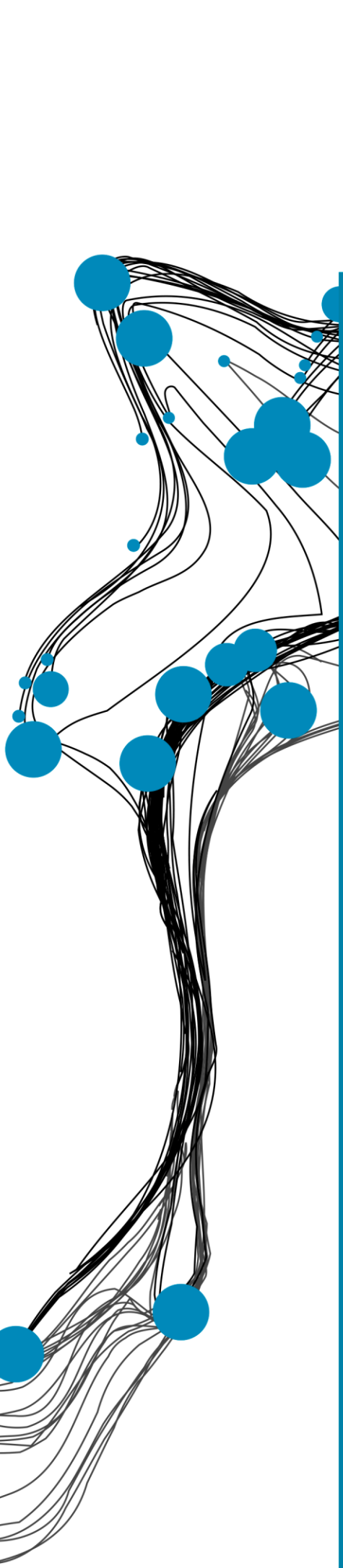


MASTER THESIS

DETECTION OF DYSFUNCTIONAL BREATHING AND UNCONTROLLED ASTHMA IN CHILDREN USING RESPIRATORY SOUND RECORDINGS

Marieke Massa

BIOMEDICAL ENGINEERING
FACULTY OF ELECTRICAL ENGINEERING, MATHEMATICS AND COMPUTER SCIENCE
BIOMEDICAL SIGNALS AND SYSTEMS

EXAMINATION COMMITTEE
Dr. Ir. M. Tabak[1,2]
M.R. van der Kamp MSc[1,3]
Dr. Ir. A.P. Berkhoff[4]
P.B. Keijzer MSc[3]

[1] Department of Biomedical Signals and Systems, University of Twente, Enschede, Netherlands
[2] Roessingh Research and Development, Enschede, Netherlands
[3] Department of Pediatrics, Medisch Spectrum Twente, Enschede, Netherlands
[4] Department of Applied Mechanics, University of Twente, Enschede, Netherlands

05-06-2021

MST Medisch Spectrum Twente een zorgvan Ziekenhuis

UNIVERSITY OF TWENTE.

ABSTRACT

Rationale Dysfunctional breathing (DB) is a common respiratory condition in children which greatly affects a child's quality of life. Symptoms of DB are often similar to symptoms encountered with exercise-induced bronchoconstriction, which is indicative of uncontrolled asthma. As these two conditions require different treatment approaches, differentiation is indispensable. At this moment, DB is only identified by exclusion of uncontrolled asthma and other possible causes and this procedure requires demanding repetitive forced breathing manoeuvres. In this study, the potential of respiratory sounds recorded during an exercise challenge test (ECT) to detect DB and uncontrolled asthma was investigated.

Methods Literature research was performed and four clinicians assessed sound recordings made with a directional microphone during ECTs to determine relevant characteristics in respiratory sounds for detecting DB and uncontrolled asthma. Test measurements with healthy individuals and pediatric patients were analyzed to assess the quality of the sound recordings and to discover what information can be extracted from the recordings. The microphone settings were changed and a microphone was added to the measurement setup to improve the signal acquisition. Machine learning algorithms were applied to 28 sound recordings in order to develop a classification algorithm. The 28 recordings were from children with either DB, uncontrolled asthma or no established respiratory diagnosis.

Results 32% of the sound recordings was correctly classified by clinicians. According to literature and clinicians, the nature of adventitious sounds and the moment at which symptoms occur in a sound recording were the most important characteristics for detecting DB and uncontrolled asthma. The test measurements resulted into signals with much ambient noise in which the discernability of a respiratory pattern varied amongst recordings. The adjustments in the measurement setup did not improve the quality of recordings. Recordings made after an ECT did show clear respiratory patterns. The machine learning approach did not result in a proper classification algorithm.

Conclusion The results of the study imply that at this point, sound recordings made during an ECT cannot be used to detect DB and uncontrolled asthma in children. Future research should focus on further improving the measurement setup to minimize disturbing sounds and extending the existing knowledge on relevant characteristics in sounds recordings. Higher quality sounds and this extended knowledge may provide objective detection of pediatric DB and uncontrolled asthma in the future.

SAMENVATTING

Achtergrond Disfunctioneel ademen (DA) is een respiratoire aandoening die frequent voorkomt bij kinderen en grote invloed heeft op de kwaliteit van leven van een kind. Symptomen van DA vertonen veel overlap met symptomen gezien bij inspanningsastma, indicatief voor ongecontroleerde astma. Beide aandoeningen vereisen een andere behandelmethode, wat correcte differentiatie essentieel maakt. Op dit moment wordt DA vastgesteld door het uitsluiten van ongecontroleerde astma en andere mogelijke oorzaken van de klachten. Deze procedure vergt herhaaldelijke geforceerde ademmanoeuvres, wat veeleisend is voor een kind. In deze studie is de mogelijkheid onderzocht om, met behulp van ademgeluiden opgenomen tijdens een inspanningstest, DA en ongecontroleerde astma vast te stellen.

Methode Literatuuronderzoek is uitgevoerd en vier clinici hebben geluidsfragmenten, opgenomen met een richtmicrofoon tijdens inspanningstesten, beoordeeld om relevante karakteristieken in ademgeluiden te kunnen identificeren voor het vaststellen van DA en ongecontroleerde astma. Testmetingen met gezonde personen en pediatrische patiënten zijn geanalyseerd om de kwaliteit van geluidsopnames te beoordelen en vast te stellen welke informatie uit de opnames verkregen kan worden. Vervolgens zijn de microfooninstellingen aangepast en is er een microfoon aan de meetopstelling toegevoegd om de signaalacquisitie te verbeteren. Machine learning algoritmes zijn toegepast op 28 geluidsopnames om een classificatiealgoritme te ontwikkelen. De 28 opnames waren afkomstig van kinderen met DA, ongecontroleerde astma of geen respiratoire diagnose.

Resultaten 32% van de geluidsfragmenten was correct geclassificeerd door clinici. De aard van bijgeluiden en het moment van voorkomen van symptomen in een geluidsopname kwamen uit zowel literatuur als bij clinici naar voren als belangrijkste karakteristieken om DA en ongecontroleerde astma vast te stellen. De testmetingen resulteerden in signalen met veel omgevingsruis waarbij de waarneembaarheid van een respiratoir patroon varieerde tussen de opnames. De aanpassingen in de meetopstelling leidden niet tot verbetering van de kwaliteit van de opnames. Respiratoire patronen waren wel duidelijk waarneembaar in opnames gemaakt direct na een inspanningstest. De machine learning aanpak resulteerde niet in een adequaat classificatiealgoritme.

Conclusie De resultaten van de studie wijzen erop dat geluidsfragmenten opgenomen tijdens een inspanningstest op dit moment niet gebruikt kunnen worden om DA en ongecontroleerde astma vast te stellen bij kinderen. Vervolgonderzoek dient zich te richten op het verder verbeteren van de meetopstelling om verstoringsgeluiden te minimaliseren en het uitbreiden van bestaande kennis over relevante karakteristieken in geluidsopnames. Geluidsopnames met hogere kwaliteit en deze uitgebreide kennis zouden objectieve detectie van pediatrische DA en ongecontroleerde astma mogelijk kunnen maken in de toekomst.

PREFACE

Before you lies my thesis which was written to complete my master assignment for Biomedical Engineering at the University of Twente. I have been working on this assignment from October 2020 until June 2021. The start of the assignment was a little earlier than expected: originally, I would be residing in Lisbon for my internship from September until the beginning of December. Unfortunately, due to reasons we are all familiar with, travelling to Lisbon was not possible and the internship had to be postponed. Luckily, it was possible to reverse the order of the internship and the master assignment. And luckily, I was welcome to perform the master assignment in collaboration with the pediatrics department in Medisch Spectrum Twente, where I also performed my multidisciplinary assignment for the Technical Medicine bachelor three years ago.

The data which was gathered during this assignment, has challenged me to think of creative ways to extract relevant information from imperfect measurements. Although the results from my study cannot be used for objective detection of respiratory diseases in children yet, I hope that I have provided the department of pediatrics with some useful insights on this topic to continue further research.

Before moving on to the real content of this thesis, I would like to thank some people who have helped me to establish this work.

First of all, I would like to thank my daily supervisors Mattiènne and Pascal for always being available for questions, discussions or help with measurements. When I needed it, it was always possible to schedule a meeting on short notice and although our regular meetings were often short, you provided me with clear directions to continue my research.
I would like to thank Monique for helping me to find a suitable assignment in a short period of time, as my study plans suddenly had to change. In addition, your feedback on (parts of) my thesis has given me valuable insights on how to bring more structure in my not so standard study approach.
I want to thank Arthur for being the external supervisor of my committee and helping me with some of the more technical questions I had during my research.
Although not an official member of my graduation committee, I would like to thank Boony for always giving me inspiration for new research directions and providing me with relevant information from the outpatient clinic, even in the weekends.

In addition, I want to thank my roommates who were my main company during this assignment. Although I often had to spend the early mornings on my own, you were always willing to give me some distraction with a cup of coffee, a walk or an episode of Temptation Island.
I want to thank my parents for their endless support and pride during my entire study but also specifically during this assignment. I especially appreciated the nice chats we sometimes had on Thursday mornings during our online coffee breaks!
Last but certainly not least, I want to thank my boyfriend Lars for providing me with unfailing support. You were always willing to help me, either with the content of my assignment as a

graduated Biomedical Engineer or with distractions from the assignment like making the weekly 'woordgraptogram' over the phone. Thank you for all the help during this assignment.

I hope you enjoy reading my thesis.

Marieke Massa

Enschede, June 2021

LIST OF ABBREVIATIONS

DB	Dysfunctional breathing
dB	Decibels
CAS	Continuous adventitious sounds
CORSA	Computerised respiratory sound analysis
DAS	Discontinuous adventitious sounds
DWT	Discrete wavelet transform
ECT	Exercise challenge test
EIB	Exercise induced bronchoconstriction
EIIS	Exercise induced inspiratory symptoms
EILO	Exercise induced laryngeal obstruction
EMG	Electromyography
FEV_1	Forced expiratory volume in 1 second
FFT	Fast Fourier transform
FIR	Finite-duration impulse response
HVS	Hyperventilation syndrome
Hz	Hertz
ILO	Inducible laryngeal obstruction
KNN	K-nearest neighbors
LDA	Linear discriminant analysis
LMS	Least mean squares
MARM	Manual Assessment of Respiratory Motion
MMSE	Minimum mean-square-error
MUAP	Motor unit action potential
MST	Medisch Spectrum Twente
ND	No respiratory diagnosis
NQ	Nijmegen Questionnaire
QDA	Quadratic discriminant analysis
RLS	Recursive least squares
RSA	Respiratory sound analysis
STFT	Short-Time Fourier Transform
SVMs	Support vector machines
UA	Uncontrolled asthma
UWT	Undecimated wavelet transform

CONTENTS

1 INTRODUCTION

Dysfunctional breathing (DB) is a common respiratory condition in both children and adults and is described as *an alteration in the normal biomechanical patterns of breathing that result in intermittent or chronic symptoms* [1]. DB has a great impact on the quality of life, as patients experience symptoms like shortness of breath and chest discomfort which often occur during exercise [2]. Children with DB may not be able to participate in physical activities on the same level as their peers. In addition to the evident reduced performance and functioning of these children, this has an emotional impact on their lives.[2, 3]

1.1 Prevalence and diagnosis of dysfunctional breathing

The symptoms pediatric DB patients present at the outpatient clinic, are very similar to the symptoms encountered with exercise induced bronchoconstriction (EIB) which is indicative of uncontrolled asthma (UA). Due to these similarities, the co-existence of DB with other respiratory conditions such as asthma and the scarcity of data from the pediatric population, the prevalence of DB in the general pediatric population is unknown.[2, 4]
Although the symptoms are often similar, the treatment of DB and EIB is not. Whereas children with EIB usually benefit from the use of reliever and controller medication [5], the best treatment for most types of DB is respiratory physical therapy or breathing retraining [1, 2, 6, 7, 8]. This difference in intended treatment makes a distinction in the diagnosis of these two conditions in the clinic indispensable. Until now, a proper method for differentiating between DB and EIB has not been established [9], which leads to DB patients receiving inappropriately prescribed medication. As symptoms do not improve for these patients, clinicians may even decide to increase the doses because the actual cause of the symptoms is not recognized. Poor recognition deprives DB patients of proper treatment and unnecessary doses of medication increase the medication burden for these patients. From a health economic point of view, it leads to unnecessary medication costs.[7]
The diagnosis of DB involves considering and excluding different causes of chronic dyspnea, as is shown in figure 1.1. In Medisch Spectrum Twente (MST), experts indicate that their diagnosis of pediatric DB also relies on ruling out other conditions like EIB. This is based on anamnesis and an exercise challenge test (ECT), including spirometry. During this ECT, both observation of the patient and the measured spirometry parameters can guide the clinician towards an objective diagnosis. Spirometry demands repetitive forced breathing manoeuvres, which may lead to exhaustion and loss of technique in pediatric patients [10]. As DB is only identified by exclusion of other causes and this procedure requires demanding repetitive forced breathing manoeuvres, there is a desire for an objective measure which can detect both DB and UA in children.

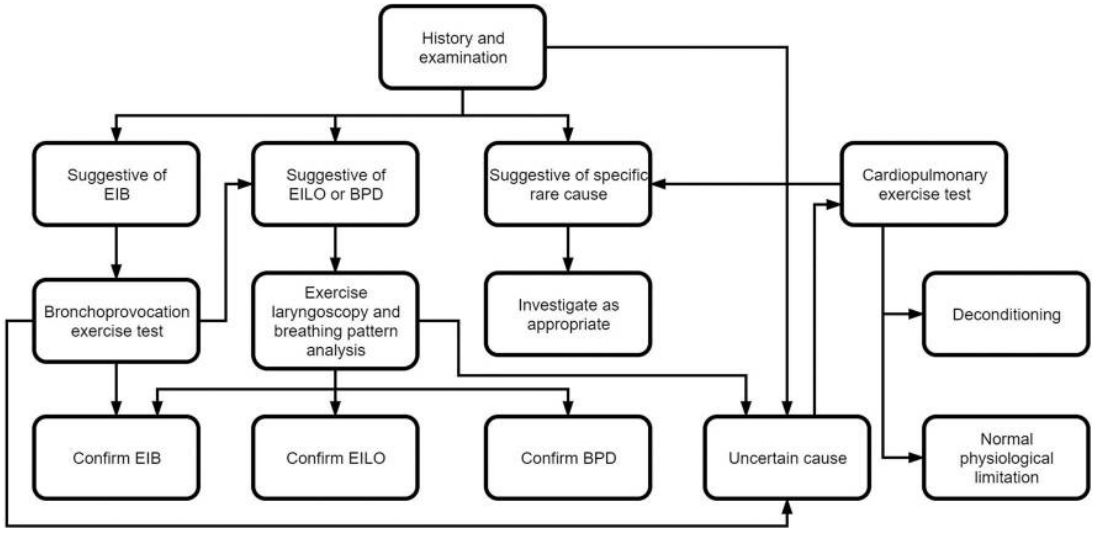


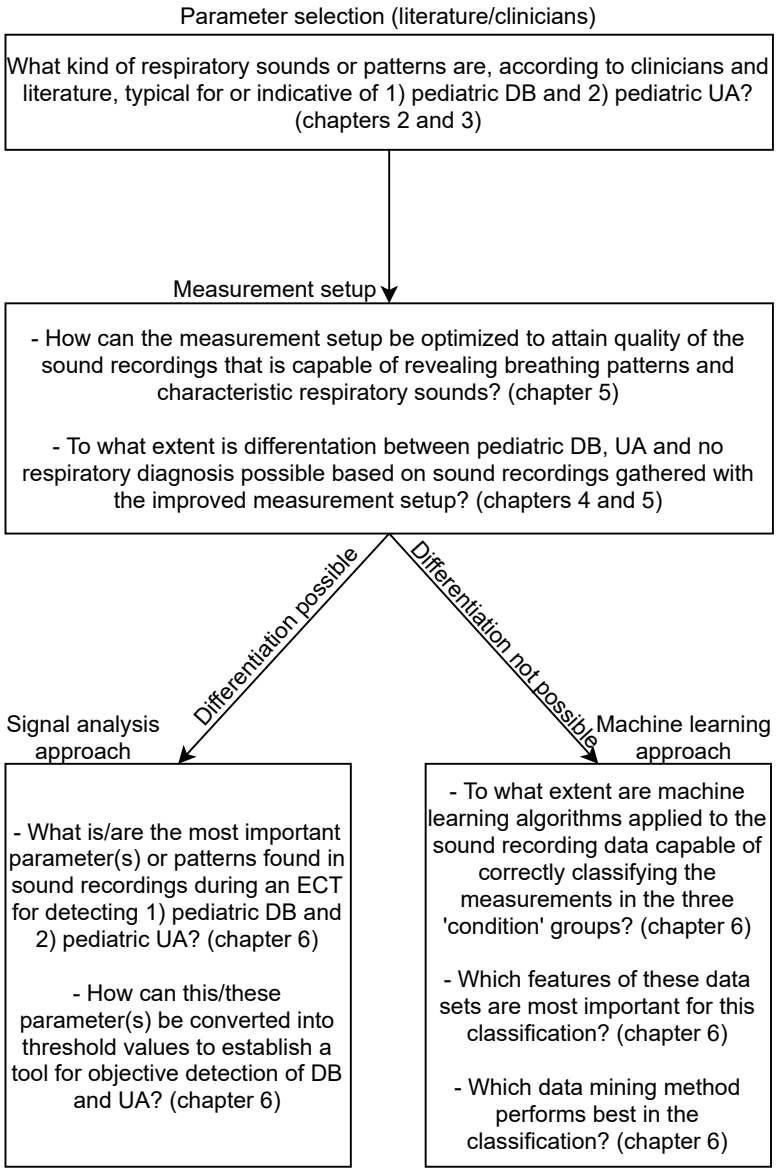
Figure 1.1: Diagnostic testing algorithm for dysfunctional breathing according to Barker et al.[2]

1.2 Potential objective diagnostic tools

A recent study from MST by Keijzer et al. has shown that in children, electromyography (EMG) measurements from the diaphragm can identify EIB and may provide insight into the objective diagnosis of DB [10]. In addition, clinicians in MST mention auditory observation sometimes supports their suspicion of UA and DB. This potential of respiratory sound analysis for the diagnosis of respiratory diseases is supported in literature [11, 12, 13, 14]. Therefore the sound recordings which are created during an ECT in the MST may provide insight into the objective diagnosis of the conditions.

1.3 Research objective and approach

This study focuses on using sound recordings as an objective diagnostic tool. The aim of the study is *to investigate the extent to which sound recordings during an ECT can be used as an objective measure to detect dysfunctional breathing and uncontrolled asthma in children when compared to the current ECT protocol involving spirometry.*
In order to investigate this, a multi-step approach will be followed in which different research questions will be answered. This approach is shown on the next page. As it is not yet known whether the sound recordings during an ECT can reveal differentiating breathing patterns and/or characteristic respiratory sounds, the approach can follow two paths at that point. If differentiation is possible with signal analysis techniques, the signal analysis approach will be followed and the three questions described in this path will be answered. If differentiation is not possible with signal analysis techniques, a machine learning approach will be followed where classification algorithms will be applied to the sound recordings to discover whether the algorithms are capable of correctly classifying the recordings in the diagnosis categories.


Parameter selection (literature/clinicians)
What kind of respiratory sounds or patterns are, according to clinicians and literature, typical for or indicative of 1) pediatric DB and 2) pediatric UA? (chapters 2 and 3)
Measurement setup
- How can the measurement setup be optimized to attain quality of the sound recordings that is capable of revealing breathing patterns and characteristic respiratory sounds? (chapter 5)
- To what extent is differentation between pediatric DB, UA and no respiratory diagnosis possible based on sound recordings gathered with the improved measurement setup? (chapters 4 and 5)
Differentiation possible
Differentiation not possible
Signal analysis approach
- What is/are the most important parameter(s) or patterns found in sound recordings during an ECT for detecting 1) pediatric DB and 2) pediatric UA? (chapter 6)
- How can this/these parameter(s) be converted into threshold values to establish a tool for objective detection of DB and UA? (chapter 6)
Machine learning approach
- To what extent are machine learning algorithms applied to the sound recording data capable of correctly classifying the measurements in the three 'condition' groups? (chapter 6)
- Which features of these data sets are most important for this classification? (chapter 6)
- Which data mining method performs best in the classification? (chapter 6)

2 BACKGROUND

In this chapter, the literature background for this research is described. First, the known symptoms of DB and asthma are listed and the current diagnosis of both conditions is explained. Next, the technique of EMG is explained and the state of the art regarding EMG as diagnostic tool for DB and asthma is described. Lastly, the emerging technique of respiratory sound analysis (RSA) is discussed. The experience of clinicians with respiratory sounds in the clinic is discussed in chapter 3.

2.1 Dysfunctional breathing

DB is a condition which is characterized by an alteration in the individual's normal breathing pattern. In literature, there is no consensus regarding the definition of DB. In 2015, however, Barker et al. proposed the following definition: *an alteration in the normal biomechanical patterns of breathing that result in intermittent or chronic symptoms which may be respiratory and/or non-respiratory* [1]. Different authors [6, 7] have since referred to this seemingly clear definition and thus when mentioning DB here, this is the definition which will be referred to.

A normal respiration is characterized by relaxed abdominal breathing, facilitated by the diaphragm. In children with DB, there is additional contraction of upper chest wall and accessory muscles. This results in a more thoracic respiration, which is often associated with an irregular respiration rate and generally involves mild hyperinflation.[1] The abnormal breathing pattern can exist in the absence of organic diseases (primary DB) or secondary to cardiopulmonary or neurological diseases (secondary DB) [15, 6]. When DB co-exists with other respiratory diseases, it is often unclear whether DB is caused by the other disease or whether it coincides [1]. It is plausible that for example with asthma, the incomplete expiration leads to an increase in expiratory reserve volume and a reduction in inspiratory reserve volume and this hyperinflation can result into an altered breathing pattern [8]. On the other hand, Depiazzi et al. mention that in their clinical experience, DB appears to be common in individuals with high expectations of themselves which provides an internal source of stress [7]. This could thus also be a trigger for the development of DB, irrespective of the existence of another respiratory disease.

Initially, such a change in breathing pattern is a natural response to stress. However, when the alteration of the relaxed, abdominal breathing pattern becomes a habit, or when it happens intermittently, symptoms start to develop and we speak of DB. DB can thus be regarded as an unconsciously learnt alteration in the regular breathing pattern, which manifests either at rest or intermittently in response to certain forms of stress. The body is not able to retrieve its original, relaxed breathing pattern.[1]

2.1.1 Classification of dysfunctional breathing

DB can be classified into different forms. In this research, the aim is not to distinguish between these different forms of DB. However, as we do want to be able to distinguish between UA and

DB, it is important to know which patterns or symptoms can be expected and to recognize that a collection of different presentations may be observed in clinical practice.
As with the definition of DB, there is no agreement regarding the classification of different forms of DB. In 2015, Barker et al. mentioned the division in thoracic and extra thoracic DB where both types can be subdivided into functional and structural. Thoracic DB describes the before mentioned pattern disordered breathing and extra thoracic DB is the upper airway involvement, in addition to pattern disordered breathing. The distinction between functional and structural is made by the absence (functional) or presence (structural) of anatomical and neurological abnormalities.[1]
Boulding et al. have proposed a classification of DB patterns in 2016 [9]. Based on literature review, they make a distinction between the following patterns with associated characteristics:

- Hyperventilation syndrome (HVS): rapid respiratory rate, tidal breathing closer to inspiratory capacity than in healthy individual.
- Periodic deep sighing: irregular breathing patterns and frequent sighing where a sigh is characterized by a tidal volume three times the normal volume. The patient can have difficulty with coordinating maximal expiratory and inspiratory manoeuvres.
- Thoracic dominant breathing: upper thorax predominantly used for breathing, lateral costal expansion lacks, large volume breaths with minimal inspiratory reserve capacity.
- Forced abdominal expiration: excessive contraction of abdominal muscles during expiration, tidal breathing at low lung volumes with minimal expiratory reserve volume.
- Thoraco-abdominal asynchrony: ineffective breathing caused by delay between rib cage and abdominal movement.

In 2020, Barker et al. revised their earlier distinction between thoracic and extra thoracic DB and expanded the information on thoracic DB with the classification proposed by Boulding et al. [2]. In other words, they explain that the breathing pattern component can be classified by the classification described by Boulding et al. and that the extra thoracic form is referred to as inducible laryngeal obstruction (ILO), where both components can occur independently of the other but may co-exist. ILO is described as an inappropriate, transient, reversible narrowing of the larynx in response to external triggers [16] where these external triggers can be either an event, situation or specific irritant [2]. This narrowing hampers the air passage into and out of the trachea, resulting in an inspiratory stridor and dyspnoea [2]. In most cases there is an asynchrony between the inspiratory phase and vocal cord movement resulting in adduction of the vocal cords. Therefore this phenomenon was formerly known as paradoxical vocal fold motion. However, nowadays the more broad term ILO is preferred, which includes pathologies affecting supraglottic structures in addition to the vocal cords. When the phenomenon occurs during exercise, we speak of exercise induced laryngeal obstruction (EILO).[16, 17]
Although this may seem a clear classification, Barker et al. emphasize that studies still need to be carried out to establish a similar classification of the breathing patterns in the pediatric population [2]. They mention that in children, there is overlap between and variation in the types of pattern, also greatly depending on the circumstances of the individual. The pattern of forced abdominal expiration is seldom seen in children with DB.

2.1.2 Symptoms

In clinic, DB patients present with a wide variety of symptoms. These are mostly respiratory symptoms, but nonrespiratory symptoms like dizziness and heart palpatations are also encountered.[6] The primary symptom is dyspnea or 'air hunger', a sense of being unable to

get a complete breath in [7], but for the different types of breathing patterns, different symptoms are common. Amongst those symptoms are shortness of breath, stridor, throat tightness, sighing, mouth breathing, chest pain, throat clearing, tingling and general fatigue.[1, 2, 7] The inspiratory stridor - which is an abnormal high-pitched respiratory sound caused by turbulent air flow due to a narrowed airway - and throat tightness are typical of the impeded air passage in (E)ILO [7]. Chest pain is likely to be of musculoskeletal origin and is therefore mostly reported during exercise. The reasoning is that at rest, respiratory muscles are used inappropriately or excessively and this overuse becomes evident when there are further demands from the muscles during exercise.[1] The chest pain can, however, have other causes like costochondritis or Tietze syndrome [18] and in this way be a trigger for the development of a disturbed breathing pattern. Barker et al. specifically focus on the pediatric population in their article and they describe the various symptoms, of which some occur when DB co-exists with other conditions [2]. They report shortness of breath to be the most common complaint, which can occur at rest but is mostly associated with exercise and is worsened by stress. Noisy breathing, stridor and air hunger are found to be most common with ILO, coughing and throat clearing are mostly associated with upper-airway inflammation. This upper-airway inflammation in turn can be associated with sensitization of the vocal cords, leading to ILO. Chest pain is also a common complaint and can either have a biochemical/biomechanical cause, or can be presented with EIB or gastro-oesophageal reflux. During an exacerbation of DB, patients may describe the feeling of having an obstruction in the throat or chest.

Comparison of symptoms with uncontrolled asthma

In order to make the distinction between DB and asthma in the clinic, it is important to not only know about the possible symptoms of DB but also of asthma and EIB, which is indicative of UA. With asthma, the lumen of the airway is decreased by smooth muscle contraction, congestion of the airway wall and increased mucus production. Consequently, the airway resistance is increased and the airflow is impeded, resulting in symptoms like coughing, wheezing and dyspnea.[10] In contrast to the inspiratory stridor in ILO, this respiratory wheezing sound is expiratory. Children may describe difficulty getting a breath out, opposed to the difficulty taking a full breath in with DB. Interval symptoms or diurnal variation of symptoms, nocturnal cough and reaction to specific triggers are also common features of asthma.[2]
Røksund et al. list three features that should be regarded when distinguishing EIB from exercise induced inspiratory symptoms (EIIS), the latter of which is caused by airflow obstruction in the upper airways, usually due to EILO [17]. Figure 2.1 summarizes these three features, being the position in the respiratory cycle, the moment in the exercise session and the time resolution. Whereas EIIS symptoms present in the inspiratory part of the respiratory cycle, EIB symptoms present in the expiratory part of the cycle. EIIS peaks during exercise and often has a typical pattern: it starts with increasing breathing difficulties with increased duration of inspiration, where coarse or high-pitched stridor sounds can be heard. This may be followed by clear-cut stridor and sometimes even hyperventilation attacks. The symptoms evolve in parallel with the increasing ventilatory demands due to increase of exercise intensity. In general, the symptoms resolve rapidly after their peak at maximum ventilation: it takes about 1-5 minutes after exercise for the symptoms to be resolved. In contrast, as EIB is a response to increased ventilation, its symptoms generally peak 3-15 minutes after exercise. However, in pediatric patients, bronchoconstriction often occurs earlier and can already start during exercise. The latter situation is known as 'breakthrough'-EIB.[19]

2.1.3 Treatment

The most commonly reported therapy for both thoracic and extra thoracic DB is breathing retraining, with the aim of restoring and maintaining a normal diaphragmatic breathing pattern

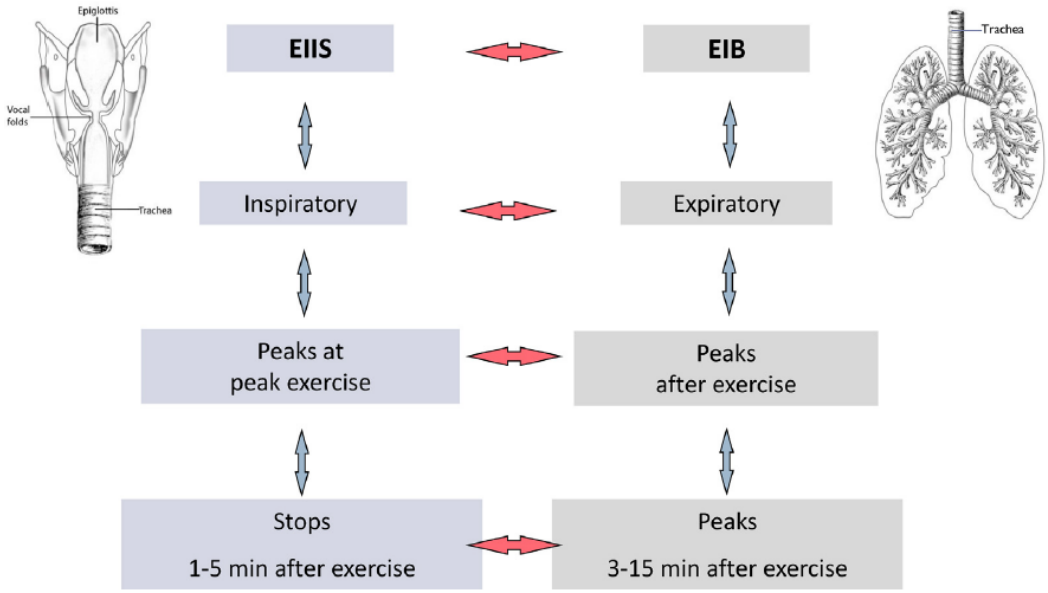


Figure 2.1: Features of EIIS versus EIB from Røksund.

[1, 15]. Cliftonsmith et al. describe that most protocols for the management of DB consist of twelve steps, of which the first step, being education on the pathophysiology of the disorder, is critical [8]. In this step, patients are reassured that their symptoms have a physiological cause and are manageable. It is also important that in this step, possible triggers for the poor breathing pattern are identified. Some report that accurate diagnosis and education are sufficient for the management of DB [9, 20], but most authors agree that although it is important, further treatment is required [1, 6, 7]. For this further treatment, two common breathing techniques are used: the Papworth method and the Buteyko method. The Papworth method focuses on diaphragmatic breathing where controlled and slow nasal breathing is emphasized. The Buteyko method also includes nasal breathing, but additionally focuses on introducing pauses in breath cycles to reduce hyperventilation.[9] In some cases, psychological treatment in addition to breathing retraining is required for managing the condition, but this differs per patient [1]. For patients with extra thoracic DB, additional treatment may be required. This treatment focuses on providing control over paradoxical movement in the vocal folds, which may require collaboration with a speech pathologist [1, 7]. Multiple studies find significant symptom improvement and improved quality of life due to breathing retraining [21, 22, 23], even after a follow-up period of five years [24]. These findings mainly describe research in the adult population, but Barker et al. describe a similar approach for the management of DB in the pediatric population [2]. They emphasize the need for individualized treatment in which both the child and family are closely involved in goal setting and planning. Again, the education on the disorder is very important to reduce fear and give the child the feeling of control. The treatment is mostly provided by a physical therapist together with a speech pathologist or psychologist, depending on the form of DB. Some children may be more in need of psychological assistance than others, for example when anxiety or performance related stress is a great contributor to the complaints. However, psychological assistance may be of help for all pediatric DB patients to help them understand how their physical and mental health are interconnected. If comorbidities are present, education may involve helping the child differentiate between symptoms so that it knows when to apply which therapeutic technique or use medication. Barker et al. report that long term outcomes with breathing retraining in the pediatric population are good and that recurrence is uncommon. They also describe the possibility of surgical interventions for supraglottic EILO, but mention that due to their morbidity, such surgical interventions are not preferred.

Alternatively or complementary to the above described approaches, EMG with biofeedback has been evaluated for the use of symptom reduction in DB. With this technique, breathing patterns can be physiologically recorded with EMG and feedback on the breathing pattern can be provided with visual signs, acoustic or haptic signals, or with virtual reality technology.[25, 26] This technique, however, has not yet been applied often.[1]
In contrast to the management of DB, asthma should in most cases be managed with appropriate use of reliever medication [1, 5]. As these approaches are substantially different, where breathing retraining also provides a drug free treatment which of course is desired when possible, it is essential to be able to differentiate between the conditions.

2.2 Diagnosis of dysfunctional breathing and uncontrolled asthma

Currently, there is no gold-standard for diagnosing DB and it is often under- or misdiagnosed due to the similarities of the symptoms with other diseases like asthma [6]. The Nijmegen Questionnaire (NQ) is often used for identifying patients with DB. The NQ assesses sixteen symptoms of various origins, such as cardiovascular, neurological, respiratory, gastro-intestinal and psychological complaints. Their frequency of incidence is indicated on a five-point ordinal scale by the patient. The instrument, however, is developed as a questionnaire for screening HVS and has only been validated for use in adults with HVS secondary to asthma.[27] One developer of the NQ has even stated that using it as gold-standard test to diagnose DB is incorrect and that it should only be used to assess the normality of subjective sensations [28]. Although the NQ may have some value in identifying some thoracic DB symptoms, it is not designed for that and it should also be recognized that it is not validated for use in children or patients with co-morbidities [1].

