Czech Technical University in Prague

Faculty of Electrical Engineering

Department Circuit Theory

SPECIFIC LANGUAGE IMPAIRMENTS AND POSSIBILITIES OF CLASSIFICATION AND DETECTION FROM CHILDREN'S SPEECH

Doctoral Thesis

Ing. Pavel Grill

Prague, August 2015

Ph.D. Programme: Electrical Engineering and Information Technology

Branch of study: Electrical Engineering Theory

Supervisor: prof. Ing. Jana Tučková, CSc.

I would like to express my deep gratitude to my supervisor prof. Ing. Jana Tučková, CSc. for guiding me throughout my research. Her friendly support, patience and theoretical and practical help have been paramount to the successful completion of my PhD study. Further, I would like to thank to my loving fiancée and whole family for all their support that made it possible for me to finish this thesis.

Acknowledgements

This work was supported by following grant:

- grant NR8287-3/2005 „**Computer Analysis of Speech and Overnight EEG in Children**"granted by IGA MH _CR agency (Science Foundation of Ministry of Health of the Czech Republic)

- grant NT11443-5/2010 „**Computer Analysis of Speech Expression, EEG Records and MR Tractography in Children with Developmental Dysphasia**"granted by IGA MH CR agency (Science Foundation of Ministry of Health of the Czech Republic).

- grant SGS12/143/OHK3/2T/13 „Algorithms and Hardware Realizations of Digital Signal Processing“

- grant GD102/08/H008 „**Analysis and modeling biomedical and speech signals**“ by the Grant Agency of the Czech Republic

- The author would like to thank to American Journal Experts for their thoughtful English corrections.

Abstract

Many young children have speech disorders. My research focused on one such disorder, known as specific language impairment or developmental dysphasia. A major problem in treating this disorder is the fact that specific language impairment is detected in children at a relatively late age. For successful speech therapy, early diagnosis is critical. I present two different approaches to this issue using a very simple test that I have devised for diagnosing this disorder. In this thesis, I describe a new method for detecting specific language impairment based on the number of pronunciation errors in utterances. An advantage of this method is its simplicity; anyone can use it, including parents. The second method is based on the acoustic features of the speech signal. An advantage of this method is that it could be used to develop an automatic detection system.

Key words: SLI, Developmental Dysphasia, pathological children speech processing, children speech database, statistical analysis of children speech, SOM analysis of pathological children speech

Contents

List of Figures

List of Tables

Abbreviations

ANN	Artificial Neural Network
CDS	Create Database Structure
DD	Developmental Dysphasia
DSP	Digital signal processing
FF	Folding Algorithm of Formants
GAPS	Grammar and Phonology Screening
HM	“Hand-made”
KSOM	Kohonen Self-Organizing Maps
LANNA	Laboratory of Artificial Neural Network Applications
LLD	Low-Level Descriptor
LPC	Linear predictive coding
MFCC	Mel-frequency cepstral coefficients
PLP	Perceptual linear predictive analysis
PS	Penalty score
RMSE	Root mean square energy
SAL	Sorting algorithm
SLI	Specific Language Impairment
SOFM	Self-Organizing Feature Map
SOM	Self-Organizing Map
WA	Web administration

1 Introduction

The ability to communicate via spoken language is one of the most important human attributes. Although there are several other means of communication, speech is difficult to replace in everyday life. Inabilities to communicate using speech can isolate individuals from society. Isolation resulting from speech impairments is significant for children with specific language impairments.

This work is part of an ongoing research project that integrates results from the fields of neurology, psychology, logopedics, MRI tractography and speech processing. The aim of this research is to further advance the diagnosis of specific language impairment in children and to develop efficient treatments for this disorder.

Our Laboratory of Artificial Neural Network Applications (LANNA) [1] in the Czech Technical University in Prague collaborates on a project with the Department of Paediatric Neurology, 2nd Faculty of Medicine of Charles University in Prague and with the Motol University Hospital, which focuses on the study of children with SLI. One aim of this project is to record children's speech (the normal and pathological speech of SLI patients) and to obtain data about SLI and speech disorders using automatic utterance analyses with self-organizing neural networks. The goal of this research is to determine which parameters correlate across the results generated from diagnostics (from a number of different specialists, e.g., speech therapists and specialists, psychologists, neurologists, and from EEG and MRI tractography) and tests. Our laboratory uses methods based on computer speech analysis to diagnose children with SLI.

This thesis addresses the issue of identifying specific language impairments in children on the basis of their speech. The aim is to develop methods and procedures that could be utilized to unambiguously classify these children. The procedures described in this thesis are intended to be used as part of a software tool for evaluating treatment progress and assisting physicians in clinical praxis and to create a simple tool for identifying this disorder that can be used anywhere by anyone.

Existing methods and their value for the diagnosis of specific language disorders were examined. The current practice uses various tests to assess children's language skills. This approach is comprehensive but certainly not simple and quick, and it cannot be used to quantitatively test children with SLI. These limitations were a significant factor in the decision to develop a simple test for detection SLI in children. Our method is based on the number of pronunciation errors in the child's utterances. The main idea to create a diagnostic tool that is accessible and easy to use. The use of a mobile device allows extreme flexibility in where the tests are conducted; the test can be performed anywhere (e.g., at home, in kindergarten, at a park, …) and not just in a speech therapist's clinic. The users of this application will be primarily the parents of the children with possible language impairments.

An alternate approach uses methods that diagnose children with SLI using acoustic analysis of the speech signal. The methods for detecting SLI are based on auditory signals that are specific to the acoustic features of speech. These features can be easily obtained, and calculations can be performed without complicated modifications of the speech signal. This method seeks to apply modern techniques to the diagnostic approach. Modern techniques allow the calculation of as many different acoustic features of an audio recording as possible. This approach reduces difficult decisions about which features and methods are relevant to the task but adds requirements for optimization methods and classification methods. The main benefit of this method resides in the possibility of developing an automatic detection system as its foundation, without the need for lengthy preparations of input data or labeling the children's utterances. Labeling utterances is difficult because of mispronunciation and various artefacts caused by fidgeting and because not all errors in the utterances clearly indicate specific language impairments.

The methods and procedures introduced in this thesis are based on different algorithms. Both approaches have one thing in common: the use of artificial neural networks, specifically Kohonen Self-Organizing Maps (KSOMs) [2], for the final classification. KSOMs determine the decision-making criteria used to classify children. Both methods, error analysis and feature analysis, were developed with intent of creating simple and reliable methods for classifying children with SLI and to help determine treatments.

Descriptions of classification experiments are an integral part of the thesis. The experiments involved the construction of evaluation methods that take into account the specific assignment of grant NT11443-5/2010, to which this dissertation also applies. It was necessary to develop such methods to classify children with SLI into three categories (mild, moderate and severe) depending on the degree of SLI without prior knowledge of this factor. The classification into three categories was carried out de facto blindly. To verify the classification results, a speech therapist's assessment was used. The algorithms are based only on the internal dependencies of data obtained from analysis without external knowledge about the children's severity.

To investigate the speech problems of children with SLI, it was necessary to create a speech database. The creation of a speech database designed for research on children with specific language impairments (especially developmental dysphasia) was one of the goals of this thesis. It was a very lengthy process. In my case, it represents almost ten years of constant work. In contrast to other speech databases, data acquisition was much more complicated because I work with children, very often preschoolers; this can create specific problems related to maintaining the children's attention and to parents' reluctance to agree to the research. Currently, the database contains 289 speech recordings from 188 children. The completion of the speech database required manually processing a total of 19,115 words. A precise description of the speech signal was created using the LABELING program [3] to segment the speech signal. Correctly processed data are critical for analyses.

1.1 State of Art

Specific Language Impairment (SLI) [4], [5], [6], [7], [8] is given as a diagnosis when a child has delayed or disordered language development for no apparent reason. It is described as a language disorder that delays the mastery of language skills in children who have no hearing loss or other developmental delays. Other names for this disorder can be developmental language disorder, language delay or Developmental Dysphasia (DD). The problem of developmental language disorder is one of the most common childhood learning disabilities. Almost 7 percent of all children aged 4 to 12 years have this disorder. The impact of the disorder in real life is that a child does not have the same speech skills as other children at the same age because the speech skills are delayed. Children with SLI thus experience the creation of a type of social barrier that separates these children from their contemporaries and disrupts their social lives. The most important factor for the development of a child's language skills and subsequent treatment is the timely diagnosis of SLI. In the following text, we describe a simple rule for classifying SLI based on speech analysis. The sooner treatment begins, the sooner child will reach a so-called "normal" level of language skills for children of their age, even if their language is delayed. According to recent studies, approximately 10 percent of all children experience some type of speech problems, with approximately 7 percent suffering from SLI and between 1 and 2 percent having serious problems with speech that require special attention [9].

SLI is characterized by a specific development of speech with signs of delay and aberration. Children with SLI have problems with the syntactic, semantic and grammatical aspects of their native language. They do not speak with a rich vocabulary, and they have problems making sentences with correct word order. SLI is a primary disease affecting other activities of a given individual (e.g., brain damage, hearing impairment), and secondary expression is reflected in speech and speech execution. Other manifestations of this defect can be a worse short-time memory, problems with attention and concentration, or problems with painting and perception of music. Additionally, these children can show abnormal EEG activity and defects in fine motor skills. The child usually knows that he has problems with speech, despite also experiencing any of these symptoms. The situation can escalate, such that the child stops using speech.

One of the more citate publication is [9], which reflecting findings and interpretations based on the hundreds of studies that have appeared since the publication of the first edition in 1997. Topics include linguistic details (descriptive and theoretical), word and sentence processing findings, genetics, neurobiology, treatment, and comparisons to such conditions as autism spectrum disorders, ADHD, and dyslexia. The epidemiologic study estimated the prevalence of specific language impairment (SLI) in monolingual English-speaking kindergarten children is describes in [10]. [11] is the chapter from comprehension Child Psychology and Psychiatry:

Frameworks for Practice (Second Edition), which introduces the research literature on specific language impairment (SLI). Authors characterize the key features, causes and the major types of SLI, and how children with SLI may be identified and differentiated from children with other developmental disorders. Fewer publications can be found under the name of Developmental Dysphasia. It is for example: A conceptual approach of the Theory Diagnosis and Treatment of Developmental Dysphasia from Speech Language Therapist A. G. Beesems is the transcript of a lecture given in Turkey at the: "disabled 07 Congress" [11]. In the Czech sources are found both names. Speech therapist and phoniatrist more use SLI (speech language impairments), but clinical neurologists rather DD (developmental dysphasia). In this thesis I will (with some exceptions) to use the name of Developmental Dysphasia. In most cases, Czech user can get an information which are available on the websites of organizations associating parents of children with SLI [12], [13].

Research groups have developed various tests for SLI, which are based on grammatical and reading skills. One of them is the Grammar and Phonology Screening (GAPS) test [14]. This test is a quick screening tool that can be administered by a professional or amateur. The grammatical abilities and key pre-reading skills of children at the age of 3.5 to 6.5 years are tested. The GAPS test was standardized on 668 children from across the UK. The goal of this test is to determine whether the children have sufficient knowledge concerning the use of grammatical rules to create sentences and whether they know the rules underlying how to add sounds together to correctly make words.

Speech analyses have already been used in patients with a number of speech difficulties. A large number of works discuss the acoustic characteristics of these speech issues. In particular, paralinguistic analysis, such as the recognition of speech emotion [15], [16], [17], [18], [19], [20], [21], and pathology[22], [23], [24], [25], [26], [27], [28] is increasingly becoming a mainstream topic in speech and language processing. The Interspeech 2013 Autism SubChallenge [29] addresses two developmental disorders with speech manifestations: autism spectrum disorders and specific language impairment. The article [30] aims to provide a broad overview of the paralinguistic analysis as a complex whole. This article describes a broad overview of the constantly growing field by defining the field, introducing typical applications, presenting exemplary resources, and sharing a unified view of the chain of processing.

1.2 Goals of the Thesis

The main motivation of this thesis is to find approaches and methods for classifying children with specific language impairments that provide relevant information to guide speech therapists' decisions. These methods should also provide information about the level of specific language impairments.

This doctoral thesis has the following interrelated goals:

- **The creation of the Children speech database**: The task was to create a specific speech database of children with SLI, with complete data processing and the processed recordings of utterances from the control group (healthy children). For further research, only part of the database, namely, the data related to children aged 6 to 11 years, was used.

- **The identification of children with SLI:** The goals were to propose a method for identifying and classifying children with specific language impairments and to develop a novel classification approach that could help speech therapists to assess children. The method should work using utterances recorded in a speech therapist's office; thus, it must be robust enough to disregard artefacts and noise present in the signal.

- **Proper formant detection**: The motivation for the proper detection of formants from the continuous speech signal is to overcome problems that might hinder the wider application of these methods. The complexity of formant calculation, along with their possible misclassification (i.e., with the Burg algorithm in PRAAT), created a tool that was unsuitable for research purposes. Thus, a neural network was implemented for the proper classification of formants.

- **Discussion of results**: We compare the results obtained from the methods with the findings of a clinical speech pathologist and discuss eventual discrepancies. The data obtained might provide a way to collect a new data and participants in research. The methods should be able to distinguish between healthy children and children with SLI. Additionally, the method is assumed to have potential for further extension to perform more precise classification.

- **Application for classification**: We propose a simple tool for assessing a person's speech skills. The results obtained from the above-described experiments and tests should lead to the development of a practical application.

2 Theory, Description and Application

2.1 Specific Language Impairments and the Relationship to Language Processing in the Brain

Specific language impairment (SLI) or named Developmental Dysphasia (DD) also ([4], [5], [6], [7], [8]) is given as a diagnosis when a child has delayed or disordered language development for no apparent reason. It is described as a language disorder that delays the mastery of language skills in children who have no hearing loss or other developmental delays. These children fail to acquire their native language properly/completely, despite having normal non-verbal intelligence, no hearing problems, and no known neurological dysfunctions or behavioral, emotional or social problems [4]. It is estimated that SLI affects about 5-7% of a kindergarten population [9]. It was demonstrated in various heritability studies, such as genetic etiology studies family evaluations and studies of twins [31] that SLI includes quite a significant genetic component. Another study showed that SLI affects boys much more frequently than girls [9]. The manifestation of impairments are mainly in reducing vocabulary development at early ages and typically exhibit difficulty in manipulating linguistic rules of inflection and derivation. This leads up to wrong syntactic structures, in their native tongue. Usually is better their language comprehension than own production. Children with SLI may have difficulties in non-linguistic cognitive skills, e.g. executive functions, mental rotation, or motor ability [32]. Other difficulties are particularly in other cognitive domains such as working memory [33], [34] and difficulties may be associated with reading impairments [35], [36]

Numerous studies deal with the questions what is underlying problem is causing the observed language difficulties. In these studies are solved directly theories of language acquisition and language representation and processing [4], [37]. Three most frequent hypotheses for the causes of SLI:

1. The children with SLI present a general processing deficit, that is reflected in slower linguistic processing, but have relatively normal linguistic representations [4], [37].
2. The children with SLI have a developmental delay in language acquisition. They have normal linguistic and other cognitive abilities but later timing in the triggering or onset of language acquisition processes [40].
3. The children with SLI have relatively intact cognitive abilities, but difficulties are with the grammar or specific subcomponents of grammar [38], [39].

The impact of the disorder in real life is that a child does not have the same speech skills as other children at the same age because the speech skills are delayed. Children with SLI thus experience the creation of a type of social barrier that separates these children from their contemporaries and disrupts their social lives.

2.1.1 Language Processing in the Brain and the Relationship between Speech and Vocal Tract

Language is processed in several areas in brain [43], [44], [45]. Three fundamental areas for human communication have a critical role in speech and language; Wernicke's area, Broca's area and Angular Gyrus (see in Figure 2.1). They are situated in the dominant hemisphere.

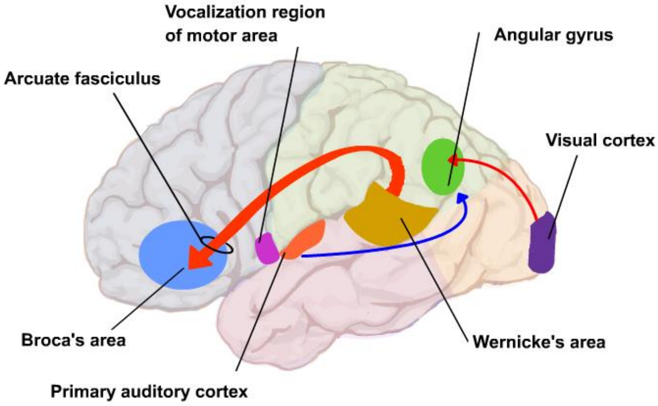


Figure 2.1 Cortical regions of some of the major language areas and pathways (according to [42]).

Broca's area is located in the left hemisphere. It is associated with realization of execution of speech production and articulation. It is also responsible to our ability to articulate ideas, as well as use words accurately in spoken and written language.

Wernicke's area is in the posterior superior temporal lobe. It is connected with Broca's area via a neural pathway. The main function is to involve in the comprehension and understanding speech.

Angular Gyrus is located in close proximity to other critical brain regions such as the parietal lobe (tactile sensation), the occipital lobe (visual analyses) and the temporal lobe (sound analyses). The angular gyrus is associated with complex language functions (i. e. reading, writing and interpretation of what is written. It allows us to associate a perceived word with different images, sensations or ideas.

The vocal cord is situated in larynx and it is the source for speech production. There are two kinds of speech sounds generated; voiced and unvoiced components of speech (see in Figure 2.2). The vibration of vocal cords produces the sound called the voicing and the unvoiced sound due to turbulence of flow of air at a constriction at all possible sites in the vocal tract. The vocal tract is divided into two parts; oral and nasal part. Oral tract is highly mobile and consists of the tongue, pharynx, plate, lips, and jaw etc. This part of vocal tract produces a different speech sounds, which we hear from the lips or nostrils. Nasal tract is immobile part but is coupled with oral part of tract by changing the position of the velum. The shape of the vocal tract responds for some basic frequency produced by vocal cord very well. This is the basic mechanism for the creation of different speech sounds. Information about the speech chain is in supplement 1.

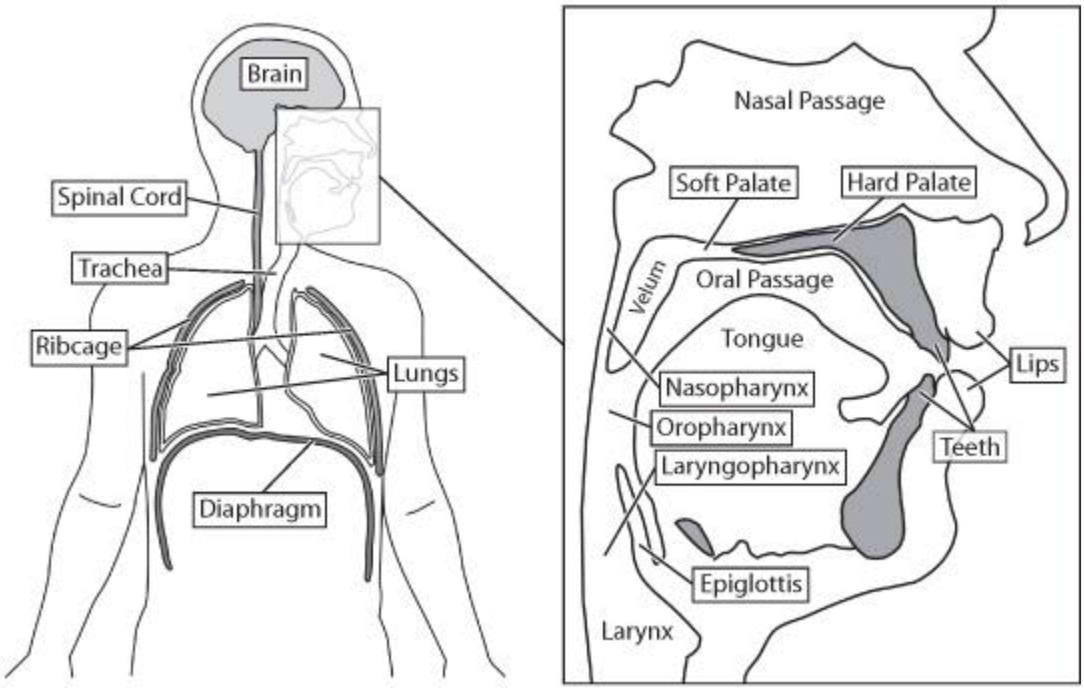


Figure 2.2 Anatomy of speech production. Full body (left) and vocal tract (right) (according to [43]).

2.2 Digital Signal Processing

The classification of the five sense organs, which have become universally accepted, belong sight, smell, taste, touch, and hearing. Primarily hearing and sight are used to communication and understanding between people. Digital signal processing (DSP) has made considerable changes in both these areas. This change allows us to communicate with computer based on human characteristics. DSP is not focused only on hearing and vision but its area of interest are all types of data obtained from sensors, called as signals. DSP is one of the most useful tools and technologies of the current science and engineering which is built on the foundations of the mathematics, algorithms, and the other techniques used to manipulate and processing these signals after they have been converted into a digital form. This affects a wide variety of fields and it is a basic skill needed not only by scientists and engineers but even other experts from a various areas, such as: space, medicine, telecommunication, commercial, military, industry and etc. Each of these areas has developed a deep knowledge of own DSP technology. Our interest is the area of audio signal processing, specifically speech processing. Therefore, there are described techniques dedicated to only this part. More detail for example in [46], [47], [49], [52].

2.2.1 Speech – Numeric Pre-Processing

After the conversion of a continuous time signal into a discrete time signal, we can start with the numeric processing of the signal. The processing usually involves the following several steps:

- **Resampling**

The input signal is usually recorded with a frequency 16 kHz and higher (it is used to select relevant data and to remove the possibility redundancy by down sampling the speech signal).

- **DC Offset**

(removal of the DC component)

- **Normalization**

(this gives us the normalization process in terms of its dynamic to the range of -1 to +1)

- **Pre-Emphasis**

(it is used to equalize the speech frequency spectrum)

- **Segmentation**

(the signal is divided up into quasi-stationary segments; the size of the quasi-stationary segments is in a range of 10-30 ms)

- **Weighting**

(the purpose of this window is to select corresponding signal samples within the analyzed segment and to equalize these signal samples by assigning a specific weight value to them). Generally is used Hamming window. The window function has nonzero value on some interval and zero value outside of that interval. Hamming window is optimized to minimize the maximum (nearest) side lobe (see Figure 2.3). The following equation contains the mathematical description describing the shape of the Hamming window.

$$w[n] = 0.56 - 0.46\cos\left(\frac{2\pi n}{N}\right) \tag{0.1}$$

N $\rightarrow$ number of samples

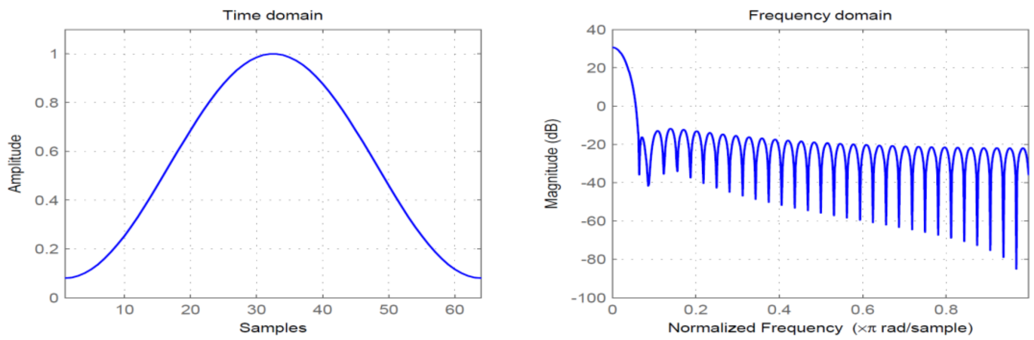


Figure 2.3 Hamming window in time plot and side lobe plot for Hamming window.

Short-Term Analysis – Parametrization

A compromise between two opposing requirements small segment with constant values and enough long segment for accuracy is reached through the use of speech segments of from 10-30 ms. When analyzing speech signals, we have to deal with a significant volume of highly redundant information. The objective of the parametrization process is to reduce the volume of data such that the signal, which is being used for subsequent processing, is not represented by the actual samples of speech collected but rather by a set of particular parameters retrieved from these samples.

2.2.2 Speech Features

It is a basic unit for distinguish of children with and without neurological disorder by using their speech. The following part describes selected speech signal processing algorithms and techniques that I have used in the course of my work. The algorithms used in speech signal processing are relatively widely known, so I will not describe them in details. More detail for example in [46], [47], [52].

2.2.3 Fundamental Frequency

Human speech is a set of audio signals (voiced and unvoiced sounds). Voiced sounds consist of fundamental frequency (F0) and its harmonic components produced by vocal cords. This excitation signal is modificated in the vocal tract, which causes formation of formants and antiformants frequencies. The fundamental frequency is defined as the lowest frequency of a periodic waveform. In this work are used autocorrelation and cepstrum analysis.

- **Autocorrelation method**

The correlation between two waveforms is a measure of their similarity. The waveforms are compared at different time intervals. The mathematical definition is (infinite discrete function $x[n]$ and finite discrete function $x'[n]$):

$$R_x(\nu) = \sum_{n=-\infty}^{\infty} x[n]x[n+\nu] \tag{0.2}$$

$$R_{x'}(\nu) = \sum_{n=0}^{N-1-\nu} x'[n]x'[n+\nu] \tag{0.3}$$

Comparing the found maximum to the zero's autocorrelation coefficinet it can be determined a voiced and unvoiced sounds.

$$R_{\max} \geq \alpha R(0) \rightarrow \text{voiced sounds} \tag{0.4}$$

$$R_{\max} < \alpha R(0) \rightarrow \text{unvoiced sounds} \tag{0.5}$$

$\alpha \rightarrow$ constant must be chosen experimentally

- **Cepstrum Analysis**

Cepstrum is defined as the inverse Fourier transform of the logarithm of the spectrum of a signal. Cepstrum of the voiced phoneme has a strong local peak corresponding to the fundamental period.

$$x_c(n) = DFT^{-1}\left\{\log\left|DFT\left\{x(n)\right\}\right|\right\} \tag{0.6}$$

The predictor coefficients are easily transformed to cepstral coefficients by the recursive relation from:

$$c(n) = a_n + \sum_{i=1}^{n-1}\left(\frac{i}{n}\right)c(i)a_{n-i} \tag{0.7}$$

2.2.4 Formants

Formants are the resonant frequencies of the vocal tract. For their identification can be used, e.g:

- **Burg Algorithms**

This algorithm is used to predict the AR parameters of a model by working directly with signal samples. The algorithm [46] [54] is derived from the lattice structure of the analyzing filter and it minimizes the total forward and backward prediction error. The calculation of the nth value of the reflection coefficient is based on following equation.

$$k_m = -\frac{2\sum_{n=m}^{l_{ram}-1}\left[\left(e^{(m-1)}[n]\right)\left(g^{(m-1)}[n-1]\right)\right]}{\sum_{n=m}^{l_{ram}-1}\left[\left(e^{(m-1)}[n]\right)^2 + \left(g^{(m-1)}[n-1]\right)^2\right]} \tag{0.8}$$

k_m → reflection coefficient

$n = 0,1,2,...,P$ → where P is the final digit in the model

$e^{(m-1)}$ → forward prediction error

$g^{(m-1)}$ → backward prediction error

- **Calculation of Formant Frequencies**

Formants concentrate of acoustic energy in the vicinity of a specific frequency. In speech frequency, there are multiple instances of such peaks (or formants) and each of them is found at a different frequency, which corresponds to roughly one thousand Hz intervals for an adult speaker. The formants are clearly visible on a voice spectrogram, where they appear as dark bands. The darker the band, the more acoustic energy is present. The spectrogram shown here (Figure 2.4) contains five formant bands that appear as red dots.

