NTNU
Norwegian University of
Science and Technology

Features for Movement based Prediction of Cerebral Palsy

Parsa Rahmanpour

Master of Science in Engineering Cybernetics
Submission date: June 2009
Supervisor: Øyvind Stavdahl, ITK

Norwegian University of Science and Technology
Department of Engineering Cybernetics

Problem Description

Description:
Shortly after birth, healthy infants exhibit so-called fidgety general movements, while infants who later develop cerebral palsy (CP) lack these movements. A clinical method known as General Movement Assessment (GMA) has been shown to accurately detect the absence (or presence) of fidgety general movements, but for practical reasons this method has not been adopted widely in the clinics.
Previous work at NTNU (Andreas Berg, 2008) has demonstrated the appropriateness of dynamic systems theory and model adaptations for the classification of infant movements as normal or abnormal, respectively. Another group (Meinecke et al., 2006) has successfully applied higher-order statistics, among other things, for the same purpose. In this assignment you will merge the two groups' approaches and investigate the result in order to obtain an optimal classification scheme for the problem at hand.

1. Give a short account of the GMA method and previous attempts at early CP diagnosis based on computer-based movement analysis.

2. Conduct a preliminary assessment of the applicability of the features of Meinecke et al. to the existing movement database, and compare the results with those obtained by Berg (2008).

3. Perform an in-depth study of both feature sets in order to find the optimal combination of features from the resulting collective pool of features. New, related features may be added to the pool if deemed appropriate.

4. Discuss the results obtained in light of the GMA method and ongoing clinical research.

Assignment given: 15. January 2009
Supervisor: Øyvind Stavdahl, ITK

Preface

This thesis is submitted in fulfilment of the degree Master of Science in Department of Engineering Cybernetics, at Norwegian University of Science and Technology (NTNU).

The project consists of relevant literature study, designing algorithms and methods and their implementation in `Matlab`. The idea behind this work is to predict the likeliness of newborn infants being at risk of having Cerebral Palsy. The main drive force behind this thesis is my personal interest in contributing to medical science and leaving a meaningful achievement behind. Typically, during such projects, some questions will be answered properly, while new hidden problems will reveal themselves, creating thrilling topics for future research.

In order to get valuable directions during any kind of research, the presence of a competent and supportive adviser is necessary. In my case, I've been lucky to have **Øyvind Stavdahl** as my supervisor, whom I'm very grateful to. Here, I want to thank him for being very helpful, enthusiastic and always having constructive suggestions.

Furthermore, I would like to thank my supportive and understanding family, particularly, my kind parents **Mohsen** and **Soraya**, my smart brother **Sam** and my uncle **Mostafa Pourbayat** for his role as my personal mentor.

Nonetheless, I should thank **Per Ferdinand Bach** and **Reza Mohseni** for being good friends and playing major roles as my discussion partners in time of need.

Trondheim, 11 June 2009

Parsa Rahmanpour

Contents

List of Figures

List of Tables

Abstract

Shortly after birth, healthy infants exhibit so-called *fidgety movements*, while infants who later develop cerebral palsy (CP) lack these movements. General Movement Assessment (GMA) which is a clinical method, has proven its accuracy in detecting the absence (or presence) of fidgety movements, but for practical reasons, this method has not been adopted widely in the clinics.

In order to create a similar but objective computer-based approach, Berg (2008) and Meinecke (2006) have studied discriminative features based on movement data collected from electromagnetic sensors and video. In this thesis, in addition to evaluation and comparison of previously introduced features, different classification methods have been applied to a suboptimal subset of these features. The results from linear and nonlinear separability analyses of features, confirm that dynamic features have better descriptive capabilities compared to statistically characterized features. Furthermore, it turns out that fidgety movements in the head (neck) and the arms show significant potential in distinguishing normal and abnormal infants, compared to signals from the trunk and the feet.

The achieved results show 86% sensitivity and 90% specificity, which are highly acceptable, but this study needs further attendance before having any clinical usability. This study contains the first step of a typical medical research, meaning that the global (generalized) validity of the implemented methods are yet to be investigated, suppose that a representative selection (data) is available.

1 Introduction

For the last decades, technical advances and improvements in obstetric and neonatal care have led to a decrease in prenatal mortality. Especially among the extremely premature infants, the chance of survival has greatly increased. Despite continued decrease in mortality rates, the incidence of neurosensory and developmental handicaps has remained constant (M.Hack, 2001). This means that a larger number of premature infants survive without major sequeale, but there are also a larger number of prematurely born survivors with a high risk of major handicaps such as cerebral palsy (CP)(L.Adde, n.d.).

In order to limit the consequences of infantile cerebral palsy (ICP), physiotherapy should start as soon as possible. This requires that infants at risk are detected at the earliest age possible. Today, diagnosis is based on visual observation by physicians and as such is influenced by subjective impressions. Objective methods, quantifying the pathological deviation from normal spontaneous motor activity would be preferable as they, for example, allow an inter -and intra-individual comparison of movements.

1.1 Background

1.1.1 Motivation

- To continue the progression in my study, in a way that it is in accordance with my background in Medical Cybernetics.
- A personal interest in the field of pattern recognition. This field stems from the need for automated machine recognition of objects, signals or images, or the need for automated decision-making based on a given set of parameters (R.Polikar, 2006). In this case, the decision-making is to predict if an infant will develop Cerebral Palsy based on collected signals indicating the child's movements.
- A personal wish to use acquired knowledge to climb one more step towards creating a technical and objective possibility for CP prediction, in order to reduce the scope of prospective motor impairment in newborn infants.

1.1.2 Previous groundwork

To be able to use any kind of pattern recognition techniques, the availability of data is absolutely necessary. Movement data from 81 infants was collected in an earlier study and has been used in this project. These movements were registered using 6 sensors. The sensors were placed on wrists, ankles, chest and head, as described in Table 1.

Table 1: Placement of sensors (Berg, 2008)

Sensor nr.	Placement
Sensor 1	Left ankle
Sensor 2	Chest
Sensor 3	Right wrist
Sensor 4	Left wrist
Sensor 5	Right ankle
Sensor 6	Head

Each sensor would measure its position in x, y and z coordinate system. In addition, each sensor would also measure rotation of its appurtenant limb as quaternions. The applied sampling frequency for data collection was 25Hz.

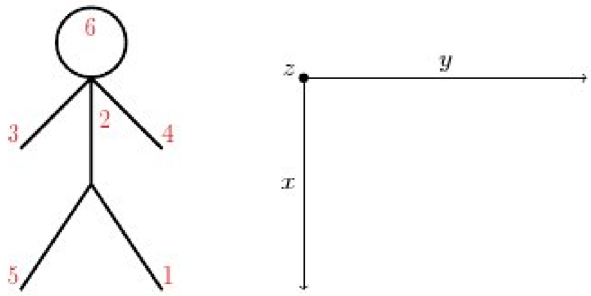


Figure 1: Coordinate system and sensor placement (Berg, 2008)

The movement data from each infant is a combination of several recordings. The interesting and relevant parts of each of recorded measurement, also referred to as *Region Of Interest* (ROI), has been analysed and selected by a physiotherapist. This means that every record contains one or more ROIs. As an example, Fig. 2 displays measured values for all 6 sensors in xy-plane for a simple ROI. It can be concluded that all the data accessed and utilized during this master thesis, is composed of ROIs from infants with normal and abnormal movements. A ROI contains time series data from all sensors in x, y and z directions, in addition to quaternions. The ROIs have different length and could vary from 25 seconds and up to 15 minutes.

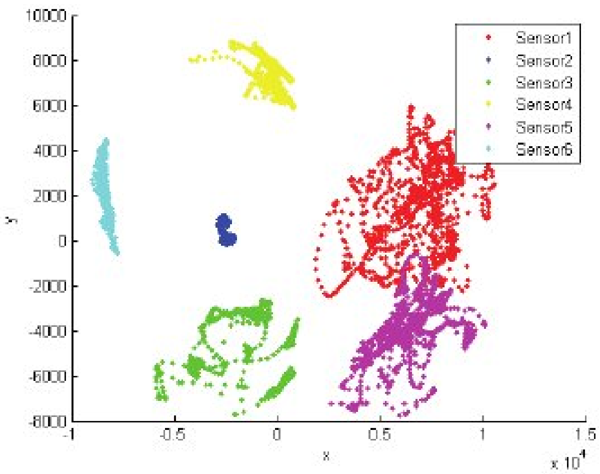


Figure 2: Measurements from different sensors (Berg, 2008). The units for the axes has been unknown and irrelevant to the author. The same reason applies for Fig. 5 and 6

1.1.3 Contribution

The goal of this project is to combine some predefined features gathered from (Berg, 2008) and (Meinecke, 2006), and apply some classification methods, using these combinations. The achieved results will be expressed using criteria like separability, sensitivity and specificity. In order to obtain the aim of the study, the author had to carry through the following points:

- Do a literature study on CP, GMS and previous attempts and comprehend the theory behind them
- Go through and understand the applied methods in Andreas Berg's work and the way they were implemented, and if necessary, modify his code
- Carry through L. Meinecke's approach and understand her procedure
- Implement L. Meinecke's approach and find all of her statistical features
- Compare the results from Berg's and Meinecke's feature extraction based on linear and nonlinear separability
- Implement an optimal or suboptimal selection and combination algorithm
- Apply the algorithm on the available pool of features
- Classification

1.1.4 Structure of the report

This documentation is organized and divided into the following chapters:

Chapter 1: Introduces the task and background for the project and enlightens the reader about the previous works and the attempts made today

Chapter 2: Gives the sufficient theoretical insight to understand the purpose of this research, while explaining related attempts in the field of this project

Chapter 3: Demonstrates construction and interpretation of the study, and clarifies the goals of this master thesis

Chapter 4: Introduces the extracted features and their origin. Furthermore, this section describes the methods and the way they have been implemented

Chapter 5: Represents the achieved results and points out the best results and observations

Chapter 6: The important choices made throughout the study and the achievements illustrated in previous chapters, are discussed in this section. In addition, possible **future works** have been suggested and emphasized

Chapter 7: The most essential works and results accomplished during the project are summarized

1.2 Cerebral Palsy

CP is a non-progressive motor impairment syndrome secondary to lesions or anomalies of the brain arising in the early stages of development. The type of motor impairment is divided into different categories according to which functions or body parts that are affected. The seriousness of CP differs from almost invisible disability to a serious handicap. Although the brain injury is a non-progressive one, the clinical picture of CP is changing with increasing age of the individual (E.Beckung, 2002).

As a consequence of the impairment, normal development and formation of the central nervous system is retarded. The general clinical symptoms of CP are characterised by very different, complex dysfunctions, such as tonus, strength, course of motion and muscle co-ordination as well as other brain functions like speech, vision or mental capabilities. CP affects approximately 1 in 500 infants. The risk of CP is highest in extremely premature infants (birth weight less than 1kg and/or gestational age less than 28 weeks) (L.Adde, n.d.).

1.3 Today's approach

As a result of the complexity of neonatal brain development and the different risk factors in different developmental stages, it is difficult to predict the neurological outcome in young infants. A diagnosis of CP is often not established until the age of 12-18 months (F.Palmer, 2002) and some of the mildest forms may still not be diagnosed before the age of four. Some authors believe that early diagnosis of neurological development disorders is important (H.F.Prechtl, 1997*a*,*b*; M.Hadders-Algra, 1996). The aim is to identify as early as possible those infants who require early intervention for suspected neurological abnormalities.

A number of techniques have been used to assess the brain at an early age. The techniques vary from clinically based methods requiring no equipment, such as various forms of neurological assessments tests, to sophisticated technical assessments, such as brain imaging (ultrasound, computer tomography and magnetic resonance imaging) and neurophysiologic tests, including electroencephalograms (EEG) and visual or sensory evoked potentials.

The introduction of ultrasound (US) techniques and magnetic resonance imaging (MRI), have contributed to a better and earlier diagnosis of brain defects in neonates. However, normal cerebral MRI and US can be found in infants who later develop abnormally and vice versa. The accuracy of the different assessment techniques to predict the neurological outcome of newborn babies at risk shows a large variation. In addition to lack of accuracy in prediction of the result, some of the methods need advanced technological equipment (L.Adde, n.d.).

As explained earlier, the spontaneous movement of an infant is dependent on development level of its motor coordination and can be related to whether or not the infant has CP. In order to use this knowledge, a special type of spontaneous movement has been studied that has been termed general movements (GM). In diagnosing a developing spasticity, the physician is usually dependent on his subjective visual observations of the baby's GMs and traditional neurological examinations. This is termed the general movement assessment technique (GMA) (see Section 2.2) and has shown promising scientific outcomes. Typically, the procedure starts by video-recording infant's movement. Then the video-record is observed by a doctor, physiotherapist and etc. Observation and classification of such movement patterns may predict later neurological results as CP as early as 3-5 months post term. In research settings, this method has resulted in a diagnosis of CP with a sensitivity and specificity of about 93% (L.Adde, n.d.).

However, most methodologies available, although some of which are quantitative in nature, do not consider the multitude of potential movement parameters available for an objective classification of patients regarding whether they are healthy or at risk. Quantitative procedures are strongly dependent on the relevant movement parameters discriminating reliably between healthy and at-risk patients (Meinecke, 2006).

2 Theory

The young human nervous system endogenously, i.e. without being constantly triggered by specific sensory input, generates a variety of motor patterns which has been known for more than a century. In the observation of infants the interest has changed from the analysis of the capacities to respond to a manifold of sensory stimulations, to the observation of the un-stimulated infant. Naturalistic observations led to the conclusion of the dominance of spontaneous behaviour, i.e. behaviour not generated by sensory stimulation (H.Prechtl, 2004).

2.1 General Movements

General movements (GMs) are a distinct movement pattern carried out spontaneously. Unlike reflexes, spontaneous movements are patterns of movements, which are not initiated by any obvious external stimuli. They occur in high frequency starting around 7 weeks of gestation and disappearing around 12 month of age (H.F.Prechtl, 1984). GMs are helpful in the early diagnosis of an impaired central nervous system and the specific prediction of later neurological deficits. Observation of the infant's GM or the so-called general movement assessment technique (GMA), has shown promising scientific results. Heinz Prechtl studied motor activity in the human fetus and newborn infants over many years. He claimed that the quality of spontaneous movements, especially the quality of GMs accurately reflects the condition of the nervous system of the fetus and the young infant. GMs involve the whole body in a variable sequence of arm, leg, neck and trunk movements. They wax and wane in intensity, force and velocity. Rotations along the axis of the limbs and slight changes in the direction of movements make them fluent and elegant and create the impression of complexity and variability (H.F.Prechtl, 1997*a*).

2.1.1 Preterm General Movements

GMs observed before term are called foetal or preterm GMs. There is no noticeable difference between foetal and preterm GMs, indicating that neither the increase of force of gravity after birth nor maturation has an influence on the appearance of GMs. The preterm GMs have typical characteristics like fast speed and large amplitudes (H.Prechtl, 2004).

