusoid shaped profile with 32 steps to facilitate a smooth transition between the temperatures. This sinusoidal thermal profile results in a so-called thermal-loop for each measurement. This measurement principle is illustrated in

fig 1.5. The response of the sensor is recorded as function of the working temperature over time. Each period of this thermal cycle is extracted and the normalized conductivity of the sensor is plotted as a function of the heater temperature, resulting in so-called 'thermal response loops' (Figure 1.5b, right panel). The shapes of these loops depend on the ambient (headspace) gas composition, type of catalyst in the metal-oxide, working temperature and heating dynamics.

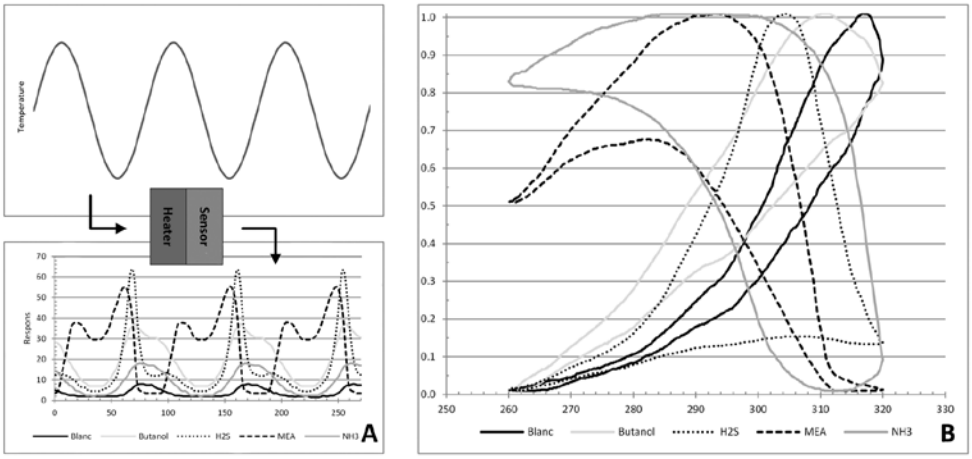


Figure 1.5. Measuring principle of the eNose. A: continuous thermal cycling in which the heater is thermally cycled (top) and the conductivity of the sensor (bottom) is recorded as function of momentary temperature. The temperature profile is applied to the heater while the response is recorded at the sensor. B: extracted normalized thermal loops of a complete thermal cycle. [MEA=methylamine, MMA=methyl methacrylate, IPA=isopropanol, Blanc=clean air].

The eNose in Microbiological Diagnosis

Culture is still the "gold standard" in the majority of diagnostic microbiology laboratories. Identification of unknown microorganisms is done by selective growth-based methods and determining differential biochemical reactivity (leading to variation in coloration, precipitation, gas production etc). For bulk analysis, this process has been automated in devices such as the Phoenix™ (Becton Dickinson) or the VITEK® (BioMérieux). After a sample has been collected from a patient, put into or inoculated on a suitable liquid or solid growth medium respectively, and allowed to grow for a period of time, usually overnight (or longer depending on the initial sample and amount and type of microorganisms) actual identification of the organism can commence. This implies the average time-to-identification of an organism to be at least between 12-24 hours after inoculation with the sample. The total time between arrival of the clinical specimen and definitive identification can vary

between 24-48 hours for “high microbial-load” samples such as urine and faeces, and up to 7 days for “low microbial-load” samples such as blood. In fact, most of the samples that are processed in an average hospital laboratory are negative for pathogens and no further processing is required.

Due to this lengthy diagnostic procedure, various efforts have been focused towards the development of (molecular) techniques that can reduce the “time-to-diagnosis” for clinical specimens. This includes approaches such as Raman spectroscopy, Matrix-assisted laser desorption/ionization - Time of Flight Mass Spectrometry (MALDI-TOF MS), polymerase chain reaction (PCR), etc. [24, 25, 26, 27]. Most of them require a certain microbial biomass that has been obtained by culturing the clinical specimen on a solid medium. PCR is an exception and can be performed directly on a sample. Furthermore, these techniques are relatively expensive and complex. Nucleic acid amplification technologies e.g. the Cepheid GenExpert system, are generally very specific and sensitive, targeting a particular microbial species and may, therefore, require separate assays and conditions for all relevant pathogen to be detected. Until now few studies have used the eNose technology to reduce the time to diagnosis.

We focus on developing and validating the applicability of the eNose technology for microbial diagnosis. The rationale for the use of eNose technology in microbial diagnostics is as follows. It has been known, that microorganisms have the ability to metabolize a very broad range of compounds needed for their growth and development. To process these compounds, microorganisms have developed a variety of metabolic pathways, with the activation of particular metabolic pathways being determined by genetics and nutrition. Additionally, if the mixture of available nutrients changes over time, e.g. by being depleted from the growth medium, microorganisms are capable of adapting their metabolic activity in response to these changes. The activity of particular metabolic pathways over time is generally quite characteristic of particular microbial genera or even species in a certain environment. And activity assessment of these pathways, via for example the measurement of resultant VOCs, could lead to the development of a rapid, broad spectrum, microbial identification system.

Alternatively the human body can be considered a sample container as well and exhaled breath as “headspace”. The lungs contain volatiles produced in the body by different processes e.g. in case of a bacterial infection the VOCs produced by the bacteria (or by the metabolic changes in the body caused by the bacteria) will be secreted through the exhaled air and can consecutively be measured with an eNose.

Experimental design

Before applying e-nose technology in microbial diagnosis, the following experimental design issues need to be considered:

- Experimental setup
- Data collection
- Pre-processing
- Pattern recognition

eNose measurements have become useful thanks to the development of pattern recognition techniques. Due to the multivariate nature of the data and the large amount collected with this type of measurement, univariate mathematical techniques are not suited to obtain useful results. Pattern recognition is a process that involves a number of consecutive steps starting from the (very important) design of the experimental set-up until ultimately the final results. This is ilustrated in figure 1.6 and will be elaborated in more detail in the next paragraph.

Figure 1.6. Schematic flow of pattern recognition flow

Experimental set-up

Headspace analysis allows for a wide range of samples to be tested. In this thesis the focus is on clinical specimens. These may vary from sputum [21, 28] and urine [19, 22, 29, 30], to wound swabs and dressings [31, 32] or blood [33, 34]. The specimens can either be measured directly or they can be used as inoculate for culturing any bacteria present.

A sample is either believed to have a “static headspace” composition that does not change significantly after the sample was collected, or a “dynamic headspace” that changes over time.

In our work we measured samples with a static headspace directly while samples with a dynamic headspace where transferred to a culture bottle containing a growth medium and measured for a longer period of time.

The small changes in a static headspace are caused by the storage of the collected samples. Gaseous samples such as (exhaled) air are typically collected in dedicated sample

bags; these bags, however, are not completely impenetrable, especially for smaller VOCs. After the sample has been collected, the composition of the sample will change due to diffusion through or binding to the sample-bag material. Exposing the bag to direct sunlight may also change the sample composition. Fortunately, these processes tend to be slow, so if the time between collecting the sample and measuring the sample is kept sufficiently short, these influences can be assumed to be negligible. When a liquid or solid sample is collected, it is usually placed in an airtight container with a sufficient volume of headspace. After a short period of time the headspace will be in molecular equilibrium with the solid or liquid, and the headspace will contain the volatile parts of the sample. Of course this equilibrium is temperature dependent. If the temperature of the sample and headspace is kept under control, the composition of the headspace will remain reasonably stable, but again the time between sampling and measuring should be kept as short as possible.

A dynamic headspace has a composition that changes over time due to the (re) activity of the sample itself. Samples containing microorganisms are cultured overnight if the identification of the species is due. This culturing provides the biomass needed for identification and removes the influence of the initial sample matrix. The experimental parameters such as initial inoculum volume, media composition, temperature, etc. have to be kept constant to ensure a reproducible headspace composition after a fixed period of time.

Data collection

After the eNose is exposed to the sample, the sample is measured during a preset period of time and then removed. In this way an entire measurement cycle consists of the exposure and recovery dynamics of the sensor when exposed to the sample. For clarity, a sensor is operated with air under ambient conditions and is in equilibrium with it, providing a baseline of the sensor measurements. When the sensor is exposed to the sample headspace the equilibrium changes due to the introduction of new substances, the so-called exposure phase. After the sample has been removed, the sensor returns to the baseline equilibrium while releasing the substances (either or not in a modified form) still adsorbed at the surface, the so-called recovery phase. When operating at a fixed temperature, this type of measurement typically result in the shark-fin shaped graph depicted in figure 1.7. When applying a thermal cycle, the result is a wave shape formed by, in our case, 32 shark-fins, each slightly different. An example is depicted in figure 1.8.

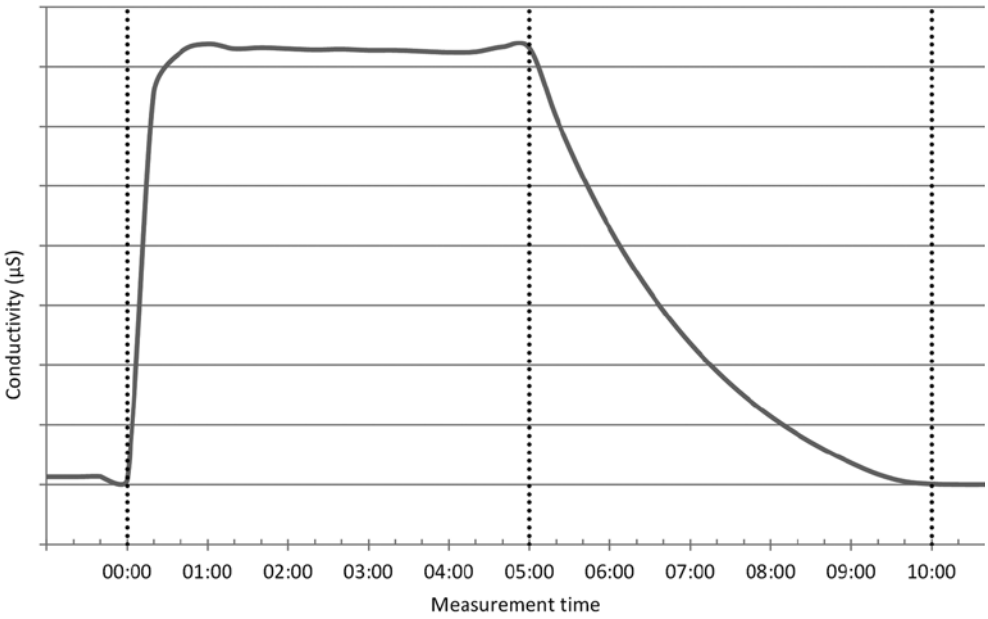


Figure 1.7. Typical Exposure-Recovery dynamics of a single point in the applied thermal cycle.

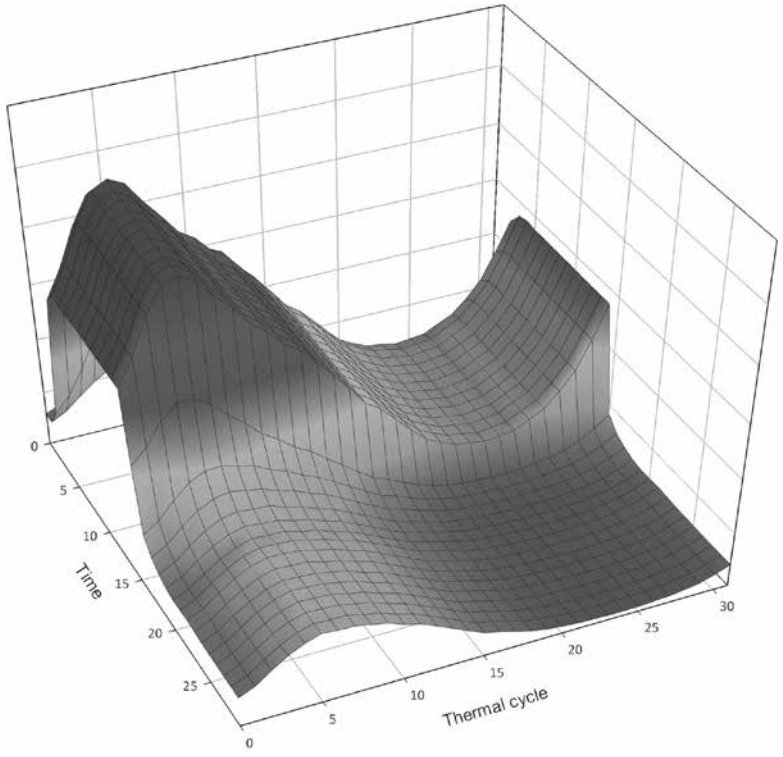


Figure 1.8. Typical Exposure-Recovery dynamics forming a thermal cycle.

Clean-up/pre-processing

Even when applying the best experimental set-up and data-collection methods, it is unavoidable that some measurements will be erroneous. With clean-up, obvious or less obvious erroneous measurements are removed from the full data set. Errors can be present in the form of partial or incomplete data sets or can be caused by experimental difficulties such as plugging of tubes, etc. The successful removal of such errors from the dataset will always improve the diagnostic results and may be a necessity to achieve any result at all. Due to the numerous possible types of errors which depend on the experimental set-up itself, clean-up often implies visual evaluation by the experimenter during the development stage, in order to spot anomalies in the raw data.

Pre-processing is applicable when a data set is processed into a form that can be used for the intended pattern recognition technique. Pre-processing can range from simple averaging and filtering to complex transformation techniques. In literature the term *feature selection* is also used. The latter kind of pre-processing is the most common one used with eNose data and is in essence a data compression technique. This data compression is usually necessary since a typical eNose signal is comprised of a few hundred measured values which contain redundant information. Feature extraction is the process of finding a small set of parameters that describe the signal in as much detail as possible. The simple, easy to calculate and therefore most often used feature extraction methods are:

1. Relative response (fig 1-9a)
2. Area under the curve (fig 1-9b)
3. Exposure or recovery area under the curve (fig 1-9c)
4. Time to maximum response (fig 1-9d).
5. Slope of the exposure or recovery curve

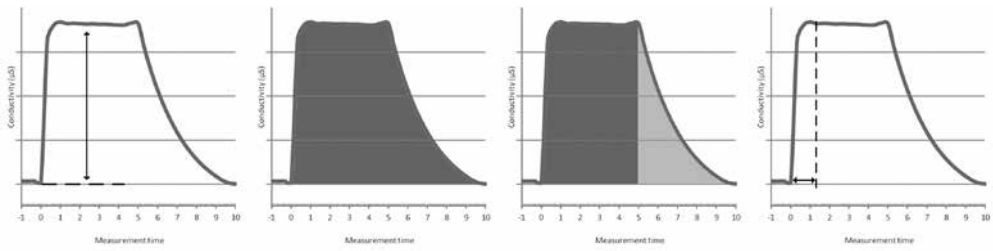


Figure 1.9. Typical Exposure-Recovery dynamics forming a thermal cycle.

These features, however, have the disadvantage of being based on the geometry of the observed signal and therewith discarding all sensor specific properties. In that way only a small amount of the information contained in the data is being used. Since these techniques are based on the operational mode commonly used for an eNose, applying a single temperature, the use of a thermal profile requires a feature for each step in the profile.

These techniques have been used successfully for simple tasks but more sophisticated methods are needed for more complex ones. Examples of the more sophisticated techniques include Fourier or Wavelet transforms [35], best-fitting parameters of the sensors parametric analytic model [36], the phase space and dynamic moments [37] or techniques using an unsupervised Kohonen neural network [38].

Pattern recognition/modeling

Pattern recognition is a process that aims to find mathematical or statistical relationships within a dataset [39]. Pattern recognition operates on different output types and includes techniques such as classification, parsing, sequence labeling and regression. With eNose data pattern recognition is mostly used for classification. Each input value is assigned to one of a given set of classes (e.g. determine whether a given sample originates from a “sick” or “healthy” subject).

Pattern recognition algorithms aim to provide suitable answers for all cases and to generate the most probable output vector for each input vector while taking their statistical variations into account. In contrast, pattern matching algorithms search for exact matches in the input with pre-defined values. Regular expression matching is a common pattern-matching algorithm. It is included in many text editors to search for patterns.

The data pertaining to a sample, represented in a vector containing all information, is formally called an ‘instance’. Within pattern recognition two different approaches are used to create a model, the difference being based on the type of learning used. These two learning methods are supervised and unsupervised learning.

Supervised learning requires a dataset consisting of a set of instances that have been properly assigned to the correct output class, the so-called training data. The selected learning algorithm tries to generate a model that resembles the best fit between two objectives. The first objective is to perform as well as possible on the training data and the second objective is to be as general (or as simple) as possible to allow new data to be correctly classified. When the main focus resides with the matching of the training data, a condition called ‘over-fitting’ occurs: the model fits the training data (almost) perfectly but the classification is based on artifacts in the training data rather than actual differences. An over-fitted model performs poorly when tested with new data. To avoid over-fitting of

the data, the data is normally split in two evenly distributed parts. After the split a model is created using the training data and the constructed model is tested with the remainder of the data. This process continues until a pre-defined performance is reached or a maximum number of attempts, whichever comes first. Well known and often used examples of supervised learning are Artificial Neural Networks (ANN) and Support Vector Machines (SVM).

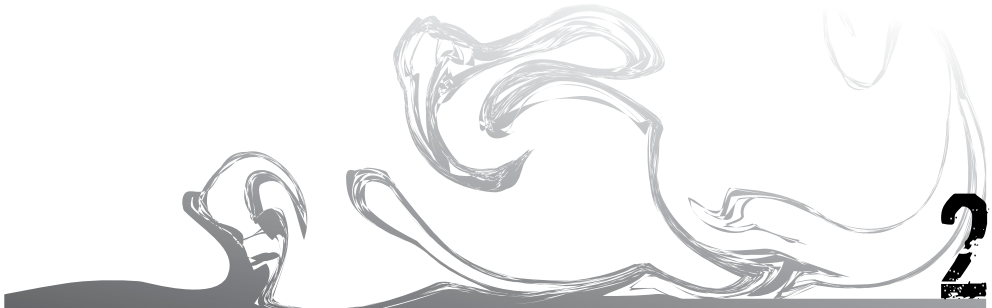
Unsupervised learning assumes no prior knowledge of the classes in the output space and tries to find classes inherent to the data itself. This makes this method less suited for classification purposes since not all relations in the data will be related to the desired classification. For exploratory analysis or data compression, the unsupervised learning methods are however very well suited. When using an unsupervised approach the data cannot be over-fitted since it only finds relations present in the data. The MOLMAP pre-processing step [38], for example, uses an unsupervised Kohonen network to determine the best possible compression of the data. Other well-known and often used unsupervised techniques are clustering and k-nearest neighbor (kNN).

When working with pattern recognition there are two other very important aspects, next to the aforementioned over-fitting of the data, which have to be taken into account: firstly, confounding variables may be present that are “unforeseen and unaccounted-for variables that jeopardize the reliability and validity of an experiment’s outcome”. It is an extraneous variable that has a relationship with the dependent variable and causes a (uncontrolled) bias in the observation. Secondly there may be voodoo correlations, a term introduced by Ed Vul and his co-workers [40]. Voodoo correlations are statistically true correlations within the data set which are only present coincidentally due to the vast number of variables measured and this type of error cannot be corrected for mathematically. The voodoo correlations are of increasing importance due to the ever increasing amount of data collected with modern eNoses. Good data analysis and pre-processing are needed to counteract this effect.

References

1. Marks E – The aphorisms of Hippocrates from the latin version of Verhoofd – Collins & Co 1817
2. Omelianski VL – Aroma producing microorganisms – J Bacteriol 1923 8(4):393–419.
3. Pavlou AK, Turner APF – Sniffing out the truth : clinical diagnosis using the electronic nose – Clin Chem Lab Med 2000 38(2):99-112
4. Leffingwell Reports, Vol. 2 (No. 1), May, 2002
5. Pavlou AK, Magan N, Sharp D, Brown J, Barr H, Turner APF – An intelligent rapid odour recognition model in discrimination of *Helicobacter pylori* and other gastroesophageal isolates in vitro – Biosensors & Bioelectronics 2000 15:333-342
6. Gibson TD, et al. – Detection and simultaneous identification of microorganisms from headspace samples using an electronic nose – Sensors and Actuators B 1997 44:413-422
7. Moens M, et al. – Fast identification of ten clinically important micro-organisms using an electronic nose – Letters in Applied Microbiology 2006 42:121-126
8. Rossi V, Talon R, Berdagué J – Rapid Discrimination of Micrococcaceae species using semiconductor gas sensors – Journal of Microbiological Methods 1995 24:183-190
9. Rock F, Barsan N, Weimar U – Electronic nose: current status and future trends – Chem.Rev 2008 108:705-715
10. Lechner M, Fille M, Hausdorfer J, Dierich MP, Rieder J – Diagnosis of bacteria in vitro by mass spectrometric fingerprinting: a pilot study – Current Microbiology 2005 51:267-269
11. Judy W, Arnold JW, Senter SD – Use of digital aroma technology and SPME GC-MS to compare volatile compounds produced by bacteria isolated from processed poultry – Journal of the Science of Food and Agriculture 1998 78(3):343-348
12. Bessa V, et al. – Detection of volatile organic compounds (VOCs) in exhaled breath of patients with chronic obstructive pulmonary disease (COPD) by ion mobility spectrometry – International Journal for Ion Mobility Spectrometry 2011 14(1):7-13
13. Messina V, de Reca NW – Applications of Electronic Nose Based on MOX and QMB Sensors – Sensors & Transducers Journal 2012 140(5):106-114
14. Arshak K, Moore E, Lyons GM, Harris J, Clifford S – A review of gas sensors employed in electronic nose applications – Sensor Review 2004 24(2):181 – 198
15. O'Sullivan CK, Guilbault GG – Commercial quartz crystal microbalances - theory and applications – Biosensors and Bioelectronics 1999 14(8–9): 663–670
16. Ramanaviciusa A, Ramanavicienea A, Malinauskasc A – Electrochemical sensors based on conducting polymer - polypyrrole – Electrochimica Acta 2006 51(27):6025–6037
17. Korotcenkov G – Metal oxides for solid-state gas sensors: What determines our choice? – Materials Science and Engineering: B 2007 139(1):1–23
18. Arshak K, Moore E, Lyons GM, Harris J, Clifford S – A review of gas sensors employed in electronic nose applications – Sensor Review 2004 24 (2):181–198
19. Kodogiannis V, Wadge E – The use of gas-sensor arrays to diagnose urinary tract infections – Int J Neural Syst. 2005 15(5):363-76.
20. Dutta R, Hines EL, Gardner JW, Boilot P – Bacteria classification using Cyranose 320 electronic nose – BioMedical Engineering OnLine 2002 1(4) ISSN 1475-925X
21. Fend R, Kolk AH, Bessant C, Buijtels P, Klatser PR, Woodman AC – Prospects for clinical application of electronic-nose technology to early detection of *Mycobacterium tuberculosis* in culture and sputum – J Clin Microbiol 2006 44(6):2039-45.
22. Pavlou AK, et al. – Use of an electronic nose system for diagnoses of urinary tract infections – Biosensors and Bioelectronics 2002 17:893-899
23. Pearce TC, Schiffman SS, Nagle HT, Gardner JW – Handbook of machine olfaction – WILEY-VCH 2003

24. Willemse-Erix DFM, et al – Optical fingerprinting in bacterial epidemiology: Raman spectroscopy as a real-time typing method – J Clin Microbiol 2009 47(3):652-659
25. Buijtels PCAM, et al. – Rapid identification of Mycobacteria by Raman spectroscopy – J Clin Microbiol 2008 46(3):961–965
26. O.Lay Jr J – MALDI-TOF mass spectrometry of bacteria – Mass Spectrometry Reviews 2001, 20:172-194
27. Lechner M, Fille M, Hausdorfer J, Dierich MP, Rieder J – Diagnosis of bacteria in vitro by mass spectrometric fingerprinting: a pilot study – Current Microbiology 2005 51:267-269
28. Pavlou AK, Magan N, Jones JM, Brown J, Klatser P, Turner AP – Detection of *Mycobacterium tuberculosis* (TB) in vitro and in situ using an electronic nose in combination with a neural network system – Biosens Bioelectron 2004 20(3):538-44
29. Aathithan S, Plant JC, Chaudry AN, French GL – Diagnosis of bacteriuria detection of volatile organic compounds in urine using an automated headspace analyzer with multiple conducting polymer Sensors – Journal of Clinical Microbiology 2001 39(7):2590-3
30. Kodogiannis VS, Lygouras JN, Tarczynski A, Chowdrey HS – Artificial odor discrimination system using electronic nose and neural networks for the identification of urinary tract infection – IEEE Trans Inf Technol Biomed 2008 12(6):707-13
31. Bailey ALPS, Pisanelli AM, Persaud KC – Development of conducting polymer sensor arrays for wound monitoring – Sensors and Actuators B 2008 131:5-9
32. Pisanelli AM, Persaud KC, Bailey A, Stuczen M, Duncan R, Dunn K – Development of a diagnostic aid for bacterial infection in wounds – 13th International Symposium on Olfaction and Electronic Nose. AIP Conference Proceedings 2009 1137:133-135
33. Yates JW, et al. – Data reduction in headspace analysis of blood and urine samples for robust bacterial identification – Comput Methods Programs Biomed 2005 79(3):259-71.
34. Trincavelli M, Coradeschi S, Loutfi A, Soderquist B, Thunberg P – Direct Identification of Bacteria in Blood Culture Samples using an Electronic Nose – IEEE Trans Biomed Eng 2010
35. Xing-Jiu H, Yang-Kyu C, Kwang-Seok Y, Euisik Y – Oscillating behaviour of hazardous gas on tin oxide gas sensor: Fourier and wavelet transform analysis – Sensors and Actuators B 2006 115:357–364
36. Carmela L, Levyb S, Lancetc D, Harela D – A feature extraction method for chemical sensors in electronic noses – Sensors and Actuators B 2003 93:67–76
37. Vergara A, et al. – Feature extraction of metal oxide gas sensors using dynamic moments – Sensors and Actuators B 2007 122:219–226
38. Ballabio D, Consonni V, Todeschini R – Classification of multiway analytical data based on MOLMAP approach – Analytica chimica acta 2007 605:134–146
39. Massart DL, Vandeginste BGM, Buydens LMC, De Jong S, Lewi PJ, Smeyers-Verbeke J – Handbook of Chemometrics and Qualimetrics – Elsevier ISBN: 978-0-444-89724-4
40. Vul E, et al. – Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition – Perspect. Psychol. Sci. 2009 4:274–90



chapter 2

Outline of this thesis

As documented in **chapter 1,** the introduction of the eNose in the 1970's also started significant research into possible medical diagnostic applications of eNose technology with, in general, promising results. Currently, however, there is no known device or application that utilizes eNose technology in medical diagnostics. The urea breath test for *Helicobacter pylori* uses the breath but only detects the products created by the splitting of introduced, labeled urea (either radioactive C^{14} or C^{13} introduced) by an enzyme (urease) produced by *H.pylori*. We first investigated the possible causes for this lack of practical applications and then we studied the applicability of metal-oxide sensor based eNoses in different aspects of the diagnosis of infectious diseases.

In **chapter 3** we address one of the major issues keeping eNoses from becoming widely used. Currently there is no method to transfer calibration models between eNoses. This lack of transferability imposes that each device needs to be calibrated for each new situation. Even with the effort to create a calibration model, that specific model will become useless if one of the sensors needs to be replaced. We studied the influence of temperature fluctuation on metal oxide sensors and showed that a temperature shift results in a shift in response 10-15 times higher than the shift in response caused by concentration variation, and therefore defines an important requirement for transferable calibration models.

In **chapter 4 and 5** we address two of the most important parts of the data analysis. In **chapter 4** we show that it is possible to obtain 100% sensitivity and specificity with purely random data, if the ratio between measured variables and subjects is large enough. However, using a leave-one-out approach works in reverse: the larger the amount of variables, the smaller the chance of finding a non-existing correlation, a so-called 'voodoo correlation'. In **chapter 5** we explain the principles of the data-compression developed for our measurement since this is returning throughout this thesis

In **chapter 6** we describe how we continuously monitored the headspace gases above a broth inoculated with a single bacterial species with the eNose. We found that different species could be identified within 6-8 h after inoculation of the broth, with an average success rate of 87% [67% - 100%], and an influence of the medium composition on the odor produced; a small change in medium composition can have a large influence on the produced VOC fingerprint. Finally we found a clear behavioral difference between gram positive and gram negative bacteria. On average the gram negatives produce VOCs that can be easily measured with a MO-sensor while the gram positive bacteria produce VOCs that are more difficult to characterize.

In **chapter 7** we describe how we evaluated the odor from vaginal discharge of cows, in the first 10 days after calving, by olfactory cognition and an electronic device. We found that the repeatability of the eNose diagnosis of acute postpartum metritis (inflammation of the wall of the bovine uterus) to be higher than the human olfactory evaluation. The eNose

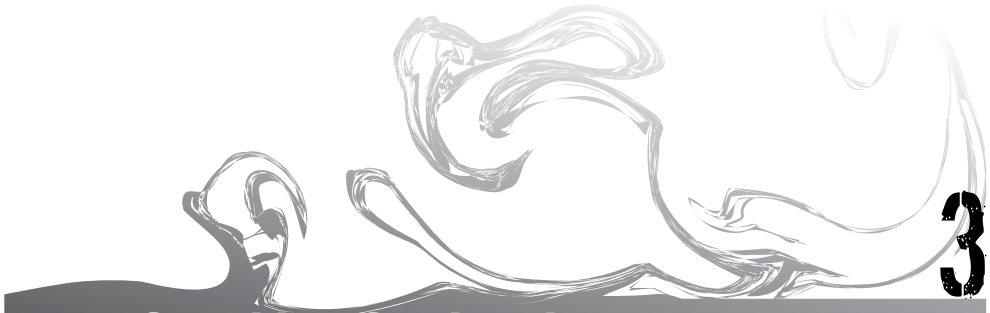
found a sensitivity and specificity of 92·0% and 100% compared to 75·0% and 60·1% for the human olfactory evaluation.

In **chapter 8** we describe how we tested the ability of the eNose to diagnose acute postpartum metritis in cows during the first 10 days after calving. There is no real tool for this diagnosis at this moment; the veterinarian checks a list of conditions to determine the absence or presence of metritis. We used vaginal swabs, collected at day 2, 5 and 10, stored at -21°C and heated to 38°C before measuring. We found a sensitivity of 100% and a specificity of 91·6% when diagnosing half of the samples with a classification model based on the other half of the samples. However the prediction of the causing pathogen was not possible, likely due to the opportunistic nature of the APM infection.

In **chapter 9** we describe how we tested the diagnostic accuracy of the eNose for active pulmonary tuberculosis (TB). We started with an exploratory study in Bangladesh to ascertain that the eNose could differentiate between healthy people and people with severe TB. After this successful study we continued with a more extensive study measuring all people enrolling at the Shyamoli Chest Disease Clinic hospital and healthy controls from the same socio-economic setting. In this follow up study we found a sensitivity of 76·5% and a specificity of 87·2% within the entire population.

Finally in **chapter 10** we summarize the research performed, define the major findings and provide an outlook on some of the additional investigation that should still be done.

Metal-Oxide Sensors – Perks & Pitfalls



chapter 3

Transferability of calibration models

M. Bruins
J.W. Gerritsen
W.W.J. van de Sande
A. Bos
A. van Belkum

Enabling a transferable calibration model for metal-oxide type electronic nose

Published in
Sensors and Actuators B **2013** 188:1187-1195

Abstract

While electronic noses have been around for over 30 years, little effort has been devoted to the development of transferable calibration models, which are models that can be applied to multiple equivalent devices without adjustment. The majority of published results limit itself to data sets gathered with a single device. This lack of insight in transferable models hampers large scale implementations of eNose-based applications as individual calibration of a multitude of devices for a specific application area is generally unrealistic due to the requirement of actual samples to be measured. For simple gases this may be do-able, but in the case of more complex samples such as biological patient material it is logistically impossible.

In this paper we show the influence of the deviation of the sensor temperature on the measurement reproducibility and by inference on the transferability of calibration models. We introduce the total inertia (φ^2) as a measure for the heterogeneity within the measured data. The total inertia is an objective measure known from linear algebra, where it is used to calculate the correspondence between matrices.

We use 5 micro-hotplate metal-oxide sensors from the same wafer, with an inter-sensor heater temperature difference of approximately 15°C in combination with 2 substances, n-butyl-acetate and hexane. This research demonstrates the increase in heterogeneity of the measured response values in relation to a temperature shift. A shift of 15°C at the sensor surface causes an increase of heterogeneity that is 10-15 times higher than the increase in heterogeneity caused by inter-sensor responses to the substances, when operated at exactly the same temperature. Some mixtures of substances will be separable by pattern recognition under virtually any condition and strict temperature control will neither improve nor deteriorate results. However, the significant contribution of temperature deviation towards data heterogeneity renders it plausible that optimized temperature control, and by inference lower data heterogeneity, is a prerequisite for transferability of a calibration model. This holds true when applied to metal-oxide sensors and for mixtures containing substances showing a fair degree of similarity.

Introduction

Ever since the introduction of the eNose in the 1980's [1] it was shown to be a useful alternative for techniques such as GC-MS and several forms of spectroscopy. eNose technology can be used for various analytical processes, quality control procedures [2-10] and disease diagnostics [11-19]. Despite the growing interest and its proven potential, no eNose-based application has been implemented in medicine on a large scale yet.

The major issues in this respect are the poor reproducibility of the measurement results between presumably identical eNose devices and the difficulty in transferring a once-developed calibration model to another, supposedly equal eNose device. Similar results can be obtained with a calibration model based on the data measured with another eNose device. Matters become complicated, however, when one aims to combine the datasets of multiple devices, which should essentially show comparable results. It turns out that most models are device-specific and using a model generated with data from another device of the same type often yields unacceptable results. This drop below acceptable quality levels indicates that there are variations between the eNose devices, or between the conditions under which the devices are used, causing the data to be strictly correlated to the specific device used to obtain the initial data. This holds even true if all individual devices find the same great diagnostic potential and applies to all commonly used sensor types.

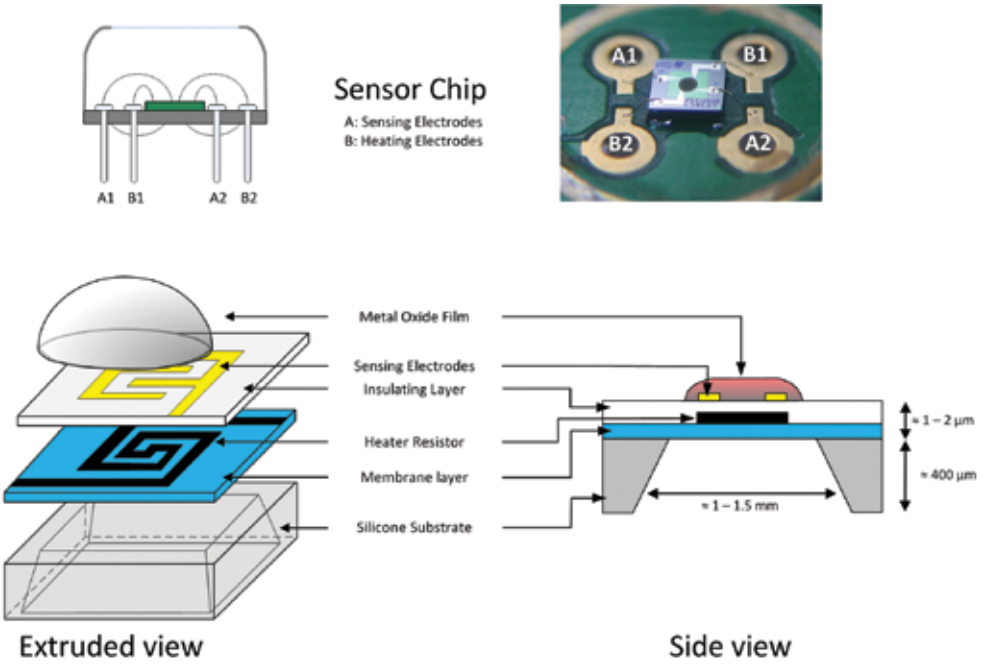


Figure 3.1. A schematic image of a metal oxide sensor.

This problem has been known for quite some time and several solutions have been proposed which all failed to effectively solve the problem [20-22]. All attempts, however, have been targeted at finding a mapping method between devices. Eliminating key differences at the core of the problem, i.e. controlling the sensor itself, has not been addressed in much detail before.

Our focus is on low-cost, mass producible eNoses. We use micro hotplate metal oxide (MO) sensors as these are low-cost, robust, and broadly applicable, hence retaining the potential for mass applications. Therefore, this paper focuses on MO sensors in particular and the name "sensor" refers to a MO sensor. The surface of the sensors used is directly mounted on an embedded heater allowing fast temperature changes (20-400°C in 2-3 milliseconds) with little loss of heat. This is advantageous since the redox reactions taking place at the sensor surface are highly temperature dependent (figure 3.1).

The unit measured with a MO-sensor is the conductivity, which generally depends on the path length between the sensing electrodes; the chemical composition and the state of the surface and the dynamics of the reaction processes (both redox-reactions at the surface of the sensor and reactions between products created) at a specific temperature. The (concentration of) oxygen species at the sensor surface, present in different states (O_2, O_2^-, O radicals, etc.), and the ratio between these species determine the conductivity. Depletion of adsorbed oxygen species results in a change of flow of electrons resulting in a change of the measured conductivity. Therefore, all sensors have only one conductivity value per temperature for a specific gas composition. Since the number of measurable substances, mixtures and concentrations thereof, largely outnumber the conductivity values, the identification capability of a single sensor at a single temperature is limited. Additional information can be obtained when operating the sensors with thermal cycles. In our eNoses we mimic a multisensory array using the well-known thermal modulation technique [23 - 25], in which a single sensor is operated at different temperatures. Due to the thermal profile applied, the single sensor can be regarded as a virtual array of sensors [26]. Operating a single sensor with a thermal profile instead of an array of sensor elements at different temperatures has two major advantages:

1. The slow redox reactions increase the correlation among neighboring virtual sensors (in terms of time in the modulation cycle). This provides additional information due to the temperature dynamics of the processes taking place: the gradually changing temperature results in concurrent changes in the equilibria in the system as depicted in figure 3.2.

2. Since it involves a virtual array, the sensor surface is the same for each of the virtual sensors. Therefore, the observed conductivity differences are caused by magnitude and dynamics of the applied thermal profile and the volatiles measured, and not by the manufacturing differences between the elements of a physical sensor array.

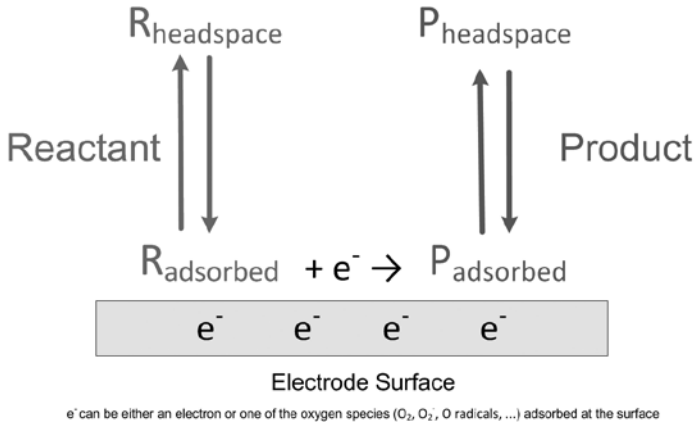


Figure 3.2. Equilibria present in a sensor/headspace environment. The reactant (R, volatiles from the headspace) adsorbs at the electrode surface while the Product (P) desorbs from the surface.