For EILO, there is a gold-standard diagnostic tool: laryngoscopy during symptomatic periods, often provoked by physical exercise. The downside of this method is that by the time the laryngoscope is inserted after the period of exercise, symptoms have often resolved and the characteristic appearances in the larynx are not seen anymore.[1] A solution for this is continuous exercise laryngoscopy, where the laryngoscope is worn during the whole period of exercise. Availability and also suitability of equipment for the pediatric population however is limited. Furthermore, the experience is that sometimes children cannot reach the exercise intensity at which they present symptoms while wearing the laryngoscope.[2]

In theory, the altered breathing pattern in DB can be recognized from simple observation by a clinician, but in practice, many clinicians do not possess this competence.[1] Manual Assessment of Respiratory Motion (MARM) is a procedure that could aid in this abnormal breathing pattern identification. With this procedure, the clinician uses palpation to estimate the motion at the posterior and lateral lower rib cage. The relative distribution of motion between upper rib cage, lower rib cage and abdomen can be used to identify the breathing pattern. In addition, the breathing rate and regularity can be determined.[29] This procedure, however, has not been widely used [1], possibly due to the subjectivity of the procedure. Moreover, MARM is not a suitable tool in children in which the altered breathing pattern or symptoms mainly present(s) during exercise, since palpation during exercise is not possible. Boulding et al. do report a good inter-examiner reliability when the procedure is carried out with the patient wearing bands across the chest to perform measurements, which could also facilitate assessment during exercise [9].
Other techniques for objectively assessing breathing patterns, like respiratory inductance plethysmography, EMG, ultrasound and optical scanning methods have been under investigation, but until this day their cost and complexity have prevented the use in clinical practice. Besides, these techniques have only been used to study normal respiratory patterns, not DB.[1]

2.2.1 Exercise challenge test

At this moment, an important part of identifying DB is excluding UA as main cause of the complaints. In MST this is done by performing an ECT to test on EIB, which is indicative of UA. During such a test, the patient is required to exercise for six minutes on a treadmill (or jumping castle for young children [30]) at a heart rate of approximately 80% the maximum heart rate in order to provoke symptoms. Spirometry is performed before exercise and after exercise to investigate possible bronchoconstriction.[31, 32] Spirometry is a lung function test that measures the volume and flow of air that can be inhaled and exhaled. The patient is required to take a full breath in and exhale with a maximum force until the lungs are emptied.[32] The forced expiratory volume in 1 second (FEV_1) is a relevant parameter for identifying EIB. A decrease in FEV_1 of more than 13% in response to exercise is regarded as a significant reaction in pediatric EIB [10] and especially in combination with reversibility in response to a bronchodilator, it is a good indicator of EIB [33]. Spirometry is usually normal with DB, although the expiratory phase might be terminated early. In ILO, sometimes a flattening of the inspiratory loop is seen.[2] In contrast to the reversibility of symptoms seen with EIB, symptoms of DB patients will not be resolved when using bronchodilators after exercise [7]. Unfortunately, the spirometry used in ECT requires repetitive forced breathing manoeuvres which may cause exhaustion and loss of technique in children and can affect respiratory physiology.[10] Moreover, studies have shown that findings might be indicative but are not diagnostic [2, 34]

2.2.2 EMG as diagnostic tool

EMG is a technique used to measure electrical activity of skeletal muscles [35]. An EMG is a representation of the activity of multiple motor units. A motor unit consists of one motor neuron and all the skeletal muscle fibers it innervates. Per human muscle, a large amount of these motor units work together to account for the contractions of the muscle. The number of motor units per muscle may range from 100 motor units for fairly small hand muscles to over 1000 motor units for large limb muscles. These motor units fire in order to create an action potential which is carried towards the muscle, to facilitate contraction of that muscle. The amount and type of motor units recruited for muscle activity determines the force of contraction. When a motor unit fires, the action potential is elicited in all of its muscle fibers. The sum of the electrical activity in all these fibres is the motor unit action potential (MUAP) and what we see in an EMG of a muscle are the superimposed MUAPs for all motor units contributing to the muscle activity.[36]

Surface EMG is a non-invasive way of measuring muscle activity, by measuring the potential differences of a muscle or muscles compared to a ground electrode. This ground electrode must be placed on a location not involved in the movement of the muscle of interest. The other electrodes are placed on the skin above the muscle and together with the ground electrode, this makes measurement of the potential differences possible.[37]

EMG is used in various medical fields and over the past years, multiple studies have proven the potential of EMG to measure electrical activity of respiratory muscles [38, 39, 40]. As abnormal breathing patterns require greater recruitment of accessory breathing muscles [6] and the maintenance of sufficient airflow with asthma also requires more muscle contraction force of both the diaphragm and accessory breathing muscles [10], it can be expected that the electrical activity of these muscles in children with DB or asthma is different compared to healthy children. Maarsingh et al. have shown that EMG signals from the diaphragm and intercostal muscles are reproducible in studying tidal breathing [41]. Later research has indicated that the change in electrical activity of the diaphragm and intercostal muscles related well to the FEV_1 and was reversible after salbutamol use, proving its potential as alternative for pulmonary function testing [42].

Keijzer et al. have studied the electrical activity of the diaphragm and accessory breathing

muscles in children who were referred for an ECT [10]. From this study, it appeared that especially increases in EMG peak amplitudes of the diaphragm were strongly related to the drop in FEV_1 in children with EIB. In the full report, Keijzer et al. also mention the additional potential of EMG to distinguish both controlled and uncontrolled asthma from DB [33]. The peak height of the EMG signal at the diaphragm in DB shows a large increase followed by a rapid decline, which is not seen with (un)controlled asthma. However, it is noted that the small population size and inability to separate individual patients from one another stresses the need for further research in order to make the distinction between DB, UA and controlled asthma.

2.2.3 Respiratory sound analysis

Another emerging method for assessing respiratory complaints is RSA. Sound occurs when a potential source of sound is set into vibratory motion and it propagates as an acoustic wave through a transmission medium. Sound is measured in terms of frequency and amplitude. The frequency, or pitch, is the number of repetitions of the acoustic wave per second. A higher frequency, expressed in its standard unit Hertz (Hz), thus produces more oscillations than a lower frequency. The amplitude is the relative strength of the acoustic waves, so the strength of the vibrations. We perceive this amplitude as loudness and it is measured in decibels (dB), referring to the sound pressure level.[43]
Respiratory sounds originate in the central airways, as a result of vibrations due to turbulence and air velocity. The sounds are both non-stationary and non-linear because of the variations in airflow rate and volumes throughout the respiratory cycle.[14] Since respiratory sounds are sensitive to airway changes, RSA is a simple method to assess airway changes [44]. We make a distinction between normal and adventitious respiratory sounds, of which the adventitious sounds can provide valuable information regarding respiratory diseases. They are additive unusual sounds which can, in theory, be detected with a stethoscope by an expert, but the correct detection is difficult. Recent studies have focused on computerised respiratory sound analysis (CORSA), of which for example Prodhan et al. have shown a computerized respiratory sound monitor to be better at detecting wheeze than medical staff [45], but at this moment there is no well-established standardised approach for CORSA [12]. Most of the studies have focused on CORSA from chest wall sounds, although some have analyzed sounds detected from the mouth [11].

Respiratory sounds detected from the mouth have a frequency range of 200 to 2,000 Hz [46], as can be derived from the table in figure 2.2. The adventitious sounds as described above can be divided in discontinuous and continuous adventitious sounds (DAS and CAS). DAS are adventitious sounds with a short duration of less than 25 ms and are not associated with asthma or DB. CAS have durations of more than 250 ms and can be further divided into high-pitched (stridor, wheeze, gasp) and low-pitched (rhonchi and squawk) sounds.[11] The characteristics of these adventitious sounds are summarized in the table in figure 2.3. Especially the high-pitched wheeze and stridor are relevant for the distinction between asthma and DB and will therefore be further investigated. Wheeze is caused by airway narrowing leading to airflow limitation and is heard mostly during expiration, although it can also be heard during inspiration or even be biphasic. Wheeze presents as sinusoid-like signals and although it is a CAS, Pramono et al. describe that smaller durations (around 80 to 100 ms) are also common. The frequency range is between 100 and 1,000 Hz with a dominant frequency of at least 400 Hz. Stridor is caused by turbulent airflow in the larynx or bronchial tree and is heard mostly during inspiration, although in rare cases it is heard during expiration or it is biphasic. Stridor presents with a dominant frequency of more than 500 Hz and has a duration of more than 250 ms.[11] Stridor typically has a harsher and louder sound than wheeze, which sounds more musical [47]. It must be mentioned that the described frequencies are established with pulmonary auscultations and the characteristic frequencies from signals detected from the mouth may differ. In a study

where respiratory sounds recorded from the mouth are used to develop a wheezing recognition algorithm, it appears that the mean frequency of the wheeze detected by the algorithm was 250 Hz when recorded over the neck compared to 900 Hz when recorded at the mouth, proving that different recording locations result into different frequencies [48].

Table 1. Normal breath sounds.

Breath Sounds	Location	Range[a]	Pitch[b]	Quality[c]	Timing (I:E ratio)[d]	Pause[e]
Vesicular	Most of lung fields	100—1,000 Hz Energy drop at 200 Hz	Low	Low-pass filtered noise like Soft Rustling sound	During inspiration and early expiration (2:1 ratio)	Pause between different breath cycle
Broncho-Vesicular	Between scapulae on posterior chest and center part of anterior chest	Intermediate between Vesicular and Bronchial	Intermediate	Intermediate intensity	During both inspiration and expiration (1:1 ratio)	N/M
Bronchial	Large airways on chest near second and third intercostal space	Similar to Tracheal	High	Loud Hollow	During both inspiration and expiration (1:2 ratio)	Short pause between inspiration and expiration phase
Tracheal	Suprasternal notch on trachea	100—5,000 Hz Energy drop at 800 Hz	High	Harsh Very loud	During both inspiration and expiration (1:1 ratio)	Distinct pause between inspiration and expiration phase
Mouth	Mouth	200—2,000 Hz	N/M	White-noise like Silent when normal	N/M	N/M

[a]Information from [8–10, 12, 13]
[b]Information from [8, 9]
[c]Information from [8, 13]
[d]Information from [8, 11]
[e]Information from [8]
Abbreviation N/M: Not Mentioned in [8–13]

Figure 2.2: Normal Breath Sounds [11].

Table 2. Types of adventitious sounds and its characteristics.

Types	Continuity	Duration[a]	Timing[b]	Pitch[c]	Quality[d]	Cause[e]	Disease Associated[f]
Wheeze	Continuous	> 80*ms*	Inspiratory, Mostly Expiratory, Biphasic	High (> 400*Hz*)	Sibilant, Musical	Airway narrowing, airflow limitation	Asthma, COPD, Foreign body
Rhonchi	Continuous	> 80*ms*	Inspiratory, Mostly Expiratory, Biphasic	Low (< 200*Hz*)	Sibilant, Musical	Secretion in bronchial, muchosal thickening	Bronchitis, COPD
Stridor	Continuous	> 250*ms*	Mostly Inspiratory, Expiratory, Both	High (> 500*Hz*)	Sibilant, Musical	Turbulent airflow in larynx or lower bronchial tree (Upper airway obstruction)	Epiglottitis, foreign body, croup, laryngeal oedema
Fine Crackle	Discontinuous	± 5 ms	Inspiratory (late)	High (650 Hz)	Non-musical, Explosive	Explosive opening of small airways	Pneumonia, Congestive heart failure, Lung fibrosis
Coarse Crackle	Discontinuous	± 15 ms	Mostly Inspiratory (early), Expiratory, Both	Low (350 Hz)	Non-musical, Explosive	Air bubble in large bronchi or bronchiectatic segments	Chronic bronchitis, bronchiectasis, COPD
Pleural Rub	Discontinuous	> 15*ms*	Biphasic	Low (< 350*Hz*)	Non-Musical, Rhythmic	Pleural membrane rubbing against each other	Inflammation of lung membrane, lung tumour
Squawk	Continuous	± 200 ms	Inspiratory	Low (200—300 Hz)	Short Musical and non-musical	Oscillation of peripheral airways	Hypersensitivity pneumonia, pneumonia
Gasp	Continuous	> 250*ms*	Inspiratory	High	Whoop	Gasping for breath	Whooping cough

[a]Information from [10, 17, 19, 20, 23]
[b]Information from [8, 10, 20, 21]
[c]Information from [10, 14, 20, 23]
[d]Information from [8, 10, 18–20]
[e]Information from [8, 15, 16, 19–21, 24]
[f]Information from [8, 10, 20, 22]

Figure 2.3: Adventitious Breath Sounds [11].

As can be seen in these tables, there is overlap in the characteristics of different adventitious sounds, making it challenging to develop a proper classification method for adventitious sounds based on frequency domain analysis. Moreover, there does not seem to be agreement in literature regarding the characteristics of the signal. Whereas we just learned from Pramono et al. that wheeze sounds may present with durations of only 80-100 ms, Enseki et al. describe that in a sound spectrogram, typical wheezing can be observed with a length exceeding 250 ms [44]. Nevertheless, some techniques and algorithms have been applied to study adventitious respiratory sounds.

Techniques and algorithms for RSA

Respiratory sound signals can be analyzed in different domains: the time domain, frequency domain and time-frequency domain. When studying the signals in time domain, the breath rate and ratio between duration of inspiration and expiration can be extracted. In addition, an impression of the loudness of the breath sound can be obtained by studying the signal amplitude. Since sound recordings are often noisy and contain a wide spectrum of frequencies, it is necessary to extract an envelope of the signal. This envelope can be seen as the visualization of the signal when interconnecting the fundamental peaks of the signals, so it is the magnitude of the analytic signal. The envelope of a signal is described by the following equation:

$$e(t) = \sqrt{x(t)^2 + \hat{x}(t)^2}$$

with $\hat{x}(t)$ the Hilbert transform of $x(t)$. The Hilbert transform is the convolution of a signal with $1/\pi t$ and is used to remove linear phase delay in the signal [49]. Extraction of the envelope can be performed with the help of MATLAB's envelope function which uses the Hilbert transform to find the magnitude of the analytic signal.

As explained before, sounds are composed of waveforms with different frequencies. Therefore studying frequencies in a sound recording is very relevant when assessing different types of sounds. To study a time signal in the frequency domain, the Fourier transform is used. The Fourier transform decomposes a signal into its frequency components. The fast Fourier transform (FFT) is an efficient algorithm to perform this mapping from time to frequency domain for discrete time signals. The periodogram, which is an estimate of the spectral density of a signal, is a version of the FFT which shows how the power of a signal is distributed over the frequencies. For noisy signals, the Welch method can be applied. This method averages over windowed periodograms to obtain a smoothed version of the periodogram. The downside of this method is a deteriorated frequency resolution.[50] In RSA researches, some frequency domain processing techniques have been used, like using quantile vector, cepstral analysis or computing (averaged) periodograms, to study frequency content. From these researches, it appears that features derived from frequency domain analysis are able to distinguish between normal and adventitious respiratory sounds. Nevertheless, amongst the different domains, RSA in frequency domain is most rarely used.[12] Due to the non-stationarity of respiratory sounds, omitting changes over time in the signal will discard relevant information in the signal. Therefore, studying the signals in time-frequency domain is more appropriate. One of the most widely used methods is computing the Short-Time Fourier Transform (STFT), which can be visualized in a spectrogram [12]. With STFT, the signal is divided into segments and for each segment, the Fourier transform is computed. In the resulting spectrogram, it can be visualized which frequencies are dominant in which time frame. The spectrogram displays amplitude of the frequency components of a signal over time, with which changes in frequency values of the components of a signal over time can also be determined. Moreover, from the STFT result, features like peak frequency, mean and median frequency and amplitude can be derived. The frequency resolution, which indicates how precise we can distinguish different frequencies, is determined by $2/windowlength$ when using a symmetric window. The time resolution is then determined by $2/frequencyresolution$, so there is a trade-off between frequency resolution and time resolution: a good (small) frequency resolution is at the expense of poor time resolution, so uncertain time localization.[50] The advantage of the STFT is that computation is easy and studying the frequency of the signal in each time frame is simple. However, the frequency resolution of the resulting spectrogram is relatively low (especially for short recordings) and the exact moment of occurrence of a specific frequency is hard to determine since the frequencies are computed over intervals.[12] The afore mentioned trade-off makes it hard to obtain information on frequency content which is precise in both time and frequency. A technique used to overcome this resolution problem is wavelet transformation. The wavelet transform is similar to the STFT, but in contrast to STFT, it

uses variable-sized windows to study the signal in time-frequency domain. For low-frequency phenomena, the time window is bigger which improves the frequency resolution and lowers the time resolution; for high-frequency phenomena, the time window is shortened which deteriorates the frequency resolution but improves the time localization.[51]
Other used signal processing techniques in the time-frequency domain are Wigner-Ville distribution, which has a high resolution but requires massive computation and the Hilbert-Huang transform, in which the signal can be decomposed based on its intrinsic characteristics (in contrast to wavelet transformation in which a mother wavelet has to be selected), but only information on certain frequencies which cannot be specified by the user is provided.[12, 52]

3 RESPIRATORY SOUNDS IN CLINIC

From the previous chapter, it appears that the presence and nature of adventitious sounds and the specific timing of symptoms could enable the distinction between different respiratory conditions. However, there are no guidelines for how respiratory sounds should be assessed and thus it is important to know from the clinic which factors are taken into consideration when assessing respiratory sounds. Together with the previous chapter, this chapter will aim to answer the question
What kind of respiratory sounds or patterns are, according to clinicians and literature, typical for or indicative of 1) pediatric DB and 2) pediatric UA?

3.1 Methods

In order to gather the experiences of clinicians with respiratory sounds, a sub-study is performed in which clinicians had to assess sound recordings. Four clinicians who regularly see children with respiratory complaints participated in this study. Two of them are pediatricians, two are technical physicians.
In the outpatient clinic, scheduled ECTs were performed and video recordings were made during these ECTs by clinicians. These ECTs were performed on a treadmill in a climate controlled room where the temperature is kept at 10°C. During an ECT, the patient exercised for six minutes on the treadmill which was set at a slope of 10°. The first two minutes were used to reach a heart rate of approximately 80% of the maximum heart rate. These two minutes were followed by four minutes of exercise at this submaximal heart rate. The velocity of the treadmill was adjusted to reach the submaximal heart rate. As the physical condition differs per patient, the treadmill velocity differs per recording.
After every ECT, the clinician who made the recording selected a short fragment (mostly between 10 and 30 seconds) of the video recording which best represented the condition of the patient. These recordings were available on a computer in the climate room. For the purpose of this sub-study, the sound was extracted with MATLAB from fifteen of these video recordings (both the complete and short fragments). Five of these recordings were from children with DB, five from children with UA and five from children with no established respiratory condition (abbreviated as ND, no diagnosis). For every short fragment, it was verified if there was no clear speech from the child which could lead to recognition of the patient by (one of the) clinicians. If there was clear speech, this part of the fragment was removed or another fragment was created from the entire ECT recording. The fragments were numbered and a random sequence generator was used to put these numbers in a random order.
The collection of fifteen sound recordings was sent to the four clinicians, together with a form containing an instruction and an assessment table. The form can be found in appendix A. The clinicians were asked to listen to all recordings and during or after a recording, write down a diagnosis solely based on the sound recording and to explain which sounds brought them to that specific diagnosis. This was entered in the row in the table which corresponded to the number of the fragment. In addition, a certainty score was written down in this table per recording: a number in the range from 1 (very insecure) to 10 (very sure) which indicated how certain they were that their diagnosis was correct. The clinicians were allowed to play the recordings as

often as was necessary for them. After all clinicians had filled out the form, the sound recordings were shuffled and they were again asked to assess the sound recordings. In this way, the intra-individual variability could be assessed.
As a third test, the video recordings from the same patients were offered to the four clinicians to see how much difference it would make when the patient could also be observed. Moreover, if a classification algorithm is developed in the remainder of the study, this algorithm could be compared to both the sound assessments by clinicians and the video assessments by clinicians to assess its true added value to the diagnosis of UA and DB. An assessment form similar to the form for the sound recordings was used to assess the video recordings. In this case, not only characteristics in the sound but also in the camera footage could be used for making a diagnosis. When a clinician recognized the patient on the footage and thus had prior knowledge, this fragment was not assessed by the clinician.
The given diagnoses for all fragments were gathered and from this data, the amount of correct diagnoses was counted. This was done separately for the sound recordings and for the video recordings. For the sound recordings, the amount of correct diagnoses was counted in total as well as per assessment round per assessor. The intra-individual variability was computed by determining what percentage of the diagnoses in the second round differed from the diagnoses belonging to the same fragments in the first round. Moreover, the average certainty scores indicated per assessor were computed for all fifteen recordings, for all correct diagnoses and for all incorrect diagnoses. Per fragment, the amount of different 'diagnosis' answers was counted as a measure of the inter-individual difference in answers. For the video recordings, the amount of correct diagnoses was counted in total and per assessor. Similar to the sound recordings, the average certainty scores were computed and the amount of different 'diagnosis' answers per fragment was counted. In addition, it was determined what the increase or decrease in correctly classified recordings was when adding camera footage to the sound recordings.
In order to answer the main research question of this chapter, the explanations given by the clinicians for the different diagnoses were studied. From these different answers, it was determined which explanations occurred most often or were most typical and could thus be useful for the diagnosis of DB and UA.

3.2 Results

3.2.1 Audio recordings

The forms for assessing the audio fragments were filled out twice by the four clinicians. Table 3.1 shows how much diagnoses were correctly made by each clinician per assessment round and in total. In addition, the intra-individual variability is indicated. The intra-individual variability is here determined by counting the amount of diagnoses in round 2 that differed from the given diagnosis in round 1 and dividing it by the total amount of fragments.

Table 3.1: Results of the assessment of fifteen audio fragments from four clinicians. The amount of correct diagnoses in the first round, second round and in total are indicated, as well as the intra-individual variability.

	Correct round 1	**Correct round 2**	**Correct total**	**Intra-individual variability**
Assessor 1	6 (40%)	5 (33%)	11 (37%)	20% (3 fragments)
Assessor 2	4 (27%)	6 (40%)	10 (33%)	53% (8 fragments)
Assessor 3	3 (20%)	2 (13%)	5 (17%)	40% (6 fragments)
Assessor 4	7 (47%)	5 (33%)	12 (40%)	33% (5 fragments)
Total	**20 (33%)**	**18 (30%)**	**38 (32%)**	**37% (22 fragments)**

From the table, it can be seen that nearly a third of all fragments was classified in the correct diagnosis category.

In table 3.2, the averages of the certainty scores which were indicated by the assessors during the assessment are shown. The scores are shown for all fragments, for all correctly diagnosed fragments and for all incorrectly diagnosed fragments.

Table 3.2: Average certainty scores indicated by the assessors. The average of all fragments is indicated in the second column. The third and fourth column show the average certainty score for the correct and incorrect diagnoses respectively.

	Average certainty score	**Average certainty score correct**	**Average certainty score incorrect**
Assessor 1	6.2	6.9	5.8
Assessor 2	6.5	6.1	6.8
Assessor 3	6.5	6.6	6.5
Assessor 4	5.4	5.1	5.7
Total	**6.2**	**6.1***	**6.2***

*This is not an average of the four values above, but the average of all scores for (in)correct answers

Table 3.3 shows, per fragment, how many answers were correct in total, in the first round and in the second round. In addition, the amount of different mentioned diagnoses is indicated per fragment. In table B.1 in appendix B, all given diagnoses per fragment per assessor are shown.

Table 3.3: Number of correct answers and different answers per fragment, displayed in total (tot) and per assessment round. r1=round 1, r2=round 2.

	Correct tot	**Different diagnoses tot**	**Correct r1**	**Different diagnoses r1**	**Correct r2**	**Different diagnoses r2**
UA_1	0 (0%)	1	0 (0%)	1	0 (0%)	1
UA_2	3 (38%)	3	2 (50%)	3	1 (25%)	3
UA_3	4 (50%)	3	2 (50%)	3	2 (50%)	2
UA_4	1 (13%)	3	0 (0%)	2	1 (25%)	3
UA_5	2 (25%)	2	1 (25%)	2	1 (25%)	2
DB_1	2 (25%)	2	1 (25%)	2	1 (25%)	2
DB_2	0 (0%)	2	0 (0%)	2	0 (0%)	1
DB_3	5 (63%)	2	3 (75%)	2	2 (50%)	2
DB_4	4 (50%)	3	2 (50%)	3	2 (50%)	2
DB_5	1 (13%)	3	0 (0%)	2	1 (25%)	2
ND_1	4 (50%)	3	2 (50%)	3	2 (50%)	2
ND_2	4 (50%)	2	2 (50%)	2	2 (50%)	2
ND_3	2 (25%)	3	2 (50%)	3	0 (0%)	2
ND_4	4 (50%)	3	2 (50%)	3	2 (50%)	2
ND_5	2 (25%)	3	1 (25%)	3	1 (25%)	2

In the assessment forms, the reasoning behind the choice for a certain diagnosis was mentioned in most cases. From these forms, it appears that different (combinations of) characteristics in the sound fragments have been mentioned as reason to make a certain respiratory diagnosis. For UA, the following characteristics were mentioned: irregular breathing pattern; swallowing mucus; soft breathing sound combined with high pitched CAS; prolonged expirium; high pitched sound at the end of expirium; putting force at expiration; holding breath before going on to next breath; high breath rate; polyphonic expiration; biphasic expiration. For DB, the mentioned characteristics were: hyperinflation; high breath rate; biphasic inspiration; vibratory respiration

sounds; pinched breath; high thoracic respiration, expressed as more sound with inspiration than with expiration; putting force at inspiration; prolonged inspiration; loud breathing sound; quick and irregular respiration (hyperventilation); irregular breaks in between breaths; groaning (due to contraction of the larynx).
Note that in this context, 'biphasic' refers to consisting of two consecutive parts, not to something occurring in both the inspiration and expiration (as is meant when using the term biphasic stridor for example).

3.2.2 Video recordings

One of the clinicians who assessed the audio fragments, did not assess the video recordings on time. A clinician who had not assessed the audio fragments, did assess the video recordings so again results were available from four clinicians. The results of the video assessments can be found in table 3.4. Assessors 1-3 indicate the same three assessors as in the audio assessments. Assessor 4 is the assessor who did not assess the audio fragments. Assessor 1 skipped one fragment because of recognition of the patient. Assessor 2 skipped two fragments for the same reason. For assessors 1-3, it was determined how the percentage of correctly classified recordings differed between the audio and video assessments.

Table 3.4: Results of the assessment of fifteen video fragments from four clinicians. The amount of assessed recordings, correct diagnoses and percentage-wise change in performance with respect to assessing audio fragments are indicated.

	Given answers	**Correct**	**Change w.r.t. audio**
Assessor 1	14	11 (79%)	+42%
Assessor 2	13	8 (62%)	+29%
Assessor 3	15	5 (33%)	+16%
Assessor 4	15	5 (33%)	NA
Total	**57**	**29 (51%)**	**+29%**

From the table, it can be seen that 51% of all fragments were classified in the correct diagnosis category.

In table 3.5, the averages of the certainty scores which were indicated by the assessors are shown. The scores are shown for all fragments, for all correctly diagnosed fragments and for all incorrectly diagnosed fragments. Assessor 3 did not indicate certainty scores.

Table 3.5: Average certainty scores indicated by the assessors. The average of all fragments is indicated in the second column. The third and fourth column show the average certainty score for the correct and incorrect diagnoses respectively.

	Average certainty score	**Average certainty score correct**	**Average certainty score incorrect**
Assessor 1	6.2	6.5	5.0
Assessor 2	7.1	7.0	7.2
Assessor 3	NA	NA	NA
Assessor 4	3.8	4.2	3.6
Total	**5.7**	**5.9***	**5.3***

*This is not an average of the three values above, but the average of all scores for (in)correct answers

Table 3.6 shows, per fragment, how many answers were given in total (since assessors 1 and 2 did not assess all fragments), how many of these answers were correct and how many different diagnoses were mentioned. In table B.2 in appendix B, all given diagnoses per fragment per assessor are shown.

Table 3.6: Number of given answers, correct answers and different answers per fragment.

	Given answers	**Correct**	**Different diagnoses**
UA_1	4	1 (25%)	3
UA_2	4	4 (100%)	1
UA_3	3	2 (67%)	2
UA_4	4	1 (25%)	3
UA_5	3	3 (100%)	1
DB_1	4	2 (50%)	2
DB_2	4	1 (25%)	3
DB_3	4	4 (100%)	1
DB_4	4	2 (50%)	3
DB_5	4	2 (50%)	2
ND_1	4	2 (50%)	2
ND_2	4	0 (0%)	1
ND_3	4	2 (50%)	2
ND_4	4	2 (50%)	3
ND_5	3	1 (33%)	3

In the assessment forms of the video fragments, again the reasoning behind the choice for a certain diagnosis was mentioned in most cases. Similar remarks about the respiratory sounds were made. In addition, characteristics appearing from the camera footage have been mentioned as reason to make a certain respiratory diagnosis, often in combination with characteristics from the sound. For DB, two frequently mentioned characteristic were shrugged shoulders and staggering. For UA, a slightly bent upper body was mentioned. The stature of the patient was also taken into consideration sometimes when determining the diagnosis. The following mentioned (loosely translated) explanations show the combination of sound and image which is taken into account:
DB: 'Shoulders move along with respiration, significantly shrugged shoulders (also in combination with biphasic inspiration)'
DB: 'Irregular respiration in combination with unsteady steps'
UA: 'Prolonged expirium in combination with regular respiration and no shrugged shoulders'
UA: 'High breath rate at average treadmill velocity, slim patient, hampered expiration'
UA: 'Uncomfortable gait and respiration, young, slightly overweight boy'

3.3 Discussion

On average, 32% of the sound recordings was correctly classified. Table 3.1 shows that the amount of correctly classified recordings differs per assessor. In addition, it can be learned from the intra-individual variability that overall, the assessors are fairly inconsistent when assessing the recordings although this clearly differs per assessor. Table 3.3 shows that almost every recording has some correct answers. Since the total 'accuracy' of the clinicians is only 32%, this means that there is a lot of variation between individuals as to which recordings they classify correctly and which they do not classify correctly. This is confirmed by the columns showing the amount of different diagnoses: for almost all recordings, there is no unanimity.
All average certainty scores shown in table 3.2 are in a range from 5 to 7. What is striking

is that the average certainty score for correct answers is not higher than for incorrect answers. This implies that in some cases, assessors were insecure about their given diagnosis when it was correct, making it likely that it was based on a guess. On the other hand, high certainty scores for incorrect diagnoses proves that judging the respiratory condition of a patient based on a sound recording only is not a reliable method.
Some characteristics mentioned on the assessment forms from the sound recordings show overlap between diagnoses and therefore these characteristics are not regarded as distinguishing characteristics. Although the hyperinflation was only noted as characteristic for DB during the test, there may be some overlap in the conditions regarding this characteristic. Keijzer et al. describe that hyperinflation is also present in asthmatic patients, creating this possible overlap [33]. Therefore hyperinflation may not be a reliable parameter for making the distinction between the conditions.
The results from the video recordings show that when camera footage is added, the percentage of correctly classified recordings increases for every assessor. Clinicians thus need more sources of information to obtain a clear view on the situation of the patient. Although more fragments were correctly classified, the certainty about the diagnoses did not increase. The average certainty score was lower for the video fragments, but the lowest scores came from the assessor who did not assess the audio fragments so a good comparison of the total certainty score is not possible. The average certainty score for assessor 1 remained the same, the average certainty score for assessor 2 increased with 0.6.
Table 3.6 shows that again, different diagnoses are given for most fragments. However, in this case there were four fragments with unanimity among the assessors and three of those were correctly diagnosed.
The characteristics mentioned when assessing video recordings prove that the posture of the patient contributes to the assessment of the patient's condition. Especially for DB, often specific characteristics in the sound (e.g. a biphasic inspiration) in combination with shrugged shoulders are decisive for the diagnosis. Moreover, the stature of the patient helps in supporting or even changing the given diagnosis. These results indicate that the camera footage (in addition to the sound recording) is of added value when assessing the respiratory condition of a patient.

3.4 Conclusions

From chapter 2, it follows that some characteristics which can be found in or derived from respiratory sounds can be used to differentiate between DB and UA. These characteristics are summarized in the upper half of table 3.7.
These findings are in line with the findings from the sound recordings test with clinicians, but the test with clinicians revealed some other possible relevant parameters. Some of the mentioned characteristics will probably not be useful since there is overlap for the different conditions. Characteristics that did not show overlap and were not found in literature, so which can be an addition to the literature, are summarized in the lower half of table 3.7.
The established characteristics are taken to the next steps of this research to see if they can indeed be used to detect DB and UA when analyzing sound recordings.

Table 3.7: Characteristics in respiratory sounds which can be used to distinguish between DB and UA found in literature (upper half) and additional characteristics resulting from test with clinicians (lower half).

	Dysfunctional breathing	**Uncontrolled asthma**
Literature	Inspiratory stridor (with a dominant frequency>500 Hz and a duration>250 ms)	Expiratory wheeze (with a dominant frequency>400 Hz and a duration>80ms)
	Difficulty getting a full breath in, associated with prolonged inspirium	Difficulty getting a breath out, associated with prolonged expirium
	Symptoms predominantly in inspiratory part of respiratory cycle	Symptoms predominantly in expiratory part of respiratory cycle
	Peak of symptoms during exercise, resolved after 1-5 minutes after exercise	Peak of symptoms 3-15 minutes after exercise, but in pediatric patients often already during exercise
	Sometimes hyperventilation	
Test clinicians	High thoracic respiration/hyperinflation	Swallowing mucus
	Hyperventilation	Soft breathing sound
	Biphasic inspiration	Biphasic expiration
	Vibratory respiration sounds	
	Pinched breath	
	Loud breathing sound	
	Groaning	

4 TEST MEASUREMENTS

In the previous chapters, potential characteristics for the auditory distinction between UA and DB were established. It appeared that the most important audible difference between the symptoms lies in the part of the respiratory cycle where abnormalities can be detected. With UA, the *expiration* is often prolonged and high pitched sounds can be heard in this part of the cycle. With DB, the *inspiration* is sometimes prolonged and high pitched sounds can be heard during inspiration. When analyzing the sound recordings, it is thus critical to be able to detect the inspirations and expirations in a fragment. Other characteristics like the loudness and frequency content of the signal should also gain attention.
This chapter discusses the original measurement setup and different test measurements which are performed with this setup. The aim of this chapter is to discover what information can be extracted from the recordings and to assess the quality of the recordings coming from this measurement setup.

4.1 Methods

The test measurements for this part of the study were gathered in two steps: first, sound recordings were made from two healthy individuals on a treadmill; second, sound recordings from children during an ECT were gathered.
All the test measurements were performed in a climate controlled room in MST where the temperature is kept at 10°C. This room contains a treadmill on which the ECTs are normally performed. The room is equipped with a camera and directional microphone, connected via an audio video bridge, to capture sound and vision during an ECT. For the purpose of this study, only the captured audio is used.

For recording the breathing sound during the ECTs, the Audio Technica ES933ML/MIC is used. This is a condenser hanging microphone with a standard angle of acceptance of 90°. The frequency response curve, which shows the ability of the microphone to transform acoustic energy into electric signals, can be flat or have a low-end roll-off. For respiratory sounds a flat frequency response is required since this means that the microphone is equally sensitive to all frequencies over the range of 20 to 20,000 Hz.[53] The signal-to-noise ratio is 70 dB, which is adequate since CORSA recommends a signal-to-noise ratio of at least 60 dB for breath sound recordings. The dynamic range is 109 dB, which also fulfills the CORSA recommendation of a dynamic range of at least 60 dB.[54] The polar pattern of a microphone shows the sensitivity of the microphone for sound coming from different angles. The CORSA recommendation is to have an omnidirectional polar pattern so that sound from all directions is picked up [54]. However, since in this case the desired sound only comes from one direction, a smaller angle of acceptance is desired. The microline cardioid polar pattern of 90°, which is default for the Audio Technica ES933ML/MIC, is therefore adequate. The specifications, polar pattern and frequency response of the microphone can be found in appendix C. In the measurement setup, the arm of the microphone is attached to the wall next to the treadmill and the microphone is directed at the mouth of the person on the treadmill.

MATLAB 2020b was used for the extraction of audio from mp4 files and for further processing of the data. The Signal Processing Toolbox was used in addition to the standard MATLAB functions. Sony Vegas Pro 13.0 was used for visualizing and simultaneously playing audio data. This software was not used for obtaining results, but merely to gain a better understanding of what audible respirations looked like in a time-amplitude plot. Microsoft Excel was used for simple calculations. IBM SPSS Statistics 27 was used to perform statistical tests on gathered data.

The protocol and analysis for the test measurements with healthy individuals will be discussed first. Afterwards, the protocol and analysis for the test measurements with patients will be discussed.

4.1.1 Test measurements with healthy individuals

For the measurements with the first healthy individual, the individual performed a regular ECT during which the sound was recorded. Because the breath was not audible in the sound recording and the signal appeared to be very noisy, the analysis was terminated at an early stage. For the other individual, various recordings were made. Measurements with only the treadmill were also included in this collection of measurements. The following recordings were made:

1. Only treadmill running, nobody on it
    (a) Microphone not properly directed at treadmill
        i. Treadmill velocity of 4, 5, 6, 7 and 8 km/h
    (b) Microphone properly directed at treadmill
        i. Treadmill velocity of 4, 5, 6, 7 and 8 km/h
2. Individual running on treadmill
    (a) Microphone not properly directed at mouth, unclear breath
        i. Treadmill velocity of 4, 6 and 8 km/h
    (b) Microphone properly directed at mouth, unclear breath
        i. Treadmill velocity of 4, 5, 6, 7 and 8 km/h
    (c) Microphone properly directed at mouth, clear breath
        i. Treadmill velocity of 8 km/h
3. Individual breathing in microphone without treadmill running

The goal of including different combinations of settings, was to be able to gain insight in the noise generated by the treadmill at different velocities. This could enable filtering noise from the recordings with someone running on the treadmill.