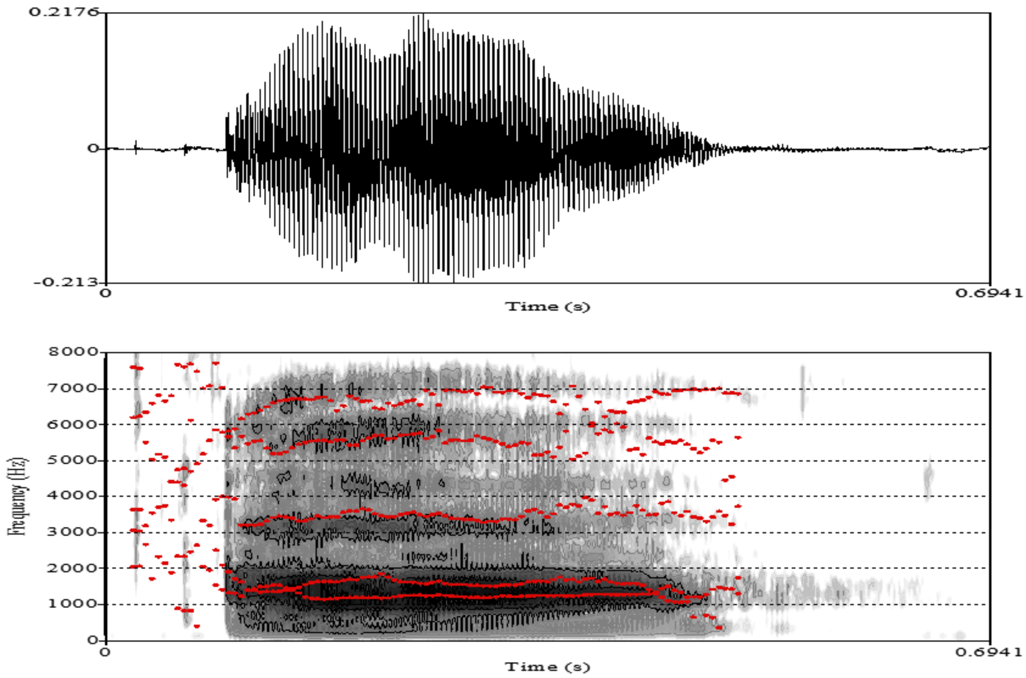


Figure 2.4 Speech signal with spectrogram and five formants bands.

Changes to the speaker's basic tone of voice are interpreted as a change in speech melody. Changes to the first formant (F1) correspond to changes in the vertical movement of the speaker's tongue; changes to the second formant (F2) correspond to changes in the horizontal movement of the tongue; while the third formant (F3) changes with actions taking place in the nasal cavity. This means that the first two-to-three formants are most important when it comes to vowels (in the order of importance F2, F1 and F3). Our research follows that the higher formants except intonation contain information about some pathologies. If we put the first two formants (F1 and F2) into context with each other, we get what we refer to as 'vowel triangle' (see in Figure 2.5).

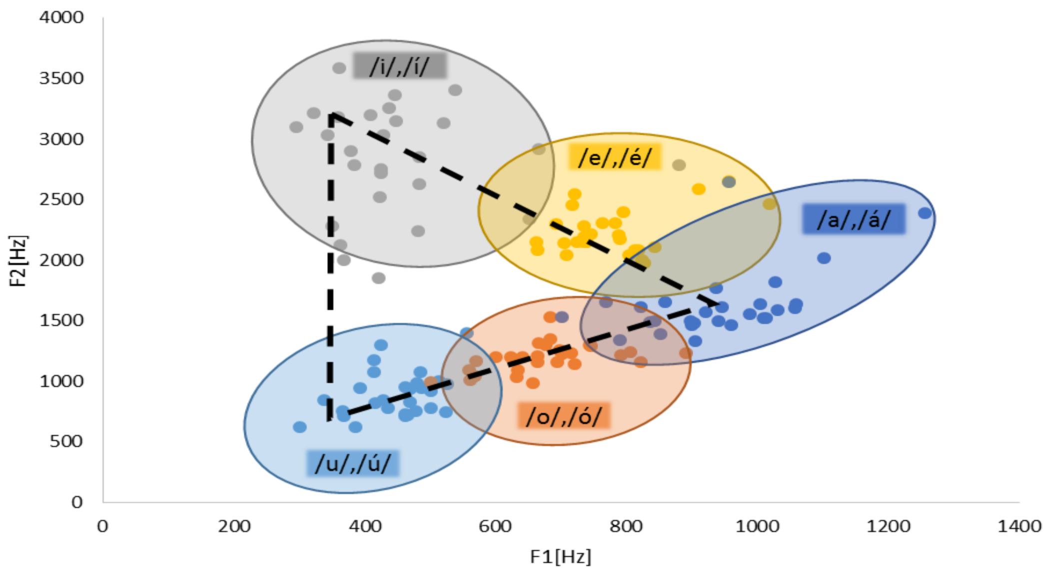


Figure 2.5 Vowel triangle (according to [55]).

Most of the methods [46] and [47] used to identify formant frequencies operate within the frequency spectrum of the given signal and they are based on the analysis of the LPC spectral envelope. There are basically two ways how one can determine formant frequency values from an LPC spectrum. The first method seeks to identify polynomial roots $A(z)$ (i.e. step response $H(z)$ poles); while the second method looks for local speaks within the spectral envelope determined using a linear predictive model. The step response poles are calculated from polynomial roots $A(z)$, which are identified by solving a mathematical equation:

$$z^Q + a_1 z^{Q-1} + a_2 z^{Q-2} + \ldots + a_{Q-1} z + a_Q = 0 \tag{0.9}$$

It is a Q-polynomial scheme equation with real coefficients, the solving of which is dominated by pairs of complex roots. Let's assume we have a single pair of complex roots: $z_i = |z_i| e^{j\phi}$ and $\overline{z_i} = |z_i| e^{-j\phi}$ spread within the z-plane (as shown in Figure 2.6a) ($\varphi(i)$ is the $z(i)[rad]$ argument). The corresponding formant frequency and the formant band width for a 3 dB reduction can then be expressed as follows:

$$F_i = \frac{f_s . \arg z_i}{2\pi} [Hz] \tag{0.10}$$

$$B_i = -\frac{f_s.\arg z_i}{2\pi}\left[Hz\right] \tag{0.11}$$

The relationship between F_i (formant frequency) and B_i (bandwidth) can then be shown as an image within the frequency spectrum (Figure 2.6b).

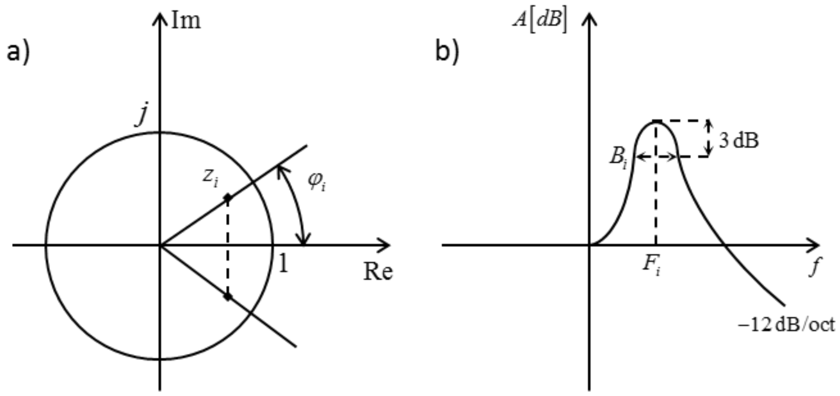


Figure 2.6 Illustration layout (a) pair of complex roots of polynomial A(z), (b) frequency spectrum (according to [46]).

Figure 2.7 shows the blocks of formant detection with the LPC method.

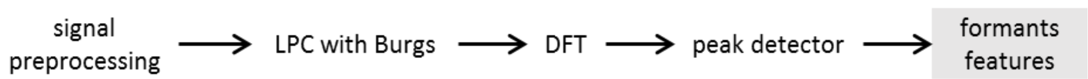


Figure 2.7 Formant detection with LPC method.

2.2.5 Mel-Frequency Cepstral Coefficients

Mel-frequency cepstral coefficients (MFCC) are one of the most used features for speech processing. The MFCCs offer through their cepstral nature abilities to model both poles and zeros. The Mel-frequency cepstrum represents the short-term power spectrum of a sound using a linear cosine transform of the log power spectrum of a mel-scale.

The speech signal is to frames (usually by the Hamming window) and amplitude spectrum $\left|S(f)\right|$ or power spectrum $\left|S(f)\right|^2$ is calculated for each segment by the formula:

$$\left|S(f)\right| = \left|\sum_{n=1}^{N} y[n] e^{\frac{-j2\pi\omega n}{N}}\right| \tag{0.12}$$

The formula for the mel-scale is:

$$f_{mel} = 2959\log_{10}\left(1+\frac{f}{700}\right) \tag{0.13}$$

f_{mel} → Mel-frequency on non-linear mel scale in mel

f → frequency on linear scale in Hz

Number of filters M should be chosen a priory and according to the characteristics of filtered signal. Recommendation for the numbers of filters M in [46].

2.2.6 Linear Predictive Coding

Linear predictive coding (LPC) is a tool used for representing the spectral envelope of a digital signal of speech in compressed form, using the information of a linear predictive model. The main advantages are encoding of good quality speech at a low bit rate and providing of extremely accurate estimates of speech parameters. Even though this method and system for speech recognition is now used less, for the analysis of pathological speech of very small children (4-6 years) has proved to be sufficient.

2.2.7 Perceptive Linear Predictive Analysis

Perceptual linear predictive analysis (PLP) was proposed by Hynek Hermansky. PLP analysis is similar to LPC. The idea of PLP is to approximate the auditory spectrum of speech by an all-pole model. Before approximation by the model, several modifications to the spectrum regards to theories of the psychophysics of hearing are made; the critical-band spectral resolution, equal-loudness curve, and intensity-loudness power law. These concepts improve performance and the parameterization is more robust for speech recognition. Also the LP model is better adapted to properties of human auditory perceptions. The algorithm is described in details in [53].

2.2.8 Signal Energy

The most common way to calculate the energy of a speech signal is the root mean square energy (RMSE) is defined in following formula:

$$E = \left[\frac{1}{N} \sum_{i=1}^{N} s_n^2 \left(i \right) \right]^{\frac{1}{2}} \tag{0.14}$$

2.2.9 Loudness

Loudness [50] is functionally related to sound pressure level, frequency, and waveform. It is a characteristic of a sound that is primarily a psychological correlate of physical strength (amplitude). The loudness level of a sound is given by a following formula:

$$P = 20 \log_{10} \frac{p_m}{P_0} \tag{0.15}$$

P → loudness level

P_m → measured sound pressure (in microbars)

P_0 → sound pressure of 0.0002 microbras

Loudness is a perceptual or subjective quality of a sound. A sone is defined as the loudness heard by typical listeners when confronted with a 1000 Hz tone at a sound pres sure level of 40 phons.

2.2.10 Voice Quality (Jitter, Shimmer)

Jitter and shimmer represent the variations that occur in the fundamental frequency. Jitter indicates the variability or perturbation of fundamental frequency and it is affected mainly because of lack of control of vocal fold vibration. Shimmer is related to amplitude of sound wave, or intensity of vocal emission and it is in relation with reduction of glottic resistance and mass lesions in the vocal folds, which are related with presence of noise at emission and breathiness. The threshold value of jitter or shimmer is used for detecting pathologies, e.g. the threshold value of jitter(absolute) to detect pathologies in adults is 83.2 μs as reported by Guimarães [51].

Jitter measurements:

Jitter (absolute) represents the average absolute difference between two consecutive periods by formula:

$$Jitter_{abs} = \frac{1}{N-1}\sum_{i=1}^{N-1}\left|T_i - T_{i+1}\right| \tag{0.16}$$

T_i → extracted F_0 period lengths

N → number of extracted F_0 periods

Jitter (local) represents the average absolute difference between two consecutive periods, divided by the average period by formula:

$$Jitter_{loc} = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|T_i - T_{i+1}\right|}{\frac{1}{N}\sum_{i=1}^{N} T_i} \times 100 \tag{0.17}$$

Jitter (rap) represents the average for the disturbance, i. e., the average absolute difference of one period and the average of the period with its two neighbors, divided by the average period.

$$Jitter_{rap} = \frac{\frac{1}{N-1}\sum_{i=2}^{N-2}\left|T_i - \left(\frac{1}{5}\sum_{n=i-2}^{i+2} T_n\right)\right|}{\frac{1}{N}\sum_{i=1}^{N} T_i} \times 100 \tag{0.18}$$

Shimmer measurements:

Shimmer (local) represents the average absolute difference of the base 10 logarithm of the difference between two consecutive periods by formula in dB:

$$Shimmer_{loc} = \frac{1}{N-1}\sum_{i=1}^{N-1}\left|20 * \log\left(\frac{A_{i+1}}{A_i}\right)\right| \tag{0.19}$$

A_i → extracted peak-to-peak amplitude data

Shimmer (relatively) represents the average absolute difference between two consecutive periods by formula in percentage:

$$Shimmer_{rel} = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|A_i - A_{i+1}\right|}{\frac{1}{N}\sum_{i=1}^{N} A_i} \times 100 \tag{0.20}$$

2.3 Artificial Neural Networks

The Artificial Neural Networks (ANNs) are a mathematical and statistical learning algorithms inspired by biological neural networks, specifically by human's brain [71], [72], [77].

Next Figure 2.8 illustrates a simple infographic depicting the human brain and its analogy in the form of an artificial neural network that performs mapping brain function. Very clearly, there is indicated a biological neuron and its mathematical adaptation in artificial neuron.

An ANN has to be designed and implemented in a way that the set of input data results into a desired output. There are three major learning algorithms: supervised learning, unsupervised learning and reinforcement learning. I used an unsupervised learning, specifically Self-Organizing Feature Map (SOFM).

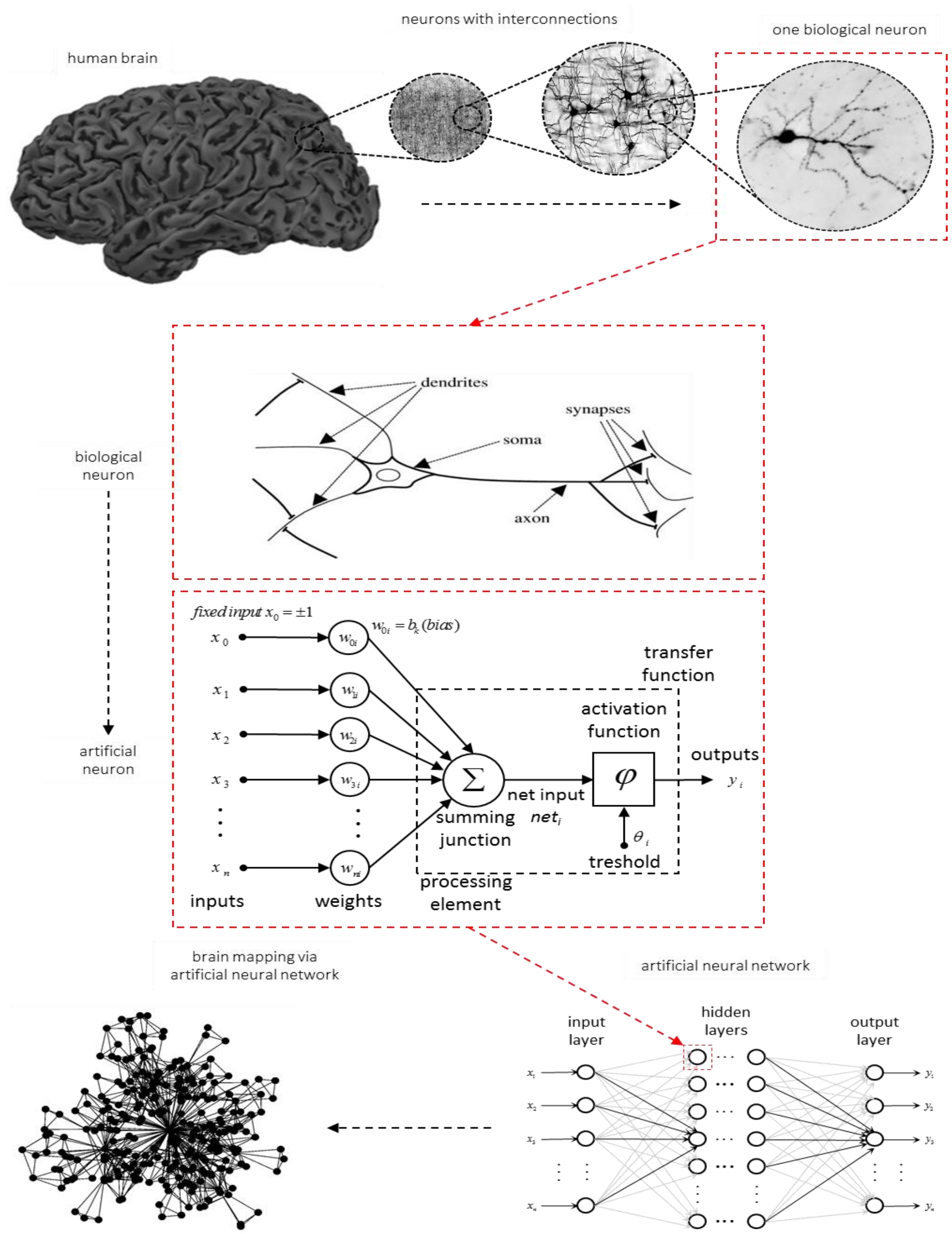


Figure 2.8 The human brain and its analogy in the form of an artificial neural network with graphical representation of biological neuron and its artificial representation [78], [79], [80], [81].

2.3.1 Self-Organizing Maps

The research and development of topologically organized networks [67], [68], [69], [70], [71], [72], [73], [74], [76] is inspired by an attempt to understand how biological neurons come to organize themselves to achieve various tasks. These aims are achieved without any instruction regarding the desired goals for each neuron. The most widely used network with this topology is the model proposed by Teuvo Kohonen, referred to as the Self-Organizing Map (SOM) or Self-Organizing Feature Map (SOFM).

SOM is a one of the most popular neural network models which are based on unsupervised learning. This method provides a data visualization technique which helps to understand high dimensional data by reducing the dimensions of data to a map. SOM also represents clustering concept by grouping similar data together. It converts complex, nonlinear statistical relationships between high-dimensional data items into simple geometric relationships on a low-dimensional display, typically into map with 2 dimensions. SOM reduces data dimensions and displays similarities among data. SOM, as well as a most ANN, operates in two modes. Training and mapping. In this process is combined a competitive learning principle with a topological structuring of nodes such that adjacent nodes tend to have similar weight vectors. In training mode is built a map with using input examples (competitive process) and in the mapping mode is automatically classified a new input vectors.

Competitive learning describes behavior of the neural network with the local connection of nodes - neurons. The most highly activated node, winner of the competition, and its neighbors move towards a sample presented to the network. The self-organizing term means that the nodes (neurons) tend to get their weights of vectors that capture characteristics of the input vector space with the neighborhood relation translating into proximity in Euclidean space. This applies to even if the initial values of weight vectors are set to random values. In clustering, the weight vectors associated with nodes in these networks are interpreted as cluster centroids. In the context of vector quantization, each weight vector is a codebook vector to which input vectors may be mapped. It can be summarized that the number of nodes with weight vectors in a given region is roughly proportional to the number of input vectors in that region.

Figure 2.9 illustrates simple model of self-organizing map. The model consists of two layers. Input layer with nodes representing each input variable and output layer (Kohonen map) that is a 2-dimensional array of nodes.

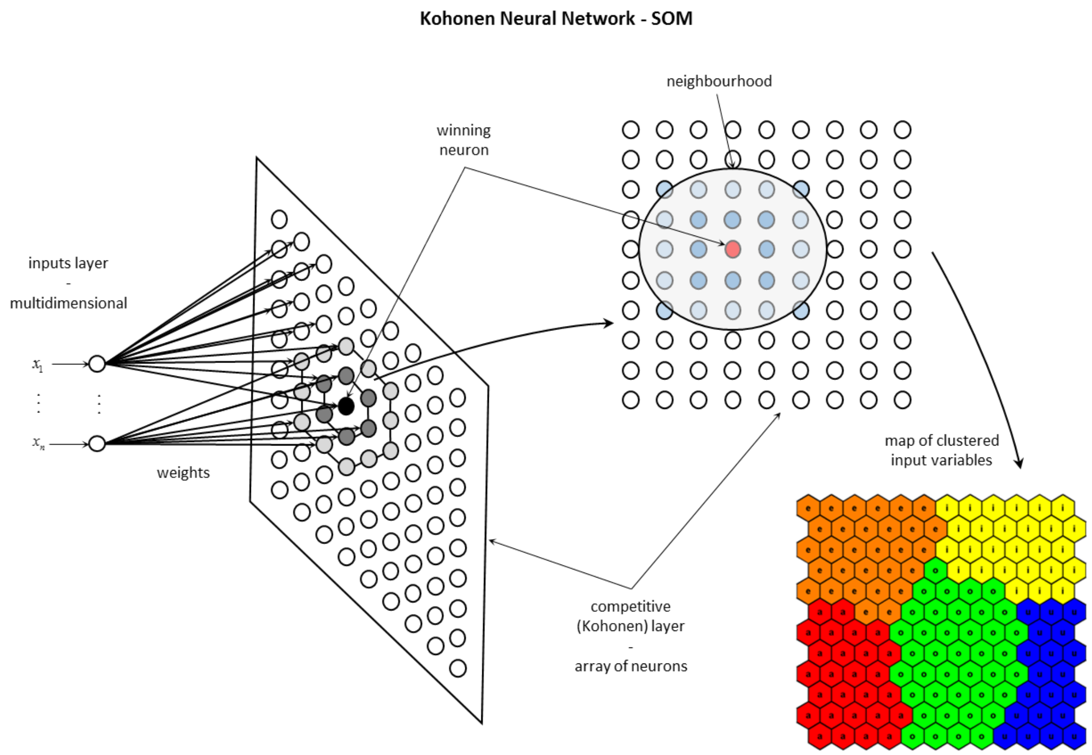


Figure 2.9 A Kohonen Self-Organizing Map illustrating with three key components: input layer, competitive layer and map of clustered variables

2.4 Statistics, analysis

Raw data, also known as primary data, is in our case data from various analysis at children speech. Raw data may have the following attributes: to contain possible errors that are based on the imperfections of methods used in analyses, possible artefacts contained in recordings, data may not be a validated, may be in different formats etc. Statistical analysis is performed to better understand to obtained raw data. Statistical data that is used to draw conclusions and inferences should be accurate and consistent. This is important in order to ensure the validity of all the inferences drawn on the basis of the data. Statistics unifies conditions for evaluation of experiments and results obtained from statistical analysis clearly proving or refuting predetermined hypothesis. Basic formulas and methods of statistical analysis used in this thesis are described in this section. Further information about the statistical methods is in supplement 2 and [56], [57], [58], [59], [60], [61], [62], [63], [64], [65], [66].

2.4.1 Statistical Inference

Statistical inference means drawing conclusions based on data. Inferential statistical analysis infers properties about a population. Inferential statistics can be contrasted with descriptive statistics. Descriptive statistics is solely concerned with properties of the observed data, and does not assume that the data came from a larger population. There are a many contexts in which inference is desirable, and there are many approaches to performing inference.

Hypothesis Testing

Hypothesis testing is a method for testing a claim or hypothesis about a parameter in a population, using data measured in a sample.

The usual process of hypothesis testing consists of four steps:

1. **State the hypotheses.** Formulate the **null hypothesis** $\boldsymbol{H_0}$ (is a claim of "no difference in the population") and the **alternative hypothesis** $\boldsymbol{H_1}$ (is a claim of "a difference in the population"). The null hypothesis is a starting point. We will test whether the value stated in the null hypothesis is likely to be true.
2. **Set the criteria for a decision.** Set the criteria for a decision. State the **level of significance α** for a test. This value refers to a criterion of judgment upon which a decision is made regarding the value stated in a null hypothesis. The criterion is based on the probability of obtaining a statistic measured in a sample if the value stated in the null hypothesis were true.
3. **Compute the test statistic.** The test statistic is a mathematical formula that allows to determine the likelihood of obtaining sample outcomes if the null hypothesis were true (Student's *t*-test, *F*-test, *Shapiro–Wilk* test, etc.).
4. **Make a decision.** The value of the test statistic is used to make a decision regarding the null hypothesis. The decision is based on the probability of obtaining a sample mean, given that the value stated in the null hypothesis is true.
   - **Reject the null hypothesis.** The sample mean is associated with a low probability of occurrence when the null hypothesis is true.
   - **Retain the null hypothesis.** The sample mean is associated with a high probability of occurrence when the null hypothesis is true.

The *p*-value is a probability: It varies between 0 and 1 and can never be negative. We apply the following recommendation:

$p_{val} > 0.10$ → the difference is "not significant"

$p_{val} \leq 0.10$ → the difference is "marginally significant"

$p_{val} \leq 0.05$ → the difference is "significant"

$p_{val} \leq 0.01$ → the difference is "highly significant"

The decision to reject or retain the null hypothesis is called **significance**:

$$p_{val} \leq \alpha \rightarrow \text{reject the null hypothesis (H0)}$$

$$p_{val} > \alpha \rightarrow \text{retain the null hypothesis (H0)}$$

Shapiro-Wilk normality test

We assume that the ability to test the hypotheses with parameters μ and σ^2 is based on the assumption of the normal distribution of the data, which corresponds to the normal Gaussian probability distribution. Thus, it is necessary to perform normality tests. For this purpose is performed the *Shapiro-Wilk* (*SW*) normality test.

Student's t-test: Comparison of two means

The two-sample *t*-test is a parametric test and it is designed to test the differences between two mean values μ. Test verifies the null hypothesis, which stated that the experimental sample is from the population, where the experimental mean value is the same as a reference file. The alternative hypothesis was defined as the exact opposite, stating that the experimental sample is not from the population.

Wilcoxon Rank-Sum test

The Wilcoxon Rank-Sum test (essentially identical to the *Mann-Whitney U* test) is a non-parametric alternative to the two sample *t*-test. Then the requirements for the *t*-test for two independent samples are not satisfied, the Wilcoxon Rank-Sum non-parametric test can often be used provided the two independent samples are drawn from populations with an ordinal distribution. It is even in the case where we accept the alternative hypothesis of the *Shapiro-Wilk* normality test, we cannot consider the statistical tests to be clearly conclusive, and we cannot characterize the data by the parameters μ and σ^2. We are not able to say that the data are from a normal distribution. Then, a non-parametric test must be used. In this test, that is much less sensitive to outliers than the two sample *t*-test, the null hypothesis, which asserts that the medians of the two samples are identical, is tested against the alternative hypothesis, which asserts that they are not

2.4.2 Correlation

Correlation and regression analysis are related in the sense that both deal with relationships among variables. Correlation analysis is used to quantify the association between two continuous variables. It is a statistical method to finding whether two variables are related and to what extent they are related. Correlation is considered as a number which can be used to describe the relationship between two variables.

Spearman Rank Correlation Coefficient is the nonparametric version of the Pearson Product-Moment Correlation. Spearman's correlation coefficient measures the strength of association between two ranked variables. The value of each variable are rank-ordered from 1 to *n*, where *n* is the number of pairs of values. Standard formula is:

$$r_s = 1 - \frac{6\sum d_i^2}{n\left(n^2 - 1\right)} \tag{0.21}$$

r_s → the correlation coefficient

d → the difference between the ranks of corresponding values ($d_i = x_i - y_i$)

2.5 Recording Devices and Software

The precipitate development of technology allows research into human speech on a new qualitative level. We are witnessing a massive utilization of voice technology. With various new devices, it is possible to very effectively plan a daily agenda or communicate with a personal assistant, mostly via mobile phone, using our voice. The human voice may be used for more sophisticated purposes, such medical or forensic applications. All of this is possible only through the use of digital technology. In my research, the recording device serves as an entry point because it is used to obtain data. On the opposite end, the software designed for analysis and classification serves as a figurative output gate.

To collect data (i.e., recorded speech), three types of digital equipment were used: an Apple iBook G4 [82] (sampling frequency fs = 44.1 kHz, 16-bit resolution in mono mode in the standardized vaw format), a SONY MiniDisc recorder (MD MZ-N710) [83] (sampling frequency fs = 44.1 kHz, 16-bit resolution in stereo mode in the standardized vaw format), and a SONY digital Dictaphone (sampling frequency fs = 16 kHz, 16-bit resolution in stereo mode in the standardized vaw format).

For speech processing, I used both professional programs and programs that were developed in the LANNA at CTU FEE in Prague. To process the recordings, I used COOL EDIT PRO 2 [84] (now known as Adobe Audition), commercial software used in the professional sphere for sophisticated work in digital audio formats. The children’s speech was separated from the therapist’s using this program. I developed the original software FORANA [104], [114] for formant analysis. It is based on the MATLAB programming environment. The development of this software was mainly driven by the need to correctly complete formant analyses. Usually, formant frequencies are extracted from speech signals using PRAAT [85] acoustic analysis software. However, because the PRAAT software produced formant classification errors during the analysis (for higher formants, pathological phenomena in the children's speech were reflected only in the higher formants), the results obtained using this approach could not be considered relevant. Another factor taken into consideration in designing the FORANA software was the need to fully automate the process of extracting formants from the recorded speech signals. This application is still being developed. The LABELING [90] program was used to segment the speech signal. It is a part of the SOMLab program system (see the SOMLab paragraph) and was developed in the LANNA at CTU FEE in Prague. The programming package MATLAB [87] can be used for a range of applications, including signal processing and communications, image and video processing, control systems, testing and measurement, and computational biology. With MATLAB, we can analyze data, develop algorithms, and create models and applications. Another program used was the Munich open toolkit Speech and Music Interpretation by Large Space Extraction (openSMILE) [88]. It is a modular and flexible feature extractor for signal processing and machine learning applications with a primary focus on audio signal features. It is written purely in the C++ language and runs on various mainstream platforms, such as Linux, Windows, and MacOS. OpenSMILE is designed for real-time online processing but can also be used on-line in batch mode to process large datasets.