2.1.2 Writhing Movements

GMs at term age and until 6 to 9 weeks post term age are called writhing movements. Writhing movements are defined as repetitive wormlike movements of the limbs and fingers due to brain lesion. They are distinguished by slow to

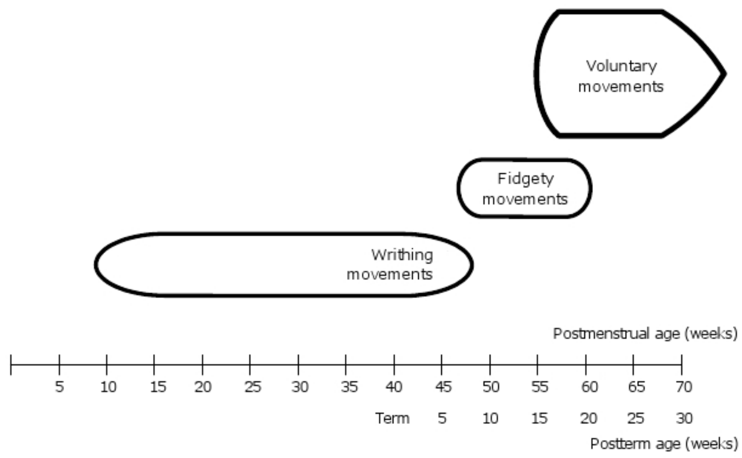


Figure 3: Developmental course of general movements (GMs). Foetal GMs do not change in form at birth (preterm or at term). There is some overlap between writhing and fidgety GMs at about 6- to 8-weeks postterm. At 12- to 15-weeks postterm infants start with voluntary movements e.g. manipulating objects, cooing vocalization, and antigravity movements (H.F.Prechtl, 1999)

moderate speed and small to moderate amplitude. Typically, they are ellipsoid in form, which creates the impression of a writhing quality (H.Prechtl, 2004).

2.1.3 Fidgety Movements

At 6 to 9 weeks post term age, writhing movements gradually disappear while fidgety GMs gradually emerge. Fidgety movements are defined by Prechtl et al as an ongoing stream of small, circular and elegant movements of neck, trunk and limbs. These movements are characterised by moderate speed and variable acceleration in all directions. Fidgety movements can be observed best when the infant is awake, alert and either lying supine or sitting reclined in a baby seat (H.Prechtl, 2004). Studies carried out by Ferrai et al and Prechtl et al, documented that the abnormal fidgety movements in preterm infants with brain lesions and in asphyxiated full term infants predicts later neurological impairment (L.Adde, 2004).

2.2 General Movement Assessment

The qualitative assessment of GMs is a method, which takes into account fully the complexity of the nervous system and at the same time fulfils the requirement of being not so time consuming at all. In addition, it is totally non-intrusive and can even be applied under intensive care conditions in very sick infants, when neurological examinations cannot be carried out.

This is an empirical method which requires physicians observing GMs of infants and predicting if there is an indication of later development of CP. It is usually achieved by video recording the infants periodically, which greatly assists detailed analysis, and searching for abnormalities in different types of GMs. Based on the observation, the physician can classify the infant as healthy or at-risk. There are two specific features of GMs that reliably predict the later neurological outcome of CP (H.F.Prechtl, 1999):

- "A persistent pattern of cramped-synchronized GMs. These GMs appear to be rigid and they lack the normal smooth and fluent character. All limb and trunk muscles contract and relax almost simultaneously. If this pattern exists over several weeks during preterm and term age, spastic CP will develop at a later age."
- "The second specific predictor is the absence of GMs of fidgety character or the so-called fidgety movements. Their absence predicts CP with a sensitivity of 95% and a specificity of 96%."

The GMA classification will then be compared with the later neurological outcome.

2.3 Previous Attempts on Computer-based Movement Analysis

Some work has been conducted in this area by Meinecke et al, as disclosed in (Meinecke, 2003) and (Meinecke, 2004). During these documents, an estimation method for predicting whether an infant is likely to have CP or not has been developed based on real-world indications from 3D movement data of the baby. In order to identify movement features that are cable of discriminating between healthy and at-risk infants, experienced physicians have been consulted. This resulted in extraction of 125 parameters based on movements or combinations thereof were determined which are relevant to CP. Such relevant features like movement speed, trajectory smoothness, periodicity, range of motion and acceleration. In this approach, in order to estimate whether or not the infant is healthy or at-risk, parameter data for babies which have already been classified by a physician are used to select optimum parameter combinations. For example, five optimum parameters are selected using cluster analysis based on Euclidian distances. To estimate the risk of having CP in an infant, these five parameters are measured for the baby, and it is then determined whether they are within the range of standard deviation for the norm collective in respect of each parameter. Depending on the number of parameters within or outside the standard deviation, classification is effected (L.Adde, n.d.).

Furthermore in 2006, Meinecke et al carries out a similar approach to her previous work, with the title "*Movement analysis in the early detection of newborns at risk for developing spasticity due to infantile cerebral palsy*" (Meinecke, 2006). Based on a 3D motion analysis system, the aim of the study was to develop a methodology which allows for the objective and quantitative description of unconstrained spontaneous movement in newborns. The identified movement parameters reflect those factors used by the clinicians during visual assessment of the baby's movement. These parameters quantitatively describe the differences between healthy and affected participants. Subsequently, optimized parameter combinations had to be found to categorize the participants' movement into homogeneous classes entitled *healthy* or *at-risk*, respectively, using an adequate classification procedure.

Twenty-two infants, 15 healthy full-term and seven high-risk pre-term infants took part in the study. All infants were clinically examined by use of ultrasonography and cranial computer tomography. Pathology of the *at-risk* patients was ensured through ultrasound-based detection of cerebral haemorrhage and/or follow-up examinations for two years.

3D motion analysis was performed using a commercially available *Vicon 370* motion analysis system. It is a passive detection system, allowing the contact-free capturing of an arbitrary number of reflecting markers with a temporal resolution of 50Hz and a high spatial precision. The kinematic biomechanical model used to describe the relation between the marker trajectories and the child's movement, is a full body model. It is based on the rigid segment approach in which each segment is assigned to one bone and consists of segments for the hand, forearm, upper arm, head, trunk, thigh, lower leg and foot.

Besides establishing a procedure for reliably retrieving the 3D movement analysis data of newborn babies, the second aim of the study was to extract those parameters of babies' movements that would best describe the difference between healthy and affected participants. To acquire the parameters, five experienced physicians in the field of neuropediatrics were asked for their methodology of visual assessment. In the next step the visual criteria stated by the physicians were reviewed for their applicability in parameter extraction and computer-based evaluation. Additionally, several physical parameters were found, which so far were not used in this clinical context. These parameters, such as skewness, cross-correlation, Moving average and periodicity are of a statistical and mathematical nature. Altogether 53 quantitative parameters could be extracted from the patient's data. In order to select the most significant parameters, all possible combinations of 8 out of the 53 parameters in total were tested for their selectivity. Cluster analysis was performed, utilising Euclidian distances, to find the one combination of parameters that best discriminates between healthy and affected participants. For classification, quadratic discriminant analysis was applied. The capability of quadratic discriminant analysis together with optimized parameter combination was tested for measurements of 11 healthy and 3 affected children that were part of an evaluation database. Measurements that were already part of the training database were detected with a specificity, sensitivity and overall detection rate of 95% each. Measurements of the evaluation database were classified with a positive predictive value of 30%. The negative predictive value reached 100%. Sensitivity also was 100%, whereas specificity only reached 70%. Finally, the overall detection rate (accuracy) was 73%.

In 2008, Andreas Berg disclosed a computer-based movement analysis in his master thesis, titled as "Model-based classification of infant's movements". The aim of the study was to use dynamic models to classify infants as healthy or at-risk, based on their movement patterns. Movements were measured using 6 sensors placed on ankles, wrists, trunk and forehead. Each sensor measured its position and rotation in 3D, but only the position data was processed during the project. As part of the feature extraction, *Principal Component Analysis* (PCA), which is a dimension reduction approach, was used to calculate parameters along the dimensions with most variance. In order to describe the movements, linear models with white noise as the input were applied.Both *Single Input Single Output* (SISO) and *Multiple Input Multiple Output* (MIMO) were considered while using *Autoregressive* (AR) and *Autoregressive Moving Average* (ARMA) as models. The parameters that were calculated by these models were considered as features during classification.

Before classification, the separability of parameters and their combinations was evaluated using *Bayes classification rule* and *Scatter matrix*. The achieved feature vector was classified using *Linear Discriminant Analysis* and *K-nearest neighbours*. The results showed that the movements of arms and head are most significant in discriminating between healthy and at-risk participants. By only considering the head's movement in y direction, side to side movement, 78.26% specificity and 78.64% sensitivity was achieved. By combining the parameters with high selectivity, a specificity of 90.91% and a sensitivity of 85.71% were accomplished.

A methodological modular framework was presented in "*Automated feature assessment in instrumented gait analysis*", by Sebastian Wolf. The goal was automated assessment of gait patterns. The processing steps of data selection, gait parameter calculation and evaluation were not limited to a specific field of application and were largely independent of case-based clinical expert knowledge. A set of 3670 parameters was ranked by relevance for classification of a group of 42 diplegic cerebral palsy patients and an age-matched reference group.

The novel approach described in this paper partly eliminates the limitations of manual procedures in evaluating gait and has produced a semi-automated modular system for objective analysis of gait data. This new method semiautomatically replicates results from previous studies which were obtained using conventional methods and features were extracted which have not been considered so far, but which emphasize additional group specific differences. The new method was put in context with the normalcy index as an important alternative approach for assessment of CP gait. The normalcy index yields a scalar number in terms of a distance measure, which makes it a very intuitive and practical tool for research and clinical applications.

Since the method here uses clinical expert knowledge on a more general level than the normalcy index does, it can be assumed that it has the potential to be transferred in an objective way to different kinds of pathologies. The new method needs to be demonstrated in more detail in future work.

2.4 The validity of the GMA

When introducing an assessment technique, the effectiveness of the method must be considered carefully. This could be measured based on the following criteria:

- "How accurate is the method when it comes to detecting disease positive, i.e. later neurological deficits."
- "How accurate is the method at excluding disease negatives, i.e. those who don not have later neurological deficits."

The conventional indicators employed to determine these points are sensitivity and specificity.

Sensitivity is the number where both disease positive and test positive individuals are divided by the number of those who are disease positive. In other words, it is the percentage of cases which are correctly identified as high-risk for later neurological impairment. Specificity is the number which both disease negative and test negative patients are divided by the number of disease negative participants. In other words, it is the percentage of cases which are correctly identified as normal (H.Prechtl, 2004).

As a simpler definition containing information from both sensitivity and specificity, the *Youden Index* is demonstrated as follows:

$$Y = Sensitivity + Specificity - 1 \tag{1}$$

3 Aim of the study

The goal of this thesis is to comprehend the importance of the General Movement Assessment (GMA) method and the GMA-based features extracted by Berg (2008) and Meinecke (2006). In addition to implementing the suggested features, they must be evaluated and compared. For this purpose, techniques like Scatter matrix and k-means clustering shall be studied and applied. Furthermore, an algorithm has to be designed in order to carefully select and combine the implemented features. The performance of the algorithm is expected not to be divergence. Subsequently, Linear Discriminant Analysis and Quadratic Discriminant Analysis have to be used as classifiers, in order to assess how optimal the feature combination has been.

Finally, the achieved classification results should be introduced and discussed. Among the discussion topics are the choice of methods, their performance and the clinical perspective and usability of this project.

4 Method description and implementation

The present project provides a method of categorising data derived from the movements of living subject, comprising: processing the data to extract information and classify the extracted information into one of a plurality of categories using a classification model, wherein the classification model is trained using data derived from the movements of other subjects whose category is known.

The extracted information relates to patterns of movement which may, or may not, be readily recognisable to a human observer. Since this approach does not involve or require these patterns to be defined or recognised as such, the method is not dependent on particular human-defined parameters. Instead, the information is extracted from the movement data in order to classify data. Thus, the present technique can take into account movement phenomena that are otherwise incomprehensible to humans (or at least not readily recognisable or describable), for example because they involve complex inter-relationships of the movements of a plurality of limbs.

This chapter describes the approaches applied in order to carry out feature extraction, feature selection and classification.

4.1 Feature Extraction

Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm which overfits the training sample and generalizes poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy (Wikipedia, 2009*a*).

Based on the achieved results from Meinecke et al. and Andreas Berg, only the most appropriate features are chosen to be implemented in this project. Meinecke et al. proposed that features like *skewness*, *cross-correlation*, *moving average* and *periodicity* were powerful choices, while Andreas Berg applied *Principal Component Analysis* and different types of *Autoregressive models*.

4.1.1 Skewness

To evaluate the distribution of velocity, the statistical parameter skewness was used. In probability theory, skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable. Therefore, in case of a normal distribution, skewness is zero. Several types of skewness are defined, the terminology and notation of which are unfortunately rather confusing. "The" skewness of a distribution is defined in (Wolfram MathWorld, 2009*b*) to be

$$\gamma_1 = \frac{\mu_3}{{\mu_2}^{3/2}} \tag{2}$$

where μ is the i^{th} central moment. This equation can be simplified using a sample of n values with g_1 as an estimator for the skewness γ_1

$$g_1 = \frac{\frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^3}{(\frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2)^{3/2}} \tag{3}$$

where x_j is the j^{th} value and $\bar{x}$ is the sample mean. Furthermore, figure 4 illustrates positive and negative skewness. Positive skewness means that the distribution has a long tail in the positive direction. The long tail in the negative direction indicates a negative skewness. Thus, skewness can be explained as if both tails of one distribution have different length.

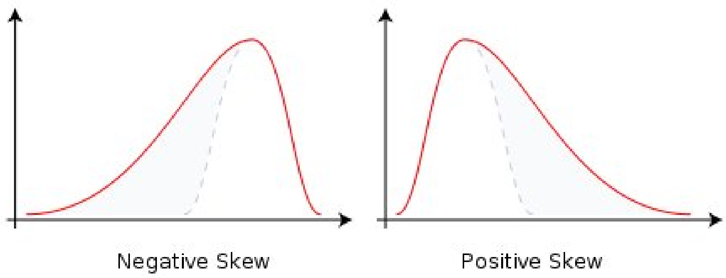


Figure 4: Skewness in a distribution (Wikipedia, 2008)

Skewness was calculated for the absolute value of velocity of head, trunk, arms and feet using the function `skewness` in Matlab.