When a thermal profile is used with slow step-speed allowing each step to reach equilibrated steady state with respect to the redox reactions occurring at the sensor surface, the hysteresis effect will be negligible and no difference will be observed between measurements at the same temperature within the cycle. There will be no difference between a temperature preceded by a lower temperature and one preceded by a higher temperature.

Although the MO-sensors are produced on wafers using micro-electro-mechanical system (MEMS) technology, it is virtually impossible to produce two sensors with exactly the same characteristics: the reactive surface, (doped) SnO_2, is normally deposited on the sensor by a printing process. This deposition process renders it impossible to ensure that each sensor contains the exact same amount of active material and identical morphology. These differences cause all sensors to behave slightly different, even when they originate from the same batch. Next to the differences in the surface material, differences in the production process of the heating element (a platinum resistor) occur as well. These differences result in different temperature values at the same heater resistance or working voltage. This effect is directly proportional to the sensor operating temperature resulting in increasing difference levels at higher temperatures.

Reproducibility between measurements is a necessity for creating a useful calibration model. For a *device-based* calibration model, the reproducibility only needs to apply to a single (set of) sensor(s). A *device-independent* calibration model, however, requires a higher degree of reproducibility. In order to be able to create a device-independent calibration model, one needs to ensure that every sensor of a specific type generates a similar response to a given substance.

In this article we will show that accurate control of the operational temperature has an extended influence on the similarity and reproducibility of the signal and is a prerequisite for transferability of calibration models between sensors. The issue of temperature control has been addressed before [27 - 29] but to our knowledge not in relation to the impact on measured data. This is shown in experiments with two different substances and 5 distinct sensors of the same type. At the same time we demonstrate that the deleterious effect of the variability of the reaction surface morphology can be reduced by normalizing the measurement results.

Material and methods

Experimental

A deviation of sensor temperature will alter signal measurements. Our hypothesis is that these deviations, mainly caused by the intrinsic differences between sensors of the same type due to the production process, will be larger than the observed differences between substances. This will potentially deteriorate the classification results.

To test our hypothesis we measured two substances to investigate the impact of sensor temperature deviation on the measurement reproducibility, and by inference on the transferability of calibration models. We used the active temperature control described below to operate the sensors at several fixed temperatures thereby simulating the “off-factory” differences present in a hypothetical batch of sensors.

Accurate temperature control is obtained using a patented high precision, programmable Wheatstone bridge setup that allows setting of a desired resistance value in two distinct phases [30]. Firstly, each sensor is calibrated using an embedded procedure allowing the microcontroller to set the heater resistance at a known value before calculating the temperature using the known relationship between the temperature and power dissipated by the heater under standard conditions. In this procedure the microcontroller creates a sensor-specific lookup-table that correlates the resistance with the actual temperature. Secondly, the microcontroller sets the temperature at the set-points required for the

thermal cycle used and keeps the resistance constant at each set-point and therewith the temperature regardless of the ambient conditions [28]. This approach results in a maximum temperature deviation of 3-5 °C between any two sensors which is significantly less than the production-related variability of maximum 50°C.

We introduce the total inertia, a similarity measure originating from correspondence analysis [31], as a measure for this deviation. We could have used a classification algorithm and shown the increase in quantitative prediction error or decrease in classification rate. However, the quality of the classification method used would camouflage the issue and blur the actual amount of deviation observed.

Although in real life most samples and situations will be composed of mixtures of volatile compounds, we decided to use 2 chemicals (n-butyl acetate, Merck, > 98% and n-hexane, Fisher Scientific, >99%) in their pure form for demonstration purposes. In this way, the variations observed are only caused by the shifts in operating temperature of the sensor and not complicated by the changing ratio between substances.

The measurement set-up is comprised of 5 individual eNose modules, fitted with a single sensor (AS-MLV sensors, Applied Sensors GmbH. Reutlingen, Germany), attached to a sample flow. The sample flow is controlled by three programmable pumps. One pump controls the main flow; this pump has a carbon filter attached to produce a zero-grade base flow (atmospheric air that has been cleaned to have less than 0.1 ppm total hydrocarbons remaining). The two other pumps control the concentrations of both substances in the main flow. During the experiments, the total flow rate of the system is kept at a constant value. All tubing is made of Viton in order to limit interactions. A schematic of the set-up is depicted in figure 3.3.

A 6000 ppm stock solution is prepared for both substances in an 8 l Tedlar® bag, filled with zero-grade air. These bags are attached to their corresponding sample pump.

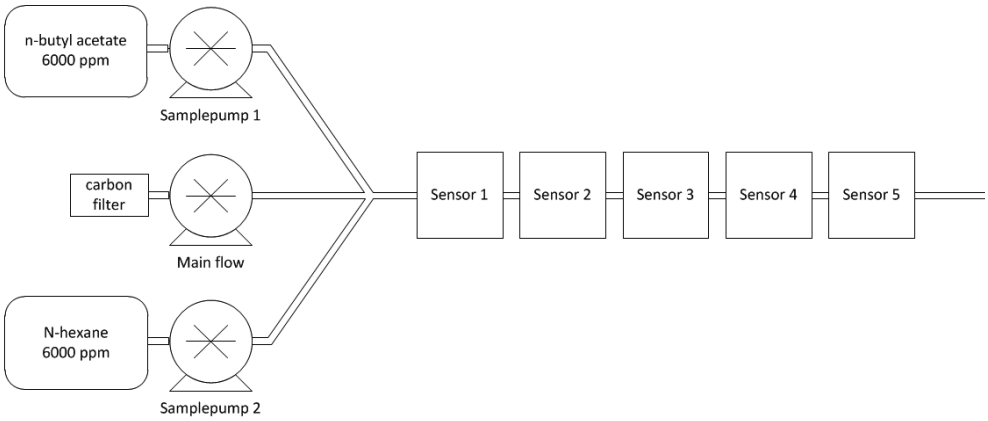


Figure 3.3. Schematic representation of the measurement setup.

A sinusoidal-shaped thermal profile with a full range of 50°C (25°C either way of the set center) and a cycle length of 7 seconds was applied to the sensors with 5 different center temperatures of 260°C, 270°C, 275°C, 280°C and 290°C. During a measurement cycle, each sensor is fixed at one of the 5 center temperatures resulting in a full range of temperatures per measurement. After each measurement the temperatures were changed in such a way that all temperature values were applied to each sensor.

Each time the sensor center temperature is changed the sensors are stabilized at their new temperature profile for 2 hours. Then n-butyl acetate is inserted into the system by turning on the appropriate sample pump while the main flow pump is altered at the same time to keep the flow passing the sensors at a virtually constant rate. The chemical substance is added to the main stream for a period of 15 minutes followed by a chemical free recovery period of 105 minutes. Each of the chemicals is added to the main flow at a rate to accomplish a total concentration of 50, 75, 100, 125 and 175 ppm respectively. Between the two substances the recovery period is extended from 105 to 180 minutes to ensure no residue is left in the system interfering with subsequent measurements.

In the experiment, all temperature, sensor, and volatile permutations were measured giving a view on the influence of temperature control and therewith the reproducibility and calibration model transferability.

In this work, we applied a sinusoidal temperature profile and recorded 32 conductivity values per cycle. Using the sinusoidal shape we measured 16 temperatures twice. With a thermal profile applied to a single sensor, the two conductivity values will differ due to the difference in reaction dynamics. For a periodic profile, this phenomenon is observed as a form of hysteresis which leads to additional information on the volatiles present.

Calculations

Since we aim to show the variance in the data recorded (the shape of the thermal loops) and not the quality of a classification algorithm we use the total inertia (ϕ^2) , a measure originating from correspondence analysis [31], to determine the amount of distortion of the thermal loop caused by the concentration and temperature variations. Correspondence analysis is based on classical, straightforward techniques in matrix algebra and mostly used when dealing with contingency matrices. However, the calculations can also be used with other matrix types. A single vector is composed of the 32 conductivity values belonging to a thermal cycle. To eliminate all amplitude information, which would introduce unwanted heterogeneity into the data all vectors are normalized using equation.3.1.

$$Normalized(x_i) = \frac{x_i - V_{min}}{V_{max} - V_{min}} \quad \text{(Eq. 3.1)}$$

Where

$$x_i = i^{th}\ value\ in\ the\ vector$$
$$V_{min} = The\ minimum\ value\ in\ the\ vector$$
$$V_{max} = The\ maximum\ value\ in\ the\ vector$$

The total inertia of a table is the weighted sum of the row inertias which in turn are defined as the product of the row's mass (its fraction of the total) and the square root of the χ^2 distance to the theoretical centroid. The χ^2 distance is a weighted Euclidean distance calculated between rows which are normalized by dividing each value by the total sum of the vector. The total inertia is also known as the Pearson's mean-square contingency denoted by ϕ^2 which is a measure for the amount of heterogeneity of the data in the matrix.

The ϕ^2 can be calculated using equation 3.2.

$$Total\ Inertia = \sum_{i=1}^{m} r_i * \sum_{j=1}^{n} \frac{\left[\frac{x_{i,j}}{r_i} - c_j\right]^2}{c_j} \quad \text{(Eq. 3.2)}$$

Where

$$m = number\ of\ rows\ in\ the\ matrix$$
$$r_i = The\ mass\ (fraction)\ of\ the\ i^{th}\ row$$
$$n = number\ of\ columns\ in\ the\ matrix$$
$$x_{i,j} = value\ at\ row\ i\ and\ column\ j\ of\ the\ matrix\ divided\ by\ the\ matrix\ total$$
$$c_j = The\ mass\ (fraction) of\ the\ j^{th}\ column$$

The row inertia is a measure between a row and the virtual centroid of the matrix. Since all values are scaled, the maximum row inertia that can be observed is 32. This value represents a vector that is orthogonal in every of the 32 dimensions. The total inertia for the matrix is the weighted average of all row inertias. Since the inertia is calculated with regard to the centroid a total inertia can never exceed 32. In practice the maximal total inertia will be 70-75% of the maximal row inertia. This value is an estimate based on the centroid principal and not a calculated amount. In our case the maximal total inertia would then be 22-24. A total inertia of 0 indicates that all vectors in the matrix are equal. In the remainder of this article we will use the term "heterogeneity" to indicate dissimilarity in the data instead of the term "total inertia". The total inertia is a measure of the dissimilarity while heterogeneity is a commonly used term to indicate a variance in data.

Results

Each of the used sensors has been calibrated using the internal temperature calibration procedure [30] and the calibration data for the 5 sensors used are plotted in figure 3.4.

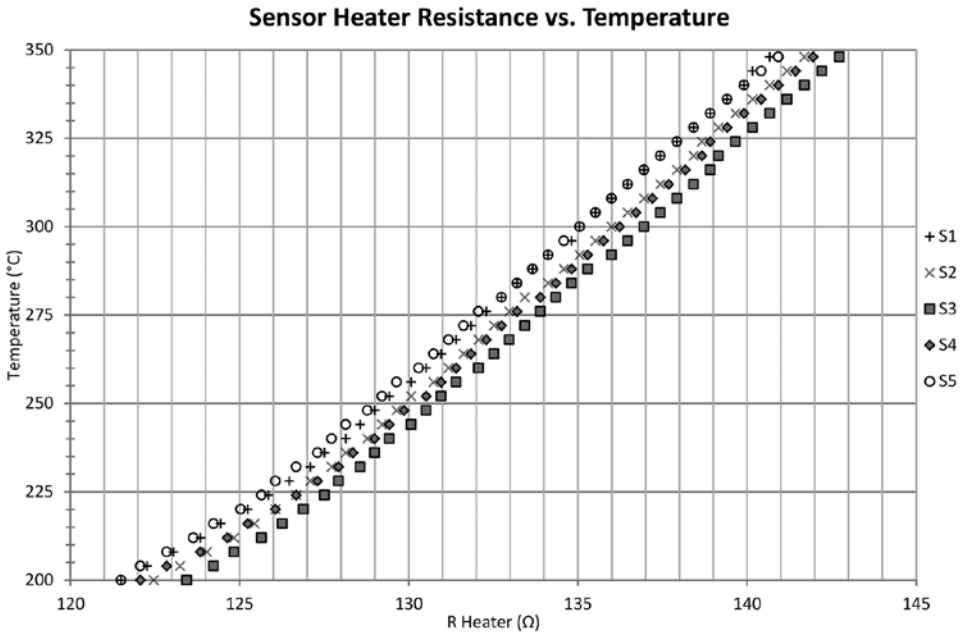


Figure 3.4. Sensor heater resistance vs. Temperature. The calibration values of the sensors used.

From this figure it can be concluded that the 5 sensors used, originating from the same wafer, show temperature differences of approximately 15°C at a given resistance setting of the heater. The variance of 15°C is used as base value for the temperature range used in our experiments.

When the conductivity is plotted against the operating temperature, a 'thermal loop' is created showing a fingerprint of a given substance. An example of a thermal loop is shown in figure 3.5.

The basic shapes used for classification purposes are the thermal loops or descriptors based on the loops. We plotted the thermal loops for a 100 ppm concentration of both substances at 3 different temperatures (260°C, 275°C, and 290°C). Figure 3.6 shows these loops in which all 32 measured values have been scaled to remove the sensor specific amplitude effect, while retaining the substance specific information. As can be seen in figure

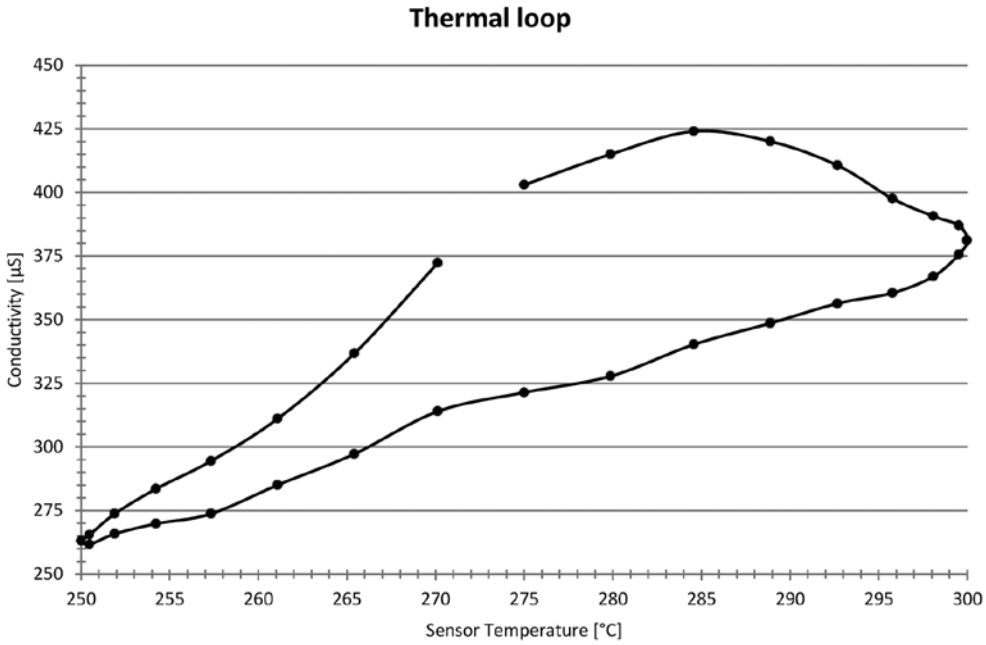


Figure 3.5. Example of a thermal loop for a thermal profile between 250°C and 300°C.

3.6, a temperature shift of 15°C, either up or down, from 275°C has limited influence when measuring butyl acetate, while the differences are much larger for hexane. Looking at the shape of the thermal loop one can imagine that the differences between both substances at 290°C are far more pronounced than at 260°C, where the hexane and butyl acetate shapes are fairly similar. It will be clear, however, that changes in temperature will heavily influence the accuracy of the classification techniques, since these are based on the shape of the thermal loop itself or descriptors derived from the shape.

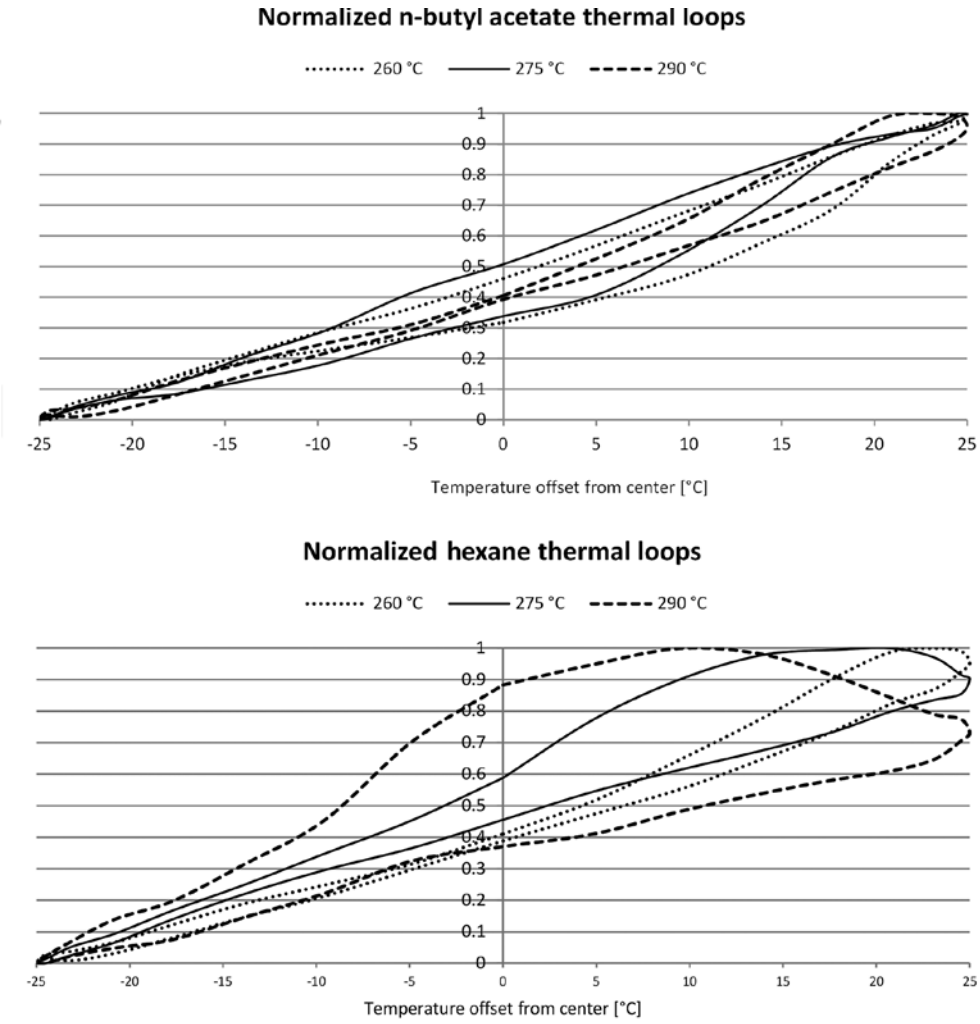


Figure 3.6. Thermal loops for n-butyl acetate and hexane at 260°C, 275°C and 290°C.

The influence of the heterogeneity of each variable

Since the sensor surface cannot be controlled and the two chemical substances are treated separately, the substance concentration and the operating temperature are the only two independent variables in this experiment. The heterogeneity caused by each variable is calculated on a per sensor basis and as an average effect to reflect the inter sensor differences. These inter-sensor differences are calculated by constructing a matrix containing the appropriate data of all sensors. This matrix has been constructed for all calculations mentioned below. The different heterogeneity graphs in figure 3.7 are all shown with the y-axis at a constant maximum value of 20 to avoid misinterpretation of the graphs and therewith the effects observed.

First we calculated the heterogeneity caused by the variance in concentration of the measured substances. A matrix is constructed for each sensor-temperature combination containing all concentrations. The heterogeneity in this matrix is, therefore, only caused by the concentration differences.

Figure 3.7a shows the calculated heterogeneity values at the different operating temperatures. It can be observed that the higher the temperature the larger the heterogeneity in the data. This is in concordance with the expected increase in reaction rate due to the increased temperature. While all separate sensors show comparable heterogeneity rates, there is some degree of variation between the sensors due to the differences in metal-oxide surface. This is reflected by the higher amount of inter sensor heterogeneity.

Next to this, we calculated the heterogeneity caused by the variance in operating temperature. A matrix is constructed for each sensor-concentration combination containing all temperatures; the heterogeneity in the matrix is only caused by the different operational temperatures.

Figure 3.7b shows the calculated heterogeneity values at the different substance concentrations. It can be observed that the influence seems related to the concentration of the substance measured; a decrease in concentration causes an increase in heterogeneity. Again this is in concordance with the expected effects since an increase in the reaction rate will cause a larger effect when there is no abundance of the measured substance and the heterogeneity decrease seems to level off with increasing concentrations. The inter sensor heterogeneity increase was considerably smaller than the one caused by the concentration variance. This is expected since the temperature of the sensor is controlled in our setup, while the differences of the sensor surfaces, largely responsible for the differences measured by the sensors, are not adjusted for. The overall increase in heterogeneity due to the

temperature influence is approximately 10-15 times as the one as caused by concentration differences when using normalized shape data.

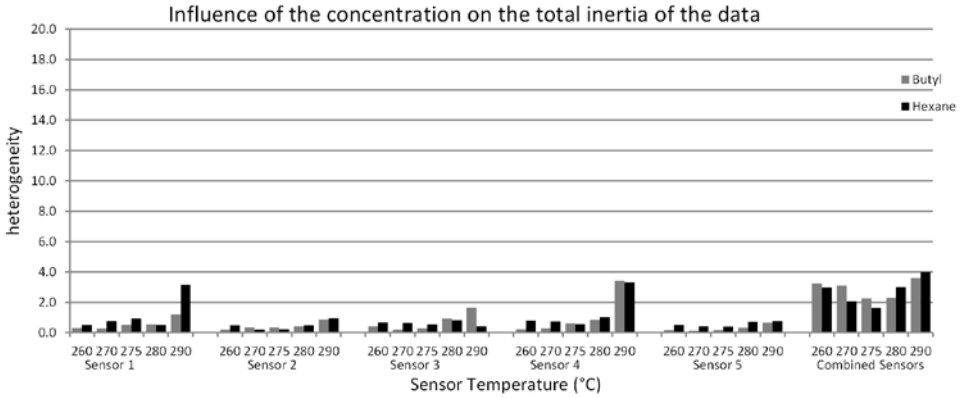
After concluding that the temperature differences have a considerably larger influence than the concentration differences, we investigated the influence of temperature shift on the data heterogeneity. Taking 275°C as the reference temperature, we calculated all possible temperature ranges. This resulted in the ranges shown in table 3.1. After that we used the 100 ppm concentration of each substance to construct a matrix for each sensor-range combination.

Range [°C]	Temperatures used [°C]
0	275
5	270-275
	275-280
10	270-275-280
15	260-270-275
	275-280-290
20	260-270-275-280
	270-275-280-290
30	260-270-275-280-290

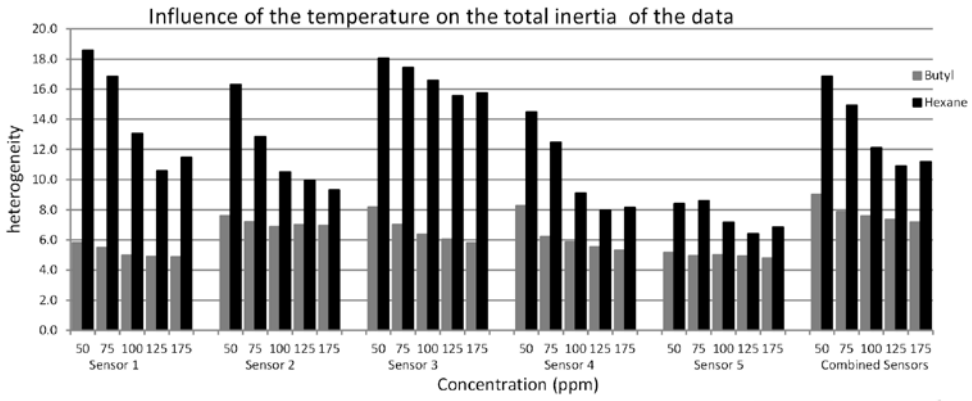
Table 3.1 The temperatures used for each temperature range.

Figure 3.7c shows the calculated heterogeneity at the different temperature ranges. It can be observed that the heterogeneity increase is roughly linear with temperature increase. On average the heterogeneity increase is 5-10% per °C. Furthermore it can be observed that the inter sensor information is elevated with an amount equal to the heterogeneity observed at a range of 5 °C. This is in concordance with the inaccuracy in the temperature control of 3-5°C.

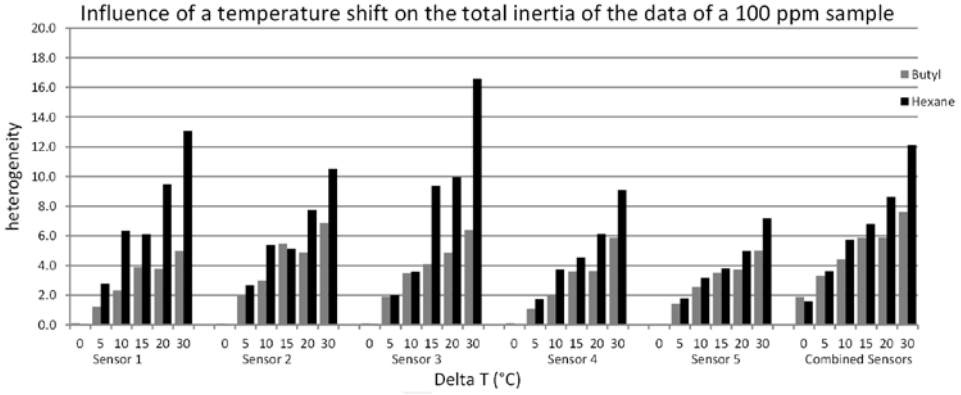
Figure 3.7.



a. The influence of the concentration on the total inertia per sensor and for the combined sensors.



b. The influence of the temperature of the total inertia per sensor and for the combined sensors.



c. The influence of a temperature shift of the total inertia per sensor and for the combined sensors.

Discussion

For nearly four decades there has been a plethora of studies describing the potential of eNose devices for various (diagnostic) applications. Although most articles promise widespread future use, no large scale implementation has been realized yet due to the difficulty in calibrating eNoses and the inability of successfully transferring calibration models between devices. There have been several efforts towards transferability of calibration models. One study compared 4 identical eNoses at 3 experimental set-ups [32] with the conclusion that the influence of both the set-up and the eNose is too large to compare the supposedly identical data. Another study compared 5 different mathematical methods [21] to translate a model for a single substance measured with 2 identical eNoses. One of the methods showed good results (23 out of 24 correctly classified) but failed to maintain the results with unknown data (only 6 out of 12 correctly classified). A third study tried to identify a model transfer function between eNoses [22]. Good results were obtained using 5 different transfer samples for 3 different substances. The number of needed transfer samples will increase when the range of expected concentrations broadens or when mixtures are measured. The calibration samples need to be measured with each new eNose making the method difficult to use in an actual production setting.

One of the largest contributors to this inability with regard to MO sensors is the lack of temperature control. Controlling the operational temperature will bring a transferable calibration model a step closer. According to data supplied by the sensor manufacturer (AppliedSensor GmbH, Reutlingen, Germany) a worst-case difference of approximately 50 °C (when operated at the same heating element resistance value) could be found between heating elements in the center and the edge of a single wafer resulting in an intrinsic variability between sensors. Since the reaction rate doubles every 10 degrees, it is obvious that the potential "off-factory" temperature difference of 50 °C will strongly affect the conductivity values and hamper transferability of calibration models especially when the sensors are operated in dynamic mode.

The results presented in this research clearly show that there is a inter sensor difference caused by the deposition process of the sensing layer, while different concentrations of analyte seem to have an effect of the same order of magnitude. The heterogeneity significantly increases due to temperature off-sets showing that the temperature is the most critical parameter to control. Fortunately accurate temperature control can be implemented by means of electronics and software.

The total inertia is a statistical measure of the differences in the data and not directly related to the separability of the data, that is how well a pattern recognition model may perform. However, it is a valuable tool as means of investigating the contributions of the variability's of various parameters to the overall variability of the measured data.

Conclusion

We show that for thermo-cycled MO sensors the intrinsic differences in heater characteristics are a major hurdle for realizing calibration model transfer and hence, large scale implementation. All reactions influencing the measured signal are strongly temperature dependent and without controlling the actual sensor-surface temperature, any combination of sensors needs to be calibrated for a specific task by using actual samples. These calibrations are time consuming and any change in sensor set-up invalidates all work done before. With simple chemicals the effort may be feasible, but more complex samples such as biological ones generally require a major effort thus defeating the purpose.

We demonstrate that strict temperature control is an essential prerequisite for obtaining reproducible results, and for developing transferable calibration models. Also, it was demonstrated that the influence of an offset in the temperature is approximately 10-15 times greater than the inter sensor morphology differences. The relationship between temperature and degree of heterogeneity is very difficult to solve using a mathematical approach only; it is substance- and mixture-specific as is demonstrated by the large differences between measurements performed on the two pure substances used in this article.

References

1. Persaud K, Dodd G – Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose – Nature 1982 299:352-355
2. Ghaddab B, et al. – Benzene monitoring by micro-machined sensors with SnO2 layer obtained by using micro-droplet deposition technique – Sensors and Actuators B 2011 152:68-72
3. Brudzewski K, Osowski S, Markiewicz T, Ulaczyk J – Classification of gasoline with supplement of bio-products by means of an electronic nose and SVM neural network – Sensors and Actuators B 2006 113:135-141
4. Eriksson A – Detection of mastitic milk using a gas-sensor array system (electronic nose) – International Dairy Journal 2005 15:1193-1201
5. Bhattacharyya N – Detection of optimum fermentation time for black tea manufacturing using electronic nose – Sensors and Actuators B 2007 122:627–634
6. Scorsone E, Pisanelli AM, Persaud KC – Development of an electronic nose for fire detection – Sensors and Actuators B 2006 116:55-61
7. Sohn, JH – Implementation of an electronic nose for continuous odour monitoring in a poultry shed – Sensors and Actuators B 2008 133:60-69
8. Berna A – Metal Oxide Sensors for Electronic Noses and Their Application to Food Analysis – Sensors 2010 10:3882-3910
9. Cano M – Rapid discrimination and counterfeit detection of perfumes by an electronic olfactory system – Sensors and Actuators B 2011 156:319-324
10. Ghasemi-Varbankhasti M, et al. – Aging fingerprint characterization of beer using electronic nose – Sensors and Actuators B 2011 159:51-59
11. Pavlou AK, Turner APF – Sniffing out the truth: Clinical Diagnosis Using the Electronic Nose – Clin Chem Lab Med 2000 38(2):99-112
12. Thaler ER, Hanson CW – Medical applications of electronic nose technology – Expert Rev Med Devices 2005 2(5):559-66
13. Ritaban D, Ritabrata D – Intelligent Bayes Classifier (IBC) for ENT infection classification in hospital environment – BioMedical Engineering On Line 2006 5:65
14. Kodogiannis VS, Lygouras JN, Tarczynski A, Chowdrey HS – Artificial odor discrimination system using electronic nose and neural networks for the identification of urinary tract infection – IEEE Trans Inf Technol Biomed 2008 12(6):707-13
15. Banday, KM, et al. – Urine volatile organic compounds could discriminate tuberculosis patients from healthy subjects – Anal. Chem 2011 83(14): 5526–34
16. Fens N, et al. – Exhaled Breath Profiling Enables Discrimination of Chronic Obstructive Pulmonary Disease and Asthma – Am J Respir Crit Care Med 2009 180: 1–7
17. Bruins M, et al. – Device-independent, real-time identification of bacterial pathogens with a metal oxide-based olfactory sensor – Eur J Clin Microbiol Infect Dis 2009 28(7): 775–80
18. Bruins M, et al. – Diagnosis of active tuberculosis by e-nose analysis of exhaled air – Tuberculosis 2012 93:232-238
19. Trincavelli M, Coradeschi S, Loutfi A, Soderquist B, Thunberg P – Direct Identification of Bacteria in Blood Culture Samples using an Electronic Nose – IEEE Trans Biomed Eng. 2010 10.
20. Oded S, Liran C, David H – On mappings between electronic noses – Sensors and Actuators B 2005 106:76–82.
21. Balaban MO, Korel F, Pdabasi AZ, Folkes G – Transportability of data between electronic noses: Mathematical methods – Sensors and Actuators B 2000 71:203-211
22. Zhang L, et al. – On-line sensor calibration transfer among electronic nose instruments for monitoring volatile organic chemicals in indoor air quality – Sensors and Actuators B 2011 160:899-909
23. An W, Yang C – Review on Temperature Modulation Technology of Gas Sensors – Applied Mechanics and Materials 2012 143-144:567-571

24. Hierlemann A – Higher-Order Chemical Sensing – Chem. Rev. 2008 108:563-613
25. Lee AP, Reedy BJ – Temperature modulation in semiconductor gas sensing – Sensors and Actuators B 1999 60:35–42
26. Pearce TC, Schiffman SS, Nagle HT, Gardner JW – Handbook of Machine Olfaction – WILEY-VCH 2003
27. Mielle P – Managing dynamic thermal exchanges in commercial semiconducting gas sensors – Sensors and Actuators B 1996 34:533-538
28. Jonda S, Fleischer M, Meixner H – Temperature control of semiconductor metal-oxide gas sensors by means of fuzzy logic – Sensors and Actuators B 34 (1996) 396-400
29. Hossein-Babaei F, Ghafarinia V – Compensation for the drift-like terms caused by environmental fluctuations in the responses of chemoresistive gas sensors – Sensors and Actuators B 2010 143:641-648
30. Bos A – Electronic Chemical Trace Detector – patent nr WO2007061294-A1
31. Yelland PM – An Introduction to Correspondence Analysis – The Mathematical Journal 2010 12:09 Wolfram Media
32. Koczuul A, et al. – Vergleich von vier baugleichen elektronischen Nasen und drei Messaufbauten – Pneumologie 2011 http://dx.doi.org/10.1055/s-0030-1256280

chapter 4

Data Analysis: Breath-profile and Biomarkers

Introduction

Dating back to the days of the ancient Greek, the smell of exhaled breath has been known for providing information of the patient's health [1]. Even today, physicians often take the smell they observe with their own nose into account when setting a diagnosis. Compared to many animals, humans have only a limited sense of smell. Therefore certain animals have been trained for identifying diseases. Dogs for example have been successfully trained to recognize different forms of cancer [2, 3] while honeybees and rats are trained to recognize tuberculosis [4, 5]. As humans, we trust their capabilities in this respect although we don't know exactly what the animal actually smells. Besides the biological nose there are eNoses which originated in the 1970's. These eNoses can be used in a similar fashion when the recorded information is processed as a whole, a so-called breath-profile or it can be used to look for specific compounds, the so-called biomarkers.

Biomarkers

Since many years, efforts have been directed at linking specific substances in exhaled air (but also in the headspace of other clinical samples such as blood, urine and wound swabs) to a specific disease, and a large number of papers have been published on this topic [6-12]. Having this direct (causal) link between volatile substances and disease potentially enables the development of accurate measurement techniques for these specific substances. Also from an industrial point of view it is valuable to elucidate the compounds that might have information in detecting disease. Such compounds, often called biomarkers, could potentially be used to improve patient treatment and can be intellectually protected in a straightforward manner with patents.

The discovery of biomarkers is a medical research field that regularly proposes new (sets of) substances as new biomarkers, but at the same time, these so-called biomarkers fail to convince in terms of sensitivity and specificity [13-15]. A physiological explanation for the occurrence of putative chemical compounds in relation to a specific disease is warranted. In many instances biomarker compounds are found for a specific condition without any explanation of a physiological or biochemical relationship between the abnormal condition monitored and the specific component. Similar substances can be related to a number of different diseases. An example of this is the presence of high levels of acetone in the breath of patients suffering from diabetes [16], but it can also be present in high levels in cases of a ketogenic diet and exercise [17] or heart failure [18].

Breath profile

The usage of a breath profile is a holistic approach circumventing the issue of identifying specific biomarkers. It searches for certain characteristic profiles that have the most discriminative information. Mapping the entire profile of metabolites and looking for unique chemical fingerprints generated by cellular and biochemical processes, is called metabolomics [9]. Similarly, an evolving field using the changes in breath is that of applying non-specific sensors [19]. In this case, the presence of a large number of VOCs in bulk causes a conductivity shift of the sensors that are used for the measurement. This approach has recently been shown by us to be successful in detecting tuberculosis using a device based upon 3 different metal-oxide sensors [20]. The shifts are thought to be related to inflammatory reaction due to the TB infection, and a sensitivity and specificity of 76·5% and 87·2 %, respectively, were found.

Deviant concentrations of chemical substances found in exhaled breath can also be indirect effects of abnormal metabolic or inflammatory conditions, e.g. a bacterial or viral infection. The body reacts on the infection and produces an inflammatory response. This, however, most likely leads to a shift in concentrations of endogenous compounds rather than the production of unique new compounds. At the same time specific compounds can be emitted by the bacteria or viruses themselves and accordingly change the breath composition. The change in breath volatile composition can be from either exogenous or endogenous origin, and thus be of importance to search for a profile or biomarker(s) with high discriminative power.

An exogenous substance, which is exclusively produced by a pathogen and does not occur endogenously, would be an unambiguous biomarker. However, to our knowledge these have not yet been reported. Most biomarkers therefore rely on shifts in exhaled concentrations of endogenous substances which are correlated with a disease by statistical methods. In this respect, both biomarker and full breath profile approaches are very sensitive regarding the methods of data handling which are full of pitfalls as has been described by Miekisch et al. [21]. With a decreasing number of subjects or increasing number of variables the chance of finding confounding variables, artefacts, and voodoo-correlations increase (a term, similar to spurious correlations, introduced by Vul et al. in 2009 [22]). In this chapter we will demonstrate how easily a non-existing correlation can be found for both cases by using pure random data, and how finding spurious correlations can be avoided.

Material and Methods

We created 10 different matrices of pure random data, uniformly distributed between 0 and 1. Each of the 10 matrices contains 256 subjects (rows) with 256 variables (columns) each, equally and randomly divided between 2 output classes. To observe the relationship between the number of subjects and amount of variables we created datasets containing 8, 16, 32, 64, 128 and 256 subjects respectively resulting in a total of 6 sets per matrix.

Since the data handling is different between both approaches we used a different method for each of the approaches. The biomarker approach is looking for one or more variables within the entire collection capable of differentiating between both classes. The breath profile approach on the other hand searches for a model that separates the classes based on the entire collection.

The dataset containing 8 subjects was converted to binary data, where 0 resembles the absence and 1 the presence of the variable, based on the generated data. This dataset was used to calculate the results for the biomarker approach.