The R Project [89] for Statistical Computing and Graphics was applied. R is an integrated suite of software facilities for data manipulation and calculation and graphical display. SOMLab (SOM Laboratory) is an original program package for conducting one part of the automatic speech analysis [34]. It is based on the Self-Organizing Maps (SOMs) [34] application for speech signal processing, which was developed in the Laboratory of Artificial Neural Networks Application (LANNA) [1]. The program takes advantage of label information from input data and neural networks. SOMLab is a complex system that was created as a user-friendly application of artificial neural networks for speech analysis. The program package consists of the tools necessary for utilizing the SOMs. Individual tools can be divided into three categories: Pre-processing tools (tools for data preparation, data creation and analysis), processing tools (tools for working with the SOMs, such as creation and training), and post-processing tools (tools for the visualization, comparison and analysis of speech). SOMLab works with the computational system MATLAB 7, Release 14 and higher. The software SOM Toolbox [2] was used to create M-files in the SOMLab. It is built using the MATLAB script language and contains functions for the creation, visualization and analysis of Self-Organizing Maps (developed in the Laboratory of Information and Computer Science [CIS] at the Helsinki University of Technology, the toolbox is available free of charge under the General Public License).

2.6 Ethical principles for human research

Scientific research [93] has produced substantial social benefits. It has also posed some troubling ethical questions. The beginning of the ethical principles were created after the Second World War, specifically after the Nuremberg trials. The Nuremberg code was drafted as a set of standards for judging physicians and scientists to research involving human subjects was carried in an ethical manner. The codes contains three principles, or general prescriptive judgments, that are relevant to research involving human subjects; boundaries between practice and research, basic ethical principles and applications.

Ethics Statement

This work is based on the analysis of human's data, specifically speech processing of children. For this purpose it is need to have agreement of ethic committee. This work was conducted at the Department of Circuit Theory of the Faculty of Electrical Engineering of the CTU in Prague. The research was approved by the Ethics Committee of Motol University Hospital in Prague, Czech Republic. All parents provided written informed consent on behalf of their children prior to participation in the study.

3 Speech Databases

To investigate the effect of speech problems of children with SLI, it was necessary to create a speech database [104]. The main of the criteria in creating the corpus were the selection of a suitable text and participants for speech recording. The words and phrases are directly selected for research of children with this disorder and they take into account the physiological and mental development of the child.

3.1 Psycholodical Examination

The children included in the research have to be examined by a clinical psychologist. The examination takes place in the Department of Pediatric Neurology at the 2nd Faculty of Medicine of Charles University in Prague. The parents are present during the examination, which takes one day. The children do not take any medications. The following tests are used during the examination [94], [95]:

- Stanford-Binet Intelligence Test, Fourth Edition
- Special test of perceptual skills, graphomotor skills and visuomotor coordination

These skills influence speech and language mechanisms, as well, which is why they are part of the examination to reveal developmental dysphasia. Further information about the psychological tests is in supplement 3.

3.2 Speaking Tasks

It is necessary to select suitable words, phrases and sentences for recordings speech of children. It is essential to select utterances that children of a given age will be able to articulate. I worked with children aged from 39 to 132 months. Younger children cannot yet read, and they repeat spoken utterances as a result, so it is necessary to maintain the same conditions for all enrolled children. A suitable text for recording contains different types of words and phrases, which were obtained from speech therapists and clinical psychologists who were using common, professionally recognized testing procedures, while also making use of their own knowledge and expertise in the subject area. The specific words, phrases and sentences that comprise the speech database are divided into 13 parts [Table 3.5].

3.3 Record of a Child Speech

The recordings were created in a doctor´s office in the faculty hospital and private speech therapist's office. Only the speech therapist and the participants were present at the recording because it is very difficult to keep a child's attention until the end and the presence of another person would be disturbing. The session took place in the following manner: the speech therapist read a chosen text and the child repeated it. The sessions were structured in this way to standardize the conditions for both groups of children (healthy or SLI children). Both of the speakers are recorded by lapel microphone into the computer. The sound recordings are saved in the standardized vaw format. The sampling frequency is set to 44.1 kHz with 16-bit resolution in mono mode.

One software packs was created to simplify the creation and management of this speech database, and it is designed for research; Create Database Structure (CDS).

- Create Database Structure (CDS):

The program for automatic creation of the database structure was developed in the MATLAB environment. The required parameters include: Name of child, Surname of child, Abbreviation of child, Date of the recording and Number of recordings. The button entitled "Create a directory" is used for the selection of the working directory in which the directory structure of the child is created.

3.4 Speech Databases

The entire database contains three subgroups of recordings of children's speech from different types of speakers. The first subgroup (healthy children, or H-CH) consists of recordings of children without speech disorders; the second subgroup (SLI-CH I) consists of recordings of children with SLI and the third subgroup (SLI-CH II) consists of children who have SLI of different degrees of severity (1 – mild, 2 – moderate, and 3 – severe). The speech therapists and specialists from Motol Hospital decided upon this classification. The children's speech was recorded in the period 2003-2013. The database has two specific parts. The first part is the recording database. These databases were commonly created in a schoolroom or a speech therapist's consulting room, in the presence of surrounding background noise. This situation simulates the natural environment in which the children live, and is important for capturing the normal behavior of children. The second part consists more of the recordings of individual children. Using this approach, we are able to compare the course of therapy.

3.4.1 SLI-CH I Database

This database contains the recordings of the pediatric patients with SLI. The trends in the development of the disease during the given time period (approximately 3 months) were the determining factors, rather than the degree of severity of the children's diagnosis. The patients were divided into two groups: one group took medication, while the other was a control group. The SLI-CH I database contains a total of 46 native Czech participants (33 boys, 13 girls) aged 39 to 132 months, and was recorded during the period 2005–2008.

3.4.2 SLI-CH II Database

This database was recorded in a speech therapist's office [105]. For the purposes of the research study, the children are separated into three groups depending upon the severity of the neurological diagnosis (mild, moderate, or severe). The SLI-CH II database contains a total of 72 native Czech participants (46 boys, 26 girls) aged 58 to 152 months, and was recorded during the period 2009–2013. The lower age limit had to be raised due to the need for MR tractography examination to be performed [97]. The recordings of some children are repeated after several months. The description of all databases is provided in Table 3.1. The recordings were created at a speech therapy office. The speech therapist reads a piece of a text to the child subject, who then repeats the text. Both of the speakers are recorded by microphone into the computer.

	H-CH		**SLI-CH I**	**SLI-CH II**
	Healthy	**With Defect**		
Girls	45	16	13	26
Number of recordings	45	16	22	45
Boys	25	17	33	46
Number of recordings	25	17	64	88
All children	70	33	46	72
All recordings	70	33	86	133
All utterances	4620	2178	5676	8819

Table 3.1 Description of All Databases.

3.5 Dataset for analysis

Source data for this experiment are the utterances obtained from the H-CH and SLI-CH II subgroup from speech database. It was determined five age categories depending on the age in months for both subgroups (A1 = 76.5 ± 6.5, A2 = 89.5 ± 5.5, A3 = 101.5 ± 5.5, A4 = 113.5 ± 5.5 and A5 = 125.5 ± 5.5).

In the H-CH subgroup was selected 44 participants (14 boys and 29 girls) at 70 to 131 months of age (mean age = 106 ± 15.4, median = 110 and range = 70 – 124 months) and in the subgroup

SLI-CH II was selected 54 participants (35 boys and 19 girls) at 70 to 131 months of age (mean age = 96 ± 16.3, median = 94 and range = 70 – 131 months). The subject details of the H-CH subgroup are listed in Table 3.2 and Table 3.3.

Age category	Subjects code	Number		
		Boys	Girls	All partcipants
A1	H26 - H30	1	4	5
A2	H31 - H37	2	5	7
A3	H38 - H44	1	6	7
A4	H45 - H61	7	10	17
A5	H62 - H69	4	4	8

Table 3.2 List of participant details from the H-CH subgroup (including age category and the number of boys, girls and total participants).

Age category	Subjects code	Number		
		Boys	Girls	All partcipants
A1	P8 - P24	11	6	17
A2	P25 - P35	8	3	11
A3	P36 - P49	7	7	14
A4	P50 - P56	4	3	7
A5	P57 - P61	4	1	5

Table 3.3 List of participant details from the SLI-CH II subgroup (including age category and the number of boys, girls and total participants).

The chosen structure of utterances included a range of words and phrases, altogether 68 different utterances (see in Table 3.4). For my research were used tasks T1 to T7.

Task code	Description	# Patterns		Utterances
[T1]	Vowels	5	Czech	"a", "o", "u", "e", "i"
			English	"a", "o", "u", "e", "i"
[T2]	Consonants	10	Czech	"m", "b", "t", "d", "r", "l", "k", "g", "h", "ch"
			English	"m", "b", "t", "d", "r", "l", "k", "g", "h", "ch"
[T3]	Syllables	9	Czech	pe, "la", "vla", "pro", "bě", "nos", "ber", "krk", "prst"
			English	pe, "la", "vla", "for", "bě", "nose", "take", "neck", "finger"
[T4]	2-Syllable Words	5	Czech	kolo, "pivo", "sokol", "papír", "trdlo"
			English	wheel, "beer", "falcon", "paper", "boob"
[T5]	3-Syllable Words	4	Czech	dědeček, "pohádka", "pokémon", "květina"
			English	grandfather, "fairy tale", "Pokemon", "flower"
[T6]	4-Syllable Words	3	Czech	"motovidlo", "televize", "popelnice"
			English	"niddy noddy", "television", "dustbin"
[T7]	5-Syllable Words	2	Czech	"různobarevný", "mateřídouška"
			English	"varicoloured", "thyme"
[T8]	Doubled Words	3	Czech	pohádková víla, "kouzelný měšec", "čarotvorný hrnec"
			English	"fairy", "magic pouch", "magic pot"
[T9]	Augmentation of Word Order	4	Czech	"voda", "živá voda", "živá a mrtvá voda", "pramen s živou a mrtvou vodou"
			English	"water", "live water", "live and dead water", "source of live and dead water"
[T10]	Compound Sentence	1	Czech	"Když šla červená Karkulka k babičce, potkala zlého vlka."
			English	"When Little Red Riding Hood went to her grandmother, she met bad wolf."
[T11]	Acoustic Differentiation	10	Czech	"pes - nes", "ten - den", "kůl - vůl", "hrát - brát", "ječí - ježí", "ble - ple", "kloč - kloč", "kvěš - kveš", "šný - šní", "vošl - vočl"
			English	Change in one phoneme in the word. For example: "pes - nes", …
[T12]	Sentence during EEG	1	Czech	"Podívej se na směšného klauna."
			English	"Look at the laughable clown."
[T13]	Describe the picture	1		A spontaneous description of the girl's picture.

Table 3.4 List of the vocal tasks

4 Classification Based on The Error Analysis (transcriptional analysis)

In this chapter, I present a new method, called error analysis, for identifying children with SLI based on the number of pronunciation errors in their utterances. Pronunciation requires the ability to distinguish the sounds of the spoken language via hearing. Children with SLI have a distinctly impaired ability to aurally differentiate phonemes, and they cannot distinguish acoustically similar words. The problems occur in the perception and processing of verbal stimuli, storage in memory and recall, including memory learning. These problems are related to acoustic-verbal processes.

Many research groups have developed various tests for SLI based on grammatical and reading skills. One is the Grammar and Phonology Screening (GAPS) test [14]. The goal of this test is to determine whether a child has sufficient knowledge concerning the use of grammatical rules to create sentences and whether they know the rules about how to add sounds together to correctly make words.

In my case, I attempted to analyze the words pronounced by children with SLI and compare them with those pronounced by healthy children (control group). I focused only on the description of errors in individual words. During my studies of children with SLI, I found that their utterances include many more errors than those of healthy children. These errors occur across all age categories (our research includes children aged 39 to 131 months).

4.1 Description of Method

The basic criterion for developing this error analysis includes a process of creating three matrices (namely the reference (*RM*), testing (*TM*) and confusion (*CM*) matrices) and two parameters for speech (text reading by speech therapist (*ut1*) and repeating text by children (*ut2*) from Table 3.4).

$$RM = \begin{pmatrix} rm_{11} & rm_{12} & rm_{13} & \dots & rm_{1n} \\ rm_{21} & rm_{22} & rm_{23} & \dots & rm_{2n} \\ rm_{31} & rm_{32} & rm_{33} & \dots & rm_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ rm_{m1} & rm_{m2} & rm_{m3} & \dots & rm_{mn} \end{pmatrix} \tag{0.22}$$

$$\sum_{i=1}^{k} rm_{kk} = k \tag{0.23}$$

$$rm_{11} = rm_{22} = rm_{33} = \ldots = rm_{kk} = 1 \tag{0.24}$$

We defined the *RM* is a square reference matrix of order *k* (where *k* is the number of phonemes in ut_1). When the number of phonemes in ut_1 is greater than that in the ut_2, then *k* is equal to the number of phonemes in the ut_1. The opposite position in the matrix represents the number of phonemes in the u_2. All of the matrix elements on the main diagonal are equal to 1, and their sum is equal to *k* (number of phonemes in the ut_1 or ut_2).

$$TM = \begin{pmatrix} tm_{11} & tm_{12} & tm_{13} & \ldots & tm_{1n} \\ tm_{21} & tm_{22} & tm_{23} & \ldots & tm_{2n} \\ tm_{31} & tm_{32} & tm_{33} & \ldots & tm_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ tm_{m1} & tm_{m2} & tm_{m3} & \ldots & tm_{mn} \end{pmatrix} \tag{0.25}$$

TM is defined as a rectangular array with *m* rows and *n* columns, where *m* indicates the speech units from the *ut1* and *n* indicates the speech units from ut_2. When the phonemes from ut_1 and ut_2 are compared, the number of errors for one ut_1 and ut_2 can be determined. The number of errors is calculated as a penalty score [Eq. 5 - Eq. 6],

$$PS = wp + up + mp \tag{0.26}$$

$$wp = u - gsu \tag{0.27}$$

$$up = w - gsw \tag{0.28}$$

$$mp = abs\left(w - u\right) \tag{0.29}$$

where PS is the penalty score, wp is the number of wrong phonemes, up is the number of unspoken phonemes, mp is the number of missing phonemes, w is the number of phonemes in ut1 and u is the number of phonemes in the ut2. The detailed description of the error analysis is illustrated in Figure 4.1, which shows the creation of *TM* and *RM* matrices and the algorithm used in the error analysis. The input data for error analysis are the recorded ut_1 and ut_2, and the output from error analysis is a *PS* of the analyzed ut_2.

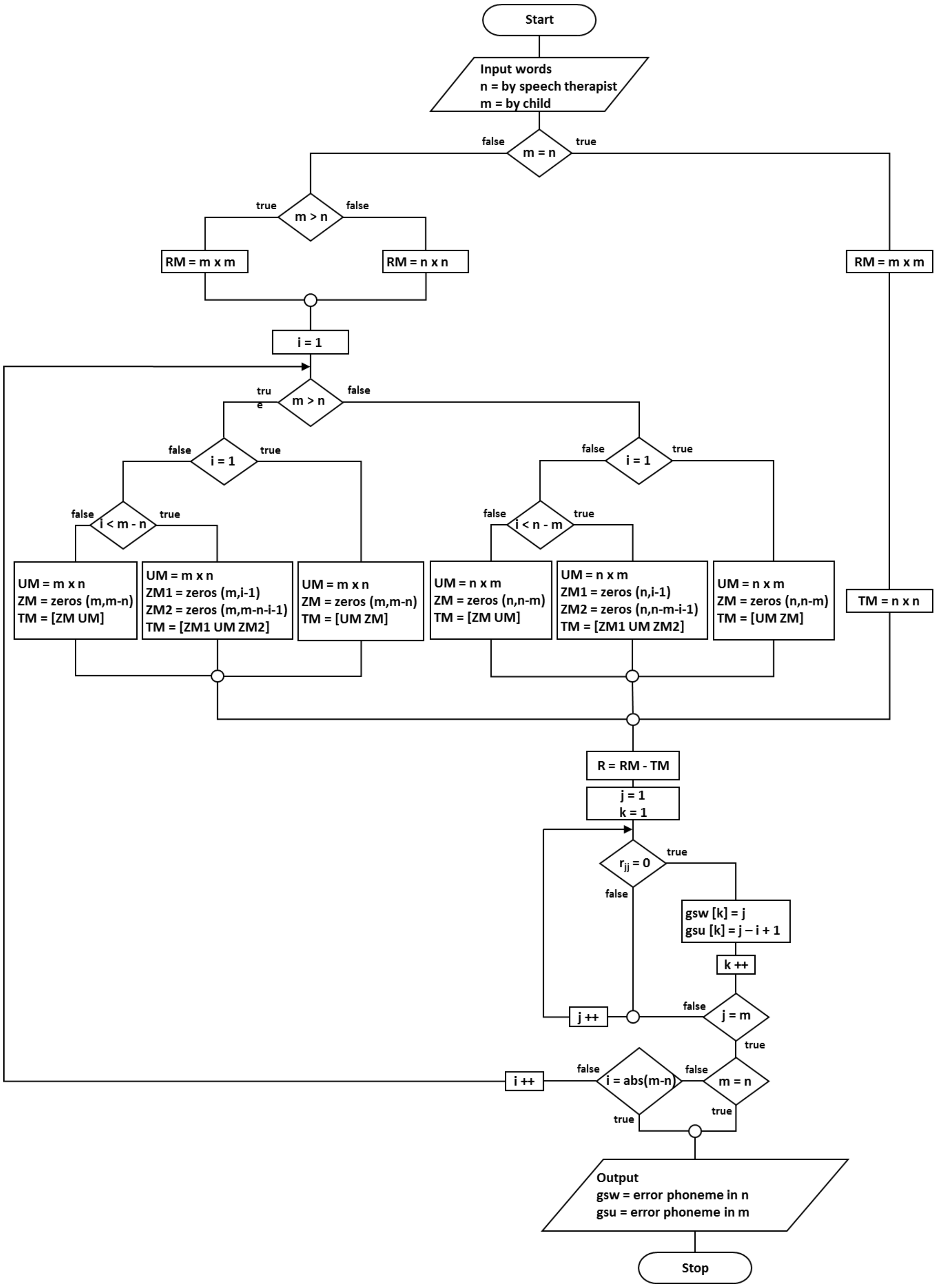


Figure 4.1 Flowchart of the error analysis. This figure shows the function of the error analysis. The input is ut_1 and ut_2, and the output is the resulting penalty score for the analyzed ut_2.

The comparison of error analyses between healthy children and children with SLI is depicted in Figure 4.2. First, the selected the monosyllabic or polysyllabic words of a child were recorded, and a precise description of the recorded words by the Labelling program [3] was performed. Next, ut_1 and ut_2 were compared, and a penalty score was calculated for each words. Finally, the sum value of the penalty scores for all the monosyllabic and polysyllabic words was taken, and the resulting value was set as the overall penalty score for a particular child. An example of the differences between a healthy child and child with SLI is shown in Figure 4.2.

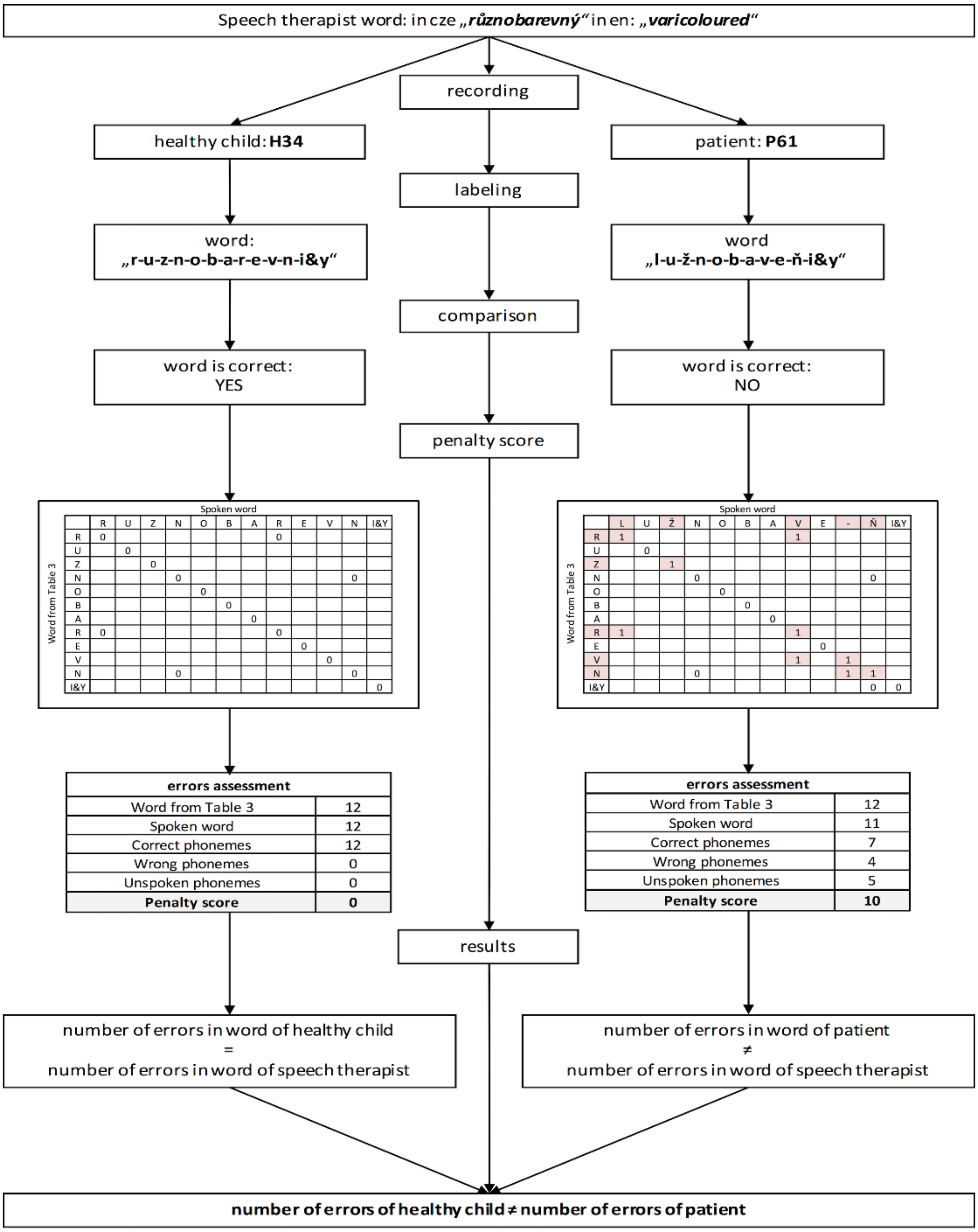


Figure 4.2 The model of classification using the error analysis. Comparison of a healthy child and a child with SLI with a final penalty score for the selected word. Red fields in the confusion matrix show errors in the utterance.

The method described in this chapter is being developed to determine whether a speaker has specific language impairment and, if so, to determine the degree of SLI and compare it with the findings of a speech therapist. For this method, a speech sample is collected and then described with the labeling program. After labeling [3], a penalty score is calculated for all words of all participants, and classification takes place.

For one part of the experiment, all of the words that met the experiment's requirements were split into two sets: one set included only the utterances of the healthy children, and the other set included only the utterances of children with SLI. For the other part of the experiment, which aimed to determine the degree of SLI, only the utterances of the children with SLI were used.

It is necessary to perform a statistical analysis to verify the validity and a certain level of significance of the results. For this purpose, I have chosen the method of statistical inference designed specifically for testing hypotheses. The role of statistical inference is to determine, based on the information obtained from random samples, whether we accept or reject a hypothesis about the data set. A result of statistical inference is considered statistically significant (i.e., it cannot be explained only by random selection) at the chosen significance level of p.

4.2 Classification of the Children with SLI

The data were divided into two groups – healthy children (p_h) and children with SLI (p_sli). I conducted individual checks of the results and found an incorrect number of errors in one healthy child. Therefore, the error value for this healthy child was not included in the statistics.

The Shapiro-Wilks test for normality was used to determine that the data were statistically normal. The obtained scores are shown in Table 4.1. The values of Wp_h and Wp_sli were too small ($p\text{-val} < 0.05$) to confirm the hypothesis that the groups had a normal distribution.

The Wilcoxon rank-sum test is a non-parametric test used as a substitute for the t-test. The obtained scores are shown in Table 4.2. The p-value was lower than the significance level of 0.05; thus, I rejected a null hypothesis of equal medians, and I concluded that there is sufficient evidence in the data to suggest that the p_h and p_sli groups are not same at the default 5% significance level, which is sufficient for significant contention. These results can be considered correct, and it can be argued that there is a significant difference in the number of errors in the speech of healthy children and children with SLI.

Healthy Children vs. Children with SLI: Shapiro-Wilk Normality Test				
Group	Classification	Results	Hypothesis	
			H0	H1
p_h	-	W = 0.9175, p-val = 0.00444	0	1
p_sli	-	W = 0.83, p-val = 2.28 e-06	0	1

Table 4.1 Error Analysis - Shapiro-Wilk Normality Test.

Healthy Children vs. Children with SLI: Wilcoxon Rank-Sum Test (w)				
Group	Classification	Results	Hypothesis	
			H_0	H_1
-	p_h vs. p_sli	(w): p-val = 1.01e-15, zval = -8.3166, ranksum = 963	0	1

Table 4.2 Error Analysis - Wilcoxon Runk-Sum Test.

4.2.1 Classification of all Children

The results of the analyses of utterance errors are displayed in Figure 4.3, which presents all of the participants. The healthy children are displayed in blue, and the children with SLI are displayed in red. Pronunciation errors are displayed in the upper graph. A higher value indicates a higher number of errors. The utterances of the children with SLI have a higher total number of errors compared with the utterances of the healthy children. The distributions of errors for the healthy children and the children with SLI are displayed in the middle graph. The distribution of the healthy children's errors was clustered around lower values compared with the error distribution of the children with SLI. Box plots representing the distributions of errors for the healthy children and the children with SLI show clear differences between these groups. The children with SLI had a higher number of errors in their utterances than did healthy children of the same age. Table 4.3 shows the difference in the average number of errors between these two groups; p_sli and p_h.

Healthy Children vs. Children with SLI - Participants					
Age Category	Average Error		Difference p_sli vs. p_h	Comparison	Difference [%]
	p_sli (2)	p_h (1)			
All	38,89	4,93	33,96	2 vs. 1	688,84

Table 4.3 Error Analysis - Comparison both groups - p_sli and p_h.

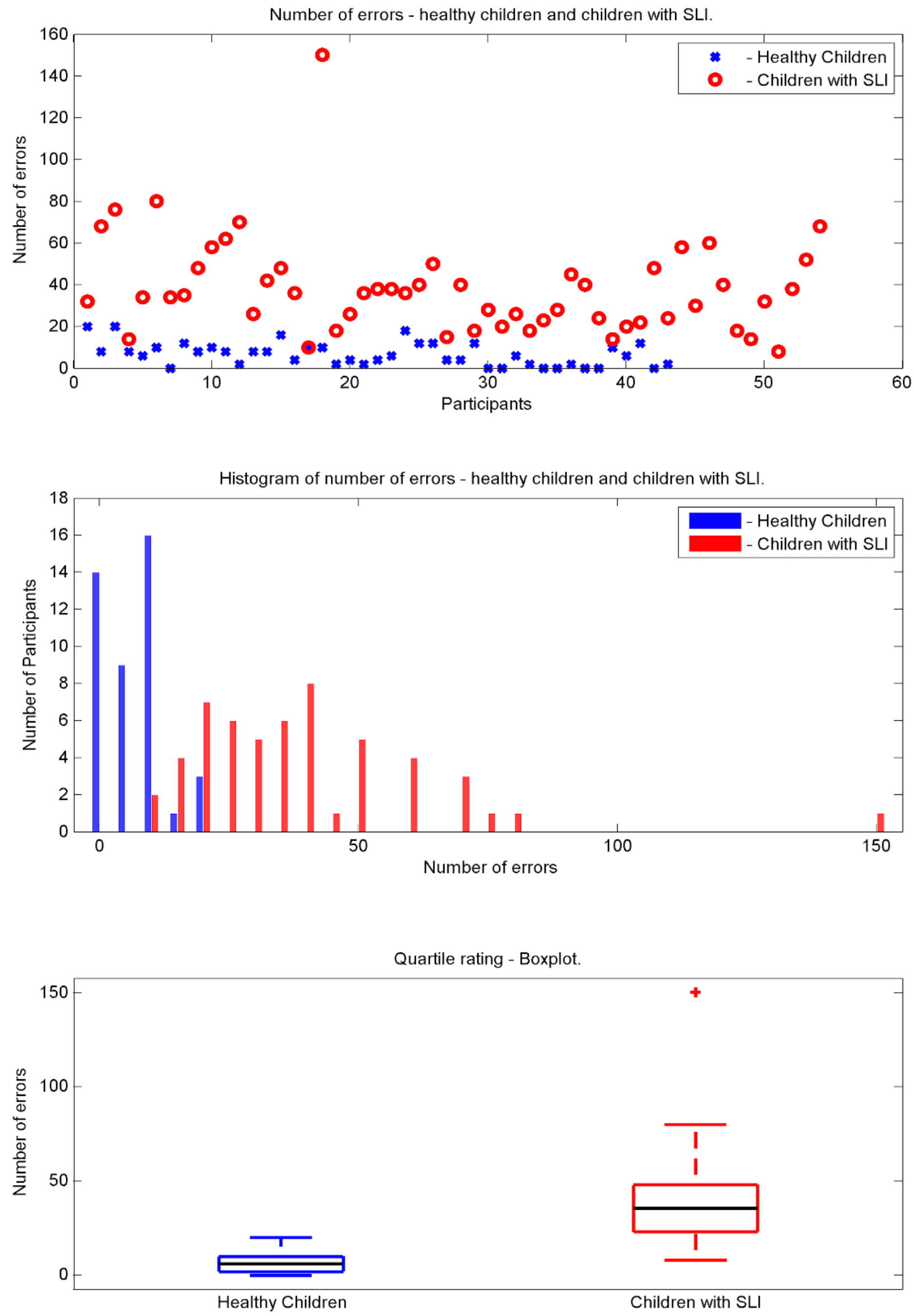


Figure 4.3 Evaluation of the error analysis. Data from healthy children are shown in blue, and data from children with SLI are shown in red. Samples with a higher number of errors are in a higher position in the first graph. A histogram of the errors of each of the participants is shown in the second graph. Box plots represent the distributions of the number of utterance errors of healthy children and children with SLI.