4.1.2 Cross-correlation

Cross-correlation is a measure of similarity of two waveforms as a function of time-lag applied to one of them. Based on the cramped synchronized movements observed by Prechtl et al. the movement of the end effectors with regard to their possible correlation was examined. The aim was to determine whether movement of certain markers proceeds in the same direction at the same time (Meinecke, 2006). In this case, the cross-correlation was calculated based on absolute value of velocity for single trajectories with zero lag. As a result, the cross-correlation coefficient between the markers of left and right foot, and left and right arm. Equation 4[1] describes cross-correlation between the markers for left foot and right foot

$$CC_{x,LFoot-RFoot} = \frac{\sigma^2_{x,LFoot-RFoot}}{\sqrt{\sigma^2_{x,LFoot} \cdot \sigma^2_{x,RFoot}}} \tag{4}$$

$$\sigma^2_{x,LFoot-RFoot} = \frac{1}{n-1} \sum_{i=1}^{n-1} (x_{LFoot_i} - \bar{x}_{LFoot})(x_{RFoot_i} - \bar{x}_{RFoot}) \tag{5}$$

$$\sigma^2_{x,LFoot} = \frac{1}{n-1} \sum_{i=1}^{n-1} (x_{LFoot_i} - \bar{x}_{LFoot})^2 \tag{6}$$

where n is the number of samples. It should be noted that equation 6 is analogous for $\sigma^2_{x,RFoot}$.

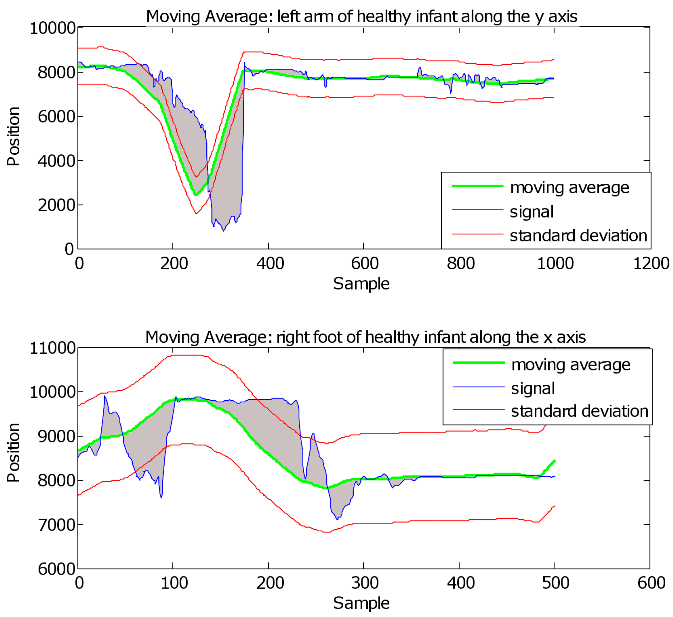
4.1.3 Area out of standard deviation of moving average and area differing from moving average

The movement of healthy children, in contrast to that of affected children, is characterised by smooth and harmonious marker trajectories of the end effectors. Since the signals contain amplitudes with many different lengths, it is not easy to compare them. Thus, by applying moving average, a continuous mean with regard to the past and future values of the signal, a smooth and comparable signal can be obtained. Usually, moving average is used to analyse a set of data points by creating a series of averages of different subsets of the full data set. So a moving average is not a single number, but it is a set of numbers, each of which is the average of the corresponding subset of a larger set of data points.

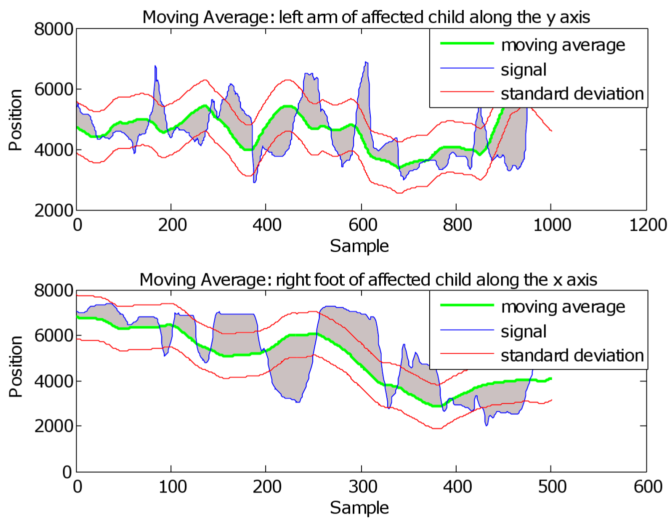
With another word, consider a window with *k* as its width. Furthermore, the average of contents of this window is continuously calculated, while the window moves along the signal. The windowing with *k* is of vital importance: it has to be large enough to calculate deviations between the marker trajectory and the moving average. On the other hand, if *k* is made too large, the analysis will become diffuse. Meinecke et al. chose $k = 99$ and the same value was used in this project.

In a further step, the standard deviation of the trajectory from its moving average within the windowing width *k* was calculated. The divergence of a trajectory's movement from that same trajectory's moving average value could then be quantified in two ways:

[1] There was a misprint in what Meinecke et al. introduce as $CC_{x,LFoot-RFoot}$ where the output of the equation 5 was in power of 3 instead of 2.



5a: Normal



5b: Abnormal

Figure 5: Area differing from moving average

- The first approach quantifies the area in which the trajectory simply differs from the moving average as illustrated in Fig. 5. As there will always be at least a small deviation of the trajectory from the moving average, this parameter tends to be greater than zero at all times.

- The second approach takes only those areas into account in which the trajectory is out of the standard deviation of the moving average as depicted in Fig. 6.

In contrast to the first approach, the parameter from the second approach will only rise as a result of high deviations from a smooth movement (near the moving average).

The moving average $\tilde{x}_j$ (here examplarily for x) of uneven order *k* is defined (Meinecke, 2006) as follows:

$$\tilde{x}_j = \frac{1}{k} \sum_{i=j-\frac{k-1}{2}}^{j+\frac{k-1}{2}} x_i \quad \text{with } j = \frac{k+1}{2}, \cdots, n - \frac{k-1}{2}. \tag{7}$$

Summing up the differences between moving average and trajectory for each sample:

$$\Delta_{s,x} = \sum_{j=\frac{k+1}{2}}^{n-\frac{k-1}{2}} |x_j - \tilde{x}_j| \quad \text{(analogous for y and z axes)} \tag{8}$$

Normalizing the differences between moving average and trajectory on measurment length:

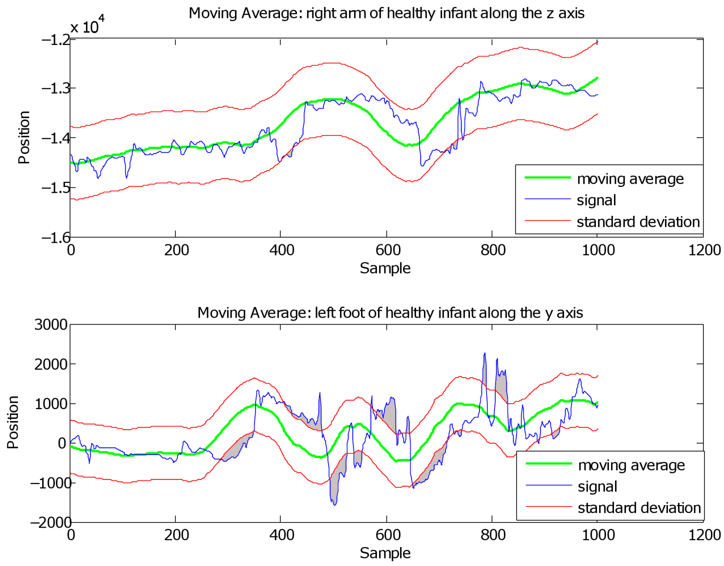
$$\eta_{s,x} = \frac{\Delta_{s,x}}{n-k} \quad \text{(analogous for y and z axes)} \tag{9}$$

Merging the calculated areas of all three spatial axes to one parameter:

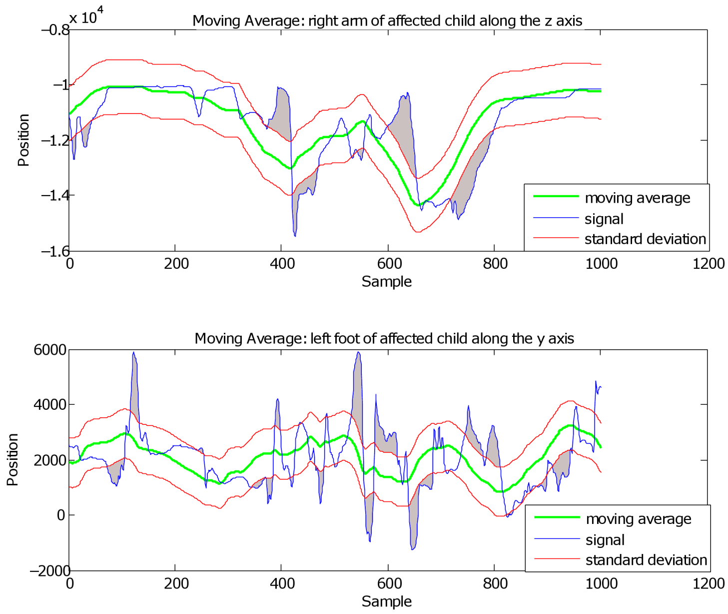
$$\eta_{s,Feet} = \sum_h \eta_{s,LFoot_h} + \sum_h \eta_{s,RFoot_h} \quad \text{with } h = [x, y, z]. \tag{10}$$

4.1.4 Periodicity

The movement of healthy children is characterized by a high degree of complexity, while an affected baby's movement is much more monotonous and lacks variation which means it has a more periodic appearance (Meinecke, 2006). In contrast to the aforementioned "area" parameters, periodicity is characterized by repetitive movements of high amplitude of the end effectors, and not by small variations or shakiness of movement.



6a: Normal



6b: Abnormal

Figure 6: Area out of standard deviation of moving average

Periodicity could be calculated by first determining the number of intersections between the signal and its arithmetic mean and then calculating the distance $\epsilon_{s,i}$ between each two intersections. After calculating the mean $\bar{\epsilon}$ and standard deviation $\sigma_{s,\epsilon}$ of these distances, the periodicity parameter can be defined (exemplarily for the x-axis) as

$$P_{s,x} = \frac{1}{\sigma_{s,\epsilon} + \bar{\epsilon}_{s,x}} \tag{11}$$

Merging the periodicities of left and right foot for each of the three spatial axes:

$$P_{s,Feet} = \sum_h P_{s,LFoot_h} + \sum_h P_{s,RFoot_h} \quad \text{with } h = [x, y, z]. \tag{12}$$

Meinecke et al. suggested to use moving average of high order where the window-width was $k > 1000$, but they could not apply this method due to gaps in the trajectories in their data. In addition, an important drawback of using moving average would have been that many samples in the beginning and end of the measurement would have been needed for calculation of the moving average, thus diminishing the available movement data.

Since the data used in current project has not the same problem as Meinecke et al. had, a similar approach to moving average is applied, using a low-pass filter. A data signal (position-time curve in motion analysis) normally has a mixture of different frequency components in it. The frequency contents of the signal and their powers can be obtained through operations such as the Fast Fourier Transform (FFT). A low-pass filter passes relatively low frequency components in the signal but stops the high frequency components. The so-called *cutoff* frequency divides the pass band and the stop band. In other words, the frequency components higher than the cutoff frequency will be stopped by a low-pass filter. This type of filter is especially useful since the random errors involved in the raw position data obtained through reconstruction are characterized by relatively high frequency contents. *Butterworth* filters are one of the most commonly used digital filters in motion analysis. They are fast and simple to use. Since they are frequency-based, the effect of filtering can be easily understood and predicted. The order and the cutoff frequency chosen for this project were respectively 5 and 0,025 Hz. The behavior of a filter can be summarized by the so-called frequency response function, *H*. The frequency response function of the Butterworth low-pass filter has the following form (J.G.Proakis, 2007):

$$|H(\Omega)|^2 = \frac{1}{1 + (\Omega/\Omega_c)^{2N}} \tag{13}$$

where *N* is the order of the filter and Ω_c is the cutoff frequency. In order to calculate the filter parameters, Butterworth filter design was applied. Furthermore, the Matlab function `filtfilt` was used for signal processing. The function `filtfilt` performs zero-phase digital filtering by processing the input data in both the forward and reverse directions. After filtering in the forward direction, it reverses the filtered sequence and runs it back through the filter. The advantage of this approach is that the result has precisely zero-phase distortion and a magnitude that is the square of the filter's magnitude response.

Periodicities were calculated for the feet and arms, using both the position trajectory and the absolute velocity trajectory.

4.1.5 Principal Component Analysis (PCA)

Andreas Berg considered several features in his master thesis (Berg, 2008), but not all of the retrieved features were helpful in the process of classification. Andreas Berg achieved an acceptable linear separability applying PCA and Single Input Single Output (SISO) Autoregressive (AR) model.

Component analysis is an unsupervised approach to find the "right" features from the data. The principal component analysis projects *d*-dimensional data onto a lower-dimensional subspace in a way that is optimal in a sum-squared error sense. First, the *d*-dimensional mean vector μ (15) and $d \times d$ covariance matrix Σ (14) are computed for the full data set.

$$\Sigma = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)(x_i - \mu)^T \tag{14}$$

$$\text{where } x_i = [x_{i_1}, \cdots, x_{i_j}]^T \quad \text{and} \quad \mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{15}$$

Next, the eigenvectors and eigenvalues are computed and stored according to decreasing eigenvalue. Then choose the *k* eigenvectors having the largest eigenvalues. Often there will be just a few large eigenvalues, and this implies that *k* is the inherent dimensionality of the subspace governing the "signal", while the remaining $d - k$ dimensions generally contain noise. Next a $d \times k$ matrix whose columns consist of the *k* eigenvectors is formed and called ***A***.

The representation of data by principal components consists of projecting the data onto the *k*-dimensional subspace according to (R.O.Duda, 2001)

$$F_1(x) = A^T(x - \mu) \tag{16}$$

As Fig. 7 describes, PCA is used to preprocess the raw data before feeding the data into the AR model, as an alternative way. In other words, all the possible inputs to the SISO AR model are the coordinate axis *x*, *y*, *z*, in addition to the projection onto the axis or plane defined by PCA [2].

In case where PCA is applied in preprocessing level, only the distance from each sensor to the axis defined by PCA is used as a feature. The reason for this choice and the parameters from the SISO AR model were selected in this project based on their performance. The selection criterion was based on a threshold

[2]This interpretation was not very well revealed through Andreas Berg's master thesis.