For the breath profile results the best model was determined using a Linear Neural Network (LNN) with and without a 'Leave-One-Out cross validation' (LOOCV) approach to evaluate the population of created models. During the LOOCV a model is built using all subjects except 1, which is predicted using the built model. This process is repeated in turn until all subjects have been left out and predicted once. Without the LOOCV a single model is created using the full dataset and all subjects are predicted using this model (straight approach).

We used a LNN since it is a pure linear technique which does not require any parameter optimization. The LNN builds *N*-1 hyper-planes, where *N* is the number of classes, that separates the different input classes as well as possible. It behaves like a Linear Discriminant Analysis (LDA)[23] but is not limited by the restrictions of a LDA regarding the singularity problem.

All datasets were calculated with an increasing amount of variables (1 – 256) with their respective method. These calculations were repeated 5 times resulting in 50 runs for each subject count; the average result of the 50 runs is reported. The results reflect the amount of correctly classified subjects. In case of the biomarker approach, 'correctly classified' was determined by comparing the 0/1 of the output class with the values of the variables while in case of the breath profile the predicted outcome was compared with a threshold value of 0.5 to determine the predicted class of the subject.

Results

Biomarker approach

For the biomarker approach we counted the number of variables that predicted 5, 6, 7 and 8 of the 8 subjects correctly. In figure 4.1 these results are shown, the x-axis denotes the increasing number of variables while the y-axis denotes the percentage of variables predicting the results correctly.

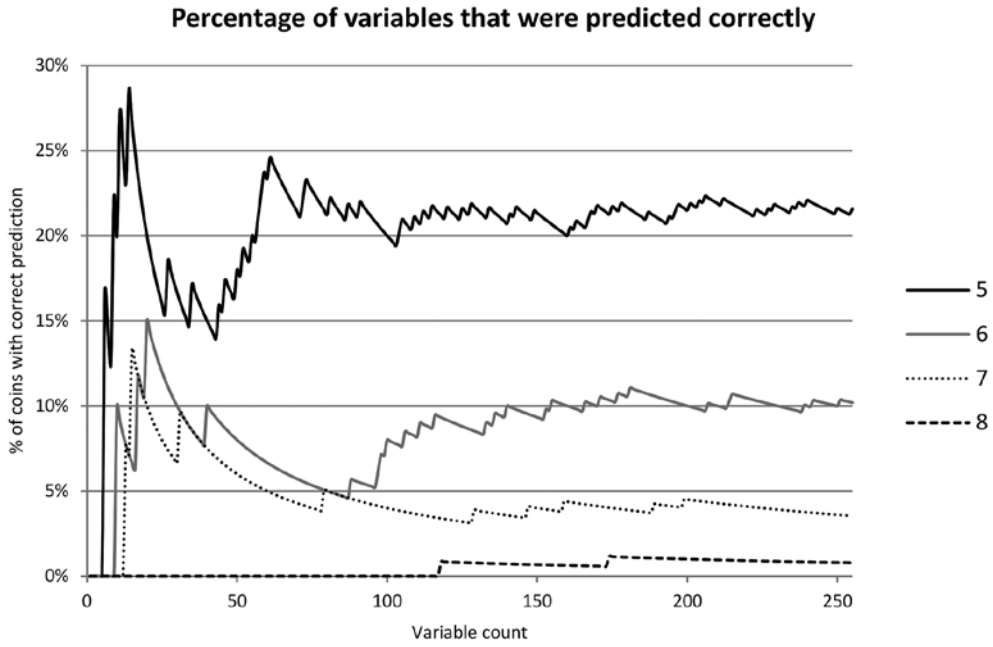
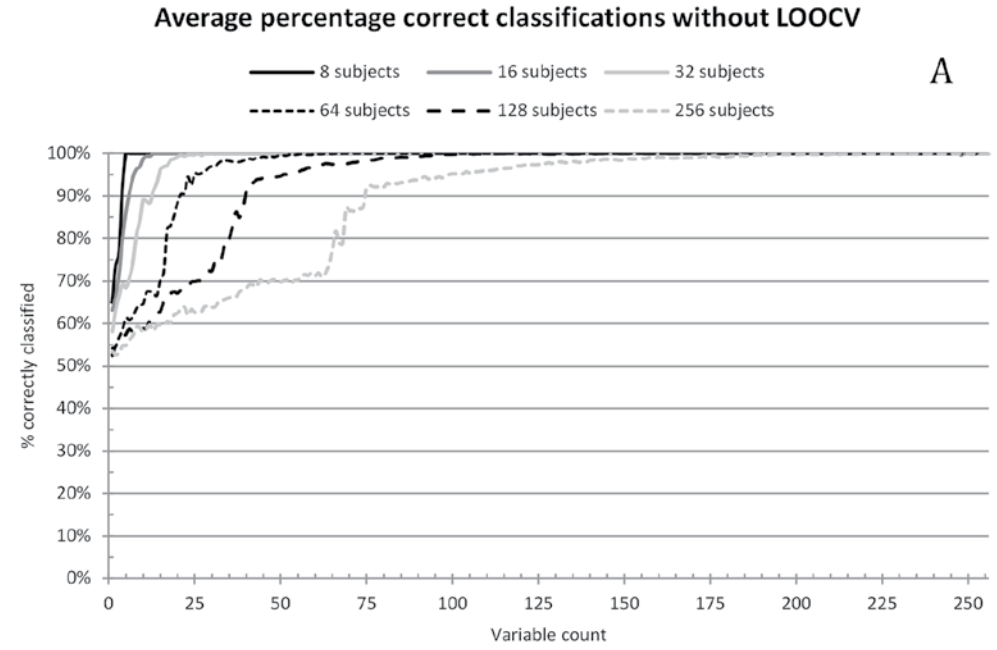


Figure 4.1. Relationship between variable count and amount of correctly predicted subjects

Breath-profile approach

All results are given as percentage correctly classified subjects so the different number of subjects can easily be compared. The chance of finding a 100% correct classification is illustrated in Figure 4.2 where the y-axis denotes the percentage of the 50 repetitions that obtained perfect separation. Figure 4.2a gives the results of the complete LNN model constructions, while the LOO-CV results are given in figure 4.2b. The graphs illustrate the relationship between the amount of variables present in the dataset and the percentage of repetitions that correctly classified all cases for a varying amount of cases per set.



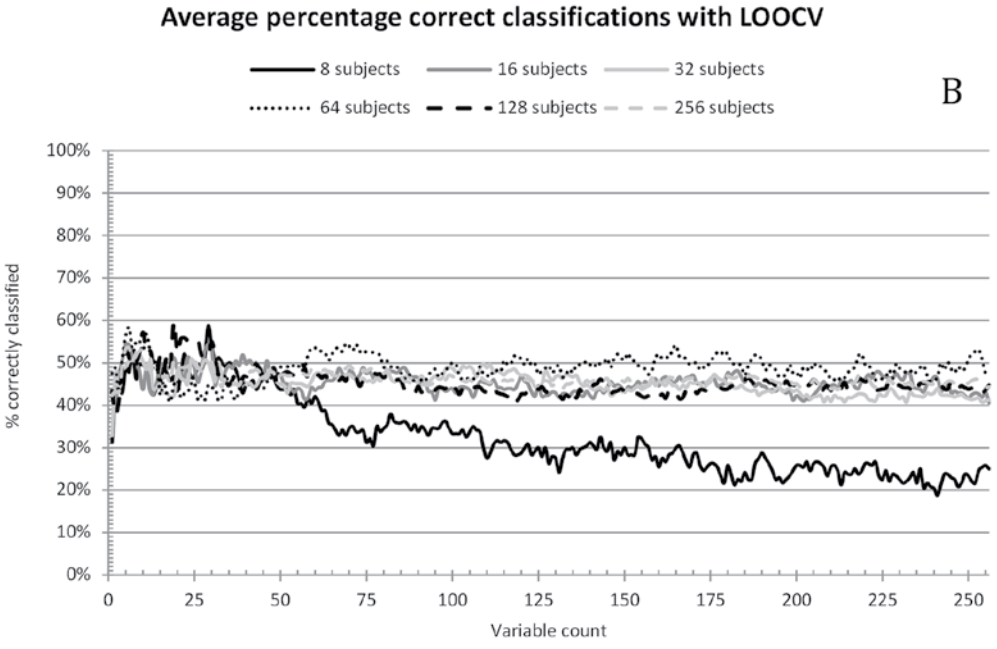


Figure 4.2. Percentage of 100% correctly classied subjects, averaged over 50 runs. Graph A shows the results without using LOOCV while graph B shows the results with LOOCV

In medical diagnostics a commonly used evaluator for a new method of proposed biomarker is the so-called Receiver Operating Characteristics Curve (ROC-curve), giving the relationship between sensitivity and specificity of a diagnostic method. From this perspective we calculated the percentage of repetitions reaching a sensitivity/specificity of 100, 95, 86 and 75% with differing number of variables. We equally divided the two classes therefore we only calculated situations where the sensitivity and specificity were equal. Figure 4.3 shows the results for the 5 different sensitivity/specificity values for LNN models performed on the entire set of 64 subjects (i.e. without CV) . The x-axis denotes the number of variables in the dataset, only the first 75 variables are given as everything above resulted in a 100% result. The y-axis in figure 4.3 denotes the percentage of the 50 repetitions with a result equal to or better than required. We used the dataset containing 64 subjects since this represented a midsized experiment. A sensitivity and specificity of roughly 90% can already be found in 50% of the cases with only 16 random variables (25% of the subjects) while the chances of finding a sensitivity and specificity of 100% occurs with 32 variables (50% of the subjects). The chances increase with an increasing ratio of numbers of variables to numbers of subjects.

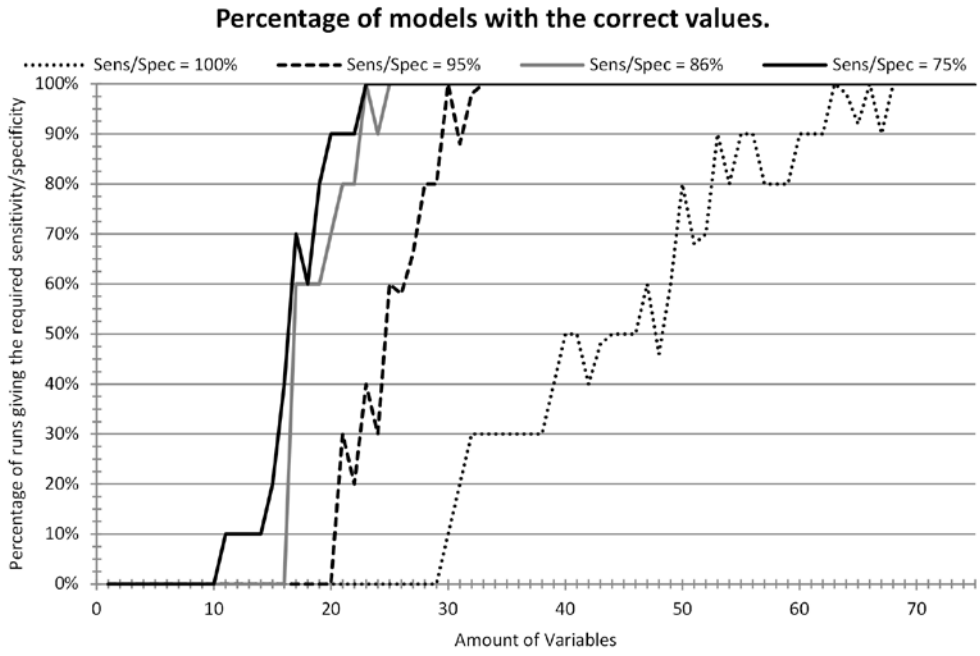


Figure 4.3. Percentage of the models with sufficient correctly classified subjects to reach the required sensitivity and specificity

Discussion

Large numbers of variables are obtained with the current technologies such as GC-MS or electronic noses with arrays of up to 32 sensors e.g. Cyranose 320. When the ratio of the number of subjects to the amount of variables measured is small, the chances of finding statistically relevant, but non-existing, correlations in the dataset (voodoo correlations) increases dramatically [22].

When one is only looking for distinctive biomarkers which can be used as identifier for a given condition, things tend to get worse [21]. The results of figure 4.1 for a set of 8 subjects demonstrate that with at least 128 random variables there is a statistical chance to find a 'magic' variable (a variable that predicts the condition perfectly). This might seem a large number of variables for just 8 subjects, but the average GC-MS spectrum easily reaches such a number of peaks. With a correct prediction rate of 87·5% (7 out of 8) a collection of 128 potential but random variables contains 4 'biomarkers' reaching this level of confidence.

The results of the calculations given in figures 4.2a demonstrate that if the number of variables is equal to 0·5 – 1 times the number of subjects, the percentage of correctly predicted subjects exceeds 90%. If the number of variables approaches the number of subjects, the percentage 'correctly predicted' cases approaches (and reaches) 100%.

Figure 4.2b demonstrates the added value of the leave-one-out approach [24]: the results found are equal or even considerably lower than 50% which is the statistical performance of simply tossing a coin to predict, and is represented by the diagonal in a typical ROC diagram [25]. For larger number of subjects the results approximate 50%. In these cases the increasing number of variables doesn't improve the results but rather stabilizes it. In other words: when dealing with random data with a leave-one-out approach, the larger the number of variables, the closer one statistically gets to the performance of tossing a coin.

This can be explained by the nature of the LNN [26,27] combined with the leave-one-out approach: a LNN builds a hyper-plane separating the input space into two parts. If more variables are introduced, all dimensions of the 'unknown' (left out) subject must be orientated on the correct side of the derived hyper-plane. With increasing number of variables, the chance of all variables being at the correct side of the hyper-plane diminishes resulting in a smaller chance of randomly finding a proper model. This may seem counter intuitive since the common belief is that the chances of finding a proper model are directly proportional to the number of variables. This belief is, however, based on the usage of LNN (or any other sophisticated algorithm) without a leave-one-out approach. Most multivariate algorithms are capable of constructing complex hyper-planes which profit from an increasing number of variables. Using leave-one-out, a perfect, complex hyper-plane based on the

training data will most likely not be able to classify an unknown sample correctly, and the chance of being misclassified increases with the number of variables taken into account. It can be concluded that a hyper-plane method combined with leave-one-out is advantageous in situations where the number of variables is in the same order of magnitude, or exceeds the number of subjects since it renders the chances of finding a 'voodoo-model' negligible. If a proper model is found when using a hyper-plane method combined with leave-one-out with a high ratio of variables with respect to the number of subjects, one can be sure there is an actual correlation underlying the separation.

The results in figure 4.3 were obtained without using a leave-one-out approach. When using leave-one-out, however, the ROC-curve of random data will never be significantly better than 50% (the diagonal in the ROC plot), independent of the ratio of number of subjects to variables.

It is very important to validate initial promising results and the performance of discovered biomarkers in independent study patients. The measurement set-up must be designed carefully in order to avoid possible confounding variables between positive and negative subjects[28]. Aspects which should be taken into account to avoid artefacts are:

1. Healthy and sick people should be recruited from the same social and economic population [20],
2. Preferably measurements should taken at different times of the day,
3. Measurements must take place at several locations in order to compensate for environmental variables (e.g. the use of Lysol in hospitals is recorded in the exhaled breath and could disturb the measurement results),
4. Several subject properties could influence the measurements: age, smoking, diet, co-morbidities etc. When these variables are however truly random distributed between both groups (negative versus positive or patients versus controls), they will principally not influence selection of profiles based on classifying algorithms.

Some diseases lack a diagnostic golden standard [29]. This will definitely influence the quality of the models obtained as the model prediction value (the result of the standard method) itself contains errors (false positives and false negatives) which hamper the derivation of correct hyper planes in classification models.

Conclusion

When using a sufficiently large number of subjects, the risk of obtaining a statistically correct but non-existing prediction model due to confounding variables, artefacts, and statistical errors can be minimized. Ideally, a study should be divided in a training population to build the model and a validation population to validate the model. If the number of subjects is small, or when the ratio of obtained measurement variables to number of subjects is relatively large, a leave-one-out approach should be used.

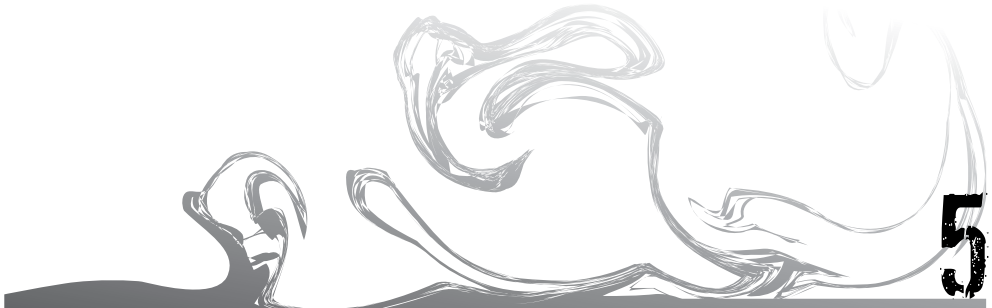
Given all of the above, performing a blind study is the proper way to ascertain that a sufficient number of subjects have been measured and that the derived model is correct.

In the present simulations, with purely random data we demonstrated that one typically needs minimally twice, but preferably four times as much subjects as measured (available) variables. Since most modern analysis techniques produce a vast number of variables, it will be next to impossible to obtain a sufficient number of subjects. We also demonstrated that using a leave-one-out approach during data processing will negate this requirement and even decrease the probability of finding false correlations with an increased number of variables.

References

1. Marks E – The aphorisms of Hippocrates from the latin version of Verhoofd – Collins & Co 1817
2. McCulloch M, et al. – Diagnostic Accuracy of Canine Scent Detection in Early- and Late-Stage Lung and Breast Cancers – Integr Cancer Ther. 2006 5(1):30-9.
3. Willis CM, et al. – Olfactory detection of human bladder cancer by dogs: proof of principle study – BMJ 2004 329:712
4. Suckling DM, Sagar RL – Honeybees Apis mellifera can detect the scent of *Mycobacterium tuberculosis* – Tuberculosis (Edinb) 2011 91(4):327-8
5. Mgode GF, et al. – Ability of Cricetomys rats to detect *Mycobacterium tuberculosis* and discriminate it from other microorganisms. – Tuberculosis (Edinb) 2012 92(2):182-6
6. Dalle-Donne I,Rossi R, Colombo R, Giustarini D, Milzani A – Biomarkers of Oxidative Damage in Human Disease – Clinical Chemistry 2006 52(4):601–623
7. Kharitonov SA and Barnes PJ – Biomarkers of some pulmonary diseases in exhaled breath – Informa Healthcare 2002 7(1):1-32
8. Meinardia S, Jinb K, Barlettaa B, Blakea DR, Vaziric ND – Exhaled breath and fecal volatile organic biomarkers of chronic kidney disease – Biochimica et Biophysica Acta 2013 1830(3):2531–2537
9. Boots AW, et al. – The versatile use of exhaled volatile organic compounds in human health and disease – J. Breath Res. 2012 6:027108
10. Penn DJ et.al – Individual and gender fingerprints in human body odour – J. R. Soc. Interface 2007 4:331–340
11. Matsumura K et.al –Urinary Volatile Compounds as Biomarkers for Lung Cancer: A Proof of Principle Study Using Odor Signatures in Mouse Models of Lung Cancer. – PLoS ONE 2010 5(1)
12. Kohl I – Non invasive detection of renal function via breath gas analysis : A potential biomarker for organ acceptance? – 6th International Conference on Proton Transfer Reaction Mass Spectrometry and its Applications 2013
13. Thompson IM et.al – Prevalence of Prostate Cancer among Men with a Prostate-Specific Antigen Level <=4.0 ng per Milliliter – N Engl J Med 2004 350:2239-2246
14. Thompson IM, et al. – Operating characteristics of prostate-specific antigen in men with an initial PSA level of 3.0 ng/ml or lower – JAMA 2005 294(1):66-70
15. Pin E, Fredolini C and Petricoin – The role of proteomics in prostate cancer research: Biomarker discovery and validation – Clinical Biochemistry 2013 46(6):524-538
16. Galassetti PR, et al. – Breath Ethanol and Acetone as Indicators of Serum Glucose Levels: An Initial Report – Diabetes Technol Ther. 2005 7(1):115-23
17. Španel P, Dryahina K, Rejšková A, Chippendale TW, Smith D – Breath acetone concentration; biological variability and the influence of diet – Physiol Meas. 2011 32(8):23-31
18. Marcondes-Braga FG et.al – Exhaled acetone as a new biomarker of heart failure severity – Chest. 2012 142(2):457-66
19. Pearce TC, Schiffman SS, Nagle HT, Gardner JW – Handbook of Machine Olfaction – WILEY-VCH Verlag GmbH Co. KGaA, Weinheim 2003 ISBN: 3-527-30358-8
20. Bruins M, Rahim Z, Bos A, van de Sande WWJ, Endtz HPh, van Belkum A – Diagnosis of active tuberculosis by e-nose analysis of exhaled air – Tuberculosis (Edinb) 2013 93:232-238
21. Miekisch W, Herbig J, Schubert JK – Data interpretation in breath biomarker research: pitfalls and directions – Breath Res 2012 6(3):036007
22. Vul E, et al. – Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition – Perspect. Psychol. Sci. 2009 4:274–90
23. Fishe r RA – The Use of Multiple Measurements in Taxonomic Problems – Annals of Eugenics 1936 7 (2): 179–188
24. Arlot S and Celisse A – A survey of cross-validation procedures for model selection – Statist. Surv. 2010 4:40-79

25. Fawcett T – An introduction to ROC analysis – Pattern recognition letters 2005 27:861-874
26. Eberhart RC and Dobbins RW – Early neural network development history: The age of Camelot – IEEE Engineering in Medicine and Biology 1990 9:15-18
27. Hinton GE – How neural networks learn from experience – Scientific American 1992 267:144-151
28. Hennekens CH, Buring JE and Mayrent SL – Epidemiology in Medicine – Little Brown & Co, London UK, 1987 ISBN:0-316-35636-0
29. Spiegelman D, Schneeweiss S and McDermott A – Measurement Error Correction for Logistic Regression Models with an "Alloyed Gold Standard" – American Journal of Epidemiology 1997 145(2):184-196



chapter 5

Data Analysis: Data compression

Introduction

Another method of reducing the chances of finding voodoo correlations, as described in chapter 4, is the data compression, part of the pre-processing step described in 1.7.3.

We used a compression method specifically developed for our data. It has been developed in collaboration with Elegant Mathematics (Ottweiler, Germany) and is based on the principle of tensor decomposition. This chapter explains the basic principles behind this principle based on a Aerekaprobe measurement with 12 sensors.

Tensor decomposition is a method which describes the data based on all relations present in the data set. These relations may also include the inter-sensor relations allowing it to use measurements comprised of multiple sensors of a device as single input instead of using each signal (sensor) separately.

Each sensor within the Aerekaprobe produces a matrix of M*32 data points per sample where M is the amount of measurement cycles conducted during the exposure and recovery phase and 32 the number of steps within each thermal cycle. Each measured sample produces thus a set of twelve matrices as output. As this technique generates a very large amount of data for a single experiment, the objective is how to extract the maximum amount of information from this full single-experiment data set (figure 5.1) and getting it in the form of a (one-dimensional) vector which can be used as input in mainstream pattern recognition algorithms.

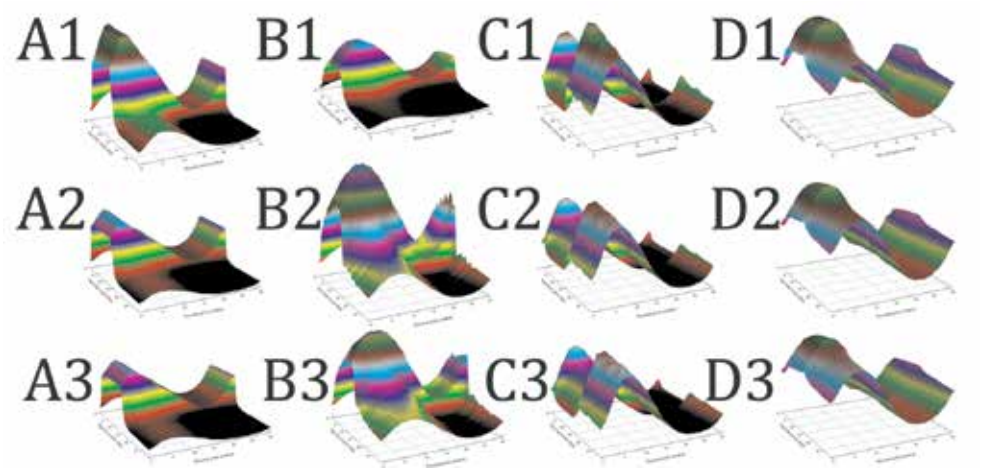


Figure 5.1. Schematic representations of the data generated by the Aerekaprobe for a single experiment. Each sensor generates a three dimensional graph as output. The x-axis of each graph depicts the 32 steps of the temperature cycle of the metal-oxide sensor, the y-axis depicts the measurements (the total of the exposure and recovery) and the z-axis depicts the response values (amplitude) of the sensors as function of temperature cycle and time. With 12 sensors, triplicates of 4 chemical types, in the Aerekaprobe, a set of 12 data files make up one measurement (sample/subject).

The most straightforward method would be to 'linearize' the data set by writing all data in one huge vector. This will, however, result in vectors of which the number of elements are orders of magnitude larger than the number of subjects in the dataset, which in turn most likely will result in 'voodoo correlations' as described in chapter 4.

Tensor decomposition

As described above, the data generated by a Aerekaprobe for each measurement form a collection of 3-dimensional figures. There is a certain degree of coherence between adjacent points in these figures and thus a fair amount of redundancy in the raw data. With the tensor decomposition the full data set is analyzed and the most likely fundamental characteristics (called the 'kernels') are estimated. The data set is then decomposed into these kernels and the 'scores' (the numbers needed for the kernel). For each kernel there will be a 'rank' (the number of scores needed to describe that characteristic). The kernels are shapes of the same dimensionality as the raw data. In the case of Aerekaprobe data, there are either 3 of 4 dimensions. Thermal cycle, time and amplitude are the first 3 dimensions, while the chemical type of the sensor may provide a fourth dimension if the samples are simultaneously measured by several sensors. In a sense one can view the kernels as a new axis system and the scores as the coordinates on these axes. However, a single axis (kernel) can have multiple scores if the rank in that dimension is larger than 1.

The decomposition generates the kernels and the scores for each of the measurements. The kernels can be stored in files for later use, for example transforming newly obtained data to the same space as the data used when evaluating the original data set. This transformation involves a set of matrix multiplications and is basic linear algebra. The scores generated for each measurement can be used as input for a pattern recognition technique as they are 1-dimensional (each measurement generates a single vector) and of a low dimensionality.

In the following example this is illustrated with a trivial two-dimensional example.

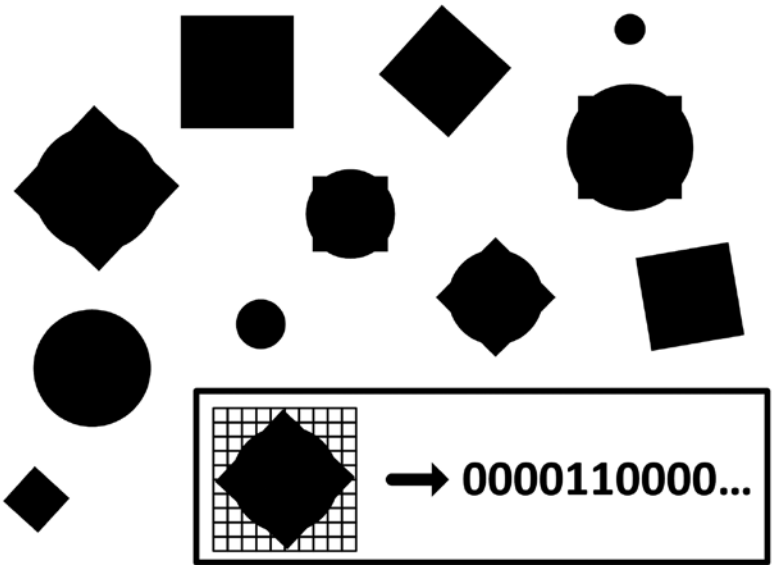


Figure 5.2. A set of shapes. The insert illustrates the rasterizing of a figure to obtain a vector.

Consider the shapes in figure 5.2. If these shapes would be our input data how could they be correlated to a target? A straight forward naïve approach would be to rasterize each shape and feed the resulting vector of 0's (empty) and 1's (filled), from left to right and top to bottom, into the pattern recognition technique as depicted in the insert. However, there are few shapes and the resulting rasterized vectors will contain many more elements than there are shapes increasing the risk of over fitting and voodoo correlation enormously. Also, the results will depend for a great deal on the chosen grid-size of the raster. A very fine grid will include the small features as in the lower right shapes but also result in a very large vector for each shape. A more course grid will reduce the size of the vector but lose the finer details.

The problem of the shapes in figure 5.2 can be solved in a more elegant way. For a human it is obvious that the shapes are all defined by two basic shapes, a square and a circle. Some shapes are mere squares, others are mere circles and some are composed by superposing a square and a circle on top of each other. In case of the squares one needs to determine the size and rotation-angle, in case of the circles only the diameter. Given these two basic shapes, each shape can thus be represented by a vector of three numbers. Two of

the numbers describe the size and rotation angle of the square part and the third describes the diameter of the circle part. For a pure square shape the number of the circle part will be zero and vice versa. Only the combined shapes will have non-zero values for all three elements of the vector.

With this approach we can reduce the information of each of the shapes of figure 5.2 into a vector of only three elements, the rotation angel of the square, the size of the square and the diameter of the circle. At the same time we have captured even the finest detail. The result in this example would be a matrix of 11 rows (11 shapes) and 3 columns. The redundancy of the data has been greatly reduced without loss of detail and thus maximizes the chance of obtaining a good and robust pattern recognition model. The square and the circle in this example are the 'kernels', the 'rank' of the square kernel is two while the rank of the circle is one. And finally the 'scores' are the size and rotation angle of the square and the diameter of the circle.

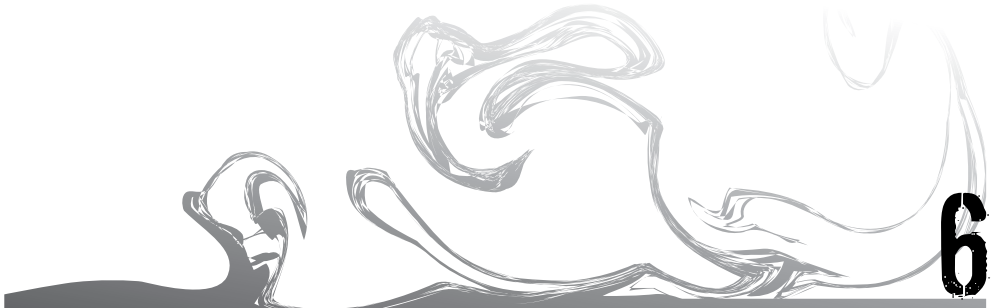
The proprietary algorithm we use is specifically developed and optimized for the eNose data generated by our devices and is a hybrid combination of PARAFAC[1] and TUCKER3[2]. The algorithm decomposes three dimensions in TUCKER space and one dimension using the PARAFAC algorithm. During the decomposition the data can be optionally scaled and noise filtered as well.

The advantages of the compression method are manifold. First of all the compression takes all possible dimensions of the dataset into account as a whole (including their inter-relations), and generates characteristic vectors with minimal loss of data. Secondly, in the decomposition scaling and removal of unwanted components (amplitude and noise) can be taken care of in a structured way. Thirdly, computation time in the subsequent pattern recognition phase is reduced enormously as the output of the compression is only a fraction of the original raw data.

References

1. Rasmus Bro, R – PARAFAC. Tutorial and applications – Chemometrics and Intelligent Laboratory Systems 1997 38:149–171
2. Henrion R – N-way principal component analysis Theory, algorithms and applications – Chemometrics and Intelligent Laboratory Systems 1994 25:1–23

Metal-oxide sensors – Diagnostic potential



chapter

6

Real-time identification of bacterial pathogens

M. Bruins
A. Bos
P.L.C. Petit
K. Eadie
A. Rog
R. Bos
G.H. van Ramshorst and
A. van Belkum

Partially published in
Eur J Clin Microbiol Infect Dis **2008** 28:775-780

Device-independent, real-time identification of bacterial pathogens with a metaloxide-based olfactory sensor

Abstract

A novel olfactory method for bacterial species identification using an electronic nose device called the MonoNose was developed. Differential speciation of micro-organisms present in primary cultures of clinical samples could be performed by real-time identification of volatile organic compounds produced during microbial replication. Kinetic measurements show that the dynamic changes in headspace gas composition are orders of magnitude larger than the static differences at the end of fermentation. Eleven different, clinically relevant bacterial species were included in this study. For each of the species 2 to 8 different strains were used to take intra-species biodiversity into account. A total of 52 different strains were measured in an incubator at 37°C. The results show that the diagnostic specificities varied from 100% for *Clostridium difficile* to 67% for *Enterobacter cloacae* with an overall average of 87%. Pathogen identification with a MonoNose can be achieved within 6-8 hours after inoculation of the culture broths. The diagnostic specificity can be improved by broth modification to improve the VOC production of the pathogens involved.

Introduction

Bacterial identification in the medical microbiology laboratory is still firmly based on old-fashioned biochemical reactions. From the ages of Pasteur and Koch onwards, medical microbiologists have relied on classical culture-based methods in order to confirm bacterial infections and identify the pathogens involved. Most infectious agents are still primarily detected using classical methods that involve liquid or solid semi-synthetic growth media [1]. Serological methods for detection of such pathogens are often lacking in sensitivity and specificity and data generated by the novel generation of expensive molecular tests need to be interpreted with equal caution [2]. Even so, given the clinical impact of bacterial infections in general, methods that reliably speed up the diagnostic process and limit costs are still eagerly awaited. Culture has the major advantage that living organisms are obtained for downstream characterization including antimicrobial susceptibility testing and epidemiological typing of the organism. This would argue in favour of new methods that combine classical culture with procedures that enhance the speed of bacterial species identification. Currently, microbiology laboratories employ classical growth-based, fermentative species-identification schemes that can be performed either manually or in an automated fashion using instruments such as the bioMérieux VITEK© or the Becton-Dickinson Phoenix©. Molecular nucleic acid amplification tests and biophysical procedures including mass spectrometry could encompass novel, more real-time methods for bacterial species identification, although this would require additional handling of positive cultures

[3]. In conclusion, real-time species identification during primary cultivation of clinical samples would be of added value for controlling costs and optimizing patient care in clinical institutions. Odour-based assays could potentially fill this diagnostic niche. Classification of micro-organisms on the basis of odour production has recently increased in general interest due to the introduction of so-called eNose devices.

Differential speciation of micro-organisms present in primary cultures of clinical samples could be performed by real-time identification of volatile organic compounds (VOCs) produced during microbial replication. Classical culture has been combined with such olfactory measurements in the past [4]. It needs to be emphasized that these experiments always involved cultivation endpoint measurements. However, by continuous sampling of a culture's headspace, the kinetics of the synthesis of VOCs can be monitored in real time. The complex bacteriological VOC signals can be defined by a gas measurement technique involving metal-oxide (MO) sensors as used in the present study.

We have developed a broadly applicable, inexpensive and highly responsive sensor system, called the MonoNose, which uses real-time VOC pattern recognition and matching of previously measured dynamic olfactory patterns with previously identified reference patterns. This system is analogous to the kinetics of human odour recognition [5]. The MonoNose technology facilitates the timely identification of bacterial species and, thereby, the clinical differentiation of medically relevant pathogens. We here show that several bacterial species can be distinguished on the basis of MonoNose-mediated identification of growth stage-specific VOC production.

Materials and Methods

eNose (MonoNose)

A set of thirty custom designed eNose devices were manufactured by The eNose Company (Zutphen, The Netherlands). These MonoNoses employ a single metal-oxide type semiconductor gas sensor. A scheme of a MonoNose and the thirty hand-manufactured devices in operation in an incubator is depicted in Figure 6.1.

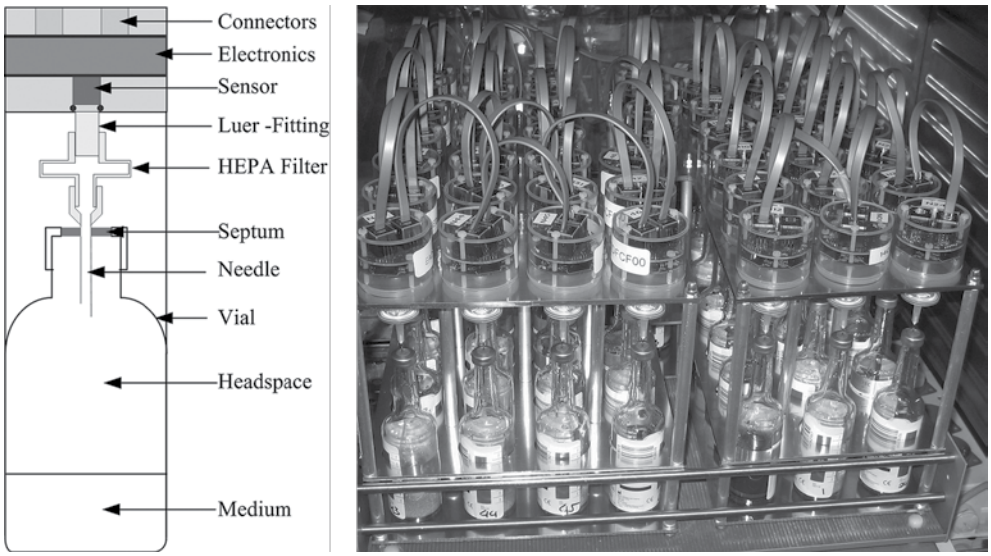


Figure 6.1.A. Schematic representation of a MonoNose device for measuring bacterial VOC production in the broth' headspace over prolonged periods of time. The sensor is a commercially available metaloxide-based micro-device. **B**: Experimental setup with thirty MonoNose devices in operation in an incubator. The sensors are serially connected and all data is assembled in a simple portable computer. The vials in the photo are standard BD-BACTEC™–Plus-Anaerobic/F disposable bottles.

For each experiment a MonoNose device was fitted with a disposable sterile syringe needle and a sterile HEPA filter with 45μ pore size to prevent cross-contamination. Humidity influences the semiconductor response values [6]. All devices were operated in an incubator at constant ambient temperature of 37°C in order to keep the relative humidity in the headspace constant during all experiments.

Bacterial species

Eleven different, clinically relevant bacterial species were included in this study. For each of the species 2 to 8 individual, genotypically distinct strains were used to take intra-species biodiversity into account. The 52 strains used are listed in Table 6.1. The strains were obtained from both commercial sources (American Type Culture Collection; ATCC) and Erasmus MC reference collections. For all strains species' nature was reconfirmed by VITEK© analysis before inclusion in the present study. Culture broths (50 ml) were inoculated with 100 µl of a physiological salt solution with a bacterial density of ±1.5 McFarland and linked to the MonoNose.