In the analysis of the signals from the second individual, different steps were taken to explore to which extent the recordings provided useful information regarding breathing pattern and frequency content. At first, the raw data with a sampling frequency of 48,000 Hz was plotted over time and periodograms were created by using the FFT. In all recordings the periodogram showed high peaks around 180-220 Hz. Since these peaks also showed up in the recordings with only the treadmill, these frequencies were expected to be noise. As is described in chapter 2, mouth breathing results in sound with frequencies between 200 and 2,000 Hz. Taking these findings into consideration, the signals were filtered with a third order high-pass filter with a cut-off frequency of 230 Hz and a low-pass filter with a cut-off frequency of 2,500 Hz.
The frequency spectra contained information until approximately 2,500-3,000 Hz. According

to Shannon's sampling theorem [50], the sampling frequency must be higher than twice the maximum frequency in the signal to recover all frequencies. Since the original sampling frequency was 48,000 Hz, it was possible to resample the signals to decrease computation time and load. The signals were resampled with a sampling frequency of 10,000 Hz to adhere to Shannon's theorem with a margin.
Various signal analysis techniques were applied to all filtered and resampled recordings. For this purpose, the MATLAB Signal Processing Toolbox was used. First, the amplitude of the signal was again plotted over time to see whether the breathing pattern, or a pattern from the treadmill sound could be recognized. The recordings were played in Sony Vegas Pro to verify if a pattern in a plot indeed belonged to a breathing or treadmill pattern in the sound. Second, the median frequency of the signal was computed with help of the FFT to gain general knowledge on the frequency content of the signal. Next, the following types of frequency spectra were plotted: the general Fourier transform to see which frequencies appear in the signal; the periodogram to visualize the power of certain frequency bands in the signal; the Welch averaged periodogram to account for changes in frequency content over time; the autoregressive power spectral density estimate with Burg's method to generate a smoother estimate of the power of frequency bands. The main peaks in these frequency spectra were determined. A spectrogram was plotted to visualize changes in frequency content over time. Lastly, an envelope was created for the positive values of the signal. It was expected that the envelope could reveal the pattern of a respiration if the noise level was low. The findpeaks function was used to detect peaks in the envelope. From the resulting envelope, a periodogram and spectrogram were again computed to possibly determine (changes in) the respiratory rate.
Apart from these techniques applied to all signals, additional analysis steps were applied to a selection of the signals. For the recordings described under 2, it was tried to apply the independent component analysis algorithm from Gvert et. al [55] in order to separate the sources of sound (breath of individual, treadmill sound and footsteps on treadmill). For the recordings described under 1b, the main peaks per velocity were used to create bandstop filters. It was thought that filtering these dominant frequencies from the recordings described in 2b with the corresponding treadmill velocities, might lead to a cleaner signal in which the breathing pattern would be clearer. Lastly, the recording described at 3 was used to compare inspirations with expirations in time and frequency domain.

4.1.2 Test measurements from the clinic during ECT

In the outpatient clinic, video recordings were made during scheduled ECTs by clinicians. After every ECT, the clinician who made the recording selected a short fragment (mostly between 10 and 30 seconds) of the video recording which best represented the condition of the patient. These recordings were available on a computer in the climate room. From these video recordings, the sound was extracted with MATLAB and these sound recordings were taken home for analysis.

For each recording, an appropriate fragment was selected for analysis. A fragment was considered appropriate when it did not contain speech and it is preferably in the last minute of exercise where symptoms peak. In most cases, the 10-30 seconds fragment created by the clinician was used for this purpose. In case this fragment contained speech, a new fragment was created from the entire recording. All fragments were filtered and resampled in the same way as in section 4.1.1.
Every fragment was played in Sony Vegas Pro to assess the quality of the recording and to detect possible adventitious sounds in the recordings. For every fragment, the median frequency was determined and Welch averaged periodograms were created of which the peak frequencies were determined. These values for the different fragments were compared with each other to discover whether there were frequency domain similarities between different recordings within the same diagnosis category. Six fragments that were regarded as good quality sounds were selected to

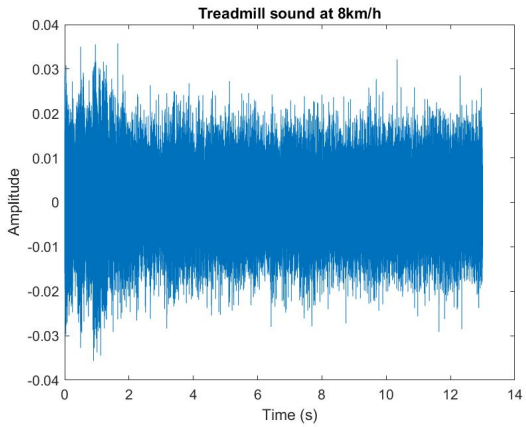
manually label inspirations and expirations. The time instances of these inspirations and expirations were entered in Microsoft Excel and from this, the duration of inspiration, duration of expiration and average ratio between those durations was computed. These ratios were compared between the different diagnosis categories, because both literature [2, 17] and clinicians describe a prolonged inspirium to be typical of DB and a prolonged expirium to be typical of UA. In addition, an envelope was created for these six fragments and MATLAB's findpeaks function was used in an attempt to automatically determine the time instances of inspirations and expirations. For the findpeaks function, different settings were used. Various values were tried for the minimal peak value, the minimal peak distance and minimal peak prominence. The values that were tried were determined by observing the peaks belonging to inspirations and expirations in time-amplitude plots. In addition, an adaptive minimal peak distance was tried which was based on the amount of full breaths that were manually counted in a certain time interval.
Lastly, from these six fragments time and frequency domain plots were created for one inspiration-expiration couple per fragment. The plots were compared to the time and frequency domain plots of an inspiration and expiration of the healthy individual described in section 4.1.1.

4.2 Results

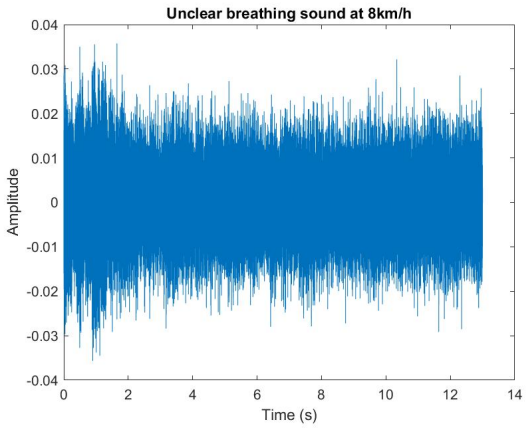
First, the results of the analysis of recordings with healthy individuals will be shown. Next, the results of the analysis of recordings with patients during ECTs will be shown.

4.2.1 Test measurements with healthy individuals

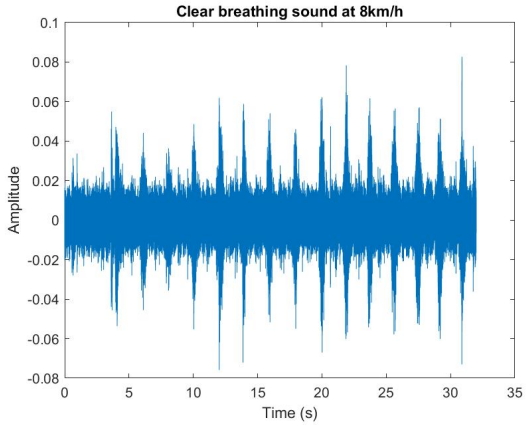
Figure 4.1 shows time domain plots for four different setups: only treadmill running with microphone directed (at 8 km/h), unclear breath when running on treadmill with microphone directed at mouth (8 km/h), clear breath when running on treadmill with microphone directed at mouth (8 km/h) and clear breath without treadmill running.
Figures 4.1a and 4.1b do not show a repeating pattern. The recordings with a clear breath and only breath in figures 4.1c and 4.1d respectively do show a repetitive pattern. Listening to these fragments in Sony Vegas Pro confirms that this repetitive pattern belongs to the respiration.



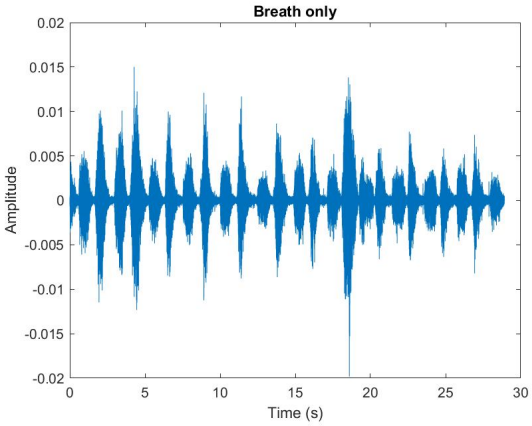
(a) Treadmill only.



(b) Unclear breath with treadmill.



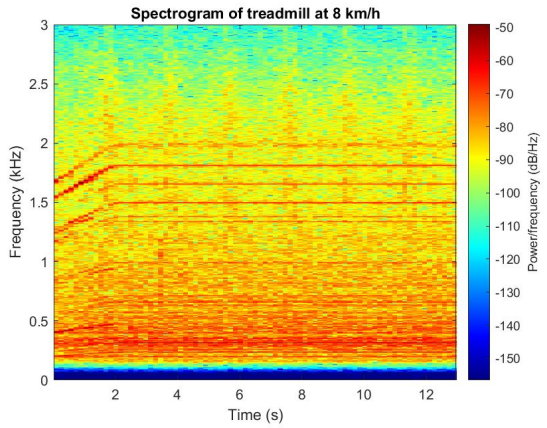
(c) Clear breath with treadmill.



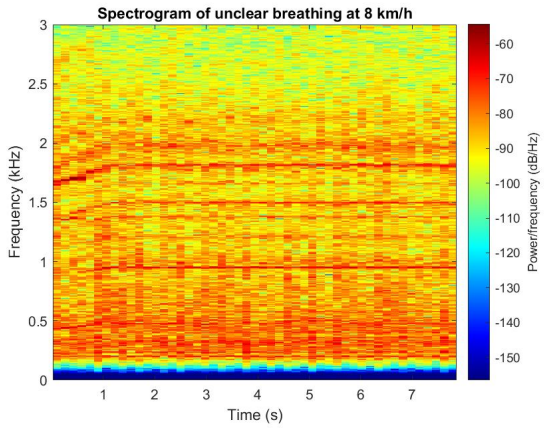
(d) Breath only.

Figure 4.1: Comparison of time domain plots with different setups. The velocity is 8 km/h whenever the treadmill is used.

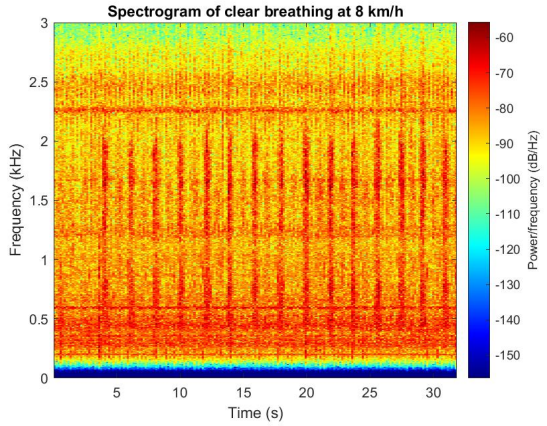
Figure 4.2 shows the spectrograms belonging to these four fragments. In figures 4.2a and 4.2b, the changes over time cannot be followed. In figures 4.2c and 4.2d, it is possible to detect changes over time but pointing out differences in frequency content between inspirations and expirations is not evident.



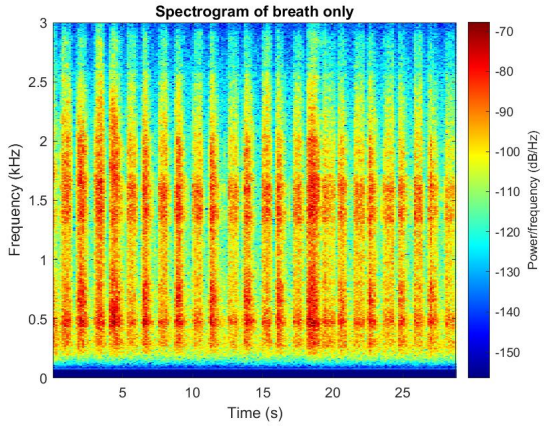
(a) Treadmill only.



(b) Unclear breath with treadmill.



(c) Clear breath with treadmill.



(d) Breath only.

Figure 4.2: Comparison of spectrograms with different setups. The velocity is 8 km/h whenever the treadmill is used.

Table 4.1 shows the frequencies that were dominant in the measurements with only the treadmill running. The entire overview with the different methods to determine peak frequencies can be found in table D.1 in appendix D. Table 4.1 also includes the cut-off values that were used to create bandstop filters with which the measurements with an unclear breath were filtered. Figure D.1 in appendix D shows a periodogram (for 8 km/h) with which the cut-off values are determined. Besides the peaks reported in table 4.1, multiple smaller peaks are visible in this periodogram.

Table 4.1: Dominant frequencies of the treadmill sound at different velocities. The cut-off values for the bandstop filters are based on these dominant frequencies.

Velocity (km/h)	**Dominant frequency/frequencies (Hz)**	**Bandstop filter cut-off values (Hz)**
4	239	230-245
5	298	290-305
6	356	350-365
7	1587, 415	1580-1595, 410-425
8	316, 1496	310-325, 1490-1505

Figure 4.3 shows an example of applying such a bandstop filter, here for a treadmill velocity of 8 km/h. Although the plots look slightly different, a clear breathing pattern is still not visible after filtering.

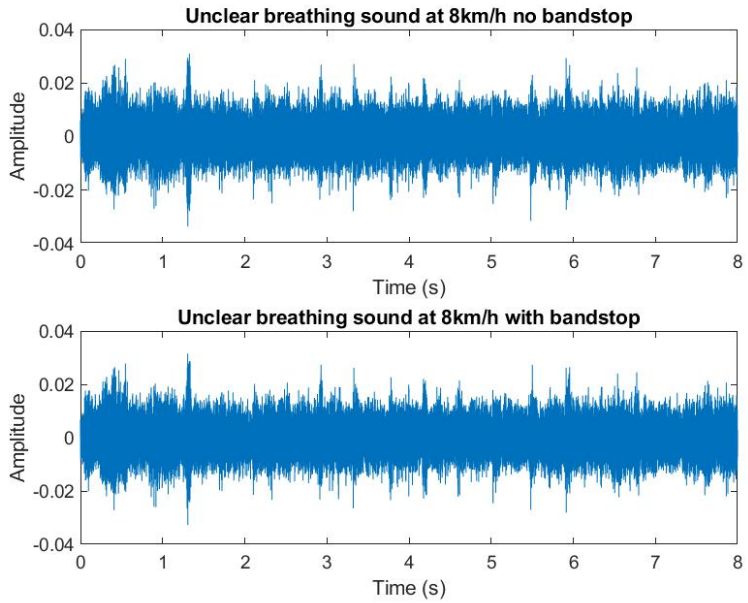


Figure 4.3: Time domain plot before (upper) and after (lower) bandstop filtering.

When listening to both signals, the peaks in the plots appear to belong to the footsteps rather than to a breathing sound in both cases. Similar results were obtained for the other velocities.

Figure 4.4 shows the envelope of the recording with only breath sound and accompanying peaks. The pattern of the respiration is followed by the envelope and the peaks correctly mark inspirations and expirations. In figure D.2a in appendix D, the periodogram of this envelope can be found. The value of 0.84 Hz has the highest magnitude in this periodogram, which corresponds to the amount of peaks in figure 4.4 divided by the duration in seconds of the fragment.

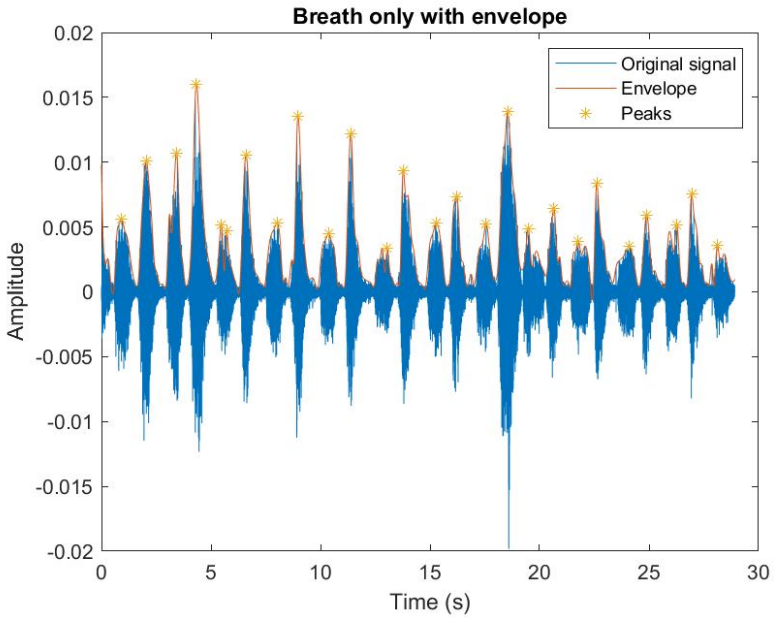


Figure 4.4: Envelope of the positive values of the recording with only breath sound. Peaks with a minimal peak prominence of 0.002 are also indicated.

Figure 4.5 shows the envelope of the recording with a clear breath (with ambient noise) and accompanying peaks. The envelope follows the pattern of the respiration but it also follows peaks in between inspirations and expirations. The peaks correctly mark inspirations and expirations in most cases, but some inspirations and expirations are skipped while other peaks do not belong to an inspiration or expiration.

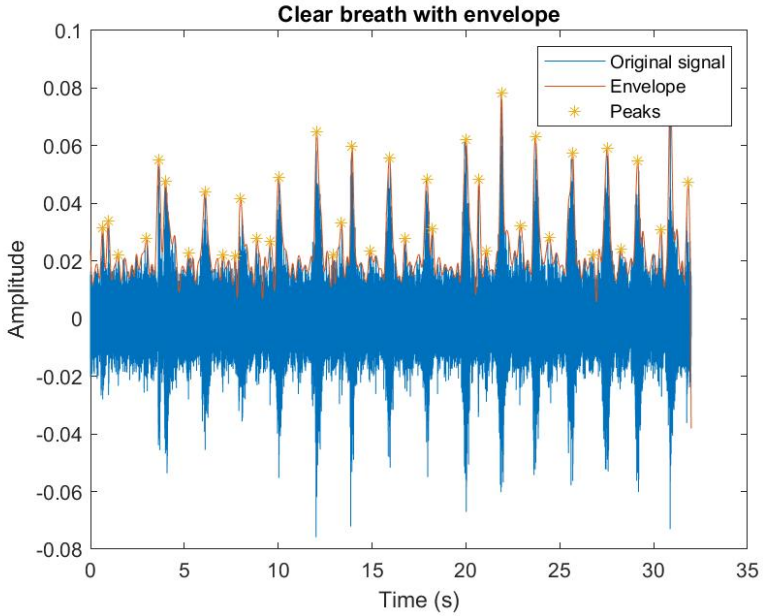


Figure 4.5: Envelope of the positive values of the recording with a clear breath while running on the treadmill. Peaks with a minimal peak prominence of 0.008 are also indicated.

In figure D.2b in appendix D, the periodogram of this envelope can be found. The value of 0.51 Hz has the highest magnitude in this periodogram, which corresponds to the amount of relatively high peaks (expirations) in figure 4.5 divided by the duration in seconds of the fragment. The fairly large peak at 1.02 Hz is expected to be the second harmonic of this 0.51 Hz.

The spectrograms of the envelopes are not shown because they showed a bad frequency resolution and did not provide additional information. The envelopes for the other measurements are not shown since no patterns could be detected in these signals.
When applying the independent component analysis algorithm, only one source of sound was detected. The algorithm did not succeed in separating the respiratory sound, treadmill sound and footstep sound.

Figure 4.6 shows the selection of a clear inspiration and subsequent expiration in the breath only sound, with a plot of these segments in the time domain.
When looking at the y-axis of the right figure, the expiration shows a higher amplitude than the inspiration. The ratio between the amplitude of the expiration and preceding inspiration has been determined for every instance and on average this ratio was 2.24. The amplitude of the inspiration looks quiet evenly distributed, while the amplitude of the expiration works towards a peak.

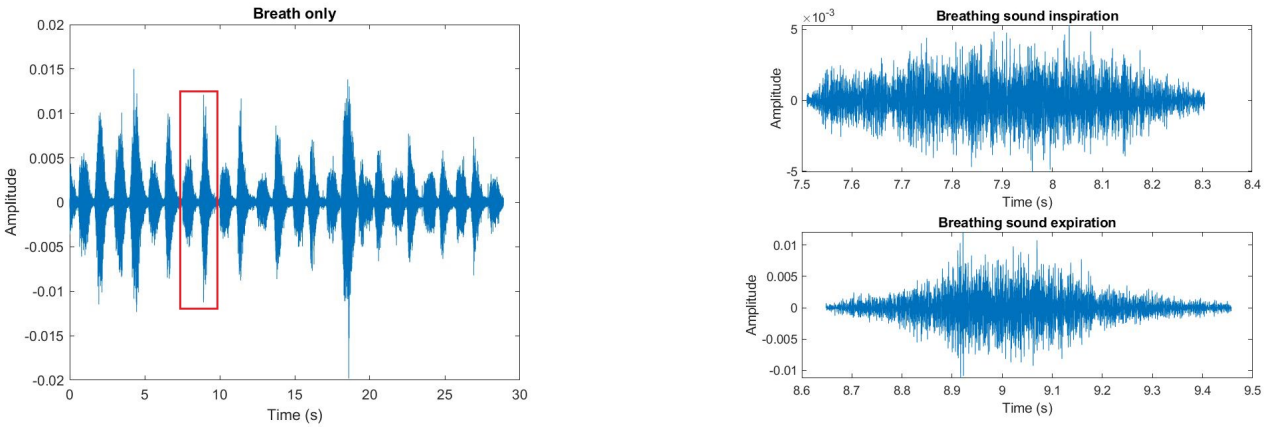


Figure 4.6: Selection of inspiration and expiration from breath only recording (left) and plot of inspiration and expiration in time domain (right).

Figure 4.7 shows the periodograms of the selected inspiration and expiration. The spectra show peaks in the same regions. The main differences between these spectra are that the frequency component around 1,500 Hz is relatively larger during inspiration than during expiration (although in absolute terms it is larger during expiration) and that the expiration shows an additional peak around 650 Hz.

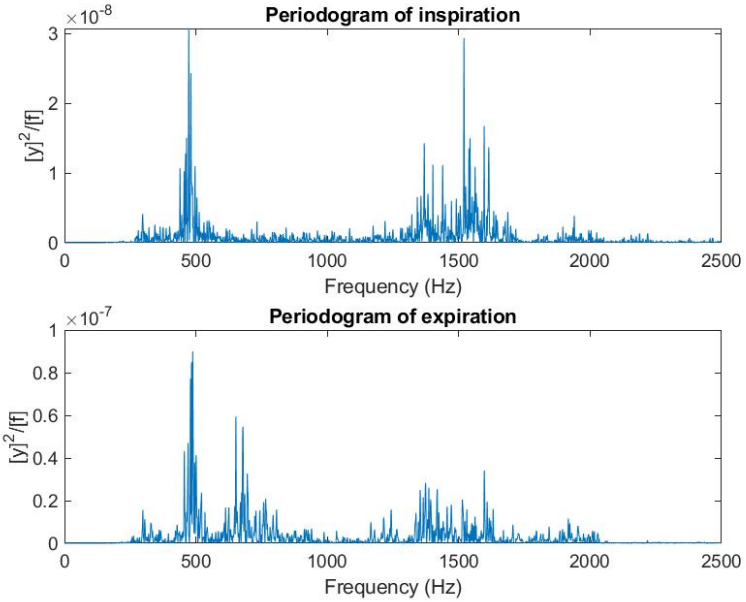


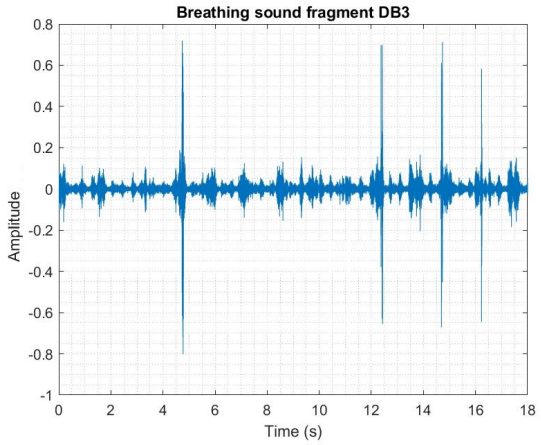
Figure 4.7: Periodogram of selected inspiration (upper) and expiration (lower).

The time domain and frequency domain plots of the other inspiration-expiration couples in this recording showed similar characteristics. Another example of these plots for an inspiration-expiration couple is shown in appendix D.2.

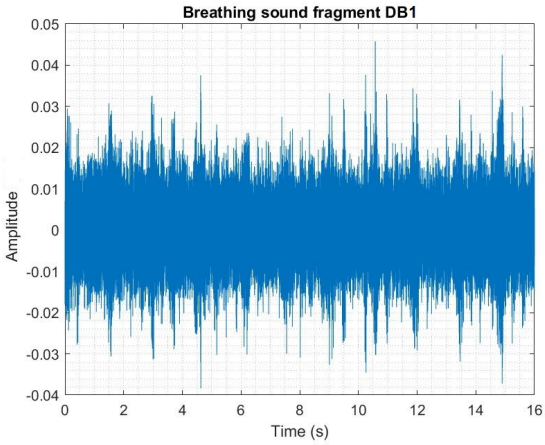
4.2.2 Test measurements with patients

Eighteen sound recordings from children during an ECT were obtained. Five of these recordings were from children with DB, seven from children with UA and six from children with no established respiratory diagnosis.
Figure 4.8 shows fragments of recordings in the time domain, belonging to the same diagnosis

category (DB). In figure 4.8a, the respiratory pattern is clear with peaks for both the inspirations and expirations. Figure 4.8b shows a noisy signal in which the respiratory pattern is not represented by distinct peaks. Listening to the signals confirms that the peaks in figure 4.8a belong to inspirations and expirations while in figure 4.8b, the peaks display a combination of the respiratory pattern and the steps on the treadmill.



(a) Clear respiratory pattern visible.



(b) No clear respiratory pattern visible.

Figure 4.8: Comparison of the breathing sound quality of two recordings belonging to the DB category.

These two figures are a good representation of the difference in quality between signals: in the total collection of fragments, some showed a clear respiratory pattern while in others the respiration could not be detected.

Performing a Kruskal-Wallis test proved that there is no significant difference between the median frequencies (P=0.521) and highest Welch peaks (P=0.908) across the different diagnosis categories. All values for the median frequencies and highest Welch peaks can be found in table D.2 in appendix D.

The six fragments of which the inspiration and expiration times were determined manually, consisted of two fragments for each category. The average ratio between the duration of the inspiration and expiration per fragment is shown in table 4.2.

Table 4.2: Average ratio between duration of inspiration and duration of expiration.

	DB	**UA**	**ND**
Fragment 1	1.48758	0.750475709	1.25758
Fragment 2	1.067811696	0.668753485	0.86985

In these fragments, the ratios for DB are higher than for UA. The ND category has one value that comes closer to the DB ratios and one value that comes closer to the UA ratios.

Figure 4.9 shows an example of the automatic peak detection with the measurement that showed the clearest respiratory pattern, in this case from a DB patient. A minimal peak distance of 0.5 was used and the inspirations and expirations were labelled manually in the figure.
Most peaks were found by the algorithm, but some were not. When the minimal peak distance was decreased in order to find more peaks, sometimes multiple peaks within one inspiration or expiration were marked as distinct peaks. Figure 4.9 shows a varying pattern: sometimes, an

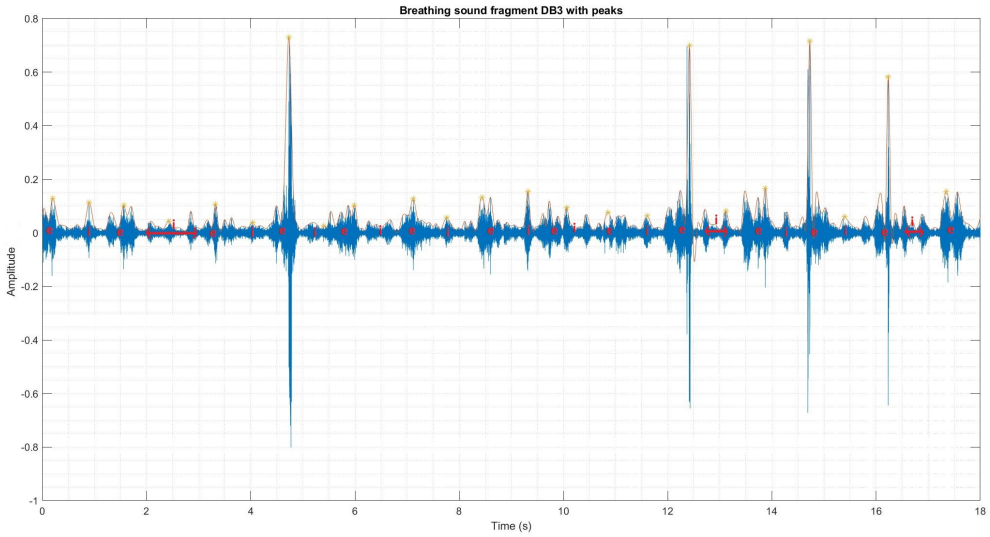


Figure 4.9: Inspirations and expirations marked with help of the envelope and findpeaks functions, and manually where an i displays an inspiration and an e an expiration.

expiration reaches a value of 0.7 while a different expiration stays below 0.1 and in some cases, an inspiration is just a short peak while at other times it takes almost a second.

Figure 4.10 shows the recording of an UA patient with a soft breathing sound. In this figure, it is not possible to recognize the inspirations and expirations by eye.

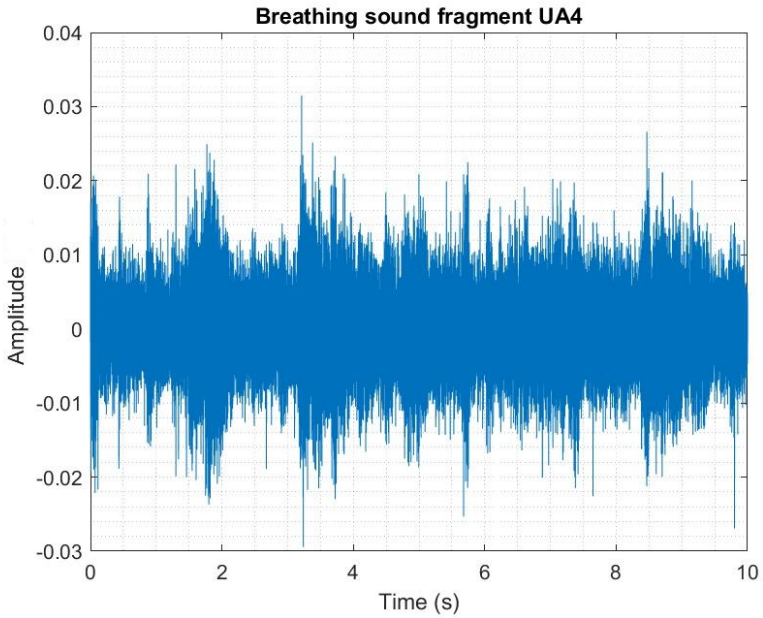


Figure 4.10: Unclear (soft) breath.

The time and frequency domain plots of the inspirations and expirations of the six fragments (one inspiration and expiration per fragment) can be found in appendix D.2. For both ND time domain plots, the amplitude during expiration is higher than during inspiration. Both expirations seem to work towards a clear peak, while the inspirations appear more even concerning

the amplitude. In the DB time domain plots, the amplitude is higher during inspiration than during expiration and the patterns seen with ND are not detected. One of the UA time domain plots has a higher amplitude during inspiration, while the other has a higher amplitude during expiration. Again, the described patterns are not seen here. For all six fragments, the peak around 1,500 Hz is not relatively larger during inspiration than during expiration as was seen with the healthy individual: in some cases it is relatively smaller during inspiration and in some cases there is no peak around 1,500 Hz. The additional peak around 650 Hz in the expiration plot of figure 4.7 is not found in the six fragments from the clinic.

4.3 Discussion

The analysis of the different test measurements gives insight in the possibilities for detecting DB and UA based on sound recordings. From the time and time-frequency domain plots of the different setups in figures 4.1 and 4.2, it was found that a clearly audible breath sound is critical for analysis of the respiratory pattern in both time and time-frequency domain. However, the spectrograms did not show obvious differences in frequency content between the inspirations and expirations. The spectrograms may thus not be the best way to visualize respiratory patterns in sound recordings.
Figure 4.3 proved that removing the treadmill sound from a recording to retrieve a breathing pattern, cannot be achieved by simply filtering the dominant frequencies generated by the treadmill. The fact that other small frequency components were visible in the periodograms of the treadmill sounds, may be the cause for the inability to remove treadmill sound by using just one or two bandstop filters. However, when more filters would be created to remove the other frequencies from the treadmill, it is plausible that relevant frequencies appearing in the respiration will be removed as well.
Figure 4.4 shows that for a breath recording without clear noise, the breathing pattern can easily be visualized with an envelope and inspirations and expirations can be determined with a peak finder. The respiratory rate can be detected from the periodogram of this envelope. However, for a breath recording with a clear breath but also ambient noise, the breathing pattern can be detected to some extent but is also disturbed by the noise. The periodogram of this recording only showed the frequency with which the large peaks, belonging to the expirations, appeared in the signal. For a signal with ambient noise, it thus seems more difficult to detect the inspirations since they stand out less than the expirations and the noise causes a certain offset in the entire signal. This stresses the need of minimizing the disturbances in the recordings.
The median frequencies and highest Welch peaks which were computed for the eighteen recordings from the clinic, did not have significantly different distributions across the diagnosis categories. Therefore the frequency information can, at this point, not be used to make a distinction between the different categories.
The ratios between the duration of the inspiration and the duration of the expiration did show some differences across the categories, although no statistic test is performed to verify this due to the low amount of samples. The ratios for DB are clearly higher than for UA which indicates a relatively long inspirium for DB and a relatively long expirium for UA. This is in line with the findings from chapters 2 and 3. From these results it could be stated that the ratio between the duration of inspiration and expiration is a relevant parameter in distinguishing DB from UA. However, the two ND fragments have divergent ratios which come close to both the DB and UA ratios. This makes it hard to differentiate between DB, UA and ND. Moreover, only two fragments per category are taken into account which makes it likely that these findings are coincidental. Unfortunately, the fact that most measurements were too noisy to determine relevant time instances in combination with the time-consumingness of this determination, hampered the inclusion of more fragments.
The influence of noise on the ability to detect a respiratory pattern was also investigated for

data from the clinic. In figure 4.9, the noise level appeared to be quiet low and a peak finder was able to correctly detect most of the peaks. Correctly detecting peaks would enable the computation of the ratio between the inspiration and the expiration, which might be useful in distinguishing the different diagnoses as was described above. However, not all peaks were found with the first threshold value of the peak finder. Even if one peak is skipped, the computation of the ratio would go wrong because an algorithm would lose track of which peak belongs to an inspiration and which peak belongs to an expiration. This could be solved by changing the threshold value to detect more peaks, but in some cases this leads to labelling multiple peaks within one inspiration or expiration as distinct peaks. Regarding an inspiration or expiration as two peaks would again derange the computation of the ratio. For this example with a fairly clear respiratory pattern, it is already not possible at this point to come up with a procedure to correctly detect all peaks belonging to an inspiration and expiration.
In figure 4.10, the inspirations and expirations cannot be detected by eye due to the soft sound and the noise in the signal. For such a low quality recording, determining peak detection thresholds would become even more difficult.
When listening to the sound recordings, it is evident that there is a lot of variation in respiratory rate between subjects. Moreover, the loudness of the recordings differs a lot, which makes using a certain amplitude threshold impossible as well. Because of this, adaptive thresholds are required. However, apart from the inter-individual variability, there is also an intra-individual variability in amplitude and durations of inspirations and expirations as was described for figure 4.9. This intra-individual variability hampers a peak detection based on for example the respiratory rate of a patient or the amplitude values of an inspiration and expiration. It is possible that this varying pattern in figure 4.9 is also caused by the respiratory condition of this specific patient. This pattern comes from a DB patient, and an inconsistent pattern is sometimes described by clinicians as a characteristic of DB. It is not verified for other recordings whether the patterns were so varying or not, but the randomness of the pattern may be a relevant characteristic in future steps.