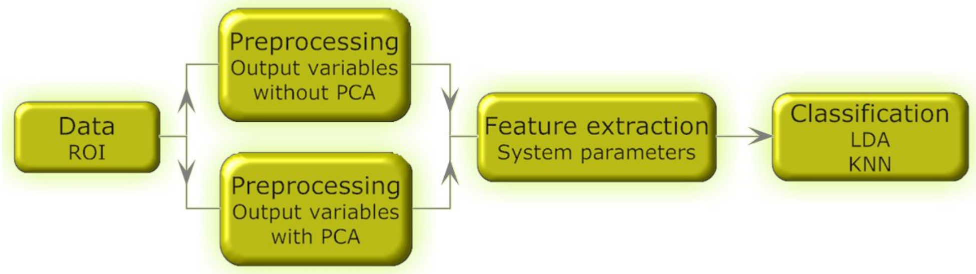


Figure 7: Application of PCA in preprocessing level

equal to 3, which was calculated using the Youden index (See Eq. 1) for features represented in (Berg, 2008).

4.1.6 Autoregressive Model (AR)

An AR-model is a type of random process which is often used to model and predict various types of natural phenomena. The AR-models are usually used in time series analysis to describe stationary time series. These models represent time series that are generated by passing the white noise through a recursive linear filter. The output of such a filter at the moment n is a weighted sum of p previous values of the filter output. The integer parameter p is called the *order* of the AR-model.

$$y_n = \sum_{k=1}^{p} a_k y_{n-k} + e_n \tag{17}$$

where e_n is the noise. As mentioned earlier, Anders Berg demonstrated in his master thesis that the SISO AR-model showed a good separability. Thus, the same approach was applied in this project with $p = 4$ as the order of the dynamic model. The features achieved by AR model are:

- The distance between each sensor and the axes x, y, z
- The distance between sensors in x-plane
- The distance from the origo of the coordinate system to each sensor in xy, xz and zy plane. In addition, the absolute distance from origin to every sensor was also considered.

4.2 Optimization of Parameter Combination

In feature selection and combination, this author specifically means selection of m features that provide the most discriminatory information, out of a possible

d features, where $m < d$. In other words, by feature selection, this author refers to selecting a subset of features from a set of features that have already been identified by some preceding feature extraction algorithms (see Section 4.2.2). The main question to answer under this setting is then "which subset of features provide the most discriminatory information?"
A criterion function is used to assess the discriminatory performance of the features, and a common choice for this function is the performance of a subsequent classifier trained on the give set of features. In essence, we are looking for a subset of features that leads to the best generalization performance of the classifier when trained on this subset. It should be noted, of course, the best subset then inevitably becomes a function of the classifier chosen. The feature selection is therefore said to be *wrapped around* the classifier chosen, and, hence, such feature selection approaches are referred to as wrapper approaches.

There is, of course, a conceptually trivial solution to this problem: evaluate every subset of features (all possible combinations of features) by training a classifier with each such subset, observing its generalization performance, and then selecting the subset that provides the best performance. Such an *exhaustive search*, as conceptually simple as it may be, is prohibitively expensive (computation wise) even for a relatively small number of features. This is because the number of subsets of features to be evaluated grows combinatorially as the number of features increase. For a fixed size of *d* and *m*, the number of subsets of features is

$$C(d,m) = \frac{d!}{m!(d-m)!} \tag{18}$$

Fortunately, more efficient search algorithms exists that avoid the full exhaustive search, such as the well-established depth-first search, breath-first search, branch and bound search, as well as hill climb searches referred to as forward and backward sequential feature selection. The latter one is applied in this project.

4.2.1 Separability of Features

In order to define how informative a feature is, the separability between the *normal* and *abnormal* data is calculated using the selected feature. Thus, the value gained from separability measurements, can be considered as a discriminating criterion between features applied in (Berg, 2008) and (Meinecke, 2006). The separability can be linear or nonlinear. The first one is achieved using *Scatter matrix* and the latter one is defined by analysis from *k-means clustering*.

4.2.1.1 Scatter Matrix Scatter matrix contains information about how the feature vectors are dispersed in a n-dimensional space. Different variations of

Scatter matrixes exist, among them *Within-class, Between-class* -and *Mixture scatter matrix,* which are defined as follows (Berg, 2008):

Within-class scatter matrix:

$$S_w = \sum_{i=1}^{M} P_i S_i \tag{19}$$

where M is the number of classes and S_i is the covariance matrix for the class w_i:

$$S_i = E[(x - \mu_i)(x - \mu_i)^T] \tag{20}$$

and P_i is the probability for the class w_i and $x \in \mathbb{R}^n$:

$$P_i \approx \frac{n_i}{N} \tag{21}$$

Between-class scatter matrix:

$$S_b = \sum_{i=1}^{M} P_i(\mu_i - \mu_0)(\mu_i - \mu_0)^T \tag{22}$$

where μ_0 is the global mean:

$$\mu_0 = \sum_{i=1}^{M} P_i \mu_i \tag{23}$$

Mixture scatter matrix:

$$S_m = S_w + S_b \tag{24}$$

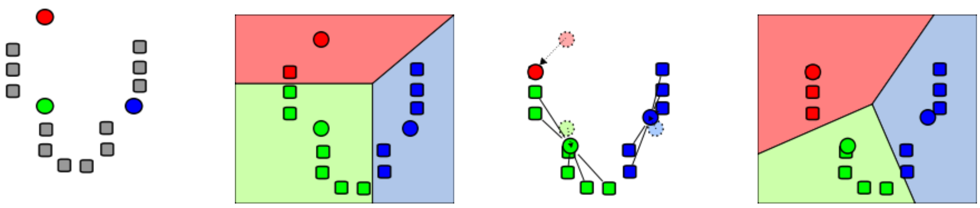
In order to measure how good the classes are separated, the criterion J is introduced:

$$J = trace(S_w^{-1} S_m) \tag{25}$$

4.2.1.2 K-means Clustering K-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume *k* clusters) fixed a priori. The main idea is to define *k* centroids, one for each cluster. These centroids should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early groupage is done. At this point we need to re-calculate *k* new centroids as barycenters of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop we may notice that the k centroids change their location step by step until no more changes are done. In other words centroids do not move any more. Finally, this algorithm aims at minimizing an *objective function*, in this case a *squared Euclidean distance* where each centroid is the mean of the points in that cluster. The algorithm for k-means clustering is defined as (R.O.Duda, 2001):

Algorithm (k-Means Clustering)
begin **initializing** n, c, $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, $\cdots$, $\boldsymbol{\mu}_i$
  do classify n samples according to nearest $\boldsymbol{\mu}_i$
    recompute $\boldsymbol{\mu}_i$
  until no change in $\boldsymbol{\mu}_i$
  return $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, $\cdots$, $\boldsymbol{\mu}_c$
end

The computational complexity of the algorithm is $O(ndkT)$ where d is the number of features and T is the number of iterations. n denotes the known number of patterns and k is the desired number of clusters. In practise, the number of iterations is generally much less than the number of samples.



8a: Step 1 8b: Step 2 8c: Step 3 8d: Step 4

Figure 8: Demonstration of the standard K-Means Clustering algorithm (Wikipedia, 2009*b*).
1) k initial "means" (in this case $k = 3$) are randomly selected from the data set (shown in color).
2) k clusters are created by associating every observation with the nearest mean. The partitions here represent the Voronoi diagram generated by the means.
3) The centroid of each of the k clusters becomes the new means.
4) Steps 2 and 3 are repeated until convergence has been reached.

The algorithm is deemed to have converged when the assignments no longer change. As it is a heuristic algorithm, there is no guarantee that it will converge to the global optimum, and the result may depend on the initial clusters. As the algorithm is usually very fast, it is common to run it multiple times with different starting conditions.

When it comes to choosing the *k* as the predefined number of clusters, two similar approaches have been implemented for the purpose of this project. The idea behind both approaches is to find the distance between created clusters and then choose the suboptimal *k* based on the largest distance. Thus, the clustering algorithm is run with several different numbers of clusters. Afterwards, for each *k*, the between-cluster distances and the mean of them are calculated. Finally, the *k* with the largest averaged distance is chosen and the k-mean clustering is run with the suboptimal *k* one more time.

Mahalanobis and *Euclidean* distances are the two techniques applied here for calculation of the between-cluster distances. It is up to the user to pick one of the techniques. Euclidean distance between two points is the length of the path connecting them. In general, the distance between the points *x* and *y* in a Euclidean space $\mathbb{R}^n$ is given by (Wolfram MathWorld, 2009*a*):

$$d = |x - y| = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2} \tag{26}$$

In Matlab, the function `silhouette` was used for measuring the distances between the clusters based on Euclidean distance.

In statistics, Mahalanobis distance is a distance measure introduced by P. C. Mahalanobis in 1936. It is based on correlations between variables by which different patterns can be identified and analysed. It is a useful way of determining similarity of an unknown sample set to a known one. It differs from Euclidean distance in that it takes into account the correlations of the data set and is scale-invariant, i.e. not dependent on the scale of measurements. The equation for the Mahalanobis distance is defined as R.O.Duda (2001):

$$r^2 = (\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \tag{27}$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_N)$ is the mean vector and Σ is the covariance matrix for a multivariate vector $\boldsymbol{x} = (x_1, x_2, \cdots, x_N)$. It is in the author's interest to find the distance between two clusters. Since Mahalanobis distance is defined as the distance from a point to a cluster, it is necessary to calculate the distance between all points in cluster A to cluster B. Then the result should be normalized by dividing it by the product of the number of points in A and B. At the final step, the mean of all the distances from A to B is found and used as the Mahalanobis distance between two clusters. In Matlab, the function `mahal` was appropriate for this purpose.

The clustering algorithm is run with normal and abnormal data, separately, using only the training data. When the suboptimal *k* is found and the k-means clustering algorithm has stopped, the separability between the normal and abnormal sets is computed. In this case, the same Mahalanobis distance approach is applied, considering the clusters from normal data and abnormal data as inputs (Appendix B.1).

4.2.2 Wrapper Method: Backward Sequential Feature Selection

As mentioned before, evaluating all subsets of features is extremely expensive in a computational manner. Thus a more clever but faster algorithm needs to be developed at cost of getting a local optimum in stead of a globally optimal feature combination. Before the applied procedure is explained, it is essential to note that the sequence and ranking of the feature in a subset plays a vital role when it comes to performance of the classifier. It can be demonstrated with a simple example. Assume that there exists a subset containing four features. Even though each of these features is very informative, individually, it does not necessary mean that any sequence of their combination together will result in good separability. So combining them as 1342 might give better result than 3214. Thus the question that comes to mind here will probably be "What feature should we start our subset with and which one should be the next in line?"

In order to overcome this problem, the first part of the algorithm should decide the sequence of features based on some weighting criteria. When the sequence of features is known, the second part of the algorithm will consider selection of those features who contribute the most, based on the some on-line classification results. For the first part of approach used in this project, the function `sequentialfs(fun,X,y)` in Matlab seemed appropriate. It selects a subset of features from the data matrix `X` that best predicts the data in `y` by sequentially selecting features until there is no improvement in prediction. Rows of `X` correspond to observations and the columns correspond to variables or features. `y` is a column vector of response values or class labels for each observation in `X`, and `fun` is a function handle to function that defines the criterion used to select features and to determine when to stop. The output of `sequentialfs` is a logical vector indicating which features are finally chosen. Since this procedure is a so-called *wrapper method*, it uses the `fun` to implement a learning algorithm. Methods like this usually apply cross-validation to select features.

The direction of the sequential search is chosen to be *backward*. This choice specifies an initial candidate set including all features and an algorithm that removes features sequentially until the criterion increases. For each candidate feature subset, `sequentilfs` performs 10-fold cross-validation by repeatedly calling `fun` with different training subsets of `X` and `y`, `XTRAIN` and `ytrain`, and test subsets of `X` and `y`, `XTEST` and `ytest` as follows (Appendix B.2):

```
criterion = fun(XTRAIN, ytrain, XTEST, ytest)
```

As the objective function for this algorithm, a linear classifier was applied, using the `classify` function in Matlab. The data fed to the algorithm contained only the predefined training set (see Section 4.3.1). This was done in order to avoid the classifier, which was used later on in the classification part, from being introduced to the test sets, thus achieving generality.

As explained here, this algorithm returns a subset containing the most discriminatory features. A critical issue that should be noticed about this algorithm is that it returns different subsets based on its initial values. In order to find out which feature has higher discriminatory value, the procedure was run several times, iteratively. The number of iterations were $10, 100, 1000$ and 10000. For each of the chosen iterations, the produced subset was stored and the numbers of times that every feature was appeared on the resulting subset were registered. A weight value for each feature was then calculated in percent demonstrating the importance of the feature based on the total number of its appearance on the subset. At the end of the first part, four subsets were achieved from $10, 100, 1000$ and 10000 iterations. Each of the resulted subsets was sorted based on the weight of the features. In other words, every final subset contained all of features, but the features were ranked based on their calculated weight value. Thus, at this stage, the sequence of the features was determined.

The second part of the feature selection algorithm, takes into account the sorted subset from the first part. Now the selection part could start which was designed by the author. Using a *forward feature selection* approach, the algorithm starts with an empty subset. The algorithm goes through the ranked subset of features and picks one by one, starting with the best features (highest weight value), and places the feature in the subset. For every selection, a classification procedure is run and the sensitivity and specificity of the classification result is calculated. Another alternative was the choosing Youden index (see Eq. 1) as the criterion, instead of sensitivity and specificity. The classification is done both with a *linear* and a *quadratic* classifier. Before adding the next feature to the subset, the previously calculated Youden Index is stored. Then after adding the next feature, a new Youden index is calculated and compared to the previous one. If there has been an increase in Youden index, the lately added feature is kept in the subset, otherwise, it is removed from the subset. Then, the algorithm moves on to evaluating the next feature in line and so on, until the very last feature has been taken into account.

The resulting subset will finally contain the suboptimal combination of the optimally selected features.

4.2.2.1 Convergence of the feature combination algorithm Due to lack of time, the convergence of the mentioned algorithm has not been proven mathematically. However, it can be shown that the algorithm will not diverge. The algorithm is intuitively expected to converge, just as indicated by Fig. 9.

The plots illustrated in Fig. 9, which are only few examples, show that the calculations are converging towards a linear line, especially after 1000 iterations. The same plots can simply be produced and investigated for all features. The code for this procedure is created by the author and is available in Appendix B.2.4. In order to gain more confidence on the convergence of the algorithm, one could continue increasing the iterations up to 10^5 or even 10^6. The idea was not realizable during this project, because of the exponentially high computation time.