Species	*Strains*
Clostridium difficile	AN59, AN63
Enterobacter cloacae	ATCC13047, B33386, B33449
Enterococcus faecalis	ATCC19433, ATCC29212, ATCC51299, ATCC7080, P794805, VanB, B33432, B33438
Escherichia coli	31972, 31995, ATCC25922, ATCC35218, B31938,ATCC35150(ETEC)
Klebsiella oxytoca	32341, ATCC700324, B33516, B33510, F54, loes
Klebsiella pneumoniae	ATCC13882, ATCC13883, ATCC35657, ATCC700603, P79789
Proteus mirabilis	ATCC25933, ATCC7002, B33183, B33505, B33546
Pseudomonas aeruginosa	ATCC10145, ATCC27853, ATCC9027, P798326
Salmonella enteriditis	ATCC13076, mrcl
Salmonella typhimurium	ATCC13311, ATCC14028, ATCC49416
Staphylococcus aureus	857S, 863S, 865S, 920S, ATCC25923, ATCC29213, ATCC29737, ATCC33862

Table 6.1. Overview of bacterial species and strains used

Culture broths and chemicals

A broad range of commercially available culture broths were tested to identify those with maximum bacterial VOC production. These explorative tests were conducted with a single strain per species. The broths were also spiked with additional chemicals. These additions consisted of 1 ml of a solution with a concentration of 0·1, 1·0 or 5·0 mM of for instance metal salts (ammonium nickel(II) sulphate, chromium(III) potassium sulphate, cobalt(II) sulphate, copper(II) chloride, copper(II) sulphate, iron(III) chloride, lithium acetate, manganese(II) sulphate, molybdenum(IV) sulphide, silver nitrate and zinc sulphate), fatty acids (butanoic acid, caprioic acid, lactic acid, propianic acid and valeric acid) and alcohols (butanol, ethanol, methanol and 2-propanol). All separate chemicals were purchased from Sigma (Dordrecht, The Netherlands). Since Escherichia coli and Klebsiella oxytoca strains produce VOCs that are efficiently detected by the MonoNose and, hence, produce "strong" signals, these two bacterial species where used to examine the influence of amino acid and NaCl concentrations by varying the concentrations between 2·5 and 7·5 $g.l^{-1}$ and 1 and 9 $g.l^{-1}$ respectively.

Data processing

During each experiment a set of 2200-2500 data points was collected (corresponding with an incubation of 12-14 hours) starting within 15 minutes from the moment of inoculation of the medium. To test the device and time independency all strains were measured three times on different days and different devices. The measured data are pre-processed and the resulting times-series data are analyzed using a Sliding Window – Minimum Variance Matching adaptation of the Dynamic Time Warping algorithm [7,8].

Bacterial Fingerprints in time

The identification of the microbial species performed in this chapter is based on so-called bacterial 'fingerprints'. The eNose generates VOC redox reaction responses in real-time, which means that looking at the entire 'odor-movie' during microbial growth may provide some specific "scenes" (specific changes in VOC patterns over time), that characterize a particular microbial genus or species present in the clinical specimen being tested. Moreover, this process may be aided by searching for specific combinations of patterns obtained at different time periods. These unique patterns may be useful in determining the 'fingerprint' of a particular microorganism present within a clinical sample. Furthermore, once these unique fingerprints have been identified, there is no need to continue with

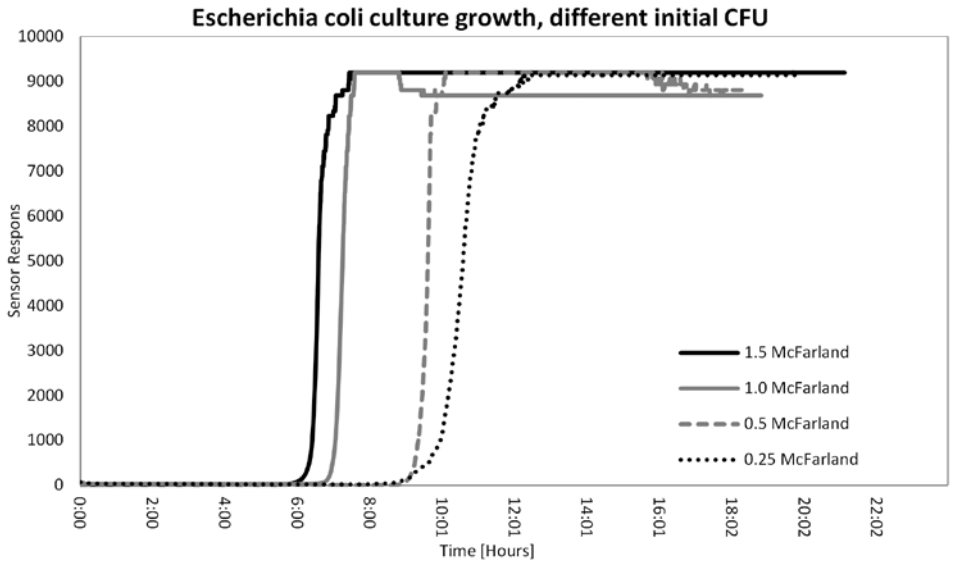


Figure 6.2. The influence of initial microbial inoculation size on VOC headspace pattern dynamics. The x-axis shows the amount of time passed since inoculation (in hours) and the y-axis shows the conductivity of the sensor in µS. A McFarland unit is a measurement of turbidity where 0.25 McFarland unit corresponds to approximately 0.75x 108 ^{cells}/ml, and 1.5 McFarland units 4.5 x 108 ^{cells}/ml (http://en.wikipedia.org/wiki/McFarland_standard).

extended headspace monitoring. The actual dynamic part of odor detection occurs over a relatively small period of time as can be seen in Figure 6.2.

This characteristic period can begin at any time throughout the measurement process, as the initial microbial inoculation concentration within the clinical specimen being tested is not known. The identification of a microorganism may however be based on specific fingerprint patterns so that the eNose monitoring device can be set up to continually scan for the emergence of relevant fingerprint patterns to ensure the shortest possible time-to-diagnosis (identification time).

The matching of a predetermined fingerprint for *E. coli* in a full experiment is shown in Figure 6.3. The insert shows the predetermined fingerprint. The sample was continually scanned for the emergence of specific motifs using dynamic time warping algorithm software. In fact, such algorithms are designed to scan for similarities instead of exact matches. This is because even a small change in temperature will generate a small change in the signal obtained, though the overall fingerprint pattern similarity may or will be maintained. This is clearly shown in Figure 6.3, where the inset shows a predetermined *E. coli* fingerprint, and the main figure shows the recoded response of a test specimen culture. The fingerprints

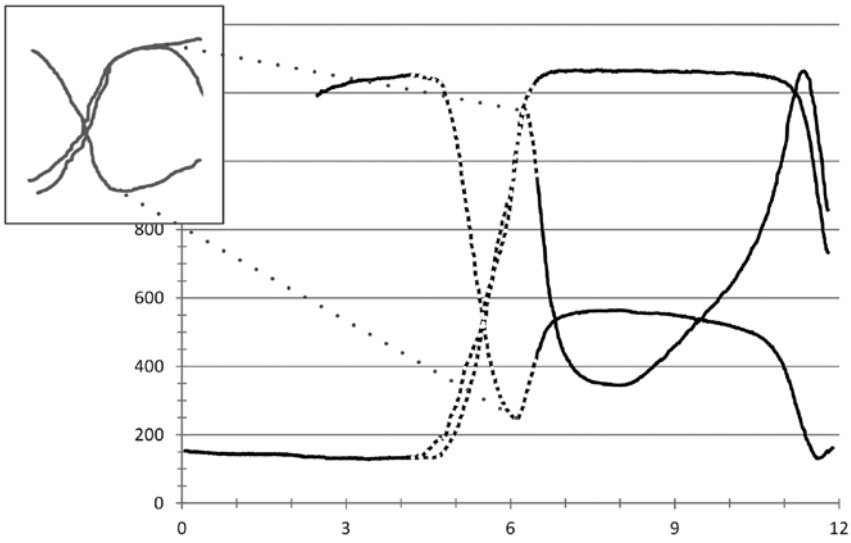


Figure 6.3. A MonoNose real-time pattern obtained using *Escherichia coli* ATCC35218 grown in Todd Hewitt Broth medium for 12 hours. The position of the predetermined fingerprint pattern for Escherichia coli (shown in the inset) is marked in the experiment. The X-axis shows the time passed since inoculation and the y-axis shows the normalized descriptor values. The normalized descriptor value represents 3 different descriptors derived from the signal recorded after applying a thermal heating profile. The descriptors used are the slope of the signal, the surface area of the signal and the difference between the maximum and minimum value in a thermal cycle. All of these values are normalized to 1.

look similar (and can be identified using dynamic time warping algorithm software) but are not identical. This fingerprint profile is composed of 3 different descriptors derived from the signal recorded after applying a thermal heating profile. The descriptors used are the slope of the curve, the surface area of the signal and the difference between the maximum and minimum value in a thermal cycle. All of these values are normalized to 1.

Results

Culture broths

The results of the exploratory tests with various broths revealed that the composition and dynamics of the headspace were co-determined by both the broth composition as well as the bacterial species in question. Although the different chemical additives caused changes in the produced VOCs, only one of the tested chemicals gave a significantly better result than was obtained in the standard broth. All other additions either generated no improvement at all or only improved the signal obtained for one species while deteriorating others, which invariably resulted in a worse overall result. These tests do, however, clearly reveal that the production of VOCs can be influenced by changes in the media as the observed features change with the different additives. The BD-BACTEC™–Plus-Anaerobic/F Medium with an addition of 0·1 mM $FeCl_3$ was selected to perform the high-throughput measurements as this medium gave the best discriminating abilities of the tested broths.

Among the bacterial species tested two response modes could be identified. Species either generated strong signals, probably due to the production of large amounts of detectable VOCs, or the signals remained low during the entire measurement process. The group of species provoking strong signals will be further referred to as the “strong” group, suffered from a squashing effect due to particulars of the measuring electronics: signal curves relatively quickly topped-off due to saturation of the measurement signals.

Amino acid and salt concentrations showed a clear optimum which differed for the species tested. The surface plots showing the results are depicted in Figure 6.4.

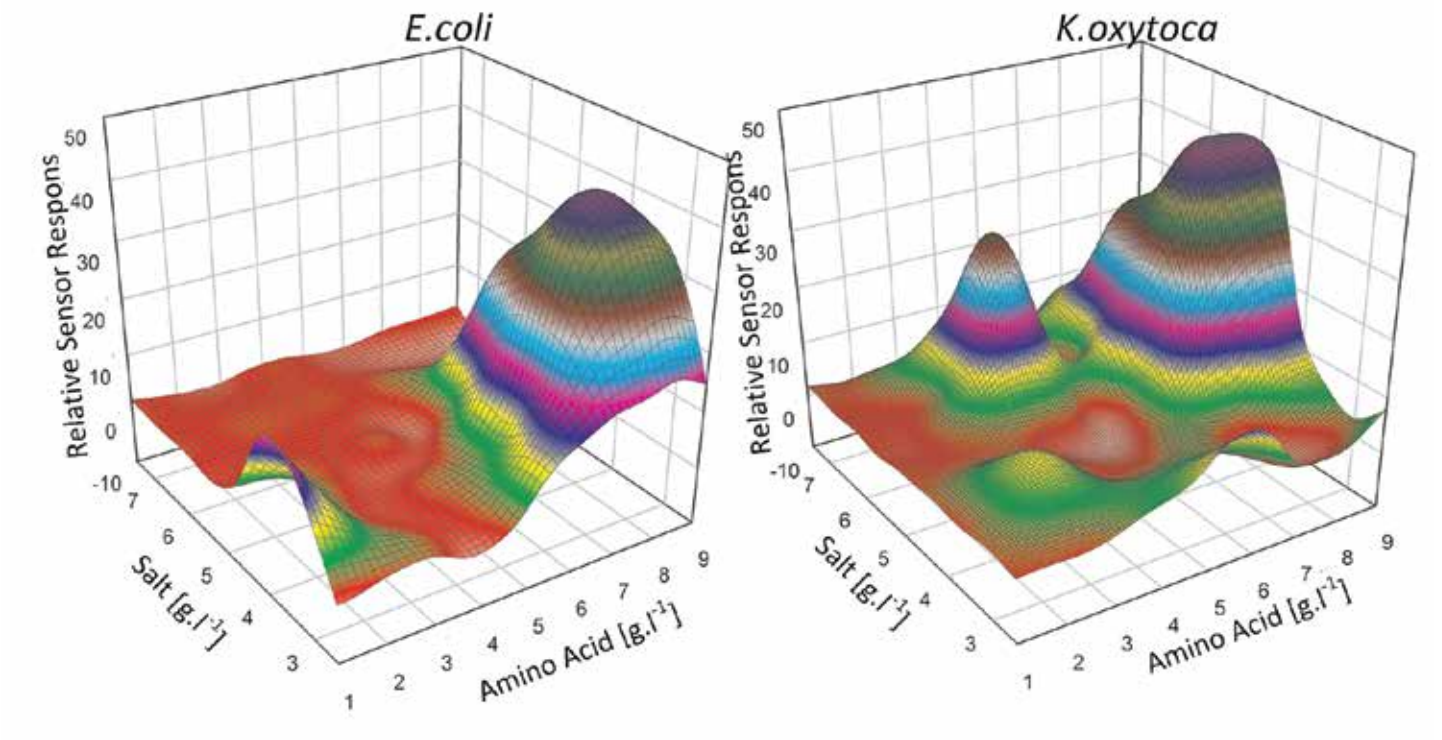


Figure 6.4. The surface plot of the maximum sensor amplitude for *E. coli* and *K. oxytoca*, as function of the combined amino acid and NaCl concentrations. Maxima and minima are clearly shown as is the difference in optimal concentrations for these two species. Note that the major difference is reflected by a bi-modal response curve for *K. oxytoca*.

Bacterial species identification

Bacteria were cultured in the BD-BACTEC™–Plus-Anaerobic/F Medium with addition of 0·1 mM $FeCl_3$ and production of VOCs was measured continuously with the MonoNose. Figure 6.5 illustrates typical experimental results. Figure 6.5-A is an example of a strong response in which the maximum amplitudes of the sensor reached the upper limits of the measuring range. Figure 6.5-B is an example of a weak response in which the maximum amplitudes of the sensor are found to remain near the lower limit of the measuring range. It must be emphasized that bacterial identification can be adequately based upon both the strong and the weak signals with comparable specificity. The group of strong bacterial reactors was found to consist of the species *Proteus mirabilis*, *Klebsiella* spp., *E. coli*, *Salmonella* spp., *E. cloacae* and *Clostridium difficile*. The weak group was found to consist of *Staphylococcus aureus*, *Enterococcus faecalis* and *Pseudomonas aeroginosa*. For each species, a characteristic sub-feature was derived from a single experiment. Sub-features for the two *Salmonella* and the two *Klebsiella* species were very similar. Therefore, these species can not yet be distinguished and had to be grouped at the genus level. Examples of these sub-features are given in Figure 6.5 C and D. The sub-features were subsequently 'generalized' by matching and averaging within a species group. With the 'generalized' sub-features, the classification process was iterated.

The results in Table 6.2 show that the test specificity varies from 100% for *C. difficile* to 67% for *E. cloacae* with an overall average of 87%. All characteristic sub-features where observed between 6 to 8 hours after initial inoculation.

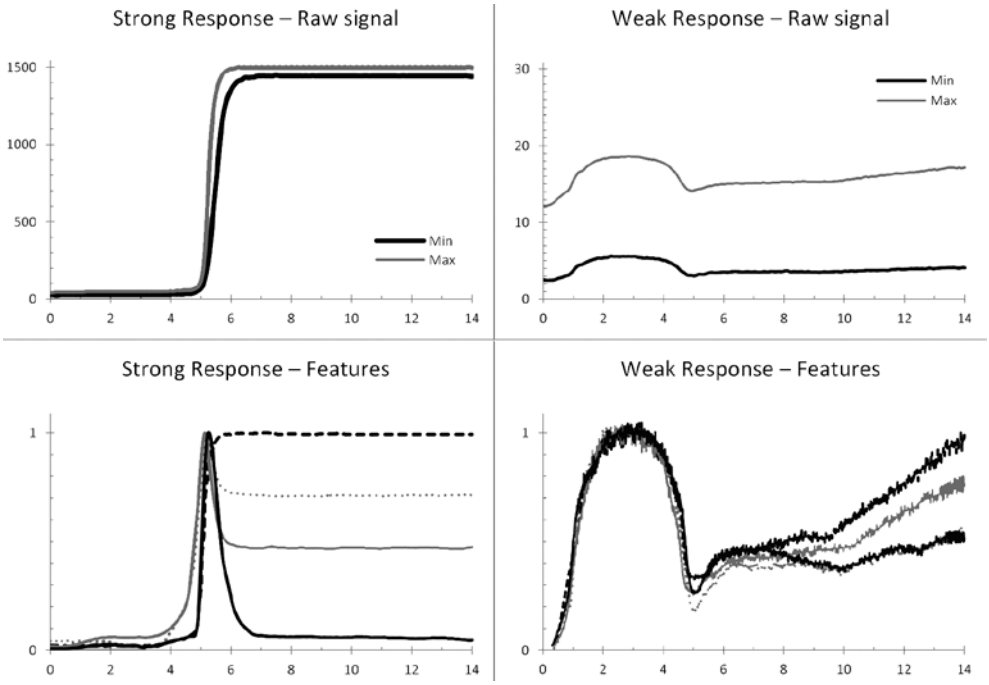


Figure 6.5. Experimental output of the VOC sensing during bacterial cultivation (panels A and B) A: Typical strong result for *E. coli*. B: Typical weak result for *S. aureus*. Note the difference in amplitudes between the strong and weak results. Derivation of species specific signatures from the VOC detection curves on the basis of Sliding Window Minimum Variance Matching and Dynamic Time Warping. C: Selected *E. coli* feature projected on a normalized E. coli experiment. D: Selected *S. aureus* feature projected on a normalized *S. aureus* experiment. Note the reproducible difference in time frames between both features.

	# Measurements	Unclassifiable	Classified		% Correct
			Correct	Error	
Clostridium difficile	4	0	4	0	100
Enterobacter cloacae	6	0	4	2	67
Enterococcus faecalis	16	2	12	2	86
Escherichia coli	12	1	9	2	82
Klebsiella spp	22	0	21	1	95
Proteus mirabilis	10	0	8	2	80
Pseudomonas aeruginosa	8	3	5	0	100
Salmonella spp	10	1	7	2	78
Staphylococcus aureus	16	0	14	2	88
Overall	**104**	**7**	**84**	**13**	**87**

Table 6.2 MonoNose based Identification Result for the bacterial species test panel.

Discussion and Conclusion

The majority of papers describing the classification of bacteria with eNose devices appeared around the turn of the century [9, 10, 11, 12]. In most cases, very good diagnostic results were achieved and bright prospects of applications were hinted at. In all cases, however, the analysis models used a single sensing device. None of the papers addressed the crucial aspect of applicability and reproducibility when using a larger series of independent devices. This seriously hampers the broad applicability and hence the commercial success for such methodologies. Although the older approaches allowed for establishing proof of principle, results came at the great expense of painstakingly calibrating for each single measuring device. In order to achieve similar results with other devices, the same amount of work needed to be performed for each unit. For practical application, it is necessary to ascertain that the reproducibility between a set of devices is sufficient to allow development of analyzing algorithms that do not depend on intensive, individual calibration of the measuring devices. If individual calibration is required to adapt an analysis model to a specific hardware device, the added effort prohibits a realistic application. The current data demonstrate that the MonoNose hardware and times-series analysis software give good results over a representative set of 30 independently produced hardware devices and nine out of eleven bacterial species. All measurements have been conducted with different MonoNoses and the results show that the measured features are interchangeable between the different MonoNoses making the need for individual calibration of each MonoNose obsolete.

This study shows that the MonoNose can identify pathogenic bacteria during the first few hours of growth on the basis of prior calibration data generated by other MonoNoses. Most identification systems in use today follow Bayes theorem and depending on the number of tests used get a score of <90% (e.g. the bioMérieux API-50 test), higher only when significant numbers of additional tests are included. When using generalized features calculated on the basis of the VOC response curves an average identification rate of 87% was achieved. It is expected that this can be optimized further when the detection limits of the MonoNose are expanded. Availability of a more universal culture broth will further enhance the possibilities of the MonoNose to become a fast, reliable and inexpensive detection method for pathogenic bacteria. The study also demonstrates that the development of the metabolizing behaviour of growing bacteria over time gives distinct and sometimes reversible features that are traceable using a single-sensor eNose employing thermal cycling combined with software analysis.

With the broth used in the present experiments not all important pathogen species can be identified. Therefore, future work will focus on optimization of the culture broth

with a new MonoNose that has a larger dynamic measuring range. The amino acid en salt variations showed different optimal mixtures for the different organisms, meaning that changing the broths can greatly improve the result but mostly the increase of one species will mean a decrease of another. Different clinical materials such as urine, blood, and wound exudates also need to be tested to determine the influence of the sample matrix on the measurement. Ideally, the next generation eNoses will not only facilitate bacterial species identification after culture, but also the characterization of bacterial cells present in non-processed clinical materials.

The development of microbial fingerprint pattern recognition such as presented in Figure 6.3, will allow eNose technology to become a rapid diagnostic tool for microbiological diagnosis of the future.

References

1. Remans JJ, Verboom P, Stijnen T, Hakkaart-van Roijen L, Goessens W, Verbrugh HA, Vos MC – Rapid identification and antimicrobial susceptibility testing reduce antibiotic use and accelerate pathogen-directed antibiotic use – J Antimicrob Chemother 2008 61(2):428-35
2. Nowakowska D, Colón I, Remington JS, Grigg M, Golab E, Wilczynski J, Sibley LD – Genotyping of Toxoplasma gondii by multiplex PCR and peptide-based serological testing of samples from infants in Poland diagnosed with congenital toxoplasmosis – J Clin Microbiol 2006 44(4):1382-9
3. Vanlaere E, Sergeant K, Dawyndt P, Kallow W, Erhard M, Sutton H, Dare D, Devreese B, Samyn B, Vandamme P – Matrix-assisted laser desorption ionisation-time-of-flight mass spectrometry of intact cells allows rapid identification of Burkholderia cepacia complex – J Microbiol Methods 2008 75(2):279-86
4. Casalinuovo IA, Pierro D, Bruno E, Francesco P, Coletta M – Experimental use of a new surface acoustic wave sensor for the rapid identification of bacteria and yeasts – Lett Appl Microbiol 2006 42(1):24-9
5. Haddad R, Khan R, Takahashi YK, Mori K, Harel D, Sobel N – A metric for odorant comparison – Nature Methods 2008 5:425-429
6. Pearce TC, Schiffman SS, Nagle HT, Gardner JW – Handbook of Machine Olfaction – WILEY-VCH 2003
7. Latecki LJ, Megalooikonomou V, Wang Q, Lakaemper R, Ratanamahatana CA, Keogh E – Elastic Partial Matching of Time Series – Fifth IEEE International Conference on Data Mining 2005 577–584
8. Berndt D, Clifford J –Using dynamic time warping to find patterns in time series – AAAI-94 Workshop on Knowledge Discovery in Databases 1994 Seattle, Washington
9. Guernin N, Ratcliffe M, Spencer-Phillips PTN, Howe RA – Identifying Bacteria in Human Urine: Current Practice and the Potential for Rapid, Near-Patient Diagnosis by Sensing Volatile Organic Compounds – Clin Cham Lab Med 2001 39(10):893-906
10. Dutta R, Hines EL, Gardner JW, Boilot P – Bacteria classification using Cyranose 320 electronic nose – BioMedical Engineering OnLine 2002 1:4
11. Lykos P, Patel PH, Morong CA, Joseph A – Rapid Detection of Bacteria from Blood Culture by an Electronic Nose – The Journal of Microbiology 2001 399(3):213-218
12. Aathithan S, Plant JC, Chaudry AN, French GL – Diagnosis of Bacteriuria by Detection of Volatile Organic Compounds in Urine Using an Automated Headspace Analyzer with Multiple Conducting Polymer Sensors – Journal of Clinical Microbiology 2001 39(7):2590-2593

chapter 7

eNose evaluation

M. Bruins
I. Sannmann
O. Burfeind
V. Suthar
A. Bos
W. Heuwieser

Evaluation of odor from vaginal discharge of cows in the first 10 days after calving by olfactory cognition and an electronic device

Journal of Dairy science **2013** 96(9):5773-5779

Abstract

The objective of this study was to determine test characteristics (i.e. intra- and inter-observer variability, intra-assay variability, sensitivity, and specificity) of an evaluation of odor from vaginal discharge (VD) of cows in the first 10 days postpartum conducted by olfactory cognition and an electronic device.

In experiment 1, 16 investigators (9 veterinary students and 7 licensed veterinarians) evaluated 5 VD samples each on 10 different days. The Kappa test revealed an agreement between investigators (inter-observer) of K = 0.43 with a Fleiss adjusted SE of 0.0061. The overall agreement was the same for students ($K = 0.28$) and veterinarians ($K = 0.28$). Mean agreement within observers (intra-observer) was K = 0·52 for all observers, and 0·49 and 0·62 for students and veterinarians, respectively. In experiment 2, the repeatability of an electronic device (Aerekaprobe) was tested. Therefore 5 VD samples from 5 cows were evaluated 10 times each. The repeatability was 0·97 determined by cronbach's alpha. In experiment 3, 20 samples collected from healthy cows and 20 of cows with acute puerperal metritis were evaluated by the 16 investigators and the Aerekaprobe using a dichotomous scale (1 = cow with APM, 0 = healthy cow). Sensitivity and specificity of olfactory evaluation was 75·0% and 60·1% compared to 92·0% and 100% for the electronic nose device.

The study revealed a considerable subjectivity of the human nose concerning the classification into healthy and sick animals based on the assessment of vaginal discharge. The repeatability of the electronic nose was higher. In conclusion, the Aerekaprobe system, although imperfect, is a reasonable tool to improve odor assessment of VD. The current system, however, is not suitable as a screening tool in the field. Further research is warranted to adapt such electronic devices to practical on-farm screening tools.

Introduction

Acute puerperal metritis (APM) is an acute systemic illness due to an infection of the uterus, usually occurring within 10 days after parturition. It is characterized by an enlarged uterus and uterine discharge varying from watery red-brown to viscous and purulent fluid which often has a fetid odor. Recently a 3 grade classification has been suggested to improve diagnosis and therapy [1]. A cow showing an abnormally enlarged uterus and a purulent uterine discharge without any systemic signs of ill health is classified as having grade 1 metritis while at grade 2 and 3 metritis, these signs are accompanied with fever ≥ 39·5 °C and signs of toxemia and a fetid watery red-brown uterine discharge. The categorization is based on the appearance of fever together with abnormal vaginal discharge (VD) indicative of a generalized infection caused by interactions between the host immune system and bacterial endotoxins [1]. Characteristics used in research and in the field to differentiate between a normal and abnormal VD include color, viscosity, and smell [1 - 3].

While VD and fever are plausible criteria and have been used in several research trials studying efficacy of different therapies there is a lack of science-based evidence for their diagnostic value [4]. Most recent studies have demonstrated that body temperature can be measured with high repeatability [5] but is subject to certain variables such as time of day, parity, and ambient temperature [6]. For a visual assessment of VD in cows suffering from clinical endometritis (CE) through vaginoscopic examination sensitivity and specificity ranged between 96·3 and 99·6% and 90·1 and 96·7%, respectively. Intra - (κ = 0·55 – 0·60) and inter-observer (κ = 0.44) repeatability of a VD score on a scale from 0 to 3, however, was only moderate [7]. To our knowledge the sensorial assessments of viscosity and smell of VD from cows with APM or CE have not been studied yet.

Odor of VD is associated with the bacterial growth density of potential pathogens in the uterus [8] and seems intuitively to evaluate without the use of additional diagnostic tools. However, data on test characteristics for the evaluation of VD (intra- and inter-observer agreement, sensitivity, specificity) are lacking [4]. There is no evidence whether the sensorial assessment of odor of VD is reliable enough to draw sound conclusions concerning the health status of the animal and the necessity of treatments.

There is science-based evidence both from accepted clinical (e.g. rectal palpation, temperature measurement) and advanced diagnostic methods (e.g. radiography, ultrasound) that the investigator is a relevant source of measurement errors [5, 7, 9]. Studies testing the repeatability of sensorial assessments of odors utilizing a graded solution of phenol in liquid paraffin and olfactometer threshold tests found a correlation between observers (n = 57 and 98, respectively) varying from r = 0·43 to 0·9 [10, 11].

Electronic sensor devices, so called eNoses [12], show possibilities for several beneficial applications to a variety of commercial industries [13] have never been implemented on a large scale.

In dairy research, electronic devices have been used for the detection of substances indicating estrus or pathological conditions such as mastitis or respiratory infection in cattle [14, 15, 16]. These electronic devices are capable of detecting different types and sources of chemical compounds and mixtures of compounds present in sampled air. Volatile organic compounds (VOC) are commonly produced and released from organic sources including living microbes and multicellular organisms [17, 18].

The overall objective of this study was to evaluate odor from VD of cows in the first 10 d postpartum by olfactory cognition and an eNose device. Specifically, we set out 1) to determine the intra- and inter-observer variability of the human olfactory cognition, 2) to determine the repeatability of odor assessment conducted with an eNose and 3) to establish sensitivity and specificity of olfactory cognition and the eNose device.

Materials and Methods

Vaginal discharge samples were collected from a total of 45 cows (22 healthy and 23 with APM in December 2011) held on a commercial dairy farm in Sachsen - Anhalt, Germany housing 1,200 Holstein dairy cows. The cows were closely monitored after calving for 10 d postpartum by daily temperature measurement and evaluation of VD (viscosity, color, and odor) according to Sheldon [1] on days in milk (DIM) 2, 5 and 10. Cows having fetid, reddish - brown, watery vulvar discharge in combination with a rectal temperature ≥ 39·5°C (i.e. fever) were characterized as having APM and sampled. All VD samples were collected with a gloved hand from the vagina after cleaning of vulva and perineum with dry paper towels. Approximately 5 ml of VD were transferred into sterile vials (Sarstedt AG & Co, Nürnberg, Germany) and stored at -20°C for later analysis. For repeatability testing, 50 ml of VD from 5 cows each (2 healthy and 3 with APM) was collected and stored in 5 ml aliquots at -20°C.

Three experiments were conducted to determine 1) intra- and inter-observer variability of olfactory cognition (i.e. human nose), 2) intra-assay variation of an electronic device (i.e. Aerekaprobe The eNose Company, Zutphen, The Netherlands), and 3) sensitivity and specificity by olfactory cognition and the electronic device, respectively, to diagnose APM based on odor alone.

Intra- and inter-observer variability

Sixteen investigators were enrolled in experiment 1, including 9 veterinary students in their 4th year and 7 licensed veterinarians working at the Clinic of Animal Reproduction, Freie Universität Berlin. All investigators consented to participate (informed consent) and had the same information about the study design (i.e. nature of substance under investigation, definition of APM) but no information about the health status of the cows.

A total of 10 appointments in 3 to 4 d intervals were made. All observers evaluated a set of 5 samples (2 healthy and 3 with APM) containing 5 ml VD. The samples were newly randomized for each appointment with the random number function of Excel (Microsoft Office 2010, Microsoft Deutschland GmbH, Munich, Germany) and labeled with capital letters A to E. At each appointment new 5 ml samples were incubated in a water bath at 38°C for 10 min before testing. To allow a semi-quantitative assessment of odor a 5 - point scale (1 = penetrating, very fetid, 2 = fetid, 3 = slightly fetid, 4 = aromatic, 5 = neutral) was used.

Intra-assay variation of an electronic device

In experiment 2, intra-assay variation of the Aerekaprobe was determined according to the descriptions of the manufacturer. The system employs an array of 12 micro-hotplate type metal oxide-sensor modules (i.e. 6 different sensors types in duplicate). The sensors consist of a heating element and a sensor element (a sintered metal - oxide). The sensors are temperature modulated under software control and allow a temperature cycling from 180 to 340°C. In this range the metal oxide-sensors behave as semiconductors. When oxygen absorbs and ionizes at the sensor surface the conductivity is low. Removal of oxygen due to reaction with other gases (redox reaction) results in a measurable change of conductivity. These types of redox-reactions are dependent on the nature of the metal-oxide catalyst, the reacting gases and the temperature [19].

To determine the repeatability of the electronic device the same 5 samples of VD (2 healthy cows and 3 with APM) were evaluated 10 times each. For calibration, the Aerekaprobe was connected to a gas filter (Respirator RT, Budapest, Hungary) to allow a clean airflow through the system for 15 min. The frozen VD samples were incubated in a water bath at 38·0°C for 10 min with closed caps to assure a constant vapor pressure and humidity over the fluid. A sample volume of 5 ml VD per 10 ml vial was chosen to get an ideal VOC saturation within the headspace proportional to the amount of fluid. After thawing the samples were thoroughly mixed (Reax top, Heidolph, Schwabach, Germany) for 2 s. The cap of the vials was punctured by two 1·6 x 25 mm sterile injection needles (Fine Ject, Henke Sass Wolf, Tuttlingen, Germany), connected to a viton tube (2·0 mm diameter,

19·0 cm length) leading into the Aerekaprobe system. This closed circuit allowed only air from the headspace of the sample to enter the Aerekaprobe. After each measurement the Aerekaprobe system was cleaned by air filtered through active carbon (Carl Roth GmbH&Co KG, Karlsruhe, Germany) for 10 min.

The raw data from the Aerekaprobe were processed using a proprietary multiway decomposition algorithm (The eNose Company/Elegant Mathematics). In brief, each individual sensor gives a 3 dimensional matrix for each measurement (time of measurement vs. thermal cycle vs. amplitude) as a result. As the Aerekaprobe unit contains multiple sensors, a set of matrices make up a single measurement. Although all sensors measure in parallel, any arbitrary selection and combination of sensors can be extracted for data analysis. For each measurement a vector was generated containing the scores of the underlying multiway kernels of the compression. These score-vectors were used for the subsequent pattern recognition as they contain the maximum amount of information of the measurement. As there were duplicate sensors (two of each type) a total of 80 patterns (20 healthy, 20 metritic cows x 2) were generated. With this data set we performed a leave–one-out cross validation [20]. The models were derived using a non-linear artificial neural network model [21].

Sensitivity and specificity for APM diagnosis.

In experiment 3, the investigators identical to the ones described in experiment 1 evaluated 20 samples collected from healthy and APM cows, using a dichotomous scale (1 = cow with APM, 0 = healthy cow) to determine sensitivity and specificity using the diagnoses made in the barn as reference. The diagnoses of the investigators had been based on the evaluation of odor alone, whereas the diagnoses in the barn were based on the clinical examination of the cows and the presence of odor and a rectal temperature ≥ 39·5°C, as described above.

Results

Data were analyzed using SPSS (Version 19.0, SPSS Inc., Munich, Germany) and Medcalc software (version 10.1.3.0, MedCalc, Mariakerke, Belgium). The inter-observer repeatability between 16 investigators was calculated using Fleiss' kappa (K) test extended with an additional syntax obtained from http://www.spsstools.net/Syntax/Matrix/CohensKappa.txt. The intra-observer repeatability was calculated with Cohen's kappa using the same syntax. The results of the kappa test were interpreted according to the classification K < 0 = poor, 0·00 to 0.20 = slight, 0·21 to 0·40 = fair, 0·41 to 0·60 = moderate, 0·61 to 0·80 = substantial and 0·81 to 1·00 = almost perfect agreement as described in [22].

Sensitivity and specificity were determined with the receiver operating characteristic (ROC) curve calculated using the presence of fever (rectal temperature ≥ 39·5) and abnormal vaginal discharge at the day of extraction of VD as a reference standard. This combination of findings has been defined as APM in [23]. Sensitivity was calculated as the proportion of positive samples correctly diagnosed as APM positive by olfactory evaluation or the Aerekaprobe, respectively. Specificity was calculated as the proportion of negative samples correctly diagnosed as negative by olfactory evaluation or the Aerekaprobe, respectively.

For the calculation of inter-observer agreement, 5 samples were evaluated 10 times by 16 investigators (i.e. total of 800 evaluations). The Kappa test revealed an agreement of K = 0·43 between different observers (inter-observer). The agreement was the same for students (K = 0·28) and veterinarians (K = 0·28). When transforming the 5 - point scale into a dichotomous scale (1, 2, 3 = APM; 4, 5 = healthy) K was 0·55. Mean agreement within each

	Odor assessment					
	5-point scale					**Dichotomous**
	penetrating very fetid	fetid	Slightly fetid	aromatic	neutral	healthy vs metritic
Interobserver repeatability						
All (n = 16)	0·31	0·23	0·15	0·32	0·39	0·55
Veterinarians (n = 7)	0·24	0·22	0·17	0·28	0·51	0·57
Students (n=9)	0·33	0·21	0·17	0·38	0·30	0·54
Intraobserver repeatability						
All (n = 16)	0·44	0·36	0·25	0·43	0·48	0·54
Veterinarians (n = 7)	0·63	0·45	0·17	0·40	0·76	0·53
Students (n=9)	0·41	0·33	0·22	0·43	0·37	0·55

Table 7.1 Inter - and intraobserver agreement (K) of odor assessment considering a 5-point and a dichotomous (healthy vs. metritic) classification for a total of 16 investigators.

observer (intraobserver) was K = 0·52 for all observers (n = 16), and 0·49 (n = 9) and 0·62 (n = 7) for students and veterinarians, respectively (Table 7.1).

For the olfactory assessment of odor we used a 5-point classification system. In a second step of the analysis, the ordinal scaled data was dichotomized; one of the two outcomes (cows with APM) being defined as the event of intervention. This was done to test whether the agreement would increase allowing a more robust distinction between diseased and healthy cows. The threshold was chosen through a survey performed at the end of the study by all participants. Investigators were asked to state a threshold for treatment intervention based exclusively on the odor of VD. The survey revealed that none of the investigators would treat a cow with a VD score of 4 (i.e. aromatic) or 5 (i.e. neutral) with antibiotics. VD with an odor score of 1 to 3 caused the investigators to consider therapeutic interventions. The kappa - test revealed a moderate overall agreement for the olfactory assessment. Agreement between observers (inter-observer) ($K = 0{\cdot}43$) was lower than within observers (intra-observer) (K = 0·52) which is in agreement with other studies evaluating clinical tests [7, 24, 25]. Differences in the perception of odors between and within observers can be accounted to various factors such as age, experience and environment [11, 25, 26].