Section 4.2.1 showed two characteristics for the time domain plots of inspirations versus expirations and two characteristics for the frequency domain plots. The first time domain characteristic, being the higher amplitude during expiration than during inspiration, was also found in the ND fragments but not in the DB fragments and partly in the UA fragments. Therefore this characteristic may be useful for distinguishing ND from DB, but it is not clear how UA would be characterized in this division. The second time domain characteristic, being the pattern of an expiration working towards a peak and an inspiration which appears more even, was found in the ND fragments but not in the other fragments. Therefore this characteristic may be useful for distinguishing ND from DB and UA. More data is required to further investigate these two time domain characteristics. The frequency domain characteristics, which were 1) the peak around 1,500 Hz being relatively larger during inspiration than during expiration and 2) the additional peak around 650 Hz for the expiration, were not found in any of the recordings from the clinic and therefore these characteristics are not useful at this point.

4.4 Conclusions

From section 4.2.1 it already became clear that a low noise level was required to perform good signal analysis and be able to detect a respiratory pattern. Section 4.2.2 confirmed this finding and showed that the quality of the majority of the recordings was too poor to detect a pattern, let alone to mark precise moments in the respiratory cycle. Moreover, the great variations in frequency contents of signals belonging to the same diagnosis category, combined with the absence of audible adventitious sounds, indicated that the frequency and time-frequency domain analysis are probably less relevant in the differentiation than the time domain analysis. Lastly,

the determination of threshold values for inspiration and expiration detection turned out to be very difficult. From these points, it is concluded that the focus in the next part of this study should be on changing the measurement setup to improve the quality of the sound recordings. Probably, a machine learning approach has to be followed afterwards because the inter- and intra-individual differences in respiratory sounds complicate the manual creation of a differentiation algorithm.

5 ADJUSTMENTS IN MEASUREMENT SETUP

Because the previously analyzed signals were too noisy to (automatically) extract relevant information from, changes will be made in the measurement setup with the aim of improving the quality of the signals. For this aim, alterations are made to the default settings of the directional microphone and a protocol is written in which an extra microphone is added to the setup. In addition, six sound recordings are made of patients directly after their ECT when the treadmill is not running anymore.
These changes in the setup will be discussed in this chapter. This chapter will aim to answer the questions:

- How can the measurement setup be optimized to attain quality of the sound recordings that is capable of revealing breathing patterns and characteristic respiratory sounds?
- To what extent is differentiation between DB, UA and no respiratory diagnosis possible based on sound recordings gathered with the improved measurement setup?

5.1 Methods

In order to improve the quality of the sound recordings, three different changes in the signal acquisition were made. First, the settings from the microphone were changed in an attempt to filter out ambient noise during recording. Second, an extra microphone from a smartphone was placed in the direction of the treadmill in order to record the noise present during a recording. The reasoning behind this second microphone is that with this recording of predominantly noise, the noise can be subtracted or filtered from the recording from the directional microphone containing the desired signal. Lastly, six patients were asked to stay on the treadmill and keep breathing in the microphone for ten seconds after exercise. In this way, breath sound recordings were gathered in which the treadmill sound was not present.
The methods for these three changes will be described separately in the next sections.

5.1.1 Changes in the microphone settings

In the microphone settings, it is possible to apply a low-pass and high-pass filter, noise cancellation and echo cancellation. Moreover, the gain of the microphone can be adjusted and an automatic gain control can be used. Different combinations of these settings were tried with test measurements during which a healthy individual was running on the treadmill.
The low-pass filter is used to reduce hissing sounds that the microphone picks up and is applied by specifying the highest frequency for the microphone to pick up. This highest frequency was set at 2,000 Hz. The high-pass filter is used to reduce low-frequency background noise that the microphone picks up and is applied by specifying the lowest frequency for the microphone to pick up. This lowest frequency was set at 250 Hz. Echo cancellation is used to subtract the far-end audio from the sound that the microphone picks up and is applied by enabling 'echo cancellation'. Noise cancellation is used to subtract background noise from the sound that the

microphone picks up and is applied by enabling 'noise cancellation'. Changing the gain is used to adjust the loudness of the signal before (pre) or after (post) the system has processed the sound. It is applied by shifting the sliders from the pre and post levels up and down. For both the pre and post level, measurements were made in which the gain was lowered and in which the gain was increased. The automatic gain control automatically adjusts for differences in volume as different people speak and is used by enabling 'Automatic Gain Control'.
The recordings with the different combinations of settings were studied in time domain to determine which combination resulted in the clearest respiratory pattern. New recordings in the clinic were made with these settings and these new measurements were analyzed to determine the extent to which the adjusted settings improved the quality of the signal.

5.1.2 Addition of microphone

In order to make a recording of the noise which could be used to remove the noise from the signal from the directional microphone, a smartphone was used. This smartphone was placed in a phone holder that was placed near the treadmill such that the microphone of the smartphone could be directed at the treadmill. The recordings were made at a constant velocity of the treadmill which was set to provoke an auditory breath. Recordings were also made at a lower pace while playing an earlier recorded breathing sound with another smartphone, to discover if this breathing sound could be retrieved with the improved measurement set up and the applied subtraction or filtering. Lastly, recordings were made with a healthy individual running on the treadmill while pacing his inspirations and expirations, to discover whether those paced in- and expirations could be seen in the recorded signal. The protocol for these measurements can be found in appendix E.2.
Afterwards, different approaches were taken with the aim to remove the noise from the recording of the directional microphone. To test the different approaches, the measurement in which the individual was running on the treadmill with audible breath and without speech in the background was used first. Before the different techniques were tried, the directional microphone recordings were downsampled to the same sampling frequency as the one of the smartphone (32,000 Hz). Moreover, the recordings were aligned in time as good as possible by making use of the 'start' and 'stop' marks in the recordings. Afterwards, the index of the maximum cross-correlation between both signals was determined. Based on the index of the maximum cross-correlation, the signals were further aligned in time.
The following approaches for noise reduction were applied: subtraction of noise signal from original signal, Wiener filtering, adaptive filtering and wavelet denoising. For each approach, a brief explanation of the theory and the implementation will be given.

Subtraction of noise signal from original signal

At first, it was tried to simply subtract the recorded noise from the signal recorded by the directional microphone. Because the amplitude of the noise in the original signal is not as high as the amplitude of the noise recording, the noise signal was first multiplied with a value between 0 and 1 before it was subtracted from the directional microphone signal. This was tried for ten different values of this 'gain'. It was also tried to tweak the gain by inspecting the power of the different signals, so by dividing the power of the original signal by the power of the noise signal. This number was then multiplied with the noise signal and this product was subtracted from the original signal. Another attempted method to tweak this gain was by looking at the values of the amplitude at certain moments in time where no respiration was registered (so only noise). The amplitude at these moments was compared between the two measurements and again the division was performed to determine the gain value.

Wiener filter

Second, Wiener filtering was applied. A Wiener filter is used to obtain an estimate of a desired signal when 1) the desired signal corrupted with noise and 2) a signal correlated with the noise in signal 1 are available. The coefficients of this filter are determined with the minimum mean-square-error (MMSE) principle: they are chosen such that they minimise the mean squared difference between the filter output and the desired signal. The assumption with a Wiener filter is that both 1) and 2) are (quasi)-stationary.[56] Figure 5.1 contains a schematic representation of how a Wiener filter can be used to filter out noise from a signal corrupted with noise.

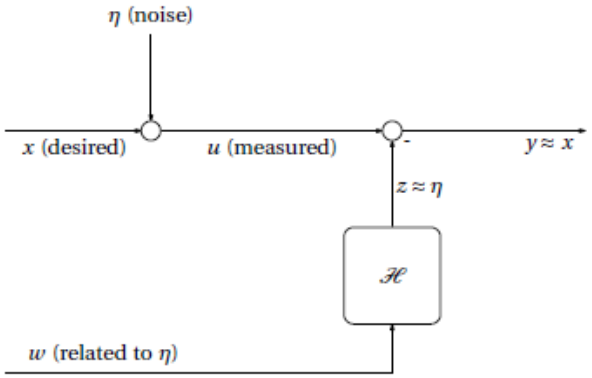


Figure 5.1: Visualization of Wiener filter for noise reduction [50].

An estimation of the noise corrupting the desired signal is made by filtering the signal correlated with the noise with a Wiener filter, after which this estimation is subtracted from the noise corrupted signal.
A Wiener filter can either be restricted to be a finite-duration impulse response (FIR), a causal infinite-duration impulse response or a noncausal infinite-duration impulse response [57]. Since FIR filters are relatively simple to compute, stable and practical, this type of Wiener filter is used for this problem. It must be noted that FIR filters do need a large number of coefficients to approximate the desired signal, compared to infinite-duration impulse response filters.[56]
For implementation of the FIR Wiener filter, a MATLAB function provided during the course Advanced Techniques for Signal Analysis at the University of Twente is used. This function requires a desired signal corrupted with noise, a signal correlated with the noise and a filter order as inputs. The cross-correlation between the noise corrupted signal and the signal correlated with noise is computed and divided by the autocorrelation of the noise correlated signal. This results in the filter coefficients, which will then be used to filter the noise corrupted signal (resulting in the z shown in figure 5.1). Lastly, this z is subtracted from the noise corrupted signal to obtain the estimate of the desired signal.
The Wiener filter was applied with different filter orders. Small orders were tried in order to minimize computation time and large orders were tried to make sure the acoustic dynamics of the environment fit in the filter. Moreover, regularization has been performed by changing the diagonal of the autocorrelation matrix. This was done by adding a small value times the diagonal value to the original diagonal value. The idea was that this would make the solution more stable.

Adaptive filters

As a next step, adaptive filtering was applied. Adaptive filters have coefficients that change automatically in response to statistical variations in the filter's operating environment [58]. In this way the filter can converge to an optimal state, in which the optimization criterion is a cost function. The earlier described MMSE principle is used most often as cost function. Figure 5.2

contains a schematic representation of how an adaptive filter can be used for noise cancellation. In this case, *s(k)* represents the signal of interest and *n(k)* represents the noise with which this signal is corrupted. The signal *n'(k)* is correlated to the noise in the corrupted signal and is used as input for the adaptive filter. The adaptive filter adjusts its coefficients to reduce the difference between *y(k)* (the filtered noise) and *d(k)* as long as the input noise *x(k)* remains correlated to the noise in *d(k)*. The filtered noise *y(k)* is subtracted from *d(k)* to obtain a clean signal, *e(k)*. In most adaptive filter applications the *e(k)* represents the error signal which should converge to zero, but with noise cancellation the error signal represents the desired signal.[59]

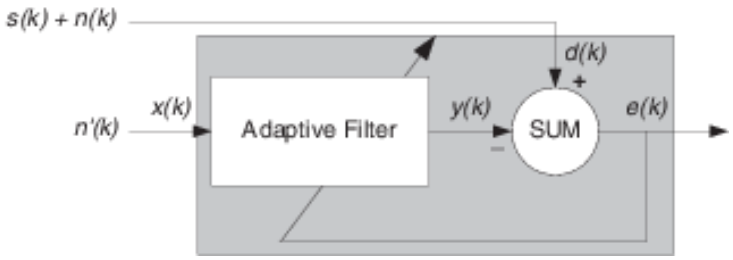


Figure 5.2: Concept of noise filtering with adaptive filter [59].

In this approach, the use of two microphones is assumed where the primary microphone records the noisy input signal and the secondary microphone records noise uncorrelated to the desired signal but correlated to the noise recorded by the primary microphone [60].
Two algorithms are tried for this adaptive noise cancellation: recursive least squares (RLS) and least mean squares (LMS). The difference between the filter types lies in the way they adapt their coefficients. The LMS filter adapts its coefficients based on the error at the current time, while the RLS filter adapts its coefficients based on the total error computed from the beginning. The RLS filter converges faster and usually has a lower error, but this comes at the cost of increased computational cost.[61]
For both algorithms, a great amount of taps (coefficients) is chosen to make sure the acoustic dynamics of the environment fit into the filter.

RLS

For this algorithm, the dsp.RLSfilter from MATLAB's DSP System Toolbox is used. The recording of the noise is used as *x(k)* input as represented in figure 5.2, the recording from the directional microphone is used as *d(k)* input. The error resulting from the filter is plotted.

LMS

For this algorithm, the dsp.LMSfilter from MATLAB's DSP System Toolbox is used. The same inputs are used as with the RLS filter. The chosen method to calculate filter weights was 'Sign-Data'. This method was chosen as it was the only one which provided an output which clearly differed from the original signal.

Wavelet denoising

An alternative denoising approach was tried for which only the recording from the directional microphone was needed. This approach is wavelet denoising. The idea behind wavelet denoising is that many real-world signals can be represented by a wavelet transform which concentrates the features of a signal in a limited number of wavelet coefficients. The coefficients that can represent the signal are usually large compared to other coefficients, like the ones from noise. With wavelet denoising, the coefficients with a value below a certain threshold are shrunk to remove noise and preserve the important information in the signal. In this approach, first the

wavelet transform is computed, then the coefficients below the threshold are shrunk and lastly the inverse wavelet transform is computed.[62, 63]
Both the discrete wavelet transform (DWT) and the undecimated wavelet transform (UWT) were tried. The DWT downsamples the coefficients during the decomposition, while the UWT does not. This makes the DWT more computationally efficient but the UWT is more accurate and smooth.[64] For both algorithms, the maximum possible decomposition level was used. This value was established by trial and error.

Expectations of noise cancellation approaches

The expectation of the different approaches is that the adaptive filters will perform best in removing the noise from the signal. The subtraction is not expected to provide proper results as the microphones are placed at different locations making the frequency transmission of sound waves in the noise to be different for both recording devices. Moreover, the microphones probably have different polar patterns resulting in a difference in the sound waves which are picked up by the microphones. The Wiener filter is not expected to provide proper results as the sound recordings are non-stationary, while Wiener filtering assumes (quasi-)stationarity [56]. The wavelet denoising is not expected to remove the noise properly as it uses only the original measurement.

5.1.3 Recordings after exercise

Usually, a patient leaves the treadmill immediately after the six minutes of exercise. Six patients were instructed to stay on the treadmill for ten more seconds after exercise and to keep breathing in the microphone. In this way, recordings were made in which the respiration could not be disturbed by the treadmill and footsteps sounds. For each of these patients, the short fragment selected by clinicians was compared to the ten seconds after exercise to investigate the difference in quality of those signals.

5.2 Results

The results of the three adjustments in the signal acquisition will be described separately in the next sections.

5.2.1 Changes in the microphone settings

Table 5.1 shows the different combinations of microphone settings that were tried to improve the signal coming from the microphone. For each combination of settings, the velocity of the treadmill during the recording is indicated and the effect on the respiratory signal is described. The complete analysis of these different setting combinations, including all relevant plots, can be found in appendix E.1. The default of the pre gain was 35 dB and for the post gain this was 0 dB. Both gains were increased and decreased with 10 dB. The combination of a high-pass and low-pass filter (cut-off at 250 respectively 2,000 Hz), echo cancellation and noise cancellation resulted in the signal with the least ambient noise. These settings were thus applied to all new recordings in the clinic. Changing the gain levels and enabling automatic gain control did not improve the signal quality.

Table 5.1: Effects of the different combinations of microphone settings on the respiratory signal.

Settings	**Velocity**	**Effect on signal**
Hp+lp filter	Low*	Clearer respiratory peaks than default
EC	6 km/h	Clearer respiratory peaks than default
EC	8 km/h	Less clear respiratory peaks than default
EC, NC	8 km/h	Respiratory peaks better discernible than default and only EC
Hp+lp filter, EC, NC	Low**	Signal improved compared to only filtering
Hp+lp filter, EC, NC	6 km/h	Quality comparable to EC and NC at 8 km/h
Hp+lp filter, EC, NC, pre +10 dB (so 45 dB)	6 km/h	Quality becomes worse, many irrelevant peaks in signal
Hp+lp filter, EC, NC, pre -10 dB (so 25 dB)	6 km/h	Pattern not altered, amplitude decreased
Hp+lp filter, EC, NC, post +10 dB (so 10 dB)	6 km/h	Clear respiratory peaks, signal sounds windy
Hp+lp filter, EC, NC, post -10 dB (so -10 dB)	6 km/h	Clear respiratory peaks, signal sounds windy
Hp+lp filter, EC, NC, automatic gain control	6 km/h	Signal is deformed, some peaks reach maximum amplitude of 1

Hp+lp filter high-pass filter with 250 Hz cut-off + low-pass filter with 2,000 Hz cut-off
EC echo cancellation, **NC** noise cancellation
* Exact velocity not known, but below 6 km/h

Figure 5.3 shows a recording of a patient (UA) with the new settings of the microphone. In this figure, the inspirations are marked with an *i* and the expirations with an *e*. The respiratory pattern is clearly visible and the peaks of inspirations and expirations can be detected. However, comparable to figure 4.9 with the old settings, some peaks were not correctly identified.

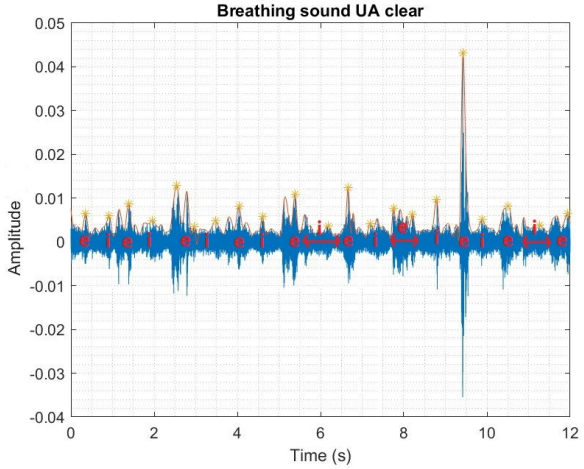


Figure 5.3: Clear respiration with new microphone settings.

Figure 5.4 shows another recording of a patient (UA) with the new microphone settings. This time, the respiration cannot be detected in the time domain plot. There are clear peaks in the figure, but the majority of those peaks belongs to footsteps rather than to respiration.

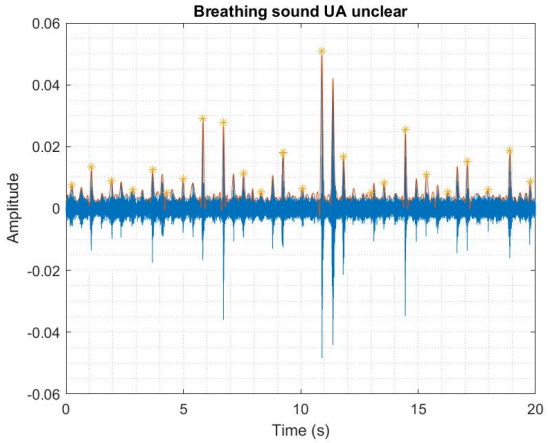


Figure 5.4: Unclear respiration with new microphone settings.

5.2.2 Addition of microphone

The different noise cancellation approaches were applied to the gathered recordings.
Figure 5.5 shows the resulting plots when the noise signal is subtracted from the original signal with different gains. This figure shows that the smaller the gain, the less noisy the signal seems to be.

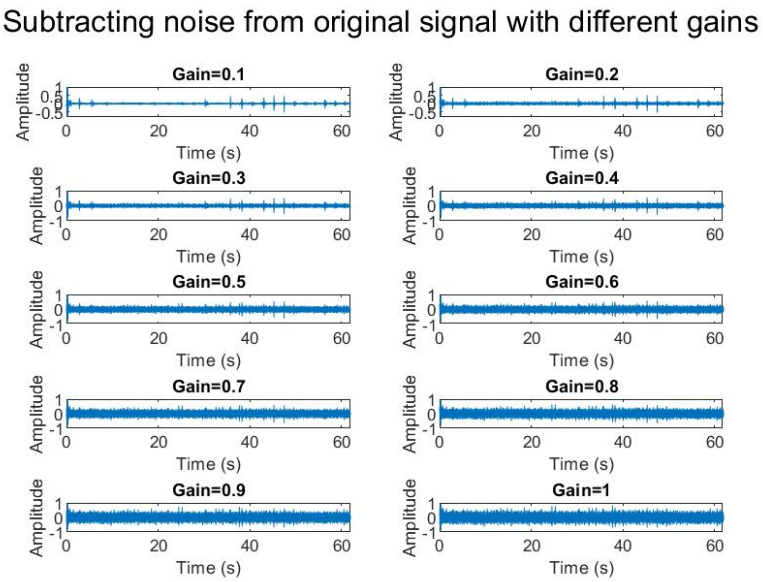


Figure 5.5: Improving signal with subtraction of noise.

The Wiener filter was first applied with a small order of 12. This did not result in an estimate that substantially differed from the original signal. The order was increased to 1,000 and even 10,000. The resulting output again did not clearly differ from the original signal (see figure E.8 in appendix E.3).
The RLS filter was applied with 250 and 500 coefficients. 1,000 coefficients was also tried, but

since the algorithm was still busy after 1.5 hours this process was terminated. The resulting output did not clearly differ from the original signal (see figure E.9 in appendix E.3). It took more than five minutes to perform the noise cancelling with 250 coefficients.

The LMS algorithm was tried with 250 coefficients, resulting in the output signal shown in figure 5.6. It took less than two seconds to compute the output.

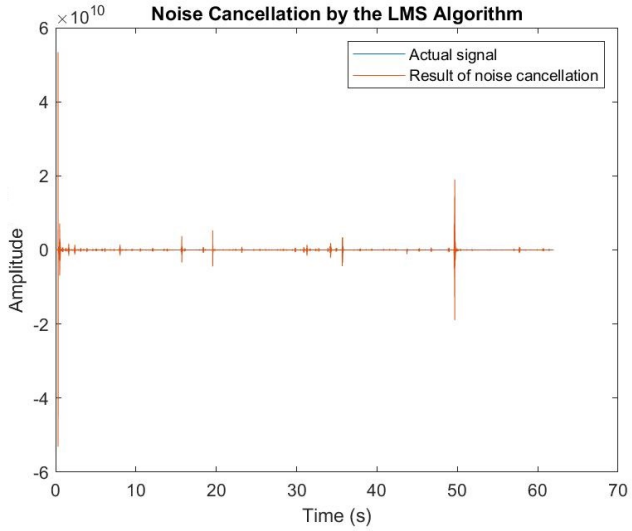


Figure 5.6: Improving signal with LMS filter, 250 coefficients.

Because the amplitude range was very large, the values were normalized and plotted in a subplot below the original signal as is shown in figure 5.7.

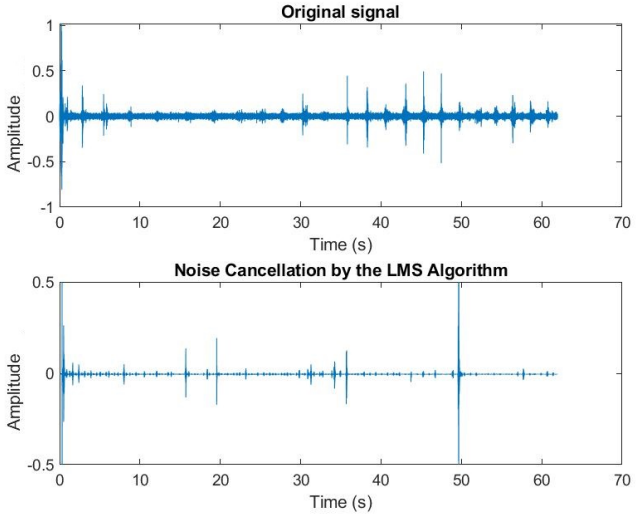


Figure 5.7: Improving signal with LMS filter, original versus normalized improved signal.

This signal looks less noisy than the original signal: there are clear peaks with less noise. The greatest peaks do not occur at the same time instances as in the original signal. When listening to the filtered signal, it sounds very high-pitched and not like a clear breath.

The DWT was applied with its maximum decomposition level of 20. The denoised signal does not clearly differ from the original signal (see figure E.10 in appendix E.3).

Figure 5.8 shows the result of the UWT approach. The main peaks are still visible in the signal while the noise level seems to be lowered. When listening to the denoised signal, mostly breath can be heard rather than the treadmill. However, the steps are still clearly audible and the complete signal sounds 'windy'.

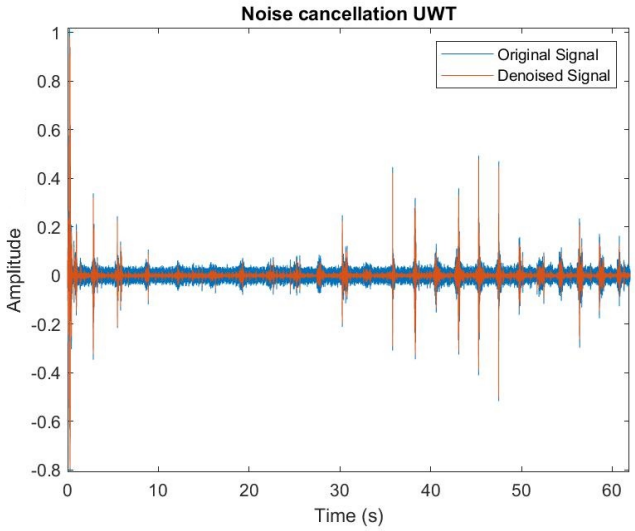


Figure 5.8: Improving signal with wavelet denoising, undecimated.

The UWT denoising was also tried on the recording of which the original breath sound was known (measurement 2 in the protocol). In figure 5.9 these two signals are compared to each other. Here, the denoised signal still shows more than just the breathing sound. There are more clear peaks in the denoised signal than in the clean signal.

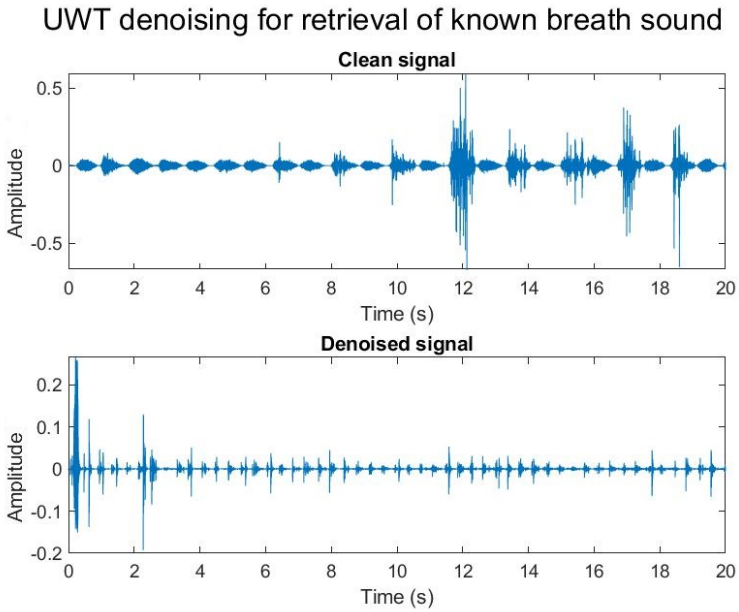
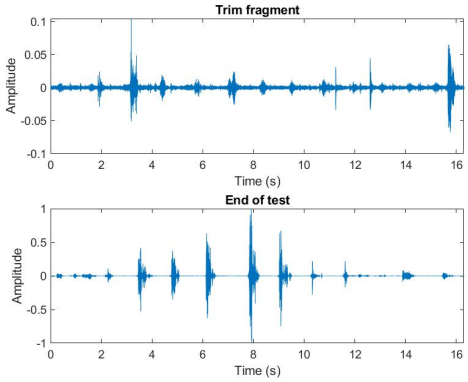


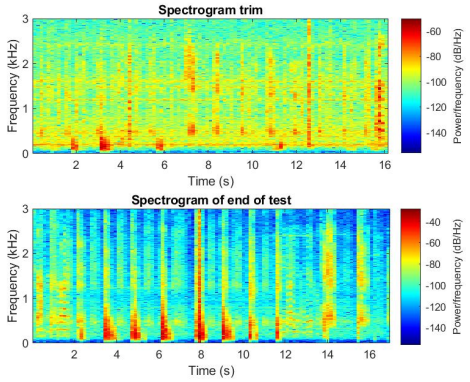
Figure 5.9: Breath recorded without (much) noise (upper) and denoised noisy recording (lower).

5.2.3 Recordings after exercise

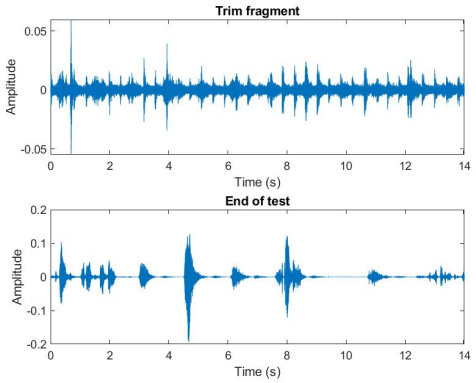
From the six recordings that were made directly after exercise, most recordings still contained speech from either the patient, a parent or the clinician. In some cases, the patient left the treadmill before the ten seconds were over. For two patients it was managed to make a recording of approximately ten seconds without speech. In figure 5.10, the fragment selected by the clinician is compared to the recording directly after exercise in both time and time-frequency domain for both patients.



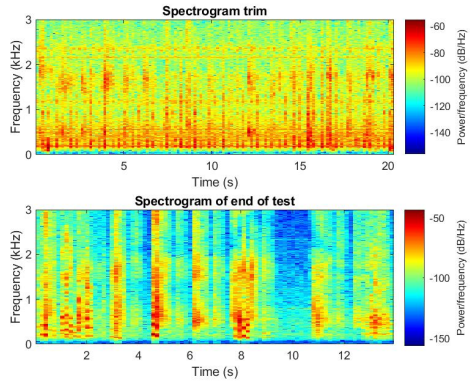
(a) Comparison of time domain plot from fragment selected by clinician with time domain plot from recording directly after exercise.



(b) Comparison of spectrogram from fragment selected by clinician with spectrogram from recording directly after exercise.



(c) Comparison of time domain plot from fragment selected by clinician with time domain plot from recording directly after exercise.



(d) Comparison of spectrogram from fragment selected by clinician with spectrogram from recording directly after exercise.

Figure 5.10: Segment selected by clinician versus recording after exercise from DB patient (a and b) and UA patient (c and d).

For both patients, the time domain plots of the recordings after exercise show a lower noise level than the fragments during exercise. The respiratory patterns can be detected in the recordings after exercise. The respiratory patterns are also better reflected in the spectrograms after exercise than during exercise.

5.3 Discussion

The analysis of the measurements from the three approaches for improving the signal acquisition gives insight in the possibilities for improving the quality of recordings.
The changes in microphone settings showed promising results for the measurements with a healthy individual. However, figure 5.4 proves that the new settings do not enable the rendering of a clear respiratory pattern for every recording. Figure 5.3 did show a clear pattern, but similarly to findings in chapter 4, not all peaks were correctly identified. The new settings thus do not enable an automatic peak detection for the respiratory sounds. More combinations of settings could be tried to discover if these combinations would improve the signal. At this point, there is no indication to try other specific settings and therefore this option is not explored further.

The different methods used for noise reduction showed divergent results: some were promising, while others did not change the signal at all.
The subtraction of the noise showed that the signal was less noisy for smaller gains. This proves that the subtraction method is not suitable for this problem, since a signal should look less noisy when subtracting the noise signal. This is in line with the expectations.
The Wiener filter did not alter the signal. A possible explanation for this is that the assumption of (quasi-)stationary signals is not met for this case where both the desired signal (breath sound) and the noise (steps on treadmill, sound of treadmill) change over time, as was mentioned in the expectation.
Although the RLS filter did not change the signal, the LMS filter resulted into a signal which appeared to contain less noise. However, the peaks did not agree with respiratory peaks in the original signal and listening to sound proved that there was no clear respiration. Therefore, in contrast to the expectations, both the RLS and LMS filter are not capable of removing the noise from the respiratory recordings. A possible explanation for the malfunctioning of both algorithms is that the two microphones are placed at different locations. Because of these different locations, their recording of the noise differs. Ideally, the recording of the desired signal plus noise and the recording of only noise would come from a microphone at one fixed place such that the reflection and deformation of the noise is similar in both recordings.
Wavelet denoising with the UWT resulted in a signal which appeared less noisy in the time domain. When listening to the denoised signal, the respiration was audible but the footsteps were also clear. Figure 5.9 showed that when the UWT was applied to a recording of which the original breath sound was known, the algorithm was not capable of retrieving the respiratory pattern. A possible explanation for this is that the algorithm has no information about the noise source as it only uses the original signal, making it plausible that the steps on the treadmill are regarded as 'desired signal' instead of noise.

The recordings after exercise were hard to acquire because there seems to be a tendency to talk and step away from the treadmill after the ECT. For the two recordings that did not contain speech and had a duration of at least ten seconds, the signal quality was better directly after exercise than during exercise. If a good protocol would be established for keeping the patient on the treadmill and keeping everyone in the room quiet (apart from the patient's respiration), these recordings after exercise could in the future be used to visualize the respiratory pattern of patients. Since only two proper measurements were available, it was not possible to investigate if the recordings could enable the distinction between the different diagnosis categories.

5.4 Conclusions

In this part of the study, the measurement setup is changed in three ways. The new microphone settings did seem to improve the quality of the test signals and therefore these settings will be used in future measurements. The different noise reduction techniques in which the noise recording was used, did not lead to an improved signal and therefore an extra microphone for noise recording will not be incorporated in the measurement setup. At this point, the improved measurement setup still does not seem to facilitate differentiation between the different diagnoses. Figure 4.9 already showed that the automatic peak detection is unattainable, and since the changes in the microphone settings did not facilitate easier peak detection as is shown in figures 5.3 and 5.4, the signal analysis approach will not provide distinguishing parameters. These findings lead to the focus on a machine learning approach in the next part of the study. If more recordings after exercise will be obtained in the future, it is relevant to investigate the possibility to detect DB and UA based on these recordings.

6 MACHINE LEARNING APPROACH

From the previous chapters, it became clear that it is not possible to differentiate between DB and UA with only signal analysis techniques. Therefore in this chapter, the machine learning approach will be discussed in which machine learning algorithms are applied to preprocessed data to discover patterns in the data. The aim of this chapter is to answer the following three research questions:

- To what extent are machine learning algorithms applied to the sound recording data capable of correctly classifying the measurements in the three 'condition' groups?
- Which features of these data sets are most important for this classification?
- Which data mining method performs best in the classification?

6.1 Background

With machine learning, computers are programmed to learn from available input so that this experience can be converted into expertise on the type of data. Computers can use algorithms and statistical models to find patterns in data sets and to act upon them. This approach is especially useful when analyzing data sets which are too large and complex for humans to understand, memorize and analyze.[65]
Machine learning problems can be divided into two categories: supervised and unsupervised learning. In this study, we have to do with supervised learning: each observation with prediction measurements is provided with an accompanying response, being the condition. The goal is to find a model that relates the response to the predictors such that the model can be used for predicting the response based on future observations.[66]
Apart from the division in supervised and unsupervised learning, problems can also be divided in regression and classification problems. Regression problems involve numerical values as response, so these are quantitative variables. With classification problems, the response can take on a value in one of x different classes, so the response is a qualitative variable.[66] In this case we have to do with a classification problem where three different classes are possible: DB, UA and ND.

6.1.1 Resampling data sets

When no test data set is available, which often is the case, the available data set has to be split in a part which will be used to train and fit the model on and a part on which the performance of the model will be tested. There are different approaches for this so called resampling of a data set.[66]
The validation set approach is the most simple approach. With this approach, the data set is randomly split into two subsets of which one will be used as training set and the other as validation set. There are two drawbacks of this method. The first is that the performance of the model can be very variable as it depends on which observations are used for the training set and which for the testing set. Secondly, the low amount of observations used for training the model (only half of the total) might result in a bad performance of the model.[66]

Cross-validation is an approach which overcomes the drawbacks of the validation set approach by averaging over model performance testing with different subsets. Usually k-fold cross-validation is used for this purpose. With this technique, the samples are divided into k groups (called folds) of approximately the same size. The first fold is regarded as validation set and the remaining folds function as training set. The model is fitted and the performance is assessed, after which the same procedure is repeated k times, so that every fold is used as validation set once. The leave-one-out cross-validation is a specific type of k-fold cross-validation where the number of folds equals the number of samples in the data set.[66, 67]
A different approach for creating subsets is bootstrapping. With this approach, the data set is randomly divided in subsets (the bootstrap samples) with replacement, so samples may occur more than once in a subset. A model is fit to every bootstrap sample and predictions can then be made by means of a majority vote: the observation of which the outcome has to be predicted is placed in the class which was predicted by the majority of the models.[66]

6.1.2 Algorithms for supervised classification

Different algorithms for solving supervised classification problems exist. The most widely used algorithms will be explained briefly.