4.3 Classification

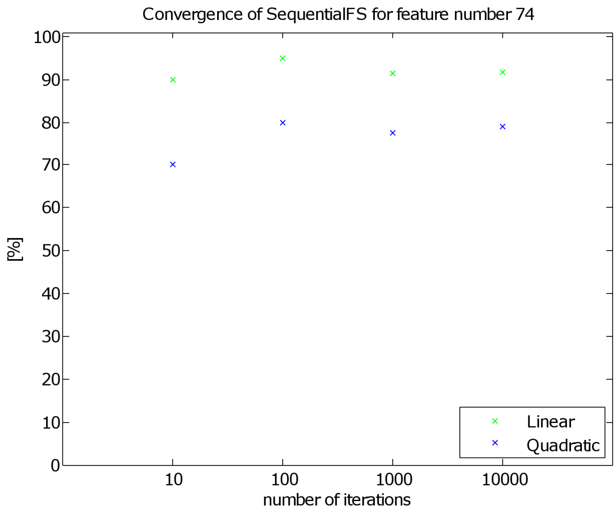
The final classification took place during the second part of the feature combination procedure. Thus, this section only explains the classifiers applied earlier, in addition to reviewing creation of *training* -and *test* sets.

4.3.1 Data Partitioning

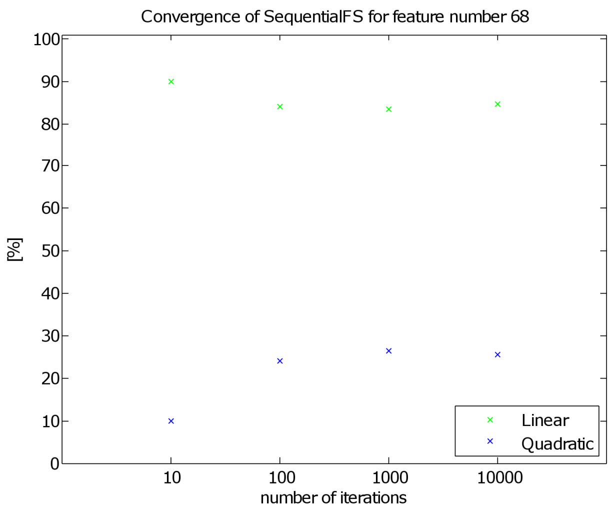
Before the implementation of feature extraction, selection, combination and classification, the collected data were divided into training and test sets. As explained in Section 1.1.2, there can exist one or several ROI-records from each baby and every participant has a unique identification number. What that makes the data partitioning more challenging is to avoid having the ROIs from one participant in both training and test sets, which is a substantial principle, and still select and place the participants randomly in training and test sets.

4.3.2 Discriminant Analysis

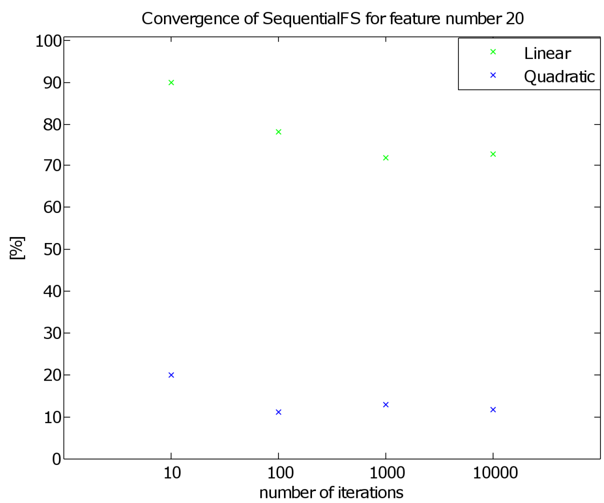
Originally developed in 1936 by R. A. Fisher, *discriminant analysis* is a classic method of classification that has stood the test of time. Discriminant analysis often produces models whose accuracy approaches (and occasionally exceeds) more complex modern methods. Discriminant analysis can be used only for classification, not for regression (P.Sherrod, n.d.). The target variable may have two (which is the case here) or more categories. For the purpose of this project, both the *linear* and *quadratic* variations of discriminant analysis were applied. Using the Matlab function `classify` with the appropriate input arguments, the



9a: Feature: head in $y - plane$, AR parameter nr. 2



9b: Feature: right arm in $x - plane$, AR parameter nr. 4



9c: Feature: area out of standard deviation for arms velocity

Figure 9: Examples of the suboptimal combination algorithm's non-divergence behaviour

classification was done during the final step of feature combination.

When linear discriminant analysis (LDA) is applied in a two-class problem, consider a set of features *x* for each sample of an object or event with known class ω. This set of samples is called the training set. The classification problem is then to find a good predictor for the class ω of any sample of the same distribution (not necessarily from the training set) given only an observation x. LDA approaches the problem by assuming that the conditional probability density function $p(\vec{x}|\omega = 1)$ and $p(\vec{x}|\omega = 0)$ are both normally distributed. Under this assumption, the Bayes optimal solution is to predict points as being from the second class if the likelihood ratio is below some threshold α, so that (Wikipedia, 2009*c*)

$$(\vec{x} - \vec{\mu}_0)^T \Sigma_{\omega=0}^{-1} (\vec{x} - \vec{\mu}_0) + ln|\Sigma_{\omega=0}| - (\vec{x} - \vec{\mu}_1)^T \Sigma_{\omega=1}^{-1} (\vec{x} - \vec{\mu}_1) - ln|\Sigma_{\omega=1}| < \alpha \quad (28)$$

Without any furthure assumptions, the resulting classifier is referred to as quadratic discriminant analysis (QDA). LDA also makes the simplifying homoscedastic assumption (i.e. that the class covariances are identical, so $\Sigma_{\omega=0} = \Sigma_{\omega=1} = \Sigma$) and that the covariances have full rank. Thus, several terms cancel and the above decision criterion becomes a threshold on the dot product $\vec{\xi} \cdot \vec{x} < \beta$ for som constant β, where $\vec{\xi} = \Sigma^{-1}(\vec{\mu_1} - \vec{\mu_0})$.

This means that the probability of an input *x* being in a class ω is purely a function of this linear combination of the known observations.

5 Results and observations

This chapter contains the results which are in accordance with the objectives of the performed study (see Section 3) and the methods applied (see Section 4). First the features from (Berg, 2008) and (Meinecke, 2006) will be compared using results from scatter matrix and k-means clustering. Then, the most optimal set of combined features are introduced, along with the classification results. At last, a ROC-plot will illustrate an evaluation of the feature combinations based on their performance during the classification.

Before we start to analyse the results, it is necessary to carry through Table 2 in order to understand the symbols used throughout this section in different tables. Since the names of some features are quite long, these symbols are used to reduce the width of columns in tables, making them perspicuous.

Table 2: Explanation of symbols

SYMBOL	DECLARATION
σ	standard deviation
$\bar{M}$	moving average
O	origin of coordinate system
S	sensor
Γ	distance between two points

5.1 Feature Selection

As discussed in Section 4.1.6, the AR model was implemented as a 4^{th} order system identification model. Thus, if the tables in this chapter contain a column named *AR Parameter Number*, it is referred to the order of the AR model. In other words, an output of the AR model will have four features within itself. For instance, the output "head sensor in $y - axis$" contains four AR parameters, each of which being considered as a feature.

5.1.1 Linear Separability

Based on Eq. 25, Table 3 contains the calculated results for only those features with good separability. The performance of all features can be found in Appendix A. As it can be observed, *periodicity* gives the highest value for separability, but the few next best features belong to the AR model.

By taking a closer look at the Eq. 25, it can easily be seen that the *J* will increase infinitely when $(\mu_i - \mu_0) \rightarrow \infty$, while the lower boundary will go toward 1 as $(\mu_i - \mu_0) \rightarrow 0$. With this scaling in mind, the periodicity is not so separable after all, even though it is the highest calculated value relative to other features. When considering the linear separability results, the focus should be on the decimal portion of the numbers and not the left hand side of the point.

Using Table 3 as a general overview of linear separability of features, it is obvious that the features suggested by Berg (2008) are more interesting than the features applied by Meinecke (2006).

Table 3: Results for Scatter Matrix

(a) Periodicity

Name of Feature	Separability
$P_{Arms,Position}$	**1.1859**
$P_{Arms,Velocity}$	**1.1489**

(b) Sensors along the y-axis

Name of Feature	Separability
S_6 in $y - axis$	**1.0921**
S_3 in $y - axis$	1.0476

(c) Distance between sensors and PCA-axis

Name of Feature	Separability
Γ_{PCA} from axis to S_6	**1.0899**
Γ_{PCA} from axis to S_3	1.0703
Γ_{PCA} from axis to S_4	1.0537
Γ_{PCA} from axis to S_2	1.0468

(d) Sensors along the x-axis

Name of Feature	Separability
S_3 in $x - axis$	1.0635
S_4 in $x - axis$	1.0520
S_6 in $x - axis$	1.0431
S_2 in $x - axis$	1.0424

(e) Distance between the sensors and O in xz-plane

Name of Feature	Separability
$\Gamma_{xz-plane}$ from O to S_3	1.0484
$\Gamma_{xz-plane}$ from O to S_4	1.0414

(f) Distance between the sensors and O in xy-plane

Name of Feature	Separability
$\Gamma_{xy-plane}$ from O to S_4	1.0515
$\Gamma_{xy-plane}$ from O to S_6	1.0426

(g) Area out of standard deviation

Name of Feature	Separability
Area out of $\sigma_{Feet,Velocity}$	1.0339

(h) Area differing from moving average

Name of Feature	Separability
Area out of $\bar{M}_{Feet,Velocity}$	1.0401

(i) Distance between arms in x-axis

Name of Feature	Separability
Γ_{x-axis} between S_3 and S_4	**1.0830**

5.1.2 Clustering Analysis

Table 4(a) and 4(b) demonstrate the non-linear separability among the top 20 features. The results are sorted and the highest values are place at the top. The features represented in Table 4(a) are ranked based on the Euclidean distance between the clusters. The most separable feature in this case is "the distance between the PCA axis and the right arm sensor". On the other hand, Mahalanobis distance was applied in the same manner as the Euclidean distance. Table 4(b) shows also the top 20 features selected and sorted based on Mahalanobis distance between the clusters. Here, the best value belongs to "the periodicity in position signal from the infant's arms".

Table 4: Clustering Analysis

(a) Euclidean between-cluster distance

NAME OF FEATURE	AR PARAMETER NUMBER	DISTANCE
Γ_{PCA} from axis to S_3	4	2.4855
Area out of $\sigma_{Feet,Velocity}$	-	1.8874
$\Gamma_{zy-plane}$ from O to S_4	4	1.2216
Γ_{PCA} from axis to S_3	3	0.9420
$\Gamma_{xz-plane}$ from O to S_4	4	0.9170
$\Gamma_{zy-plane}$ from O to S_6	4	0.6666
Area out of $\sigma_{Arms,Position}$	-	0.6643
Γ_{PCA} from axis to S_4	4	0.6422
Γ_{PCA} from axis to S_3	1	0.6218
$\Gamma_{xz-plane}$ from O to S_6	3	0.5146
Γ_{PCA} from axis to S_3	2	0.4352
Γ_{x-axis} between S_3 and S_4	1	0.4348
S_6 in $x-axis$	3	0.4117
Γ_{x-axis} between S_3 and S_4	2	0.3215
$\Gamma_{Absolute}$ from O to S_3	3	0.3186
S_3 in $x-axis$	3	0.3140
$\Gamma_{Absolute}$ from O to S_3	4	0.2441
Area out of $\bar{M}_{Arms,Velocity}$	-	0.2357
$\Gamma_{xz-plane}$ from O to S_6	2	0.2011
S_3 in $x-axis$	4	0.1957

(b) Mahalanobis between-cluster distance

NAME OF FEATURE	AR PARAMETER NUMBER	DISTANCE
Periodicity$_{Arms,Position}$	-	5.7060
Area out of $\bar{M}_{Feet,Velocity}$	-	3.4041
$\Gamma_{xz-plane}$ from O to S_4	4	3.0031
Area out of $\sigma_{Feet,Velocity}$	-	2.2721
Area out of $\sigma_{Feet,Position}$	-	1.7555
$\Gamma_{xy-plane}$ from O to S_6	1	1.4939
Γ_{PCA} from axis to S_3	1	1.2522
Γ_{PCA} from axis to S_3	2	1.1393
Γ_{PCA} from axis to S_3	4	0.9473
Γ_{x-axis} between S_3 and S_4	1	0.7696
Γ_{PCA} from axis to S_3	3	0.7487
$\Gamma Absolute$ from O to S_3	4	0.6974
$\Gamma_{zy-plane}$ from O to S_4	4	0.5724
S_3 in $x-axis$	1	0.5508
S_3 in $x-axis$	4	0.5238
Γ_{PCA} from axis to S_4	4	0.5148
$\Gamma_{xz-plane}$ from O to S_6	3	0.5146
S_6 in $x-axis$	3	0.5119
$\Gamma_{zy-plane}$ from O to S_6	4	0.4196
Γ_{x-axis} between S_3 and S_4	2	0.3980

Considering both tables, the features from (Berg, 2008) and (Meinecke, 2006) are well included, both in quantum and quality. Furthermore, 15 of 20 features are mutually selected using both procedures, but the sequence and ranking of the features are different. It can be observed that features like "the distance from the PCA axis to arms and head", "the area out of standard deviation for the feet sensors" and "the distance from the origin to head and arms in different planes" are often selected. This could be an indication for how informative these features are, compared to the rest of them.

5.2 Feature Combination and Classification

In total, there were produced up to eight different combinations regarding the 10, 10^2, 10^3, 10^4 and criteria like Youden Index and sensitivity and specificity. Table 5 only demonstrates the top four combinations which showed the best classification capability. It is noticeable that size of combinations are not restricted and Table 5(c) contains the largest number of features. Furthermore, three of the four presented tables apply Youden Index as their selection criterion while only one of the tables uses QDA as its classifier. It is essential to take notice of the sequence of features in each combination. If the AR parameter numbers of a feature are separated by comma, this means that most left digit has higher ranking than the following numbers. Despite the relatively good performance of the features suggested by Meinecke (2006) as in separability value, very few of them are taken into account in these combinations. Specially in Table 5(d), there are none features belonging to Meinecke (2006).

In addition to explanation of the source of every feature in Section 4.1, there is a simple way to distinguish them in represented tables. By considering the column for AR parameter number in every table, it can be seen that only the features belonging to Berg (2008) have such parameter number. Thus, features who do not have an AR parameter number could be said to be originated from (Meinecke, 2006). So it is easy to observe that the dynamic features extracted from AR model are much more interesting than the statistic features. Furthermore, features related to the head sensor are highly informative and discriminative, thus supporting Andreas Berg's conclusion. A surprising observation is the difference between the number of times the arms and feet sensors have participated in combinations. Both during the separability analysis and feature combination, the AR parameters from arms sensors play a much more important role, compared to the signals from baby's feet.