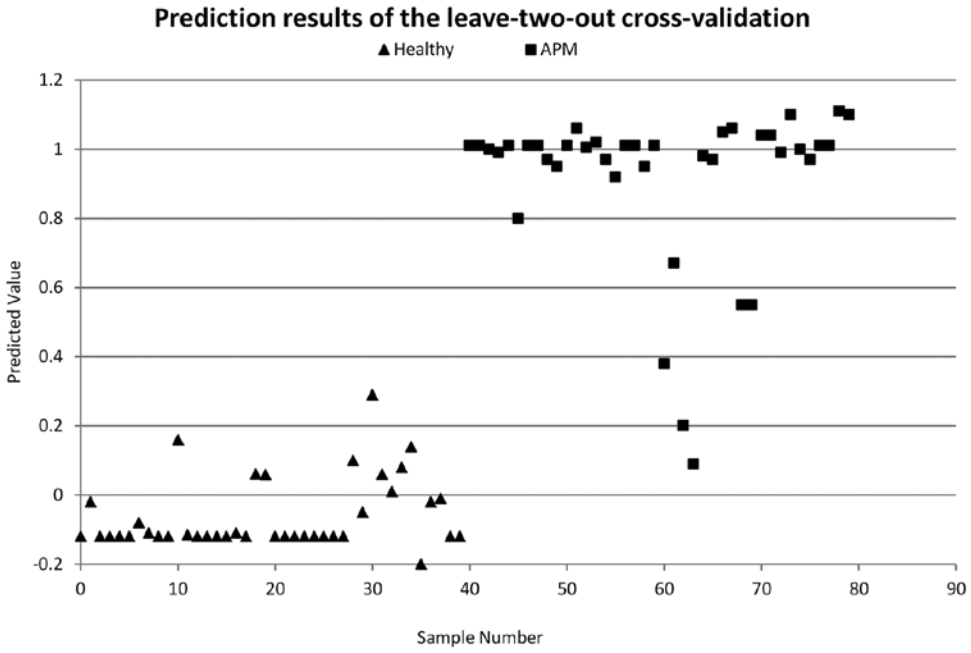


Figure 7.1. Results of the leave-two-out cross validation from the Aerekaprobe system considering 20 healthy cows (black asterisks) and 20 cows with acute puerperal metritis (black squares). Each two consecutive points relate to one measurement as there were two identical sensors in the measurement system. The artificial neural network was trained to give a low value if the cow was healthy and a high value if the cow was sick.

The veterinarians were more experienced diagnosing APM compared to the students and had higher agreement within observers (intra-observer). Agreement between observers (inter-observer), sensitivity and specificity, however, did not differ between veterinarians and students. Even though experience had an influence on agreement, the ability to distinguish healthy cows from cows with APM did not differ between the two groups.

The leave-one-out results of experiment 2, calculated with the artificial neural network, are presented in figure 7.1.

The artificial neural network was trained to give a 0 if the cow was healthy and a 1 if the cow was sick. A ROC curve was calculated (Figure 7.2) from the prediction results of the leave-two-out cross-validation [27]. Sensitivity was 92·0% and specificity 100% with an AUC of 0·99 (P = 0·00) when using 0·5 as the threshold value. To test the repeatability of measurements, the relative conductivity response of each sensor (n = 12) was used in microSiemens (μS) for further analysis. The relative response of the sensor was calculated by where G_{max} and G_{min} are the maximum and minimum conductivity response of the sensor, respectively. This method utilizes only one parameter (i.e. the relative sensor response) of the measurement at a given time and thus only a small part of the total available measurement information. Subsequently we performed a ROC analysis with the results of the relative responses for each sensor pair. The sensor with the highest values, sensitivity of 100%, specificity of 73·3% and an AUC - value of 0·87, was chosen for further analysis. The within-class correlation coefficient was 0·76 and 0·97 for single and average (n = 10) measures, respectively. The repeatability was tested using cronbach's alpha. The resulting α was 0·97 for this sensor.

Odor measurements from VD of healthy cows and cows with APM utilizing the electronic device were different (P < 0·05 based on the area under the calculated ROC curve) as shown in figure 7.2. This difference provides first evidence that the Aerekaprobe could be used as a screening tool to identify cows with APM based on an electronic assessment of odor from VD. In our study, however, the electronic device used was restricted to a laboratory setting. In a pre-test with 18 VD samples, Pearson' s correlation coefficient for native and frozen samples was calculated. Based on the satisfying results (r = 0·88, P = 0·01) we decided to store the VD samples after collection at – 20 °C until analysis. In order to reduce a potential bias as far as possible we implemented a standardized thawing procedure and examined all samples at the same temperature (38 °C). Another potential factor that might have influenced the measurements could be the composition of the VD itself. Although all VD samples were thoroughly shaken the fluids were not all homogenous. VD often consists of different compounds such as blood, pus and cell fragments [28]. This composition could have contributed to limited agreement within observers (intra-observer) as it is possible that the composition of blood, cell fragments and pus was not identical in all of the aliquots used.

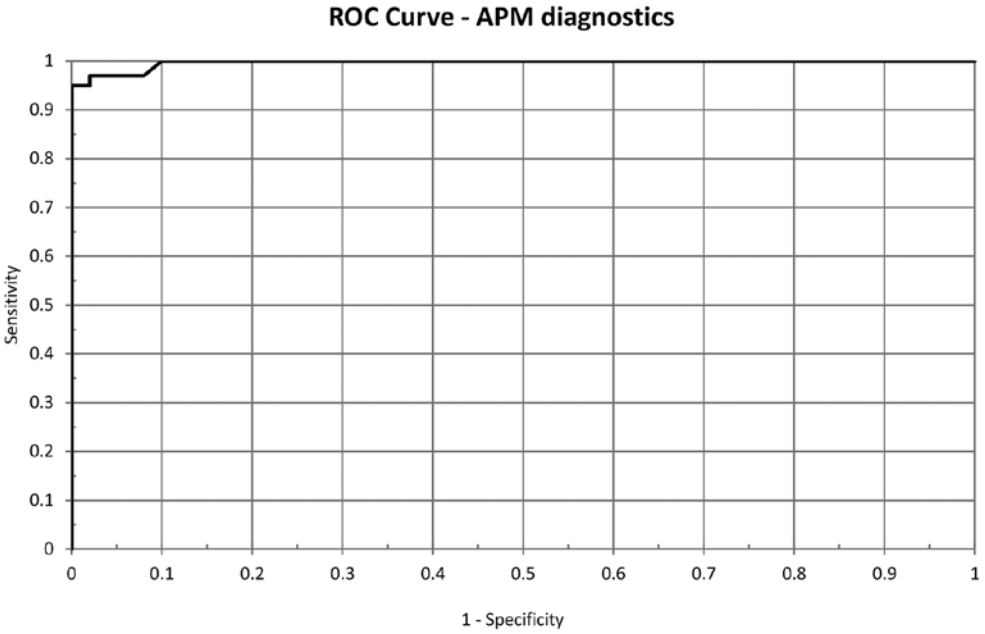


Figure 7.2. Receiver operator characteristic (ROC) curve constructed from the leave-two-out cross validation from the Aerekaprobe measurements for 20 healthy and 20 cows with acute postpartum metritis.

Discussion and Conclusion

A direct comparison between the olfactory and electronic assessment is challenging because data from the olfactory assessment is ordinal while those obtained by eNose are continuous (µS) and consequently different statistical tests were applied. Sensitivity and specificity of the eNose device, however, were higher than those of the olfactory assessment. Data of all 12 sensors resulted in twelve 3-dimensional matrices. This multi-dimensional data is not eligible for repeatability analysis. Therefore, we chose the sensor, with the best results based on a ROC-analysis of each individual sensor for the repeatability analysis. A considerable amount of the totally available information is therewith lost in the process. Notwithstanding, the repeatability was 0·97 (cronbach's alpha) which is excellent [29].

Furthermore, repeatability of odor assessment utilizing the eNose was high while the olfactory assessment with multiple observers showed considerable variability. The intra-observer variability was lower than the inter-observer variability. In agreement with previous studies evaluating palpation both in vivo and in vitro and ultrasound examinations the experience of the observers did not influence the results [30, 31]. While several applications of electronic measuring systems for odors related to estrus, mastitis, and respiratory diseases in cattle have been described previously [14, 15, 16], data on repeatability and test characteristics to compare with our data are not available.

Sensitivity and specificity of olfactory evaluation was 75·0% and 60·1%, respectively considering presence of fever and abnormal vaginal discharge at the day of VD collection as a reference. These criteria have been used in several previous studies to distinguish between healthy and metritic cows [1, 2, 3]. Because a sensory evaluation of odor of VD was part of the on-site diagnosis (i.e. healthy vs. APM) the sensitivity and specificity determined in the laboratory was biased by the inherent limitations of olfactory cognition. There is not a gold standard for APM [23]. Therefore, confounding through type I and type II errors cannot be fully excluded. Furthermore, the nature of VD even varies in healthy cows up to a certain degree. Abnormal VD without a fever could occur in healthy cows as a result of the opening of the cervix at DIM 7 to 10 [32] and cause type I errors by the assessment of VD alone. To minimize the risk of such errors only VD samples collected at DIM 5 were used to avoid a potential bias caused by this natural phenomenon for the evaluation of sensitivity and specificity.

The relatively low sensitivity (75·0%) and specificity (60·1%) for the olfactory assessment could have had different reasons. It is possible that the color and viscosity of the VD samples distracted the investigators. All observers were familiar with the definition of APM which includes odor, color, consistency, and a rectal temperature ≥ 39·5°C and therefore may have unconsciously included these attributes into their conclusion. This could be a potential bias, since color and consistency of lochia can differ widely in a physiological range [28, 33, 34]. Disregarding color and consistency may be an advantage of the Aerekaprobe system measuring only the odorous gases. Furthermore, investigators did not have any information on rectal temperature of these cows, which might have biased the diagnosis as well.

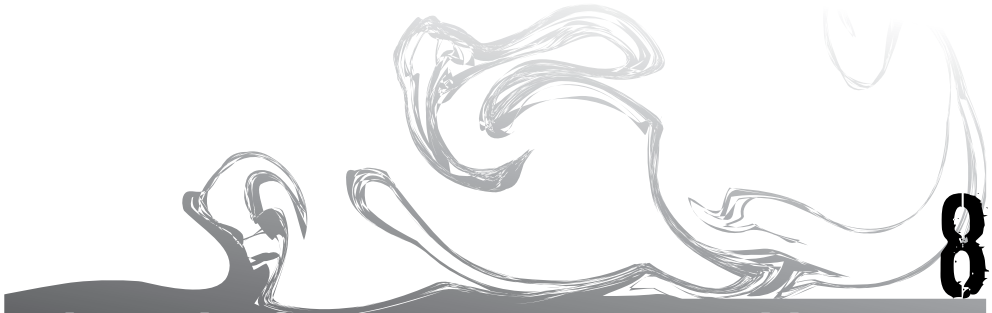
The sensitivity (92%) and specificity (100%) of the Aerekaprobe system was higher than the olfactory assessment. Figure 7.1, however, shows three values obtained from metritic cows below the threshold of 0·5. These are either false negative diagnoses by the Aerekaprobe (i.e. type I error) or a misclassification of the cows as having APM at the time of VD collection by the authors.

This is the first validation of a measuring system to electronically assess odor of VD from cows with the objective to distinguish between healthy and metritic cows. Our data provide first evidence that the Aerekaprobe system, although imperfect, is a useful tool to improve odor assessment of VD. The current system, however, is not suitable as a screening tool in the field. Further efforts are warranted to 1) adapt such electronic devices to on farm screening tools providing real time findings and 2) to determine test characteristics considering fertility as a gold standard. Furthermore, our data show that the assessment of odor from VD by the human nose is highly subjective and might be a confounder of a correct diagnosis regarding the health status of the uterus.

References

1. Sheldon IM, Cronin J, Goetze L, Donofrio G, Schuberth HJ – Defining postpartum uterine disease and the mechanisms of infection and immunity in the female reproductive tract in cattle – Biol. Reprod 2009 81:1025-1032
2. Sheldon IM and Dobson H – Postpartum uterine health in cattle – Anim. Reprod. Sci 2004 82-83:295-306
3. Benzaquen ME, et al. – Rectal temperature, calving-related factors, and the incidence of puerperal metritis in postpartum dairy cows – J. Dairy Sci 2007 90:2804-2814
4. Sannmann I, Arlt S, Heuwieser W – A critical evaluation of diagnostic methods used to identify dairy cows with acute postpartum metritis in the current literature – J. Dairy Res 2012 11:1-9
5. Burfeind O, von Keyserlingk MAG, Weary DM, Veira DM, Heuwieser W – Short communication: Repeatability of measures of rectal temperature in dairy cows – J. Dairy Sci 2010 93:624-627
6. Burfeind O, Suthar VS, Heuwieser W – Effect of heat stress on body temperature in early postpartum dairy cows – Theriogenology 2012 78:2031-2038
7. Leutert C, Von Krueger X, Plöntzke J, Heuwieser W – Evaluation of vaginoskopy for the diagnosis of clinical endometritis in dairy cows – J. Dairy Sci 2012 95:206- 212
8. Williams EJ, et al. – Clinical evaluation of postpartum vaginal mucus reflects uterine bacterial infection and the immune response in cattle – Theriogenology 2005 63:102-117
9. Andermann P, et al. – Intra- and interobserver variability of thyroid volume measurements in healthy adults by 2D versus 3D ultrasound – Nuklearmedizin 2007 46:1-48
10. Fordyce D – Olfaction tests – Brit. J. Ind. Med 1960 18:213-223
11. Doty R, McKeown D, Lee W, Shaman P – A study of the test-retest reliability of ten olfactory tests – Chem. Senses 1995 20:645-656
12. Gardner JW, Bartlett PN – A brief history of electronic noses – Sens. Actuat. B. 1994 18:211-220
13. Wilson AD, Baietto M – Applications and advances in electronic-nose technologies – Sensors 2009 9: 5099-5148
14. Eriksson Å, et al. – Detection of mastitic milk using a gas-sensor array system (electronic nose) – Int. Dairy J. 2005 15:1193-1201
15. Knobloch H, Schroedl W, Turner C, Chambers M, Reinhold P – Electronic nose responses and acute phase proteins correlate in blood using a bovine respiratory infection – Sens. Actuat. B. 2010 144:81-87
16. Wiegerinck W, et al. – BOVINOSE: Pheromone-based sensor system for detecting estrus in dairy cows – Proc. Comp. Sci. 2011 7:340-342
17. Barsan N, Koziej D, Weimar U – Metal oxide-based gas sensors research: How to? – Sens. Actuat. B. 2007 121:18-35
18. Wilson D, Baietto M – Advances in electronic-nose technologies developed for biomedical applications – Sensors 2011 11:1105-1176
19. Bruins M, van Belkum A, Bos A – The use of electronic nose devices in clinical microbiology (The role of new technologies in medical microbiological research and diagnosis – Bentham eBooks 2012 90-110
20. Airola A, Pahikkala T, Waegeman W, DeBaets B, Salakoski T – An experimental comparison of cross- validation techniques for estimating the area under the ROC curve – Comput. Stat. Data Analysis 2010 55:1828-1844
21. Malmgren H – Artificial neural networks in medicine and biology; a philosophical introduction. Opening lecture at the Artificial Neural Networks in Medicine and Biology (ANNIMAB)-1 conference, Göteborg, Sweden, 2000 May 13e16. ISSN: 1652-0459. Available at: http://gupea.ub.gu.se/handle/2077/19466
22. Landis JR, Koch GG – The measurement of observer agreement for categorical data – Biometrics 1977 33:159-174
23. Sheldon IM, Lewis GS, LeBlanc S, Gilbert RO – Defining postpartum uterine disease in cattle – Theriogenology 2006 65:1516-1530

24. Thomsen PT, Baadsgaard NP – Intra- and inter-observer agreement of a protocol for clinical examination of dairy cows – Prev. Vet. Med 2006 75:133-139
25. Pedersen KS, Toft N – Intra- and inter-observer agreement when using a descriptive classification scale for clinical assessment of faecal consistency in growing pigs – Prev. Vet. Med. 2011 98:288-291
26. Hadley K, Orlandi RR, Fong KJ – Basic anatomy and physiology of olfaction and taste. Otolaryngol – Clin. North Am. 2004 37:1115-1126
27. Fawcett T – An introduction to ROC analysis – Pattern recognition letters 2005 27: 861-874
28. Olson J, Bretzlaff K, Mortimer RG, Ball L – The metritis- pyometra complex. Current Therapy in – Theriogenology 1986 227-288
29. George D, Mallery P – SPSS for windows: Step by step: A simple guide and reference, 11.0 Update. Edition 4 – Allyn and Bacon 2002 ISBN 978-0205375523: 231
30. Steyn M, Becker PJ, L'Abbé EN, Scholtz Y, Myburgh J – An assessment of the repeatability of pubic and ischial measurements – For. Sci. Int. 2011 214: 210e1-210e4
31. Leutert C, Suthar VS, Heuwieser W – Evaluation of transrectal examination of cervical diameter by palpation in dairy cows – J. Dairy Sci. 2013 96:1063-1070
32. Wehrend A, Failing K, Bostedt H – Cervimetry and ultrasonographic observations of the cervix regression in dairy cows during the first 10 days postpartum – J Vet. Med. A 2003 50:470–473 Elliott L, McMahon KJ, Gier HT, Marion GB – Uterus of the cow after parturition: Bacterial content – Am. J. Vet. Res. 1968 29:77-81
33. Kumar NS, et al. – Nature and properties of lochia collected from crossbred cows maintained under farm condition – Ind. J. Anim. Prod. and Manag. 2009 25:27-32



8

Diagnosis of acute postpartum metritis

M. Bruins
W. Heuwieser
A. Bos
I. Sannmann, and
O. Burfeind

Diagnosis of acute postpartum metritis by electronic nose analysis of vaginal discharge.

Submitted with *Theriogenology* **21-10-2013**

Abstract

The objective of this study was to estimate the diagnostic accuracy of an electronic nose device using vaginal discharge samples to diagnose acute postpartum metritis in dairy cows. Uterine fluid was sampled manually with a gloved hand and under sterile conditions for electronic nose analysis (DIM 2, 5, and 10) and bacteriological examination (DIM 5), respectively as well as on additional days if acute postpartum metritis was diagnosed during the daily clinical examinations. A dataset containing samples from 70 cows was used to build and train a model and to validate the predicted APM status based on the electronic nose measurement, respectively. Half of the dataset (n = 35; 14 healthy, 21 metritic cows) was provided with information regarding the APM diagnosis and contained all 3 measurements (DIM 2, 5 and 10) for each cow and was used as a training set while the second half was blinded (n = 35; 14 healthy, 21 metritic cows) and contained only the samples collected on DIM 5 of each cow and was used to validate the created prediction model.

A ROC curve was calculated using the prediction results of the validation test. The best observed sensitivity was 100% with specificity of 91·6% when using a threshold value of 0·3. The calculated P-value for the ROC curve was < 0·01.

Overall, *E. coli* was isolated in 8 of 28 (28·6 %) and 22 of 42 (52·4 %) samples collected from healthy and metritic cows, respectively (P = 0·08). *T. pyogenes* and *F. necrophorum* were isolated in 14 and 6 of 28 (50·0 %; 21·4 %) and 17 and 16 of 42 (40·5 %; 38·1%) samples collected from healthy and metritic cows, respectively. The prevalence of *E. coli* and *T. pyogenes* was similar in the samples obtained from metritic cows used for the training set and the validation test. The results are promising especially due to the objective nature of the measurements obtained by the electronic nose.

Introduction

Acute puerperal metritis (APM) is an acute systemic illness with fever ≥ 39·5°C and signs of toxemia due to an infection of the uterus, usually occurring within 10 days after parturition characterized by an enlarged uterus and a watery red - brown fluid to viscous off-white purulent uterine discharge, which often has a fetid odor [1,2].

In the past several intervention studies [3-5] comparing efficacy of different therapies for APM have primarily used the appearance of fever together with abnormal vaginal discharge (VD) as inclusion criteria. These signs are indicative of a generalized infection caused by interactions between the host immune system and bacterial endotoxins [2]. Characteristics used in research and in the field to differentiate between normal and abnormal VD include color, viscosity, odor and amount of VD [2,6,7]. Odor of VD is associated with the bacterial growth density of potential pathogens (i.e. *Escherichia* (E.) *coli*, *Trueperella* (T.) *pyogenes*, *Fusobacterium* (F.) *necrophorum*) in the uterus [8]. Therefore abnormal VD is a plausible criterion which is commonly used in the field because it is easy and intuitively assessable without the use of additional diagnostic tools. A most recent evidence-based review of diagnostic methods for APM, however, points out that information on the diagnostic value of the assessment of VD is lacking [4]. Furthermore, it has been demonstrated that sensorial assessments such as the olfactory evaluation of odor of VD or the visual evaluation of vaginal fluid through vaginoscopy can be confounded by only moderate intra- as well as interobserver repeatability [9,10].

In human medicine electronic nose devices have been used for different purposes including diagnostics of urinary tract infection [11], wound infection [12], and tuberculosis [13]. These devices are capable of detecting different volatile organic compounds (VOC) commonly produced and released from organic sources such as living microbes and multicellular organisms [14-16]. In bovine research electronic noses were utilized for the detection of estrus, mastitis or respiratory disease [17-19]. Data on sensitivity and specificity, however, were not described.

In the past we have determined a moderate inter-observer (κ = 0·43) and intra-observer repeatability (κ = 0·52) for the classification of healthy versus metritic animals based on the assessment of VD by olfactory cognition [10]. These data were generated in a laboratory setting evaluating identical frozen-thawed aliquots of 5 VD samples in 10 replicates and a panel of 16 observers. Interestingly, the intra-assay repeatability of an electronic nose was higher (cronbach's alpha = 0.97) compared to the human nose. Differences in the perception of odors between and within observers can be accounted to various factors such as age, experience and environment [20-22].

The objective of this study was to determine sensitivity and specificity of an electronic nose device to diagnose APM in dairy cattle utilizing VD samples collected from cows diagnosed by a veterinarian as metritic or healthy as reference.

Materials and Methods

The study was conducted between October and December 2011 on a commercial dairy farm in Sachsen-Anhalt, Germany housing 1,200 Holstein dairy cows with an average 305 d milk production of 10,147 kg (3·98 % fat and 3·33 % protein). The main objective of the overall study was to compare the efficacy of 2 treatment protocols for APM in dairy cattle. Details of the study design have been described previously [5]. In brief, a total of 222 cows were enrolled, clinically examined by a graduated veterinarian affiliated with the Clinic of Animal Reproduction and diagnosed as having APM (i.e. fetid, reddish-brown, watery vulvar discharge in combination with a rectal temperature ≥ 39.5°C) or being healthy. Cows with APM received a systemic antimicrobial treatment of 6·6 mg/kg ceftiofur crystalline-free acid (Naxcel, Pfizer Limited, Kent, United Kingdom) in 2 different treatment protocols as described elsewhere [5].

Cows were managed according to the guidelines set by the International Cooperation on Harmonisation of Technical Requirements for Registration of Veterinary Medicinal Products [23]. The experimental procedures reported herein were conducted with the approval of the Institutional Animal Care and Use Committee. Lactating cows were housed in a free-stall barn with cubicles equipped with rubber mats and slotted floors. Early-postpartum cows were fed a total mixed ration consisting of 34·1% corn silage, 20·5% grass silage, 4·2% barley straw, and 41·2% concentrate mineral mix on a days in milk (DIM) basis distributed with a conveyer belt system up to 10 times per day. Cows were milked 3 times a day (0600 h, 1400 h, 2200 h). Milk yield was recorded daily by using the parlor software (Fullexpert Software, Lemmer Fullwood, Lohmar, Germany).

Cows entered the experiment one day after calving at 0700 h. Cows that received anti-inflammatory drugs or antimicrobial drugs for purposes not related to the study (e.g. acute mastitis, lameness) or suffered from other inflammatory diseases than APM were excluded from the trial (n = 29). Rectal temperature was measured daily with a digital thermometer (Microlife AG, Heerbrugg, Switzerland) until DIM 10. Measurements were performed at the same time of the day (0700 h to 0900 h) and at the same insertion depth (8 cm) to minimize any bias due to the measuring process [24].

Vaginal discharge was collected through manual vaginal examination with a gloved hand. To minimize contamination of the vagina, the tails of the cows were held and the vulva and perineum were cleaned with dry paper towels before the discharge was collected into screw-cap vials (Rotilabo, Carl Roth GmbH & Co KG, Karlsruhe, Germany) and evaluated regarding color, proportion of pus, consistency, and smell by one of the investigators. Cows without VD or physiologic lochia were classified as healthy and cows with fetid, red-brown, watery VD in combination with a rectal temperature ≥ 39.5°C (fever) were classified as having APM [1,25]. Cows with fever or with fetid, red-brown, watery VD were not considered as metritic but further monitored.

Sampling

Sampling for electronic nose analysis was performed immediately after routine clinical examination and diagnosis of APM and in all cows routinely on DIM 2, 5, and 10. Therefore, the vulva was cleaned with dry paper towels and vaginal discharge collected manually with a gloved hand, and transferred into screw-cap vials. A bacteriological examination of uterine fluid was conducted on DIM 5 in all cows and on the day of diagnosis of APM. The vulval lips were parted, and a sterile plastic catheter (insemination pipette for horses, Minitueb GmbH, Tiefenbach, Germany), protected through a hygienic sheath (Minitueb GmbH, Tiefenbach, Germany) was advanced under manual control into the vagina and through the cervix. Uterine fluid was aspirated using a 20 ml sterile disposable syringe (Soft-Ject, Luer, Kenke, Sass, Wolf, Tuttlingen, Germany). The fluid was transferred into sterile milk tubules (Sarstedt AG & Co, Nürnberg, Germany), captured by a sterile swab and inoculated into a BD BBL Prepared culture medium (BBL Port-A-Cul Tubes, Becton, Dickson and Company, Sparks, MD 21152 USA) within 2 h of sampling. Samples were sent to a commercial laboratory (Synlab vet GmbH, Leipzig, Germany) and analyzed for relevant pathogens such as *E. coli*, *T. pyogenes*, *F.necrophorum*, and *Prevotella spp.* [8,26,27].

Electronic nose measurements

The electronic nose device (Aerekaprobe, The eNose Company, Zutphen, The Netherlands) employs an array of 12 (i.e. 5 different chemical types, 2 in triplicate and 3 in duplicate) micro-hotplate type metal oxide-sensor modules. The sensors consist of a heating element and a sensor element (a sintered metal - oxide). The sensors are temperature modulated under software control and use a temperature cycle from 260 to 320°C. In this range the metal-oxide sensors behave as semiconductors. The redox reaction taking place at the sensor surface results in a measurable change of conductivity. These types of redox reactions are dependent on the nature of the metal-oxide catalyst, the reacting gases and the temperature [13].

To establish a zero-grade air baseline (atmospheric air that has been cleaned to have less than 0.1 ppm total hydrocarbons remaining) the Aerekaprobe was connected to a gas filter (Respirator RT, Budapest, Hungary) for 15 min before the first sample was measured. The frozen VD samples were incubated in a water bath at 38·0°C for 10 min with closed caps to assure a constant vapor pressure and humidity over the fluid. According to the manufacturer's instruction a sample volume of 5 ml VD per 10 ml vial was chosen to get an ideal VOC saturation within the headspace proportional to the amount of fluid. After thawing, the samples were thoroughly mixed (Reax top, Heidolph, Schwabach, Germany)

for 2 s. The cap of the vials was punctured by two 1·6 x 25 mm sterile injection needles (Fine Ject, Henke Sass Wolf, Tuttlingen, Germany), connected to a viton tube (2·0 mm diameter, 19·0 cm length) leading into the Aerekaprobe system. This closed circuit allowed only air from the headspace of the sample to enter the Aerekaprobe. After each measurement the Aerekaprobe system including the viton tubing was cleaned by air filtered through active carbon (Carl Roth GmbH&Co KG, Karlsruhe, Germany) for 10 min and the needles were disposed.

Selection of samples

A dataset containing samples from 70 cows was used to build and train a model as well as to validate the predicted APM status based on the electronic nose measurement, respectively. Therefore the raw data of the Aerekaprobe measurements were provided by the Clinic of Animal Reproduction, FU Berlin to the manufacturer. Half of the dataset (n = 35; 14 healthy, 21 metritic cows) was provided with information regarding the APM diagnosis and contained all 3 measurements (DIM 2, 5 and 10) for each cow and was used as a training set while the second half was blinded (n = 35; 14 healthy, 21 metritic cows) and contained only the samples collected on DIM 5 of each cow and was used to validate the created prediction model. Here, a consistent DIM was chosen because it has been described that bacteriological findings change within the first 2 weeks postpartum [28].

Raw data analysis

The raw data from the Aerekaprobe were processed using a proprietary multiway compression algorithm (Elegant Mathematics, The eNose Company, Zutphen, The Netherlands) based on tensor decomposition. In brief, each individual sensor gives a 3 dimensional matrix for each measurement (time of measurement vs. thermal cycle vs. amplitude) as a result. As the Aerekaprobe unit contains multiple sensors, a set of matrices make up one measurement. Although all sensors are measured in parallel, any arbitrary selection and combination of sensor matrices can be extracted for data analysis. For each measurement a vector was generated containing the scores of the underlying multiway kernels of the compression. These score vectors were used for the subsequent pattern recognition as they contain the maximum amount of information of the measurement.

The data of the training set was split into 3 different groups based on their APM status. Besides the clinical diagnoses on a given day (i.e. healthy, sick) a third group (i.e. pre-sick) was introduced containing the measurements of samples collected from cows before they were diagnosed with APM (e.g. a sample collected at DIM 2 while the diagnosis APM first

emerged with the sample of DIM 5). Assigning this sample to either the healthy or sick group would be disputable since one cannot know if the cow was suffering from an early stage APM infection or was not yet infected at that point. Since the quality and robustness of the created model depends on quality of the input groups we discarded all measurement belonging to the pre-sick group (n = 23 samples) during the model development stage. With this data set we performed a leave–one-out cross validation [29]. A non-linear artificial neural network model [30] was used to create the models and predict the blinded samples.

The predicted results of the blinded samples where used to calculate sensitivity and specificity and to create a receiver operating characteristic (ROC) curve based on the presence of fever (rectal temperature ≥ 39.5°C) and abnormal VD at the day of extraction of VD as a reference standard. This combination of findings has been defined as APM by Sheldon et al. [1]. Sensitivity was calculated as the proportion of positive samples correctly diagnosed as APM positive by the Aerekaprobe. Specificity was calculated as the proportion of negative samples correctly diagnosed as APM negative. Besides the diagnosis of APM it was also investigated if it would be possible to distinguish an *E. coli* infection because of its predominant occurrence in the uterus of clinically diseased animals and endometrial pathogenicity [28,31].

Results

The training dataset (n = 105) contained 1 invalid measurement due to corrupted data resulting in 42, 39 and 23 samples from the healthy, sick and pre-sick group, respectively. All possible sensor combinations were evaluated using an artificial neural network in a leave-one-out approach. The artificial neural network was trained with a 0 representing a healthy sample and a 1 representing a sick sample. The leave–one–out approach means removing all vectors belonging to the measurement of a given sample. For this dataset a 2-sensor combination yielded the best result, defined as the highest combination of sensitivity and specificity achieved with the lowest number of different sensors. Since there are at least duplicates of each sensor this resulted in 4 vectors per measurement (A1B1, A1B2, A2B1 and A2B2). These 4 vectors represent virtual measurements conducted with 4 different devices. The combining of the data in this manner assures a higher degree of device independency. The usage of 4 vectors per measurement resulted in a training set containing 416 vectors (35 cows x 3 samples x 4 vectors minus the 4 vectors of the invalid measurement) and a validation set of 140 vectors (35 cows x 1 sample x 4 vectors). Using the selected sensor combination a final model was created using the data from all healthy and sick cows in the training set. The APM status for both the pre-sick and the samples of the validation test were determined using this model and visualized in Figures 8.1 and 8.2, respectively

which show the calculated values using the model on the y-axis. The values (representing the similarity of the sample compared to the two groups [0 or 1] defined in the model) do not have a unit as they are the result of a mathematical model. A low value indicates a healthy cow while a high value indicates a sick cow and a threshold value acts as a cut-off between healthy and sick.

Figure 8.1 shows the predicted values of the samples classified as pre-sick grouped by cow. The first part of the graph shows cows which were diagnosed with APM between DIM 2 and 5 and have only one predicted value (cow 1 to 7) while the second part shows the cows diagnosed with APM between DIM 5 and 10 (cow 8 to 15) and therefore have 2 values presenting both pre-sick measurements.

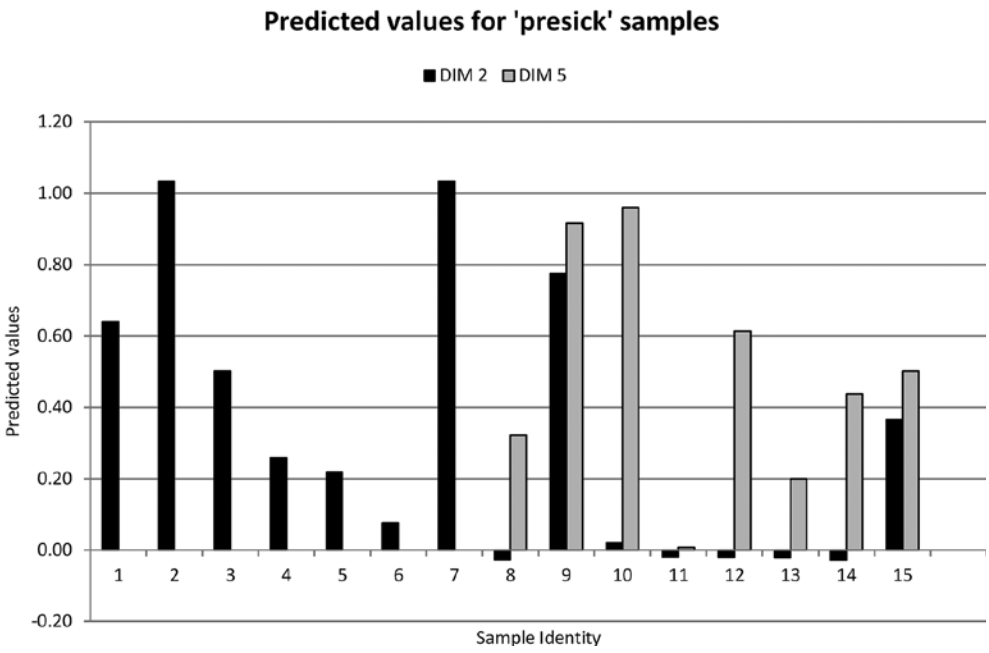


Figure 8.1. Calculated values for the samples from the pre-sick group. A high value indicates a sick cow while a low value indicates a healthy cow. Cows 1 to 7 were diagnosed with APM between DIM 2 and 5 and have only one predicted value while cows 8 to 15 were diagnosed with APM between DIM 5 and 10 and therefore have 2 values presenting both pre-sick measurements.

Figure 8.2 shows the samples classified by their status where the initial sick group is sub-divided into 3 different sub-groups (i.e. sick, pre-sick, and treated). The treated group consists of cows diagnosed with APM before DIM 5 that received antibiotic treatment immediately after the diagnosis.

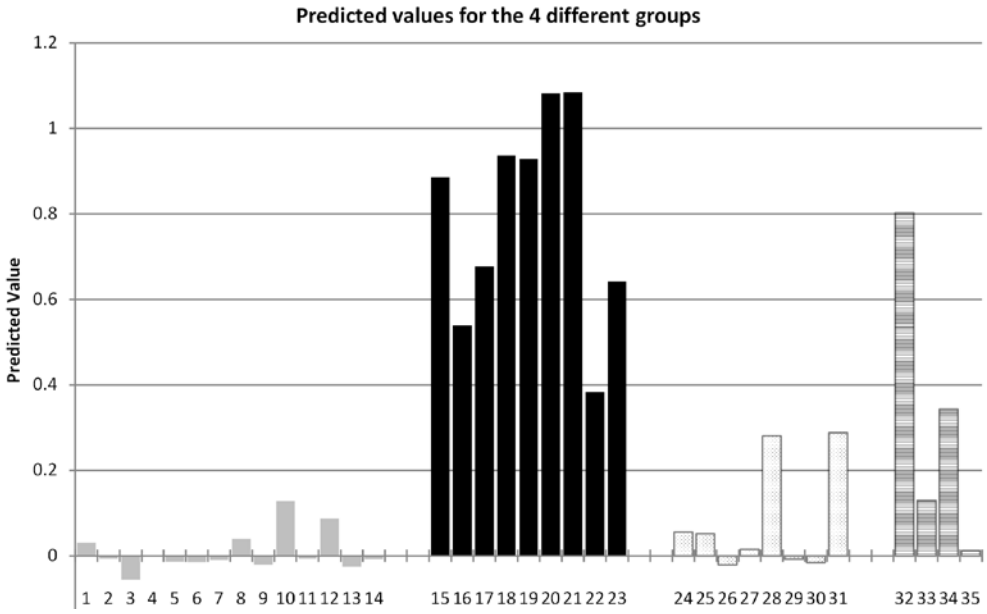


Figure. 8.2. Calculated values for the samples from the validation test collected at 5 DIM. Cows were classified as healthy (▒), sick (■), pre-sick (▨) or treated (▤) based on the clinical diagnoses and disease occurrence. A high value indicates a sick cow while a low value indicates a healthy cow.

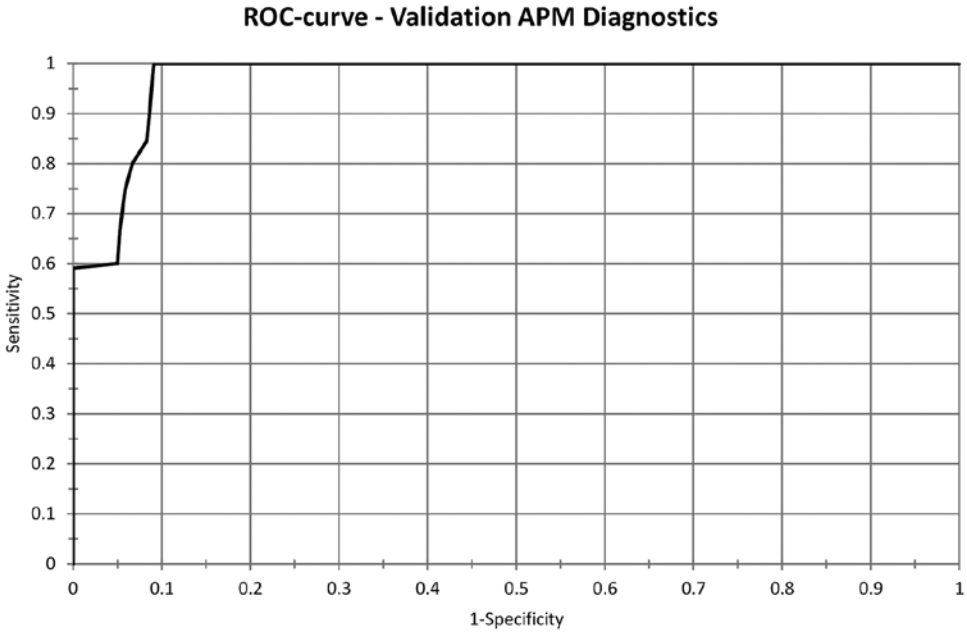


Figure. 8.3. Receiver operating characteristics curve for 35 samples used to validate the prediction model.

A ROC curve was calculated [32] using the prediction results of the validation test (Figure 8.3). The best observed sensitivity was 100% with a specificity 91·6% when using a threshold value of 0·3. The calculated P-value for the ROC curve was < 0·01.

Overall, *E. coli* was isolated in 28·6 % (8 of 28) of the healthy samples and in 52·4% (22 of 42) of the metritic samples (P = 0·08). *T. pyogenes* was isolated in 50·0 % (14 of 28) of the healthy samples and in 40·5 % (17 of 42) of the metritic samples (P = 0·59). *F. necrophorum* was isolated in 28·6 % (8 of 28) of the healthy samples and in 52·4 % (22 of 42) of the metritic samples (P = 0·23).