Table 4.4 presents the methods used to distinguish these two groups and their percent success rate. I achieved a 93.81% success rate using the best method. Because it was necessary to determine the limits value for every category, I used two different approaches to classify the participants into these groups. In the first method, HM ("hand-made"), I determined the threshold values for each group using the minimum and maximum values. The values located outside these limits were identified as misclassifications (misclass in Figure 4). These values were then evaluated according to the penalty score (PS) criterion. The PS is the sum of all errors for each word. The criterion for classification into each group was as follows: for the SLI group, (seven or more words contained an error), and for the healthy group, The other three methods are based on Artificial Neural Networks (ANN) [71], specifically Self-Organizing Maps (SOMs) [2], which are a standard part of the Neural Network Toolbox in MATLAB. The default ratios for training, testing and validation are 0.7, 0.15 and 0.15, respectively. The data were chosen randomly. In the first method, the weights were set to the original default value. In the second method, the weights were set to the threshold values (both minimum and maximum values) for these groups, and in the third method, the weights were set to the mean values of these groups. Figure 4.4 provides an overview of this process.

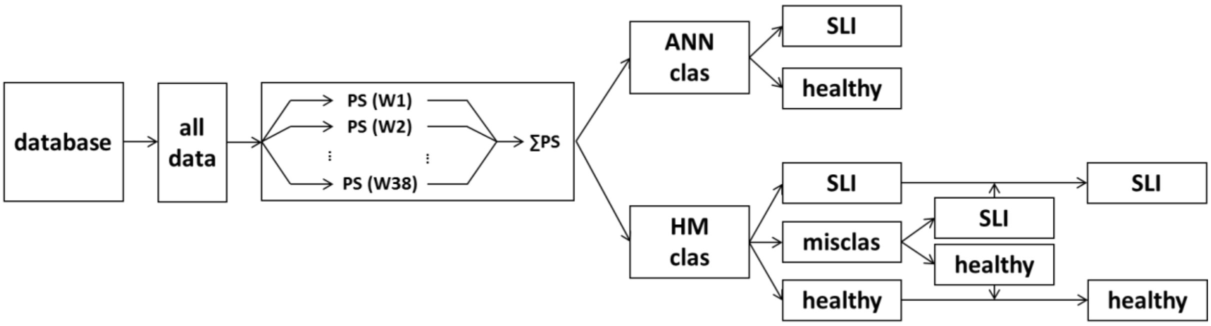


Figure 4.4 Error Analysis - Overview of the classification through ANN and HM classifier.

Methods	HM	SOM orig	SOM min-max	SOM mean
Success of classification *p_h* as *p_h*	**95.35 %**	81.40 %	97.67 %	97.67 %
Success of classification *p_sli* as *p_sli*	**92.60 %**	88.89 %	83.33 %	87.04 %
Final percent success rate	**93.81 %**	85.14 %	90.50 %	92.36 %

Table 4.4 Error Analysis - The success of classification. The percentage rates of correct classification of the methods used to distinguish the two groups.

Both classifiers, ANN and HM, offer approximately the same success rate for classification. For further development and utilization of this method (especially for parents), the HM classifier is easier to use.

This result proved the premise that children with SLI have a higher number of speech errors compared with healthy children of the same age.

4.2.2 Classification of the Age Categories

In this experiment, the differences were compared for each age category separately. The results of the error analyses for each category are displayed in Fig. 4. Age categories in months were determined as follows: A1 = 76.5 ± 6.5, A2 = 89.5 ± 5.5, A3 = 101.5 ± 5.5, A4 = 113.5 ± 5.5 and A5 = 125.5 ± 5.5. Healthy children are displayed in blue, and children with SLI are displayed in red. Individual pronunciation errors are displayed on the left. Box plots representing the error distributions of the healthy children and those of the children with SLI show clear differences between these groups for all age categories. Only the A1 category showed misclassifications (3 participants, of which 2 children were participants from p_h and 1 was from p_sli). In absolute numbers, 3 out of 22 children were misclassified. Table 4.5 shows the differences in average errors for each age category for these two groups, and Table 4.6 shows the misclassification of the individual participants.

Healthy Children vs. Children with SLI – Participants					
Age Category	**Average Error**		**Difference p_sli vs. p_h**	**Comparison**	**Difference [%]**
	p_sli (2)	**p_h (1)**			
A1	46,06	10	36,06	2 vs. 1	360,6
A2	44,27	6	38,27	2 vs. 1	637,83
A3	27,43	6	21,43	2 vs. 1	357,17
A4	35,14	3,53	31,61	2 vs. 1	895,47
A5	40	3	37	2 vs. 1	1233,33

Table 4.5 Error Analysis - Comparison both groups for all age categories.

Misclassification of the Participants.						
Age Category	**p_h**	**p_h in p_sli**	**% misclassification**	**p_sli**	**p_sli in p_h**	**% misclassification**
A1	5	2	**40**	17	1	**5,9**
A2	6	0	0	11	0	0
A3	7	0	0	14	0	0
A4	17	0	0	7	0	0
A5	8	0	0	5	0	0

Table 4.6 Error Analysis - Misclassification for all age categories.

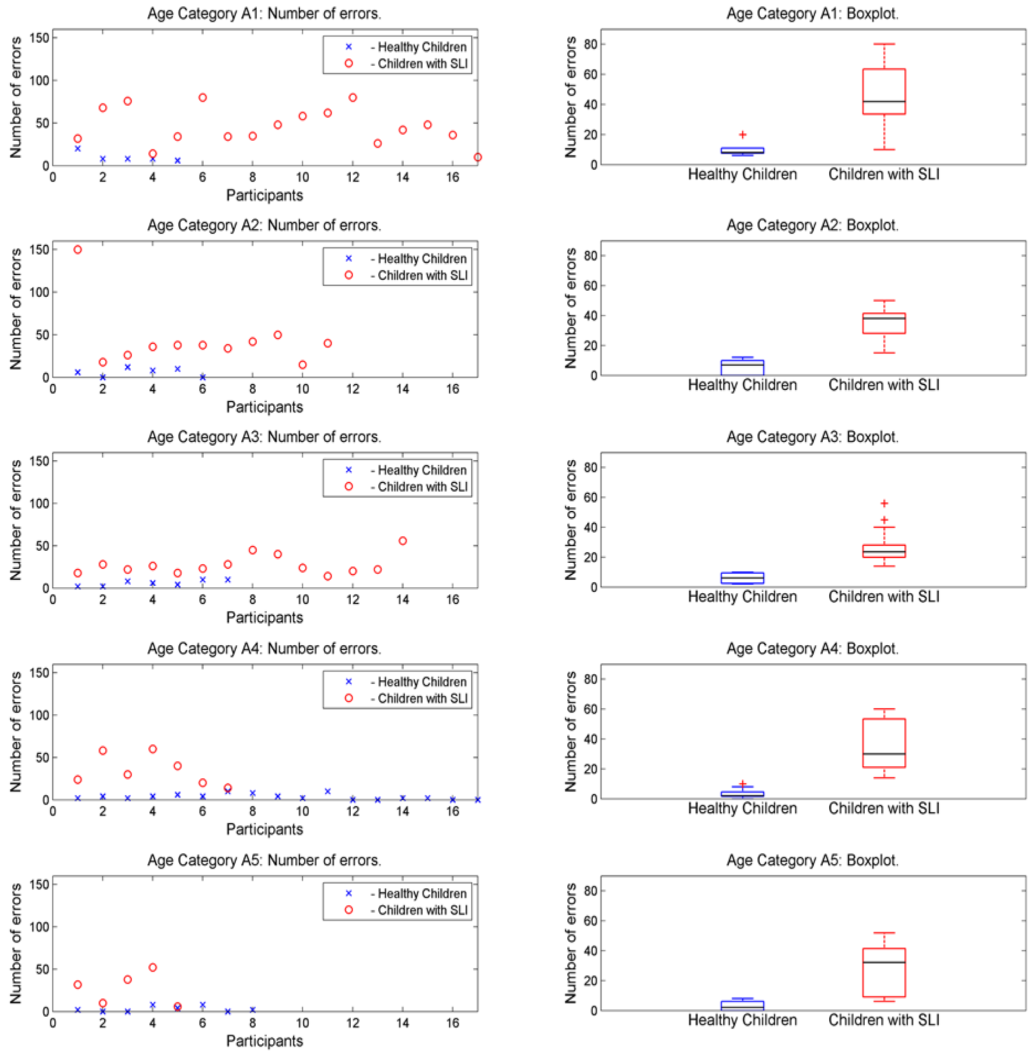


Figure 4.5 Evaluation of the error analysis. Data from healthy children are shown in blue, and data from children with SLI are shown in red. Samples with a higher number of errors are in a higher position in the left graphs. Box plots represent the distributions of the number of utterance errors of healthy children and children with SLI.

These results indicated apparent differences in error rates between healthy children and children with SLI in every age category. The average error rates for healthy children were characterized by a decrease (i.e., the average number of errors decreased with increasing age), while no such pattern emerged for the children with SLIt (see Table 4.5, Average Errors column).

4.2.3 Classification of the suitable words for classification

Another experiment focused on the differences between individual words. In my research, I examine various types of words, from very simple words (i.e., those with isolated vowels) to complex words (i.e., multisyllabic words that contain all of the vowels). All of the words used in the experiments are described in chapter 3.1. The aim of the experiment is to determine whether all words are suitable for identifying children with SLI and to determine which words have the greatest ability to distinguish children with SLI from healthy children. In this experiment, I compared the error rate for each word.

The following procedure was used: Each word was evaluated using a coefficient that was calculated as the difference between the average error rates of p_h and p_sli. If the difference was greater than one (see formula X1), the word was identified as suitable for an unambiguous classification of children with SLI.

$$W_n E_{sli} - W_n E_h > CL \tag{0.30}$$

where $W_n E_{sli}$ is the average error rate for a word in group *p_sli*, $W_n E_h$ is average error rate for a word in group *p_h* and CL is the critical line for evaluation. In my case, the line is based on the value 1.

The results for the suitable words experiment are displayed in Figure 4.6, which presents all of the words. The average error rates for the healthy children is in blue, and that for the children with SLI is in red. The horizontal line is the critical line with a value of 1. A black curve represents the differences between the average error rate of the healthy children and that of the children with SLI. The X-axis represents all of the words, and the y-axis represents the average errors. There are five graphs for individual categories and one graph for all participants independent of age. Table 4.7 shows a comparison of the average error rates and their differences between the healthy children and the children with SLI for all words as a group and for selected words. The penultimate and final columns were very important for determining which words best classified children with SLI. The penultimate column shows the differences between the average error rate when selected words are used and when all words are used. As we can see, using only selected, suitable words yields greater differences between healthy children and children with SLI. The last column shows a number of words that satisfy the critical line condition. These words are suitable for decisively identifying children with SLI.

All Words vs. Selected Words						
Age Category	**All Words – Average Error Rate**			**Selected Words – Average Error Rate**		
	p_sli (2)	**p_h (1)**	**p_sli vs. p_h**	**p_sli (2)**	**p_h (1)**	**p_sli vs. p_h**
All	1,02	0,13	0,89	2,11	0,27	1,84
A1	1,21	0,26	0,95	2,19	0,33	1,87
A2	1,17	0,16	1,01	4,38	0,46	3,92
A3	0,72	0,16	0,56	1,26	0,2	1,06

Table 4.7 Error Analysis - Comparison of results for all words and selected words depending on the critical line.

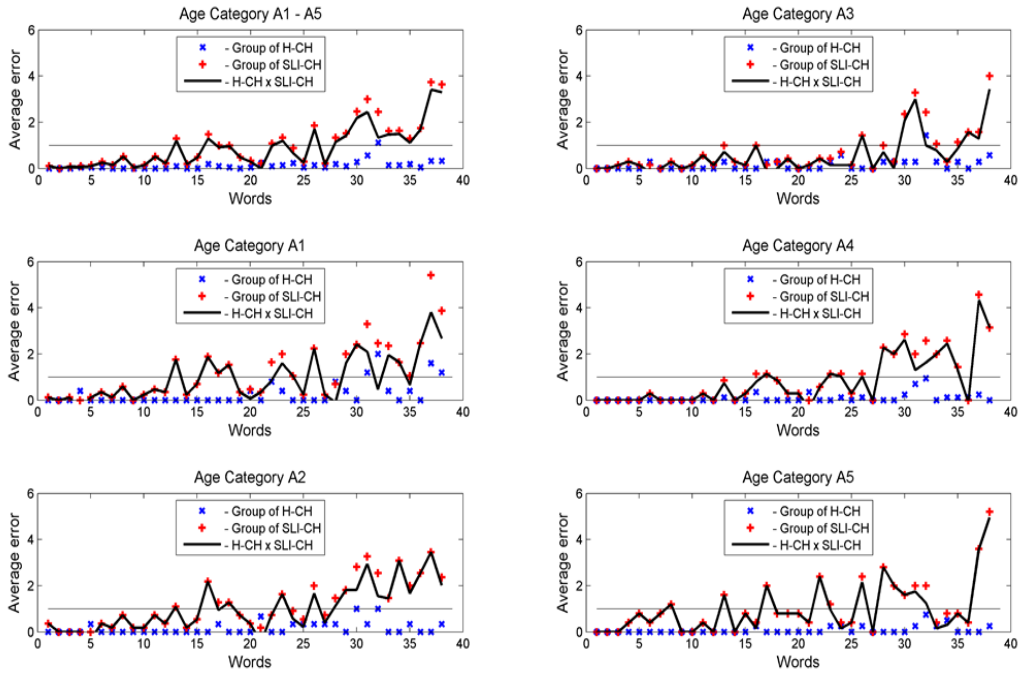
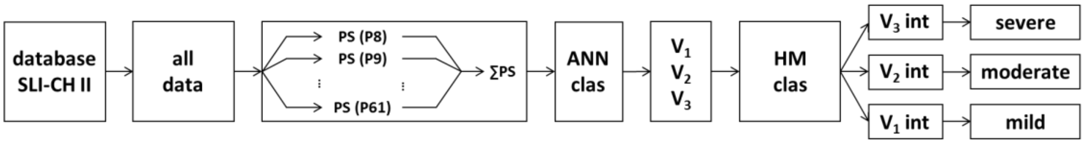


Figure 4.6 Evaluation of the classification of the suitable words. Data from healthy children are shown in blue, and data from children with SLI are shown in red. Horizontal line is the critical line. A black curve represents a differences between average error rate from healthy children and children with SLI. The x-axis constitutes all words, the y-axis constitutes average errors.

4.3 Differentiate children with SLI under three categories

This experiment focused on identifying and classifying three degrees of specific language impairment: specifically mild, moderate and severe. To determine severity, we used only the error rate (i.e., a greater number of errors indicated a higher degree of SLI severity). The results were compared with a speech therapist's assessment. The aim of the experiment was to determine whether there is a relationship between the error rate and the degree of SLI.

The procedure was as follows: Only the data from the group of children with SLI (SLI-CH II) were classified. An error rate was assigned to each word. Then, the resulting penalty score (the sum of the error rates for all words) was calculated for each participant. To determine the particular degree of SLI, we used ANN (SOMs with default ratios for training, testing and validation of 0.7, 0.15 and 0.15, respectively) and HM classifiers. SOM was used to determine three values (V_1, V_2 and V_3) around which most of the other values occurred (penalty score). From these values, we determined intervals for classifying errors into a particular degree of SLI. The final classification was performed using HM ("hand-made") classifiers according to the classification intervals (V1 int, V2 int and V3 int). An overview of the process of classifying the children according to the three degrees of SLI, including intervals, is provided in Figure 4.7.



$$V_3 \,\text{int}: V_3 - \frac{V_3 - V_2}{2} < PS$$

$$V_2 \,\text{int}: V_2 - \frac{V_2 - V_1}{2} \le PS \le V_3 - \frac{V_3 - V_2}{2}$$

$$V_1 \,\text{int}: PS < V_2 - \frac{V_2 - V_1}{2}$$

Figure 4.7 Overview of the classification into three types of degree of SLI.

This method is based on the distribution of the participants according to error rate. The results obtained with this method were compared in the following manner: A penalty score was assigned to each child. The degree of severity based on error analysis was determined using the penalty score, while the speech therapist's determination of severity was based on his or her own assessment. Individual classifications of degrees of SLI were compared according to the following criteria:

$$mean\,val_{\text{mild}} < mean\,val_{\text{moderate}} < mean\,val_{\text{severe}} \qquad (0.31)$$

In other words, the mean value for the "mild" classification must be less than the mean value for the "moderate" classification, and so on. If this relationship was valid for both the speech therapist's classification and the classification based on error analysis, then the classification based on error analysis corresponds approximately to the classification of a speech therapist.

The data were divided into three groups (mild, moderate and severe) depending on the error rate. This was necessary to verify that the data sets could be compared.

The Shapiro-Wilks test for normality was used to determine that the data were statistically normal. The obtained scores are summarized in Table 4.8 and Table 4.9. Table 4.8 summarizes the classifications obtained from the error analysis, and Table 4.9 summarizes the classifications of the speech therapist using the penalty score. Only the W values for the mild and moderate categories (p-val > 0.05) based on error analysis (see Table 4.8) met the requirement of normalityA t-test was used to compare these two classifications (see Table 4.10 for the classification of mild vs. moderate). Because the other values did not meet the requirement for normality, they were compared using the Wilcoxon rank-sum test (see Table 4.10 and Table 4.11).

The p-values for all comparisons (mild vs. moderate, moderate vs. severe and mild vs. severe) were lower than the significance levels (t-Test: alpha = 0.01 and Wilcoxon: alpha = 0.05). Thus, I rejected a null hypothesis of equal medians for the Wilcoxon rank-sum test and equal means for Student's t-test, and I concluded that there is sufficient evidence in the data to suggest that the mild, moderate and severe groups are not same at the default 5% (Wilcoxon rank-sum test) or 1% level of significance (Student's t-test; see Table 4.10 and Table 4.11). These results can be considered correct, and it can be argued that the number of word errors differs significantly among the mild, moderate and severe groups.

Error Analysis Classification: Shapiro-Wilk Normality Test				
Group	Classification	Results	Hypothesis	
			H_0	H_1
p_sli	mild	W = 0.9281, p-val = 0.18	1	0
	moderate	W = 0.9327, p-val = 0.08041	1	0
	severe	W = 0.6634, p-val = 0.0008636	0	1

Table 4.8 Error Analysis - Shapiro-Wilk Normality Test for the classification by the error analysis.

Speech Therapist Classification: Shapiro-Wilk Normality Test				
Group	Classification	Results	Hypothesis	
			H_0	H_1
p_sli	mild	W = 0.786, p-val = 0.004674	0	1
	moderate	W = 0.921, p-val = 0.02852	0	1
	severe	W = 0.7475, p-val = 0.001928	0	1

Table 4.9 Error Analysis - Shapiro-Wilk Normality Test for the classification by the speech therapist.

Error Analysis Classification: Student's t-Test (t) and Wilcoxon Rank Sum Test (w)				
Group	Classification	Results	Hypothesis	
			H_0	H_1
p_sli	mild vs. moderate	(t): t = -10.0699, df = 42.566, p-val = 7.831e-13	0	1
	moderate vs. severe	(w): p-val = 2.3481e-05, zval = -4.2289, ranksum = 378	0	1
	mild vs. severe	(w): p-val = 6.8209e-05, zval = -3.9824, ranksum = 171	0	1

Table 4.10 Error Analysis - Conformity testing for classification based on the error analysis through the Student's t-Test and Wilcoxon Rank-Sum Test.

Speech Therapist Classification: Wilcoxon Rank Sum Test (w)				
Group	Classification	Results	Hypothesis	
			H_0	H_1
p_sli	mild vs. moderate	(w): p-val = 0.0205, zval = -2.3166, ranksum = 198	0	1
	moderate vs. severe	(w): p-val = 0.0039, zval = -2.8855, ranksum = 531.5	0	1
	mild vs. severe	(w): p-val = 5.5082e-04, zval = -3.4547, ranksum = 102.5	0	1

Table 4.11 Error Analysis - Conformity testing for classification of speech therapist based on the error analysis through the Wilcoxon Rank-Sum Test.

The resulting classifications into three categories depending on the degree of SLI are displayed in Figure 4.8. There are 2 box plot graphs. The left graph is for classification based on error analysis, and the right graph is for classification based on the assessment of a speech therapist. The mild category is displayed in blue, the moderate category is displayed in red, and the severe category is displayed in black. Both graphs have approximately the same layout. Box plots representing the distributions of utterance errors for all three classes show clear differences among these groups.

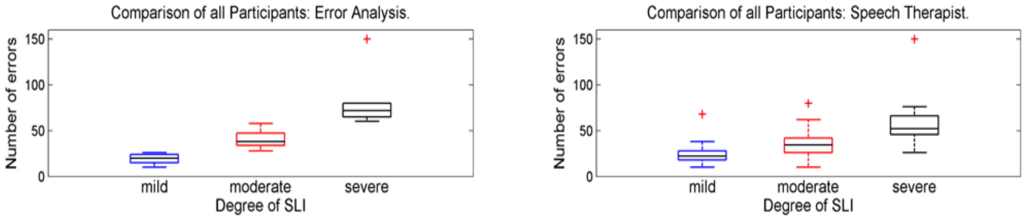


Figure 4.8 Comparison classifications - error analysis and assessment of speech therapist - depending on the error rate.

The last experiment focused on the differences among individual words. The aim of the experiment was to determine whether all words are suitable for classifying children with SLI as mild, moderate or severe and to determine which words are the most useful for classification based on error rate. This experiment compared the differences in error rates for each word depending on the speaker's degree of SLI.

The procedure was as follows: Each word was evaluated using the three coefficients representing the error rate for each class (mild, moderate and severe). The basic condition was the differences (err_{mm}, err_{sm}) between error rates for the moderate ($err_{moderate}$) and mild (err_{mild}) class and between the severe (err_{severe}) and moderate ($err_{moderate}$) classes (those differences must be greater than zero, and the sum of these differences must be greater than two) using these formulas:

$$\begin{aligned} &err_{mm} : err_{\text{moderate}} - err_{\text{mild}} > 0 \\ &err_{sm} : err_{\text{severe}} - err_{\text{moderate}} > 0 \\ &\Sigma(err_{mm}, err_{sm}) > 2 \end{aligned} \tag{0.32}$$

The results are summarized in Table 4.12 and Table 4.13. Table 4.12 contains the error analysis classification, and table Table 4.13 contains the speech therapist's evaluation based on error rate. The tables show the comparison of the average values of the error rates and the differences in error rates when all words were used and when selected words were used. The last column shows the differences between the average error rates for selected words and for all words. As we can see, using selected suitable words yields greater differences among the three classes. With this method, the results are evident even for people without related technical training.

Error Analysis Classification								
ID	Group	All Words			Selected Words			All vs. Selected words
		Average Error	Comparison	Difference [%]	Average Error	Comparison	Difference [%]	
1	healthy	4,99			1,73			
2	mild	19,21	2 vs. 1	**285,31**	7,37	2 vs. 1	**325,16**	**+ 39,84 [%]**
3	moderate	40,73	3 vs. 2	**112,02**	22,15	3 vs. 2	**200,62**	**+ 88,60 [%]**
4	severe	81,39	4 vs. 3	**99,84**	49,14	4 vs. 3	**121,85**	**+ 22,02 [%]**

Table 4.12 Error analysis classification: Comparison of mild, moderate and severe degree of SLI.

Speech Therapist Classification								
ID	Group	All Words			Selected Words			All vs. Selected words
		Average Error	Comparison	Difference [%]	Average Error	Comparison	Difference [%]	
1	healthy	4,99			1,12			
2	mild	25,96	2 vs. 1	**420,65**	6,46	2 vs. 1	**474,85**	**+ 54,20 [%]**
3	moderate	36,61	3 vs. 2	**41,04**	13,47	3 vs. 2	**108,41**	**+ 67,37 [%]**
4	severe	61,75	4 vs. 3	**68,65**	21,51	4 vs. 3	**59,72**	**- 8,93 [%]**

Table 4.13 Speech therapist assessment: Comparison of mild, moderate and severe degree of SLI.

Using these experiments, I tried to find specific words that were appropriate, i.e., words that have the best results for a given task. These include the detection of children with SLI and the determination of SLI severity. The main aim was to find words (in our database) that can be successfully used for both types of tasks. For a word to be selected as suitable for classifying children with SLI, it had to meet the criterion of suitability for all ages (in the Classification of Age Categories experiment). For a word to be selected as suitable for severity classification, it had to meet the criteria of suitability based on both error analysis and a speech therapist's assessment. Table 4.14 shows the words that were suitable for specific tasks. The last column shows the words that are optimal for both tasks. Table 4.15 shows a final evaluation of the speech tasks (T1 – T7). It is quite obvious that simple words are not suitable for classification. The usefulness for classification increases with the increasing complexity of words.

Results of Words.			
Words	**Healthy vs. SLI**	**Degree of SLI**	**Final**
paper in cze „papír“	**1**	0	0
grandfather in cze „dědeček“	**1**	**1**	**1**
flower in cze „květina“	**1**	0	0
niddy noddy in cze „motovidlo“	0	**1**	0
television in cze „televize“	0	**1**	0
thyme in cze „mateřídouška“	**1**	**1**	**1**
multicoloured in cze „různobarevný“	**1**	**1**	**1**

Table 4.14 Error Analysis - The Final evaluation of words.

Results of Group of Task Code.				
Task Code	**Healthy vs. SLI**		**Degree of SLI**	
	Number of Words	**Success rate**	**Number of Words**	**Success rate**
T1	0	0	0	0
T2	0	0	0	0
T3	0	0	0	0
T4	**1**	**20**	0	0
T5	**2**	**50**	**1**	**25**
T6	0	0	**2**	**66,7**
T7	**2**	**100**	**2**	**100**

Table 4.15 Error Analysis - The Final evaluation of words.

4.4 Discussion

This section describes the first steps toward obtaining relevant information about language problems from the utterances of children with SLI. The method is based on the number of errors in utterances, which is a transcriptional task. The initial goal for developing this method originated when we solved the problems associated with creating a children's speech database. The main problem was with the quality of the speech recordings, which varied because the recordings were conducted in the real-world environment of speech therapy clinics. All of these problems are described in [104].

The results of the error analysis show that the number of errors in a speech sample provides important but simple information that can be used as a threshold for distinguishing healthy children from children with SLI or for determining the degree of SLI.

The first experiments showed that this method can distinguish between the utterances of healthy children and those of children with SLI. This was true when all children were considered as a group and when they were divided into different age groups.

Another part of the experiment was devoted to finding suitable words to use for classification. This experiments used various types of words, from very simple words (i.e., those with a single vowel) to complex ones (i.e., multisyllabic words that contained all of the vowels). The assumption that not all words are suitable for classification purposes was confirmed.

The most recent experiments showed that the error analysis method can distinguish between different degrees of SLI (mild, moderate and severe). The resulting classifications were compared with a speech therapist’s assessment. The final part of the experiment was devoted to determining which words were suitable for distinguishing the degree of SLI.