Table 5: The most suboptimal feature combinations achieved

(a) **Combination A** with sensitivity and specificity as selection criterion and linear classifier as @fun (see Section 4.2.2)

FEATURE	AR PARAMETER NUMBER
S_6 in $y-axis$	3
Γ_{x-axis} between S_3 and S_4	3
Γ_{x-axis} between S_3 and S_4	4
S_4 in $z-axis$	3, 2, 4 [3]
S_6 in $y-axis$	1
$\Gamma_{zy-plane}$ from O to S_6	2
S_4 in $x-axis$	2
S_4 in $x-axis$	4
$\Gamma_{xz-plane}$ from O to S_6	4
Γ_{x-axis} between S_3 and S_4	1
$\Gamma_{Absolute}$ from O to S_4	3
$\Gamma_{xz-plane}$ from O to S_4	4
$\Gamma_{Absolute}$ from O to S_3	4
Γ_{PCA} from axis to S_3	2
Area out of $\bar{M}_{ArmsPosition}$	-
Γ_{PCA} from axis to S_6	1
Cross-correlation$_{Arms}$	-

(b) **Combination B** with Youden Index as selection criterion and linear classifier as @fun (see Section 4.2.2)

FEATURE	AR PARAMETER NUMBER
S_3 in $x-axis$	3
S_6 in $y-axis$	3
Γ_{x-axis} between S_3 and S_4	1
$\Gamma_{zy-plane}$ from O to S_4	2, 4 [3]
S_3 in $x-axis$	4
S_4 in $z-axis$	1
Γ_{x-axis} between S_3 and S_4	4
Γ_{PCA} from axis to S_4	4
$\Gamma_{Absolute}$ from O to S_3	3
$\Gamma_{xz-plane}$ from O to S_6	2
S_4 in $z-axis$	4
$\Gamma_{Absolute}$ from O to S_4	1
$\Gamma_{xz-plane}$ from O to S_6	1
Cross-correlation$_{Arms}$	-
$\Gamma_{xz-plane}$ from O to S_4	4
Cross-correlation$_{Feet}$	-
Area out of $\sigma_{FeetPosition}$	-

(c) **Combination C** with Youden Index as selection criterion and linear classifier as @fun (see Section 4.2.2)

FEATURE	AR PARAMETER NUMBER
S_6 in $y-axis$	3, 4 [3]
S_6 in $x-axis$	2, 3 [3]
$\Gamma_{zy-plane}$ from O to S_6	4
S_4 in $x-axis$	4
$\Gamma_{zy-plane}$ from O to S_6	2, 3 [3]
S_4 in $z-axis$	4
S_6 in $y-axis$	1
S_4 in $z-axis$	2
$\Gamma_{xy-plane}$ from O to S_6	2
S_3 in $x-axis$	4
S_4 in $z-axis$	1
$\Gamma_{Absolute}$ from O to S_4	2
$\Gamma_{xy-plane}$ from O to S_6	1, 3 [3]
$\Gamma_{zy-plane}$ from O to S_4	3
S_4 in $x-axis$	2
$\Gamma_{xy-plane}$ from O to S_6	4
$\Gamma_{xz-plane}$ from O to S_6	1, 4 [3]
Area out of $\sigma_{ArmsVelocity}$	-
Area out of $\bar{M}_{ArmsVelocity}$	-
Γ_{PCA} from axis to S_6	3, 4 [3]
Area out of $\sigma_{FeetVelocity}$	-
Γ_{PCA} from axis to S_3	4
Area out of $\sigma_{ArmsPosition}$	-
Γ_{PCA} from axis to S_4	1
Cross-correlation$_{Arms}$	-
Skewness$_{S_3}$	-

(d) **Combination D** with Youden Index as selection criterion and quadratic classifier as @fun (see Section 4.2.2)

FEATURE	AR PARAMETER NUMBER
S_6 in $y-axis$	2
Γ_{x-axis} between S_3 and S_4	1
S_6 in $x-axis$	3, 2 [3]
S_4 in $z-axis$	1
S_6 in $y-axis$	3, 1 [3]
S_4 in $z-axis$	2, 4 [3]
Γ_{x-axis} between S_3 and S_4	4, 4 [3]
S_6 in $x-axis$	1

Each of combinations has been denoted by a capital letter which is later on used to link them to the appurtenant classification result in Table 6. The most appropriate results are achieved by using LDA as the classifier, where the highest Youden Index acquired was 0.76 as shown in Table 6. Number of iterations for the top result is known to be 10^3, which indicates the number of iteration during the combination procedure (see Section 4.2.2).

Table 6: Classification

Combination	Classifier	Sensitivity	Specificity	Youden Index	Number of Iterations
A	LDA	0.86	0.90	0.76	10^3
B	LDA	0.90	0.84	0.74	10^1
C	LDA	0.86	0.88	0.74	10^2
D	QDA	0.76	0.90	0.66	10^2

Figure 10 illustrates Table 6 in a more standardized and comprehensive way, using a ROC-plot.

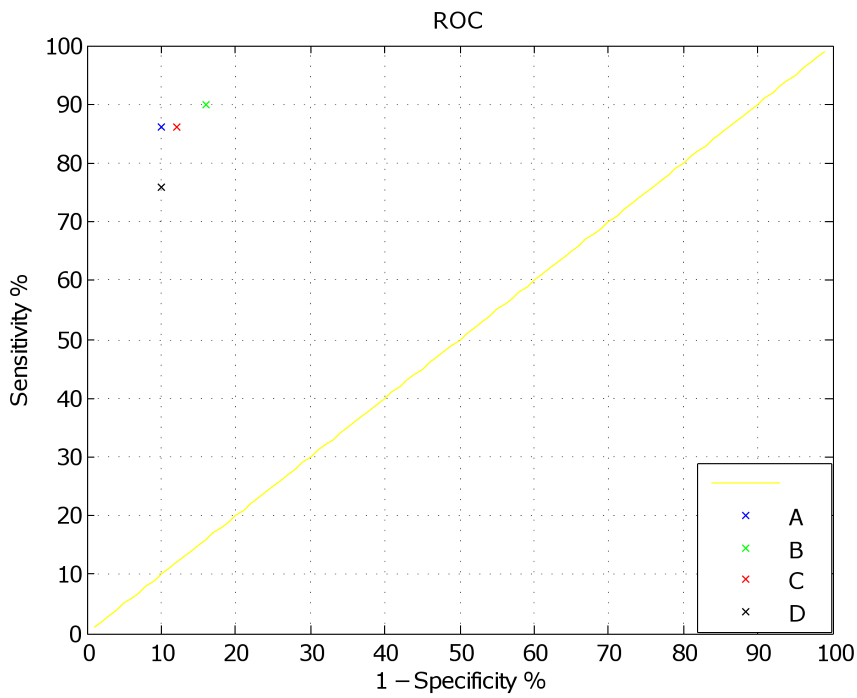


Figure 10: Illustration of Table 6

[3]It is essential to take notice of the sequence of these AR parameter numberings.

6 Discussion

6.1 Linear Separability of Features

Based on the values displayed in Table 3, features like "*periodicity*", "*distance between arms along the $x - axis$*", "*distance between PCA axis and sensors*" and "*sensors along x and $y - axis$*" give the highest linear separability values. Specially, the penultimate and last one are the most repetitive ones. Furthermore, the sensor signals from head and arms are quite informative, substantiating the results from Berg (2008). Surprisingly, the signals from feet sensors are not as interesting as they were assumed to be. On the other hand, movements along the x and $y-axis$ are much more enlightening than the movements in direction of $z - axis$, which is explainable by considering Figure 1. It is obvious that most of the movement variation happens in x and $y - axis$, since the infant is lying on his/her back, particularly the head movements, which are mostly from right to left and vice versa.

Even though that the features "*periodicity in position signal from arms*" and "*periodicity in velocity signal from arms*" show the best linear separability achieved by Scatter Matrix, they are the only top feature which belong to the subset defined by Meinecke (2006). All the remaining top features demonstrated in Table 3 belong to Berg (2008).

If the same "belonging" pattern of features repeats itself both in analysis from *k-means clustering* and feature combination, it will be a very strong indicator for supporting features from Berg (2008) in preference to features from Meinecke (2006).

As explained in Section 5.1.1, the lower boundary for Eq. 25 is 1, while the upper boundary is unlimited. With that in mind, a linear separability value like 1.1859 (consider only the right hand side of the point) is not an extreme value. Thus, the achieved results in Table 3 illustrate a poor linear separability among the features included in this study. It would be interesting to see if a non-linear classifier could also support this behaviour. A not so time-consuming attempt should be done in future, defining the rate of dependency and connection between the value of linear separability and a classifier's performance (like Youden Index). The main part of such a future work shall focus on the ratio and find out whether the relation between the classifier's perfomance and the linear separability of a feature is linear or exponential. In other words, how good should the linear separability of a feature be in order to achieve an acceptable performance in classification.

Still, it can not be concluded that a subset of these features will necessarily show poor classification performance, using a linear classifier. Usually, by combining two weak features, a better classification result could be obtained, since each feature is independent from the other one and adds new information to the classification procedure. One alternative would be to use the linear separability value as a criterion for feature selection and combination. But in this project, it has only been used to evaluate every feature individually and compare them with each other.

During the calculation of Scatter matrix for the outputs of AR model, the AR parameters were not considered as individual features. As an improvement, this should be recalculated separately, so that the effect of each parameter can be correctly evaluated.

6.2 Clustering Analysis

During the clustering procedure, first, the intra-class clusters were created within each of classes. Then, the clusters from the Normal and the Abnormal classes were analysed all together. For both of mentioned stages, Euclidean and Mahalanobis distance were applied. Table 4 demonstrates the final inter-class clustering analysis. As mentioned earlier, 15 of the top 20 features are common in both Table 4(a) and Table 4(b), but the sequence and ranking of the mutual features are quite different. Features from Meinecke (2006) constitute 20% and 40% of the $\underline{\text{top } 10}$ features, respectively in Table 4(a) and Table 4(b). On the other hand, when using the Euclidean distance, only 1 out of 5 top features belongs to Meinecke (2006), while the ratio is 4 out of 5 when applying Mahalanobis distance. Furthermore, 85% and 80% of the $\underline{\text{top } 20}$ features respectively in 4(a) and Table 4(b) belong to Berg (2008). It seems that the clustering analysis also confirms and supports the conclusion from linear separability analysis. Obviously, the outputs of AR model are more fitted and suitable for describing the dynamic of the infant's movements, than the statistic features introduced by Meinecke (2006). The reason for this phenomenon will be overviewed later on in Section 6.3.

Compared to the results from linear separability analysis, features like "*distance between arms along the $x - axis$*", "*distance between PCA axis and sensors*" and "*sensors along x and $y - axis$*" are well represented and quite contributive. In addition, "*periodicity in arms*" and "*the area out of standard deviation of the moving average of feet's velocity*" are of great value, too. It can not be assumed that features with significant linear separability value, will necessarily achieve high performance in nonlinear separability. Therefore, there should not be any expectation about having common features appearing both in Table 3 and Table 4. On the contrary, if some features take part in both tables, then it's an indication for how versatile and essential these features are.

It is also interesting to take notice of the AR parameters. Apparently, parameters of 3^{th} and 4^{th} order are more informative and have better discriminative capabilities, than the parameters of lower orders. This observation is based on how often each parameter is appeared in Table 4. Since the infant's movements have a highly sophisticated nature, it might be more precise to describe and distinguish them by a more complex AR model. An idea could be to try a higher order than 4 in the AR model or take advantage of other heavy system identification models. But as it is reviewed in (Berg, 2008), principles like *final prediction error* and *Akaike information criterion* were used in order to define an appropriate order for the AR model. Both of these criteria penalize high orders.

A suggestion for future work could be to somehow use k-means clustering results as a feature selection and combination criterion, specially, if it has been decided to implement a nonlinear classifier.

Investigating both procedures, it can not simply be decided which of the applied distances obtains better results and is more suited for problems of this kind. As we know, Mahalanobis distance assesses the scattering and dispersion of the data within each cluster, in addition to taking the distance between the centers of clusters into account. So, it can be said that the Mahalanobis distance is more complex and complete, compared to the Euclidean distance. Thus, it is not so strange that numerical values for distances have different ranges.

The author has spent some time and tried to compute a scale for the values achieved in Table 4 without any success. It would be interesting to know how good an inter-cluster distance with 2.48 or 5.70 as the distance value is. Does this mean that the clusters show good separability or do they have any overlap? Questions like this can be suggested as a part of future work in this area. The idea is to calculate the distance between clusters of two different classes, and then perform a classification, using same clusters. The Youden Index from the classification and the calculated distance can be used to create a scale for the nonlinear separability of a feature. The data used by the author for creating a scale was generated randomly in order to hold a general view. K-means clustering and LDA was used as before. Since the K-means clustering is usually used for calculation of *nonlinear* separability and the LDA considers the *linear* relations among the data, the attempt was neither appropriate nor successful. Thus in the future work, care must be taken to apply a nonlinear classifier, instead of a linear one, when creating a scale for distance between clusters.

6.3 Feature Comparison

Both feature evaluation techniques, the linear separability achieved by scatter matrix and nonlinear separability calculated by clustering analysis, indicate that the features produced by Berg (2008) are much more reliable and show better discriminative abilities, compared to the features introduced by Meinecke (2006).

One essential point is to notice the amount of statistically characterized features in the work of Meinecke (2006). This means that just the majority of her features were statistical. Meinecke (2006) applied also feature with dynamical properties, like moving average, which functions as a low-pass filter. In a more precise point of view, it is the outputs of the autoregressive model that contribute the most. The AR model was originally used to describe a *linear* model, because of the lack of information and knowledge about any physical model, which could interpret the child's unpredictable movements.

In most system identification and estimation techniques, it is necessary to assume that the signal is stationary. This requires that the underlying statistics and the model parameters that characterize the process are not dependent on time. However, this assumption is often incorrect for many physical signals encountered in speech processing, EEG analysis, and seismology. No general mathematical framework exists, and in practice, the problem of time dependency is circumvented by presuming that the process is locally stationary over a relatively short time interval, but globally nonstationary.

An important reason for the good performance of AR is the generality that its poles give to the model, while the statistical features, like cross-correlation coefficient, lack this vital capability. It is a well-known connection between the poles of a model and its time constants. The complexity of a multivariable system can be decomposed to simpler monovariable subsystems. Each of these subsystems has its own unique time constant describing the components behaviour, individually. It is the (linear) relation between the subsystems that builds up the multivariable system. .From a technical point of view, it is the eigenvalues or poles of the system that express the functionality and performance of the system. These eigenvalues can be extracted from the system matrix. The interesting part is that different system matrixes can be obtained from the same set of eigenvalues, thus explaining how general, complete and descriptive these poles can be.

Furthermore, the AR model takes the sequence of samplings in to account, thus producing a smooth output. On the other hand, the statistical approaches might only consider the distribution of samples without caring about the order of samplings.