The prevalence of *E. coli* and *T. pyogenes* was similar in the samples obtained from metritic cows used for the training set and the validation test (Table 8.1).

The distinction between AMP samples with E. coli present (pure or mixed) and the samples with no or other pathogens resulted in a sensitivity of 63·3% and specificity of 56·5%. When trying to distinguish samples collected from metritic cows with E. coli from those collected from healthy cows without E. coli sensitivity and specificity were 77·8% and 71·4%, respectively. For samples with A. pyogenes sensitivity and specificity of 53·8% and 48·1% were reached when compared to all other AMP samples.

	Total				Training set				Validation test			
	healthy (n = 28)		metritis (n = 42)		healthy (n = 14)		metritis (n = 21)		healthy (n = 14)		metritis (n = 21)	
Pathogen present	n	%	n	%	n	%	n	%	n	%	n	%
E. coli												
-in pure culture	2	7.1	10	23.8	1	7.1	7	33.3	1	7.2	3	14.3
-in mixed culture	6	21.4	12	28.6	1	7.1	4	19.0	5	35.7	8	38.1
-Total	8	28.6	22	52.4	2	14.3	11	52.4	6	42.9	11	52.4
T. pyogenes												
-in pure culture	6	21.4	5	11.9	3	21.4	5	23.8	3	21.4	0	0
-in mixed culture	8	28.6	12	28.6	3	21.4	3	14.3	5	35.7	9	42.9
-Total	14	50.0	17	40.5	6	42.9	8	38.1	8	57.1	9	42.9
F. necrophorum												
-in pure culture	3	10.7	1	2.4	3	21.4	1	4.8	0	0	0	0
-in mixed culture	3	10.7	15	35.7	2	14.3	4	19.0	1	7.2	11	52.4
-Total	6	21.4	16	38.1	5	35.7	5	23.8	1	7.2	11	52.4
No pathogens isolated	9	32.1	7	16.7	4	28.6	3	14.3	5	35.7	4	19.0

Table 8.1 Frequency distribution of pathogens in pure and mixed cultures from 70 postpartum cows used in the training set and validation test of the electronic nose device considering uterine health status

Discussion

The rationale of employing an electronic nose device for APM screening in dairy cows was threefold. The intra-assay repeatability of odor assessment of VD by means of an electronic nose was higher compared to the human nose [10]. The composition of the bacterial flora in the postpartum uterus differs between healthy and metritic cows [33]. Previous research in human medicine had demonstrated that electronic nose devices can be used for identification of bacterial pathogens such as *E. coli*, *Proteus spp.* and *Staphylococcus spp.* [11,15].

There is evidence that odor of VD at DIM 21 and 28 is associated with the bacterial growth density of potential pathogens in the uterus [8]. For the first 10 DIM the period during which the samples of our study were collected such information is not available. Our VD samples were collected 6 ± 3 d (mean ± SD) after calving. Prevalence of *E. coli* was higher in the samples collected from metritic cows (both training and validation) compared to healthy cows. This observation is consistent with previous reports which classified *E. coli* and *T. pyogenes* combined with certain gram negative anaerobic bacteria (e.g. *F. necrophorum*) as the main responsible pathogens in this complication [7,8,34]. Considering both pure and mixed cultures *E. coli* was present in 52·4 % and 28·6 % (P = 0·08) of the samples of metritic and healthy cows, respectively. In the first 10 d after calving the uterine lumen, however, is usually contaminated with a lot of other bacteria [28] that were not examined in this study. Furthermore, it has been shown that traditional culture methods of uterine pathogens which were also employed in our study might underestimate the microbial complexity of the intrauterine environment of postpartum cows [33]. Under standard laboratory conditions less than 1% of the microorganisms in many environments are readily cultured [35,36]. Bacterial diversity in healthy and metritic uteri was greater and more complex utilizing metagenomic analysis than described previously for traditional culture methods [33]. These factors might have contributed to the less frequent use of traditional culture methods in research on APM compared to other methods such as measurement of rectal temperature or evaluation of VD [4]. Also, the limitations of traditional culture methods might help to explain, that despite overlapping bacteriological culture findings between healthy and metritic cows sensitivity and specificity of the electronic nose was high. Our data demonstrate that APM can be detected accurately by the Aerekaprobe while it is not possible to determine the causing pathogen. We assume that the opportunistic nature of the APM infection makes is difficult to determine the causing pathogen especially in mixed cultures which made up a large proportion of the bacteriological results. The APM infection causes an odor change which might not be directly related to the type of pathogen. Therefore further research is warranted to determine if more advanced analyses of pathogens considering pathogenicity

traits, inflammation markers and mixed cultures can improve the predictive value of odor composition analyses.

Manual vaginal examination of VD has been validated and does not cause uterine bacterial contamination, provoke an acute phase protein response, or delay uterine involution [37]. To avoid contamination of our samples used for bacteriological analysis with pathogens present in the vagina we collected uterine fluid using disposable plastic catheters protected through a hygienic sheath. This approach was also important as the presence of endometritis, vaginitis, and cervicitis may not be differentiable by looking only at the vaginal discharge [38].

The introduction of the pre-sick group and excluding these samples from the model building process was plausible to assure performance and robustness of the model as it is impossible to determine in advance if a given cow was still healthy at that point or suffering from an early stage APM. This approach seems justified when looking at the predicted APM statuses for the cows in this group (Figure 8.1). Some of the samples would be classified as being sick (e.g. cows 2, 7, 9) while others would be determined as healthy (e.g. cows 8, 10, 12) considering a certain cut-off point. Although there is an obvious difference between the predicted values of these samples a clear pattern can be observed. The values predicted by the Aerekaprobe of cows sampled at DIM 2 and 5 and diagnosed with APM later increased in 7 out of 8 cases. This observation shows an increasing resemblance with the sample of a sick cow the closer it gets to being diagnosed as metritic. A similar pattern has been demonstrated for rectal temperature in cows with APM which increased linearly beginning at 48 h before diagnosis of APM [6].

Until now the accepted definition of APM both for research and practice is an elevated rectal temperature (≥ 39.5°C) and fetid, watery, red VD and has been used in 39 and 21 of 48 peer reviewed research papers, respectively, addressing APM [4]. In lack of a gold standard [1] this definition has to be considered as the best available reference standard. Nevertheless, we emphasize that this definition is empirically based and requires refinement. There is first evidence, however, that VD and body temperature are related (unpublished results) and might therefore indeed be useful to classify healthy and metritic cows. However, it has to be noted that a considerable proportion (i.e. 15·6 to 55·0%) of cows defined with APM using this definition self-cured from this condition [5, 39]. Therefore a screening tool that is able to discriminate between cows with APM that require antibiotic treatment and cows that do not would be of great value considering both prudent antibiotic drug use and animal welfare particularly if such a tool could be employed timely and as point of care diagnostics. Further research is necessary to prove the relationship between the odor development and the APM stage and to study if the Aerekaprobe is able to help discriminating these 2 subpopulations of cows with APM.

In the validation test the APM status of the 35 unknown cows was predicted based on the measurements conducted at DIM 5. The rational of only including samples collected at DIM 5 was to have a consistent interval from calving to sampling trying to eliminate time as a bias for disease development and bacteriological dynamics [28]. The initial results showed a specificity of only 62·5%. After un-blinding and re-evaluating of the clinical data it turned out that 7 cows had been classified as sick while they actually would have belonged to the pre-sick group. As there is no pre-sick option when using a gold standard to determine test characteristics these samples were re-classified from the sick to the healthy group because that would have been the correct status at that moment. The corrected classification was used to calculate the ROC curve (Figure 8.3) and as a consequence, specificity increased to 91·6%.

Three cows were incorrectly classified as healthy based on their validation samples while APM was diagnosed at DIM 2. Two of these 3 cows were treated with a single injection of 6·6 mg/kg BW ceftiofur crystalline-free acid (Naxcel, Pfizer Limited, Kent, United Kingdom) after the APM diagnosis. Presumably, this treatment affected bacterial growth and consequently the composition of the VOC in the headspace of the samples. It remains speculative if these cows could be considered as cured at DIM 5. If this could indicate that a Aerekaprobe measurement might be used to determine the efficacy of an antibiotic treatment needs to be determined.

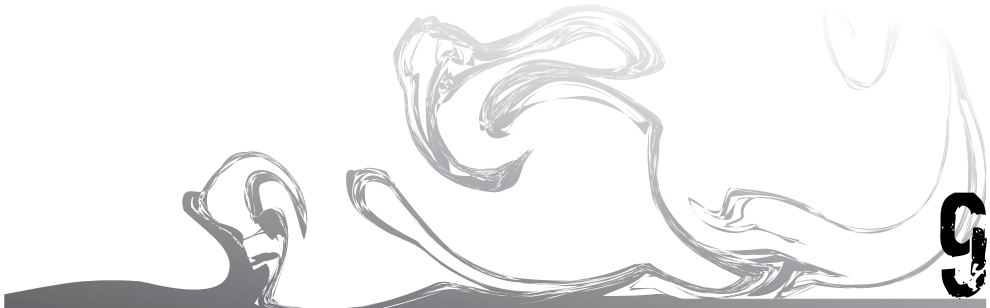
Conclusion

The results are promising especially due to the objective nature of the measurements obtained by the Aerekaprobe. A more extended study needs to be performed with more frequent sampling to determine disease dynamics more accurately, to confirm the actual detection threshold, and to monitor the effectiveness of the administered antibiotic treatment.

References

1. Sheldon IM, Lewis GS, LeBlanc S, Gilbert RO – Defining postpartum uterine disease in cattle – Theriogenology 2006 65: 1516-1530.
2. Sheldon IM, Cronin J, Goetze L, Donofrio G, Schuberth H-J – Defining postpartum uterine disease and the mechanisms of infection and immunity in the female reproductive tract in cattle – Biol Reprod 2009 81: 1025-1032.
3. Drillich M, Beetz O, Pfutzner A, Sabin M, Sabin HJ, Kutzer P, Nattermann H, Heuwieser W – Evaluation of a systemic antibiotic treatment of toxic puerperal metritis in dairy cows – J Dairy Sci 2001 84: 2010-2017.
4. Sannmann I, Arlt S, Heuwieser W – A critical evaluation of diagnostic methods used to identify dairy cows with acute post-partum metritis in the current literature – J Dairy Res 2012 79: 436-444.
5. Sannmann I, Burfeind O, Voigtsberger R, Heuwieser W – Comparison of two monitoring and treatment strategies for cows with acute puerperal metritis – Theriogenology 2013 79: 961-969.
6. Benzaquen ME, Risco CA, Archbald LF, Melendez P, Thatcher MJ, Thatcher WW – Rectal temperature, calving-related factors, and the incidence of puerperal metritis in postpartum dairy cows – J Dairy Sci 2007 90: 2804-2814.
7. Sheldon IM, Dobson H – Postpartum uterine health in cattle – Anim Reprod Sci 2004 82–83: 295-306.
8. Williams EJ, Fischer DP, Pfeiffer DU, England GCW, Noakes DE, Dobson H, Sheldon IM – Clinical evaluation of postpartum vaginal mucus reflects uterine bacterial infection and the immune response in cattle – Theriogenology 2005 63: 102-117.
9. Leutert C, von Krueger X, Plöntzke J, Heuwieser W – Evaluation of vaginoscopy for the diagnosis of clinical endometritis in dairy cows – J Dairy Sci 2012 95: 206-212.
10. Sannmann I, Burfeind O, Suthar V, Bos A, Bruins M, Heuwieser W – Technical note: Evaluation of odor from vaginal discharge of cows in the first 10 days after calving by olfactory cognition and an electronic device – J Dairy Sci 2013 96: 5773-5779.
11. Pavlou AK, Magan N, McNulty C, Jones JM, Sharp D, Brown J, Turner APF – Use of an electronic nose system for diagnoses of urinary tract infections – Biosens Bioelectron 2002 17: 893-899.
12. Šetkus A, Galdikas A-J, Kancleris Ž-A, Olekas A, Senuliene D, Strazdiene V, Rimdeika R, Bagdonas R – Featuring of bacterial contamination of wounds by dynamic response of sno2 gas sensor array – Sensor Actuat B-Chem 2006 115: 412-420.
13. Bruins M, Rahim Z, Bos A, van de Sande WWJ, Endtz HP, van Belkum A – Diagnosis of active tuberculosis by e-nose analysis of exhaled air – Tuberculosis 2013 93: 232-238.
14. Barsan N, Koziej D, Weimar U – Metal oxide-based gas sensor research: How to? – Sensor Actuat B-Chem 2007 121: 18-35.
15. Bruins M, Bos A, Petit PL, Eadie K, Rog A, Bos R, van Ramshorst GH, van Belkum A – Device-independent, real-time identification of bacterial pathogens with a metal oxide-based olfactory sensor – Eur J Clin Microbiol Infect Dis 2009 28: 775-780.
16. Wilson AD, Baietto M – Advances in electronic-nose technologies developed for biomedical applications – Sensors (Basel) 2011 11: 1105-1176.
17. Eriksson Å, Persson Waller K, Svennersten-Sjaunja K, Haugen JE, Lundby F, Lind O – Detection of mastitic milk using a gas-sensor array system (electronic nose) – Int Dairy J 2005 15: 1193-1201.
18. Knobloch H, Schroedl W, Turner C, Chambers M, Reinhold P – Electronic nose responses and acute phase proteins correlate in blood using a bovine model of respiratory infection – Sensor Actuat B-Chem 2010 144: 81-87.
19. Wiegerinck W, Setkus A, Buda V, Borg-Karlson AK, Mozuraitis R, de Gee A – Bovinose: Pheromone-based sensor system for detecting estrus in dairy cows – Procedia Comput Sci 2011 7: 340-342.
20. Doty RL, McKeown DA, Lee WW, Shaman P – A study of the test-retest reliability of ten olfactory tests – Chem Senses 1995 20: 645-656.

21. Orhan KS, Karabulut B, Keles N, Deger K – Evaluation of factors concerning the olfaction using the sniffin' sticks test – JAMA Otolaryngol Head Neck Surg 2012 146: 240-246.
22. Hadley K, Orlandi RR, Fong KJ – Basic anatomy and physiology of olfaction and taste – Otolaryngol Clin North Am 2004 37: 1115-1126.
23. Hellmann K, Radeloff I – International cooperation on harmonisation of technical requirements of veterinary medicinal products (vich) – Bruessel/Belgien 2000.
24. Burfeind O, von Keyserlingk MAG, Weary DM, Veira DM, Heuwieser W – Short communication: Repeatability of measures of rectal temperature in dairy cows – J Dairy Sci 2010 93: 624-627.
25. Drillich M, Voigt D, Forderung D, Heuwieser W – Treatment of acute puerperal metritis with flunixin meglumine in addition to antibiotic treatment – J Dairy Sci 2007 90: 3758-3763.
26. Dohmen MJW, Joop K, Sturk A, Bols PEJ, Lohuis JACM – Relationship between intra-uterine bacterial contamination, endotoxin levels and the development of endometritis in postpartum cows with dystocia or retained placenta – Theriogenology 2000 54: 1019-1032.
27. Sheldon IM, Dobson H – Postpartum uterine health in cattle – Animal Reproduction Science 2004 82-83: 295-306.
28. Földi J, Kulcsár M, Pécsi A, Huyghe B, de Sa C, Lohuis JACM, Cox P, Huszenicza G – Bacterial complications of postpartum uterine involution in cattle – Anim Reprod Sci 2006 96: 265-281.
29. Airola A, Pahikkala T, Waegeman W, De Baets B, Salakoski T – An experimental comparison of cross-validation techniques for estimating the area under the roc curve – Comput Stat Data Anal 2011 55: 1828-1844.
30. Malmgren H – Artificial neural networks in medicine and biology - a philosophical introduction – Artificial Neural Networks in Medicine and Biology (ANNIMAB-1) Conference, Göteborg, Sweden, 2000 1-21.
31. Sheldon IM, et.al – Specific strains of escherichia coli are pathogenic for the endometrium of cattle and cause pelvic inflammatory disease in cattle and mice – PLoS One 2010 5: e9192.
32. Fawcett T – An introduction to roc analysis – Pattern Recognit Lett 2006 27: 861-874.
33. Santos TMA, Gilbert RO, Bicalho RC – Metagenomic analysis of the uterine bacterial microbiota in healthy and metritic postpartum dairy cows – J Dairy Sci 2011 94: 291-302.
34. Miller ANA, et.al. – The effects of arcanobacterium pyogenes on endometrial function in vitro, and on uterine and ovarian function in vivo – Theriogenology 2007 68: 972-980.
35. Amann RI, Ludwig W, Schleifer KH – Phylogenetic identification and in situ detection of individual microbial cells without cultivation – Microbiol Rev 1995 59: 143-169.
36. Whitman WB, Coleman DC, Wiebe WJ – Prokaryotes: The unseen majority – Proc Natl Acad Sci USA 1998 95: 6578-6583.
37. Sheldon IM, Noakes DE, Rycroft AN, Dobson H – Effect of postpartum manual examination of the vagina on uterine bacterial contamination in cows – Vet Rec 2002 151: 531-534.
38. Dubuc J, Duffield TF, Leslie KE, Walton JS, LeBlanc SJ – Definitions and diagnosis of postpartum endometritis in dairy cows – J Dairy Sci 2010 93: 5225-5233.
39. McLaughlin CL, et.al. – Evaluation of two doses of ceftiofur crystalline free acid sterile suspension for treatment of metritis in lactating dairy cows – J Dairy Sci 2012 95: 4363-4371.



9 Diagnosis of active tuberculosis

M. Bruins,
Z. Rahim,
A. Bos,
W.W.J. van de Sande,
H.P. Endtz, and
A. van Belkum

Diagnosis of active tuberculosis by e-nose analysis of exhaled air

Published in
Tuberculosis **2012** 93:232-238

Abstract

Tuberculosis, a highly infectious airborne disease, remains a major global health problem. Many of the new diagnostic techniques are not suited for operation in the highly-endemic low-income countries. A sensitive, fast, easy-to-operate and low-cost method is urgently needed.

We performed a Proof of Principle Study (30 participants) and a Validation Study (194 participants) to estimate the diagnostic accuracy of a sophisticated electronic nose (Aerekaprobe, The eNose Company) using exhaled air to detect tuberculosis. The Aerekaprobe uses a measurement method that enables transfer of calibration models between devices thus eliminating the most common pitfall for large scale implementation of electronic noses in general. Aerekaprobe measurements were validated using traditional sputum smear microscopy and culture on Löwenstein-Jensen media.

We found a sensitivity of 95.9% and specificity of 98.5% for the pilot study. In the validation study we found a sensitivity of 93.5% and a specificity of 85.3% discriminating healthy controls from TB patients, and a sensitivity of 76.5% and specificity of 87.2% when identifying TB patient within the entire test-population (best-case numbers).

The portability and fast time-to-result of the Aerekaprobe enables a proactive screening search for new TB cases in rural areas, without the need for highly-skilled operators or a hospital center infrastructure.

Introduction

It is estimated that one third of the world's population is infected with *Mycobacterium tuberculosis*, often leading to active tuberculosis (TB) [1]. In 2010 there was an estimated incident case count of 8·8 million active TB infections, resulting in 1.5 million deaths. The primary detection technique is the 125 year old Ziehl-Neelsen (ZN) staining [2] combined with microscopy. The major drawback of microscopy is that it only allows for detection of pulmonary disease cases in an advanced stage, meaning that often the disease has already been transmitted to close contacts. Chest X-ray and microbiological culture have been added to the diagnostic arsenal in developing countries while other, more advanced techniques such as nucleic acid amplification, serology, and cytokine release assays [3] are becoming available in the developed countries while research is still ongoing. In 2010 the World Health Organization (WHO) endorsed the use of the Cepheid Xpert MTB/RIF system for use in endemic areas. This polymerase chain reaction system, however, relies on a constant power supply making it not ideally suited for portable operation.

Using exhaled air as potential diagnostic indicator is an emerging activity [4-8]. In previous studies gas chromatography combined with mass spectrometry (GC/MS) was used to identify volatile biomarkers unique to *M. tuberculosis* [9-12]. From these studies it appeared that volatile biomarkers may be used to discriminate TB-patients from healthy controls. However, GC/MS is not viable as a diagnostic tool due to the complex settings of the equipment and operation skills needed and using animals [7, 8] introduces a whole other set of challenges.

A viable diagnostic tool which uses volatile biomarkers to differentiate between people with and without TB should be based on an easy-to-use method. ENoses have already been used for different medical purposes [13-15] including the diagnostics of asthma [4], chronic obstructive pulmonary disease [4, 5], urinary tract infection [16], wound infection [17] and even cancer [18, 19]. In the past, we used an Aerekaprobe for the laboratory-based identification of bacterial pathogens [20].

In this study, we used three independently produced Aerekaprobe units and determined the device-independent diagnostic accuracy in detecting TB. We started with a proof of principle (PoP) study to determine the possibility to discriminate between severely TB-infected and healthy people. Subsequently, we conducted a validation study (VS) to substantiate the PoP results and extend the variance between the examined groups.

Methods

Experimental equipment

The Aerekaprobe (The eNose Company, Zutphen, The Netherlands) is an eNose device incorporating 12 metal-oxide sensors, being 4 different sensor types (AS-MLC;AS-MLN;AS-MLK;AS-MLV, Applies Sensors Gmbh) in triplicate.

The Aerekaprobe is equipped with a pump, where the inlet is controlled by a solenoid switching between 2 different inlets, to create an active airflow across the sensors. One inlet is connected to an active carbon filter to provide a baseline free from environmental influence while the second inlet is attached to the sample-bag. The Aerekaprobe measures the air composition every 20 seconds using a 32-step sinusoidal modulation of the sensor surface temperature between 260 and 340 °C, thus resulting in a vector of 32 values each 20 seconds for each of the twelve sensors.

Study design

The studies were conducted at the Tuberculosis laboratory of the International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B) in collaboration with the National TB Control Program between October 2009 and October 2010. The PoP study is conducted by the author and a Dutch trainee while the VS is conducted by a local Bangladeshi technician, trained by the author.

The PoP study was setup as a retrospective study and conducted in October 2009; all patient samples were collected at the National Institute of Disease of Chest and Hospital (NIDCH). The Sputum Smear Microscopy (SSM) was used to select the patients meeting the TB status inclusion criteria. The PoP study was approved by the Research Review Committee (RRC) and Ethical Review Committee (ERC) of ICDDR,B under number PR-09034.

The VS was setup as a prospective study and conducted from May-October, 2010. All patients provided 3 samples at 2 consecutive days at Shyamoli Chest Disease Clinic (SCDC). The SSM result of the first sample formed the basis to approach neighbors to volunteer as socio-economic healthy controls. The VS study was approved by the RRC and ERC of ICCDR, B under number PR-10023. In the VS, patients with a non-TB medical condition were included.

For both studies, the breath samples were measured on-site and analyzed immediately. All sputum samples were transported to the TB laboratory at ICDDR,B where SSM and cultivation on solid Löwenstein-Jensen media were performed within 3 hours. The culture results obtained were used as diagnostic reference standard since this is still the current gold standard in TB-diagnostics.

Patient characteristics and case definitions

The PoP study consisted of two groups. The first group contained 15 patients enrolled at NIDCH meeting the inclusion criteria; the second group contained 15 employees of ICDDR,B.

The inclusion criteria for the PoP study were as follows:

1. aged over 18,
2. agreement to participate through informed consent,
3. ability to produce sputum and air sample,
4. not having received anti-TB-medication for at least 7 days,
5. having two of the three provided sputum samples indicated as 3+ by SSM,
6. controls should have no health complaints.

The VS consisted of 4 groups. The first two groups consisted of patients enrolling at SCC and discriminated on basis of the TB culture results. The first group (PT, Patient confirmed TB) contained all TB positive patients while the second group (PN, Patient No TB) contained all TB culture negative patients. The third group (HC, Healthy controls) consisted of neighbors having no health complaints of the first 20 patients that were diagnosed with a TB confirmation as defined in the PoP study. These neighbors most likely have comparable socio-economic status and act as socio-economic control. The fourth group (EC, Employee control) consisted of randomly selected healthy employees of ICDDR,B.

The inclusion criteria for the VS were as described above for the PoP with the differences that patients should be aged over 15 and that the criterion number (5) mentioned above was not applicable.

Sample collection and measurement

All participants were asked to breathe through the sampling set-up that consisted of a non-rebreathing T-valve with an active-carbon filter attached to the inlet. During sampling, a nose clamp was placed on the nose of the participant to avoid entry of non-filtered air.

The first 5 minutes were used to flush the environmental influences from the lungs after which the sample bag was attached and filled with at least 1 L of exhaled air. After collection, the sample bag was attached to the sample inlet of the Aerekaprobe.

Each patient visited the hospital on two consecutive days and provided three sputum samples. The first sample was taken during the first visit, the second one was collected by the patient in the early-morning at home and a third one was taken during the second visit. These three samples are needed for the routine TB diagnostics (smear microscopy and culture). During the PoP, we first needed the smear results to select eligible patients and

therefore we could only obtain one breath sample per person during their second visit while during the VS the patients provided a sample on both consecutive days. The content of the sample bag was measured twice during the PoP to act as duplicate while during the VS both samples were measured independently but only once.

A single measurement took 10 min in which the first five minutes consisted of the exposure of the sensors to the breath sample, actively taken from the sample bag, and the last five minutes consisted of the recovery phase where the adsorbed chemicals are released from the system under influence of the clean reference air. Thus each single measurement comprises of the chemical adsorption and desorption dynamics at the sensor surface, caused by the sample, during these 10 minutes.

The Aerekaprobe sample collection and measurements during the PoP were conducted by the authors and prior to the VS a local research associate was trained by the authors. For both studies trained staff members of ICDDR,B performed the sputum collection, sputum smear preparation, SSM, and culture according to well accepted microbiological guidelines [21] without any knowledge of the Aerekaprobe results.

Aerekaprobe measurements and analysis

A single measurement of the Aerekaprobe results in 12 matrices of exposure/recovery dynamics, one for each of the sensors. These 12 sets can be regarded as a multi-way dataset in which the first dimension consists of the temperature cycle applied to the sensor, the second dimension is the time (the exposure and recovery phase), the third dimension is the chemical sensor type and finally the fourth dimension, being the amplitude of the sensors. We used a proprietary 'hybrid' tucker-tensor decomposition to compress the data in these four dimensions using PARAFAC [22] and TUCKER3 [23] kernels. The software used was customized for this specific data set-up. The compressed data is thereafter used to build a classification model using an artificial neural network (ANN) and the 95% confidence interval is calculated using the Wilson score interval.

A binary decision such as presence or absence of a disease, as in this study, always has 4 possible outcomes, true positive (TP), false-positive (FP), true-negative (TN) and false-negative (FN). The relationship between these outcomes, the sensitivity and specificity of the test, determine the usability of the test. The result of the ANN is a similarity indication and the chosen threshold value determines the ratio between the outcomes. All possible outcomes are plotted in a so-called ROC curve. A diagonal line is plotted in the ROC curve depicting the result when the outcome would be completely random and is called the 'line of no-discrimination'. The larger the distance between the observed results and this line, the better the classification results. The best result for a particular test, however, depends

on the equilibrium between benefits of a true-positive and the cost/risks of a false-positive result.

Results

The studies were conducted in two parts during a time span of 14 months in Dhaka, Bangladesh. For the PoP study all possible permutations of the 4 sensor types where evaluated and a combination of 2 sensors yielded the optimal result. Within the VS study these findings were first validated before conducting the full data-analysis based on the measurements of only two of the 4 sensor types. All possible sensor permutations for the triplicate of these 2 types result in 9 unique combinations per measurement. All subjects were measured twice, resulting in 18 vectors per patient ensuring the sensor independency of the created data model.

Proof of Principle

A total of 15 patients from the Dhaka area, enrolling at NIDCH, and 15 healthy ICDDR,B employees were sampled and diagnosed with both classical microbiological technology and a Aerekaprobe (see Table 9.1). The included patients all had at least a ZN 3+ SSM for two out of the three samples.

	PoP		VS			
	Patient	Control	Patients TB	Control non TB	Control Neighbor	Control Employee
Amount - count (%)	15 (50·0)	15 (50·0)	34 (17·5)	114 (58·8)	20 (10·3)	26 (13·4)
Male - count (%)	13 (86·7)	12 (80·0)	22 (64·6)	80 (70·2)	13 (65·0)	18 (69·2)
Female - count (%)	02 (13·3)	03 (20·0)	12 (35·4)	34 (29·8)	7 (35·0)	8 (30·8)
Age - average [range]	32 [21-58]	30 [18-58]	31 [16-70]	35 [18-69]	30 [16-55]	35 [22 - 59]

Table 9.1. Population characteristics for the two study populations.

All 18 vectors representing a patient were classified using the ANN software which was trained with the remaining vectors. The ANN is trained with a target value of 1·0 for the ZN 3+ SSM cases and a target value of 0·0 for the healthy employee cases. The ANN determines the similarity between the sample and both target groups, resulting in a predicted position of the sample relative to both targets. The average of these calculated predictions for each of the 30 participants is depicted in Figure 9.1. As can be clearly observed, there is a 100% correct classification when looking at the average values. Most vectors of a subject are

located tightly together showing a high resemblance between their measured data while there are some subjects with less resemblance, causing them to be likely outliers, but still being classified within the correct class.

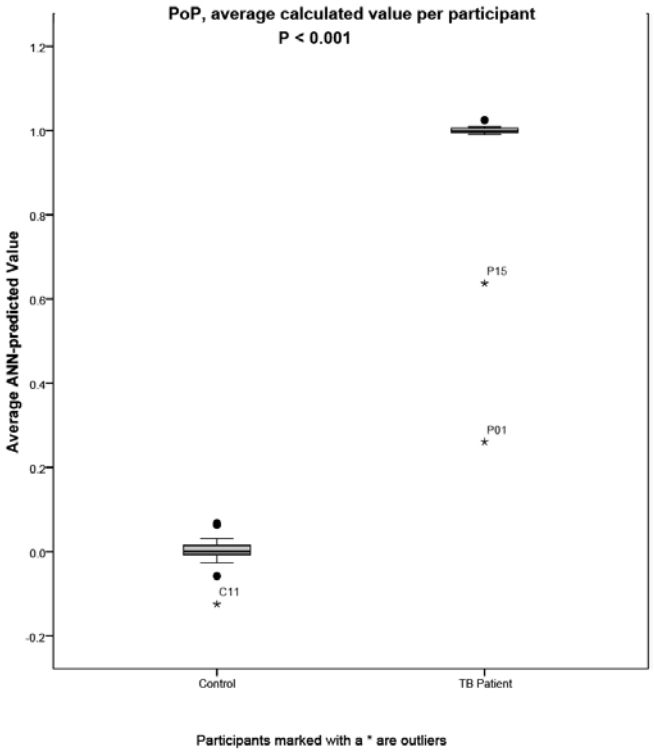


Figure 9.1. Box plot showing the distribution of the average calculated values for the Control and the Patient group of the PoP. Outliers are depicted with a circle while extreme values are depicted with an asterisk.

However, if the data would be evaluated as 540 (30*18) separate measurements (as if there had been 18 different devices used) there were some misclassified measurements. One employee would be classified as TB patient in 4 out of the 18 instances and two patients would be classified as healthy in respectively 5 and 6 of the 18 instances. The sensitivity of the Aerekaprobe would be 95·9% [92·9%-97·7%] and its specificity 98·5% [96·2%-99·4%] when these partially misclassified cases were taken into consideration.

Validation Study

The above results suggest that the Aerekaprobe can differentiate between people with and without TB. However, additional validation studies were clearly required since the Aerekaprobe does not measure pathogens directly as with culture. For instance, all

the healthy controls had a regular income while most of the patients were unemployed. This could mean that the PoP study might have been biased based on other criteria than the TB-infection status of the participants. Therefore, 200 patients, enrolling at SCDC, were asked to participate in our VS without having prior knowledge of their TB status. Of these 200 patients, 52 did not want to participate or did not meet our inclusion criteria. The remaining 148 participants were divided in 2 groups based on their culture results. These two groups consisted of patients with culture-confirmed TB infection (PT) and patients with TB culture-negative results (PN). The PT-group consisted of 34 participants while the PN-group contained 114 participants. A flowchart showing the separation into these two groups is given in figure 9.2. This chart illustrates the separation into the different groups when the calculations are performed using all the subjects rather than the partial values shown in the chart.

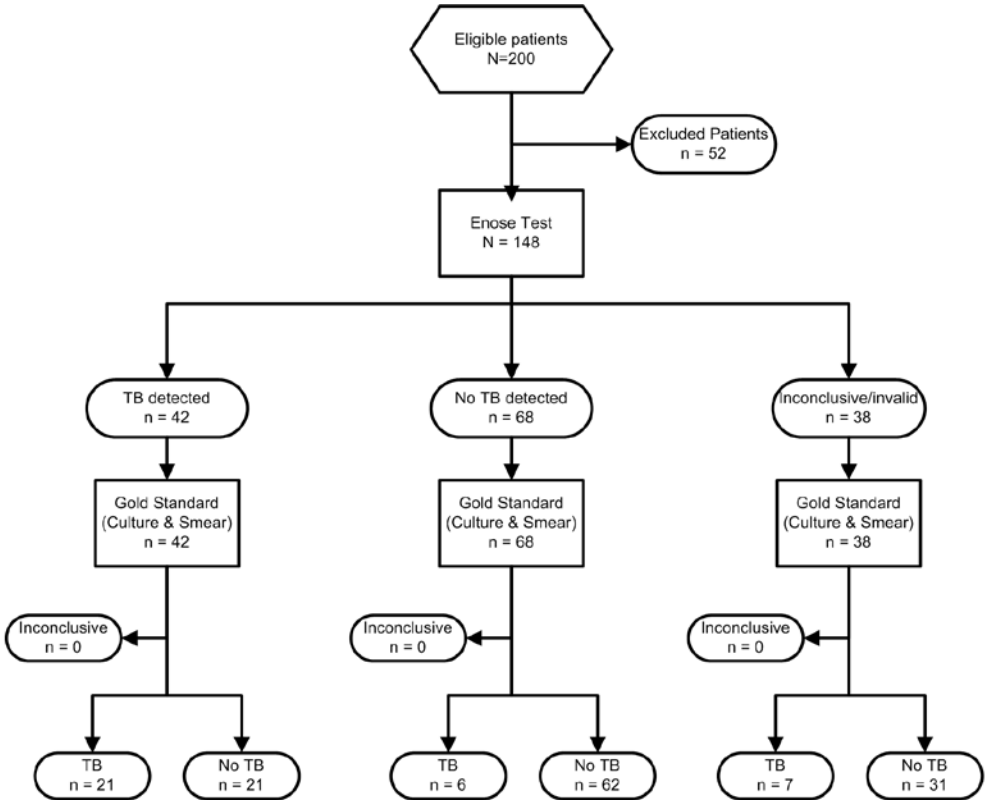


Figure 9.2. Flowchart showing the combination combined result of the Aerekaprobe classification and the reference test and their distribution amongst the eligible patients. The controls are not included as these were all TB negative with the reference test.

Together with these groups, we also included 20 participants (the estimated number of culture positive patients) for a group of socio-economic healthy controls (HC) and 26 employees into the employee control (EC) group. These 2 additional groups where introduced to account for general differences between persons with and without, lung related, health problems (HC) and the influence of the socio economic status of the participants (EC). Participants of both groups were asked to provide a sputum sample to validate their TB condition. An inability to provide a sample was considered as not suffering from active TB. The 3 groups of people not suffering from an active Tuberculosis infection reflect the most common geographical related differences. The characteristics of the 194 participants, spread over the four groups, are shown in Table 9.1.

The collected data was first analyzed to ensure the measurements were all complete and without errors. Out of the 148 patients a total of 38 (7 from the PT group and 31 from the PN group) could not be used. This was due to the fact that both measurements had to be complete and in the beginning we experienced some difficulties with the storage of the carbon filters causing signal fluctuations due to saturation. After this issue had been solved, invalid measurements due to insufficient sample volume or technical issues (such as the operator forgetting to charge the battery overnight or damaged tubing) occurred only occasionally. All of the 46 control measurements were complete and without errors.

The multi-way data was compressed using the measurements from all groups together to avoid the introduction of a compression artifact. As this compression is unsupervised it does not introduce any bias towards the analysis results. In accordance with the first study, an artificial neural network (ANN) was trained in a leave-one-out cross validation approach. As in the PoP, the results according to the gold standard need to be known before the analysis can be conducted. The PT group always had a target value of 1·0 and the three other groups always had a target value of 0·0 independent of the combination of sets. The graphical representations (Figure 9.3 A-D) of the average of the calculated predictions for the different groups used are given to visualize the relations/differences between the different control groups. The ANN used only classifies between 2 classes (TB and non-TB).

The PT group was compared with the EC group to validate the results obtained in the PoP. In the VS these groups gave a sensitivity of 96·5% [93·5% - 98·0%] and a specificity of 99·1% [97·5% - 99·7%] (Data not shown), a slightly better result than found in the PoP study thus confirming the reproducibility of the results over a timeframe of approximately one year.

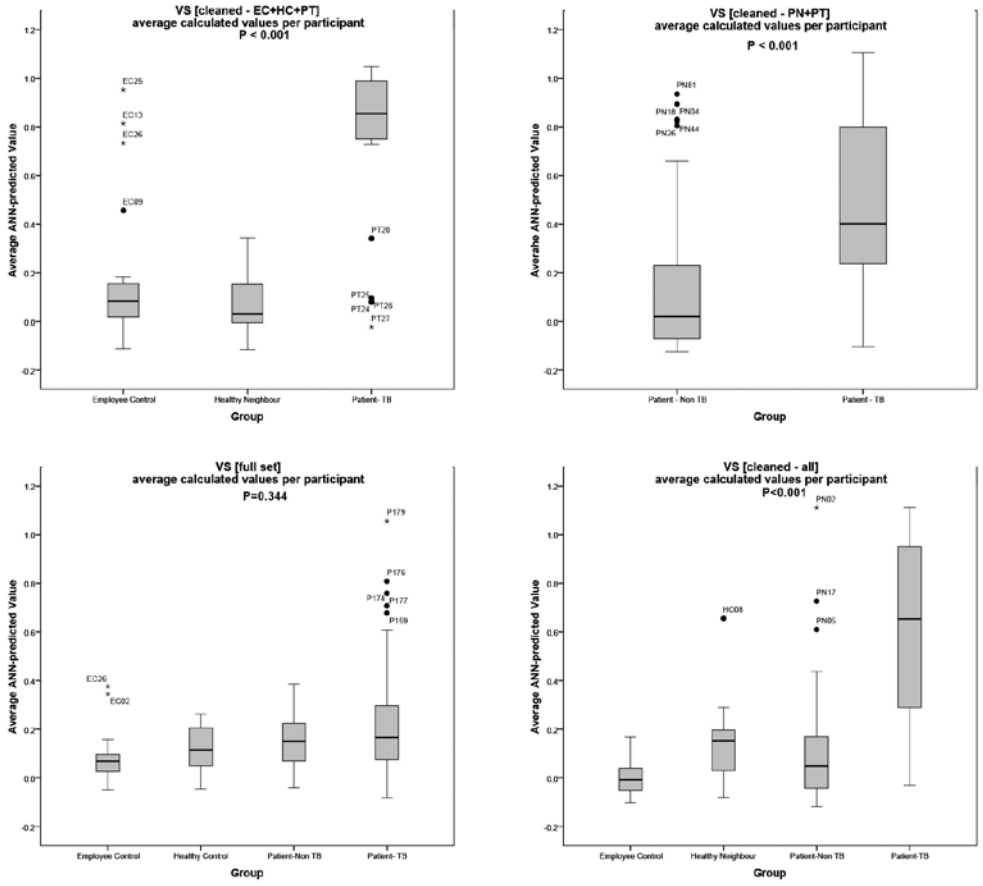


Figure 9.3. Box plot showing the distribution of the average calculated values for different combinations of groups. Outliers are depicted with a circle while extreme values are depicted with an asterisk. Graph A shows the comparison between the distributions of employee controls, healthy neighbors and reference positive patients for the cleaned dataset. Graph B shows the comparison between the distributions of reference negative patients and reference positive patients for the cleaned dataset. Graph C shows the comparison between the distributions of employee controls, healthy neighbors, reference negative patients and reference positive patients for the full dataset. Graph D shows the comparison between the distributions of employee controls, healthy neighbors, reference negative patients and reference positive patients for the cleaned dataset.