K-Nearest Neighbors

K-nearest neighbors (KNN) is a method which takes into account the K points in the training data set which are closest to the observation that needs to be classified. The observation is then classified in the class which is most common among the K nearest neighbors.[68] When K is chosen to be small, the decision is based on a small number of data samples so the algorithm may find patterns in the data that do not apply to the whole data set. The bias will be low, but the variance very high. When increasing the K, there is less flexibility in the model resulting in a lower variance but higher bias.[66] A weighted KNN takes into account the distance between the observation and a neighbor: the closer the neighbor is to the observation, the more weight is put on this neighbor.[69]

Support Vector Machines

Support vector machines (SVMs) are in default linear binary classifiers, but can be extended to become non-linear multiclass classifiers. The objective of this approach is to establish a hyperplane which separates the data according to their class label.[70] Figure 6.1 shows such a hyperplane for the linear, binary case.

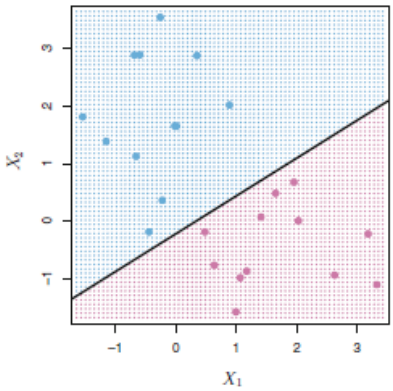


Figure 6.1: SVM for the linear, binary case (adapted from *An Introduction to Statistical Learning* [66])

The hyperplane should be designed such that it has a maximum distance to the data points. The data points that fall within a certain range from the hyperplane are called support vectors.

Because these are relatively close to the hyperplane, they have a greater influence on the position of the hyperplane.[66] Since a linear model often does not suffice for classification problems, the kernel trick is applied which converts the input data space into a higher-dimensional space. This higher-dimensional space enables a non-linear separation of the data points.[71] Figure 6.2 shows a simple example of this kernel trick. In the left part of the figure, so the untransformed data space, it is not possible to create a hyperplane that provides a good separation between the data points of the different classes. After transforming the data space, a hyperplane can easily be established. Popular choices for kernels are the polynomial kernel of degree d (with d a positive

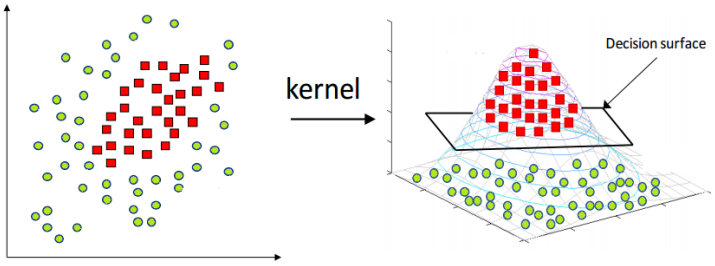


Figure 6.2: Visualization of kernel trick.

integer) and radial kernel. In order to apply the SVMs on multiclass classification problems, the problem is broken down to multiple binary classification problems.[66] SVMs have proven to work well in data sets with few samples but many variables [70].

Discriminant analysis

Discriminant analysis is a method in which, similar to SVM, a data space is divided into regions where every region belongs to a specific class. In this case, the lines between regions are drawn by maximizing the distance between the means of the classes while minimizing the scatter within the classes. To do so, it is assumed that the observations within each class come from a (multivariate) Gaussian distribution with a class-specific mean. For linear discriminant analysis (LDA), it is assumed that all classes have a common variance. Quadratic Discriminant Analysis (QDA) is a more flexible classifier which assumes that every class has its own covariance matrix.[66]

Naïve Bayes classifiers

Naïve Bayes classifiers use probability theory for the classification of data samples. The prior probability is the probability before any features are taken into account, so it is the number of observations coupled to a certain class divided by the total number of observations. The conditional probability of a class is the probability given a certain value for a feature. Both the prior and conditional probabilities for all classes and features can be computed using a training set. Bayes rule, in which both the prior and conditional probabilities are combined in one formula, can then be applied to compute the probability of each classification given the observation. The observation is assigned to the class with the highest probability. The term naïve refers to the assumption in the approach that the effect of the value of one feature on the probability is independent of the values of other features.[67] Different types of naïve Bayes classifiers are known. The Gaussian naïve Bayes classifier assumes a normal distribution of the values associated with each class [72]. The kernel naïve Bayes classifier makes a non-parametric estimation of the distribution of values associated with the classes and can be used when the distributions are far from normal [73].

Logistic regression

Logistic regression is a statistical approach that uses a logistic function to approximate the relationship between features and the classes. The idea behind this approach is to find coefficients with which the probability of a certain outcome given the input can be modelled according to the formula:

$$p(X) = \frac{e^{\beta_0+\beta_1 X}}{1 + e^{\beta_0+\beta_2 X}}$$

in which β_0, β_1 and β_2 denote the coefficients. The maximum likelihood estimator is used to find these coefficients and thus fit the logistic model. Usually, logistic regression is only used for binary classification problems, but it can be extended to be applied to multiclass classification problems by breaking the problem down in multiple binary classification problems.[66]

Decision trees

Decision trees assign a class to observations by testing them on specific features in a certain order. With these tests, the predictor space is segmented into a number of regions. A decision tree contains decision nodes which specify the so called attribute on which an observation is tested. After every decision node, the data set is split in two subsets. So after each split, two branches go from the decision node towards two different decision nodes. At the end, a new observation is assigned to the class that occurs most often among the training observations in the region to which it belongs. In order to grow a good tree, the quality of possible splits should be assessed and the split with the best outcome measure at that particular step is selected. The Gini index and entropy are common measures for assessing the quality of a split. Decision trees are often pruned after construction, meaning that the size of the tree is reduced by removing regions which are non-critical for classification. In this way, computation time is reduced but most important, overfitting of the model to the training data is prevented.[66]

Decision trees show good results for highly non-linear and complex relationships between the features and the classes. On the other hand, decision trees often show lower accuracy than previously mentioned approaches and decision trees can be non-robust, meaning that a small change in the training data can result in large changes in the tree. To overcome these drawbacks, methods are developed which aggregate many trees to improve the performance. These methods are bagging, random forests and boosting.[66]

Bagging, random forests and boosting

The main goal of bagging is to reduce variance of a classifier, in this case a decision tree classifier. The training set is bootstrapped into different training sets and a decision tree is constructed for every new training set. These trees are not pruned which results in a high variance for every tree, but a low bias since the tree is fitted very well for the specific training set. As averaging reduces variance, 'averaging' the classification outcomes leads to a classification which suffers less from variance and has low bias since the individual trees had low bias. To attain this averaging, the classification of each individual tree for a single observation is recorded and a majority vote is taken.[74, 75]
The random forest algorithm also makes use of averaging multiple trees, but in this approach another step is taken to guarantee the individual trees do not look alike. Every time a new split is added to the tree, only a predefined amount of features are considered as candidates for this split. The candidate features are randomly selected from the entire collection of features. This technique prevents that each tree starts in the same way because there is one very distinguishing feature for example. The result is that there is less correlation between the individual trees, making the average of the trees less variable.[76]

Unlike bagging and random forest, the boosting method does not use bootstrapped training data but grows different trees sequentially. The first tree is fitted to the original training data. The next tree is fitted to the residuals from the previous tree. The original tree is then updated by adding a shrunken version of the second tree to the original tree. This procedure is iterated, so at every step a new tree is fitted to the residuals of the previous model and a shrunken version of the new tree is added to the previous model.[74] The sequential construction method enables the construction of smaller and hence more interpretable trees. The model construction might take longer due to the sequential growth, but the reduced size of the final tree reduces computation time when the model has been established.[66]

6.1.3 Expectation of algorithms

The signals which will be used for the machine learning problem are complex signals with much inter- and intra-individual variability. Because of this complexity, it is expected that non-parametric algorithms perform best on the data. Moreover, due to this complexity the observations are not assumed to have Gaussian distributions and class-specific means (not verified as the data first needs to be prepared). As the LDA classifier assumes Gaussian distributions and linearity, this classifier is not expected to perform well. The QDA classifier is expected to perform a little better, but still not well since it also assumes Gaussian distributions. For this reason, the Gaussian naïve Bayes classifier is also not expected to perform well. However, as the kernel naïve Bayes classifier makes a non-parametric estimation of the distribution of values, this classifier might perform well. KNN, SVM with kernel trick and decision trees are also non-parametric classifiers so these are also expected to perform well on the available data. As was mentioned before, SVMs have proven to work well in data sets with few samples but many variables, so especially this approach is expected to show good performance on the available data.
In general, ensemble methods are expected to perform well since they are usually improved versions of 'simpler' methods. However, since these classifiers use subsets of a data set and the available data set is relatively small, the subsets on which they train become very small and the averaged performance might therefore not be good.
Although some classifiers would in theory perform well on the data, the overall expectation is that the performance of machine learning algorithms is not good for this data. The results from chapters 4 and 5 did not provide clear directions for features which could make the distinction between DB, UA and ND. However, as was mentioned in the beginning of this chapter, computers can find patterns in data sets which are too complex for humans and therefore it is relevant to explore this machine learning approach.

6.2 Methods

Sound recordings were gathered during scheduled ECTs in MST. The majority of the recordings was made before the new microphone settings were introduced. To maximize the size of the data set but to overcome problems due to usage of recordings with different settings, the recordings with the old microphone settings were used in machine learning. In total, 43 recordings with a duration of approximately six minutes each were available for analysis.
The main disturbances in the audio signals come from the treadmill: the sound of the treadmill and the sound of footsteps on the treadmill. It is expected that the more the microphone is directed at the treadmill, the more the signal is disturbed because more sound from the treadmill is captured by the microphone. The microphone is always directed at the mouth of the child. Therefore it is expected that in younger and particularly smaller children, the quality of the respiratory sound is worse than in taller children because the distance between the mouth and treadmill is smaller. To test this hypothesis, short fragments of ten seconds were selected from each of the 43 recordings which best represented the symptoms of the patient and did not

contain speech. These short fragments were listened to in Sony Vegas Pro and every recording was labelled as having either a good, OK or bad quality. The quality represents how well the respiratory pattern can be detected in the recording. The accompanying heights and ages of the patients were derived from the electronic health records and it was statistically tested if there were differences in age and/or height across the different 'quality' groups. Afterwards, only the recordings with 'good' and 'OK' quality (28 recordings) were used for machine learning purposes.

The amount of samples required for a proper machine learning approach depends both on the amount of classes in which the samples can be classified and on the amount of features taken into account. Because of the small amount of samples in the data set, different techniques were used for handling the data which included enlarging the data set in different ways. These techniques will be described in the next section.

6.2.1 Preparation of data

Before applying the different techniques for handling the data, all ten-seconds and six-minutes recordings were high-pass and low-pass filtered and resampled in the same way as the recordings in chapter 4. Next, the data was prepared in the following ways:

1. The ten-seconds fragments from 28 patients were selected. Eleven recordings were from DB patients, eleven from UA patients and six from ND patients.
   (a) For every category, two recordings were excluded from training to be able to test the model on. Because this data set contained few samples, only two features were used for the classification: the mean and standard deviation of the signal.
   (b) The training set was balanced by using as many samples for DB and UA as were available for ND and again the mean and standard deviation were used as features.
2. 75% overlapping windows of one second were created
   (a) from the ten-seconds fragments, for five fragments per category. From the resulting segments, mostly statistical features were determined: mean, standard deviation, 0.25-0.5-0.75 quantiles and the median frequency of the signal. The model was tested on the overlapping windows of the remaining fragments.
   (b) from the entire six-minutes recordings. The same features were used as with the ten-seconds fragments.
3. Per class, four ten-seconds fragments were split up in segments of 2.5 seconds and these shorter segments were randomly recombined to 500 new ten-seconds fragments. These new fragments were used as training data. The remaining intact ten-second fragments were used as test data.
   (a) The training was first performed with the following features: amount of peaks with a minimal distance of 0.4 seconds, the average peak prominence, the average peak interval, the median frequency of the signal and the median frequency of the signal envelope.
   (b) Next, the training was performed with the following statistical features: mean, standard deviation, 0.25-0.5-0.75 quantiles and the median frequency of the signal.
4. Every entire six-minutes recording was split up in 36 ten-seconds segments. These segments were used in the following ways:

(a) 144 segments, which was 2/3th of the amount of ND samples, were used as training data for each class. These segments were randomly selected to make sure that they came from different time instances in the recordings. The model was tested on the remaining samples. The following features were used: amount of peaks with a minimal distance of 0.4 seconds, the average peak prominence, the average peak interval, the median frequency of the signal and the median frequency of the signal envelope.

(b) For each class, the segments belonging to the last 1.5 minute of the recording were used as training data. The reason for this is that it is expected that in the last one or two minutes of the ECT, the symptoms are most prominent. The model was tested on the segments from the remaining minutes of the recordings. The following features were used: amount of peaks with a minimal distance of 0.4 seconds, the average peak prominence, the average peak interval, the median frequency of the signal and the median frequency of the signal envelope. For every test measurement, it was determined what percentage of the segments was correctly classified.

(c) For each class, the segments belonging to the last 1.5 minute of four recordings (2/3th of ND recordings) were used as training data. The following features were used: amount of peaks with a minimal distance of 0.4 seconds, the average peak prominence, the average peak interval, the median frequency of the signal and the median frequency of the signal envelope. The model was tested on:

   i. the entire recordings that were not used for testing. For a few test measurements, it was determined how the classification differed over time by displaying the segments in order of time and indicating per segment whether it was correctly classified or not.

   ii. the remaining segments belonging to the last 1.5 minutes of the recordings. In addition, it was verified for every segment whether it contained speech since speech usually results in peaks which might affect the classification capability based on peak information.

(d) For each class, the segments belonging to the last 1.5 minute of four recordings (2/3th of ND recordings) were used as training data. Because of the possible speech issue mentioned above, here statistical features were used: mean, standard deviation, 0.25-0.5-0.75 quantiles, standard deviation in the mean of 75% overlapping windows, standard deviation in the standard deviation of 75% overlapping windows. The model was tested on the remaining segments belonging to the last 1.5 minutes of the recordings.

6.2.2 Implementation of algorithms

After the data was prepared in MATLAB, The Classification Learner app that comes with MATLAB's Statistics and Machine Learning Toolbox was used to construct the different models. This app includes the following classifiers: decision trees, discriminant analysis, naïve Bayes classifiers, SVMs, KNN classifiers and ensemble classifiers. The ensemble classifiers are boosted trees, bagged trees (with random forest algorithm), RUSBoosted trees (algorithm for skewed, imbalanced data), subspace discriminant analysis and subspace KNN. The subspace classifiers combine the outcomes of their algorithms trained on different subsets of the data set. Logistic regression, which was described in the background, can only be used for binary classification in the app. Before implementing the different algorithms, the resampling method can be set to protect against overfitting. The default for this is the method which is determined to be the best approach based on the size of the data set. For the different data sets in this study, this was the 5-fold cross-validation.

6.2.3 Model selection and testing on unseen data

After training all classifiers, their performance was studied in terms of the accuracy and the confusion matrix. In principle, the model with the highest accuracy was selected as best model. However, in some cases the highest accuracy was reached by a very imbalanced classification (finding derived from the confusion matrix). In these cases, the next best model which showed a more balanced classification was selected.
Per method described in section 6.2.1, the best performing model was exported to the MATLAB workspace. From the data that was set apart to test the models on, the required features were determined and the model was used to classify these samples in a diagnosis category. In each case, it was counted how many data points were correctly classified. From these results, it was determined which method performed best on this data set.

6.3 Results

Table F.1 in appendix F.1 contains the baseline characteristics of the 43 fragments with the original microphone settings. Eighteen fragments received the quality label 'good', nine received the label 'OK' and thirteen received the label 'bad'. One recording fell between the 'good' and 'OK' categories and two fell between the 'OK' and 'bad' categories. These were excluded from the statistical test. The age and height values from the different categories were normally distributed (for proof, see appendix F.1) and thus a one-way ANOVA test was performed to compare the means across the quality groups. The means of the age and height do not significantly differ across the quality groups (P=0.842 and P=0.949 respectively).
The thirteen 'bad' and two 'OK/bad' fragments were excluded from use in machine learning.

Table 6.1 shows the results of training the different data sets with the Classification Learner. Per data preparation method described in section 6.2.1, the best performing classifier is indicated along with the training accuracy. The confusion matrices belonging to the resulting models can be found in appendix F.2.

Table 6.1: Results of model training. For every preparation method, the best performing model with its accuracy is indicated.

Data preparation method	**Number of training samples**	**Features**	**Best performing algorithm**	**Training accuracy**
1a	DB: 9, UA: 8, ND: 4	Mean, sd	QDA	57.1%
1b	DB, UA, ND: 4	Mean, sd	Cubic SVM, Fine KNN	33.3%
2a	DB, UA, ND: 180	Mean, sd, quantiles, med. freq.	Weighted KNN	82.6%
2b	DB: 6917, UA: 7062, ND: 7176*	Mean, sd, quantiles, med. freq.	Bagged trees	60.9%
3a	DB, UA, ND: 500	Amount of peaks with a minimal distance of 0.4 seconds, average peak prominence, average peak interval, med. freq. of signal, med. freq. of signal envelope	Gaussian SVM	82.3%
3b	DB, UA, ND: 500	Mean, sd, quantiles, med. freq.	Bagged trees	94.4%
4a	DB, UA, ND: 144	Amount of peaks with a minimal distance of 0.4 seconds, average peak prominence, average peak interval, med. freq. of signal, med. freq. of signal envelope	Kernel Naïve Bayes	43.1%
4b	DB, UA, ND: 54	Amount of peaks with a minimal distance of 0.4 seconds, average peak prominence, average peak interval, med. freq. of signal, med. freq. of signal envelope	Weighted KNN	58.6%
4c	DB, UA, ND: 36	Amount of peaks with a minimal distance of 0.4 seconds, average peak prominence, average peak interval, med. freq. of signal, med. freq. of signal envelope	KNN	68.0%
4d	DB, UA, ND: 36	Mean, sd, quantiles, med. freq., std. in mean of overlapping windows, std. in std. of overlapping windows	Coarse tree	68.0%

sd=standard deviation, med. freq.=median frequency
numbers differed because some recordings were slightly shorter than six minutes, resulting in less windows

Table 6.2 shows the classification accuracy when the models are applied to unseen data. In appendix F.3 the confusion tables, which visualize the complete performance of each model on the unseen data, can be found. The model resulting from method 1b is not used for testing because the training accuracy was extremely low. For method 4b, the testing accuracy is determined per recording. The results of this can be found in table 6.3.

Table 6.2: Result of testing models on unseen data.

Testing method	Testing accuracy
1a	66.7%
1b	Not tested due to extremely low training accuracy
2a	22.4%
2b	35.5%
3a	43.8%
3b	37.5%
4a	46.9%
4b	Indicated per recording
4ci	32.6%
4cii	28.2%
4d	25.9%

Table 6.3: Results of testing the model based on all segments from the last 1.5 minutes on the segments from the remaining minutes of the ECTs. Per recording, it is indicated what percentage of segments is classified in which category. The correct category is highlighted.

	DB	UA	ND
DB1	33%	22%	44%
DB2	11%	44%	44%
DB3	59%	26%	15%
DB4	15%	35%	50%
DB5	5%	33%	62%
DB6	19%	19%	63%
DB7	11%	56%	33%
DB8	44%	26%	30%
DB9	7%	22%	70%
DB10	30%	22%	48%
DB11	63%	26%	11%
UA1	17%	17%	67%
UA2	19%	48%	33%
UA3	22%	41%	37%
UA4	44%	15%	41%
UA5	26%	44%	30%
UA6	30%	37%	33%
UA7	20%	40%	40%
UA8	22%	15%	63%
UA9	23%	42%	35%
UA10	19%	44%	37%
ND1	7%	44%	48%
ND2	33%	7%	59%
ND3	59%	26%	15%
ND4	23%	22%	56%
ND5	19%	41%	41%
ND6	15%	48%	37%

With the test described at 4(c)i, eight recordings were used to determine how the classification correctness differed over time. For each ten-seconds segment, it was determined whether it was classified correctly or incorrectly. In total, there were 36 segments per recording. Per time instance, it was determined in how many of the eight recordings the segment belonging to that time instance was classified correctly. Figure 6.3 shows this sum of correctly classified segments over time.

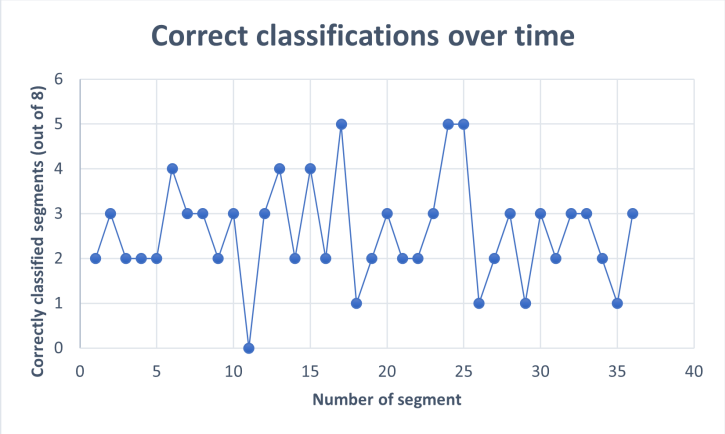


Figure 6.3: Amount of correctly classified segments per time instance. The numbers on the x-axis show the segments over time, so segment 1 indicates the first 10 seconds of the recording.

In total, there were 243 segments belonging to the last 1.5 minutes of the ECTs. 148 (61%) of these segments contained speech.

6.4 Discussion

The ANOVA test performed on the 43 fragments to compare the age and height across different quality categories, proved that there was no significant difference in age and height across the quality categories. A relation between the age or height of the patient and the quality of the signal is thus not established. This does not agree with the expectation that a smaller value for the age and especially the height would lead to lower quality signals. The reliability of this test is disputable, since putting a quality label on the fragments was based on a subjective judgement by one person.

In order to create an algorithm for the classification of DB, UA and ND, the available data was handled in different ways. Table 6.1 shows that different preparation methods resulted into divergent training accuracies. The best training results came from the recombination method with statistical features. However, the testing accuracy belonging to this model was only 37.5% which is very poor for a distinction between three classes. The best testing accuracy (66.7%) came from model 1a in which 22 fragments were used for training and 6 for testing. This accuracy is still not very high and it is assumed that the higher testing performance of this model in comparison to other models is coincidence because of the low amount of samples used for both training and testing. This assumption is enhanced by the training accuracy of only 57.1% of this model.
In most cases, the best performing algorithms were algorithms of which it was expected that they would perform well on the data but the QDA and Gaussian SVM were unexpected since they assume Gaussian distributions. However, the expected loss of performance due to incorrect distribution assumptions pales into insignificance compared to the overall loss of performance due to the small and low quality data set. Table 6.3 shows that it differs a lot per recording how many segments are correctly classified according to the model built with method 4b. These variations are seen in all three diagnosis classes.
Because the symptoms during an ECT are most prominent during the last 1-2 minutes of exercise, it was expected that there would be more correct classifications in the segments from these last minutes than from the earlier minutes. However, when looking at figure 6.3, an increase in correctly classified segments over time is not found. The fact that a large part of the segments

from the last 1.5 minutes contain speech, could be an explanation for this.

In the different data preparation methods, only a few (combinations of) features are tried for classifying the samples. In theory, many more features can be thought of to use in the models. However, there are a few reasons why only the features summarized in table 6.1 are included. Firstly, no good starting points for features resulted from the previous chapters. There was too much inter- and intra-individual variation in the shapes of respiratory peaks and in chapter 4 it was found that the frequency information did not significantly differ across the different categories. Therefore creating more features would come down to trial and error and an ending point for this could not be indicated. Secondly, the available data of 28 patients is not enough to create a good classification algorithm and to test it on unseen data. To overcome this problem, the different data preparation methods were applied. However, these methods imply that more data with less typical sounds (from earlier during the ECTs) was included. When only taking the segments in the last minutes of exercise, it turned out that 61% contained speech which results in high peaks that do not correspond with the respiration. If these speech containing segments were left out, the amount of data would again be too small. Lastly, the data contains too many disturbances apart from the speech. As was discovered in the previous chapters, the sound of the treadmill and footsteps on the treadmill often drown out the respiration sound. These disturbing sounds in general do not have a relation with the diagnosis category of a fragment and thus derange the classification based on respiratory sound.

6.5 Conclusions

The available data does not enable the creation of an algorithm which can make the distinction between DB, UA and ND. The combinations of training and testing accuracies were too low to choose a data preparation method which showed an acceptable performance and could thus be used for the classification of the sound recordings in the diagnosis categories. Since no model is selected, it is not determined which features are most important for classification and which data mining model performs best. Before further machine learning can be applied to sound recordings during ECTs, the amount of samples must be increased and the quality of the data must be improved.

7 DISCUSSION

In this study, we investigated the extent to which sound recordings during an ECT can be used as an objective measure to detect DB and UA in children when compared to the current ECT protocol involving spirometry.
Our main finding was that breath sound recordings may contain characteristic sounds but that the quality of the signals is too low at this moment for proper analysis. The measurement setup needs to be improved to acquire higher quality signals and the existing data can be used to label typical sounds. Once better signals are acquired, the next steps can be taken to identify the typical characteristics in quality data.

This chapter will discuss the research findings from the different chapters with respect to each other and to related work. In addition, strengths and limitations of the study will be discussed and recommendations for future research will be provided.

7.1 Interpretation of results

7.1.1 Characteristic respiratory sounds and patterns

From our literature search, RSA seemed promising for the detection of DB and UA. Two main differences in symptoms were established: a difference in the nature of adventitious sounds and a difference in the moment at which symptoms occur. Wheeze and inspiratory stridor were not detected in the available sound recordings. A possible explanation for this is that not all children with complaints produce these adventitious sounds. Literature findings suggest that not all children with UA wheeze [77] and inspiratory stridor is mostly heard only with extrathoracic DB, not with thoracic DB [1]. In addition, there is a lack of information regarding typical frequencies and durations of the adventitious sounds in mouth breathing [11]. Establishing reference values for the adventitious sounds could be achieved in future work by labelling and analyzing individual cases where the sounds occur. This information can then be used to detect adventitious sounds in new recordings, but should not be used as only criterion for the detection of DB and UA because of the described absence of these sounds in many cases.

The results of the literature search were combined with input from clinicians to obtain a view on differentiating characteristics supported both by theory and clinical experiences. The findings from literature research were confirmed by the tests with clinicians, but other differentiating characteristics were also established. The most important additional characteristic was the loudness of the respiratory sound: a soft breathing sound was often mentioned when diagnosing UA while a loud breathing sound was mentioned when diagnosing DB. Although this difference in loudness of respiration between the respiratory conditions is not found in literature, Pramono et al. describe that stridor is normally louder than wheeze [11]. The adventitious sounds related to the conditions thus do exhibit this difference in loudness.
Although a collection of gathered characteristics contributes to the potential of differentiation, the established characteristics were often not well-defined. In our literature search only one article [46] defined the range of frequencies found in mouth breathing and the duration and pitch

of wheeze and stridor differed amongst different articles [11, 44]. Moreover, the results from the test with clinicians showed overlap in symptoms: a high respiratory rate and hyperinflation were mentioned more than once as characteristic sounds when diagnosing DB, but sometimes these characteristics were also mentioned for UA. Lastly, the mentioned characteristics in the test with clinicians were subjective sometimes. For example, the term 'abnormal' was used, but it is not defined when a pattern is normal or abnormal. The lack of clarity regarding the nature of characteristic sounds in both literature and amongst clinicians complicates the differentiation between the respiratory conditions. A first step for future research would thus be to establish differentiating characteristics with unambiguous features.

7.1.2 Performance of clinicians

The tests with clinicians revealed that the assessment of a patient's condition based solely on sound recordings is very hard: only a third of the recordings was correctly classified. Moreover, there was a considerable inter-individual variability between the assessors and the two assessment rounds resulted in an average intra-individual variability of 37%. Findings from literature suggest that discrepancies between independent opinions among physicians are common in many clinical domains [78], but the magnitude of the variation is very application dependent. Burger et al. have researched the differences between first and second opinions in general internal medicine at the University Medical Center in Utrecht and found that a new diagnosis was established in 13% of the patients [79]. Thavendiranathan et al. report an interobserver variability for ejection fraction estimation from echocardiograms of 14% [80]. In contrast, Haritoglou et al. conclude that their interobserver variability for tumour height measurements was low when using an appropriate scanning method [81]. These findings and our considerable inter-individual variability stress the need for a more reliable and objective tool to assess respiratory conditions.

Addition of camera footage to these sound recordings led to a clear increase in correct classifications and revealed that a combination of respiratory sounds and the posture or stature of the patient is often used to make a diagnosis. Despite this increase, the percentage of correct classifications was only 51% which again highlights the necessity of an objective diagnostic tool. The fact that clinicians combine different sources of information to come to a diagnosis, proves that differentiation based on sound recordings alone is a more complex task compared to when more information sources are available. This finding lowers the expected accuracy of an automatic classification algorithm. In previous studies, successful classification of respiratory sounds is performed [82, 83, 84, 85]. However, these studies aim to detect specific sounds (especially wheeze) instead of detecting a respiratory condition and we have seen that such adventitious sounds are not presented by all children with respiratory conditions.

7.1.3 Quality and differentiating potential of sound recordings

The analysis of test measurements showed that with the original measurement setup, differentiation in respiratory conditions based on sound recordings is not possible. The main cause for this is the presence of disturbances in the signals. The measurements with a healthy individual showed that good positioning of the microphone is necessary for recognizing a respiratory pattern in the time domain. When the microphone was properly directed and there were no disturbances from the treadmill and footsteps, it was possible to obtain information on the respiratory rate and the moments at which inspirations and expirations occurred. However, this was not possible when disturbances were present in the signal. The attempt to remove a pattern or frequency content of the treadmill from these signals with disturbances was unsuccessful. The treadmill recordings alone did not show a repeating pattern and in the frequency domain, these recordings showed various frequency peaks. Many of these peaks occurred in the range of

frequencies of mouth breathing, so filtering all of these frequencies would remove relevant information in the signal. It is questionable whether the attempted method to remove noise is the optimal method, since no other studies were found in which noise removal from such machinery has been described.
If the treadmill sound cannot be removed from these test measurements with constant velocity, it can certainly not be removed from recordings during an ECT since in these recordings, the velocity of the treadmill often varies throughout the recording. A possible method to purely filter the treadmill sound in future recordings, would be to make recordings for a wide variety of treadmill velocities such that these patterns can be subtracted from the recording made during an ECT. However, these recordings should then be made for each patient before or after the ECT as the exact positioning of the microphone influences the transmission of sound waves [86]. This would be a time-consuming task and therefore it is better to focus on reducing noise before the sound is recorded, which will be described in detail in section 7.3.2.

The measurements from the outpatient clinic all contained disturbances because they were recorded during the ECTs. There was a lot of variation in the quality of these signals: in some signals, the respiratory pattern could be detected while in other signals, peaks were only caused by disturbances. Significant differences in frequency content across the diagnosis categories were not found. This was expected, as noise is often dominant and adventitious sounds were not detected in the recordings. Moreover, this absence of significant differences complies with the finding that frequencies in lung and tracheal sounds depend on gender and height [87, 88], making it likely that frequencies in mouth breathing also depend on these factors. This dependence results in expected differences in mouth breathing frequencies within a single diagnosis category.
The prolonged inspirium or expirium, which were both described by clinicians and resulted from literature search [2, 17] to be relevant for the differentiation between DB and UA, was studied in the test measurements by analyzing the ratio between the duration of inspirations and the duration of expirations. In the few samples that were studied, the ratio was indeed different for these two respiratory conditions. However, it was not evident where the ratio of the 'no respiratory condition' should be placed and the amount of samples was too low to determine statistically significant differences. Automatic peak detection could make it possible to determine the ratio for more samples. Unfortunately, a general peak detection algorithm could not be established due to the inter-individual variability in signals. Developing an adaptive peak detection algorithm was unsuccessful because in addition to the inter-individual variability, there was a considerable intra-individual variability in e.g. breath rate and amplitude of the signals. Even for relatively good quality signals the peak detection algorithm did not detect all peaks from inspirations and expirations correctly. This was seen in figures 4.10 and 5.3, where no adequate threshold values were established due to the great variations in breath rate and amplitude throughout the recordings.

7.1.4 Improvement of measurement setup

The improvements in the measurement setup did not facilitate differentiation between the diagnosis categories based on sound recordings. The altered microphone settings showed improvement in quality of the test measurements, but not when these settings were applied during scheduled ECTs. The adaptive noise cancelling approaches that were applied to improve the signal quality resulted in changes in the time domain plot of the signals but it was not beneficial to the quality of the signals. This result is not in line with the expectations: various studies were found in literature which have proven the adaptive RLS and LMS algorithms to be successful for noise cancellation in audio signals [89, 90, 91] and these adaptive filters are also applied to physiological signals [92, 93]. A reason for the malfunctioning in this typical application, could be the placement of the microphones: the noise in the recording with the desired signal

was recorded from a different angle than the noise recorded in the pure noise recording. This results in a different signal transmission [86] which makes it plausible that the algorithms did not have a correct view on the nature of the noise. The wavelet denoising approach showed promising results for the first test measurement, but in the end this algorithm was also not capable of retrieving a breathing pattern from measurements. This is probably due to the fact that wavelet denoising does not use a measurement of only the noise, so it has no information on which components of the signal should be regarded as noise and which components belong to the desired signal. Besides using the wavelet denoising on its own to remove noise from the signal, wavelet denoising can also be applied as preprocessing stage for blind source separation. This is not tried in this study, but a study on atrial fibrillation analysis proves the potential of this approach by showing an improved ICA performance with wavelet denoising as preprocessing step [94].
Two of the recordings that were made right after the ECT showed a clear respiratory pattern in both the time and time-frequency domain. From these signals, the respiratory rate and the changes in dominant frequencies over time could be derived. As only two of these recordings were available, it was not possible to investigate if these signals had a differentiating capability. It appeared that acquiring recordings after exercise without much speech and in which the child remained on the treadmill for approximately ten seconds, was challenging.

7.1.5 Machine learning

Machine learning algorithms applied to the data set of sound recordings did not result in a usable classification algorithm and it was not determined which features and data mining method performed best. The classification algorithms that were established overall had a low testing accuracy. In some cases, especially when recombining signals, the training accuracy was acceptable but when testing these resulting models on unseen data, the accuracy was below 50%. Acceptable values for performance metrics depend on the clinical context, as the importance of proper treatment can be in a range from desirable to critical [95, 96]. For this application, acceptable values are not established but with common sense it can be reasoned that an accuracy below 50% is unwanted since this would lead to more than half of the pediatric respiratory patients receiving inappropriate treatment. Because the quality of the signals was low and the size of the available data set was small for machine learning, it was expected beforehand that the machine learning approach would not lead to good classification algorithms. However, it was expected that the models using only data from the last 1.5 minutes of the ECTs would perform better than the other models because respiratory symptoms are most prominent in the last 1-2 minutes of exercise [17, 19]. This was not the case.
No classification model was selected as best performing model because of the low training and testing accuracies. Therefore the performance of the clinicians in assessing the sound recordings could not be compared to the performance of a classification algorithm. However, it can be stated that at this point the performance is similar as the accuracy of the clinicians was 38% and the testing accuracies of the different models (shown in table 6.2) were in the same region. The performance of clinicians when assessing video recordings is better than both audio assessment by clinicians and by machine learning, although there is a considerable inter-individual variability.

7.2 Strengths and limitations

This study has followed different approaches to investigate the potential of sound recordings for detection of DB and UA thoroughly. As sound recordings during ECTs are not analyzed in previous studies, this study covers a new area of research contributing to the existing knowledge on objective differentiation between respiratory conditions. The main limitation of this research is the limited amount of good-quality sound recordings. The presence of disturbances greatly impacts various aspects of the study, which will be described below. In addition, strengths and limitations that do not relate to the quality of the data set will be discussed.