Linear autoregressive models (AR) have a broad spectrum of applications ranging from identification, prediction and control of dynamical systems, but a problem with this method lies with the appropriate model order selection. As one can surmise, the AR model can be of any order as desired. However, it should be as accurate as possible. From our intuition, we know that a model order, which is too small will not represent the properties of the signal, whereas a model order which is too high will also represent noise and inaccuracies and thus, will not be a reliable representation of the true signal. Therefore, methods that will determine the appropriate model order must be used. As mentioned earlier, criteria like *final prediction error* and *Akaike information criterion* has been applied by Berg (2008) in order to ensure an relatively optimal choice of order. As mentioned in Sec. 6.2, there could be some benefit in evaluation of an AR model of higher order. But, as explained here, it might turn out to be useless. The main idea to look at would be the criterion selected to choose the order. Is it certain that final prediction error and Akaike information criteria are the most appropriate criterion for the purpose of the current study? Are there other type of criteria that could fit the objective of this project better? These are questions that can be answered during a closer investigation of the AR model, as part of a future study.

6.4 Classification

The highest achieved performance in Table 6 belongs to the LDA with a 86% and 90% as sensitivity and specificity, respectively. The combination which is the source to such interesting and valuable result contains features that mostly, if not all, belong to the work of Berg (2008). The combination spoken of (combination **A**), has a total of 19 features, where only 2 of them has been suggested by Meinecke (2006). As discussed earlier, such behaviour was predicted and supported by both linear and nonlinear separability analysis. Thus it can be concluded that features produced by autoregressive model have a powerful and rich influence in classification of healthy and at-risk infants.

By reviewing the discussion chapter in (Berg, 2008), the best result is contributed by movement signals from child's head in $y - direction$ and the distance between infant's arms along the $x - axis$. This observation has been supported in the current work, too. Considering combination **A**, which is the highest ranked one, the top 3 features are the third AR parameter of infant's head in $y - direction$, the third and the fourth AR parameter from the distance between baby's arms in $x - direction$.

The overlapping performance and connexion between this thesis and the Andreas Berg's thesis could be an indication, suggesting a completely acceptable convergence of the designed feature combination algorithm. Both, the features selected for this combination and the weighting (ranking) procedure applied, confirm a quite optimal performance from the constructed algorithm.

When it comes to the complexity of classifiers, it can be suggested as a future work to apply a powerful classifier like Artificial Neural Network or Support Vector Machine, even though the LDA used in this study showed better performance than QDA.

6.4.1 Feature Combination

During the feature combination approach, there was no restriction on the number of chosen features. In other words, a subset could involve up to 84 features. It might be helpful to try and reduce the number of features down to only those features with the most information. First of all, one should be able to measure the contribution from every single feature. With that in mind, the algorithm can start removing features one by one, while monitoring the classifiers performance. A weighting function will be appropriate for this part. Since the subsets, originally, are suboptimal, only features with minimal performance decrease shall be removed.

As a role, this feature elimination has something to do with the reduction of processing and calculation time. One should be reminded that these feature and subsets will be used in a real-time approach, if their performance is approved by technician, physicians and doctors. Thus, collecting, processing and classification of the data online should not take much time, in order to make the approach user-friendly.

Most of the features observed in the locally optimal feature subset [Table 5(a)] are in accordance with Table 3 and Table 4. This illustrates that linear and nonlinear separability analysis could effectively be used as feature selection criteria. All these evidence illustrate the great importance of monitoring the movements of infant's head and arms, when classifying them.

One conclusion can be that combining signals from different body parts will help us get closer to the optimal solution, instead of considering limbs, individually. This statement is fairly intuitive and not so surprising at all, since characteristic movements can appear in different body parts at various points in time.

6.4.2 Clinical Perspective

In order to gain a clinical point of view, an interview has been conducted with Lars Adde, who is a physiotherapist (physician). He has been working with diseases related to children for more than ten years, specially with CP among newborn babies with premature birth. In addition, he has had a few publications in this field.

During the current section, it has clearly been reflected that signals recorded from head and arms have been more informative than the signals collected from the feet sensors. To start with, the observed conclusion sounded quite surprising to Lars Adde. But he could neither reject nor completely accept the suggestion. Lars Adde has taken some courses in visual recognition of Fidgety Movements (see Section 2.1.3) and has become an expert in the area. By studying video records from an infant, he can predict if the child is at-risk of developing CP or not. The way he explained it, he looks for fidgety movements as an overall movement pattern. That means that he does not search for detailed limb movements, thus considering a combination of movements in all limbs (head, arms, trunk and feet). This means that his brain has not been trained to weight movements in limbs differently. Such behaviour can be explained by Gestalt Theory of Visual Perception [4]. Therefore, he could not agree or disagree whether movement patterns in head and arms are more informative than movement patterns in feet. However, he was quite positive to the results achieved in this project.

An interesting future work could be to find out what a physician searches for when he/she is observing fidgety movements. An idea would have been to somehow track physician's eyes and locate the points on the recorded video which the physician mostly focuses on. This could after all give us some information about the distinction and influence of different limbs, regarding prediction of CP.

The sensor placement has also been investigated. Sensors that belong to feet are placed above the ankle, so they are actually located on the leg and not the foot. Thus, the movements of feet around the ankle-point are not registered. According to Lars Adde, the variation in feet movements contribute a lot as part of fidgety movements. Similarly, sensors on the arms are placed between the wrist and elbow and will not register any movements from the hand and fingers. Furthermore, it is probably higher degree of activity (motion) in arms than in feet for an infant. This might be a reason for better response from arms. Based on what Lars Adde has experienced, he stated that fidgety movements from child's head are very clear and certain. This statement is in accordance with the results obtained here.

Since the best sensors have been presented, the worst one should also be introduced. During this study, the contribution from the trunk sensor has been minimal. One logical explanation could be that the movements from arms and head (via neck) are damped through muscles and joints before affecting the chest, making the trunk movements steady and peaceful. In addition, contraction and expansion of the lungs during respiration can generate some noise, making the signal less informative.

[4] a psychologic phenomenon is perceived as a total configuration or pattern, rising from the relationships among its constituent elements, rather than as discrete elements possessing attributes of their own, and that the pattern, or Gestalt, cannot be derived from the summation of its constituents (Dictionary, 2009).

Medically, a research study can be comprised of two main steps. The first step starts with collection of data. After that, a method is evaluated and selected in connection with the purpose of the study. At last, the chosen methodology is applied, using the collected data and the results are introduced. A way to evaluate the result is to calculate the sensitivity and specificity. At the second step of the research, the presented system/method is tested with data from random patients. The performance of the method is measured by Positive Predictive Value and Negative Predictive Value. The positive predictive value tells you how likely it is that you actually have a disease if you test positive for it. It is defined as the number of true positives (people who test positive who have the disease) divided by the total number of people who test positive, and it varies with test sensitivity, test specificity, and disease prevalence. The negative predictive value of a test is the probability that the patient will not have the disease when restricted to all patients who test negative.

Some of the major differences between these two stages in medical research, are in the data collection part. During the first step, data is recorded from a group of participants. These participants, are known in advance to have some essential features (like a disease), and are chosen based on the predefined characteristic and not randomly. In our case, the infants contributed in data collection have been selected from two divided groups. One group contains infants who are highly at risk of developing CP, based on results from MRI. On the contrary, the other group covers infants that have small chances of being at risk. When it comes to data recoding at the second stage, the participants are selected from a representative assortment (a random selection). That means that for the purpose of this thesis, infants that are brought to hospital for CP examination are a suitable representative population.

In addition, during a 2-years follow-up program, the outcome of the CP prediction for infants in this project (step one) has been known, thus giving the possibility to calculate sensitivity and specificity for the implemented method. Calculation of Positive Predictive Value and Negative Predictive Value is dependent on the number of at-risk patients involved in the test group, while the number of participants with CP has no influence on measurement of sensitivity and specificity.

So, this study contains only the first step of a typical medical research, meaning that the global (generalized) validity of the implemented methods are yet to be investigated in a later project, suppose that a representative selection is available.

7 Conclusion

During the conducted study, unfamiliar fields have been examined, new methods and techniques have been implemented and meaningful results have been achieved.

At the beginning, the concept of Fidgety Movement and General Movement Assessment has been studied and understood. This has led to confirmation of the valuable predictive capabilities of the GMA approach. The idea behind the current project has been to create a computer-based method with hopefully acceptable ability to predict CP among infants, accomplishing an objective diagnosis, rather than the subjective solution produced by GMA method. This goal has successfully been achieved, but more studies are needed on well-defined techniques for feature extraction, selection, combination, and classification. Furthermore, the produced procedure needs to be carried through several tests, in order to define a standardized level of predictive functionality.

Along the way, introduced features from previous studies has been implemented and evaluated, using linear and nonlinear separability values. Later on, a suboptimal algorithm has been designed to select and combine the extracted features. The performance of the introduced algorithm was experimentally shown to converge. Afterwards, the calculated combinations were assessed applying discriminant analysis. At the end, the performance of all mentioned elements of the project has been discussed in-depth.

As a conclusion, the author can certainly state that comparing the sensors, movements in head and arms contained the most discriminative information and the results indicated that features produced by dynamic models were more powerful than statistically characterized features. Furthermore, LDA combined with the relatively optimal feature subset resulted in 86% sensitivity and 90% specificity.

From the clinical perspective, the results are highly interesting and acceptable, but this study needs further attendance, before having any clinical usability. The final and futuristic purpose and objective is to get the reliability and credibility of the introduced procedure confirmed. Afterwards, this study can be an essential part of a sophisticated system, which will assist many doctors and physicians in prediction of CP.

8 Bibliography

Berg, A. (2008), Modellbasert klassifisering av spedbarns bevegelser, Master's thesis, Norwegian University of Science and Technology, Trondheim, Norway.

Dictionary, T. F. (2009), 'Gestalt psychology'.
URL: *http://medical-dictionary.thefreedictionary.com/Gestalt+perception*

E.Beckung, E.Brogren, B. (2002), *Sjukgymnastik før barn och ungdom. Teori och tillampning*, Narayana Press, Denmark.

F.Palmer (2002), *Editorial: First, observe the patient*, Arch Pediatr Adolesc Med 156, pp. 422-423.

H.F.Prechtl (1997*a*), *Editorial: State of the art of a new functional assessment of the young nervous system. An early prediction of cerebral palsy*, Early Hum Dev, 50, pp. 1-11.

H.F.Prechtl (1999), 'General movement assessment as a method of developmental neurology: new paradigms and their consequences'.

H.F.Prechtl, C. (1997*b*), *An early marker for neurological defictis after perinatal brain lesions*, Lancet, 349(9062), pp. 1361-1363.

H.F.Prechtl, J. (1984), *Fetal motility in the first half of pregnancy*, Blackwell Scientific Publicatinos, Oxford.

H.Prechtl, C. (2004), 'Prechtl's method on the qualitative assessment of general movements in preterm, term and young infants'.

J.G.Proakis, D. (2007), *Digital Signal Processing, Fourth Edition*, Pearson Prentice Hall.

L.Adde (2004), General movement assessment predicting cerebral palsy in infants, Master's thesis, Norwegian University of Science and Technology, Trondheim, Norway.

L.Adde, S. (n.d.), 'Categorising movement data(wo//2007/029012)'.
URL: *http://www.wipo.int/pctdb/en/wo.jsp?wo=2007029012*

Meinecke (2003), *Movement analysis in early diagnosis of a developing spasticity in newborns with infantile cerebral palsy*, Gait and Posture Vo. 18, Suppl. 2, 96-98.

Meinecke (2004), *Procedure for the classification of movement analysis data in early diagnosis of a developing spasticity in newborns with ICP*, Gait and Posture Vo. 20, Suppl. 1, 61-112.

Meinecke (2006), 'Movement analysis in the early detection of newborns at risk for developing spasticity due to infantile cerebral palsy'. Human Movement Science 25, pp. 125-144.

M.Hack (2001), *The outcome of neonate intensive care, 5th edition*, W.B Saunders Company, Philadelphia, USA.

M.Hadders-Algra (1996), *The assessment of general movements is a valuable technique for the detection of brain dysfunction in young infants. A review.*, Acta Pediatr Suppl 416, pp. 39-43.

P.Sherrod (n.d.), 'Linear discriminant analysis'.
URL: *http://www.dtreg.com/lda.htm*

R.O.Duda, P.E.Hart, D. (2001), *Pattern Classification, second edition*, Wiley-Interscience Publication.

R.Polikar (2006), 'Pattern recognition'.

Wikipedia, T. F. E. (2008), 'Skewness'.
URL: *http://en.wikipedia.org/wiki/File:Skewness_Statistics.svg*

Wikipedia, T. F. E. (2009*a*), 'Feature extraction'.
URL: *http://en.wikipedia.org/wiki/Feature_extraction*

Wikipedia, T. F. E. (2009*b*), 'K-means clustering'.
URL: *http://en.wikipedia.org/wiki/K-means_clustering*

Wikipedia, T. F. E. (2009*c*), 'Linear discriminant analysis'.
URL: *http://en.wikipedia.org/wiki/Linear_discriminant_analysis*

Wolfram MathWorld, t. w. m. e. m. r. (2009*a*), 'Euclidean distance'.
URL: *http://mathworld.wolfram.com/Distance.html*

Wolfram MathWorld, t. w. m. e. m. r. (2009*b*), 'Skewness'.
URL: *http://mathworld.wolfram.com/Skewness.html*

Appendix A Tables

Table 7: Definition of GMs and their abnormal appearance officially agreed upon by the GM Trust (H.F.Prechtl, 1999)

Period	Normal general movements	Abnormal general movements
Prenatal and preterm age	Gross movements, involving whole body. They may last from a few seconds to several minutes or longer. Variable sequence of arm leg, neck, and trunk movements. Wax and wane in intensity, force, and speed, and have a gradual beginning and end. Majority of sequences of extension and flexion movements of arms and legs are complex, with superimposed rotations and often slight changes movement. These added components in the direction of the make the movements fluent and elegant and create the impression of complexity and variability.	Poor repertoire of general movements: the sequence of the successive movement components is monotonous and the movements of the different body parts do not occur in the complex way as seen in normal GMs. Cramped-synchronized general movements: these appera rigid and lack normal smooth and fluent character; all limb and trunk muscles contract and relax almost simultaneously.
Term age until 8 weeks' postterm age	Writhing movements are characterized by small-to-moderate amplitude and by slow to moderate speed. Fast and large extension movements may occasionally break through, particularly in the arms. Typically, such movements are elliptical in form; this component creates the impression of writhing quality of movement.	Chaotic general movements: movements of all limbs are of large amplitude and occure in a chaotic order with no fluency nor smoothness. They consistently appear to be abrupt.
6 to 20 weeks' postterm age	Fidgety movements are circular movements of small amplitude and moderate speed and variable acceleration of neck, trunk, and limbs in all directions. They are continual in the awake infant, except during focused attention. They may be concurrent with other gross movements, such as kicking, wiggling-oscillating and swiping of the arms or pleasure bursts. Fidgety movements may be seen as early as 6 weeks postterm but usually occur around 9 weeks and are then present until 15 to about 20 weeks.	Absent fidgety movements: fidgety movements are never observed from ages 6 to 20 weeks postterm. Other movements can, however, be commonly observed. Abnormal fidgety movements: look like normal fidgety movements except that their amplitude, speed, and jerkiness are moderately or greatly exaggerated.