The potential influence of the socio-economic status was determined by combining the EC and the HC group. This combined group was compared with the PT group. As can be seen in Figure 9.3a the combined groups of healthy controls were clearly differentiated from TB patients with a sensitivity of 93·5% [91·1% - 95·4%] and a specificity of 85·3% [82·7% - 87·5%].

To determine if the differentiation was based on TB infection or just a general lung problem, we compared the PN with the PT group. As can be seen in Figure 9.3b the

differentiation was more difficult between these two groups than between the healthy controls and the TB patients. Still, the groups could be differentiated with a sensitivity of 76·5% [57·98% - 88·5%] and a specificity of 74·8% [64·5% - 82·9%].

Finally, we assessed the differentiation between the PT group and all other groups including a group of consistent outliers. As can be seen from Figure 9.3c the separation between TB and all non-TB cases was rather poor with a P value of 0·344 resulting in a sensitivity of 50·0% [45·6% - 54·4%] and a specificity of 60·0% [58·0% - 62·0%] (worst case). Since a total of eleven cases were consistently misclassified as false positives regardless of the analysis method used, we removed these cases and recalculated the results (Figure 9.3d) to determine the best-case potential of this diagnosis. This figure shows that under these conditions there was a clear differentiation between TB and non-TB cases, with a sensitivity of 76·5% [72·6% - 80·1%] and a specificity of 87·2% [85·8% - 88·6%]. To show the full potential of the differentiation between TB and non-TB using a Aerekaprobe, the ROC curves for the full range of different subsets are given in Figure 9.4. To construct the ROC curve different threshold were applied to the predicted values resulting in a TP, TN, FP or FN result for each of the samples. Each applied threshold provides a different ratio between sensitivity and specificity, hence a ROC curve.

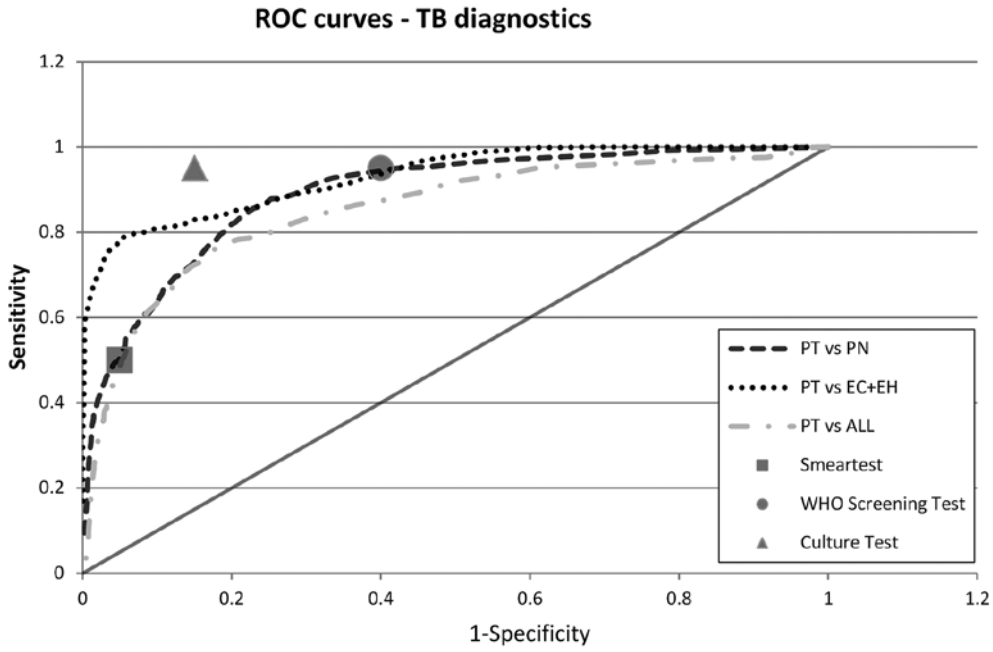


Figure 9.4.

Receiver Operator Curve (ROC) for the different calculated subset. The specified values for the Culture, Smear microscopy and Screening test as defined by the WHO [2] are included for comparison purposes.

Discussion and Conclusion

In the present study we tested the diagnostic accuracy of Aerekaprobe breath analysis for the detection of TB in Bangladesh. In a PoP study we established that the Aerekaprobe was able to differentiate almost perfectly (sensitivity 95·9% and specificity 98·5%) between healthy controls and TB patients. However, the control group was without health complaints while the group of patients was obviously sick, possibly making the differentiation one of healthy versus sick individuals. Besides this, there was also a significant difference in the socio-economic statuses between members of these two groups. The socio-economic influence was covered in the VS by the inclusion of neighbors of participants diagnosed with TB while the general health influence was addressed by the inclusion of all persons enrolling at the hospital without prior knowledge of their TB status. All people enrolling at the SCDC had some form of lung problems but only a fraction of them turned out to be suffering from pulmonary TB. However, we were not able to check if the people diagnosed with TB were also suffering from another, lung, disease.

The less adequate differentiation observed in the VS clearly shows that many factors influence the breath sample of a person as can be seen from the lower sensitivity and specificity results for the larger datasets. Although the differentiation established in the VS was not as good as the differentiation established in the first study (sensitivity of 76·5% vs. 95·9 and a specificity of 87·2% vs. 98·5%) it is still capable of differentiating between TB and non-TB cases with a reliability generally exceeding the classical SSM.

A total of 11 cases, which were marked as non-TB cases by conventional TB culture results, where consistently misclassified as false positives during the different analysis attempts. This fact that there was no distribution of the outliers in false negatives as well as false positives indicates that the source of these outliers is not likely a degree of error in the analysis model, but should be attributed to an unknown external factor. It could imply that these patients were actually infected with TB but the mycobacteria were not detected by the culture analysis. The sensitivity of the reference method causes 10-20% of the pulmonary TB cases to be missed, which would translate to 4-7 false negative cases of the reference test in our study. Besides these cases, also extra-pulmonary cases cannot be detected by the reference method as this form of TB-infection. The prevalence of extra-pulmonary TB in Bangladesh is about 11% [1] which would translate to another 4-5 false negative cases in the present study bringing the total number of expected false-negative cases somewhere between 8 and 12. The total of 11 cases continuously misclassified in our result is in full concordance with this. Alternatively, these 11 cases could have another unknown confounding factor such as HIV co-infection or another condition. Since we were unable to track these cases to determine their actual TB status in a later stage we tentatively removed these cases as moving them to the TB positive group would bias the result. After

the removal of these 11 cases, the quality of the results increased significantly to the reported sensitivity of 76·5% and a specificity of 87·2%. These percentages are however only an indication for the potential diagnostic accuracy of the Aerekaprobe. The actual diagnostic accuracy of the Aerekaprobe can only be determined when a study is conducted that enables the TB status to be determined with 100% certainty since the mathematical model used is completely dependent on the correct subdivision of the samples in the training set. The ANN searches for the most optimal separation between groups based on the results provided. This is the case with every classifier being developed: prior knowledge of the data during the development is required (also known as training stage). This is why training and a validation sets are always needed. With a large population the data is divided into a training and a validation set while with a smaller population, like this study, a Leave-one-out cross validation is used where only one subject is used for validation purposes while the others are used for training. In this case all permutations are tested. The certainty of the current golden standard, as used by us, however, is less than 100% hence hampering the results observed. In a future study the accuracy of the reference method used should be improved, e.g. by following each subject during a prolonged period of time and including multiple samples. This is however only necessary during the acquisition of the dataset used for the development of the algorithm. The more accurate the training set, the more robust the final classifier will be.

In this study 9 permutation-combinations of the two selected sensor types in triplicate were used to include small fabrication differences of the sensors into the data set. The created calibration model is independent of sensor manufacturing variations and can be used for newly manufactured devices without adjustment. The present study shows that an optimized version of the Aerekaprobe, using only 2 sensor types, can be used as the first point of care screening test, in the battle against TB. The Aerekaprobe is an innovative and revolutionary method for TB detection exceeding the sensitivity of smear microscopy combined with only a slightly lower specificity (when compared to the best-case scenario for the full subject set). The Aerekaprobe is fast and does not require the highly-skilled operators as needed for microscopy. This allows application in rural areas where little or no medical infrastructure is available. With the currently reported values the Aerekaprobe would not be suited as diagnostic tool, but neither is the SSM. When looking at the setting of Bangladesh it could be used to proactively search for TB cases locally. Although the specificity is slightly lower than that of the SSM the screening would still be boosting the prevalence by 300-400% in a short time period (<10 min per subject) by untrained personnel while more cases would be detected through the increased sensitivity. When improved data would become available, the Aerekaprobe has the potential for becoming a diagnostic tool in the near future.

References

1. Global tuberculosis control : epidemiology, strategy, financing : WHO report 2009. ISBN 978 92 4 156380 2 WHO/HTM/TB/2009.411
2. Bishop PJ, Neumann G – The History of the Ziehl-Neelsen Stain – Tubercle 1970 51(2): 196–206
3. Diagnostics for tuberculosis : global demand and market potential / TDR, FIND SA. ISBN 92 4 156330 3
4. Fens N, et.al – Exhaled Breath Profiling Enables Discrimination of Chronic Obstructive Pulmonary Disease and Asthma – Am J Respir Crit Care Med 2009 180: 1–7
5. Velásquez A, Durán CM, Gualdron O, Rodríguez JC, Manjarres L – Electronic Nose To Detect Patients with COPD From Exhaled Breath – Olfaction and Electronic Nose: Proceedings of the 13th International Symposium; American Institute of Physics, Melville, N.Y., USA, 2009
6. Nawrath T, et al. – The volatiles of pathogenic and nonpathogenic mycobacteria and related bacteria – Beilstein J. Org. Chem. 2012 8:290–299
7. Mgode GF, et al. – Ability of Cricetomys rats to detect Mycobacterium tuberculosis and discriminate it from other microorganisms – Tuberculosis (Edinb) 2012 92(2):182-6
8. Suckling DM, Sagar RL – Honeybees Apis mellifera can detect the scent of Mycobacterium tuberculosis – Tuberculosis (Edinb) 2011 91(4):327-8
9. Syhre M, Chambers ST – The scent of Mycobacterium tuberculosis – Tuberculosis (Edinb) 2008 88: 317–323
10. Syhre M, Manning L, Phuanukoonnon S, Harino P, Chambers ST – The scent of Mycobacterium tuberculosis – Part II breath – Tuberculosis (Edinb) 2009 89(4): 263–6
11. Phillips M, et al. – Volatile biomarkers of pulmonary tuberculosis in the breath – Tuberculosis (Edinb) 2007 87: 44–52
12. Banday KI.M, et.al – Urine volatile organic compounds could discriminate tuberculosis patients from healthy subjects – Anal. Chem. 2011 83(14): 5526–34
13. Pavlou AK, Turner APF – Sniffing out the Truth: Clinical Diagnoses Using the Electronis Nose – Clin Chem Lab Med 2000 38(2): 99–112
14. Pavlou AK, et al. – Detection of Mycobacterium tuberculosis (TB) in vitro and in situ using an electronic nose in combination with a neural network system – Biosens Bioelectron 2004 20: 538–544
15. Fend R, et al. – Prospects for Clinical Application of Electronic-Nose Technology to Early Detection of Mycobacterium tuberculosis in Culture and Sputum – J Clin Microbiol. 2006 44(6): 2039–2045
16. Pavlou AK – Use of an electronic nose system for diagnoses of urinary tract infections – Biosens Bioelectron 2002 17: 893–899
17. Setkus A – Featuring of bacterial contamination of wounds by dynamic response of SnO2 gas sensor array – Sens Actuators B Chem 2006 115: 412–420
18. Dragonieri S – An electronic nose in the discrimination of patients with non-small cell lung cancer and COPD. – Lung Cancer 2009 64(2): 166–170
19. Kateb B – Sniffing out cancer using the JPL electronic nose: a pilot study of a novel approach to detection and differentiation of brain cancer – Neuroimage 2009 47 :Suppl 2:T5-9. Epub
20. Bruins M, et al. – Device-independent, real-time identification of bacterial pathogens with a metal oxide-based olfactory sensor – Eur J Clin Microbiol Infect Dis 2009 28(7): 775–80
21. WHO, Guidelines on Standard Operating Procedures for Microbiology, Chapter 17 Tuberculosis
22. Rasmus Bro R – PARAFAC. Tutorial and applications – Chemometrics and Intelligent Laboratory Systems 1997 38: 149–171
23. Henrion R – N-way principal component analysis Theory, algorithms and applications – Chemometrics and Intelligent Laboratory Systems 1994 25: 1–23

Part IV

General Discussion & Concluding Remarks

chapter 10

Summarizing discussion

Introduction

The concept of an eNose originated in the 1970s and in 1995 the first medical diagnostic tools were reported [1]. Ever since, there has been research into possible medical diagnostic applications of eNose technology with, in general, promising results [2-15]. Nevertheless, there is no commercially available device or application yet that utilizes eNose technology in medical diagnostics.

An important aspect of the eNose concept, independent of the sensor type used, is the registration of a sensor response caused by a substance, or mixtures of substances, which can only be recognized after a calibration phase. In order to identify a complex odor pattern, it is necessary to develop a database first that defines one or more of the "odor patterns". The impact is that all eNoses currently available rely on the use of a searchable (digital) database where characteristic odor signal patterns obtained during previous measurements have systematically been stored. The current lack of a generic, device independent database is the major cause for the current deficit in real-world applications.

In order to create such a generic, device independent database we need to be able to create device independent (i.e. transferable) calibration models. There are several issues causing the inability to generate such transferable models, even between eNoses of the same type. In this thesis we focus on metal oxide (MO)-sensors since these are incorporated in the eNose used. The proposed solutions, temperature control and data compression, however can also be used by CP and QMB sensors with exception of the temperature control.

Transferability of calibration models

Absence of transferable models has largely blocked commercial diagnostic application of eNoses, especially when high-throughput assays are required. The major issues in this respect are the poor reproducibility of the measurements between presumably identical eNose devices and the difficulty in transferring a once-developed calibration model to another, supposedly equally well functioning, eNose device. This issue has been reported by Koczulla, et.al [16] who conducted an experiment with four identical eNoses at three different locations which showed that the result was highly dependent on the local set-up.

Matters become even more complicated when one aims to combine the datasets of multiple devices, which theoretically should generate comparable results. It turns out that most models reported in literature are device-specific and a model generated with data from another device of the same type usually yields unacceptable results [17-18]. Balaban, et al. [19] used five different mathematical approaches to transfer calibration models between two eNoses and the best approach was only able to correctly classify 8 out of the 12 samples

included in the comparison. This is below acceptable quality levels and indicates that there are variations between the eNose devices, or the conditions under which the devices are used, causing the data to be highly correlated to the specific device used to obtain the initial data.

Due to the high correlation between a specific device and its data it is not worthwhile to create a large set of generic calibration models. These could only be used with that specific device and become obsolete if even only one sensor needs replacement. As a result, at this moment the major area of use for eNoses is in the field of quality and process control. They are being used for detecting deviations of normal situations triggering an investigation into that specific situation and thus only requiring a single device. This trigger can either be an "error" situation where the process is disturbed or a "time for action" situation e.g. the best moment to initiate a next step or stop the process. This specific type of application does not require a database containing calibration models: it only needs an initial period to "learn" the specific base-line situation and can then monitor deviations [20-25].

In our view one of the main causes for the device specific calibrations models is the lack of temperature control of the sensor surface. Therefore, we took a different approach described in chapter 3 towards the issue of device dependent calibration models. Instead of trying to solve the issue mathematically, we increased the reproducibility of the data by controlling the sensor surface temperature. The conductivity of a MO sensor is determined by the (concentration of) oxygen species present in different states (O_2, O_2^-, O radicals, etc.), adsorbed at the surface of the sensor and the ratio between these species [26]. Depletion of adsorbed oxygen species results in a change of flow of electrons resulting in a change of the measured conductivity [27]. The reaction rates in the system depend on the reaction rate constants which in turn are directly related to the temperature. The reaction rate is calculated using the Arrhenius equation.

$$k = Ae^{\frac{-E_a}{RT}}$$

$$k : Reaction\ rate\ constant$$
$$A : Pre-exponential\ factor$$
$$E_a : Activation\ energy$$
$$R : Universal\ gas\ constant$$
$$T : Absolute\ temperature\ in\ °K$$

A variation in temperature will, therefore, generate a different odor profile for the same substance making it almost impossible to create a generic odor profile for a population of sensors which is related to a specific substance. Precise temperature control is thus of utmost importance when employing MO sensors.

To illustrate the influence of temperature control, we conducted an experiment fixing the noses at specific temperatures and measuring similar samples. This experiment is described in chapter 3. We introduced the total inertia ($\phi 2$) as a measure for the heterogeneity within the measured data. Using the total inertia we showed that the influence of a temperature shift of only 15° (the manufacturing process results in a variance of up to 50° within a single batch) results in a variance within the data which is 10-15 times larger than other inter-sensor differences related, among others, to surface morphology. This large difference is certainly a major issue hampering a transferable calibration model. With the experiments conducted in chapter 6 and chapter 9 we showed that the data generated by our eNoses are interchangeable with respectively 30 and 2 different eNoses. The results clearly show that controlling the temperature is a necessity for the creation of a transferable calibration model.

In this thesis we use eNose modules (Figure 1.4) developed by The eNose Company (Zutphen, The Netherlands). These eNose modules have a patented on-board temperature control and are fitted with a single MO-sensor. Due to the temperature control these modules can be used for large scale implementations. We use 4 distinct chemical types of sensors (Applied Sensors Gmbh, Reutlingen, Germany). Three of these sensors use SnO2 as reactant with either Pd or Pt in varying concentrations as embedded catalyst while the fourth type applies WO3 as reactant. These 4 types are employed in the Aerekaprobe in triplicate bringing the total number of sensors to 12. The use of a triplicate of identical types enables a better selection of the best-suited sensor(s) for a given type of sample and reduces the chances of creating calibration models which rely on sensor-specific information as the model is forced to generalize for the same sensor type. We form a set of unique measurements by creating all unique permutations of the different sensor types. When using just one type of sensor we would form only 3 unique measurements for each sample (A1, A2, A3) and with all 4 types we can create 81 (34) unique measurements for each sample (A1B1C1D1, A1B1C1D2, …., A3B3C3D2, A3B3C3D3). When building a calibration model using these measurements (part of) the inter-sensor differences will be eliminated.

Although the use of a single sensor within an eNose module tends to limit the specificity of VOC detection, the application of a thermal cycle heating program to the sensor compensates for this lack of chemical specificity, facilitating the creation of a “virtual array” of sensors [27]. The sensor thermal cycle used in the work performed in this thesis varies between 220°C and 320°C, following a sinusoid shaped profile to facilitate a smooth transition between the temperatures.

Amount of data collected

The creation of generic calibration models requires a large amount of data collected with several physically distinct but operationally identical devices to compensate for or eliminate the inevitable inter-device differences. Most published articles only employ a single device and if more than one device is used the results are reported separately per device [28]. The aforementioned lack of transferability combined with the limited availability within a location makes it virtually impossible to collect sufficient data to develop generic calibration models. We used a set of 30 eNoses in our experiment described in **chapter 6** showing that it is possible to combine the information of eNoses and at the same time collect a sufficient amount of data to create a transferable calibration model.

Data handling

In every experiment the issue of so-called confounding variables exists. These are variables which are not accounted for but do in fact influence the measurements and thereby the observed and/or calculated relationships between cause and effect. Most likely also in our TB study, described in **chapter 9**, where the socio-economic status difference between both groups in the "proof of concept" (PoC) study will have influenced the results. The healthy controls in the PoC study all have a regular income and place to live while the patients mostly have no place to stay and no regular income. This may influence their breath profile. Therefore, we included the "healthy neighbour" group in the following "validation study" (VS) study. We clearly showed that the socio-economic status of the subject acts as a confounding variable in this study, the result of the VS study were even better for the same populations while they became less when the new groups where introduced. The overall results show that a good choice of the test population is important but a check for potential confounding variables is good practice.

With the ever growing amount of data collected by eNoses there is also another, statistical, problem one needs to be aware of: voodoo correlations. The term 'voodoo correlations' was coined by Vul et al. [29] and represents statistically relevant correlations observed within the data which just exists by coincidence. In contrast to a spurious correlation, which is caused by the aforementioned confounding variables, the voodoo correlations cannot be removed by changing the experimental set-up. If the dataset is large enough the statistical chance of finding a variable or set of variables that classifies the samples perfectly but unjustified is rather large and special care needs to be taken when analyzing this data.

In **chapter 4** we show the impact of the amount of measured variables on the classification results by creating a dataset consisting of pure white-noise data and tested

this with a varying number of subjects and variables. If the number of variables is 0·5-1 times the amount of subjects a chance of 90% exists of finding a perfect separation between two target classes, even if all data are completely random. When the number of variables increases to 2-3 times the number of subjects this chance increases to 100%.

When one tries to find a specific component that can be used as a so-called biomarker the process is different. The search for a biomarker does not focus on the entire profile recorded but it tries to find a (small set of) variable(s) that correctly classify(ies) the samples. To mimic this, we created a new white-noise data set containing 8 subjects and we showed that we need 126 random variables to find a single variable capable of predicting all 8 cases correctly, whereas a prediction rate of 87·5% can already be reached with only 14 variables. One needs at least the same number of subjects as the number of measured/used variables, preferably even double this number to avoid voodoo correlations.

In **chapter 4** we also presented 2 different methods to reduce the chance of finding voodoo correlations. The first method is leave-one-out cross validation (LOO-CV). When all data are used to create the model, a perfect separation between the samples may be found. This perfect separation is in almost every case only perfect for that specific dataset and a new sample will not fit perfectly within the model. The LOO-CV approach removes 1 subject from the dataset, creating a model with the remainder of the data and using the removed subject to validate the model. This is iterated in turn until all subjects have been used for validation at least once. Paradoxically, we have found that the reliability of the model results obtained using LOO-CV improved with an increasing number of variables. Without the LOO-CV approach the result rapidly go to 100% correct classification while the results with LOO-CV show a (small) decrease below 50% when the number of variables increases.

The second method involves the reduction of the number of variables. The 'normal' procedure involves the extraction of features from the signal as we did with the pathogen detection. This kind of variable reduction also implies a loss of information and the selection of the features may be biased by the knowledge and expectations of the researcher. We developed a compression algorithm, in collaboration with Elegant Mathematics (Ottweiler, Germany), which uses proprietary tensor decomposition as feature extraction technique to compress the complex multi-way data into simple 1-dimensional vectors which can be used by most mainstream pattern recognition techniques. This technique is not biased and retains almost all data. The amount of data to be retained can be chosen and will influence the compression rate of the algorithm. On average a single sample Aerekaprobe measurement contains 4 (sensor types) * 32 (thermal cycle points) * 30 (10 min @ 3 cycles per minute) = 3840 variables. Using the tensor compression algorithm this is reduced to 10-20 variables while retaining 99% of the information originally present in the data.

Sample treatment

Using an eNose, samples are measured in a consistent manner. The sensors are exposed to a sample during a defined period of time after which sensors are allowed to recover for another defined period of time resulting in the typical shark-fin like shaped measurement curves as shown in figure 1.7.

For most samples this approach works fine as long as effects such as diffusion [30] and decomposition by sunlight [31] are taken into consideration e.g. by keeping a fixed time period between sample collection and measurement. The influence of storage time and choice of sample bags on the stability of volatile breath constituents has been assessed by Mocahlski et.al [32]. This has shown that a sample stored in a Tedlar® bag, the best performing one by far, should be used within 6 hours if the bag is filled to its maximum. Biggest issue is the loss of water (up to 10%) and heavier volatiles (mass > 90, 20-40%). Matters get more complicated when dealing with dynamic samples in which the headspace composition is actively changing over time which is the case with bacterial samples.

In **chapter 6** we introduce a new measurement method especially suited for dynamic samples. When working with active samples it is next to impossible to control the conditions in such a way that the headspace composition at a set point in time remains at a constant value. This continuous variance in headspace compositions poses the question whether an observed differentiation is really based on actual differences between samples or merely on differing experimental conditions. Measuring the same sample on consecutive days could result in a better differentiation than observed between different samples on the same day. Our new method follows the sample during the entire process instead of picking a single, fixed moment in time. This approach compensates for variances in medium compositions, initial inoculum size, etc. by continuously watching for emerging fingerprints that are unique for a certain species.

Using our eNose modules we studied 3 different applications in order to develop potential solutions for the issues regarding transferability of calibration models as mentioned above.

Identification of bacterial pathogens

Significant research has been performed on the identification of micro-organisms using an eNose [2, 33-35]. Al these experiments however share the same problems which keep them from being translated into real diagnostic tools. Firstly, all reported studies used just a single eNose and hence were carefully avoiding the problem of the development of a transferable calibration model. Secondly, all studies only included a single strain per bacterial species, therewith disregarding the biodiversity within the species. Finally, all studies concern the analysis of samples after a fixed period of time allowing for confounding variables due to species dependent timing characteristics.

In **chapter 6** we simultaneously addressed all three issues listed above. We used a total of 30 independently produced MonoNose devices (shown in figure 5.1) to collect sufficient data to create a generic model. The requirement of transferability was supported by the embedded temperature control and the samples were monitored continuously during culturing allowing for time-warping matching algorithms. Next to the identification of the pathogens we also studied the influence of the growth conditions and initial microbial concentration.

We used a set of 11 clinically relevant species with two to eight individual, genotypically different strains to account for intra-species diversity [table 5.1].

During the measurements we observed 2 generic groups of patterns. One group of species developed almost no measureable odor throughout the monitored period while the other group showed a very large increase in measurement signal somewhere during cultivation. We designated these two groups as the “weak” and “strong” signal producers respectively. Although we did not manage to define the actual cause of this relationship it appears that the weak signal group contains the gram-positive species while the gram-negatives include the strong signal producers. Both the weak and the strong odor producing species can be identified with comparable specificity.

We also studied the influence of the medium composition and initial CFU counts on the measurements. We created a set of media with different amounts of amino acid and salt concentrations. We measured a set of media with a single standardized initial inoculum and recorded the maximum observed conductivity during the entire experiment for each of the compositions. The surface plot for the set measured with E. coli and K. oxytoca is shown in figure 5.2 and clearly shows that both develop differently when the growth media is altered.

The influence of a varying initial inoculum was tested with a single E. coli strain. We created 4 different initial CFU counts, the average results of these measurements are shown in figure 5. It can be clearly seen that the initial inoculums concentrations influence the headspace composition in time as the amount of biomass increases

It is obvious that a single measurement taken at an arbitrary point in time cannot be used when medical samples are involved. To use the odor developing capacities of a species to identify that species, one needs to monitor the headspace continuously and match the observed kinetic odor profile against a set of known, pre-recorded profiles. The characteristic part of a profile will be visible during a fairly short period of time which can occur throughout the entire measurement period depending on the initial CFU count.

For the identification of the samples we derived features from the signal over time and created a generalized feature set for each of the species and associated time-warping matching algorithms. The extracted features for both Salmonella and Klebsiella species where very similar within the genus and species could, therefore, only be identified at the genus level. An example of a weak and a strong odor and their respective feature set is shown in figure 5.6.

We used this generalized feature set for all species tested. The results in Table 5.2 show that the test specificity varied from 100% for C. difficile to 67% for E. cloacae with an overall average of 87%. All characteristic sub-features in this work where observed between 6 to 8 hours after initial inoculation at maximum.

A major advantage of this approach is its non-destructive nature, the biomass developed during the measurement remains available for classic determination, and it can improve the time-to-detection for a selection of species while it does not disturb the current work flow.

Acute Postpartum Metritis Detection

In bovine disease diagnostics research has been performed towards the use of an eNose for the detection of changes in perineal odors in cattle [36]. A recent literature review on diagnostics of acute postpartum metritis (APM) [37] showed there are still no reliable methods for the detection of APM In practice, a veterinarian visually inspects the samples and the physical appearance combined with the body temperature of the cow is then used as an indication of a potential APM.

In **chapter 7** we first determined the reliability of the Aerekaprobe by comparing the Aerekaprobe with the olfactory evaluation of 16 knowledgeable investigators. The results generated by this panel represented a sensitivity of 75·0% and a specificity of 60·1% while the Aerekaprobe showed a sensitivity of 92·0% and specificity of 100% on the same samples. The study also revealed a considerable subjectivity of the human panel while the Aerekaprobe showed a larger repeatability between the samples when measured multiple times.

In **chapter 8** we conducted a second study where a total of 70 cows were sampled at days 2, 5 and 10 after calving. During the sampling period the APM status was also determined.

For the purpose of this experiment we introduced a third group besides 'healthy' and 'sick', the so-called 'pre-sick' group. This group contained the samples taken when there was no diagnosis of APM by the veterinarian while the following sample revealed APM. After introducing the concept of the 'pre-sick' cows, the results showed a sensitivity of 74·0% and a specificity of 95·0%. There were 2 sick cows diagnosed as being healthy by the Aerekaprobe. However these two cows where already diagnosed with APM at day 2 and were receiving an antibiotic treatment as of that day. A differentiation based on the causing pathogen was not possible with these samples. Most likely is the opportunistic nature of the APM infection and the cause of this in combination with the large number of mixed cultures.

Tuberculosis Diagnosis

Diagnosis of tuberculosis based on exhaled air has been proven to be possible by Phillips et al. in 2007 [38] where the breath of the subjects was collected on a sorbent tube and analyzed with GC/MS. The need for laboratory equipment such as GC/MS renders this method unsuited for use as a screening tool despite the good results (82·6% sensitivity and 100% specificity). In 2011, van Beek et.al [15] performed an experiment using a handheld device to measure the nitric oxide levels in the exhaled air. The results were not as good as in the 2007 GC/MS study (sensitivity 78% and specificity of 62%) but the major flaw in this study is the fact that the control group contained healthy workers from a different socio-economic environment than the TB patients. So it is likely that confounding variables were introduced in that study. Also, all TB patients were a-priory classified with at least 1 smear positive result.

In **chapter 9** we describe a breath analysis experiment in Dhaka, Bangladesh. During these experiments we asked the subjects, either suffering from TB or 'healthy' (defined as not having TB), to provide a breath sample by filling a Tedlar™ bag with exhaled breath. The sample was then measured with one of the two employed Aerekaprobe devices. We documented a sensitivity of 95·9% and specificity of 98·5%. After these promising results we initiated a new and much larger study with an extended population of possible TB patients and their volunteering 'healthy' neighbours. In this validation study (VS) we found a sensitivity of 93·5% and a specificity of 85·3% discriminating healthy controls from TB patients, and a sensitivity of 76·5% and specificity of 87·2% when identifying TB patients within the entire test-population.

One of the major issues with a study like this one is the relatively poor results of the gold standard diagnostic technology. Besides its low sensitivity, the gold standard is also unable to detect extra-pulmonary TB. For the PoC we only measured people with a 3+ SMM indication, for the VS we enrolled a large group in the hospital without prior knowledge of their TB

status. The “gold standard” used misses about 10-20% of all TB cases and within our setting there was also an 11% prevalence of extra pulmonary TB. The quality of a computational model depends on the reliability of training data and with the current method we cannot provide a 100% error-free training set. A further study needs to be conducted in a setting where the patient is tracked during an extended period of time to ensure the real TB status at the time breath samples were taken.

At this point the results show a good potential for development of a screening tool for active pulmonary TB; this tool will be easy to use and the results are based on differentiation between groups actually present at a screening location.

Future Perspectives

The research described in this thesis brought us closer to the realization of a large scale eNose application. We have solved a major issue regarding the transferability of calibration models by controlling the temperature, we have shown the necessity of continuous monitoring when working with samples containing live micro-organisms and we have determined the need of either variable reduction or the use of LOO-CV when handling data in order to improve the robustness of the interpretative models. One of the major difficulties with continuous measurements concerns the group of microorganisms which produce an odor which is hard to record by the sensors. This group mainly consists of gram-positive bacteria which contain the clinically relevant species *Staphylococcus aureus*. It can either be that this group produces substances which are hard to record or it could be that there is a rather small amount of odor being produced, an amount so small that it quantitatively reacts at the sensor surface at the same rate in which it is produced without causing a signal increase. One way to manage this would be to decrease the interval in which the sensor is being exposed to the headspace allowing an accumulation of produced VOCs to a higher amount. We conducted an experiment to validate this hypothesis. During this experiment the sensor was intermittently exposed to the headspace during a fixed period of time. The first nose measured the sample headspace continuously, the second nose measured clean air during 20 minutes allowing the sample headspace to accumulate and then measured the sample headspace for 5 min after which the sequence repeated. The third nose uses the same sequence as nose two but with an accumulation period of 45 minutes between two exposures. Figure 10.1 shows the results of these measurements.

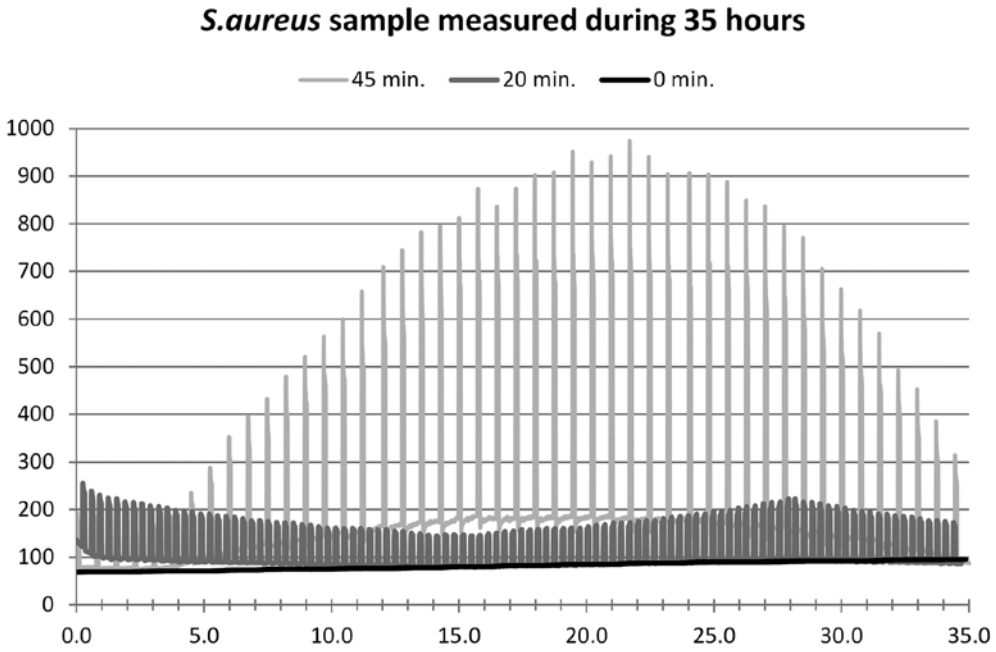


Figure 10.1. S.aureus sample measured during 35 hours with eNoses using different exposure intervals.

This experiment indicates that there is a relationship between the production rate of the compounds and the measurement interval, and that the reactions taking place at the surface of the eNose seem to "consume" or alter the compounds. This experiment was only conducted with a single strain of *S.aureus* and more work needs to be done to prove that this holds true for other species that seem to produce no measurable odor with the current measurement principle. The influence of this change in measurement principle on the measured signal also needs to be verified for other species. It might be possible that the distinction between the measurable odor producing species (the strong odor group) is lost due to this alteration. Another point of concern regarding this change is the increase in time-to-detection: the accumulation and measurement intervals also dictate the smallest possible detection interval.

Figure 10.2. The Aeonose, developed for breath analysis

The Aerekaprobe has been developed into a new version optimized for direct breath analysis. This new device, the Aeonose, reduces the operations needed to obtain and measure exhaled-breath samples and the possible errors related to these operations. The Aeonose takes the sample directly without needing the subject to fill an intermediary sample bag, figure 10.2 shows the Aeonose.

With this Aeonose several different studies are being conducted at this moment. The TB study has been extended and is conducted in The Netherlands, Kenya and Indonesia at this point. New studies to investigate the diagnostic potential are started for Head & Neck Cancer, Dehydration, Lung Transplants, Angina Pectoris, Complex Regional Pain Syndrome, Diabetes and Acute Mountain Disease. For some of these conditions there are previous studies implying the possibility to differentiate based on odor while the others are based on fact that the people treating these patients noticed a difference in odor between patients.

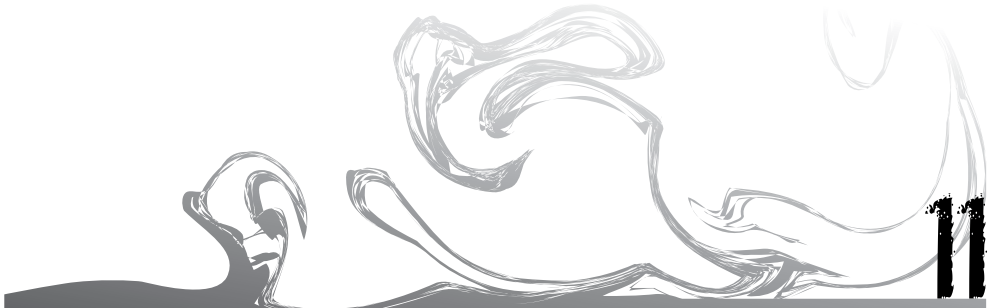
All these studies are pilot studies since there is no way of predicting if our eNose will be capable of “smelling” the same variations as either the other eNose or the humans.

The metritis study is being continued while a parallel study towards the diagnostics of mastitis is started as well. In contradiction to metritis, where there is no actual standard, there is a diagnostic gold standard for mastitis. This standard is however not really reliable and is not able to distinguish between the different causes of mastitis.