7.2.1 Differentiation between respiratory conditions

As DB is an under-diagnosed condition but requires tailored treatment, this study on objective detection of both DB and UA contributes to future improvements in patient care. In our literature search, a distinction was found between thoracic DB and extra thoracic DB and different DB patterns were explained. Since the different patterns are placed under the same heading of DB in the outpatient clinic, there was no differentiation between these patterns in this study. This makes it difficult to establish characteristics which can make the distinction between DB, UA and ND because DB can represent a variety of patterns so there is not one characteristic pattern for DB. When more sound recordings of DB patients are gathered in the future, it is desirable to label the patterns to create a database of the different patterns in respiratory sound. In addition, if differentiating characteristics would have been established for the three diagnosis categories, the different forms of DB could not have been distinguished in clinic because they were placed under the same heading. In this situation, it would therefore be critical that a physical therapist can determine the exact underlying problem for the respiratory complaints to assure tailored treatment.
In MST, three diagnosis categories were distinguished, apart from DB, UA and ND: EILO, a mix of conditions and controlled asthma. Moreover, in the UA category there was a subgroup with 'no significant decrease in FEV_1 at t=1'. Because the amount of recordings for these categories and the subgroup was limited, these categories were not taken into account although it is relevant to make a distinction between all categories.

7.2.2 Dynamics of symptoms

The symptoms of DB evolve in parallel with increasing ventilatory demands due to increase of exercise intensity. Our literature search showed that the symptoms of UA generally peak after exercise but with breakthrough-EIB, symptoms already occur during the last minutes of an ECT.[17] The moment in time at which a fragment is taken thus makes a difference for the presence of symptoms. Not all selected ten-seconds fragments came from the same moment during the ECT, so the severity of symptoms could be unequal between the different patients. Moreover, this selection makes it impossible to compare symptoms between patients at a specific point in time during the ECT. However, the most representative moments during the ECT (selected by clinicians) are used and sometimes certain fragments could not be used due to speech which would cause recognition in the test with clinicians and unwanted peaks for signal analysis or machine learning. Therefore with the available data, this was the best choice for the selection of fragments.
Because the severity of symptoms generally increases during exercise, the dynamics of the symptoms are relevant. Studying how the symptoms evolve over time could thus provide insight in the patterns of the different respiratory conditions. Due to the low amount of good quality data, especially from the earlier minutes of the ECTs, these dynamics were not studied.

7.2.3 Analysis

A variety of signal analysis techniques was applied to the audio signals in order to study the signals extensively. The sound recordings were high-pass filtered with a cut-off frequency of 230 Hz to remove unwanted low-frequency noise from the treadmill. Our literature search showed that mouth breathing is found in a frequency range from 200 to 2,000 Hz [46] so it is possible that frequencies in the respiration are removed by this high-pass filter. In section 2.2.3, other time-frequency domain analysis techniques were described besides the STFT. These techniques were not applied to the available recordings because, based on the analysis techniques that were performed, it was not expected that these techniques would facilitate extraction of relevant information from the recordings. However, since the techniques are not applied it is uncertain whether these techniques could lead to useful insights in the analysis.

7.2.4 Measurement setup

A strength of this study is the exploration of possible improvements to the measurement setup to optimize signal acquisition. In addition, different algorithms for noise reduction were tried on the acquired test data. A limitation regarding the improvement of the setup is that the optimal distance between the microphone and the mouth of the patient was not investigated. The use of a fixed microphone is in itself a limitation as the patient moves during exercise, changing the relative position of the microphone with respect to the mouth. Moreover, although the audio technica ES933ML/MIC suffices according to the discussed CORSA guidelines for breath sound recordings, it is not validated for the use of recording mouth breathing.

7.2.5 Machine learning

Before applying machine learning algorithms to the gathered data, the quality of the recordings was assessed subjectively. Only recordings with an acceptable or good quality were included for machine learning. In this way, the proportion of recordings from which it was expected that they could enable classification was maximized. By omitting low quality recordings, the data set which was already small for machine learning purposes became even smaller. To combat the negative effects of using a small data set, various data preparation methods were explored to increase the amount of samples.

The performance of various machine learning algorithms, each with their advantages and disadvantages, was investigated by using MATLAB's Classification Learner app. However, the app does not include all existing classification algorithms. At this point, this has probably not influenced the overall performance of machine learning for this problem because all other algorithms could not generate good classification models. However, when quality and size of the data set would be improved and constructing a proper classification algorithm would be possible, it would be a limitation not to be able to explore all classification algorithms.

When preparing the training data, different features were created and tested for the classification. Although different combinations were tried, not all possibilities of feature engineering were explored. For example, more thresholds for peak detection could have been explored. Moreover, we have not determined which combinations of features resulted in the best training and testing accuracies.

7.3 Recommendations for future research

In order to further explore the possibilities for using sounds recorded during ECTs for the objective detection of DB and UA, different steps should be taken in future research. The main recommendations for future research are summarized in the box at the end of this section. The first two steps of the process can be performed simultaneously. These steps are 1) creating more clarity regarding characteristic sounds by studying and labelling existing sound recordings and 2) improving the measurement setup to acquire signals with higher quality. After these two steps have been taken, characteristic sounds can be identified in quality recordings and opportunities of combining these characteristics with parameters from other objective information sources can be explored.

7.3.1 Identifying characteristics

Until better quality signals are acquired, it is advised to take a step back and zoom in on individual recordings instead of trying to classify all available data. In all available recordings with an acceptable quality, characteristic sounds should be labelled to define different features for every diagnosis category. Templates of these features can then be used to detect similar characteristics in other recordings. The recordings made directly after an ECT can also be used for this purpose, as two of these recordings resulted in very clear time domain and time-frequency domain plots. To determine whether the quality of a recording is acceptable, it is advised to develop an algorithm which determines the noise level and decides whether the recording is suitable for further analysis.
Stridor and wheeze are typical sounds but they are not often clearly recognized in the recordings. Because of this and the lack of ambiguous information regarding these CAS in mouth breathing, it is not known how these sounds could be recognized during analysis of sound recordings. Therefore it could be relevant to create samples of stridor and wheeze to investigate what these sounds look like in the different domains.

7.3.2 Improving measurement setup

The most prominent disturbances in the sound recordings arise from the treadmill and the footsteps on the treadmill. The main recommendation for improving the measurement setup is to overcome these disturbances by using a stationary bicycle or cross-trainer for exercise. As these devices are not motor-driven, the sound generated by the device is expected to be less prominent. Moreover, a patient does not step on these devices so footsteps will not be a disturbing noise anymore. At this moment, the protocol for performing an ECT includes exercise on a treadmill (or jumping castle for young children). Therefore it must first be investigated if the provocation of symptoms with a stationary bike or cross-trainer is similar to the provocation of symptoms with a treadmill, before this could be used in practice.

Secondly, it is advised to investigate the possibilities of using a different microphone for recording respiratory sounds. This is especially recommended when the machinery for exercise is not changed. The current microphone is admittedly a directional microphone, but it still has a relatively large angle of acceptance of 90°. A parabolic microphone is a directional microphone that is placed in the focal point of a parabolic reflector. The parabolic reflector focuses the incoming sound at the position of the microphone, resulting in a great sensitivity to sounds coming from one direction.[97] The addition of a parabolic reflector is thus expected to improve the quality of the recorded signals. Another option is to use a microphone which can be worn by the patient. In this way, the microphone can be placed closer to the mouth and the position of the microphone with respect to the patient's mouth remains the same throughout the ECT. A third option is to

use a throat microphone (also known as laryngophone). This type of microphone is attached to the user with a strap around the neck. A transponder absorbs the vibrations generated by the larynx, which are turned into electronic audio signals. As this type of microphone only picks up vibrations from the larynx, ambient noise is not captured. However, the laryngophone may have limitations, such as damping of specific sounds, user comfort and usability during exercise, which need to be further explored.[98]

Thirdly, it is recommended to apply noise-damping material in the climate controlled room. Again, this is especially necessary when exercise on the treadmill is continued. No previous applications of noise-damping material when recording respiratory sounds were found, probably because other RSA studies record respiratory sounds in already quiet environments. However, in gyms, rubber mats are often placed underneath treadmills to prevent damage to the treadmill but also to reduce noise by dispersing vibrations [99]. Although no scientific proof was found for the noise reduction of these mats, findings from literature suggest that rubber can indeed be used for noise reduction in sound [100, 101, 102]. It is therefore recommended to study the noise reduction when applying a rubber mat underneath the treadmill during ECTs. In addition, damping material may be placed around the treadmill (or other exercise device) to prevent reflection of the sound waves by walls. The best way to remove the treadmill noise with damping material would be to place the material somewhere between the mouth of the patient and the treadmill such that the noise cannot reach the microphone. However, coming up with a setup in which this is applied while not limiting the freedom of movement for the patient, is difficult.

7.3.3 Additional research steps

After the above mentioned research steps have been taken, the detection of characteristics indicative of the different diagnoses should be applied to new recordings. In addition, it is recommended to study the dynamics of the sound characteristics. The best approach would be to make a breath sound recording at baseline (so before the ECT), during the ECT, directly after the ECT and a few minutes after the ECT. The recording a few minutes after the ECT is relevant because our literature search showed that with DB, the symptoms peak during exercise while with UA, they usually peak after exercise [17]. From these recordings, for example the ratio between the duration of inspiration and the duration of expiration can be determined. The expectation is that the ratio increases during exercise for DB and that the ratio decreases for UA, especially after exercise. For children with no respiratory condition, the respiratory rate goes up during exercise but the duration of the inspiration and expiration should be approximately the same.

Our study results in section 3 showed that aspects from the sound recordings and images of the patients are often combined for diagnosis. Therefore it is relevant to study the characteristics in sound in combination with parameters from different information sources. Our literature search proved that EMG has the potential to differentiate between the different conditions [10, 33] and researchers in MST plan to conduct a study in which this potential will be further explored. It is recommended that findings from the future EMG study are combined with findings from future sound recording studies to investigate the combination of sound and EMG for the detection of DB and UA. It may then be determined whether sound recordings are an addition to EMG parameters in the detection of respiratory conditions. Moreover, the position of the shoulders during the ECT can be a relevant parameter as shrugged shoulder was sometimes mentioned as characteristic for DB in our study results. Image processing techniques can be applied to the camera footage to determine the position and displacement of the shoulders. Alternatively, accelerometers and gyroscopes can be placed on the shoulders of the patient to determine motion and positioning of the shoulders.

7.3.4 Future use in home-monitoring

At this moment, the development of an objective tool is not in a stadium at which it can be used in clinical practice. If characteristic sounds are established and a classification algorithm is developed in the future, the algorithm may be used for home-monitoring of respiratory complaints. Especially in these times where all physical contact is limited, it would be very useful if parents could make a sound recording of their child during a respiratory exacerbation which could be objectively assessed with a classification algorithm. Before this could be used in practice, it should be explored if smartphones are capable of revealing similar characteristic sounds in sound recordings. Additionally, the diagnosis made by such an algorithm should always be verified by a clinician.

Recommendations

- Create more clarity regarding characteristic sounds & improve measurement setup for higher quality signals
- Identify characteristic sounds in new recordings and study dynamics
- Combine sound characteristics with information from other sources (e.g. EMG and accelerometer data)

8 CONCLUSION

This report describes a first, explorative research to analyze the potential of sound recordings during ECTs to detect DB and UA in children. To investigate the extent to which the sound recordings can be used as an objective measure, relevant characteristics in breathing sound were established and the sound recordings were extensively analyzed.

The results of the study indicate that at this point, the sound recordings alone cannot be used for objective assessment of the respiratory condition in children. The characteristics in sound recordings which could have the potential to differentiate between the conditions are the nature of adventitious sounds and the moment at which respiratory symptoms occur, both during the ECT and during one respiratory cycle. The quality of the sound recordings was too poor to identify these characteristics in the available sound recordings and the described features were often ambiguous. Changing the measurement setup did not lead to sufficient improvement of the quality of the sound recordings.

Following a machine learning approach to differentiate between the respiratory conditions did not result in a well-performing classification algorithm. The inability to create such a classification algorithm is expected to be caused by the poor quality of the data and the small size of the data set. Clinicians showed similar performance when assessing sound recordings: only a third of the sound recordings was correctly classified. When adding camera footage to the sound recordings, the classification accuracy of the clinicians increased to 51%. This accuracy is still low and stresses the importance of further development of an objective diagnostic tool.

Future research should focus on improving the quality of the sound recordings by reconsidering the measurement setup with regard to the used exercise machinery and microphone. With good quality recordings, more knowledge can be acquired regarding the characteristics of the respiratory conditions in sound recordings. Simultaneously, existing recordings should be used to label characteristic sounds and thus create more clarity regarding differentiating features in respiratory sounds. As a next step, characteristics should be identified in new recordings and comparison of good quality recordings before, during and after the ECT may give insights in the dynamics of symptoms. Studying characteristics in sound recordings in combination with parameters from different information sources, like EMG, should also gain attention the follow-up steps.

If these research steps lead to faciliation of objective detection of DB and UA in children, the possibility of revealing characteristic sounds with smartphone recordings during respiratory exacerbations can be investigated. These smartphone sound recordings may provide an elegant way of assessing respiratory complaints in the home-situation.

REFERENCES

[1] Nicki Barker and Mark L. Everard. Getting to grips with 'dysfunctional breathing'. *Paediatric Respiratory Reviews*, 16(1):53–61, 2015.

[2] Nicki Barker, Ravi Thevasagayam, Kelechi Ugonna, and Jane Kirkby. Pediatric Dysfunctional Breathing: Proposed Components, Mechanisms, Diagnosis, and Management, jul 2020.

[3] Nicola J. Barker, Heather Elphick, and Mark L. Everard. The impact of a dedicated physiotherapist clinic for children with dysfunctional breathing. *ERJ Open Research*, 2(3), 2016.

[4] Mike Thomas, Robert McKinley, Elaine Freeman, Chris Foy, and David Price. The prevalence of dysfunctional breathing in adults in the community with and without asthma. *Primary care respiratory journal : journal of the General Practice Airways Group*, 14:78–82, 05 2005.

[5] Jonathan P Parsons, Teal S Hallstrand, John G Mastronarde, David A Kaminsky, Kenneth W Rundell, James H Hull, William W Storms, John M Weiler, Fern M Cheek, Kevin C Wilson, and Sandra D Anderson. American Thoracic Society Documents An Official American Thoracic Society Clinical Practice Guideline: Exercise-induced Bronchoconstriction Diagnosis Measuring and Quantifying EIB Exercise Challenge Testing to Identify EIB Surrogates for Exercise to Identify EIB Treatment Questions and Recommendations General Comments Regarding Therapy Screening for EIB Exercise, Asthma, and Doping. *Am J Respir Crit Care Med*, 187(9):1016–1027, 2013.

[6] Laís Silva Vidotto, Celso Ricardo Fernandes de Carvalho, Alex Harvey, and Mandy Jones. Dysfunctional breathing: What do we know? *Jornal Brasileiro de Pneumologia*, 45(1):1–9, 2019.

[7] Julie Depiazzi and Mark L. Everard. Dysfunctional breathing and reaching one's physiological limit as causes of exercise-induced dyspnoea. *Breathe*, 12(2):120–129, 2016.

[8] Tania CliftonSmith and Janet Rowley. Breathing pattern disorders and physiotherapy: inspiration for our profession. *Physical Therapy Reviews*, 16(1):75–86, 2011.

[9] Richard Boulding, Rebecca Stacey, Rob Niven, and Stephen J. Fowler. Dysfunctional breathing: A review of the literature and proposal for classification. *European Respiratory Review*, 25(141):287–294, 2016.

[10] Pascal Keijzer, Mattiènne van der Kamp, Boony Thio, Frans de Jongh, and Jean Driessen. Assessing paediatric exercise-induced bronchoconstriction using electromyography. *ERJ Open Research*, 6(2):00298–2019, 2020.

[11] Renard Xaviero Adhi Pramono, Stuart Bowyer, and Esther Rodriguez-Villegas. *Automatic adventitious respiratory sound analysis: A systematic review*, volume 12. 2017.

[12] Achmad Rizal, Risanuri Hidayat, and Hanung Adi Nugroho. Signal domain in respiratory sound analysis: Methods, application and future development. *Journal of Computer Science*, 11(10):1005–1016, 2015.

[13] Matthew Amper-West, Reza Saatchi, Nicola Barker, and Heather Elphick. Respiratory Sound Analysis as a Diagnosis Tool for Breathing Disorders. 2019.

[14] Rami J. Oweis, Enas W. Abdulhay, Amer Khayal, and Areen Awad. An alternative respiratory sounds classification system utilizing artificial neural networks. *Biomedical Journal*, 38(2):153–161, 2015.

[15] Mandy Jones, Fiona Troup, John Nugus, Michael Roughton, Margaret Hodson, Charlotte Rayner, Frances Bowen, and Jennifer Pryor. Does manual therapy provide additional benefit to breathing retraining in the management of dysfunctional breathing? A randomised controlled trial. *Disability and Rehabilitation*, 37(9):763–770, 2015.

[16] Thomas Halvorsen, Emil Schwarz Walsted, Caterina Bucca, Andrew Bush, Giovanna Cantarella, Gerhard Friedrich, Felix J.F. Herth, James H. Hull, Harald Jung, Robert Maat, Leif Nordang, Marc Remacle, Niels Rasmussen, Janet A. Wilson, and John Helge Heimdal. Inducible laryngeal obstruction: an official joint European Respiratory Society and European Laryngological Society statement. *The European respiratory journal*, 50(3), 2017.

[17] Ola Drange Røksund, John Helge Heimdal, Hege Clemm, Maria Vollsæter, and Thomas Halvorsen. Exercise inducible laryngeal obstruction: diagnostics and management. *Paediatric Respiratory Reviews*, 21:86–94, 2017.

[18] Surendranath R. Veeram Reddy and Harinder R. Singh. Chest pain in children and adolescents. *Pediatrics in Review*, 31(1):e1–e9, jan 2010.

[19] Janneke C van Leeuwen, Jean M M Driessen, Frans H C de Jongh, Wim M C van Aalderen, and Boony J Thio. Monitoring pulmonary function during exercise in children with asthma. *Archives of Disease in Childhood*, 96(7):664–668, 2011.

[20] M A Howell and H R Guly. A comparison of glucagon and glucose in prehospital hypoglycaemia. *Emergency Medicine Journal*, 14(1):30–32, 1997.

[21] M Thomas, R K McKinley, E Freeman, C Foy, P Prodger, and D Price. Breathing retraining for dysfunctional breathing in asthma: a randomised controlled trial. *Thorax*, 58(2):110–115, 2003.

[22] A. R. Kraft and C. A. L. Hoogduin. The hyperventilation syndrome: A pilot study on the effectiveness of treatment. *British Journal of Psychiatry*, 145(5):538–542, 1984.

[23] P. Grossman, J.C.G. De Swart, and P.B. Defares. A controlled study of a breathing therapy for treatment of hyperventilation syndrome. *Journal of Psychosomatic Research*, 29(1):49 – 58, 1985.

[24] Carina Hagman, Christer Janson, and Margareta Emtner. Breathing retraining - a five-year follow-up of patients with dysfunctional breathing. *Respiratory Medicine*, 105(8):1153 – 1159, 2011.

[25] Erik Peper and Vicci Tibbetts. Effortless Diaphragmatic Breathing, 1997.

[26] Thomas Ritz, Andreas Von Leupoldt, and Bernhard Dahme. Evaluation of a respiratory muscle biofeedback procedure-effects on heart rate and dyspnea. *Applied Psychophysiology Biofeedback*, 31(3):253–261, sep 2006.

[27] Jan Van Dixhoorn. Efficiacy of Nijmegen questionnaire in recognition of the hyperventilation syndrome of chronic , habitual or acute of a becoming chronic and disabling . is the fact that clinically a large number of can be Provocation HVS , and was found capable of discri. *Journal of Psychosomatic Research*, 29(2):199–206, 1985.

[28] Jan Van Dixhoorn and Hans Folgering. The nijmegen questionnaire and dysfunctional breathing. *ERJ Open Research*, 1(1):3–6, 2015.

[29] Rosalba Courtney, Jan Van Dixhoorn, and Marc Cohen. Evaluation of breathing pattern: Comparison of a manual assessment of respiratory motion (MARM) and respiratory induction plethysmography. *Applied Psychophysiology Biofeedback*, 33(2):91–100, 2008.

[30] Janneke C. Van Leeuwen, Jean M.M. Driessen, Frans H.C. De Jongh, Sandra D. Anderson, and Boony J. Thio. Measuring breakthrough exercise-induced bronchoconstriction in young asthmatic children using a jumping castle. *Journal of Allergy and Clinical Immunology*, 131(5):1427–1429.e5, 2013.

[31] R. O. Crapo, R. Casaburi, A. L. Coates, P. L. Enright, J. L. Hankinson, C. G. Irvin, N. R. MacIntyre, R. T. McKay, J. S. Wanger, S. D. Anderson, D. W. Cockcroft, J. E. Fish, and P. J. Sterk. American Thoracic Society guidelines for methacholine and exercise challenge testing -1999. *American Journal of Respiratory and Critical Care Medicine*, 161(1):309–329, 2000.

[32] M R Miller, J Hankinson, V Brusasco, F Burgos, R Casaburi, A Coates, R Crapo, P Enright, C P M Van Der Grinten, P Gustafsson, R Jensen, D C Johnson, N Macintyre, R Mckay, D Navajas, O F Pedersen, R Pellegrino, G Viegi, and J Wanger. Standardisation of spirometry. 26(2):319–338, 2005.

[33] P. Keijzer, B. Thio, F. de Jongh, and J. Driessen. Assessment of Asthma in Children Using Electromyography. pages A6106–A6106, 2019.

[34] Maura A. Watson, Christopher S. King, Aaron B. Holley, David L. Greenburg, and Jeffrey A. Mikita. Clinical and lung-function variables associated with vocal cord dysfunction. *Respiratory Care*, 54(4):467–473, 2009.

[35] G John Gibson, William Whitelaw, Nikolaos Siafakas, Gerald S Supinski, Jean Will Fitting, François Bellemare, Stephen H Loring, Andre De Troyer, and Alex E Grassino. ATS/ERS Statement on respiratory muscle testing. *American journal of respiratory and critical care medicine*, 166(4):518–624, 2002.

[36] Roberto Merletti and Philip Parker. *Electromyography: Physiology, Engineering, and Non-Invasive Applications*. John Wiley & Sons, Hoboken, NJ, 2004.

[37] Roberto Merletti and Dario Farina. *Surface Electromyography: Physiology, Engineering, and Applications*. John Wiley & Sons, Hoboken, NJ, 2016.

[38] Eric J.W. Maarsingh, Mireille Oud, Leo A. van Eykern, Maarten O. Hoekstra, and Wim M.C. van Aalderen. Electromyographic monitoring of respiratory muscle activity in dyspneic infants and toddlers. *Respiratory Physiology Neurobiology*, 150(2):191 – 199, 2006.

[39] Ivanize Mariana Masselli Dos Reis, Daniela Gonçalves Ohara, Letícia Bergamin Januário, Renata Pedrolongo Basso-Vanelli, Ana Beatriz Oliveira, and Mauricio Jamami. Surface electromyography in inspiratory muscles in adults and elderly individuals: A systematic review. *Journal of Electromyography and Kinesiology*, 44:139 – 155, 2019.

[40] ML Duiverman, LA van Eykern, PW Vennik, GH Koeter, EJW Maarsingh, and PJ Wijkstra. Reproducibility and responsiveness of a noninvasive emg technique of the respiratory muscles in copd patients and in healthy subjects. *Journal of Applied Physiology*, 96(5):1723–1729, May 2004.

[41] E. J.W. Maarsingh, L. A. Van Eykern, A. B. Sprikkelman, M. O. Hoekstra, and W. M.C. Van Aalderen. Respiratory muscle activity measured with a noninvasive EMG technique: Technical aspects and reproducibility. *Journal of Applied Physiology*, 88(6):1955–1961, 2000.

[42] Eric J.W. Maarsingh, Leo A. Van Eykern, Rob J. De Haan, Rupino W. Griffioen, Maarten O. Hoekstra, and Wim M.C. Van Aalderen. Airflow limitation in asthmatic children assessed with a non-invasive EMG technique. *Respiratory Physiology and Neurobiology*, 133(1-2):89–97, 2002.

[43] Charles E. Speaks. The Nature of Sound Waves. In *Introduction to Sound: Acoustics for the Hearing and Speech Sciences*. Plural Publishing, San Diego, fourth edition, 2018.

[44] Mayumi Enseki, Mariko Nukaga, Hiromi Tadaki, Hideyuki Tabata, Kota Hirai, Masahiko Kato, and Hiroyuki Mochizuki. A breath sound analysis in children with cough variant asthma. *Allergology International*, 68(1):33–38, 2019.

[45] Parthak Prodhan, Reynaldo S. Dela Rosa, Maria Shubina, Kenan E. Haver, Benjamin D. Matthews, Sarah Buck, Robert M. Kacmarek, and Natan N. Noviski. Wheeze detection in the pediatric intensive care unit: Comparison among physician, nurses, respiratory therapists, and a computerized respiratory sound monitor. *Respiratory Care*, 53(10):1304–1309, 2008.

[46] Paul Forgacs, A. R. Nathoo, and H. D. Richardson. Breath sounds. *Thorax*, 26(3):288–295, 1971.

[47] Sami Spencer, Belinda H. Yeoh, Peter P. Van Asperen, and Dominic A. Fitzgerald. Biphasic stridor in infancy. *Medical Journal of Australia*, 180(7):347–349, 2004.

[48] Plamen Bokov, Bruno Mahut, Patrice Flaud, and Christophe Delclaux. Wheezing recognition algorithm using recordings of respiratory sounds at the mouth in a pediatric population. *Computers in Biology and Medicine*, 70:40 – 50, 2016.

[49] Howard Austerlitz. Chapter 10 - data processing and analysis. In Howard Austerlitz, editor, *Data Acquisition Techniques Using PCs (Second Edition)*, pages 222 – 250. Academic Press, San Diego, second edition edition, 2003.

[50] Gjerrit Meinsma, Ciska Heida, and Michel J.A.M. van Putten. Advanced Techniques for Signal Analysis. pages 1–93, 2011.

[51] Wavelet Transforms in MATLAB - MATLAB & Simulink.

[52] P.F. Pai. 14 - space wavenumber and time–frequency analyses for vibration and wave-based damage diagnosis. In Fuh-Gwo Yuan, editor, *Structural Health Monitoring (SHM) in Aerospace Structures*, pages 393 – 426. Woodhead Publishing, 2016.

[53] Yasemin P. Kahya. Breath Sound Recording. In *Breath Sounds From Basic Science to Clinical Practice*, chapter 8, pages 119–137. Springer, 2018.

[54] Laura Vannuccini, J.E. Earis, Panu Helistö, Barry Cheetham, M. Rossi, A.R.A. Sovijärvi, and J. Vanderschoot. Capturing and preprocessing of respiratory sounds. Technical report, Eur Respir Rev 10, 2000.

[55] Hugo Gvert, Jarmo Hurri, Jaakko Srel, and Aapo Hyvrinen. FastICA for Matlab 7.x and 6.x, 2005.

[56] Saeed V. Vaseghi. Wiener Filters. In *Advanced Digital Signal Processing and Noise Reduction*, chapter 6, pages 178–204. John Wiley & Sons Ltd, second edition, 2000.

[57] Alan Oppenheim and George Verghese. Wiener Filtering. In *Signals, Systems and Interference*, chapter 11, pages 195–210. Prentice Hall, 2010.

[58] Weifeng Liu, José C. Principe, and Simon Haykin. Background and Preview. In *Kernel Adaptive Filtering: A Comprehensive Introduction*, pages 1–26. Wiley, 2010.

[59] Overview of Adaptive Filters and Applications - MATLAB & Simulink.

[60] Adaptive Noise Cancellation Using RLS Adaptive Filtering - MATLAB & Simulink.

[61] Compare RLS and LMS Adaptive Filter Algorithms - MATLAB & Simulink.

[62] Sorin Zoican. Audio signals noise removal real time system. In *2010 8th International Conference on Communications, COMM 2010*, pages 25–28, 2010.

[63] Wavelet Denoising - MATLAB & Simulink.

[64] Undecimated Wavelet Transform (Advanced Signal Processing Toolkit) - LabVIEW 2010 Advanced Signal Processing Toolkit Help - National Instruments.

[65] Shai Shalev-Shwartz and Shai Ben-David. Introduction. In *Understanding Machine Learning: From Theory to Algorithms*, chapter 1, pages 1–10. Cambridge University Press, 2014.

[66] Witten D Hastie T Tibshirani R James, G. *An Introduction to Statistical Learning - with Applications in R | Gareth James | Springer*. 2013.

[67] Max Bramer. *Principles of Data Mining*. Undergraduate Topics in Computer Science. Springer, London, 3 edition, 2013.

[68] Harikumar Rajaguru and Sunil Kumar Prabhakar. KNN Classifier. In *KNN Classifier and K-Means Clustering for Robust Classification of Epilepsy from EEG Signals. A Detailed Analysis*, chapter 3, pages 31–38. Anchor Academic Publishing, Hamburg, 2017.

[69] Weighted K-NN - GeeksforGeeks | https://www.geeksforgeeks.org/weighted-k-nn/.

[70] Alexander Statnikov, Constantin F. Aliferis, Douglas P. Hardin, and Isabelle Guyon. *A gentle introduction to support vector machines in biomedicine: Volume 1: Theory and methods*. World Scientific Publishing Co., jan 2011.

[71] Hucker Marius. Multiclass Classification with Support Vector Machines (SVM), Dual Problem and Kernel Functions | Towards Data Science.

[72] Kevin P Murphy. Naive Bayes classifiers | https://www.cs.ubc.ca/ murphyk/Teaching/CS340-Fall06/lectures/naiveBayes.pdf.

[73] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. Kernel smoothing methods, 2009.

[74] Lior Rokach. Introduction to Ensemble Learning. In *Pattern Classification Using Ensemble Methods*, chapter 2, pages 19–64. World Scientific, Singapore, 2010.

[75] Pierre Geurts, A. Zighed Djamel, and Jan Komorowski. Some Enhancements of Decision Trees. In *Prinicples of Data Mining and Knowledge Discovery: 4th European Conference, PKDD, 2000, Lyon, France, September 13-16, 2000 Proceedings*, pages 136–147. Springe rScience & Business Media, Lyon.

[76] Karthikrajan Senthilnathan, Balamurugan Shanmugam, Dinesh Goyal, Iyswarya Annapoorani, and Ravi Samikannu. *Deep Learning Applications and Intelligent Decision Making in Engineering*. Advances in Computational Intelligence and Robotics. IGI Global, Hershey, 2020.

[77] R. J. Kurukulaaratchy, M. Fenn, R. Twiselton, S. Matthews, and S. H. Arshad. The prevalence of asthma and wheezing illnesses amongst 10-year-old schoolchildren. *Respiratory Medicine*, 96(3):163–169, 2002.

[78] Geva Vashitz, Joseph S. Pliskin, Yisrael Parmet, Yona Kosashvili, Gal Ifergane, Shlomo Wientroub, and Nadav Davidovitch. Do first opinions affect second opinions? *Journal of General Internal Medicine*, 27(10):1265–1271, 2012.

[79] Pascal M. Burger, Jan Westerink, and Bram E.L. Vrijsen. Outcomes of second opinions in general internal medicine. *PLoS ONE*, 15(7 July):1–17, 2020.

[80] Paaladinesh Thavendiranathan, Zoran B. Popović, Scott D. Flamm, Arun Dahiya, Richard A. Grimm, and Thomas H. Marwick. Improved interobserver variability and accuracy of echocardiographic visual left ventricular ejection fraction assessment through a self-directed learning program using cardiac magnetic resonance images. *Journal of the American Society of Echocardiography*, 26(11):1267–1273, 2013.

[81] C. Haritoglou, A. S. Neubauer, H. Herzum, W. R. Freeman, and A. J. Mueller. Interobserver and intraobserver variability of measurements of uveal melanomas using standardised echography. *British Journal of Ophthalmology*, 86(12):1390–1394, 2002.

[82] Jen-Chien Chien, Huey-Dong Wu, Fok-Ching Chong, and Chung-I Li. Wheeze detection using cepstral analysis in gaussian mixture models. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3168–3171, 2007.

[83] R. Riella, P. Nohama, and J. Maia. Method for automatic detection of wheezing in lung sounds. *Brazilian journal of medical and biological research = Revista brasileira de pesquisas medicas e biologicas*, 42 7:674–84, 2009.

[84] Chizu Habukawa, Naoto Ohgami, Naoki Matsumoto, Kenji Hashino, Kei Asai, Tetsuya Sato, and Katsumi Murakami. A wheeze recognition algorithm for practical implementation in children. *PLoS ONE*, 15(10 October):1–12, 2020.

[85] Jiarui Li and Ying Hong. Wheeze detection algorithm based on spectrogram analysis. In *2015 8th International Symposium on Computational Intelligence and Design (ISCID)*, volume 1, pages 318–322, 2015.

[86] Zena El Hajj. Analysis of Sound Localization and Microphone Arrays. *Ingenium Revista de la facultad de ingeniería*, 15(29):49, 2014.

[87] Volker Gross, Anke Dittmar, Thomas Penzel, Frank Schüttler, and Peter von Wichert. The relationship between normal lung sounds, age, and gender. *American Journal of Respiratory and Critical Care Medicine*, 162(3):905–909, 2000. PMID: 10988103.

[88] Ignacio Sanchez and Hans Pasterkamp. Tracheal sound spectra depend on body height. *American Review of Respiratory Disease*, 148(4_pt_1):1083–1087, 1993. PMID: 8214929.

[89] Shubhra Dixit and D. Nagaria. Design and analysis of cascaded lms adaptive filters for noise cancellation. *Circuits, Systems, and Signal Processing*, 36:742–766, 2017.

[90] Sayed. A. Hadei and M. Lotfizad. A Family of Adaptive Filter Algorithms in Noise Cancellation for Speech Enhancement. *International Journal of Computer and Electrical Engineering*, 2(2):307–315, 2010.

[91] A. Zerguine, M. Bettayeb, and C.F.N. Cowan. Hybrid LMS-LMF algorithm for adaptive echo cancellation. *Image (Rochester, N.Y.)*, 146(4):173–180, 1999.

[92] Mohammad Zia Ur Rahman, Rafi Ahamed Shaik, and D V Rama Koti Reddy. Adaptive noise removal in the ecg using the block lms algorithm. In *2009 2nd International Conference on Adaptive Science Technology (ICAST)*, pages 380–383, 2009.

[93] A.Bhavani Sankar, D Kumar, and K Seethalakshmi. Performance Study of Various Adaptive filter algorithms for Noise Cancellation in Respiratory Signals. *Signal Processing : An International Journal (SPIJ)*, 4(5):267 – 278, 2010.

[94] Sánchez C., Rieta J.J., Vayá C., Perez D.M., Zangróniz R., and Millet J. Wavelet Denoising as Preprocessing Stage to Improve ICA Performance in Atrial Fibrillation Analysis. In *Independent Component Analysis and Blind Signal Separation*, pages 486–494. Springer, Berlin, 2006.

[95] Abdul Ghaaliq Lalkhen and Anthony McCluskey. Clinical tests: sensitivity and specificity. *Continuing Education in Anaesthesia Critical Care Pain*, 8(6):221–223, 12 2008.

[96] Daniël A. Korevaar, Gowri Gopalakrishna, Jérémie F. Cohen, and Patrick M. Bossuyt. Targeted test evaluation: a framework for designing diagnostic accuracy studies with clear study hypotheses. *Diagnostic and Prognostic Research*, 3(1):1–10, 2019.

[97] Francis Rumsey and Tim McCormick. Parabolic microphone. In *Sound and recording*, chapter 3. Microph, page 60. Elsevier/Focal, Amsterdam, 6th edition, 2009.

[98] IASUS Concepts. About throat mics. https://iasus-concepts.com/about-us/about-throat-mics/.

[99] Simon Gould. Do I Need A Treadmill Mat>.

[100] Partheban Manoharan and Kinsuk Naskar. Recycling of Tire Rubbers and Their Reusability. In *Rubber Recycling: Challenges and Developments*, chapter 5, pages 102–125. Royal Society of Chemistry, 2018.

[101] Valiya Parambath Swapna and Ranimol Stephen. Recycling of Rubber. In *Recycling of Polymers: Methods, Characterization and Applications*, pages 141–160. John Wiley & Sons, 2016.