Table 8: Result for Scatter Matrix

(a) Skewness

Name of Feature	Separability
J_{Chest}	1.0040
$J_{LeftFoot}$	1.0014
J_{Head}	1.0013
$J_{LeftFoot}$	1.0007
$J_{RightArm}$	1.0004
$J_{RightArm}$	1.0002

(b) Cross-correlation

Name of Feature	Separability
$J_{Acceleration,Arms}$	1.0205
$J_{Acceleration,Feet}$	1.0041

(c) Area differing from moving average

Name of Feature	Separability
$J_{Velocity,Feet}$	1.0401
$J_{Position,Feet}$	1.0282
$J_{Velocity,Arms}$	1.0096
$J_{Position,Arms}$	1.0024

(d) Area out of standard deviation

Name of Feature	Separability
$J_{Velocity,Feet,std}$	1.0339
$J_{Velocity,Arms,std}$	1.0112
$J_{Position,Arms,std}$	1.0032
$J_{Position,Feet,std}$	1.0020

(e) Periodicity

Name of Feature	Separability
$J_{Position,Arms}$	1.1859
$J_{Velocity,Arms}$	1.1489
$J_{Position,Feet}$	1.0379
$J_{Velocity,Feet}$	1.0367

Table 9: Result for Scatter Matrix, all sensors and axes

(a) Separability of sensors along the x-axis

Name of Feature	Separability
$Sensor3_x$	1.0635
$Sensor4_x$	1.0520
$Sensor6_x$	1.0431
$Sensor2_x$	1.0424
$Sensor1_x$	1.0345
$Sensor5_x$	1.0243

(b) Separability of sensors along the y-axis

Name of Feature	Separability
$Sensor6_y$	1.0921
$Sensor3_y$	1.0476
$Sensor4_y$	1.0394
$Sensor2_y$	1.0339
$Sensor1_y$	1.0185
$Sensor5_y$	1.0032

(c) Separability of sensors along the z-axis

Name of Feature	Separability
$Sensor4_z$	1.0383
$Sensor6_z$	1.0315
$Sensor3_z$	1.0293
$Sensor1_z$	1.0093
$Sensor5_z$	1.0042
$Sensor2_z$	1.0035

Table 10: Result for Scatter Matrix, distance between sensors and origo

(a) Distance calculated in xy-plane

NAME OF FEATURE	SEPARABILITY
$Sensor4_{xy}$	1.0515
$Sensor6_{xy}$	1.0426
$Sensor3_{xy}$	1.0389
$Sensor5_{xy}$	1.0262
$Sensor1_{xy}$	1.0245
$Sensor2_{xy}$	0

(b) Distance calculated in xz-plane

NAME OF FEATURE	SEPARABILITY
$Sensor3_{xz}$	1.0484
$Sensor4_{xz}$	1.0414
$Sensor6_{xz}$	1.0312
$Sensor1_{xz}$	1.0275
$Sensor5_{xz}$	1.0054
$Sensor2_{xz}$	0

(c) Distance calculated in zy-plane

NAME OF FEATURE	SEPARABILITY
$Sensor4_{zy}$	1.0347
$Sensor6_{zy}$	1.0262
$Sensor3_{zy}$	1.0240
$Sensor1_{zy}$	1.0141
$Sensor5_{zy}$	1.0031
$Sensor2_{zy}$	0

(d) Distance calculated in absolute value

NAME OF FEATURE	SEPARABILITY
$Sensor4_{abs}$	1.0390
$Sensor3_{abs}$	1.0381
$Sensor6_{abs}$	1.0272
$Sensor1_{abs}$	1.0230
$Sensor5_{abs}$	1.0017
$Sensor2_{abs}$	0

Table 11: Result for Scatter Matrix, distance between sensors and PCA-axis

NAME OF FEATURE	SEPARABILITY
$Sensor6_{PCA}$	1.0899
$Sensor3_{PCA}$	1.0703
$Sensor4_{PCA}$	1.0537
$Sensor2_{PCA}$	1.0468
$Sensor5_{PCA}$	1.0282
$Sensor1_{PCA}$	1.0078

Appendix B Source code from Matlab

Comment: All my programs are tested in `Matlab R2008a` and require the `Statistics Toolbox`.

Appendix B.1 K-means Clustering

Appendix B.1.1 `k-mean-Separability.m`

```
% Should be used as the starting function. Calculates the separability
% of clusters based on Euclidean and Mahalanobis distances. The clustering
% algorithm applied is k-means clustering, which can be found in Help.

%% Modification History
% When          Who                 What
%-----------------------------------------------------
% 2009.05.01    Parsa Rahmanpour    Original version
%-----------------------------------------------------
load  '../Saved/Features/Normal_Features_for_Clustering.mat'
load  '../Saved/Features/Abnormal_Features_for_Clustering.mat'

columnN=size(Normal_Features_for_Clustering,2);
columnA=size(Abnormal_Features_for_Clustering,2);

% Define the criteria function for selecting the number of clusters
% func = 'mahalanobis';
% func = 'euclidean';

Result_clustering = zeros(columnN,1);
% Separability for each feature
for i = 1:columnN
    n = Normal_Features_for_Clustering(:,i);
    ab = Abnormal_Features_for_Clustering(:,i);
    current_dist = mahal_kmean(n,ab,i,func);
    Result_clustering(i)=current_dist;
end
[d,i] = sort(Result_clustering, 'descend');
```

Appendix B.1.2 `mahal-kmean.m`

```
% Starts by finding the clusters and calculating the centroids of each
% cluster. The input consists of data from normal and abnormal babies, in
% addition to the features previously found. "func" is used to define the
% appropriate distance method, chosen by the user. This function returns
% the final mean distance between clusters
```

```
%% Modification History
% When          Who                 What
%--------------------------------------------------
% 2009.05.01    Parsa Rahmanpour    Original version

function mahal_average = mahal_kmean(normal,abnormal,feature,func)

% Clustering only the training sets from Normal and Abnormal data
if(strcmp(func, 'mahalanobis'))
    [cidx1, cnt1, k1] = run_kmean_Mahala(normal);
    [cidx2, cnt2, k2] = run_kmean_Mahala(abnormal);
end
if(strcmp(func, 'euclidean'))
    [cidx1, cnt1, k1] = run_kmean_Euclid(normal);
    [cidx2, cnt2, k2] = run_kmean_Euclid(abnormal);
end

% Mahalanobis distance between all the points within each cluster and
% all the points inside the opposite cluster
columnN=size(normal,2);
columnA=size(abnormal,2);

for i = 1:k1
    for j = 1:columnN
        eval(['cluster_' num2str(i) '_n(:,j)= [normal(cidx1==i,j)];'])

    end
end
for i = 1:k2
    for j = 1:columnA
        eval(['cluster_' num2str(i) '_ab(:,j)= [abnormal(cidx2==i,j)];'])
    end
end
for i = 1:k1
for j = 1:k2
try
 eval(['tmp = mahal(cluster_' num2str(i) '_n, cluster_' num2str(j) '_ab);' ])
 eval(['Norm = length(cluster_' num2str(i) '_n) * length(cluster_' num2str(j) '_ab);' ])
 MEAN = mean(sqrt(tmp))/Norm;
 c = isnan(MEAN);
 if(c == 1)
     eval(['mahal_' num2str(i) '_n_' num2str(j) '_ab = 0;'])
 else
     eval(['mahal_' num2str(i) '_n_' num2str(j) '_ab = MEAN;'])
```

```
 end

catch ME
 errmsg = ME.identifier;
 if(strcmp(errmsg,'stats:mahal:TooFewRows'))
 sprintf('**Error: Singelton cluster')
 else
     disp(errmsg)
 end
end

end
end

for i = 1:k1
  for j = 1:k2
      eval(['TableOfDistance(i,j) = mahal_' num2str(i) '_n_' num2str(j) '_ab;'])
  end
end
mahal_average = mean(mean(TableOfDistance));
sprintf('Separability of feature %d is %d',feature,mahal_average)
end
```

Appendix B.2 Sequential Feature Selection

Appendix B.2.1 SeqFeatSelect.m

```
% Using the function "kombiner.m", the "sequentialfs" is run as many times
% as the variable "iter" defines it. Remember that the computation time
% appears to be exponential.
% This code calculates the number of times that each feature has appeared
% in the subset
%% Modification History
% When         Who                 What
%------------------------------------------------------
% 2009.05.01    Parsa Rahmanpour    Original version

load '../Saved/Features/Features_Tr.mat' Features_Tr;
TrainingData = Features_Tr;
[Xrow,Xcolumn] = size(TrainingData);
% Choose iter to be the number of iteration you wish
iter = 10;
method = {'linear', 'diagquadratic'};
result_lin = zeros(Xcolumn,1);
result_quad = zeros(Xcolumn,1);
```

```
for t = 1:length(method)
    clear comb;
    for s = 1:iter
        [fs, c] = kombiner(TrainingData, cell2mat(method(t)));
        comb(s,:) = fs;
    end

    [row, column]=size(comb);
    if(t == 1)
        for i = 1:row
            for j = 1:column
                if(comb(i,j)==1)
                    result_lin(j) = result_lin(j) + 1;
                end
            end
        end
    end

    if(t == 2)
        for i = 1:row
            for j = 1:column
                if(comb(i,j)==1)
                    result_quad(j) = result_quad(j) + 1;
                end
            end
        end
    end
end
```

Appendix B.2.2 kombiner.m

```
% Using the function "sequentialfs" defined in Matlab (see Help) to
% remove the poor features from the subset. The selection criterion is
% the performance of a classifier
%% Modification History
% When          Who                 What
%-----------------------------------------------------
% 2009.05.01    Parsa Rahmanpour    Original version

%%
function [fs, c ] = kombiner(X, Func)
% **************************FEATURE SELECTION***************************%
n='Normal';
m='Abnormal';
```

```
% N = 198;
% M = 48;
N = 101;
M = 27;
for i=1:N
    Y(i)=cellstr(n);
end
for i=1:M
    Y(i+N)=cellstr(m);
end
y = Y';

% c = cvpartition(group,'holdout',p)randomly partitions observations into a
% training set and a test set with stratification, using the class
% information in group; that is, both training and test sets have roughly
% the same class proportions as in group. The parameter p must be a scalar.
% When 0 < p < 1, cvpartition randomly selects approximately p*n
% observations for the test set. When p is an integer, cvpartition
% randomly selects p observations for the test set.
% The default value of p is 1/10.
c = cvpartition(y,'holdout',1/3);
opts = statset('display','iter');
fun = @(XT, yT, Xt, yt)(sum(~strcmp(yt, classify(Xt,XT,yT,Func))));
fs = sequentialfs(fun,X,y,'cv',c,'direction','backward','options',opts);
end
```

Appendix B.2.3 run-classify.m

```
% Calculating sensitivity and specificity which will be used as
% selection criteria in "Manueal_Kombinering"
%% Modification History
% When          Who                 What
%---------------------------------------------------
% 2009.05.01    Parsa Rahmanpour    Original version

function [Sensitivitet, Spesifisitet] =
run_classify(Features_Tr, Features_Te, y_Tr,func)

test_sett_norm = 97;
test_sett_abnorm = 21;

[C,err,P,logp,coeff] = classify(Features_Te, Features_Tr, y_Tr, func);

TN=sum(C(1:test_sett_norm));
FP=test_sett_norm - TN;
```

```
FN=sum(C(test_sett_norm + 1: test_sett_norm + test_sett_abnorm));
TP=test_sett_abnorm - FN;

Sensitivitet=TP/(TP+FN);
Spesifisitet=TN/(TN+FP);
end
```

Appendix B.2.4 convergence.m

```
% Plots the convergence of the "SeqFeatSelect.m" based on the number
% appearance of each feature during every iteration of the algorithm.
%% Modification History
% When          Who                 What
%----------------------------------------------------
% 2009.05.01    Parsa Rahmanpour    Original version

%%
clear all
clc
close all
load  '../Saved/Combination/New/10/result_lin.mat' result_lin;
load  '../Saved/Combination/New/10/result_quad.mat' result_quad;
[data_lin_10, index_lin_10]=sort(result_lin, 'descend');
[data_quad_10, index_quad_10]=sort(result_quad, 'descend');

load  '../Saved/Combination/New/100/result_lin.mat' result_lin;
load  '../Saved/Combination/New/100/result_quad.mat' result_quad;
[data_lin_100, index_lin_100]=sort(result_lin, 'descend');
[data_quad_100, index_quad_100]=sort(result_quad, 'descend');

load  '../Saved/Combination/New/1000/result_lin.mat' result_lin;
load  '../Saved/Combination/New/1000/result_quad.mat' result_quad;
[data_lin_1000, index_lin_1000]=sort(result_lin, 'descend');
[data_quad_1000, index_quad_1000]=sort(result_quad, 'descend');

load  '../Saved/Combination/New/10000/result_lin.mat' result_lin;
load  '../Saved/Combination/New/10000/result_quad.mat' result_quad;
[data_lin_10000, index_lin_10000]=sort(result_lin, 'descend');
[data_quad_10000, index_quad_10000]=sort(result_quad, 'descend');

% Choose your wanted feature ranging from 1 to 84
feature = 20;
count = 1;
for i = [10 100 1000 10000]
```

```
    eval(['ind_l = find(feature==index_lin_' num2str(i) ');'])
    eval(['ind_q = find(feature==index_quad_' num2str(i) ');'])
    eval(['percent_l(count) = 100 * data_lin_' num2str(i) '(ind_l) /i;'])
    eval(['percent_q(count) = 100 * data_quad_' num2str(i) '(ind_q) /i;'])

    sprintf('Linear: Feature nummer %d has %d percent for %d iterations'
     ,feature,percent_l(count),i)
    sprintf('Quadratic: Feature nummer %d has %d percent for %d iterations'
     ,feature,percent_q(count),i)
    count = count + 1;
end

% Plot
xRange = [0 5];
yRange = [0 101];
x = [1 2 3 4];
y1 = percent_l;
y2 = percent_q;
plot(x,y1,'gx'); hold on
plot(x,y2,'bx');
set(gca,'XTick',[1:1:4])
set(gca,'XTickLabel',{'10','100','1000','10000'})
xlim(xRange); ylim(yRange);
ylabel('[%]'); xlabel('number of iterations');
title('Convergence of SequentialFS for feature number 20');
legend('Linear','Quadratic','location','SouthEast');
```

Appendix C DVD

Here is a short explanation of what is put on the DVD.

References Contains most of the references as PDF files, and the BibTeX file `bibliography.bib`.

LaboratoryWork Contains all files from the laboratory.