Finally, the suitable words were compared, and three specific words were found that are suitable for both diagnosing children with SLI and distinguishing the severity of SLI. These three words are: Grandfather (in Czech: dědeček), Thyme (in Czech: mateřídouška) and Varicolored (in Czech: různobarevný).

5 Classification Based on The Feature Analysis

In this chapter, I present a method, called feature analysis, for identifying children with SLI based on the auditory signal features specific to the acoustic features of speech. Children with SLI show impaired perception and production of spoken language and can also present with motor, auditory, and phonological difficulties.

This thesis focuses on children with SLI. One inspiration was achievements in acoustic analyses related to the recognition of emotions. An analogy between the analysis of emotions and the analysis of SLI can be made: Children without pathological changes in speech and no diagnosis of any disease can be compared with a neutral emotion, while children with SLI can be compared with some unspecified emotion, e.g., fear or anger.

In my case, I attempted to compare the words spoken by children with SLI with those spoken by healthy children (control group). I focused on acoustic speech parameters, which were extracted from individual words without using the Labeling program [3]. This method of classification could provide a basis for the future development of a completely automatic detection system. The aim of this chapter and the use of this method is to identify acoustic features that can be used to uniquely identify children with SLI.

The main challenge was to maximize the variability between classes, i.e., healthy vs. SLI or mild vs. moderate vs. severe SLI, while minimizing the variability within the individual classes so that the classes are well separated. Nevertheless, the features indicating different classes may overlap, and there may be multiple ways to identify the same class. Modern techniques allow us to classify as many different acoustic features of an audio recording as possible. Optimization methods can be used to reduce the number of features and to select the features that offer the best classification. These selected data sets can then be used to classify children with SLI or to determine the degree of SLI. This processing method reduces the need to make difficult decisions about which features and methods are relevant for the type of tasks, but it introduces the need to develop optimization methods.

5.1 Description of Method

Both of the examined issues (the classification of children with SLI and the ability to distinguish between different degrees of SLI) followed the same implementation structure. The implementation work can be divided into four parts whose respective key components can be described as follows:

1. Input data: Describes the data used for experiments.
2. Feature extraction: Presents the tool for audio feature extraction and summarises its use in the project.
3. Feature selection: Presents methods and procedures used to reduce of the number of features.
4. Classification: Presents the decision-making procedures used for the final classification.

1. Input data

The data used to identify children with SLI were selected from our speech database, particularly from the H-CH and SLI-CH II subgroups. The H-CH speech subgroup includes data from healthy children and contains 44 speech recordings from children in five age categories. The SLI-CH II speech subgroup includes data from children with SLI and contains 54 speech recordings from children in five age categories. The age categories were determined using months of age. The data for determining the degree of SLI were selected from the SLI-CH II subgroup. More information is provided in the chapter about the speech databases.

2. Feature extraction

The feature extraction methods are among the main components of the detection system used to identify children with SLI based on their speech. To extract the features, we used the OpenSMILE toolkit [88] (freely available under the terms of the GNU General Public License). The OpenSMILE toolkit is a flexible feature extractor for speech processing and machine learning applications that is capable of producing a wide range of acoustic speech features. In this research, the feature set used to analyze the data from the speech database contains 1582 acoustic features obtained by applying statistical functionals to the utterance contours of 34 low-level descriptors (LLDs) and their deltas, which were estimated from the speech signal every 25 ms (default settings The names of the 34 low-level descriptors as they appear in the output file are shown in Table 5.1, and the names of the 21 functionals, as they appear in the output file, are shown in Table 5.2. The description is taken from the openSMILE toolkit tutorial [88], as shown in Table 5.1 and Table 5.2.

NAME	NUMB. OF COEFFICIENT	DESCRIPTION
PCM_LOUDNESS	1	Loudness as the normalized intensity raised to a power of 0.3.
MFCC	15	Mel-Frequency cepstral coefficients 0-14.
LOGMELFREQBAND	8	Logarithmic power of Mel-frequency bands 0 - 7 (distributed over a range from 0 to 8 kHz)
LSPFREQ	8	The 8 line spectral pair frequencies computed from 8 LPC coefficients.
F0FINENV	1	The envelope of the smoothed fundamental frequency contour.
VOICINGFINALUNCLIPPED	1	The voicing probability of the final fundamental frequency candidate. Unclipped means that it was not set to zero when it falls below the voicing threshold.

Table 5.1 The acoustic feature set used in the issue of the detection of the children with SLI. Descriptions of 34 low-level descriptors used for experiments from the openSMILE toolkit [88].

NAME	DESCRIPTION
MAXPOS	The absolute position of the maximum value (in frames).
MINPOS	The absolute position of the minimum value (in frames).
AMEAN	The arithmetic mean.
LINREGC1	The slope (m) of a linear approximation of the contour.
LINREGC2	The offset (t) of a linear approximation of the contour.
LINREGERRA	The linear error computed as the difference of the linear approximation and the actual contour.
LINREGERRQ	The quadratic error computed as the difference of the linear approximation and the actual contour.
STDDEV	The standard deviation of the values in the contour.
SKEWNESS	The skewness (3rd order moment).
KURTOSIS	The kurtosis (4th order moment).
QUARTILE1	The first quartile (25% percentile).
QUARTILE2	The second quartile (50% percentile).
QUARTILE3	The third quartile (75% percentile)
IQR1-2	The inter-quartile range: quartile2-quartile1.
IQR2-3	The inter-quartile range: quartile3-quartile2.
IQR1-3	The inter-quartile range: quartile3-quartile1.
PERCENTILE1.0	The outlier-robust minimum value of the contour, represented by the 1% percentile.
PERCENTILE99.0	The outlier-robust maximum value of the contour, represented by the 99% percentile.
PCTLRANGE0-1	The outlier robust signal range 'max-min' represented by the range of the 1% and the 99% percentile.
UPLEVELTIME75	The percentage of time the signal is above (75% * range + min).
UPLEVELTIME90	The percentage of time the signal is above (90% * range + min).

Table 5.2 The acoustic feature set used in the issue of the detection of the children with SLI. Description of 21 statistic functionals applied to the LLDs from the openSMILE toolkit [88].

3. Feature selection

A large number of features (in this case, 1582 acoustic features were extracted from the speech recordings) increases the ability to more accurately characterize the analyzed speech; however, it also increases the possibility that some of the data will be irrelevant or redundant. The classification algorithms are able to attain high classification accuracy when "the curse of dimensionality" is minimized [98], i.e., when a large number of weakly relevant and redundant features are eliminated. Another disadvantage of high-dimension feature sets is the computational complexity of the resulting algorithms. Dimensionality reduction via feature selection reduces dimensionality and removes redundant and irrelevant data by discarding features according to certain criteria. The two tasks addressed in this chapter required different approaches to feature reduction.

Regarding the classification of children with SLI, everything was known about the data, i.e., which data came from healthy children and which came from children with SLI, and the analyzed utterances are clearly identified. This fact enabled the use of a method called “maximization of gain” for feature selection. For each word, we selected only those features that satisfy the requirement for the highest classification accuracy, i.e., those that had the highest success rate for distinguishing the two groups based on their data.

The procedure was as follows: All of the data were evaluated according to the following criteria: If the samples from group 1 had a value lower than the minimum value for group 2, then the samples were assigned a value of 1; otherwise, they were assigned a value of 2. If the samples from group 2 had a value higher than the maximum value for group 1, the samples were assigned a value of 1; otherwise, they were assigned a value of 2. The value of 1 indicated correct classification, and the value of 2 indicated incorrect classification. An overview of the classification procedure is provided in Figure 5.1.

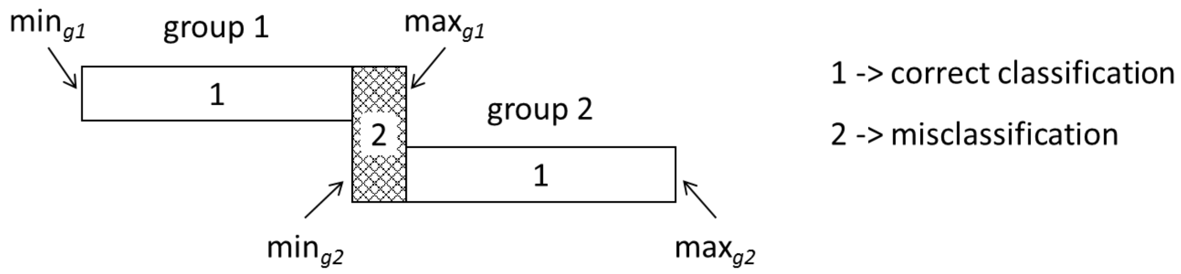


Figure 5.1 Overview of the classification individual groups.

The determination of the degree of SLI had completely different attributes. The speech recordings used for this classification were taken only from the SLI-CH II database and came from children who were diagnosed with SLI of different degrees of severity (mild, moderate and severe); however, the children’s individual severities had not been determined. It was necessary to use this feature selection method to allow the data to find the dependencies and select the appropriate features without specific knowledge of the output. Feature selection was based on Spearman's rank correlation coefficient, a nonparametric measure of the statistical dependence of two variables. This calculation does not require the data to be normally distributed.

The procedure was as follows: The data were evaluated using Spearman's rank correlation coefficient. All of the data were compared using this correlation. For each feature, the sum of the correlation coefficients was calculated, and the 30 features with the highest values were selected. The selected features were those that contained the best information about the internal dependencies of the data.

4. Classification

The assignment of the children to a specific class, i.e., healthy vs. SLI, was relatively simple. Each participant was evaluated on the basis of the selected features of several words (usually 38). Thus, each participant was evaluated by of all the selected features from all the words. The resulting classification was based on the number of classifications, i.e., 1 for a correct classification and 2 for a misclassification, and a class is assigned based on the number of classifications. The winning class was based on a larger number of classifications.

A comparative method was used for feature reduction. For each word, we selected the features with the best ability to distinguish categories, i.e., the features that best differentiated between healthy children and children with SLI. Features could be selected in two ways: For each word, we chose the feature set with either the same number of features or a variable number of features (see Figure 2, feature selection - FS: constant and FS: variable). First, it was necessary to determine the rate of successful classification. First class features distinguished healthy children from children with SLI with a success rate higher than 90%; second class features had a success rate between 80% and 90%, and so on. For each word, we selected the properties with the highest accuracy rate. The participants were classified using a total of 268 features obtained from 38 words. If a constant number of features was used, each participant was classified using a total of 760 features.

A comparison of these two selection methods is shown in Figure 5.2, which presents all of the words. The average success rate of correct classification using a constant number of features is shown in blue, and method using a variable number of features is in red. The horizontal line is the critical line, with a value of 90%. The x-axis represents all of the words, and the y-axis represents the percent success rates. Table 5.3 shows the total number of words that met the criterion for success (> 90%) and the average success rates for both feature selection methods. It is evident that variable number method had greater accuracy (i.e., an average success rate higher than 91%) and more successful classification (i.e., 25 words for classification). The entire process used to identify the children with SLI is shown in Figure 5.3.

Healthy Children vs. Children with SLI						
Age Category	**Improving of Feature Selection Method**					
	Group	**Words (> 90%)**	**Change**	**Average Success [%]**	**Change [%]**
All	Constant (1)	8		80,81	
	Variable (2)	25	**17**	91,11	**10,3**
	Group	Task Code (> 90%)	Change	Average Success [%]	Change [%]
	Constant (1)	2		83,24	
	Variable (2)	4	**2**	89,29	**6,04**

Table 5.3 Feature Analysis - Improving of Feature Selection Method.

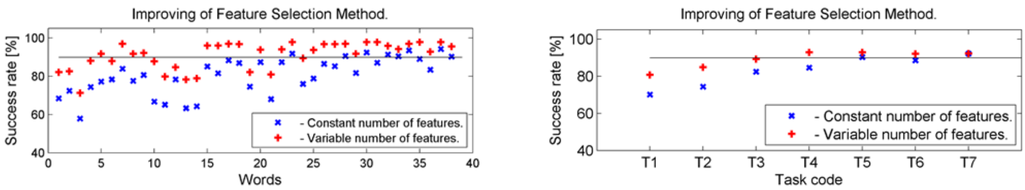


Figure 5.2 Feature Analysis - Improving of Feature Selection Method.

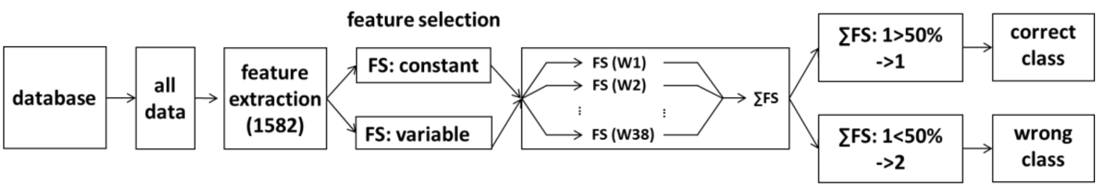
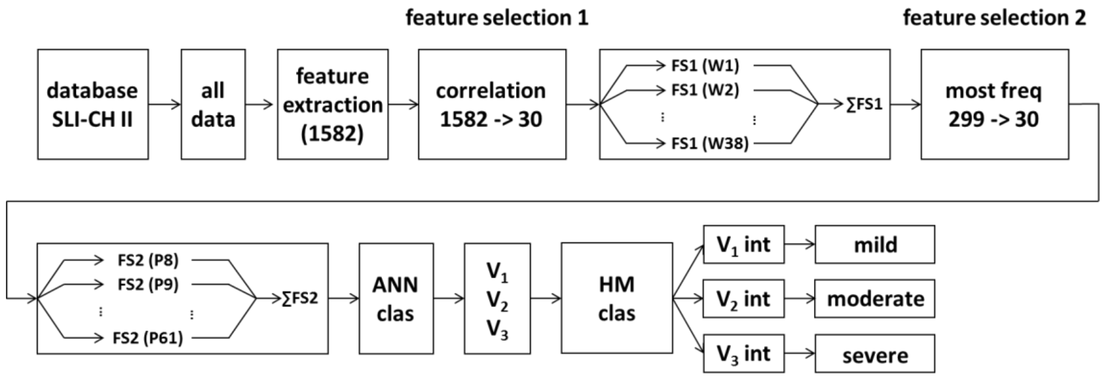


Figure 5.3 Overview of the classification children with SLI through acoustic features.

For the final classification of the degree of SLI, the same evaluation method used for error analysis, the ANN and HM classifiers, was used. SOM was used to determine three values (V1, V2 and V3) around which most of the other values occurred (i.e., the participant's classification coefficients). The intervals used to determine the specific degree of SLI were based on these values. The final classification was determined using an HM ("hand-made") classifier and the classification intervals (V1 int, V2 int and V3 int).

The first feature reduction (“feature selection 1”) selected the features that best described the intrinsic properties of the class based on the correlation (1582 features were reduced to 30 features). The second reduction of features (“feature selection 2”) selected the 30 features that were most frequently represented in all words (299 selected features were reduced to the final 30 features). These parameters were used to analyze the words and the participants. The entire process used to classify the three degrees of SLI is shown in in Figure 5.4.



$$V_3\ \mathrm{int}: V_3 - \frac{V_3 - V_2}{2} < PS$$

$$V_2\ \mathrm{int}: V_2 - \frac{V_2 - V_1}{2} \leq PS \leq V_3 - \frac{V_3 - V_2}{2}$$

$$V_1\ \mathrm{int}: PS < V_2 - \frac{V_2 - V_1}{2}$$

Figure 5.4 Overview of the classification into three types of degree of SLI through acoustic features.

The method described in this chapter is being developed to determine whether a speaker has a specific language impairment. It is based on the acoustic speech parameters extracted from individual words without using the Labeling program. This allows the creation of an automatic detection system based on this method.

The experiments had the same structure that was described for the error analysis method. For one part of the experiments, all of the words that met the experiment's requirements were split into two sets: one based on the utterances of the healthy children and the other based on the utterances of children with SLI. For other part of the experiments, we used only the utterances of the children with SLI.

It is necessary to perform a statistical analysis to verify the validity and level of significance of the results. For this purpose, I used the statistical inference method to test the hypotheses. The role of statistical inference is to determine, based on the information obtained from random samples, whether to accept or reject a hypothesis about the data set. The result of the statistical inference is considered statistically significant at the chosen significance level of p.

5.2 Classification of the Children with SLI

The data were divided into four groups depending on the classification, i.e., correct or incorrect classification for healthy children (p_h) and correct or incorrect classification for children with SLI (p_sli). The number of classifications was based on the evaluation of features.

Statistical tests evaluated the correct versus incorrect classification of selected features for both groups of children.

The scores of the Shapiro-Wilks test for normality are shown in Table 5.4. The *p* values were too small (*p-val* < 0.05) to use to confirm the hypothesis that the groups have a normal distribution.

he scores of the Wilcoxon rank-sum test, which was used as a substitute for the t-test, are shown in Table 5.5. The p values were lower than the significance level of 0.05. Thus, I rejected a null hypothesis of equal medians, and I concluded that there is sufficient evidence in the data to suggest that the groups (correct and incorrect classifications for both groups of children) are not same at the default 5% significance level, which is sufficient for significance. These results can be considered correct, and it can be argued that there are a significant differences in the number of classifications between wrong and correct evaluations for both healthy children and children with SLI.

Healthy Children vs. Children with SLI: Shapiro-Wilk Normality Test				
Group	**Classification**	**Results**	**Hypothesis**	
			H0	**H1**
p_h	correct	W = 0.5969, p-val = 8.965e-10	0	1
	wrong	W = 0.5678, p-val = 3.567e-10	0	1
p_sli	correct	W = 0.7825, p-val = 1.598e-07	0	1
	wrong	W = 0.7898, p-val = 2.344e-07	0	1

Table 5.4 Feature Analysis - Shapiro-Wilk Normality Test.

Healthy Children vs. Children with SLI: Wilcoxon Rank Sum Test				
Group	**Classification**	**Results**	**Hypothesis**	
			H0	**H1**
p_h	correct vs. wrong	p-val = 1.7510e-15, zval = 7.9578, ranksum = 2911	0	1
p_sli	correct vs. wrong	p-val = 3.3145e-19, zval = -8.9577, ranksum = 1485	0	1

Table 5.5 Feature Analysis - Wilcoxon Runk-Sum Test.

5.2.1 Classification of all Children

The results of the feature analyses for all of the participants are displayed in Figure 5.5. The “healthy” classifications are displayed on the left, and the “SLI” classifications displayed on the right. The healthy children’s results are displayed in blue, and children with SLI’s results are displayed in red. The total numbers of classifications are displayed in the upper graphs. A higher value indicates more obvious inclusion of the participant in his group (i.e., blue in the left graph and red in the left graph). The distributions of rating of classifications for healthy children and children with SLI are displayed in the middle graphs. Box plots represent the distributions of the classification of “healthy” and “SLI” for both groups. The results show that the healthy children had a higher number of classifications as “healthy” (values in blue on the left chart), and the children with SLI had a higher number of classification as “SLI” (values in red on the right chart). Table 5.6 shows the difference between the average correct classification and misclassification for both groups of participants, i.e., for healthy children (group p_h) and children with SLI (group p_sli). The last column shows the rate of correct classification, i.e., how many features on average were correct for a given classification.

Healthy Children vs. Children with SLI - Participants				
Age Category	Average Number of Classification			Average Success [%]
	Group	Correct	Wrong	
All	SLI-CH II (2)	242,77	24,55	91,3
	H-CH (1)	242,8	23,41	91,23
	∑(1 + 2)	243,58	23,31	91,27

Table 5.6 Feature Analysis - Comparison average success of classification of healthy children and children with SLI.

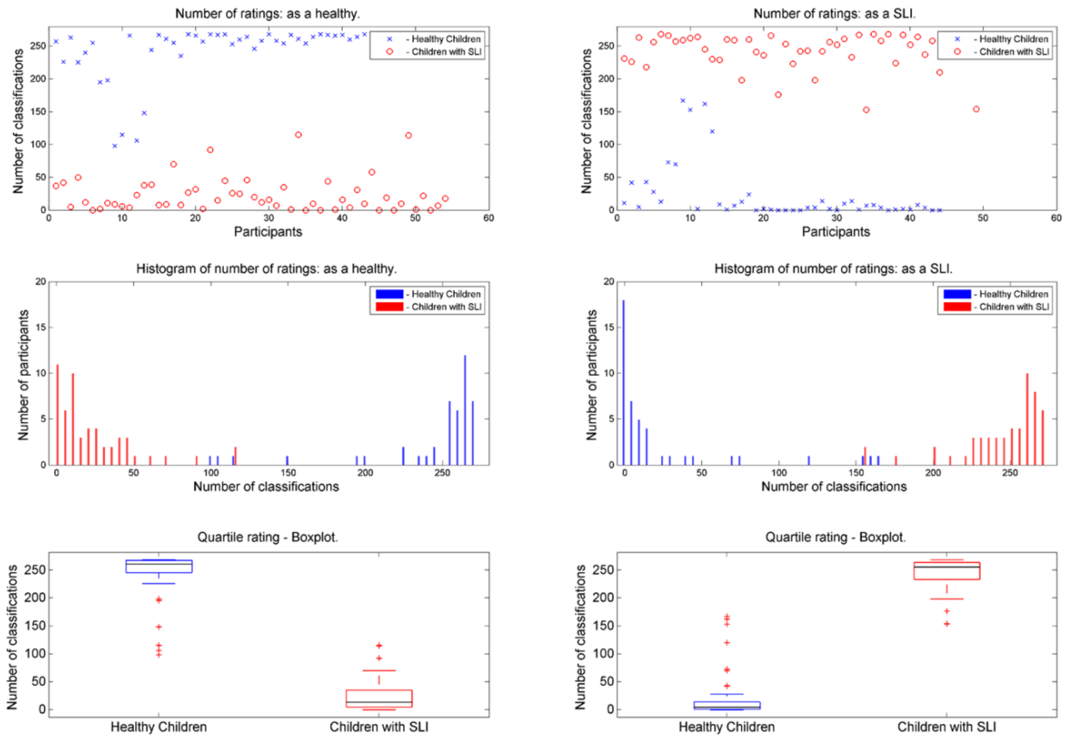


Figure 5.5 Evaluation of the feature analysis. Data from healthy children are shown in blue colour, and data from children with SLI are shown in red colour. Samples with a higher number of errors are in a higher position in the first graph. A histogram of the errors of each of the participants is shown in the second graph. Box plots represent the distributions of the number of utterance errors of healthy children and children with SLI.

Table 5.7 presents the final results used to distinguish the two groups and their percent success rate. I achieved a 96.94% success rate for both groups. Only 3 participants out of 98 were misclassified. These participants were from the p_h group, i.e., the healthy children.

Healthy Children vs. Children with SLI – Final Percent Success Rate.				
Age Category	**Classification of Participants**			Success Rate [%]
	Group	Correct	Wrong	
All	P_SLI (2)	54	0	100
	H-CH (1)	44	41	93,18
	∑(1 + 2)	**98**	**95**	**96,94**

Table 5.7 The success of classification. The percentage rates of correct classification of the method used to distinguish the two groups.

This result proved that it is possible to find acoustic features that can distinguish healthy children from children with SLI with high accuracy.

5.2.2 Classification of the Age Categories

Individual numbers of classifications are displayed on the left. Box plots represent the distributions of classifications as "healthy" (in Figure 5.6) and "SLI" (in Figure 5.7) for both groups. The results show clear differences between these groups for all age categories. Only the category A2 had a misclassification (3 participants). These participants were from the p_h group, i.e., the healthy children. In absolute numbers, 3 out of 6 children were misclassified. Table 5.8 Table 7 shows the difference between the average correct classification and misclassification for each age category for both groups, i.e., healthy children (group p_h) and children with SLI (group p_sli). The last column shows the rate of correct classification, i.e., how many features on average were classified correctly. Table 8 presents the misclassifications for the individual participants.

Healthy Children vs. Children with SLI - Participants				
Age Category	**Average Number of Classification**			Average Success [%]
	Group	Correct	Wrong	
A1	P_SLI (2)	246,53	21,47	91,99
	H-CH (1)	242,2	25,8	91,17
	∑(1 + 2)	245,96	22,04	91,77
A2	P_SLI (2)	234,55	30,73	88,37
	H-CH (1)	176,14	91,43	60,89
	∑(1 + 2)	209,29	56,76	78,67
A3	SLI-CH II (2)	247,14	20,86	92,22
	H-CH (1)	239,71	24,86	90,65
	∑(1 + 2)	244,67	22,19	91,7
A4	SLI-CH II (2)	237,86	30,14	88,75
	H-CH (1)	260,76	4,18	98,42
	∑(1 + 2)	254,08	11,75	95,6
A5	SLI-CH II (2)	258,4	9,6	96,42

	H-CH (1)	266	2	99,25
	$\sum(1+2)$	263,08	4,92	98,16

Table 5.8 Feature Analysis - Comparison both groups for all age categories.

Misclassification of the Participants.						
Age Category	H-CH	H-CH in SLI-CH II	% misclassification	SLI-CH II	SLI-CH II in H-CH	% misclassification
A1	5	0	0	17	0	0
A2	6	3	**50**	11	0	0
A3	7	0	0	14	0	0
A4	17	0	0	7	0	0
A5	8	0	0	5	0	0

Table 5.9 Feature Analysis - Misclassification for all age categories.

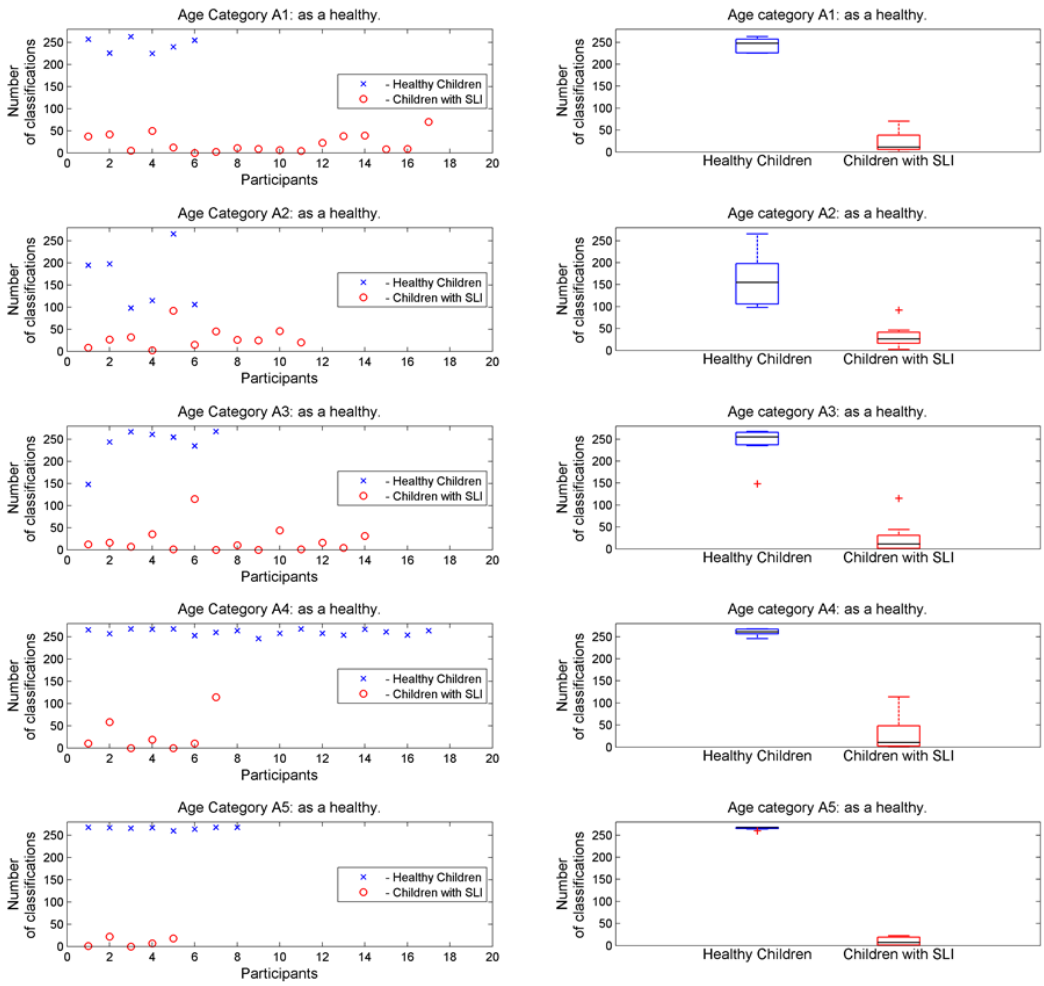


Figure 5.6 Evaluation of the feature analysis.