[102] R Gayathri, R Vasanthakumari, and C Padmanabhan. Sound absorption, Thermal and Mechanical behavior of Polyurethane foam modified with Nano silica, Nano clay and Crumb rubber fillers. 4(5):301–308, 2013.

A FORMS FOR TESTS WITH CLINICIANS

Below, the forms that were offered to clinicians for assessing the audio and video fragments can be found. The first form is for assessing the audio recordings, the second form for assessing the video recordings. The forms are in Dutch.

Uitleg beoordeling geluidsfragmenten

In dit zipbestand staan 15 geluidsfragmenten van kinderen tijdens een inspanningstest. Dit zijn opnames van kinderen met ongecontroleerde astma, disfunctioneel ademen en zonder respiratoire aandoening. De bedoeling is dat u alle bestanden luistert en gaat letten op de ademhaling die te horen is. De kenmerken van de ademhaling dienen te worden genoteerd in de tabel die op de volgende pagina te vinden is. Daarin dient ook te worden genoteerd welke diagnose volgens u bij deze ademgeluiden past, met een toelichting. Als laatste dient er met een cijfer van 1 t/m 10 aangegeven te worden hoe zeker u bent van de gestelde diagnose, waarbij 1 heel onzeker is en 10 heel zeker. Hieronder een aantal aspecten waar u aan kunt denken bij het invullen van de verschillende kolommen.

Beoordeling ademhalingspatroon

- Frequentie (rustig/snel)
- Constante ademhaling?
- Verhouding in duur inspirium:expirium t.o.v. normaal
- Tweede fase inademing?

Bijgeluiden (indien aanwezig)

- Karakter (bijv. piepend, gierend, luid, zacht, hoesten, bifasisch)
- Moment van optreden in ademhalingscyclus

Diagnose

- Ongecontroleerde astma
- Disfunctioneel ademen
- Geen ademhalingsdiagnose

Toelichting

- Welke aspecten in het geluidsfragment hebben tot deze diagnose geleid?

Geluidsfragment	Beoordeling ademhalingspatroon	Bijgeluiden	Diagnose	Toelichting	Zekerheidsschaal 1 (onzeker) t/m 10 (zeker)
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					

Beste beoordelaar,

Voor mijn afstudeeropdracht van Biomedical Engineering ben ik aan het onderzoeken of de geluidsopnames die gemaakt worden tijdens een ECT kunnen bijdragen aan het stellen van de juiste diagnose bij kinderen met respiratoire klachten. Een onderdeel van mijn onderzoek is dat ik wil achterhalen waar normaal gesproken op gelet wordt tijdens een ECT voor het stellen van een diagnose. Daarom staan er in deze map 15 videofragmenten waarbij ik van u zou willen weten welke diagnose u zou stellen en welke beelden/geluiden u tot deze diagnose gebracht hebben. De fragmenten zijn van kinderen met ongecontroleerde astma, disfunctioneel ademen of geen diagnose. Uiteraard is het de bedoeling dat er enkel beoordeeld wordt op basis van de fragmenten, het is niet de bedoeling dat bijbehorende dossiers geraadpleegd worden.
In het document 'Beoordelingsformulier' staat een tabel waarin deze beoordelingen ingevuld kunnen worden. Voor u dat bestand invult, zou ik willen vragen om het bestand op te slaan als 'Beoordeling_*je initialen*', dus dat zou voor mij 'Beoordeling_MM' worden. Op die manier blijft het originele bestand leeg zodat de volgende beoordelaar het weer kan gebruiken.
In de kolom 'kenmerken' dient te worden genoteerd wat er opvalt bij het kijken en luisteren naar de opname. In de kolom 'diagnose' wordt vervolgens de diagnose ingevuld die u stelt op basis van het fragment, waarbij gekozen kan worden tussen ongecontroleerde astma (OA), disfunctioneel ademen (DA) en geen diagnose (GD). In de 'toelichting' kolom geeft u aan wat u gebracht heeft tot deze diagnose en in de kolom 'zekerheidsscore' geeft u met een cijfer van 1 t/m 10 aan hoe zeker u bent van de gestelde diagnose. Uiteraard mogen de kolommen groter gemaakt worden indien u meer ruimte nodig heeft. Daarnaast opent Wordpad het bestand automatisch in portretstand, dus het kan handig zijn voor het invullen om bij de 'page setup' de 'landscape orientation' aan te vinken.
Indien er een opname voorbij komt van een patiënt die u zelf hebt gezien/beoordeeld, vul dan voor elke kolom in die rij een 'X' in.

Hopelijk is alles zo duidelijk, alvast bedankt voor het invullen!

Marieke Massa

Geluidsfragment	Kenmerken	Diagnose	Toelichting	Zekerheidsschaal 1 (onzeker) t/m 10 (zeker)
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				

B ADDITIONAL RESULTS RESPIRATORY SOUNDS IN CLINIC

Table B.1 shows the diagnoses that were given by each assessor in the two different audio assessment rounds, per fragment. The certainty scores are indicated between brackets. The correct diagnoses are highlighted.

Table B.1: Overview of all 'answers' given by the four assessors for the audio recordings. The number behind the diagnosis is the certainty score for that fragment. The first row shows from which assessor in which round the answers come, so A1_R1 means assessor 1 in round 1.

	A1_R1	**A1_R2**	**A2_R1**	**A2_R2**	**A3_R1**	**A3_R2**	**A4_R1**	**A4_R2**
UA_1	ND(6)	ND(6)	ND(8)	ND(4)	ND(7)	ND(7)	ND(4)	ND(7)
UA_2	UA(2)	ND(2)	DB(7)	UA(6)	ND(6)	DB(6)	UA(4)	ND(4)
UA_3	UA(10)	UA(9)	UA(5)	UA(6)	ND(5)	ND(8)	DB(4)	ND(7)
UA_4	DB(8)	UA(3)	DB(7)	DB(7)	ND(7)	ND(8)	ND(6)	ND(5)
UA_5	ND(8)	ND(8)	UA(7)	UA(5)	ND(7)	ND(7)	ND(5)	ND(5)
DB_1	DB(5)	DB(7)	ND(8)	ND(7)	ND(6)	ND(7)	ND(5)	ND(7)
DB_2	ND(4)	ND(3)	UA(7)	ND(6)	ND(5)	ND(7)	ND(7)	ND(6)
DB_3	DB(6)	DB(8)	DB(6)	ND(5)	DB(7)	ND(6)	ND(4)	DB(5)
DB_4	DB(10)	DB(10)	ND(4)	DB(6)	UA(5)	ND(7)	DB(5)	ND(4)
DB_5	UA(7)	UA(4)	ND(9)	DB(7)	UA(7)	UA(7)	UA(8)	UA(7)
ND_1	ND(6)	DB(5)	UA(6)	DB(8)	DB(5)	ND(6)	ND(3)	ND(5)
ND_2	UA(9)	UA(8)	UA(7)	UA(7)	ND(6)	ND(7)	ND	ND(6)
ND_3	DB(5)	DB(7)	ND(8)	UA(7)	UA(6)	UA(6)	ND(5)	DB(7)
ND_4	UA(5)	UA(7)	DB(8)	ND(5)	ND(7)	UA(6)	ND(5)	ND(7)
ND_5	DB(3)	DB(5)	DB(6)	DB(7)	UA(5)	DB(9)	ND(5)	ND(6)

Table B.2 shows the diagnoses that were given by each assessor for the video recordings, per fragment. Assessors 1-3 are the same assessors as 1-3 in the audio recordings, assessor 4 is a new assessor. The certainty scores are indicated between brackets. Assessor 3 has not indicated certainty scores. The correct diagnoses are highlighted.

Table B.2: Overview of all 'answers' given by the four assessors for the video recordings. The number behind the diagnosis is the certainty score for that fragment.

	A1	**A2**	**A3**	**A4**
UA_1	ND(5)	UA(7)	DB	ND(4)
UA_2	UA(8)	UA(6)	UA	UA(3)
UA_3	UA(10)	X	ND	UA(5)
UA_4	UA(7)	DB(8)	ND	DB(4)
UA_5	X	UA(8)	UA	UA(5)
DB_1	DB(9)	UA(7)	UA	DB(4)
DB_2	UA(4)	ND(5)	DB	UA(4)
DB_3	DB(9)	DB(7)	DB	DB(4)
DB_4	DB(10)	DB(8)	ND	UA(3)
DB_5	DB(4)	DB(8)	ND	ND(4)
ND_1	ND(3)	ND(6)	DB	DB(3)
ND_2	UA(6)	UA(9)	UA	UA(6)
ND_3	ND(6)	ND(6)	UA	UA(3)
ND_4	ND(2)	DB(7)	ND	UA(2)
ND_5	ND(4)	X	DB	UA(3)

C SPECIFICATIONS AUDIO TECHNICA ES933ML/MIC

ES933ML/MIC & ES933WML/MIC

MicroLine® Condenser Hanging Microphones

Engineered Sound® Microphones

Features

- **MicroLine® polar pattern provides narrow 90° acceptance angle**
- **Low-profile design with low-reflectance finish for minimum visibility**
- **Superior off-axis rejection for maximum gain before feedback**
- **UniGuard® RFI-shielding technology offers outstanding rejection of radio frequency interference (RFI)**
- **Available interchangeable elements permit angle of acceptance from 90° to 360°**
- **Steel hanger positions microphone over choirs, instrumental groups and theater stages**
- **Available in two colors: black (ES933ML/MIC) and white (ES933WML/MIC)**

Description

The ES933ML/MIC is a wide-range miniature condenser microphone with a MicroLine® (line cardioid) polar pattern. It is designed for quality sound reinforcement, professional recording, television and other demanding sound pickup applications. The combination of small size and excellent response makes the microphone ideal for suspension over choirs, instrumental groups or theater stages.

The microphone requires a compatible Audio-Technica power module (not included) for operation.

The microphone is equipped with UniGuard® RFI-shielding technology, which offers outstanding rejection of radio frequency interference (RFI).

The microphone's MicroLine® (line cardioid) polar pattern provides a 90° angle of acceptance. Additional interchangeable elements with omnidirectional (360°), cardioid (120°) and hypercardioid (100°) pickup patterns are available.

The microphone include a 15.2 m (50') permanently attached miniature cable. Its free end connects to a compatible Audio-Technica power module (not included) via a special TA3F-type connector designed to optimize RFI immunity.

The microphone comes equipped with a vinyl-coated steel hanger for positioning over a choir/orchestra/stage, and a two-stage foam windscreen. The microphone is enclosed in a rugged housing with a low-reflectance black finish. It is also available with white housing, cable, hanger and windscreen as the ES933WML/MIC.

Installation and Operation

The ES933ML/MIC requires a compatible Audio-Technica power module (not included) for operation.

A uniform 90° angle of acceptance provides well-balanced audio pickup. The microphone should be located forward of the front-most source, above the rear-most source, and "aimed" between them (Fig.1). Increasing the height of the mic above the sources will tend to equalize sound levels between them, but may also increase background/ reverberant sound pickup. When possible, the distance from the mic to the rear-most source should be no more than twice the distance to the front source, to maintain front-to-rear balance (Fig. 1).

Width of pickup is approximately 1.5 times the distance to the closest source. If additional mics are needed for wide sources, they should be positioned apart laterally at least 1.5 times the distance to the front source, to avoid phase cancellation (Fig. 2).

To orient the microphone in the proper direction, twist the housing slightly in its wire holder. (Clockwise rotation moves the microphone to the right; counterclockwise rotation moves it to the left.)

The provided two-stage foam windscreen simply snaps over the head of the microphone, effectively reducing noise from wind or ventilation air currents.

Avoid leaving the microphone in the open sun or in areas where temperatures exceed 110° F (43° C) for extended periods. Extremely high humidity should also be avoided.

Note: Audio-Technica has developed a special RFI-shielding mechanism, which is an integral part of the connectors in the Engineered Sound® line. If you remove or replace the connector, you may adversely affect the unit's RFI immunity. Audio-Technica offers a crimp tool (ATCT) and RFI shields that enable you to shorten the cable and correctly reinstall the connector while maintaining the highest level of RFI immunity.

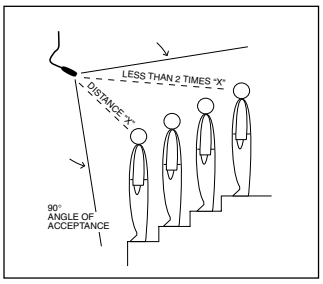


Figure 1

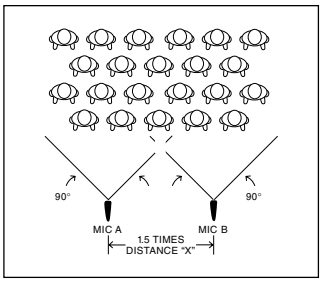


Figure 2

audio-technica

ES933ML/MIC & ES933WML/MIC

To reduce the environmental impact of a multi-language printed document, product information is available online at www.audio-technica.com in a selection of languages.

Afin de réduire l'impact sur l'environnement de l'impression de plusieurs langues, les informations concernant les produits sont disponibles sur le site www.audio-technica.com dans une large sélection de langue.

Para reducir el impacto al medioambiente, y reducir la producción de documentos en varios leguajes, información de nuestros productos están disponibles en nuestra página del Internet: www.audio-technica.com.

Para reduzir o impacto ecológico de um documento impresso de várias linguas, a Audio-Technica providência as informações dos seus produtos em diversas linguas na www.audio-technica.com.

Per evitare l'impatto ambientale che la stampa di questo documento determinerebbe, le informazioni sui prodotti sono disponibili online in diverse lingue sul sito www.audio-technica.com.

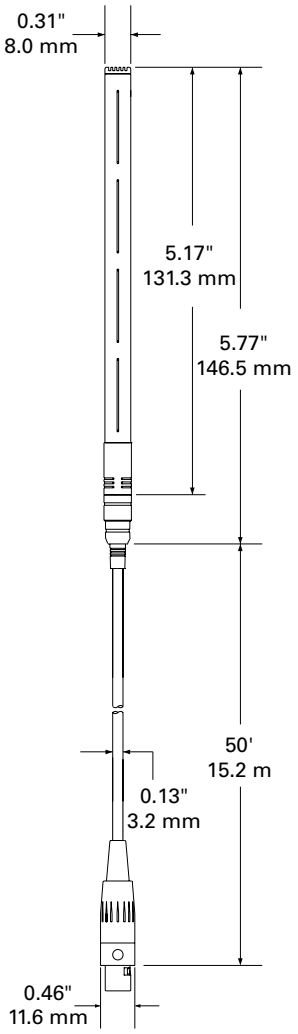
Der Umwelt zuliebe finden Sie die Produktinformationen in deutscher Sprache und weiteren Sprachen auf unserer Homepage: www.audio-technica.com.

Om de gevolgen van een gedrukte meertalige handleiding op het milieu te verkleinen, is productinformatie in verschillende talen "on-line" beschikbaar op: www.audio-technica.com.

本公司基於減少對環境的影響，將不作多語言文件的印刷，有關產品訊息可在 www.audio-technica.com 的官方網頁上選擇所屬語言及瀏覽。

本公司基于减少对环境的影响，将不作多语言文檔的印刷，有关产品信息可在 www.audio-technica.com的官方网页上选择所属语言和浏览。

자원절약, 환경보호를 위해 국문 사용 설명서는 인쇄하지 않았습니다.
제품정보는 www.audio-technica.com 에서 원하는 언어 선택 후에 다운로드 받으실 수 있습니다.



Specifications

Element	Fixed-charge back plate, permanently polarized condenser
Polar pattern	MicroLine® (line cardioid)
Frequency response	30-20,000 Hz
Open circuit sensitivity	–35 dB (17.7 mV) re 1V at 1 Pa
Maximum input sound level	133 dB SPL, 1 kHz at 1% T.H.D.
Dynamic range (typical)	109 dB, 1 kHz at Max SPL
Signal-to-noise ratio[1]	70 dB, 1 kHz at 1 Pa
Weight	9.5 g (0.3 oz)
Dimensions	146.5 mm (5.77") long, 8.0 mm (0.31") head diameter
Output connector	TA3F-type
Cable	15.2 m (50') long (permanently attached to microphone), 3.2 mm (0.13") diameter, 2-conductor shielded cable, terminated with TA3F-type connector
Optional interchangeable elements	ESE-O omnidirectional (360°) ESE-C cardioid (120°) ESE-H hypercardioid (100°)
Audio-Technica case style	M14
Accessories furnished	
ES933ML/MIC	AT8452 steel hanger; AT8138 two-stage foam windscreen
ES933WML/MIC	AT8452(WH) steel hanger; AT8138(WH) two-stage foam windscreen
Compatible power modules	AT8534 wall/ceiling plate power module; AT8538 power module; ATND8734a network audio microphone power module

Specifications derived by using AT8538 power module.

In the interest of standards development, A.T.U.S. offers full details on its test methods to other industry professionals on request.

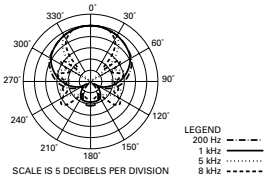
1 Pascal = 10 dynes/cm^2 = 10 microbars = 94 dB SPL

[1] Typical, A-weighted, using Audio Precision System One.

Specifications are subject to change without notice.

frequency response: 30–20,000 Hz

polar pattern



audio-technica

Audio-Technica Corporation
www.audio-technica.com
©2018 Audio-Technica

D ADDITIONAL RESULTS TEST MEASUREMENTS

D.1 Peak frequencies for different setups

Table D.1 shows the peak frequencies determined with different methods for all setups described in chapter 4.
Figure D.1 shows an example of a periodogram for which the peak values were determined and used to create bandstop filters.

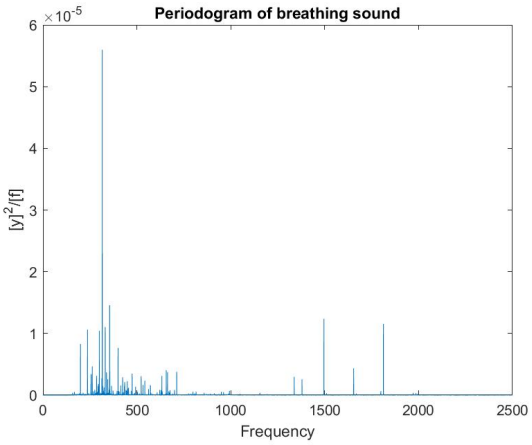
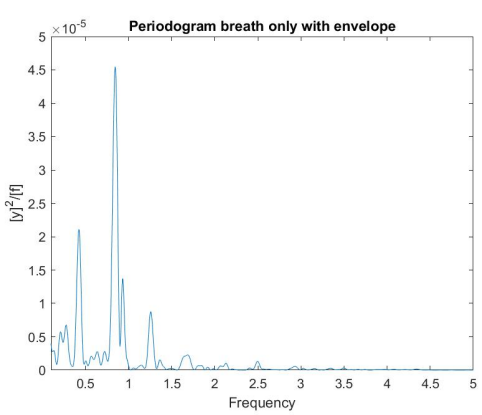
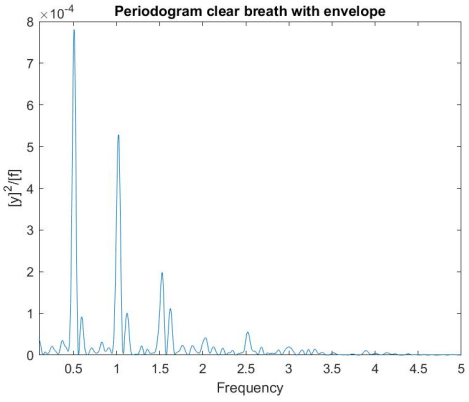


Figure D.1: Periodogram of the treadmill sound at a velocity of 8 km/h.

Figure D.2a shows the periodogram of the envelope of the breath only recording. The peak value of 0.84 Hz stands out. Figure D.2a shows the periodogram of the envelope of the clear breath recording with ambient noise. The peak values of 0.51 and 1.02 Hz stand out.



(a) Periodogram of envelope of the positive values of the recording with only breath sound.



(b) Periodogram of envelope of the positive values of the recording with a clear breath while running on the treadmill.

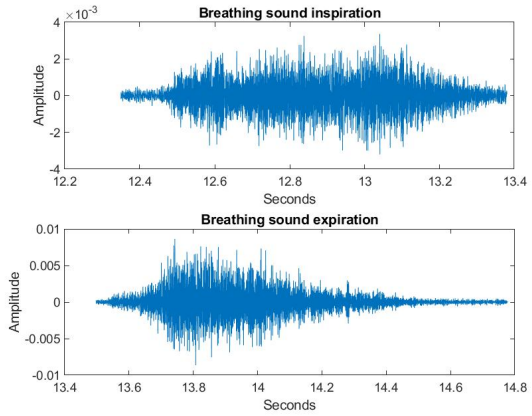
Figure D.2: Periodograms of envelopes of breath only and clear breath recording.

Table D.1: Peak frequency determined with different methods for all 20 measurements. The dominant frequencies are displayed in order of height (with the highest peak first). Fmed stands for the median frequency of the signal. Only for the last two measurements, the peaks for the periodogram of the enveloped signal are determined.

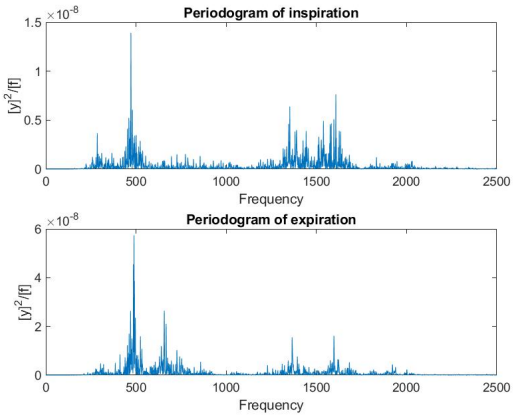
	Velocity (km/h)	**Fmed**	**Peaks periodogram**	**Peaks Welch**	**Peaks AR**	**Peaks periodogram enveloped signal**
Treadmill not directed	**4**	405	239, 957	239, 957	234, 469	
	5	585	597, 1136, 298, 938	596, 1138, 298, 938	469, 938, 1133	
	6	357	357, 1362	356, 1362	352, 1484	
	7	446	416, 1728, 208, 1589	1729, 415, 210, 1587	430, 195, 1719, 1211	
	8	475	475, 950, 1496, 1813	474, 947	469, 234	
Treadmill directed	**4**	474	239	239	234	
	5	298	298	298	313, 1133	
	6	357	357	356	352	
	7	473	1589, 416, 208	1587, 415, 210	430, 313, 1211	
	8	365	316, 1496, 1813	317	313, 1680	
Individual on treadmill unclear not directed	**4**	725	241, 730	239, 732	234, 1133	
	6	372	359, 1365	356, 1362	352, 1367	
	8	555	477	479, 952, 1816	469	
Individual on treadmill unclear directed	**4**	509	241	239	234	
	5	576	300	298, 601	313, 625	
	6	363	359, 1365	356, 1362	352, 1367	
	7	1006	1732, 200, 1593, 418	1733, 1592, 420	1719, 313	
	8	695	202, 953, 1814, 1499	952	313, 469, 938, 1797	
Individual on treadmill clear directed	**8**	848	594, 200, 295	596, 444, 298, 1636	469, 313, 1641	0.51, 1.02, 1.53
Individual breathing in microphone	**-**	788	489, 657, 1400-1600	488, 654, 1400-1600	469, 664, 1400-1600	0.84, 0.42, 1.26

D.2 Comparison inspiration and expiration

Figure D.3a shows the time-domain plots of an inspiration and subsequent expiration from the breath-only recording from the healthy individual. A different inspiration-expiration couple is selected than in figure 4.6. Figure D.3b shows the accompanying periodograms.



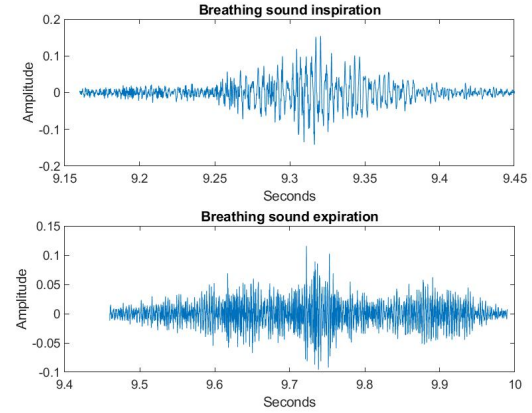
(a) Time-domain plot.



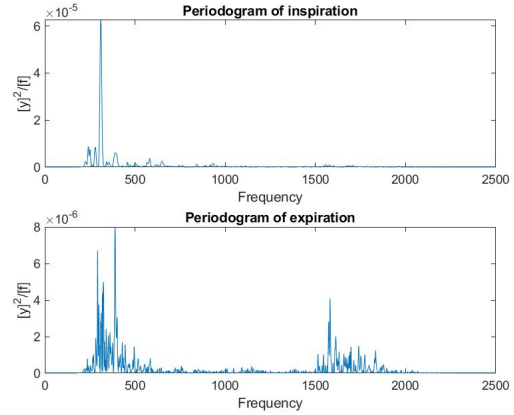
(b) Frequency domain plot.

Figure D.3: Time and frequency domain plots of selected inspiration-expiration couple from the breath only sound of the healthy individual.

Figures D.4a until D.9b show the time and frequency domain plots of two DB segments, two UA segments and two ND segments respectively.

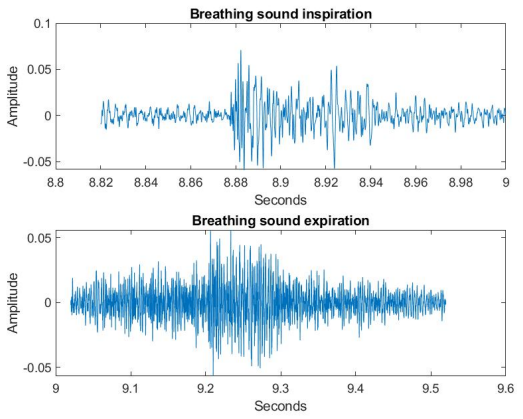


(a) Time-domain plot.

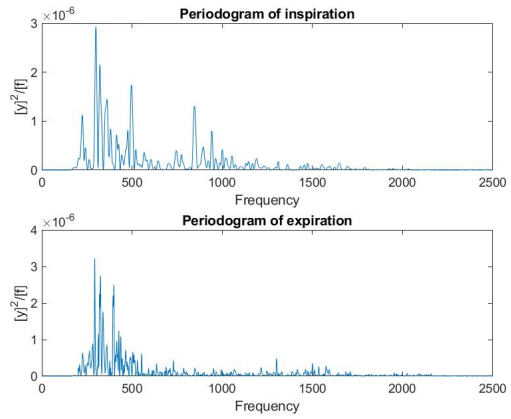


(b) Frequency domain plot.

Figure D.4: Time and frequency domain plots of selected inspiration-expiration couple from first DB segment.

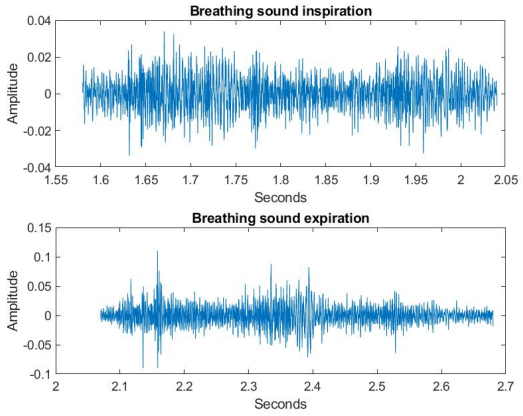


(a) Time-domain plot.

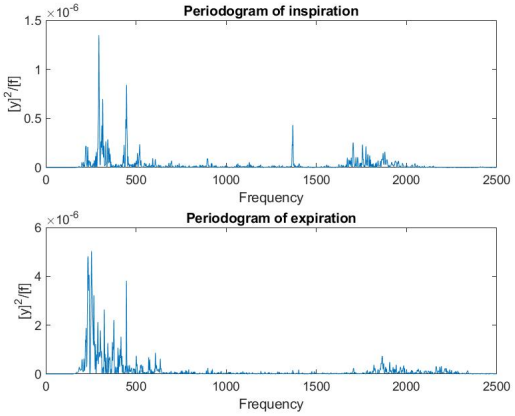


(b) Frequency domain plot.

Figure D.5: Time and frequency domain plots of selected inspiration-expiration couple from second DB segment.

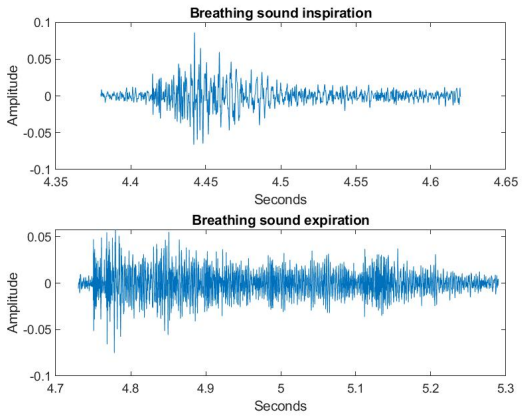


(a) Time-domain plot.

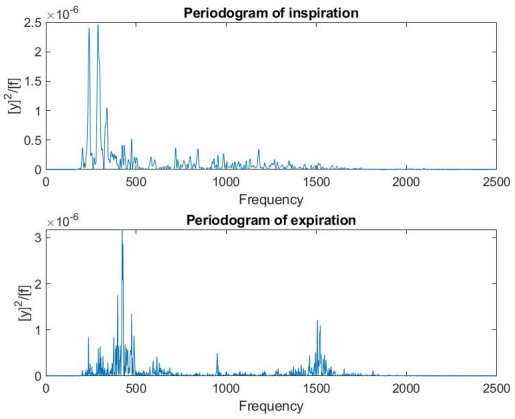


(b) Frequency domain plot.

Figure D.6: Time and frequency domain plots of selected inspiration-expiration couple from first UA segment.

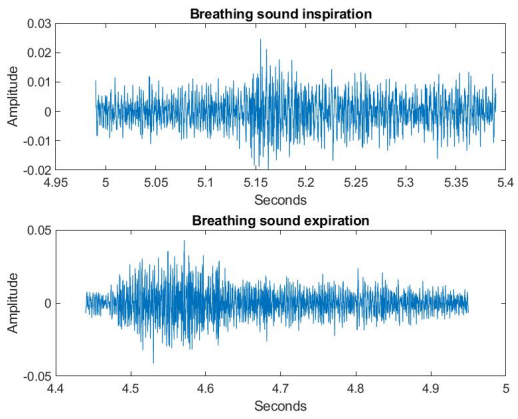


(a) Time-domain plot.

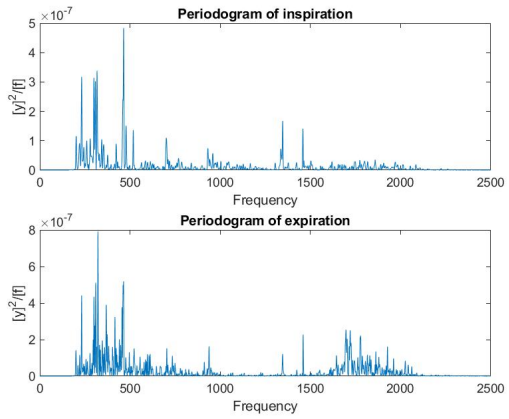


(b) Frequency domain plot.

Figure D.7: Time and frequency domain plots of selected inspiration-expiration couple from second UA segment.

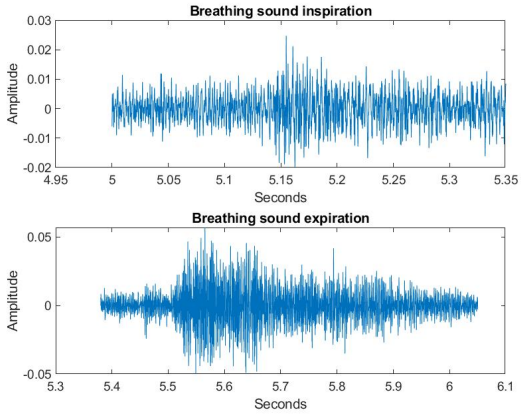


(a) Time-domain plot.

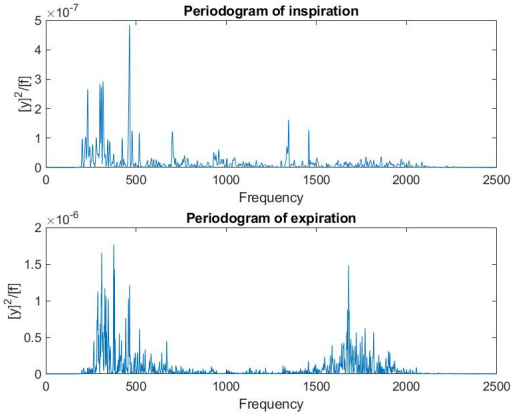


(b) Frequency domain plot.

Figure D.8: Time and frequency domain plots of selected inspiration-expiration couple from first ND segment.



(a) Time-domain plot.



(b) Frequency domain plot.

Figure D.9: Time and frequency domain plots of selected inspiration-expiration couple from second ND segment.

D.3 Median and peak frequencies

In table D.2, the median frequencies and frequencies of the highest peak in the Welch averaged periodogram are shown per segment. DB1 to DB5 refer to dysfunctional breathing patients, UA1 to UA7 refer to uncontrolled asthma patients and ND1 to ND6 refer to patients without an established respiratory diagnosis (no diagnosis).

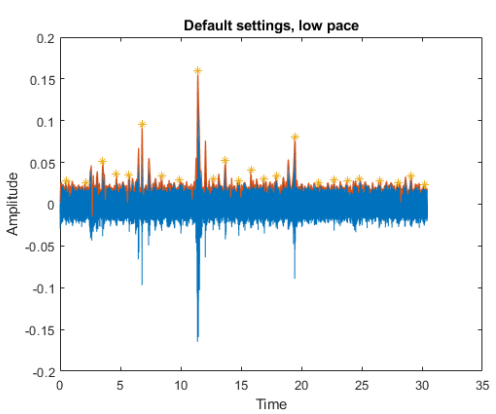
Table D.2: Median frequencies and frequencies belonging to highest peaks in Welch averaged periodogram of all 18 segments.

Patient	Median frequency (Hz)	Highest Welch peak (Hz)
DB1	610	1768
DB2	508	400
DB3	462	308
DB4	896	1704
DB5	527	317
UA1	329	342
UA2	572	181
UA3	1241	444
UA4	933	1602
UA5	599	1724
UA6	827	474
UA7	1346	1768
ND1	452	430
ND2	1010	464
ND3	680	313
ND4	914	1699
ND5	1372	1768
ND6	386	322

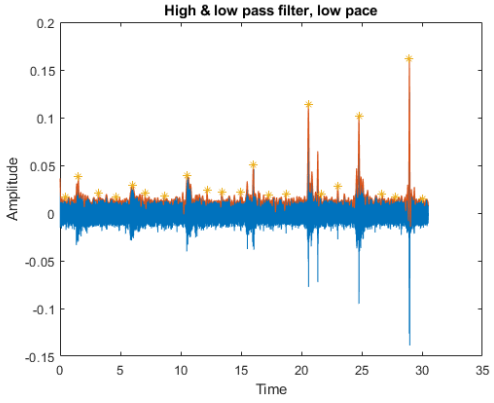
E ADDITIONAL RESULTS ADJUSTMENTS IN MEASUREMENT SETUP

E.1 Results of changing microphone settings

To establish an optimal set of microphone settings, different settings were tried and plots in time domain were compared. Figure E.1 compares the default settings to applying a highpass and lowpass filter, both with a slow treadmill velocity at which the individual on the treadmill could walk. Applying these filters results in a plot with clearer respiratory peaks.



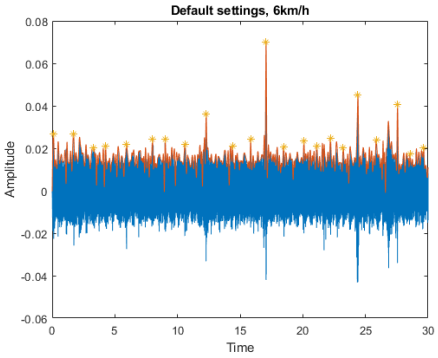
(a) Default microphone settings.



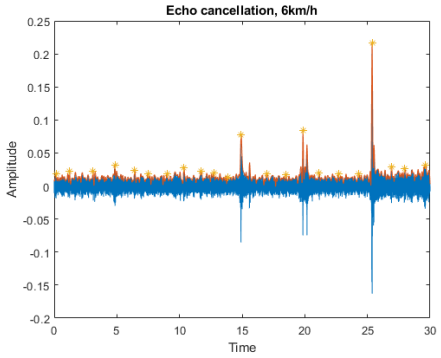
(b) Highpass filter with cut-off at 250 Hz and lowpass filter with cut-off at 2,000 Hz applied.