References

1. Rossi V, Talon R, Berdagué J – Rapid Discrimination of Micrococcaceae species using semiconductor gas sensors – Journal of Microbiological Methods 1995 24:183-190
2. Moens M, Smet A, Naudts B, et al. – Fast identification of ten clinically important micro-organisms using an electronic nose – Letters in Applied Microbiology 2006 42:121-126
3. Pavlou AK, Turner APF – Sniffing out the truth : Clinical Diagnosis Using the Electronic Nose – Clin Chem Lab Med 2000 38(2):99-112
4. Gibson TD, Prosser O, Hulbert JN, et al. – Detection and simultaneous identification of microorganisms from headspace samples using an electronic nose – Sensors and Actuators B 1997 44:413-422
5. Pavlou AK, Magan N, Sharp D, Brown J, Barr H, Turner APF – An intelligent rapid odour recognition model in discrimination of Helicobacter pylori and other gastroesophageal isolates in vitro – Biosensors & Bioelectronics 2000 15:333-342
6. Bessa V, et al. – Detection of volatile organic compounds (VOCs) in exhaled breath of patients with chronic obstructive pulmonary disease (COPD) by ion mobility spectrometry – International Journal for Ion Mobility Spectrometry 2011 14(1):7-13
7. Kodogiannis V, Wadge E – The use of gas-sensor arrays to diagnose urinary tract infections – Int J Neural Syst 2005 15(5):363-76
8. Pavlou AK, Magan N, Jones JM, Brown J, Klatser P, Turner AP – Detection of Mycobacterium tuberculosis (TB) in vitro and in situ using an electronic nose in combination with a neural network system – Biosens Bioelectron 2004 04(3):538-44
9. Aathithan S, Plant JC, Chaudry AN, French GL – Diagnosis of Bacteriuriaby Detection Of Volatile Organoc Compounds in Urine Using an Automated Headspace Analyzer with Multiple Conducting Polymer Sensors – Journal of Clinical Microbiology 2001 July:2590-2593
10. Ritaban D, Ritabrata D – Intelligent Bayes Classifier (IBC) for ENT infection classification in hospital environment – BioMedical Engineering OnLine 2006 5:65
11. Yates JW, et al. – Data reduction in headspace analysis of blood and urine samples for robust bacterial identification – Comput Methods Programs Biomed 2005 79(3):259-71
12. Kateb B – Sniffing out cancer using the JPL electronic nose: a pilot study of a novel approach to detection and differentiation of brain cancer – Neuroimage 2009 47(Suppl 2):5-9
13. Setkus A – Featuring of bacterial contamination of wounds by dynamic response of SnO2 gas sensor array – Sens Actuators B Chem 2006 115:412–420
14. Banday KM, et al. – Urine volatile organic compounds could discriminate tuberculosis patients from healthy subjects – Anal. Chem. 2011 83(14):5526–34
15. Van Beek SC, Nhung NV, Sy DN, Sterk PJ, Tiemersma EW, Cobelens FG – Measurement of exhaled nitric oxide as a potential screening tool for pulmonary tuberculosis – Int J Tuberc Lung Dis. 2011 15(2):185-92
16. Koczulla AR, Hattesohl A, Biller H, Hofbauer J, Hohlfeld J, Oeser C, Wirtz H, Jörres RA – Vergleich von vier baugleichen elektronischen Nasen und drei Messaufbauten – Pneumologie 2011 65(8):465-470
17. Oded S, Liran C, David H – On mappings between electronic noses – Sensors and Actuators B 2005 106:76–82
18. Zhang L, et al. – On-line sensor calibration transfer among electronic nose instruments for monitoring volatile organic chemicals in indoor air quality – Sensors and Actuators B 2011 160:899-909
19. Balaban MO, Korel F, Pdabasi AZ, Folkes G – Transportability of data between electronic noses: Mathematical methods – Sensors and Actuators B 2000 71:203-211
20. Haugen JE, et al. – Rapid control of smoked Atlantic salmon (Salmo salar) quality by electronic nose: Correlation with classical evaluation methods – Sensors and Actuators B 2006 116:72–77
21. Romain AC, Godefroid D, Kuske M, Nicolas J – Monitoring the exhaust air of a compost pile as a process variable with an e-nose – Sensors and Actuators B 2005 106:29-35
22. Falasconi M, Pardo M, Sberveglieri G, Ricc`o L, Bresciani A – The novel EOS835 electronic nose and data analysis for evaluating coffee ripening – Sensors and Actuators B 2005 110:73–80

23. Aparicio R, Rocha SM, Delgadillo I, Morales MT – Detection of rancid defect in virgin olive oil by the electronic nose – J Agric Food Chem. 2000 48(3):853-60
24. Zondervan C, Muresan S, de Jonge HG, van Velzen EU, Wilkinson C, Nijhuis HH, Leguijt T – Controlling Maillard reactions in the heating process of blockmilk using an electronic nose – J Agric Food Chem. 1999 47(11):4746-9
25. Collins MA, Moy L – The Electronic Nose for Process Control – Neural Networks: Artificial Intelligence and Industrial Applications 1995 297-302
26. Delgado RD – Thesis Tin Oxide Gas Sensors:An Electrochemical Approach – University of Barcelona 2002
27. Pearce TC, Schiffman SS, Nagle HT, Gardner JW – Handbook of Machine Olfaction – WILEY-VCH 2003
28. Kolk A, et al. – Electronic-nose technology using sputum samples in diagnosis of patients with tuberculosis – J Clin Microbiol. 2010 48(11):4235-8
29. Vul E, et al. – Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition Perspect – Psychol. Sci. 2009 4:274–90
30. Beauchamp J, Herbig J, Gutmann R, Hansel A – On the use of Tedlar® bags for breath-gas sampling and analysis – J Breath Res. 2008 2(4):046001
31. Sturini M, Speltini A, Maraschi F, Profumo A, Pretali L, Fasani E, Albini A – Sunlight-induced degradation of soil-adsorbed veterinary antimicrobials Marbofloxacin and Enrofloxacin – Chemosphere 2012 86(2):130-7
32. Mochalski P, King J, Unterkofler K, Amann A – Stability of selected volatile breath constituents in Tedlar, Kynar and Flexfilm sampling bags – Analyst. 2013 138(5):1405-18
33. Dutta R, Hines EL, Gardner JW, Boilot P – Bacteria classification using Cyranose 320 electronic nose – BioMedical Engineering OnLine 2002 1:4
34. Lykos P, Patel PH, Morong C, Joseph A – Rapid Detection of Bacteria from Blood Culture by an Electronic Nose – The Journal of Microbiology 2001 39-3:213-218
35. Trincavelli M, Coradeschi S, Loutfi A, Söderquist B, Thunberg P – Direct Identification of Bacteria in Blood Culture Samples using an Electronic Nose – IEEE Trans Biomed Eng. 2010 57(12):2884-90
36. Lane AJ, Wathes DC – An electronic nose to detect changes in perineal odors associated with estrus in the cow – J Dairy Sci. 1998 81(8):2145-50
37. Sannmann I, Arlt S, Heuwieser W – A critical evaluation of diagnostic methods used to identify dairy cows with acute post-partum metritis in the current literature – J Dairy Res. 2012 79(4):436-44
38. Phillips M, et al. – Volatile biomarkers of pulmonary tuberculosis in the breath – Tuberculosis 2007 87:44–52



11 Nederlandse samenvatting

M.Bruins

Partially published in
Analyse **2013** 7:196-198

Een neus voor medische diagnostiek

Een neus voor medische diagnostiek

Iedereen die ooit bacteriën heeft gekweekt weet dat bepaalde bacteriesoorten een specifieke geur hebben tijdens groei. Een ervaren analist of microbioloog kan bij een kweek in een aantal gevallen op geurkenmerk vaststellen om welk micro-organisme het gaat. Geur wordt al sinds de oudheid gebruikt als een diagnostisch hulpmiddel bij het opsporen van infectieziekten. Hippocrates noemde bijvoorbeeld de diagnostische waarde van geur al in zijn werk genaamd "Aforismen" [1] geschreven in 400 voor Christus. Hij schreef in deel IV: 81 "Als een patiënt bloed, schilfers en pus in de urine uitscheidt, en een sterke geur heeft, is dat een indicatie voor een ulceratie van de blaas". In 1923 is er een publicatie door Omlanski verschenen die de geur producerende eigenschappen van micro-organismen beschrijft [2]. In het huidige tijdperk van moderne immunologische en moleculaire diagnostiek, is de kunst van de geurdiagnose langzaam aan het verdwijnen. Het biedt echter perspectieven voor real-time identificatie van infecties in de microbiologische diagnostiek[3].

Tijdens de groei produceren micro-organismen een breed scala van verbindingen. De beschikbare componenten in de omgeving en de (genetisch bepaalde) mogelijkheden van het micro-organisme bepalen welke stoffen specifiek geproduceerd worden. Het micro-organisme is in staat om zijn metabolische routes aan te passen als de beschikbaarheid van de componenten wijzigt, bijvoorbeeld door uitputting van het groeimedium. De (variatie in) metabolische activiteit is een kenmerk van bepaalde microbiële geslachten of soorten en kan dienen als een soort vingerafdruk.

Elektronische neus al sinds jaren 70

Het concept van een elektronische neus (eNose) ontstond in de jaren zeventig van de vorige eeuw [4]. De beschikbaarheid van computers maakte het mogelijk patronen te herkennen in metingen opgebouwd uit complexe signalen. In feite is dit analoog aan de manier waarop mensen karakteristieke geuren herkennen, vandaar de naam "eNose". Een eNose registreert geurprofielen en net als een biologische neus weet men niet wat de exacte chemische stoffen zijn die men ruikt maar toch herkent men de geur.

Een belangrijk aspect van het eNose concept, onafhankelijk van het gebruikte sensortype, is dat de sensorrespons veroorzaakt door een stof of mengsel van stoffen alleen herkend kan worden na een kalibratiefase. Er moet eerst een database gebouwd worden met een verzameling geurprofielen. Het gevolg hiervan is dat alle eNoses afhankelijk zijn van zo'n database. Het gebrek aan een *generieke* database is een van de grootste oorzaken voor het (tot nu toe) ontbreken van grootschalige diagnostische toepassingen. In **hoofdstuk 3** hebben we laten zien dat de temperatuur controle een essentieel onderdeel is voor het overdraagbaar maken van de kalibratie modellen. De door ons gebruikte eNoses beschikken

over een goede temperatuur controle waardoor de kalibratie modellen overdraagbaar zijn. Dit is in alle experimenten aangetoond door het gebruik van een variabel aantal neuzen voor ieder experiment.

Er bestaan verschillende types eNoses waarvan Conducting-Polymer (CP), Quartz-Micro-Balance (QMB) en Metal-Oxide (MO) de meest gebruikte zijn. Al ons werk is uitgevoerd met MO-sensoren van The eNose Company omdat hiermee, in tegenstelling tot andere sensoren, wel de mogelijkheid bestaat om een apparaat onafhankelijke kalibratiemodel te maken. De gebruikte temperatuurcontrole is hierbij cruciaal, een module met sensor is te zien in figuur 1.4.

Huidige toepassingen eNose

Op dit moment wordt de eNose vooral commercieel gebruikt in de voedingsmiddelen-industrie en milieubewaking. Dit zijn relatief eenvoudige toepassingen omdat hier het belangrijkste doel de detectie van “abnormale” situaties is. Dit soort applicaties is niet afhankelijk van de hierboven genoemde database. De eNose meet continu een proces en detecteert een afwijking van de normale situatie. Deze afwijking fungeert vervolgens als een trigger voor het opstarten van een onderzoek naar de oorzaak van de afwijking. Op dit moment worden er geen eNoses gebruikt in medische microbiologische laboratoria, hoewel onderzoek naar het mogelijke gebruik van eNoses voor medische diagnostische toepassingen veelbelovend is.

Headspace - en ademanalyse

Voor de eNose zijn er twee hoofdvormen van monstersoorten. De eerste soort is het klassieke monster met een potentiele bacteriële infectie, het alom bekende monster op de microbiologische laboratoria. Op het lab worden deze monsters opgekweekt waarna identificatie plaatsvindt aan de hand van differentiële biochemische reacties en microscopie. De identificatie van het monster wordt gestart nadat een zichtbare microbiële groei heeft plaatsgevonden (meestal 16-24 uur na inoculatie). Hierdoor komt de totale doorlooptijd, van binnenkomst monster tot identificatie, uit tussen 48 uur voor “hoge microbiologische load” monsters zoals urine en feces en 168 uur voor “lage microbiologische load” monsters zoals bloed. Het tweede soort monster voor de eNose betreft uitgeademde lucht. Beide soorten worden hierna kort beschreven.

Geurfilm in plaats van momentopname voor identificatie

De eNose-experimenten, die in de literatuur beschreven worden om micro-organismen aan te tonen en/of te identificeren gaan uit van een initiële kweek van biomassa. Het opkweken van de biomassa zorgt er voor dat de neuzen minder hinder ondervinden van de initiële monstermatrix. Tevens wordt het aantal aanwezige micro-organismen, en daarmee de gevoeligheid, verhoogd. Het grootste probleem van deze experimenten is dat ze uitgaan van een "momentopname" een aantal uren na inoculatie. Deze methode is gebaseerd op de standaard werkwijze in de traditionele klinische microbiologie. De ontwikkeling van de geur is afhankelijk van in het medium aanwezige nutriënten en de snelheid waarop en volgorde waarin deze geconsumeerd worden. Andere parameters zijn onder meer de initiële microbiële load van het monster, de replicatiesnelheid en het soort organisme.

Een momentopname kan voldoende informatie bevatten zodat het idee ontstaat dat een identificatie op basis van deze methode mogelijk zou zijn. Er wordt bij deze methode echter geen rekening gehouden met factoren zoals het initiële inoculum, de samenstelling van het medium en de incubatietemperatuur. Al deze factoren beïnvloeden de dynamica van het proces waardoor de kans op het vinden van een specifieke geur voor een organisme een stuk kleiner is dan het vinden van een dynamica artefact. Hieronder worden de invloeden van de bovengenoemde factoren en de invloed op het geurpatroon kort uitgelegd.

Als er een hogere initiële bacteriële load in een monster aanwezig is, zullen de beschikbare nutriënten sneller geconsumeerd worden. Hierdoor moet er sneller overgeschakeld worden naar een andere metabolische route waarbij de geproduceerde stoffen wijzigen. In figuur 6.2 wordt een viertal experimenten getoond waarbij alle variabelen gelijk zijn gehouden met uitzondering van het begin inoculum. In de grafiek is de tijd sinds inoculatie uitgezet tegen het gemeten signaal. Het is duidelijk te zien dat de kenmerkende steile flank op een ander tijdstip plaatsvindt waardoor een meting op x uur na inoculatie zowel voor als na de flank kan liggen. De hier gebruikte *E.coli* produceert een relatief eenvoudig signaal. Als een monster een complexer en/of vaker wisselend signaal genereert wordt het belang van het tijdstip nog groter.

Een hogere replicatiesnelheid als gevolg van een hogere incubatietemperatuur heeft als direct gevolg dat de nutriënten sneller geconsumeerd worden en er een afwijkende geurdynamica ontstaat.

Het beschikbare nutriëntenmengsel bepaalt welke metabole processen plaatsvinden evenals de volgorde van route die door een specifiek micro-organisme wordt gevolgd. In figuur 6.4 is een aantal experimenten weergegeven voor *E. coli* en *K. oxytoca* bij een variërende hoeveelheid zout en aminozuren in het medium. Het is duidelijk te zien dat beide

organismen op andere wijze reageren op de verschillen en dat de samenstelling een grote invloed heeft op de geproduceerde geur.

We hebben onze experimenten, beschreven in **hoofdstuk 6**, om het identificatie potentieel van de eNose te bepalen niet uitgevoerd met een momentopname, maar de monsters vanaf het moment van inoculatie gevolgd. Hierdoor is het niet van belang op welk tijdstip een geurprofiel optreedt. In tegenstelling tot een "momentopname" wordt het gemeten signaal nu continu vergeleken met bekende patronen. We hebben gedurende dit experiment met 30 eNoses een set van 11 verschillende, klinisch relevante, soorten gemeten waarbij van elk soort minimaal 2 verschillende stammen zijn gebruikt. De gevonden specificiteit varieerde van 100% voor *Clostridium difficile* tot 67% voor *Enterobacter cloacae* met een gemiddelde van 87%. De gemiddelde tijd tot identificatie bedroeg hier 6-8 uur in tegenstelling tot de standaard 48 uur.

Metritis detectie in melkvee

Naast het gebruik van de geurfilm zoals hiervoor beschreven kunnen monsters met een potentiele bacteriële inhoud ook direct gemeten worden. Deze monsters worden dan niet eerst aan kweek onderworpen maar worden direct na afname gemeten of opgeslagen bij -20°C om op een later tijdstip gemeten te worden. Door de opslag bij -20°C staat de bacteriële groei stil en worden de aan groei gerelateerde problemen voorkomen. Op deze manier hebben we een aantal vaginale afscheiding monsters van runderen vlak na de bevalling, het moment waarop de kans op een metritis het grootst is, gemeten. Op dit moment is de diagnose van metritis nog gebaseerd op subjectieve waarnemingen van de dierenarts, deze beoordeeld de afscheiding op samenstelling, uiterlijk en geur om te bepalen of er sprake is van metritis. In het onderzoek beschreven in hoofdstuk 7 hebben we de eNose vergeleken met de diagnose uitgevoerd door een aantal veeartsen. Hieruit bleek dat de eNose zowel specifieker als reproduceerbaarder werkte dan de veeartsen. Hierna hebben we een nieuwe set monsters gemeten om de sensitiviteit en specificiteit van de eNose voor het diagnosticeren van metritis te bepalen. Hoewel er goede resultaten zijn behaald qua metritis detectie (sensitiviteit van 74·0% en specificiteit van 95·0%) bleek het niet mogelijk om de veroorzakende bacteriesoort te identificeren. Dit werk is beschreven in hoofdstuk 8.

Ademanalyse

In de uitgeademde lucht bevinden zich een groot aantal (500 – 1000) verschillende vluchtige verbindingen. Hierbij zoeken we niet naar specifieke stoffen (de zogenoemde biomarkers), maar vergelijken het adempatroon van gezonde en zieke personen. Het idee is dat we niet rechtsreeks de metabole gassen van de bacteriën meten, maar het veranderde metabole adempatroon van het menselijk lichaam onder invloed van de ziekte. Dit laatste wordt aannemelijk als we bedenken dat ook veel andere ziektebeelden een veranderend adempatroon ten gevolge hebben, ook als geen sprake is van een bacteriële infectie. Voorbeelden daarvan zijn diverse soorten kanker en diabetes.

In het geval van tuberculose (TB) hebben we in 2010 een studie uitgevoerd in Bangladesh, beschreven in **hoofdstuk 9**. In de eerste fase is een proof-of-principle studie uitgevoerd met 30 deelnemers (15 patiënten en 15 gezonde controlepersonen). Bij deze studie moesten de proefpersonen een zak vullen met uitgeademde lucht. De lucht werd vervolgens over de sensoren van een eNose geleid. Na data-analyse bleek de eNose zeer goed in staat te zijn deze twee groepen personen van elkaar te onderscheiden.

Vervolgens is er een validatiestudie uitgevoerd met ongeveer 200 personen. Deze groep bevatte ook mensen met andere ziektebeelden en gezonde personen uit vergelijkbare sociale klasse als de patiënten. Ook in dit geval was een goede scheiding mogelijk tussen TB-patiënten enerzijds en de andere groepen anderzijds. Er bleek echter ook een groep personen te zijn, die volgens de eNose TB positief was, maar kweek-negatief bleek. Mogelijk was sprake van extra-pulmonale TB, maar het was niet meer mogelijk de patiënten te traceren: na het onderzoek gingen ze terug naar hun dorpen en waren ze niet meer bereikbaar.

Perspectieven voor nieuwe diagnostiek

De resultaten in Bangladesh waren zo veelbelovend dat we een eNose hebben ontwikkeld specifiek voor ademanalyse, de Aeonose. Het nieuwe apparaat vertoont gelijkenis met de apparatuur die de politie gebruikt voor alcoholcontrole doordat de patiënt niet meer een zak hoeft te vullen met uitgeademde lucht, maar direct in het apparaat ademt (zie figuur 10.2). Tijdens het onderzoek dient de patiënt gedurende 5 minuten rustig via het mondstuk te ademen. In het (eenmalig te gebruiken) mondstuk bevindt zich een koolstoffilter zodat de patiënt schone lucht inhaleert. In dit tijdsbestek wordt de uitgeademde lucht over de sensoren geleid en bemonsterd. Een bacteriefilter voorkomt besmetting van de sensoren.

Op basis van deze Aeonose zijn nieuwe TB-studies gestart op locaties waar de patiënten goed gevolgd kunnen worden om zaken als extra-pulmonaire TB achteraf uit te kunnen

sluiten. Er zijn nog diverse openstaande vragen zoals: wat is de invloed van co-infecties (bijvoorbeeld HIV) op het adempatroon van patiënten met TB? Is het adempatroon van TB-patiënten (geografische) regio afhankelijk? Kunnen we inderdaad extra-pulmonale TB of wellicht zelfs latente TB detecteren? Door de uitgebreidere opzet van de nieuwe TB studies hopen we deze vragen te kunnen beantwoorden.

Momenteel worden de eerste experimenten uitgevoerd in het tuberculose centrum van het UMCG (Beatrixoord) en worden asielzoekers gescreend in het asielzoekerscentrum in Ter Apel. In Kenia is een studie gestart naar TB bij kinderen en in Yogyakarta zijn we op 5 verschillende locaties begonnen met het verzamelen van data.

Naast het TB onderzoek zijn we ook studies naar onder andere longziekten, hartproblemen, diabetes, en tumoren gestart om de potentiele bruikbaarheid van de Aeonose voor deze aandoeningen vast te stellen.

Op grond van de huidige resultaten kan de conclusie getrokken worden dat het gebruik van de eNose in de klinische diagnostiek veelbelovend is. Dit geldt zowel voor het versnellen van de klassieke microbiologische diagnostiek als voor niet-invasieve detectie van TB en wellicht nog een aantal andere ziekten.

Referenties

1. Marks E – The aphorisms of Hippocrates from the latin version of Verhoofd – Collins & Co 1817
2. Omelianski VL. Aroma producing Microorganisms J Bacteriol. 1923 July; 8(4): 393–419
3. Pavlou AK, Turner APF. Sniffing out the truth : Clinical Diagnosis Using the Electronic Nose. Clin Chem Lab Med 2000 38(2):99-112
4. Pearce TC, Schiffman SS, Nagle HT, Gardner JW. Handbook of Machine Olfaction WILEY-VCH, 2003
5. Bruins M, Bos A, Petit PL, et al. Device-independent,real-time identification of bacterial pathogens with a metaloxide-based olfactory sensor Eur. J. Clin. Microbiol. Infect. Dis., vol.28, pp. 775–780, Jul 2009
6. Marcel Bruins, Zeaur Rahim, Albert Bos, Wendy W.J. van de Sande, Hubert Ph. Endtz, and Alex van Belkum - Diagnosis of active tuberculosis by e-nose analysis of exhaled air - Tuberculosis 2012 93:232-238

chapter 12

Curriculum Vitae

Marcel Bruins was born on April 14th 1973 in Almelo. In 1989 he finished the MAVO at the Ichthus College in Vriezenveen. After this he followed the study Chemical Engineering & Laboratory, differentiation Analytical Chemical Techniques at the vocational college "Twents MBO College" in Hengelo, which he completed in 1994. He started the study Chemistry & Technology, Laboratory Education, differentiation Analytical Chemistry at the "Hogeschool Enschede" college in Enschede which he concluded in 1998 with the project "Differential evolution strategy as selection method for demulsifier blends" conducted at Servo BV in Delden. For this company he moved to Dubai as trainee oilfield engineer. Mid 1999 he returned to The Netherlands and started working for TAS Toptalent BV in Baarn as software programmer stationed at the ABN-AMRO headquarters in Amsterdam. He started working for C-it BV early 2000 as R&D employee focusing at electronic nose technology. After a few sidesteps as IT-consultant and C# programmer at the same company he returned to the eNose activities and part of the work involved experiments at Erasmus Medical Center. These experiments turned out to be the start of his PhD research on the medical diagnostic potential of an eNose at the Department of Medical Microbiology and Infectious Diseases of the Erasmus Medical Center in Rotterdam, under supervision of Prof.dr.dr. A. van Belkum, Dr.ing W.W.J. van de Sande and Dr.ir. A.Bos. After completion of his PhD he will continue to work for The eNose Company (started after C-it BV split up in a eNose and an IT company in May 2013) and be involved in the exploration and development of the medical diagnostic potential of the eNose.

Marcel Bruins werd geboren op 14 april 1973 te Almelo en groeide op in Westerhaar. In 1989 haalde hij zijn MAVO diploma aan het Ichthus college in Vriezenveen. Hierna volgde hij de MLO opleiding Analytische Chemie aan het Twents MBO College in Hengelo dat hij afrondde in 1994. Meteen aansluitend is hij de opleiding Analytsiche Chemie gaan volgen aan de Hoge School Enschede dat hij in 1998 afrondde met het afstudeeronderzoek "Differentiële evolutie strategie als selectie methode voor demulsifier blends" uitgevoerd bij Servo BV te Delden. Voor Servo is hij daarna naar Dubai vertrokken als trainee oilfield engineer. Halverwege 1999 is hij terug gekeerd naar Nederland waar hij bij TAS-toptalent in Baarn is gaan werken als programmeur en gedetacheerd werd bij de ABN-AMRO in Amsterdam. Hij is in het voorjaar van 2000 in diest getreden bij C-it BV als R&D medewerker voor de elektronische neus. Via een korte omweg als IT-consultant en C# programmeur bij hetzelfde bedrijf is hij teruggekeerd naar de eNose activiteiten. Onderdeel van deze activiteiten betrof een aantal experimenten bij het Erasmus Medisch Centrum. Deze experimenten bleken de start te zijn van een promotie onderzoek naar het medische diagnostisch potentieel van de eNose op de afdeling Medische Microbiologie & Infectie Ziekten bij het Erasmus Medisch Centrum in Rotterdam onder supervisie van Prof.dr.dr. A. van Belkum, Dr.ing. W.W.J. van de Sande en Dr.ir. A. Bos. Na het voltooien van zijn promotie onderzoek zal hij blijven werken bij The eNose Company (ontstaan na het splitsen van C-it BV in een eNose en een IT bedrijf in mei 2013) en daar betrokken zijn bij het onderzoeken naar en ontwikkelen van het medisch diagnostisch potentieel van de eNose.

chapter 13

List of publications

A. W. Heuwieser, **M. Bruins**, A. Bos, I. Sannmann, and O. Burfeind - Diagnosis of acute postpartum metritis by e-nose analysis of vaginal discharge - *Theriogenology* **2013** submitted

B. **M.Bruins** - Een neus voor medische diagnostiek - Analyse **2013** 7:196-198

C. **M. Bruins**, J.W. Gerritsen, W.W.J. van de Sande, A. van Belkum, and A. Bos - Enabling a transferable calibration model for metal-oxide type electronic noses - *Sensors and Actuators* B **2013** 188:1187-1195

D. I. Sannmann, O. Burfeind, V. Suthar, A. Bos, **M. Bruins**, and W. Heuwieser - Evaluation of odor from vaginal discharge of cows in the first 10 days after calving by olfactory cognition and an electronic device - *Journal of Dairy science* **2013** 96(9):5773-5779

E. **M. Bruins**, Z. Rahim, A. Bos, W.W.J. van de Sande, H.P. Endtz, and A. van Belkum - Diagnosis of active tuberculosis by e-nose analysis of exhaled air - *Tuberculosis* **2012** 93:232-238

F. W.B.van Leeuwen, K. Maquelin, **M. Bruins**, and others - Chapter 10 - "Biofysische Methoden" of *Moleculaire Diagnostiek* – Heron-reeks, Syntax media **2011** 421-426

G. **M. Bruins**, A. van Belkum ,and A. Bos - Chapter 5 - "The Use of Electronic Nose Devices in Clinical Microbiology" from *The Role of New Technologies in Medical Microbiological Research and Diagnosis* – Bentham Publishers **2011** 90-101

H. **M. Bruins**, A. Bos, P.L.C. Petit, K. Eadie, A. Rog, R. Bos, G.H. van Ramshorst, and A. van Belkum - Device-independent, real-time identification of bacterial pathogens with a metaloxide-based olfactory sensor - *Eur J Clin Microbiol Infect Dis* **2008** 28:775-780

chapter 14

PhD portfolio

Name:	Marcel Bruins
Institute:	Erasmus University Rotterdam
Department:	Medical Microbiology and Infectious Diseases
Promotor:	Prof.dr.dr. A. van Belkum
Copromotors:	Dr.ing W.W.J. van de Sande and Dr.ir A.Bos

Conferences and Seminars

- International Symposium on Olfaction and Electronic Nose 2009, Brescia, Italy
- Scientific spring meeting of the Dutch Society for Medical Microbiology (NVMM) 2009, Papendal, The Netherlands
- Research day, Medical Microbiology and Infectious Diseases (MMIZ), 2008, 2009 & 2010, Rotterdam, The Netherlands

Research abroad

- 2 weeks Tuberculosis study at ICDDR,B, Dhaka, Bangladesh. Supervision Prof.dr. H.P. Endtz and Dr. Z. Rahim.

Teaching

- Lectures during the summer course of the Research Master in Infection and Immunity of the Erasmus University Rotterdam. 2010, 2011 and 2013
- Supervisor of bachelor students 2008 – 2011
- Training of field engineer in Dhaka, Bangladesh

Presentations

- *An Electronic Nose as a tool for pathogen detection and identification* - Research day, Medical Microbiology and Infectious Diseases 2008
- *Device-independent, real-time identification of bacterial pathogens with a metaloxide-based olfactory sensor* - Scientific spring meeting of the Dutch society for Medical Microbiology 2009
- *Device independent bacteria identification* - International Symposium on Olfaction and Electronic Nose 2009
- *Bacteria detection & identification with an electronic nose* - ICDDR,B 2009
- *Identificatie zonder [exact] te weten* - Research day, Medical Microbiology and Infectious Diseases 2009
- *Hoe herken je een Teringlijder, Tuberculose (snel) diagnostiek* - onderwijsdag voor de internisten-infectioloog in opleiding 2009
- *Identification without Knowing, Diagnostics with an Electronic Nose* - FU Berlin, Department of Veterinary Medicine 2013

15 Dankwoord

Het laatst geschreven en veelal eerst en meest gelezen hoofdstuk van een proefschrift. Iedereen wil weten "*sta ik er in*". Voor een (te) groot deel van de lezers zal hierop het antwoord "*Nee*" zijn. Er zijn in de loop van dit onderzoek heel veel mensen geweest die op deze of gene wijze hebben bijgedragen om dit eindresultaat tot stand te laten komen. Ik weet zeker dat ik hier mensen ga vergeten. Als jouw naam ontbreekt, is dat zeker niet omdat ik jouw bijdrage niet gewaardeerd heb maar hooguit door een te beperkte omvang van mijn geheugen. Bij deze speciaal voor jou: BEDANKT voor alles, en als je graag persoonlijk bedankt wil worden neem dan even contact met me op.

Beste **Alex** zonder jou was dit boekje er nooit geweest. Het lijkt, en is, alweer een eeuwigheid geleden dat jij samen met Pieter de trip naar die rare gasten met hun elektronische neus in Zutphen hebt gemaakt. Wie had toen kunnen vermoeden dat het nog hierop zou uitlopen. Na de eerste positieve resultaten op mijn bureau stond je ineens weer zelf op je eigen lab 'echt' werk te doen in plaats van (elektronische) stapeltjes data te verschuiven. Nadat ook die resultaten er veelbelovend uit zagen was het ineens gegroeid van een experimentje naar een promotie onderzoek en was ik ineens bezig de halve huisraad van Zutphen naar Rotterdam te verhuizen. Had je de consequenties van je "*Ja*" op mijn vraag of ik dat onderzoek dan niet kon gaan doen maar van te voren kunnen overzien…

Nu staan we dan een paar jaar later hier om er een eind aan te breien. Dank je wel voor je geduld met mijn microbiologische onwetendheid. De overstap naar BioMérieux heeft je ongelofelijk snelle en grondige correcties niet beïnvloed. De vol gekalkte pagina's vallen nog steeds per omgaande op mijn bureau. In de loop van de tijd heb ik zelfs de vulpen leren waarderen maar geloof me dat een vulpen niet het handigste correctie middel is gedurende een onrustige vlucht… Dank je wel voor alles en ik hoop dat we contact blijven houden in de toekomst.

Beste **Albert**, het is alweer meer dan 16 jaar geleden dat ik bij Servo Delden kwam binnenlopen om bij jou stage te lopen en te gaan afstuderen. Na afronding vertrok ik naar Dubai maar we hielden contact en ik heb nog steeds geen spijt van mijn bevestigende antwoord op jouw vraag "*Zou je wel eens weer echt aan het werk willen*?" nadat ik weer terug was waardoor ik bij C-it aan het werk ging. We hebben een hoop meegemaakt sindsdien. Dank je wel voor jouw unieke manier van tegen dingen aankijken, reden voor menig discussie maar met een onveranderd positief eindresultaat. Mede dankzij jou heb ik straks ook het T-shirt….

Beste **Wendy**, dank je wel voor al je hulp de afgelopen jaren. Toen ik met mijn onderzoek begon kreeg jij ineens een eNose onderzoek op je bordje. Ik had me geen betere rots in de medisch, microbiologische branding kunnen wensen. Jouw snelheid van corrigeren kan zeker wedijveren met die van Alex, soms leek het wel of jullie daadwerkelijk een wedstrijdje deden wie het snelst een gecorrigeerde versie retour kon krijgen.

Henny & Arjen, dank je wel voor de mogelijkheid die jullie me gegeven hebben om dit promotie onderzoek onderdeel van mijn dagelijks werk te maken. De combinatie van beide werelden was niet altijd even gemakkelijk maar mede dankzij jullie heb ik daar mijn weg in kunnen vinden.

Jan-Willem, dank je wel voor het duwtje en het behouden van het hierdoor gecreëerde momentum. Ik hoop dat we het momentum kunnen behouden en versnellen in de toekomst.

Andre, het is nog maar iets meer dan een jaar geleden dat jij in beeld kwam. Wat kan er veel veranderen in een jaar tijd. Bedankt dat het afronden van mijn promotie tussen alles door heeft kunnen laveren.

Bert, al die jaren tegenover elkaar en de laatste tijd in afzonderlijke hokjes maar ook dat staat onze gesprekken niet in de weg. Dank je wel voor je luisterend oor, frisse invalshoeken en vooral stabiele handen. Ik moet er niet aan denken dat ik zelf dat priegelwerk met de printplaatjes had moeten uitvoeren, dan was er zeker niets zinnigs gemeten.

Beste "Handjes", **Aimee, Remco, Erwin** en **Evelien**, zonder jullie was dit boekje er nooit geweest. Dank je wel voor al het werk dat jullie gedaan hebben, discussies die we hebben gevoerd en vooral het geduld dat jullie met me hebben gehad als ik eens per week weer binnen kwam vallen.

Marianne, gaat deze simpele jongen uiteindelijk nog echt promoveren ook…

Door de jaren heen ben je een bijzonder "handje" geworden. Je begon gewoon aan je stage en voor je het door hebt zit je ineens 2 weken in Bangladesh. En alsof dat nog niet genoeg is wordt je jaren later ook nog eens in je vakantie naar het Vlietland gesleept om daar van alles voor me te doen. Dank je wel voor alles…

Pieter, dank je wel dat je in den beginne samen met Alex de hele reis naar Zutphen hebt gemaakt. Door de jaren heen hebben we allerlei uithoeken van het potentieel van de micro-organismen verkend en zonder jouw kennis en hulp zou ik daar zeker verdwaald zijn. Laten we hopen dat we in de toekomst nog een paar onbekende uithoeken kunnen verkennen.

Wolfgang, meinen herzlichen Dank an dich, dass du mich in die veterinärmedizinische Diagnostik eingeführt hast. Einen speziellen Dank dafür, dass du nach Rotterdam gekommen bist, um Argumentationsgegner während mein Disputation zu sein. Ich hoffe, dass wir in der Zukunft gemeinsam große Schritte in der Metritis Diagnostik machen werden.

Hubert, dank je wel voor je begeleiding tijdens de TB studies in Bangladesh, een ervaring die ik niet had willen missen en dat je secretaris van mijn kleine commissie wilde zijn. Ik hoop je in de toekomst nog vaker tegen te komen.

Dear **Zeaur**, thanks for your collaboration in our TB-study. Thanks to you Dhaka became more than just another foreign hotel.

Beste **MMIZ'ers**, dank jullie wel voor de gastvrijheid waarmee ik al deze jaren door jullie ben ontvangen. Ik heb erg genoten van alle gesprekken als ik weer eens een keer met

mijn laptop in de gang zat te werken. Ik was telkens weer aangenaam verrast door jullie betrokkenheid, alle vragen "*Zou dit ook kunnen…*", ideeën "H*eb je hier wel aan gedacht..*" en informatie "*Ik kwam dit artikel tegen en volgens mij is dat wel iets voor jou*" heb ik met dank aangenomen. Iedereen stond altijd klaar om te helpen als ik weer eens iets bedacht had wat toch lastig te realiseren bleek, hetzij door hiaten in mijn microbiologische kennis hetzij tijd technisch omdat het in die ene dag dat ik er was niet allemaal paste.

Beste **collega's**, dan hier eindelijk een overzicht van alle dingen waar ik me de afgelopen jaren heb bezig gehouden. Dank jullie wel voor de getoonde interesse in alles wat ik aan het doen was. Mochten er na het lezen van dit boekje nog vragen zijn kom dan maar een keer binnen lopen en dan zal ik proberen ze te beantwoorden.

Lieve **Chantal**, dank je wel voor alle gesprekken en WhatsApp sessies. Nooit hadden ze te maken met mijn onderzoek maar juist daardoor waren ze extra waardevol. Ik ben erg blij dat je dit ook met me wilt delen door mijn paranimf te zijn. Ik weet zeker dat we nog heel wat gesprekken gaan hebben in de toekomst, ik kijk er nu al naar uit…

Lieve **Axel**, dank je wel voor alle afleiding waardoor ik ook nog andere dingen heb gedaan de laatste tijd. Alvast sorry voor het feit dat het pinguïn pakje dat ik straks draag toch niet het pinguïn kostuum, inclusief maki op de schouder, is dat jij voor me had bedacht…

Lieve **Marlea**, de laatste plek in het rijtje gereserveerd voor de belangrijkste (volwassen) persoon in mijn leven. Toen een aantal jaar geleden de kans zich voordeed om een promotie onderzoek te gaan doen was jij de eerste die zei: "*Dat moet je gaan doen*". Dat heb ik uiteindelijk ook gedaan en dankzij jou niet aflatende support is het nog gelukt ook. Dank je wel voor het uitlenen van je oren als ik weer eens wat kwijt moest en je ogen als er weer eens iets gelezen moest worden. Zonder jouw steun en de ruimte die je me gegeven hebt betwijfel ik of het me gelukt zou zijn. Meestal is dit de plek waarop er beloofd wordt dat er vanaf nu weer meer tijd zal zijn voor andere dingen, ik durf het niet aan om je een dergelijke belofte te doen. We weten beiden dat "rust" en "tijd" dingen zijn die niet echt in mijn woordenboek staan. Maar ik kan je wel beloven dat ik, voorlopig, niet aan iets zal beginnen dat net zoveel impact heeft als dit promotie onderzoek.

Eindelijk kan je het leuke jurkje kopen dat je jezelf beloofd hebt toen ik hieraan begon en ik ben erg blij dat je, in dat leuke jurkje, ook de laatste stap van dit traject met me wilt delen door mijn paranimf te zijn.