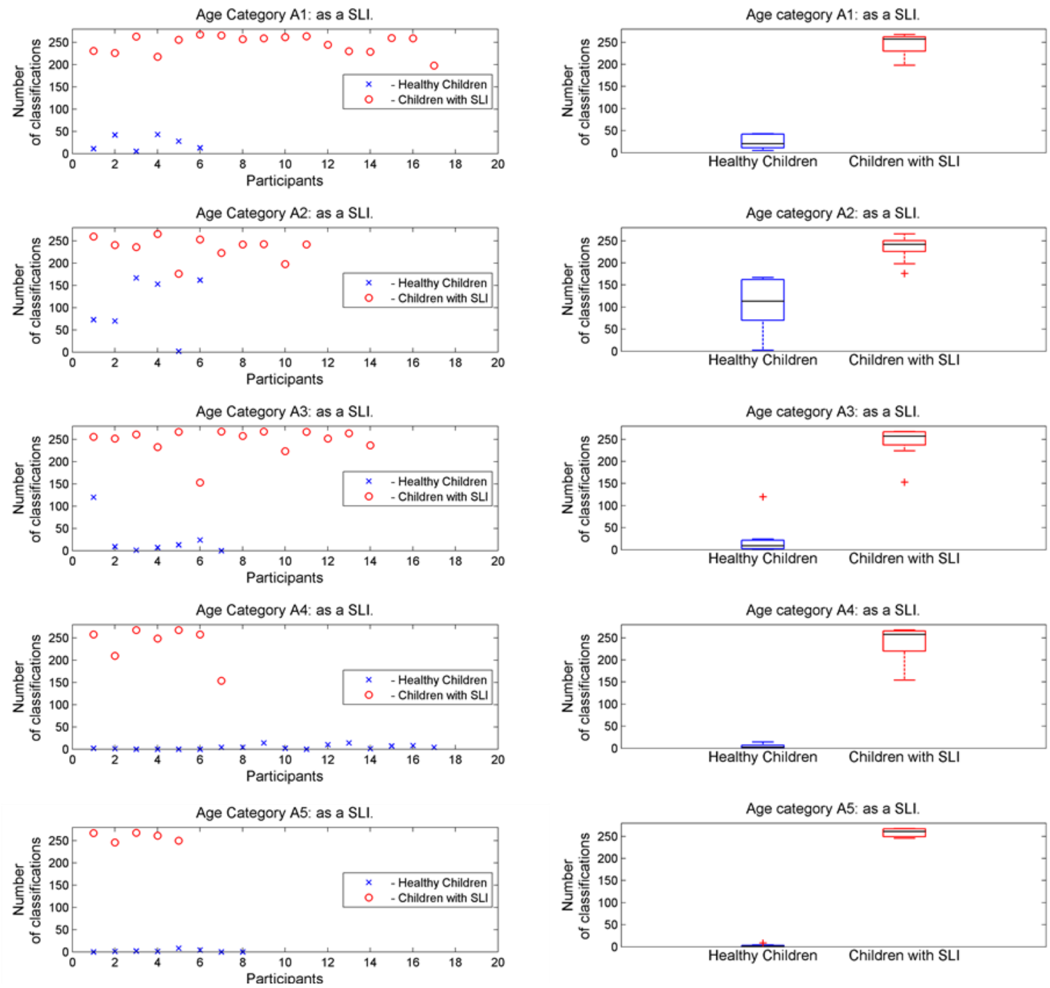


Figure 5.7 Evaluation of the feature analysis.

These results proved that a feature analysis of healthy children and children with SLI could effectively classify children in all age categories. Misclassifications of healthy children and children with SLI was detected only in the second age category, A2, and only for participants from group p_h. Such misclassification may have been the result of low-quality speech recordings.

5.2.3 Classification of an Appropriate words for classification

Another experiment focused on the differences between individual words. All of the words used in the experiments are described in the chapter 3.1. The aim of the experiment was to determine whether all words are suitable for identifying children with SLI and to identify the words with the greatest ability to distinguish these two groups. This experiment compared the efficiency of classifying each word using the average correct classification based on acoustic features.

The procedure was as follows: Each word was evaluated according to its coefficient, which was calculated as the average percent success rate of correct classification for both groups of participants, i.e. groups of *p_h* and *p_sli*. If the success rate was greater than 90% (see in formula 0.33), the word is identified as a suitable for an unambiguous classification of children with SLI.

$$W_{np} = \frac{W_{nc}}{W_{nm}} * 100 > CL \tag{0.33}$$

Where W_{np} is the average percent success rate of the selected word for all features and all participants, W_{nc} is all correct classifications of the selected word and W_{nm} a s all misclassifications for selected word. CL is the critical line for evaluation. In this case, the formula is based on the value of 90%.

The results of this suitable words analysis are displayed in Figure 5.8, which presents all of the words. The average success rate of correct classification of healthy children is in blue, and the average rate for correctly classifying children with SLI is in red. The horizontal line is the critical line, with a value of 90%. A black curve represents the average percent success rate of correct classification for both groups of participants. The x-axis represents all words, and the y-axis represents the percent success rates. There are five graphs for individual categories and one graph for all participants independent of age. Table 5.10 shows the total number of correct and incorrect classifications and the average rate of correct classification. The penultimate and final columns were very important for deciding which method best classifies children with SLI. The penultimate column shows the differences between the average error rates for the selected words and all words. As we can see, using only selected suitable words yields a higher value of differences between healthy children and children with SLI. The last column shows the number of words that satisfy the critical line. These words are suitable for the clear detection of children with SLI.

Healthy Children vs. Children with SLI - Words (38)					
Age Category	**Number of Classification**			Average Success [%]	Words (> 90%)
	Group	Correct	Wrong		
All	SLI-CH II (2)	13188	1254	89,16	21
	H-CH (1)	10683	1030	88,71	25
	∑(1 + 2)	23871	2284	88,98	25
A1	SLI-CH II (2)	4191	365	90,32	22
	H-CH (1)	1211	129	87,88	23
	∑(1 + 2)	5402	494	89,77	26
A2	SLI-CH II (2)	2580	338	84,95	17
	H-CH (1)	1233	640	59,85	1
	∑(1 + 2)	3813	978	75,2	1
A3	SLI-CH II (2)	3460	292	90,38	24
	H-CH (1)	1678	174	88,83	22
	∑(1 + 2)	5138	466	89,88	25
A4	SLI-CH II (2)	1665	211	86,33	13
	H-CH (1)	4433	71	96,67	33

	∑(1 + 2)	6098	282	93,61	30
A5	SLI-CH II (2)	1292	48	95,79	32
	H-CH (1)	2128	16	97,93	35
	∑(1 + 2)	3420	64	97,1	34

Table 5.10 Comparison of results for all words and selected words depending on the critical line.

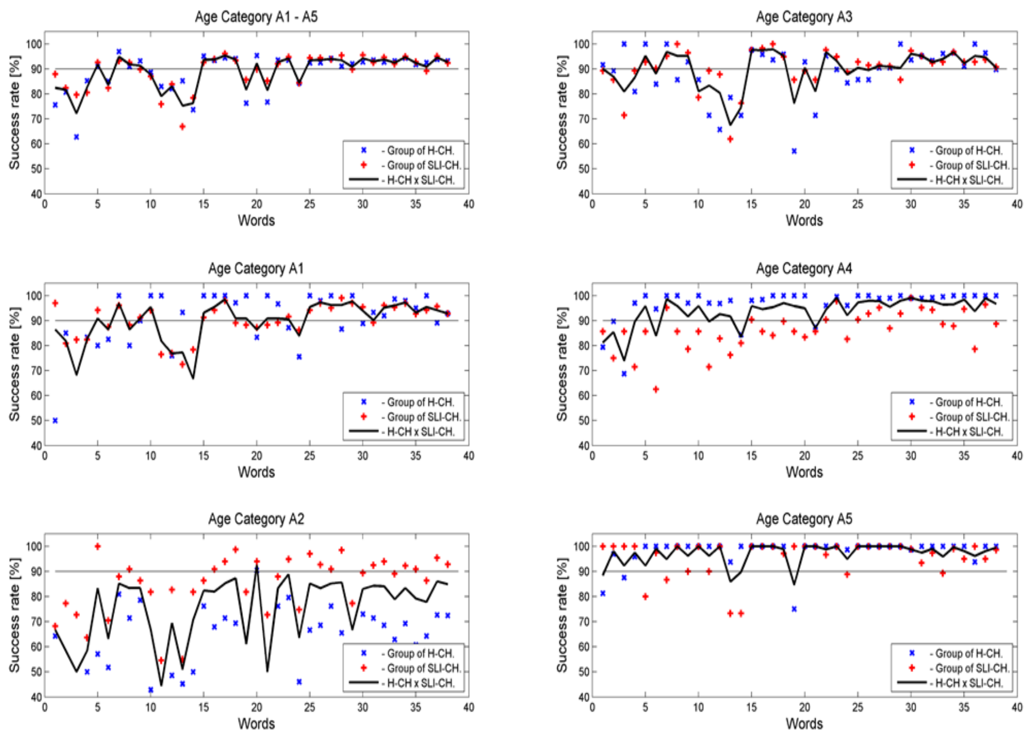


Figure 5.8 Evaluation of the classification of the suitable words. Data from healthy children are shown in blue colour, and data from children with SLI are shown in red colour. Horizontal line is the critical line. A black curve represents a differences between average error rate from healthy children and children with SLI. The x- axis constitutes all words, the y- axis constitutes average errors.

5.2.4 Differentiate children with SLI under 3 categories

This experiment focused on identifying and classifying three degrees of specific language impairment severity: mild, moderate and severe. This evaluation was based solely on the analysis of acoustic features (i.e., the higher the classification coefficient of the acoustic features, the greater the SLI severity). The results were compared with a speech therapist's assessment of severity. The aim of the experiment was to determine whether there is a relationship between the classification coefficient of acoustic features and the degree of SLI.

The procedure was as follows: Only the data from the group of children with SLI (p_sli) were classified. Each word was characterized using the same acoustic features. A classification coefficient was calculated using the average values obtained from these acoustic features and was assigned to each word. Then, the resulting classification coefficient (the average value for all words) was calculated for each participant. The following procedure is identical to the one used for error analysis.

To establish the parameters for a particular degree of SLI, we used ANN (the SOMs with default ratios for training, testing and validation were 0.7, 0.15 and 0.15, respectively) and HM classifiers. SOMs were used to determine the three values (V1, V2 and V3) around which most of the other values occurred (i.e., the final classification coefficients). The intervals used to classify the coefficients as a particular degree of SLI were determined using these values. The final classification was performed with a HM ("hand-made") classifier according to the classification intervals (V1 int, V2 int and V3 int). An overview of the classification into the three degrees of SLI, including the intervals, is shown in Figure 5.9.

This method is based on the distribution of the participants based on the coefficient of acoustic features. The results obtained from this method were compared in the following manner: A classification coefficient was assigned to each child. The degree of severity based on the feature analysis was determined according to the coefficient of the acoustic features, while the speech therapist's determination was based on his or her own assessment. Individual classifications of degrees of SLI were compared according to the following criteria:

$$mean\,val_{\text{mild}} < mean\,val_{\text{moderate}} < mean\,val_{\text{severe}} \qquad (0.34)$$

In other words, the mean value for the "mild" group must be lower than the mean value for the moderate class, and so on. If this relationship is comparable for both the speech therapist's assessments and the values obtained from the feature analysis, then the classification of severity based on feature analysis corresponds approximately to the speech therapist's classification.

The data were divided into three groups depending on the classification coefficient – mild, moderate and severe. It was then necessary to verify that the data sets could be compared with one another. The Shapiro-Wilks test for normality was used to determine whether the data were statistically normal. The obtained scores are summarized in Table 5.11and Table 5.12. Table 5.11 summarizes the classification obtained via feature analysis, and Table 5.12 summarizes the classification of the speech therapist using the classification coefficients obtained with feature analysis. All three classes, mild, moderate and severe, met the requirement of normality (p-val > 0.05) for classification via feature analysis; therefore, the t-test could be used to compare the three classes (see Table 5.11). Other values and the classification by speech therapist did not meet the requirement of normality, and the Wilcoxon rank-sum test was used for comparison (see in Table 5.14).

The p-values for all comparisons (mild vs. moderate, moderate vs. severe and mild vs. severe) were lower than the significance levels (t-test: alpha = 0.01 and Wilcoxon: alpha = 0.05).

Thus, I rejected a null hypothesis of equal means for Student's t-test and medians for the Wilcoxon rank-sum test, and I concluded that there is sufficient evidence to suggest that the mild, moderate and severe groups were not same at the default 1% (Student's t-test) or 5% significance level (Wilcoxon rank-sum test; see Table 5.11 and Table 5.12). These results can be considered correct, and it can be argued that there is a significant difference in the values of the classification coefficients obtained from the acoustic features of words produced by the mild, moderate and severe groups.

Feature Analysis Classification: Shapiro-Wilk Normality Test				
Group	Classification	Results	Hypothesis	
			H_0	H_1
p_sli	mild	W = 0.8584, p-val = 0.09222	1	0
	moderate	W = 0.9145, p-val = 0.3489	1	0
	severe	W = 0.8889, p-val = 0.1945	1	0

Table 5.11 Feature Analysis - Shapiro-Wilk Normality Test for the classification by the feature analysis.

Speech Therapist Classification: Shapiro-Wilk Normality Test				
Group	Classification	Results	Hypothesis	
			H_0	H_1
p_sli	mild	W = 0.8906, p-val = 0.2025	0	1
	moderate	W = 0.948, p-val = 0.6681	0	1
	severe	W = 0.9245, p-val = 0.4309	0	1

Table 5.12 Feature Analysis - Shapiro-Wilk Normality Test for the classification by the speech therapist.

Feature Analysis Classification: Student's t-Test				
Group	Classification	Results	Hypothesis	
			H_0	H_1
p_sli	mild vs. moderate	t = -9.3411, df = 12.852, p-val = 4.323e-07	0	1
	moderate vs. severe	t = -7.4632, df = 10.947, p-val = 1.292e-05	0	1
	mild vs. severe	t = -13.0022, df = 14.852, p-val = 1.62e-09	0	1

Table 5.13 Feature Analysis - Conformity testing for classification based on the error analysis through the Student’s t-Test and Wilcoxon Rank-Sum Test.

Speech Therapist Classification: Wilcoxon Rank Sum Test				
Group	Classification	Results	Hypothesis	
			H_0	H_1
p_sli	mild vs. moderate	p-val = 0.0024, ranksum = 53	0	1
	moderate vs. severe	p-val = 4.1135e-05, ranksum = 45	0	1
	mild vs. severe	p-val = 4.1135e-05, ranksum = 45	0	1

Table 5.14 Feature Analysis - Conformity testing for classification of speech therapist based on the error analysis through the Wilcoxon Rank-Sum Test.

The resulting classifications into three classes depending on the degree of SLI are displayed in Figure 5.9. There are 2 box plot graphs. Left graph is for classification through the feature analysis and right graph is for classification based on the assessment of speech therapist of values obtained from feature analysis. Mild category is displayed in blue, moderate category is displayed in red and severe category is displayed in black. Both graphs have approximately the same layout. Box plots representing the distributions of classification coefficients for all three classes show clear differences between these groups for all participants.

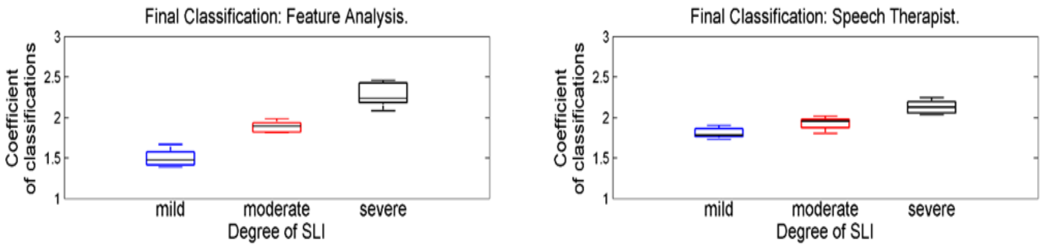


Figure 5.9 Comparison classifications - feature analysis and assessment of speech therapist - depending on the coefficients of classification.

The last experiment focused on the differences between individual words. The aim of this experiment was to determine whether all words were suitable for classifying children with SLI according to severity (mild, moderate and severe) and to identify the words with the greatest ability to correctly determine severity based on the error rate. This experiment compared the differences in error rates for each word according to the degree of SLI.

The procedure was as follows: Each word was evaluated in terms of the three coefficients. These coefficients represent the average classification for each class (mild, moderate and severe) based on feature analysis. The basic condition was the difference (cla_{mm}, cla_{sm}) between the moderate ($coef_{moderate}$) and mild ($coef_{mild}$) class and between the severe ($coef_{severe}$) and moderate ($coef_{moderate}$) classes(must be greater than zero) based on these formulas:

$$\begin{aligned} cla_{mm} &: coef_{\text{moderate}} - coef_{\text{mild}} > 0 \\ cla_{sm} &: coef_{\text{severe}} - coef_{\text{moderate}} > 0 \end{aligned} \tag{0.35}$$

The results are summarized in Table 5.15 and Table 5.16. Table 5.15 shows the feature analysis evaluation, and Table 5.16 shows the speech therapist's evaluation based on the coefficients of classification from the feature analysis. The tables compare the average classifications and the differences between the results for all words and for selected words. The last column shows the differences between the average classifications for the selected words and all for words. As we can see, the use of selected suitable words yields higher values for the differences between the three classes.

Feature Analysis Classification								
ID	Group	All Words			Selected Words			All vs. Selected words
		Average Error	Comparison	Difference [%]	Average Error	Comparison	Difference [%]	
1	mild	1,53	2 vs. 1		1,48	2 vs. 1		
2	moderate	1,86	3 vs. 2	**21,32**	1,87	3 vs. 2	**26,56**	**5,24**
3	severe	2,17	4 vs. 3	**16,88**	2,26	4 vs. 3	**20,55**	**3,68**

Table 5.15 Feature analysis classification: Comparison of mild, moderate and severe degree of SLI.

Speech Therapist Classification								
ID	Group	All Words			Selected Words			All vs. Selected words
		Average Error	Comparison	Difference [%]	Average Error	Comparison	Difference [%]	
1	mild	1,87	2 vs. 1		1,81	2 vs. 1		
2	moderate	1,88	3 vs. 2	**0,78**	1,91	3 vs. 2	**5,61**	**4,83**
3	severe	1,99	4 vs. 3	**5,85**	2,12	4 vs. 3	**10,89**	**5,04**

Table 5.16 Speech therapist assessment: Comparison of mild, moderate and severe degree of SLI.

Using these experiments, I tried to find specific words that are appropriate, i.e., that lead to the best classification for a given task, such as identifying children with SLI and determining the degree of SLI severity. The main aim was to find words in our database that can be successfully used for both types of tasks. Words selected to identify children with SLI had to meet the criterion of suitability for all participants regardless of age (because of the very poor results for category A2 in the Classification of Age Categories experiment, we specified that selected words had to show a classification success rate greater than 90% in at least 75% of experiments, i.e., at least 9 of the 12 classifications were above 90%). For a word to be selected for the determination of severity, it had to meet the criteria of suitability for feature analysis and for classification by a speech therapist. Table 5.17 shows the final evaluation for the speech tasks (T1 - T7) and the number of words that were suitable for a specific type of task. Compared to the error analysis results, the feature analysis expanded the range of words that were suitable for uniquely identifying children with SLI and distinguishing the degree of SLI.

Results of Group of Task Code.				
Task Code	**Healthy vs. SLI**		**Degree of SLI**	
	Number of Words	**Success rate**	**Number of Words**	**Success rate**
T1	0	0	0	0
T2	**2**	**20**	0	0
T3	**6**	**66,7**	**3**	**33,3**
T4	**6**	**100**	**1**	**16,7**
T5	**4**	**100**	**3**	**75**
T6	**3**	**100**	**2**	**66,7**
T7	**2**	**100**	0	0

Table 5.17 Feature Analysis - The Final evaluation of words.

5.3 Discussion

This section describes different perspectives regarding children with SLI. The method for identifying children with SLI was based on the audio-signal features specific to acoustic features of speech features. These features can be easily obtained and can be calculated without complicated modifications of the speech signal. This method seeks to apply a modern techniques to this issue. Modern techniques permit the calculation of as many different acoustic features of audio recording as possible. This approach reduces the need for difficult decisions about which features and methods are relevant to the task but adds the need for optimization and classification methods. The main benefit of this method is its foundation for the possible development of an automatic detection system.

The first presented experiments show that the method can distinguish between the utterances of healthy children and those of children with SLI. This ability was demonstrated for all children as a group and according to age group. Only 3 out of 98 participants were incorrectly classified.

Another experiment was devoted to finding suitable words to use for classification. These experiments used various types of words, from very simple words (i.e., with isolated vowels) to complex ones (i.e., multisyllabic words containing all of the vowels). This experiment confirmed the assumption that not all words are suitable for classification. This method increased the number of words identified as suitable for classifying children with SLI compared with the results of error analysis.

Similar to the results for error analysis, these experiments show that feature analysis can distinguish different degrees of SLI (mild, moderate and severe). These results were compared with the assessments of a speech therapist. The final experiment was devoted to finding suitable words for determining the degree of SLI.

6 Duration as a tool for the evaluation of analysis

The children with language impairment, regardless of severity, had reduced processing and response speeds on a range of tasks. This issue was resolved and confirmed in this work [99]. Generally, it can be assumed that analogies to this issue will be related to questions about the average duration of spoken utterances.

This chapter addresses the comparison of the duration of the words spoken by healthy children with that of children with SLI. The first assumption is that the utterances of children with SLI have a longer duration than the same utterances spoken by healthy children. The second assumption is related to the degree of SLI; namely, the duration of utterance correlates directly with the degree of SLI severity, i.e., more severe SLI should lead to a longer utterance duration. The aim of the present experiment was to verify that the duration of words can be used to verify the results of the error analysis and feature analysis.

6.1 Classification of the Children with SLI

The experiment involves all participants included in the speech database, specifically, the H-CH (44 participants) and SLI-CH II (54 participants) groups, and all words (n = 38). The methods of the experiment were as follows: For each word, the average duration of all participants was calculated, and the results were divided into two groups. The first group contained the data from healthy children, and the second contained the data from children with SLI. Finally, we compared the average duration of each word for both groups.

Figure 6.1 ws the evaluations of both groups. The y-axis represents the duration [s], and the x-axis represents all words used in this experiment. The healthy children's values are displayed in blue, and children with SLI's values are displayed in red. Table 6.1 illustrates the average duration of all words for both groups. The result is an average duration for the children with SLI that is approximately 27.56 percent higher than that of the healthy children.

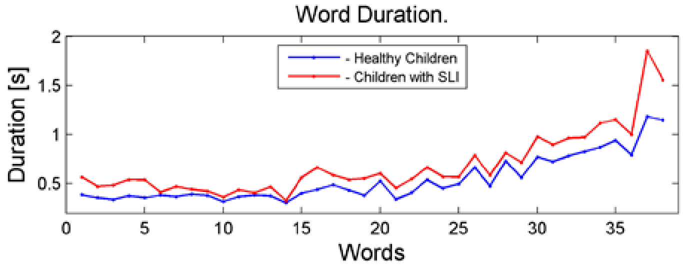


Figure 6.1 Average duration of words at healthy children and children with SLI.

Healthy Children vs. Children with SLI				
ID	Group	Average Duration [s]	Comparison	Difference [%]
1	Healthy Children	0,54		
2	Children with SLI	0,69	2 vs. 1	**27,51**

Table 6.1 Comparison of average duration at children with SLI and healthy children.

The table and figure show that the children with SLI had a longer duration of words than did the healthy children. This experiment verified the hypotheses about the speed of processing and response for a range of tasks.

6.2 Comparison of Error Analysis, Feature Analysis and Speech Therapist Classifications

The following experiment is divided into three parts and was used to verify the results of the speech therapist's classification, error analysis and feature analysis. These methods were described in previous chapters.

The experiment involved only children with SLI, specifically those from the SLI-CH II group. The children were divided into three groups depending on the degree of SLI severity: mild, moderate or severe. The first part examined the classification using error analysis, the second part examined the classification using feature analysis, and the third part examined the classification resulting from the speech therapist's assessment.

The design of the experiment was as follows: Each word was evaluated in terms of three coefficients representing the average duration of the word when it is spoken by children from the different severity levels. The first class included the durations of words spoken by the children classified with mild SLI; the second class contained the durations of words spoken by the children with moderate SLI, and the third class contained the durations of words spoken by children with severe SLI. The basic idea was that children with mild SLI would have a shorter average word duration compared with the children with moderate SLI, and the children with moderate SLI would have a shorter word duration than the children with severe SLI. This experiment compared the word durations for the mild, moderate and severe groups, and the results obtained using all of the words were compared with the results obtained using only selected words i.e., the words that provided the best classification.

The basic conditions for designating words as suitable were the differences in their duration for each class, which was determined with the following formula:

$$duration_{\mathrm{mild}} < duration_{\mathrm{moderate}} < duration_{\mathrm{severe}} \qquad (0.36)$$

In other words, the mean value of word duration for the mild class must be less than the mean value of word duration for the moderate class, and so on.

The resulting classifications into three groups according to the severity of SLI are displayed in Figure 6.2 through Figure 6.4. Figure 6.2 shows the classifications based on error analysis,Figure 6.3 shows the classification based on feature analysis, and Figure 6.4 shows the classification based on the speech therapist's assessment. The upper (for all words) and middle (for selected words only) graphs show the evaluation of word duration in relation to the degree of SLI. The mild category is displayed in blue, the moderate category is displayed in red, and the severe category is displayed in black. The x-axis represents the words, and the y-axis represents the average duration [s] of the spoken words. The lower graph shows a comparison of the average duration of each class, i.e., the average duration for the mild, moderate and severe classes. The average duration of all words is displayed in blue, and the average duration of selected words is displayed in red. The results of comparison among these three assessment methods are summarized in Table 6.2 through Table 6.4. Table 6.2 contains average duration of the words for the classification using error analysis, Table 6.3 contains the average duration of the words for the classification using feature analysis, and Table 6.4 contains average duration of the words for classification based on a speech therapist's assessment. The tables show a comparison of the average duration values and the differences between the all-words analysis and the selected-words analysis. The last column shows the differences between the average duration for selected words and for all words. Using selected words yielded higher values for the differences among the three classes.

- **error analysis**

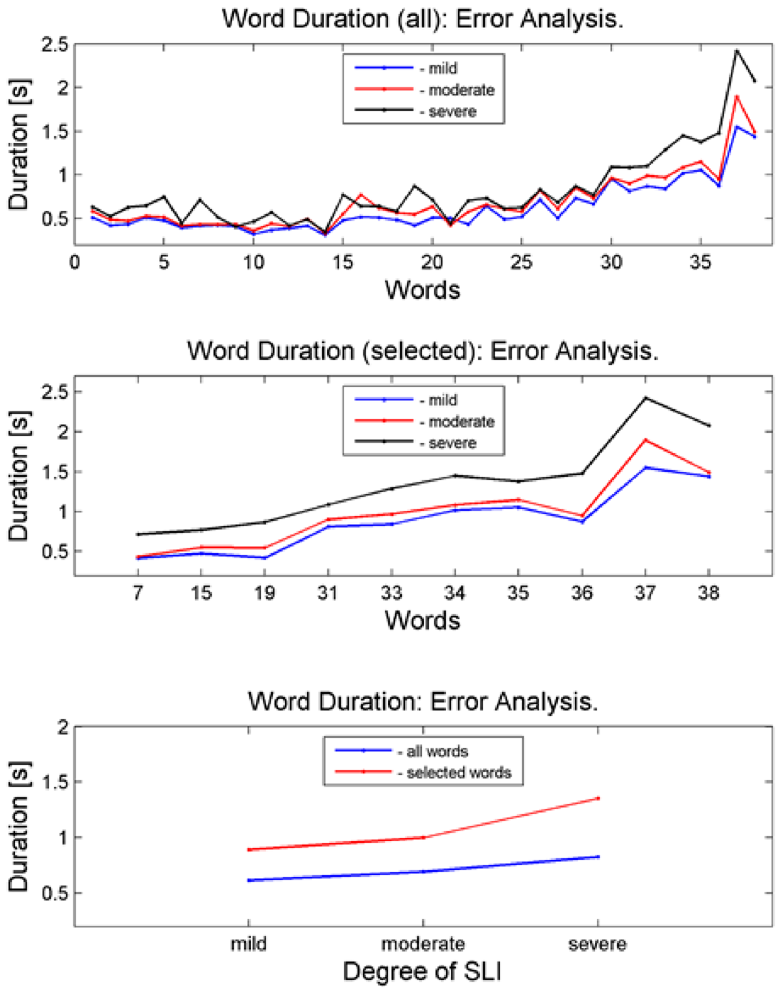


Figure 6.2 Average duration for all words and selected words in error analysis.