Figure E.1: Changes in time domain plot with highpass and lowpass filter.

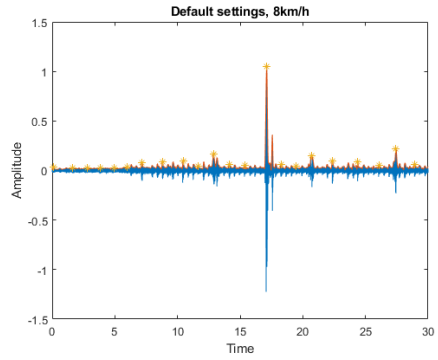
Figure E.2 compares the default settings to applying echo cancellation for different velocities. At 6 km/h, the plot with echo cancellation shows clearer peaks than the plot with the default settings. However, when looking at the 8 km/h plots, the default settings result in a clearer respiratory pattern than when applying echo cancellation. Therefore the performance of the echo cancellation settings is disputable.



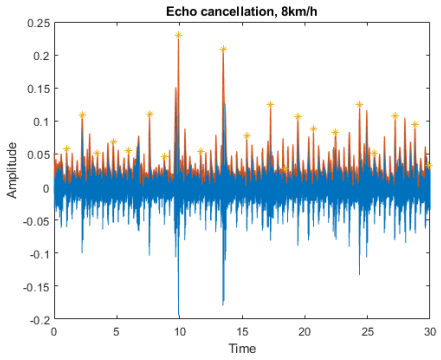
(a) Default microphone settings, 6 km/h.



(b) Echo cancellation applied, 6 km/h.



(c) Default microphone settings, 8 km/h.



(d) Echo cancellation applied, 8 km/h.

Figure E.2: Changes in time domain plot with echo cancellation for velocity of 6 and 8 km/h.

However, when adding noise cancellation to the echo cancellation at 8 km/h, the respiratory pattern becomes even better discernible as is shown in figure E.3.

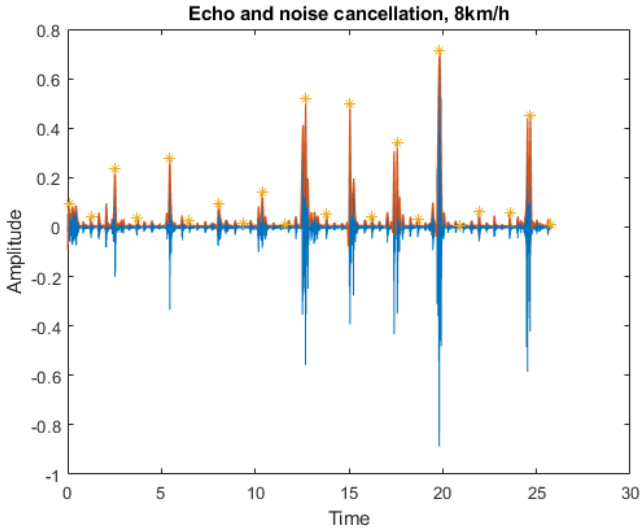
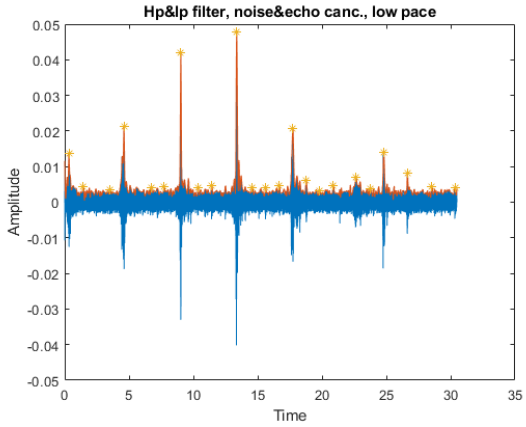
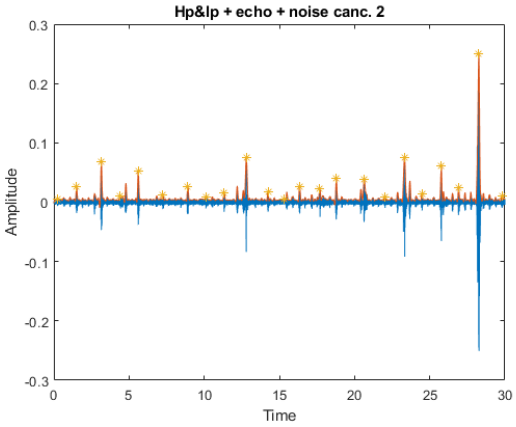


Figure E.3: Echo and noise cancellation applied at 8 km/h.

Since both filtering and applying noise and echo cancellation improved the discernability of the respiratory pattern, they were combined. Figure E.4 shows the result at a low pace and at a velocity of 6 km/h.



(a) Highpass filter, lowpass filter, echo cancellation and noise cancellation applied at low pace.

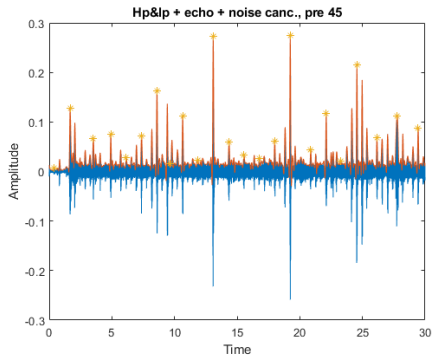


(b) Highpass filter, lowpass filter, echo cancellation and noise cancellation applied at 6 km/h.

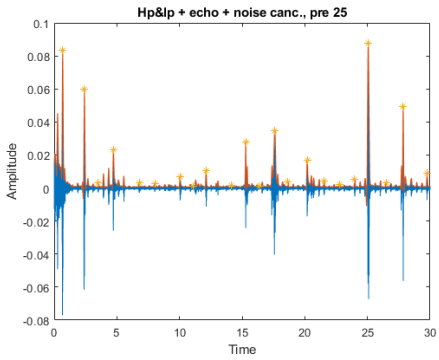
Figure E.4: Result of applying high- and lowpass filters and echo and noise cancellation.

When comparing figure E.4a to E.1b, the application of echo and noise cancellation in addition to the filtering seems to improve the signal. The signal quality of E.4b seems comparable to that of E.3, but a good comparison cannot be made since the velocities differ. However, since E.1 shows improvements due to filtering, it is assumed that filtering in addition to echo and noise cancellation improves signal quality.
Next, different gain levels were tried by changing the 'pre' and 'post' dB values. The default settings were pre=35 dB and post=0 dB. Figure E.5 shows the result of slightly increasing or decreasing these gain levels.

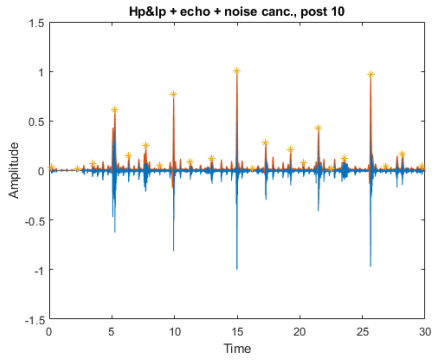
Increasing the pre gain seems to worsen the respiratory pattern. Decreasing this gain does not seem to alter the pattern, but the amplitude is a lot lower so this is not favourable either. Changing the post gain to both sides seems to render a good plot. However, when listening to the resulting signals they sound really 'windy'. Therefore it is assumed that changing this value modifies the sound too much.



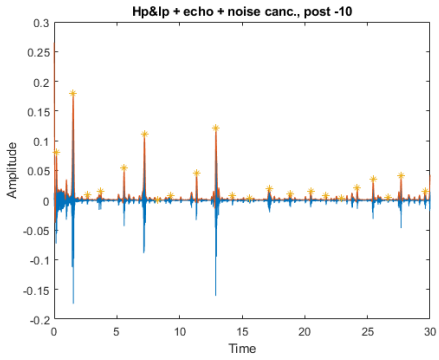
(a) Pre gain increased with 10 dB (to 45 dB).



(b) Pre gain decreased with 10 dB (to 25 dB).



(c) Post gain increased with 10 dB (to 10 dB).



(d) Post gain decreased with 10 dB (to -10 dB).

Figure E.5: High- and lowpass filtering, echo cancellation and noise cancellation with different gain settings at 6 km/h.

Lastly, the 'automatic gain control' was enabled which normally automatically adjust for differences in volume as different people speak. Figure E.6 shows the result of this setting at 6 km/h.

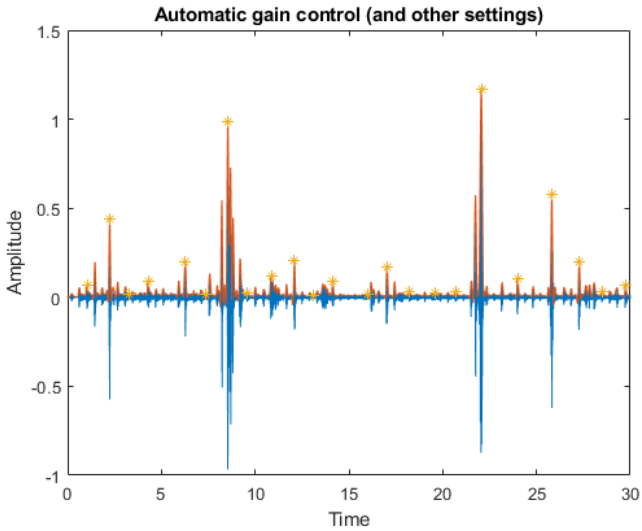


Figure E.6: Automatic gain control in addition to high- and lowpass filtering, echo cancellation and noise cancellation at 6 km/h.

Although the plot seems alright, this setting will not be used for the measurements since the maximum amplitude is reached sometimes, resulting in a deformation of the sound.

E.2 Protocol for improving signal quality

Before performing the measurements, approximately 20 seconds of clear respiration are recorded with a smartphone. This sound recording serves as reference signal for the second measurement. Materials required for the measurements:

- Setup with treadmill and directional microphone
- Smartphone 1 for playing the recorded respiration sound and pacing the respiration
- Smartphone 2 for recording the sound of the treadmill and footsteps on the treadmill
- Smartphone holder
- Stopwatch or extra smartphone with stopwatch function

The phone holder with smartphone 2 placed in it will be mounted to the rod of the treadmill as shown in figure E.7, or to an object placed very close to this rod. The microphone of the smartphone will be directed at the treadmill and feet.

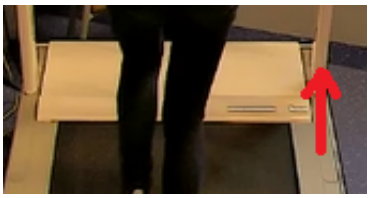


Figure E.7: Placement of phone holder with smartphone 2, indicated by the red arrow.

The following three measurements will be performed:

1. An individual with audible respiration during exercise runs on the treadmill for 2x1 minute, at a velocity of the treadmill which will make the individual exhausted so that the respiration is clearly audible (in this case, 10 km/h). At the start of the experiment, the individual on the treadmill says 'Start' clear and loud. In this way it will be easy to recognize the start of the measurements from both recording devices during analysis. At the same moment, a second individual starts a timer. During the first minute, everyone is quiet (except for the sound of the respiration of the exercising individual). In the next minute, the non-exercising individuals talk now and then so that the influence of speech in the background on the recorded signal can be examined.

2. An individual with inaudible respiration during exercise runs on the treadmill with smartphone 1 in one hand, which will be kept at the same height as the mouth. At the start of the experiment, the individual on the treadmill says 'Start' clear and loud and at the same time starts playing the recorded sound signal on smartphone 1. It is important for this individual to make sure that his or her own respiration is inaudible. When the recording has finished, the individual says 'Stop'. The recording is played another time with again a 'Start' said by the individual on the treadmill at the beginning. This time, the non-exercising individuals talk now and then during the measurement. At the end of the recording, the person on the treadmill says 'Stop' again. This entire measurement will be performed once at a treadmill velocity of 6 km/h and once at 8 km/h.

3. An individual with audible respiration during exercise runs on the treadmill for 1 minute at a velocity of 6 km/h to make sure the respiratory rate will not be too high. At the start of the experiment, the individual on the treadmill says 'Start' clear and loud and

at the same time, the individual starts a stopwatch. The individual will pace his or her own respiration during the experiment, by capturing the time instances of the end of an inspiration and end of an expiration using lap times.

E.3 Results of noise cancellation

Figure E.8 shows the result of applying a Wiener filter with an order of 10,000 to remove noise from the signal. Similar results were obtained for an order of 12 and 1,000.

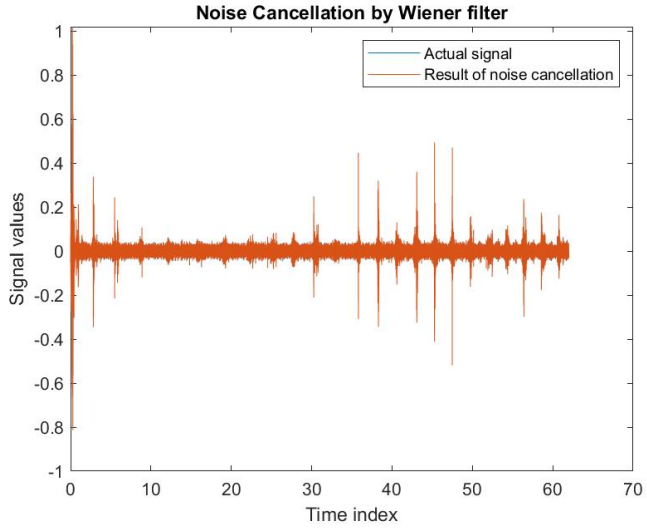


Figure E.8: Improving signal with Wiener filter (order M=10,000)

Figure E.9 shows the result of applying an RLS filter with 250 coefficients to remove noise from the signal. Similar results were obtained with 500 coefficients. Figure E.10 shows the result of

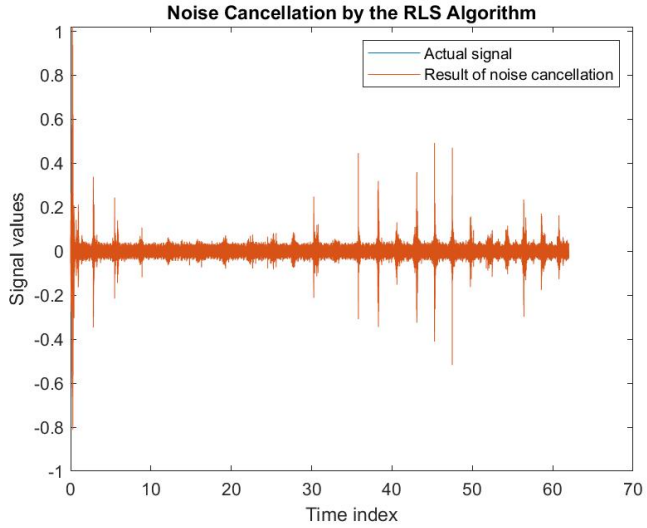


Figure E.9: Improving signal with RLS filter, 250 coefficients.

applying the DWT with a decomposition level of 20.

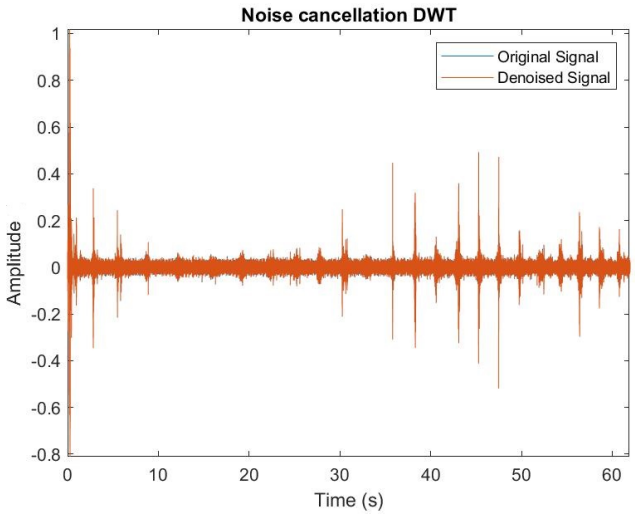
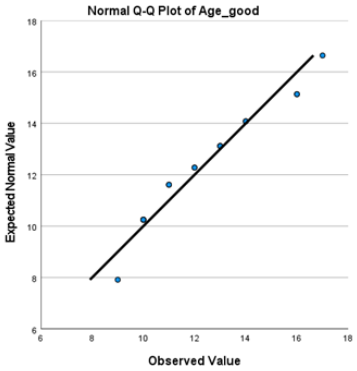


Figure E.10: Improving signal with wavelet denoising

F ADDITIONAL RESULTS MACHINE LEARNING APPROACH

F.1 Baseline characteristics fragments

Table F.1 shows the baseline characteristics of the 43 recordings which were available for machine learning.
To test whether the age and height values in the three quality categories were normally distributed, Q-Q plots were generated and a Shapiro-Wilk test on normality was performed. The results are shown in figures F.1a-F.6b.



(a) Q-Q plot.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Age_good	,162	18	,200*	,917	18	,116

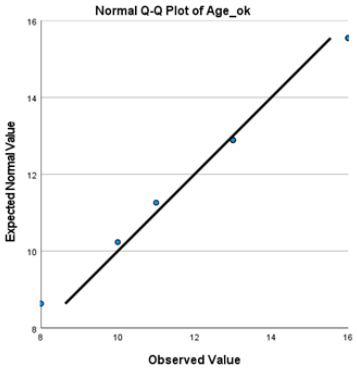
*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

(b) Shapiro-Wilk test.

Figure F.1: Q-Q plot and Shapiro-Wilk test of the age of the patients belonging to the quality=good group. The values in the Q-Q plot follow the y=x line, suggesting a normal distribution. N=18, =5%, c=0.897. To reject the normality hypothesis, the statistic should be smaller than c. This is not the case, so this suggests a normal distribution.

Table F.1: Age, height and determined quality of the 43 recordings with the old microphone settings

Fragment	Quality	Age	Height
DA1	OK/bad	16	157
DA2	OK	16	178
DA3	Good	10	148
DA4	Bad	13	157
DA5	Bad	17	182
DA6	Good	10	153
DA7	OK	10	148
DA8	OK	13	163
DA9	Good	12	176
DA10	Good	16	177
DA11	Bad	11	150
DA12	Good	11	146
DA13	Bad	15	181
DA14	Good	13	161
DA15	OK	16	175
DA16	Good	14	170
UA1	Bad	10	143
UA2	Good	9	145
UA3	OK	11	140
UA4	Good/OK	9	137
UA5	Good	13	156
UA6	Bad	9	137
UA7	OK/bad	9	145
UA8	Bad	10	147
UA9	OK	13	162
UA10	Bad	9	139
UA11	Good	10	140
UA12	OK	13	146
UA13	Good	14	154
UA14	Good	10	145
UA15	Good	11	146
UA16	Good	13	156
ND1	Good	16	186
ND2	Bad	13	158
ND3	OK	8	133
ND4	Good	17	184
ND5	Bad	14	166
ND6	Bad	15	157
ND7	Good	10	152
ND8	Good	12	164
ND9	Bad	15	175
ND10	OK	16	182
ND11	Bad	10	150



(a) Q-Q plot.

Tests of Normality

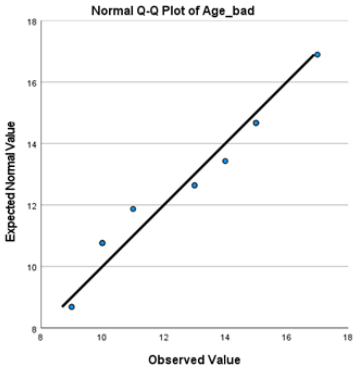
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Age_ok	,196	9	,200*	,901	9	,257

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

(b) Shapiro-Wilk test.

Figure F.2: Q-Q plot and Shapiro-Wilk test of the age of the patients belonging to the quality=ok group. The values in the Q-Q plot follow the y=x line, suggesting a normal distribution. N=9, =5%, c=0.829. To reject the normality hypothesis, the statistic should be smaller than c. This is not the case, so this suggests a normal distribution.



(a) Q-Q plot.

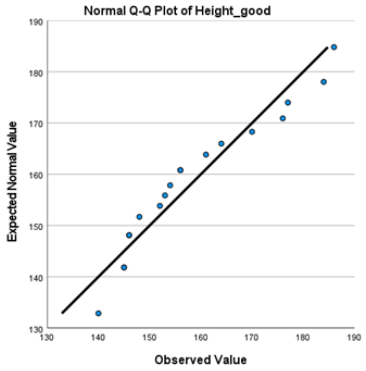
Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Age_bad	,197	13	,180	,908	13	,170

a. Lilliefors Significance Correction

(b) Shapiro-Wilk test.

Figure F.3: Q-Q plot and Shapiro-Wilk test of the age of the patients belonging to the quality=bad group. The values in the Q-Q plot follow the y=x line, suggesting a normal distribution. N=13, =5%, c=0.866. To reject the normality hypothesis, the statistic should be smaller than c. This is not the case, so this suggests a normal distribution.



(a) Q-Q plot.

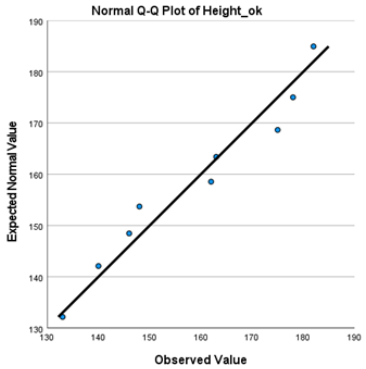
Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Height_good	,190	18	,086	,909	18	,084

a. Lilliefors Significance Correction

(b) Shapiro-Wilk test.

Figure F.4: Q-Q plot and Shapiro-Wilk test of the height of the patients belonging to the quality=good group. The values in the Q-Q plot follow the y=x line, suggesting a normal distribution. N=18, =5%, c=0.897. To reject the normality hypothesis, the statistic should be smaller than c. This is not the case, so this suggests a normal distribution.



(a) Q-Q plot.

Tests of Normality

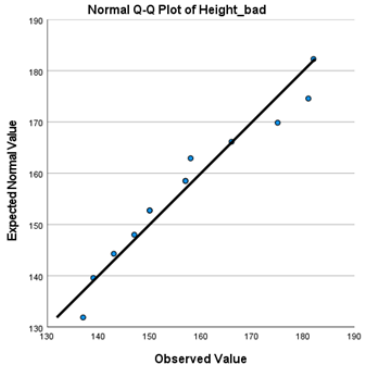
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Height_ok	,169	9	,200*	,936	9	,542

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

(b) Shapiro-Wilk test.

Figure F.5: Q-Q plot and Shapiro-Wilk test of the height of the patients belonging to the quality=ok group. The values in the Q-Q plot follow the y=x line, suggesting a normal distribution. N=9, =5%, c=0.829. To reject the normality hypothesis, the statistic should be smaller than c. This is not the case, so this suggests a normal distribution.



(a) Q-Q plot.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Height_bad	,168	13	,200*	,928	13	,324

*. This is a lower bound of the true significance.

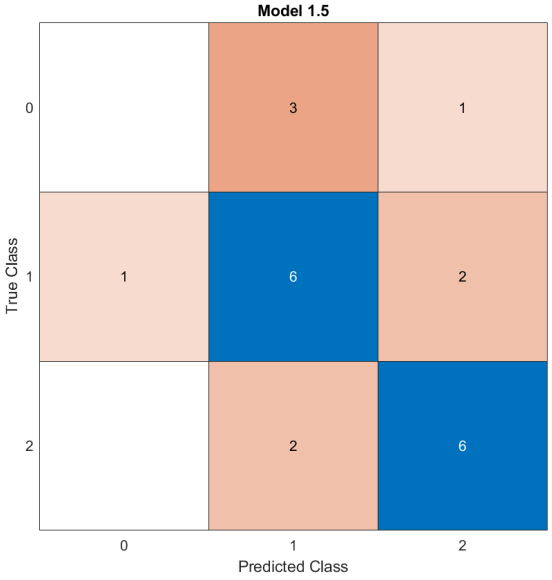
a. Lilliefors Significance Correction

(b) Shapiro-Wilk test.

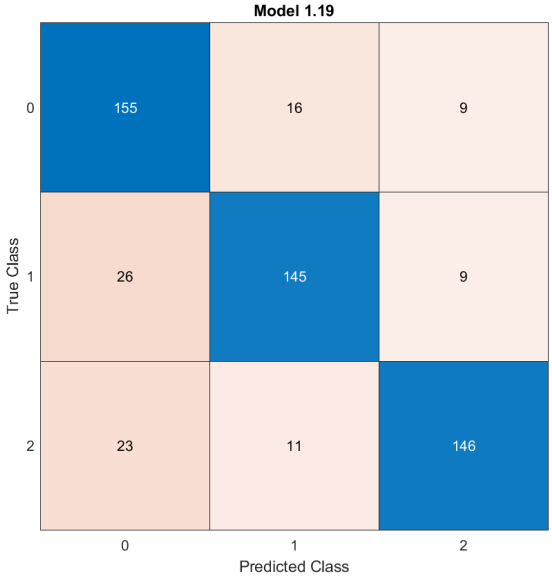
Figure F.6: Q-Q plot and Shapiro-Wilk test of the age of the patients belonging to the quality=bad group. The values in the Q-Q plot follow the y=x line, suggesting a normal distribution. N=13, =5%, c=0.866. To reject the normality hypothesis, the statistic should be smaller than c. This is not the case, so this suggests a normal distribution.

F.2 Confusion matrices training results

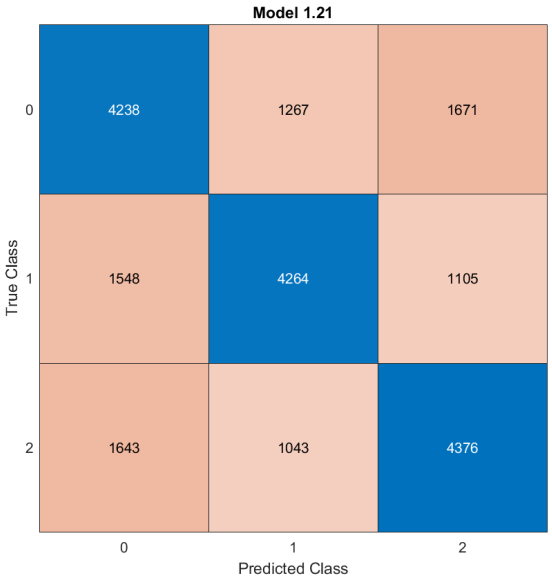
Figures F.7-F.9 show the confusion matrices belonging to the models with the best performance for each data preparation method.



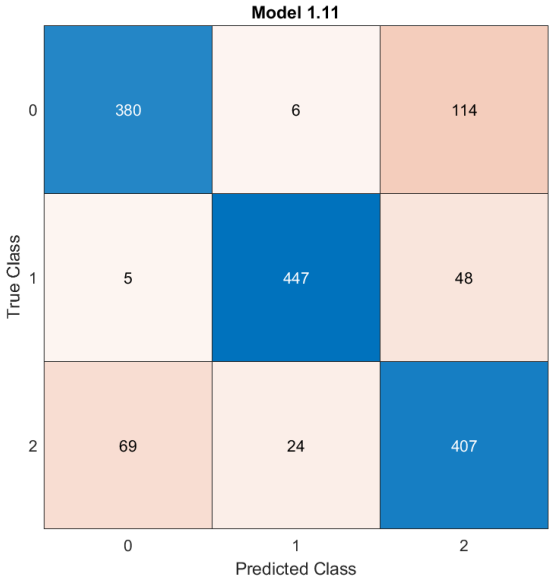
(a) QDA classifier method 1a.



(b) KNN classifier method 2a.

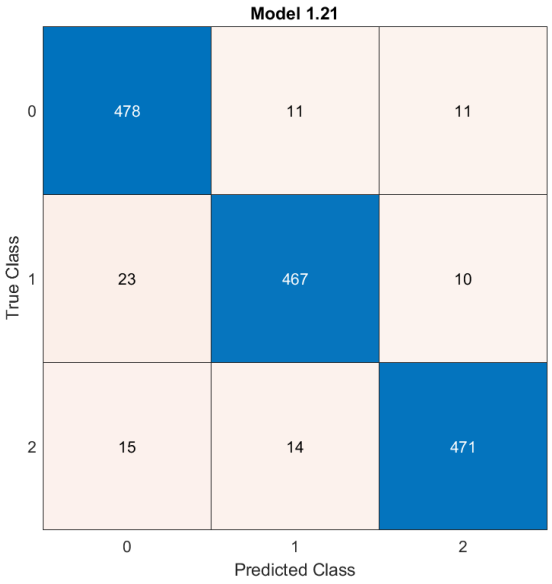


(c) Bagged trees classifier method 2b.

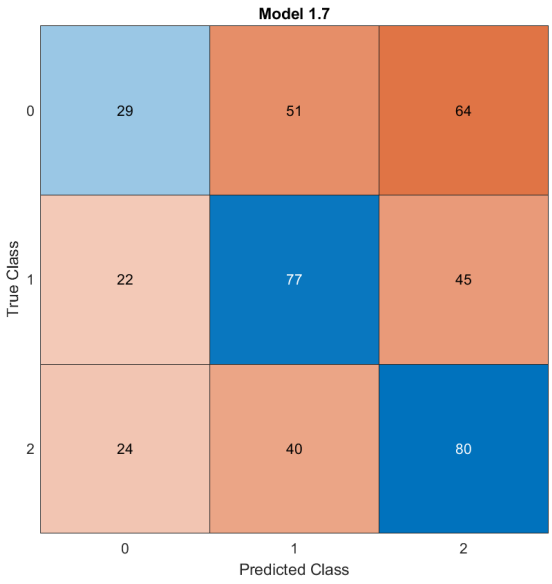


(d) Gaussian SVM classifier method 3a.

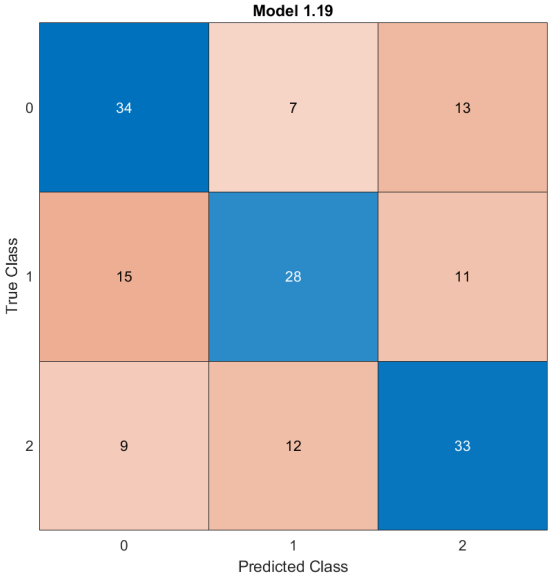
Figure F.7: Confusion matrices for classifiers from data preparation methods 1a, 2a, 2b and 3a.



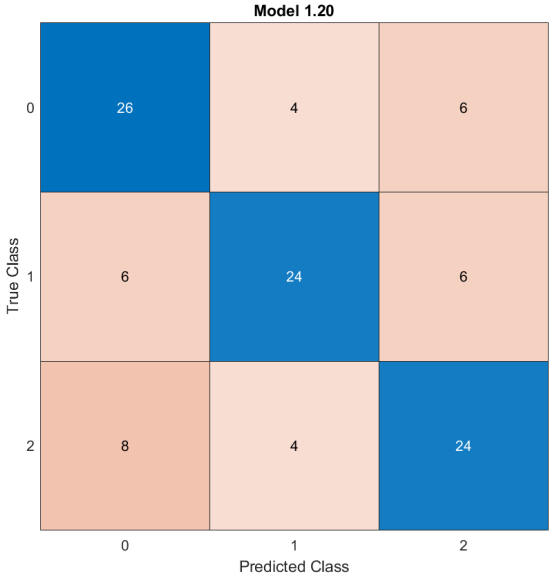
(a) Bagged trees classifier method 3b.



(b) Kernel naive Bayes classifier method 4a.



(c) KNN classifier method 2b.



(d) Boosted trees classifier method 4c.

Figure F.8: Confusion matrices for classifiers from data preparation methods 3b, 4a, 4b and 4c.

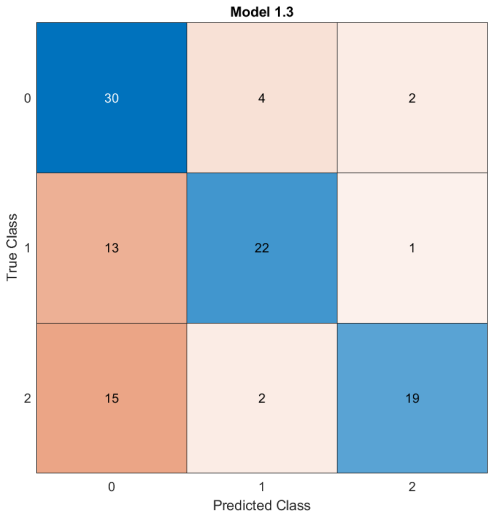


Figure F.9: Confusion matrix of coarse tree classifier on data preparation method 4d.

F.3 Test results

Tables F.2-F.10 visualize the performance of each selected model on the unseen data. Method 1b is excluded because the training accuracy was extremely low. For method 4b, the performance is not shown as this was determined per recording.

Table F.2: Test results of testing model belonging to method 1a on unseen data.

True class → / Predicted class ↓	DB (2)	UA (2)	ND (2)
DB	**2 (100%)**	1 (50%)	1 (50%)
UA	0 (0%)	**1 (50%)**	0 (0%)
ND	0 (0%)	0 (0%)	**1 (50%)**

Table F.3: Test results of testing model belonging to method 2a on unseen data.

True class → / Predicted class ↓	DB (8606)	UA (7053)	ND (1437)
DB	**46 (21%)**	94 (44%)	13 (36%)
UA	81 (38%)	**46 (21%)**	10 (28%)
ND	89 (41%)	76 (35%)	**13 (36%)**

Table F.4: Test results of testing model belonging to method 2b on unseen data.

True class → / Predicted class ↓	DB (8606)	UA (7053)	ND (1437)
DB	**2663 (31%)**	1780 (25%)	719 (50%)
UA	3033 (35%)	**3049 (43%)**	357 (25%)
ND	2910 (34%)	2224 (32%)	**361 (25%)**

Table F.5: Test results of testing model belonging to method 3a on unseen data.

True class → / Predicted class ↓	DB (7)	UA (7)	ND (2)
DB	**3 (43%)**	3 (43%)	0 (0%)
UA	2 (29%)	**3 (43%)**	1 (50%)
ND	2 (29%)	1 (14%)	**1 (50%)**

Table F.6: Test results of testing model belonging to method 3b on unseen data.

True class → / Predicted class ↓	DB (7)	UA (7)	ND (2)
DB	**3 (43%)**	4 (57%)	0 (0%)
UA	0 (0%)	**2 (29%)**	1 (50%)
ND	4 (57%)	1 (14%)	**1 (50%)**

Table F.7: Test results of testing model belonging to method 4a on unseen data.

True class → / Predicted class ↓	DB (245)	UA (210)	ND (72)
DB	**121 (49%)**	72 (34%)	39 (54%)
UA	98 (40%)	**113 (54%)**	20 (28%)
ND	26 (11%)	25 (12%)	**13 (18%)**

Table F.8: Test results of testing model belonging to method 4ci on unseen data.

True class → / Predicted class ↓	DB (108)	UA (108)	ND (72)
DB	**29 (27%)**	32 (30%)	23 (32%)
UA	37 (34%)	**40 (37%)**	24 (33%)
ND	42 (39%)	36 (33%)	**25 (35%)**

Table F.9: Test results of testing model belonging to method 4cii on unseen data.

True class → / Predicted class ↓	DB (63)	UA (54)	ND (18)
DB	**9 (14%)**	5 (9%)	0 (0%)
UA	25 (40%)	**14 (26%)**	3 (17%)
ND	29 (46%)	35 (65%)	**15 (83%)**

Table F.10: Test results of testing model belonging to method 4d on unseen data.

True class → / Predicted class ↓	DB (63)	UA (54)	ND (18)
DB	**13 (21%)**	17 (31%)	0 (0%)
UA	19 (30%)	**6 (11%)**	2 (11%)
ND	31 (49%)	31 (57%)	**16 (89%)**