Error Analysis Classification - Evaluation of Duration								
ID	Group	All Words			Selected Words			All vs. Selected words
		Average Duration [s]	Comparison	Difference [%]	Average Duration [s]	Comparison	Difference [%]	
1	healthy	0,54			0,76			
2	mild	0,62	2 vs. 1	**14,34**	0,89	2 vs. 1	**16,77**	**2,44**
3	moderate	0,69	3 vs. 2	**12,69**	1	3 vs. 2	**12,14**	**-0,55**
4	severe	0,83	4 vs. 3	**19,37**	1,35	4 vs. 3	**35,45**	**16,07**
	severe vs. mild	-	4 vs. 2	**34,53**	-	4 vs. 2	**51,89**	**17,37**

Table 6.2 Comparison of average duration for all words and selected words in error analysis.

- **feature analysis**

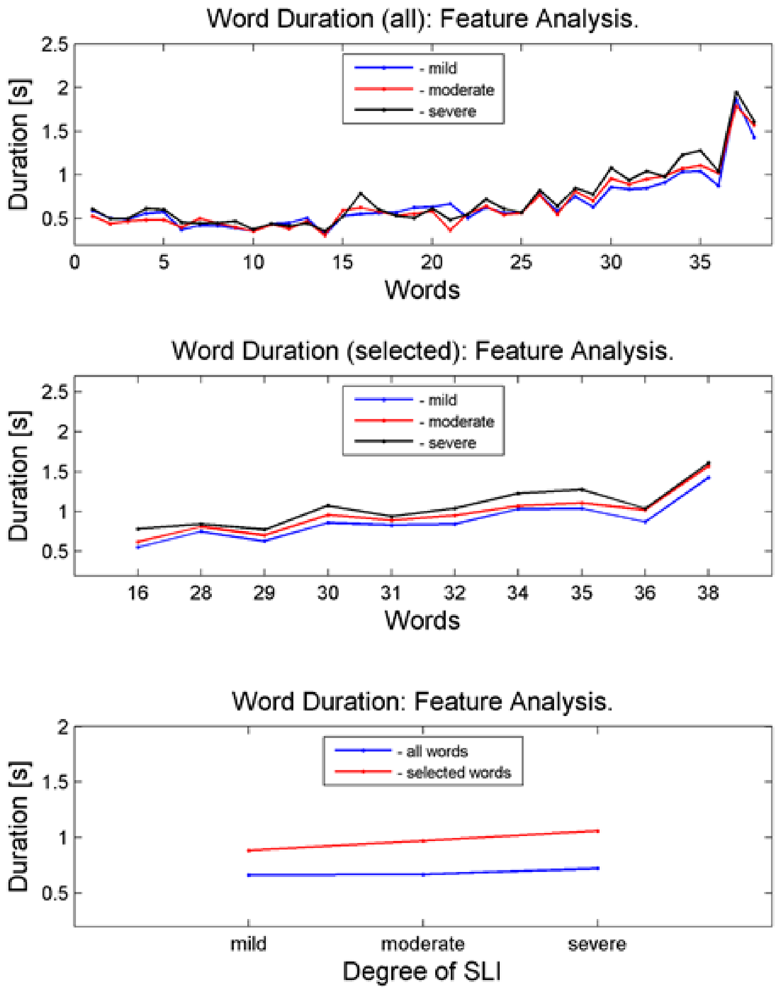


Figure 6.3 Average duration for all words and selected words in feature analysis.

Feature Analysis Classification - Evaluation of Duration								
ID	Group	All Words			Selected Words			All vs. Selected words
		Average Duration [s]	Comparison	Difference [%]	Average Duration [s]	Comparison	Difference [%]	
1	healthy	0,54			0,78			
2	mild	0,67	2 vs. 1	**23,61**	0,89	2 vs. 1	**14,16**	**-9,46**
3	moderate	0,67	3 vs. 2	**0,74**	0,97	3 vs. 2	**9,75**	**9,01**
4	severe	0,72	4 vs. 3	**7,79**	1,06	4 vs. 3	**9,32**	**1,53**
	severe vs. mild	-	4 vs. 2	**8,59**	-	4 vs. 2	**19,98**	**11,4**

Table 6.3 Comparison of average duration for all words and selected words in feature analysis.

- **speech therapist**

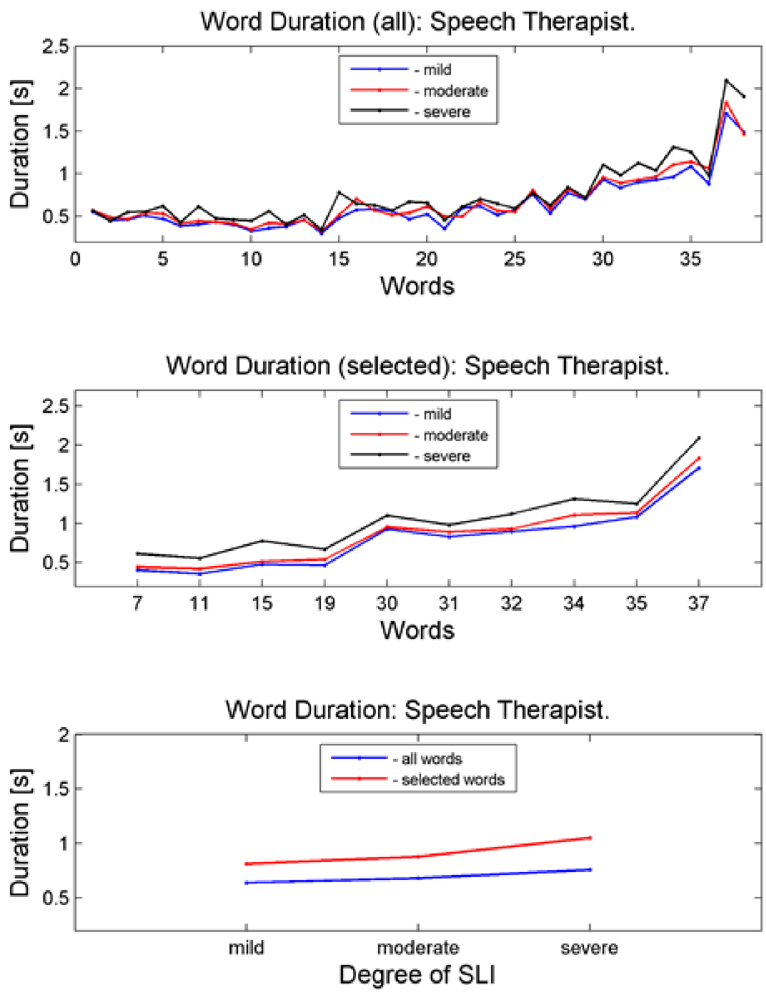


Figure 6.4 Average duration for all words and selected words in speech therapist classification.

Speech Therapy Classification - Evaluation of Duration								
ID	Group	All Words			Selected Words			All vs. Selected words
		Average Duration [s]	Comparison	Difference [%]	Average Duration [s]	Comparison	Difference [%]	
1	healthy	0,54			0,68			
2	mild	0,64	2 vs. 1	**18,77**	0,81	2 vs. 1	**20,04**	**1,28**
3	moderate	0,68	3 vs. 2	**6,42**	0,88	3 vs. 2	**7,97**	**1,55**
4	severe	0,76	4 vs. 3	**11,28**	1,05	4 vs. 3	**19,42**	**8,14**
	severe vs. mild	-	4 vs. 2	**18,43**	-	4 vs. 2	**28,94**	**10,51**

Table 6.4 Comparison of average duration for all words and selected words in speech therapist classification.

6.3 Summary

This section describes the comparison of the duration of words spoken by healthy children and by children with SLI. This comparison was used as a simple test to verify the methods used to analyze children with SLI. The main goal of the present experiment was to verify that the duration of words can be used to verify the results obtained using error analysis and feature analysis.

The experiment was conducted to determine whether the methods; i.e., error analysis and feature analysis; podmět are suitable for identifying children with SLI and for distinguishing three degrees of SLI. The results of both of methods were similar to a speech therapist's classifications based on professional assessment. This independent test clearly showed that it is possible to use both methods to differentiate the degree of SLI into three categories: mild, moderate and severe.

The tables and figures show the duration of the words for all categories of SLI (mild, moderate and severe) and for the individual severity groups. The resulting classifications were more accurate when selected words were used than when all words were used; the selected words showed a greater difference between groups.

7 Detection of Formants

The ability to produce and perceive speech originates in certain parts of the human brain. SLI is described as a neurological disorder of the brain [95], [103], [116]. The formant is a parameter that possesses a physical dimension (i.e., the presence of acoustic energy across the spectrum of speech sounds). The existence of formants can be related to human brain activity and to the movement of the articulatory system. This fact satisfies a condition for the use of formants in the classification of children with SLI. One of the prerequisites for the use of formants as descriptive parameters is the ability to obtain formant values with a minimum number of errors. To acquire suitable formants, we used the software tool FORANA [114]. The development of this software mainly driven by the need to correctly complete a formant analysis. Originally, formant frequencies were extracted from speech signals using PRAAT [85] acoustic analysis software. However, because the use of the PRAAT, specifically the use of Burg's algorithm to compute formants (method "To Formants (burg) ..."), the software produced formant classification errors in the course of the analysis, and results obtained using this approach could not considered relevant. More about this issue can be found in [100], [101], [109], [118].

7.1 Optimal Frequency Analysis

In order to be able to complete a proper classification of formant frequencies, one must first address the problems that are encountered when analyzing these frequencies. More specifically, this includes the following:

Determine an optimal formant analysis frequency for each speech sound. One major problem that arises as part of a formant analysis is that it is highly frequency dependent; and, a different frequency must be selected for each vowel and subsequently used for the analysis. This has to do with the underlying physiological aspects of the analyzed material. If an optimal frequency is not selected for the analysis of the given vowel, it often leads to an incorrect classification of formants, which eventually leads to the following two basic problems:

- Not all of the formants that are present in the given speech signal are included in the analysis.
- Replacement of one formant by a different formant.

The success rate achieved in detecting formants is directly dependent on the accuracy of the setting of the given parameters prior to the completion of the formant analysis. This is mainly tied into the width of the speech frequency band, which is used to detect the formants. The general rule is that, for an adult speaker, each 1000 Hz measured corresponds to 1 formant. When identifying a set of 5 formants, a sampling frequency of about 10 kHz is used (this corresponds to a band width of 5 kHz). However, this rule doesn't apply for child subjects. This is because a child's voice is constantly evolving and developing and it isn't fully developed until roughly age 7. This is one reason why the above rule doesn't apply. The other reason is the fact that a child's vocal tract is much smaller in size. Another important factor, which impacts bandwidth, is the character of the subject's speech. When articulating words, movement occurs in the person's vocal tract, which results in a change in the size of the laryngeal ventricles, causing a change in the spread of formants across the frequency spectrum of the given speech sound. This is why it is recommended to use a different bandwidth for the formant analysis of each individual speech sound. In our case, we are interested in an optimal analysis frequency for each speech recording. In finding out whether we have in fact chosen what we can consider to have been an optimal analysis frequency, we can make use of what is referred to as the 'vocalic triangle'. If a correct frequency has been used, a vocalic triangle will be present; and, if a correct frequency hasn't been used, then this vocalic triangle will not be present.

7.2 Correct Analysis Frequency

Even if we select a correct frequency for the formant analysis, this in and of itself can't always be taken as a guarantee that the formants will be classified correctly. The correct classification of formant frequencies is achieved through two separate but mutually complementary processes. The first of these processes involves the use of an internal sorting algorithm (SAL), which has been incorporated into the FORANA software (see the next section of this paper for a description of the SAL algorithm). The final classification of the formant frequencies being analyzed is done through the use of an artificial neural network in which certain modifications were made to the SOM. More specifically, these modifications involved alterations to the initial strength (weights) of the connections in the network. Under normal circumstances, the weights are set to have small random values; but, in our case, the weights were preset to average values of the identified formants. This SOM modification approach had a notable impact on the progression of the network's learning process and the final classification of the given formants. Once sorted in this fashion, the formants no longer deviate from their bandwidths and it can therefore be assumed that they have been properly classified (aside from a few exceptions, which however have no impact on the average value that we are looking for). The following charts show the difference between incorrectly and correctly classified formants (Figure 7.1). The 'incorrectly' classified formants are shown in 'red'. This problem occurs because we can't find the required number of sound peaks, which represent our formants, in every segment of a speech signal. This leads to a situation in which all of the formants that we need to analyze are not included in the analysis. In such situations, the missing formants have been replaced by the formant, which immediately follows. This can lead to an outcome in which it is impossible to analyze a particular formant because the formant has been replaced by the one that follows and it is this replacement formant that is classified instead of the missing formant (the one that should be classified). This error manifests itself as a shift in the speech frequency spectrum. In order to be able to properly classify formants, we have to make sure all formants, which are included in the analysis, correspond to their original peaks.

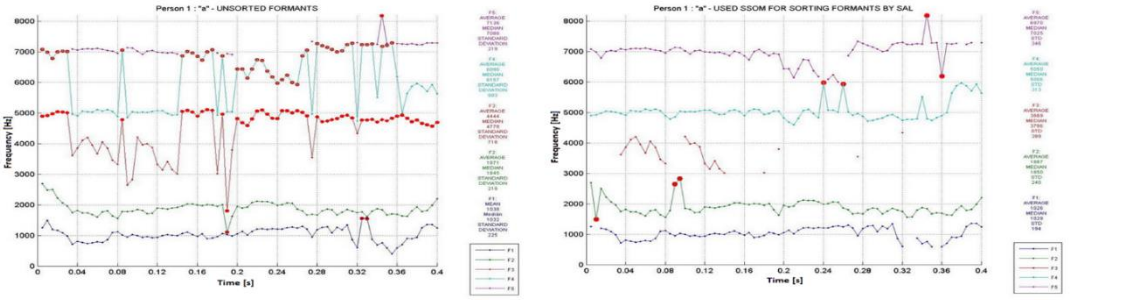


Figure 7.1 Incorrectly classified formants for vowel in the left chart and correctly classified formants for the same vowel in the right chart.

In the following chart, we can verify that the frequency that has been selected for the formant analysis (displayed in the vocal registration system) is correct. Figure 2 (vocalic triangle in red colour) illustrates the interdependency between formants F1 and F2 (the proper location of these formants would normally lead to the existence of the vocalic triangle). In this case, the formants were obtained by using the same frequency for all of the speech sounds included in the analysis; and, the vocalic triangle was therefore never obtained. In contrast, Figure 2 (vocalic triangle in blue colour) shows a vocalic triangle, which was created because all of the sounds included in the analysis were analyzed at their own optimal frequency. This clearly illustrates how the vocalic triangle is achieved.

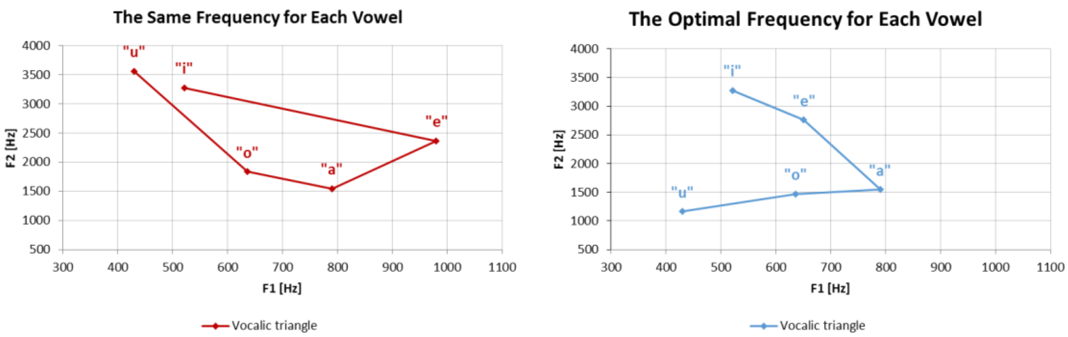


Figure 7.2 Red vocalic triangle with same frequency for each vowel and blue vocalic triangle with optimal frequency for each vowel in the right chart.

7.3 SAL Algorithm (Sorting Algorithm)

The SAL algorithm [101] is a tool used for the proper classification of identified formants. The algorithm was built into the FORANA software, which is designed to study formant frequencies as part of speech signal analyses. The algorithm sorts through formant frequencies by using a base nominal formant frequency and the neighboring frequencies. The algorithm also works with the bandwidth of the formants. Using these specific parameters, the algorithm assigns the analyzed formants into five different classes (for five formants). The algorithm works its way through a series of steps, which I will try to briefly describe; and, to give the reader an overall picture, I will use a highly simplified MATLAB programming structure (pseudo-code). In order to be able to properly classify our formants, we have to establish two specific conditions.

Condition 1

For the all of F_i shall apply:

$$F_i > F_{ai} \wedge F_i < F_{bi} \tag{0.37}$$

$F_i \rightarrow$ *i*-th formant

$F_{ai} \rightarrow$ *a*-th minimal frequency for the *i*-th formant

$F_{bi} \rightarrow$ *b*-th maximal frequency for the *i*-th formant

Condition 2

For the all of F_i shall apply:

$$\left|F_{in} - F_{in+1}\right| < X_a \tag{0.38}$$

$F_{in} \rightarrow$ *i*-th formant in the *n*-th segment of the speech signal

$X_a \rightarrow$ *a*-th maximal difference between two *i*-th formants of adjacent segments

The first condition is defined frequency band in which the formants can be. The second condition it is defined by the relationship between the formants from the segments of speech signal that are contiguous.

Function algorithm has several steps. Here I will describe only the most important of them:

Step 1

Calculated formants (in the the matrix matA) that satisfy both conditions for split, is stored in the matrix matB. Other formants are stored in the matrix matC. Both matrix have the same dimension.

```
for i = 1:n;
   for j = 1:m;
      if matA(i,j) == con1 && matA(i,j) == con2;
         matB(i,j) = matA(i,j);
      else
         matC(i,j) = matA(i,j);
      end
   end
end
```

Step 2

In the second step is to calculate the absolute value of difference between the formants, which don't meet the conditions for a division and average value of formant for which it was originally classified.

```
if matC(i,j) > 0;
   matE(i,j) = abs(matC(i,j) - matp(1,j));
end
```

Step 3

In this Step is calculated absolute value of difference between the formants and the average values as in step 2, but this applies to the formants and their neighboring formants , which satisfy the conditions.

```
if matC(i,j) > 0 && matB(i,j) > 0
   matD(i,j) = abs(matB(i,j) - matp(1,j));
   matD(i,j-1) = abs(matB(i,j-1) - matp(1,j));
   matD(i,j+1) = abs(matB(i,j+1) - matp(1,j));
end
```

Step 4

Now will be a comparison of values between them and selected the smallest value that it is stored in a matrix with the formants, which satisfy conditions.

```
for i = 1:n
   for j = 1:m
```

```
            minimal_value = [matC(i,j) matB(i,j) matB(i,j-1) matB(i,j+1)];
            matB(i,j) = min (minimal_value);
        end
    end
```

Step 5

The following steps of the algorithm will involve an accuracy check of the processes completed in Steps 1 through 4 using formant bandwidths (Bi) for a random replacement of formants. The average values of the formants are once again recalculated. The algorithm performs a subtle correction to the parameters used for the sorting of the formants because it now has more accurate average values to work with. This accuracy check and subsequent correction is included in the algorithm for the following reason: if we wanted to strictly adhere to the basic set of sorting criteria, it could lead to an outcome in which the formant frequencies that were correctly calculated and classified would be thrown out as unsuitable by the rough filtering mechanism that is part of this algorithm.

7.4 Problems Occurring When Identifying Formants in Children Suffering from a Neurological Disorder

The ability to produce and perceive speech originates in specific parts of the human brain. Developmental dysphasia or specific language impairment is described as a neurological disorder, which is why the use of formant analysis, which works with the physiological aspects of speech signal generation, is a suitable approach for research involving child subjects with this condition. It is assumed that the speech signal produced by children affected with this condition is marked by a shift in the frequency spectrum compared with the speech signal produced by healthy children. Although the original assumption was that this frequency spectrum shift occurred in any speech signal that the subject produced, the initial experiments, which examined only isolated vowels, failed to confirm this conjecture. Such speech signal recordings revealed no difference between healthy children and children with SLI. However, once the experiments began to include more complex speech patterns that required the extraction of individual vowels, the results started to confirm the initial assumptions. This change is clearly shown in the following chart (Figure 7.3), which depicts two vocalic triangles from a single speaker diagnosed with developmental dysphasia. The vocalic triangle representing isolated vowels is shown in blue, and the vocalic triangle representing vowels that were extracted from the Czech word “různobarevný” is shown in red. This particular word contains all of the vowels, thus allowing a comparison of the different vowels. When isolated vowels are analyzed, the resulting vocalic triangle is not the same as the vocalic triangle that results from the analysis when all of the vowels are contained in a sample spoken word. This particular example can be used to demonstrate a relationship between the complexity of the words being spoken and the shift in the speech sound frequency spectrum demonstrated by children with developmental dysphasia. This shift can be confirmed by listening to the recordings of the children's speech.

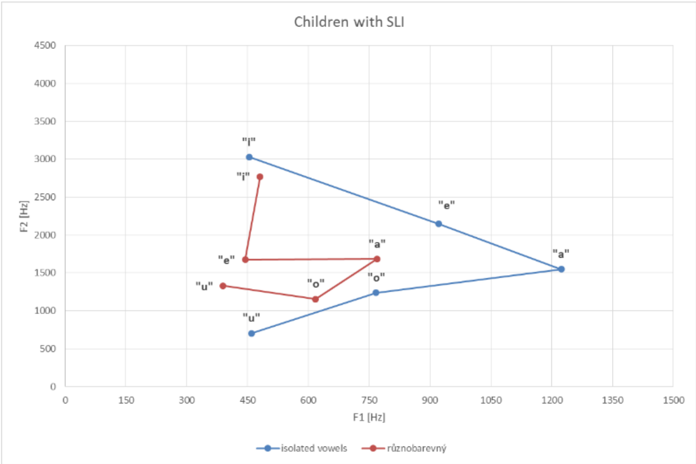


Figure 7.3 Vocalic triangles from word „různobarevný“ (en: „varicoloured“) and isolated vowels.

For the purposes of this research, the above-described experiment, which shows the vocalic triangle for speech recordings with different levels of complexity, was completed for all of the individuals being tested. To do this, I performed a speech signal analysis for the following two types of speakers: Speaker A was a randomly chosen child subject (age 10 years) diagnosed with SLI (from the SLI-CH II group); Speaker B was a randomly chosen child of the same age who did not have SLI (i.e., a healthy child from group H-CH). Both participants were tested using the same speech (text) to allow a proper comparison of the results. The first chart (on the left in Figure 4) represents Speaker A. It shows two vocalic triangles, a blue one for the isolated vowels and a red one for the vowels in “různobarevný” (“varicolored”). For simple speech patterns that can be used to obtain isolated vowels, the vocalic triangle is present. However, when analyzing the vowels retrieved from a more complex speech pattern, the triangle is absent. The arrows point to the positions where the vowels should be located under ideal circumstances. Figure 4 shows the same results as those obtained from the first subject who was tested (i.e., a child diagnosed with SLI who was tested as part of the earlier experiment). Figure 4 shows the vocalic triangles obtained from Speaker B; the triangle is present for both situations (simple speech and more complex speech). The results of these experiments confirmed the original expectations and pointed us in a direction that should be pursued in future research.

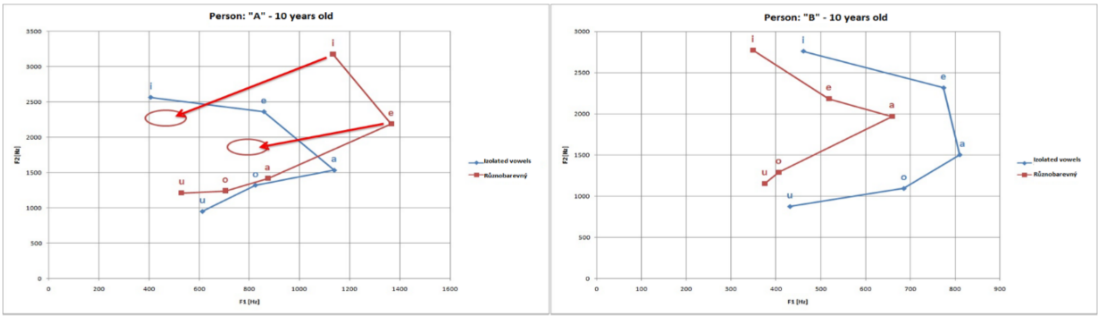


Figure 7.4 Vocalic triangles,left chart, obtained from child with SLI and vocalic triangles, right chart, obtained from healthy child. Both children were at age 10 years.

7.5 Experiment

The experiment involved only children with SLI, specifically those from the SLI-CH II group. Twenty-four children were randomly selected. Some of the children had one speech recording on record, and some had several. A total of 58 recordings were analyzed. The words chosen for the analysis included one multisyllabic word ("různobarevný": "varicolored") and all isolated vowels ("a", "e", "i", "o", "u"). The experiment was based on the comparison of two different vocalic triangles: One for isolated vowels and the other for vowels from a multisyllabic word. A prerequisite of this method is that the vocalic triangle from isolated vowels has the correct shape, while the vocalic triangle for vowels from a multisyllabic word is misshapen. We determined the three possible classification: correct, wrong and not produced. The results obtained from the vocalic triangle classification method are shown in Table 7.2.

Classification		
vocalic triangle	**correct**	**wrong**
isolated	1	1
multisyllabic	0	1

Classification	
number of classification	**not**
correct = wrong	1
correct ≠ wrong	0

Table 7.1 Vocalic Trinagle Classification - Evaluation of classification

Vocalic Triangle Classification			
Participants	Classification		
	correct	wrong	not
24	19	2	3
success rate [%]	79.17 %	8.33 %	12.50 %

Table 7.2 Success rate of method based on the Vocalic Trinagle Classification.

7.6 Detection of Formants in a Continuous Signal

The frequency dependence of vowels does not allow to use the method with Burg's algorithm in its current form for formants analysis in the continuous speech signal. Each vowel has a different position of articulation organs. It follows that each vowel has a different frequency of its analysis. This fact shows a distribution of the vowels in the vocalic triangle. The same situation is for other phonemes that are suitable for formants analysis. From this reason has each analysed phoneme a different frequency of its analysis. In the Figure 7.5 are displayed all vowels with their frequency spectrum and with position of the lips and articulation organs during speaking.

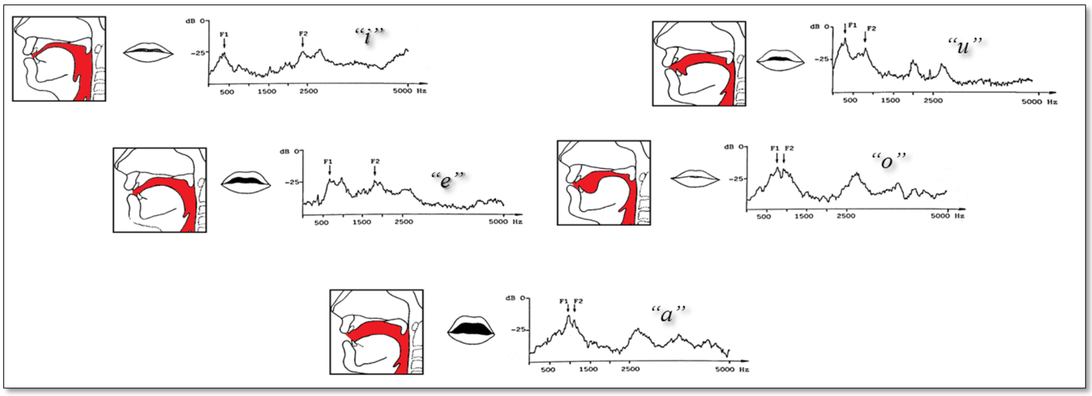


Figure 7.5 Frequency spectrum of the vowels and position of the lips and articulation organs during speaking [102].

7.6.1 Folding Algorithm of Formants (FF)

A new formant analysis method using a continuous signal solves problems with the frequency of analysis. The method can by very simply described in the following four steps:

1) Dividing the signal into *n* equally long consegments. Each consegments contains more segments.
2) Multiple formant analysis. Each consegments is subjected to more formant analysis.
3) Selection of the appropriate frequency.
4) Compositions of the formants.

A graphic illustration of the FF algorithm is shown in Figure 7.6. In this case, we tested a speech signal for the word "mateřídouška". We can see all of the steps of the algorithm. In this experiment, we calculated five formant bands for the whole word. In particular, a multiple formant analysis is shown. For the five formant bands, we calculated the values of the formants for cases from 3 to 7. These values were compared to indicate the values for 5 formant bands. The basic prerequisite was the use of the proper formant analysis, which was introduced in the first part of this chapter.

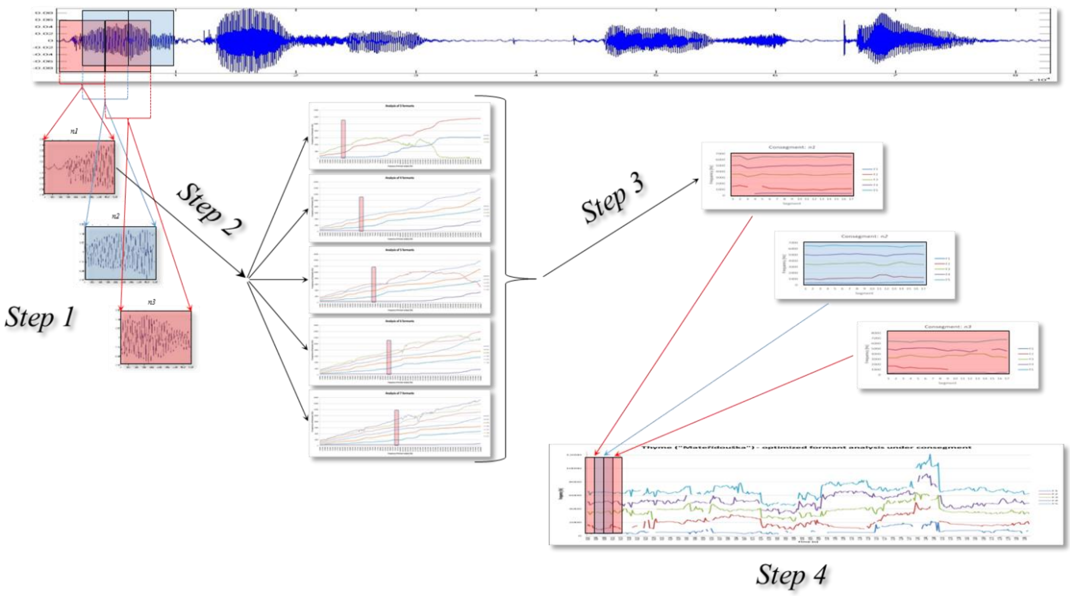
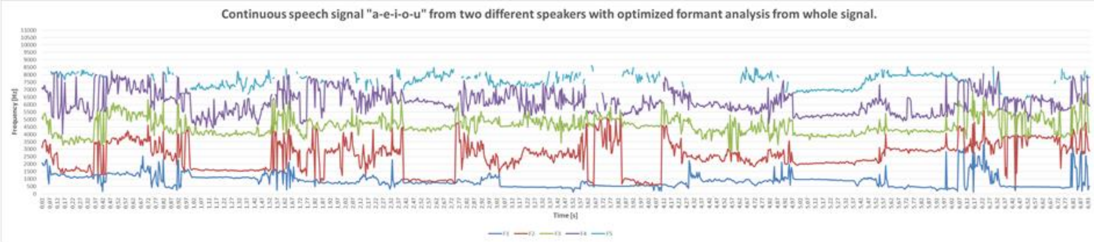


Figure 7.6 Graphic illustration of Folding algorithm of formants

7.6.2 Comparisson of the new FF algorithm and classical approach

The experiment involved a continuous speech signal with this text: "a-o-u-e-i". The recordings include two different speakers: One speaker is a child (child with SLI) and second speaker is an adult (speech therapist). The recording is from their session. We compared the two different approaches and examined their differences in classification. The first was a new FF algorithm (marked as FF) and the second was the classical approach used in PRAAT via the method "To Formant (burg)..." (marked as CA). The exact procedure for obtaining formants from PRAAT using this method is reported in [100] or on the PRAAT website. Comparative data sets (marked as references), i.e., formants, were obtained from the vowel extracted from the continuous signal. Formants were calculated for each vowel separately. For proper formant analysis, we obtained information from the speech signal using a method described at the beginning of this chapter and in [101]. This method is based on the Burg's algorithm. The correctness of the calculated formants were verified using a vocalic triangle.

The results of the analyses are displayed in Figure 7.7. The upper chart shows the formants calculated using the CA method, and the lower chart showed the formants obtained using the FF algorithm. The formant analysis frequency used for the CA method was set at the level with the smallest number of errors, i.e., the most calculated formants. The appropriate frequencies for the FF algorithm analyses were set completely automatically. The differences are evident. The curves in the upper chart (generated using the CA method) cannot be clearly distinguished for the entire signal. The curves in the lower chart (established with the FF algorithm) can be clearly distinguished for the entire signal.



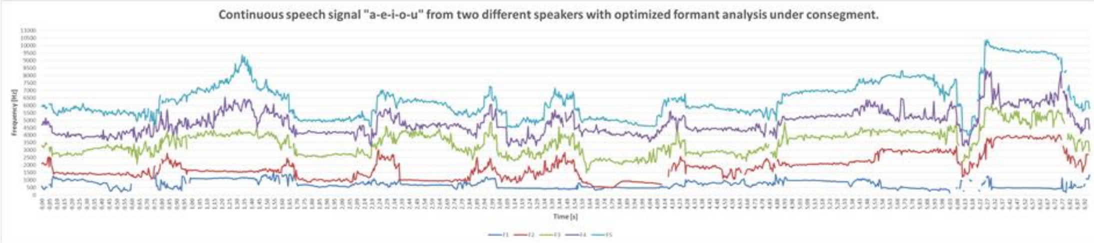


Figure 7.7 Difference between FF algorithm and CA method.

Table 7.3 contains final evaluation of percent difference for all vowels and overall average difference for used methods. Final difference between FF algorithm and reference data set is only 1.96 %, while difference between CA method and reference data set is 24.78 %. We set the criterion for acceptance of the method used as a successful that the level of difference between reference data set and data from evaluated method is not greater than 5%. FF algorithm satisfies this condition but CA method does not satisfy it. Complete results from this analysis are in supplement 4.

Evaluation Method of Analysis							
Success Rate [%]		Type of Vowel					Ø all
ID	Method	/a/	/e/	/i/	/o/	/u/	
1 vs. 2	FF	2,37%	2,41%	2,75%	1,31%	0,95%	1,96%
1 vs. 3	cla	0,47%	4,15%	17,49%	20,48%	81,34%	24,78%

Table 7.3 Final evaluation of succes rate of FF algorithm.

7.7 Summary

In the first part of this chapter, we described a method for correctly classifying formants using our FORANA software. A new algorithm, SAL, can be used for the final classifications and had to be developed and integrated into the software. More specifically, the approach involved creating a SAL sorting algorithm and an SOM-type artificial neural network. In our experiments, we successfully classified children with SLI using a method that uses two different vocalic triangles to identify differences between healthy children and children with SLI. Children with SLI have problems correctly saying difficult (multisyllabic) words. Formant analysis clearly verifies whether the vowels in utterances are pronounced correctly.

In the second part of this chapter, we presented an way to detect formants in a continuous signal. The method is based on the frequency dependence of individual speech sounds. This property does not allow the efficient use of formant analysis. In our experiment, we showed that if we combine correct formant analysis with method of detecting formants in the continuous signal, we can obtain much more relevant information from the signal, such as the ability to distinguish between different types of speakers (typically adults and children). This information can be used to classify types of spoken utterances, but more importantly, it offers the possibility of identifying children with SLI.

8 Applications

This chapter deals with application and software which is based on the methods from this thesis. Attention was paid to design and implementation of algorithm from Error Analysis in the form of a simple test. This test is used to find children predominantly with SLI. The main aim is to possess a simple tool for detection children with this disorder, which can be used anywhere by anyone. We present a mobile application SLIt Tool which is designed for iPad.

8.1 SLIt Tool

The Test of Specific Language Impairments (SLIt Tool) is a tablet application that calculates very simple test for detection of children with SLI on iOS platform (Apple, Inc.), specifically for using on iPad (iPad 3rd generation or newest). The application is based on the Error Analysis, which was presented in previous chapters and is the original method of testing children. The main idea was the creation of the application that is very friendly and easy to use. A mobile device enables extreme flexibility in where the tests are carried out, making it possible to perform the test anywhere (e.g., at home, in kindergartner, in the park, …) and not just in speech therapist clinic. The users of this application will be primarily parents of children. The uniqueness of this application is the possibility to perform timely diagnostic test for determining SLI; in the result this means faster and more successful speech therapy.

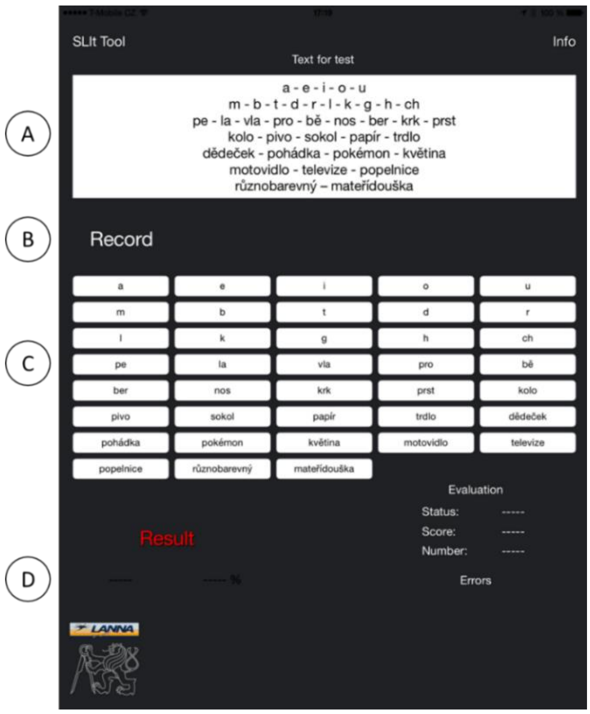


Figure 8.1 Screen shot of SLIt Tool application on iOS platform. This figure shows a basic window for testing of children.

Functional design of SLIt Tool (see in Figure 8.1) is devided onto a 4 basic part. Part A is for text for testing of children (list of the vocal tasks from Table 3). Part B is for recording of child speech. Part C is for correction of utterances and part D is for result and evaluation of test.

We compared two participants from our current research. Both are from the same age category. Healthy child (see in the left part Figure 8.2) and child with SLI (see in the right part of Figure 8.2). While healthy child said all utterances correct (one wrong utterance is red (i.g. one error), result is in green with a success rate and final status is "healthy") then child with SLI said utterances with many errors (wrong utterances are red, result is in red with a success rate and final status is "non-healthy").

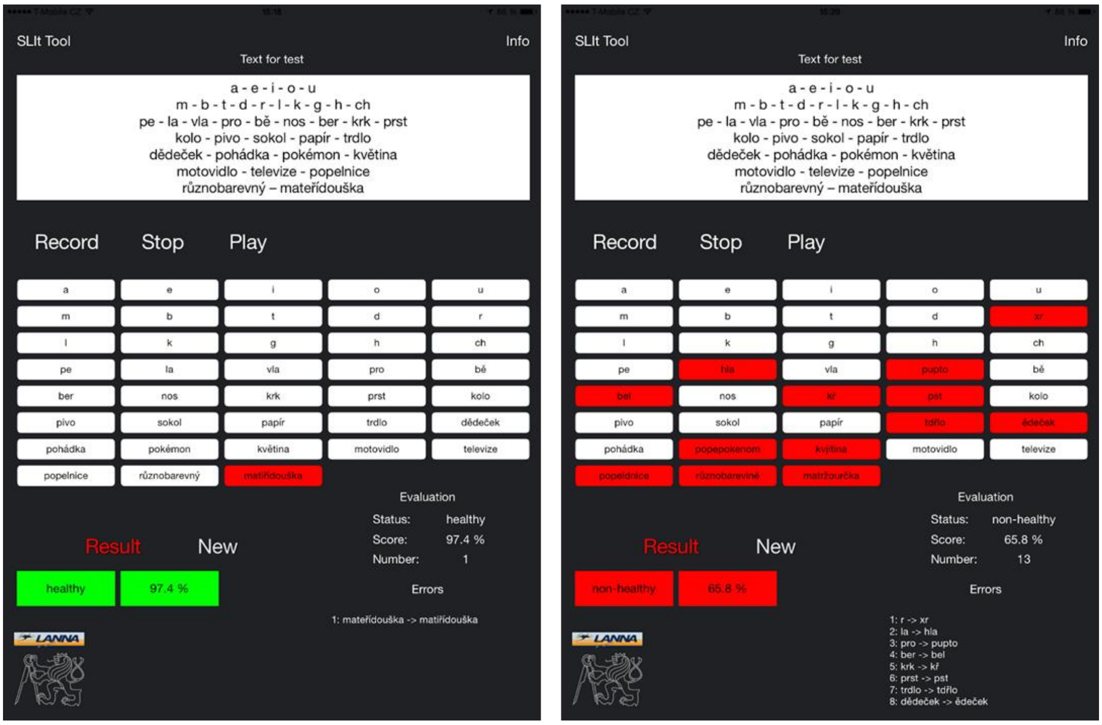


Figure 8.2 Screen shot of SLIt Tool application on iOS platform (a) A part in the left: In this case we tested a healthy children with this results; error analysis is equal to 97.4 % success rate and the final state is healthy. (b) A part in the right: This case shows a real test of child with diagnosis of SLI with this result; error analysis is equal to 65.8 % success rate and the final state is non-healthy.

Test procedure is a very simple and its course is identical to the way of recording of children's speech from our research. The test runs in 4 steps:

1. An adult reads the text. The tested child listens and repeats the same text. Text is in the part "Text for test."
2. The course of the test is possible to record by click on the "Records" button; for later replay of speech recording.
3. The correction of spoken words is done using text boxes, which were pre-filled text. If the child repeats the word wrong, it is necessary to edit the text depending on the spoken

words, eg the wrong spoken word "pro". Change from "pro" to "pupto" (in our real example).

4. The evaluation of this test is done by clicking on the "Result" button. Results can be sent to speech therapist via email for more precise classification.

The result of the output from application is the recommendation for children which obtain a status as a "non-healthy" to visit a speech therapist for deeper classification of their speech problems.

The application, SLIt Tool, is stored in iTunes. It is designed for iPad users and is free for everyone. Downloading applications is possible either by inserting the name of the application "slit tool" in the search field applications in the iTunes Store or by clicking on the link: https://itunes.apple.com/WebObjects/MZStore.woa/wa/viewSoftware?id=994893998&mt=8.

Currently I collaborate with clinical speech therapist which intesively tests our application. Together tune the final version of the application that everyone can use.

8.2 Web administration (WA)

Another application that is created is an online administration of the developed database SLI-CH II. It is a web supporting application (see in Figure 8.3) [115], and access is made through a login window. The administrator assigns a name and password and the login is secured by a cryptographic hash function, type SHA-1 [96]. The reasons for creating an online application are to provide immediate access to the data records of the children and to maintain the ability to create the text that the children repeat for the recording. We found that children with SLI are able to remember this text. It is possible to choose the order of recorded speeches (for repeated recordings), and the application is able to generate three variants of text:

- Original word order
- Random word order within sub-groups
- Random order of all words

This application saves the history of the recordings that includes some supplementary information. Here, we can find information about the patient (name, surname, date of birth, diagnosis and its degree, and eventually other supplementary information) and detailed information about the recordings (number of recordings, their serial number, date of recording, the type of the recorded text, list of all speeches, etc.). This complete information is able to be shown immediately, to be edited or to be saved as txt files for all researchers involved. All the information is available only for appointed medical staff or researchers.

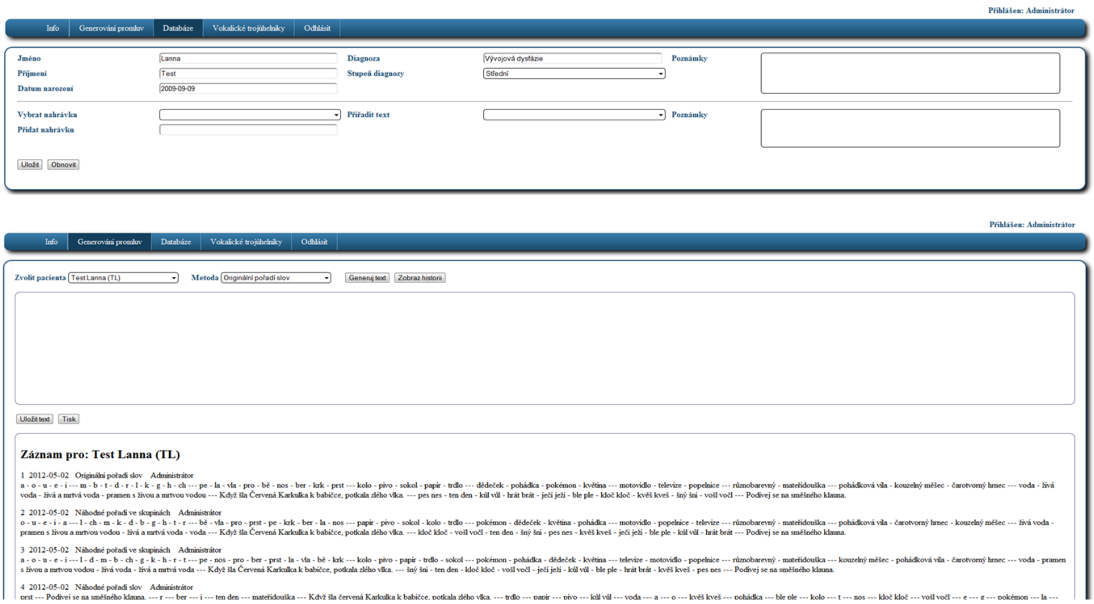


Figure 8.3 Screen shots of the online database management system [115].

8.3 Summary

In this chapter, we introduced a mobile application for detection children with SLI. The main benefit is a very simple operation and the ability to perform a test with immediate display of result. The second application, Web Administration supporting tool, is primarily intended for researchers and other staff who collaborate on research of children with this disorder. The application is focused on the on-line management of speech database that is created as a part of the research dealing with the children with specific language impairments. Although is it a relatively simple and basic property of each research project, the creation and processing of speech database is a very lengthy process and this application simplifies this work.

9 Final Discussion and Conclusion

A method described in this thesis was developed to analyse disordered children speech. I focused on children with specific language impairments and believe that the utilization of this method in clinical practice will bring more insight into progression and treatment of the disease and help to think out efficiently treatment of the disease.

This work is a part of the research project that is focused on treatment of SLI. The project integrates results from the fields of neurology, psychology, logopedics, MRI tractography and speech processing. In cooperation with the department of Paediatric Neurology in 2nd Faculty of Medicine of Charles University in Prague we develop methods for utterance analysis that further advance diagnosis of the children with SLI and help to find the most efficient therapy.

Laboratory of Artificial Neural Network Applications is focused on development of detection method using the artificial neural network. ANN, namely Kohonen Self-Organizing Maps, were chosen for their robustness to artefacts and noise which is present in the signal. Another important characteristic is the ease of implementation. This specific feature helps to develop methods that might be later used directly in clinical practice without any special and expensive equipment. The results of this research are a combination of many methods based not only on speech processing, but also on medical diagnoses (e.g., EEG and MR tractography), methods that are not discussed in this thesis. Children's speech processing with regard to normal and pathological speech is also a combination of standard and nonstandard approaches. In this way, I try to eliminate weaknesses of these methods (e.g. formants are very susceptible to the quality of the recordings) and ensure correct classification, for example, while using their strengths (robustness of the ANN and their resistance to the quality of recordings). The combination of standard methods (formants analysis, MFCC, and so on), ANN methods (Supervised Self-Organizing Maps) and non-standard methods (number of errors in utterances) has resulted with a high degree of successful classification for the children with SLI.

In this thesis, I described almost ten years long time of work with data collection comprising speech signals of children with or without specific language impairments. I did not focus only on the description of the entire children's speech database and its subgroups. This thesis is mainly about methods that was developed to classify children with SLI and based on direct database processing. I believe that the utilization of the method in clinical practice or in real life will bring more insight into progression and treatment of the disease and help to think out efficiently treatment of the disease. Especially the possibility to the earlier diagnosis of children with SLI.

- **Speech database**

The creation of a speech database designed for research focusing on children SLI (especially developmental dysphasia) is a very lengthy process. In contrast to the other speech databases, data acquisition is much more complicated, because I work with children, very often of preschool age, and this can create specific problems. The most complex problem is how to maintain the children's attention throughout the recording. Children are very often distracted and may not focus on the recording: they may speak during the speech therapist’s speech, asking him about various things or they may play with the microphone. Therefore, the recordings contain many unwanted sound remnants. Another problem is the reluctance of parents, of both healthy and sick children, to agree to the research. The research process involves many examinations of their children and parents have to fill out a medical questionnaire, which is one reason why the dataset is not larger. It is important to note that these are biological data which have their own particular specificity and the quantity of the data is not the most significant of their properties. The last problem is related to the technique and location of the recordings. The rooms were not soundproofed and normal urban noises can be overheard in recordings. The same applies to the quality of the recordings. At the beginning of our research study we had a different equipment for recordings than at the end of the research. The recordings are with a variable quality and with a different level of background noise. It is very complicated to compare recordings of low quality and those of high quality. Despite these problems and shortcomings, I managed to find several methods for identifying children with SLI.

- **Error analysis**

Error analysis is based on the number of errors in the utterances. The advantage of this access is that its function does not require complex computational methods and can be performed by anyone. This approach provides complete information about the most common errors and substitutions between speech sounds in the utterances of the children with SLI. The results of the error analysis show that the number of errors in utterances is an important cue that can be used as a border line to determine healthy children and children with SLI.

This approach enabled the creation of a simple test to detect children with SLI, specifically mobile application SLIt Tool. The initial goal of creating a test originated when we solved the problems with the creation of a children's speech database. The main problem is in the quality of the speech recordings, which is variable because the recording was carried out in real environment of speech therapy clinics. All of these problems have been described in [7]. The criteria used to create the simple detection test included noise immunity in the recordings, and the newly developed SLIt Tool for testing that we presented in this article fulfills this condition.

The result of the output from application is the recommendation for children which obtain a status as a "non-healthy" to visit a speech therapist for deeper classification of their speech problems. The indisputable advantage is the simplicity of this method, which can be used anywhere by anyone. To use this test, no special technical knowledge that would prevent its wider use by speech therapists or other language therapists is required.

- **Feature analysis**

This method, called feature analysis, for identifying children with SLI based on the auditory signal features specific to the acoustic features of speech. These features can be easily obtained and can be calculated without complicated modifications of the speech signal. This method seeks to apply a modern techniques to this issue. Modern techniques permit the calculation of as many different acoustic features of audio recording as possible. This approach reduces the need for difficult decisions about which features and methods are relevant to the task but adds the need for optimization and classification methods. The main benefit of this method is its foundation for the possible development of an automatic detection system. Similar to the results for error analysis, these experiments show that feature analysis can distinguish healthy children and children with SLI and also can distinguish different degrees of SLI (mild, moderate and severe). These results were in agreement with the assessments of a speech therapist.

- **Formants**

Formant analysis provides information about the individual vowels in the frequency spectrum. Each vowel has a clearly defined location in the vocalic triangle when two conditions are fulfilled. First, the formants must be correctly classified. Second, the utterance must be properly spoken. The whole point of using formants in our research is to verify the correctness of the spoken utterances. Children with SLI have problems with correctly speaking difficult (multisyllabic) words. It is supposed that developmental dysphasia can influence a shift of formant frequencies in spectral characteristics compared with the formant frequencies of healthy children. Formant analysis clearly verifies whether the vowels contained in utterances are pronounced correctly. Otherwise, if there are any errors in the analyzed vowel in the utterance, there is a shift in the frequency spectrum. This fact means that the speakers have the articulatory organs in a bad position and the distribution of articulatory cavities has the wrong shape for a vowel. In the final result, it leads to the malfunction of speech control in the brain. The disadvantages of using formant analysis include a notable susceptibility of the quality of the recordings and the demand for correct classification of the calculated formants.

- **Shortcomings**

The aim of this thesis was to determine an auxiliary method of classification of children with SLI. These methods are based on the different approaches, e.g. based on the number of pronunciation errors in utterances or based on the formants analysis, MFCC, and so on. The results of these methods were satisfactory. However, complications were with the task used. The description of utterances was done by many students and that caused many errors in the labelling of utterances, particularly in healthy children. I have shown that a small number of errors has a negligible influence on the resulting classification.

One of the biggest problems with this method is a number of participants included in our research. Our data are based on real human data, and their specificity is that their quantities belong to the minority of their properties. The group size of participants included depended on the willingness of the parents of either healthy or sick children to connect to the research. The study included many examinations of their children and a medical questionnaire. Compliance with these parameters is the main reason why the dataset is not larger.

Another fact that we could not control is difference between the number of girls and boys in each of the subgroups. We have a higher number of recordings of boys than girls in the SLI-CH subgroup (children with SLI) compared with the H-CH subgroup (healthy children). The whole H-CH subgroup contains recordings of 45 girls and 25 boys. In contrast, the subgroup of children with SLI, the SLI-CH II subgroup, contains recordings of 26 girls and 46 boys. This fact corresponds with the diagnosed disability in boys and girls. SLI is detected much more frequently in boys than in girls; however, more girls than boys who participated in our study attended the primary school and kindergarten.

The words used in the analyzes were limited by entering the previous grant NR8287-3/2005. This grant dealt only to distinguish children with SLI. Task with focusing on determining degree of SLI, mild moderate and severe, was solved in grant NT11443-5/2010. Entering the grant was modified, but the selection of words remained the same. It was confirmed that the simple words have minimal classification ability to compared with more complex words. Another shortcoming based on the grant is the determining the children into three classes by only one clinical speech therapist with 20 years experience. Despite this fact the results of experiments coincided with a speech therapist’s assessment.

- **Conclusion**

These results prove that it is possible for children with specific language impairments to be clearly identified and distinguished from healthy children, based on their speech and speech skills. I combined traditional and alternative approaches to this issue and obtained a resistance tool that is not dependent on the quality of the captured recordings. I found several different classification methods for children with SLI. All of these methods can be used separately for the classification of these children. Each method yields a high level of success in classifying children with SLI, but each has its particular limiting factors and shortcomings. Using these methods together, we are able to eliminate these shortcomings and obtain a powerful tool for diagnosing children with SLI.

The possibility of the classification of children with SLI using their speech analysis is very significant. The first factor is the opportunity to detect more children with this disorder. For these children, therapeutic treatment can be performed earlier. Another important factor is the financial aspect. Diagnostics that use speech are less demanding on hardware, and are more financially feasible compared with other specific medical examinations.

- **Further Development**

Error analysis and iPad application: Future work will be concentrate on two main areas. The first is the creation of reference lists of utterances in collaboration with speech therapists that are suitable for each individual age category. It is necessary to account for the verbal skills of the children. The goal is to determine utterances that we can use to determine the degree of SLI and that can contribute to the earlier diagnosis of SLI. This approach relates to the long-term goal of creating a specific application for automatic testing children through their speech. Everyone, especially parents, will be able to use the application to detect specific language impairments.

Feature analysis: Nowadays used data set contains almost 6.5 million of acoustic features and takes approximately 14 days on powerful computers. The whole data set contains a huge amount of data (60 million acoustic features); i.e. data sets with 1582, 4134 and 6670 acoustic features. In further work would be useful to explore other data sets (with 4134 and 6670 acoustic features) and find the top acoustic features. This set of selected features will be classify more accurately especially the task of classification into 3 categories depending on the degree of SLI.

The current situation is that the algorithms are implemented in the tens of files for Matlab and in the batch files. This is advantageous when dealing with development, since it allow the problems very easily debug and modify. Before possibilities of extensive testing must be optimized number of acoustic features and rewritten to further decrease computation demands.

It is advisable to rewrite algorithms into a unified system, for example C++ language, for the possibility of computation in multithreaded way on many cores in one time. This will reduce the time required for analysis. This optimization allows to use this approach in automatic detection system.

10 References

[1] LANNA - Laboratory of Artificial Neural Network Applications, Department of Circuit Theory at the FEE-CTU in Prague. http://ajatubar.feld.cvut.cz/lanna/

[2] Kohonen, T. (2001). Self-Organizing Maps. Third, extended edition. Springer.

[3] Tučková J, Bártů M, Zetocha P (2009) Aplikace umělých neuronových sítí při zpracování signálů. (lecture notes in the Czech) Prague: Czech Technical University in Prague. ISBN 978-80-01-04400-1.

[4] Leonard L B (1998). Children with specific language impairment. Cambridge, MA: MIT Press. ISBN 9780262621366

[5] Bishop, D V M (2014). Uncommon Understanding (Classic Edition): Development and Disorders of Language Comprehension in Children. England Psychological Press. ISBN 9780203381472.

[6] Klein, L.T., Amato, V. (2013). Languages and Linguistics. Nova Science Publishers, Incorporated. ISBN 9781629483337.

[7] Zwitserlood, R. L. M. (2014). Language Growth in Dutch School-Age Children with Specific Language Impairment (Proefschrift). LOT. Netherlands. ISBN 9789460931383.

[8] Specific Language Impairment. Available: http://www.nidcd.nih.gov/health/voice/pages/specific-language-impairment.aspx Accessed 03 November 2014

[9] Tomblin, J.B., Records,N.L., Buckwalter,P., Xuyang Zhang, Smith,E. and O'Brien, M.: Prevalence of Specific Language Impairment in Kindergarten Children. Journal of Speech, Language, and Hearing Research, December 1997, Vol. 40, 1245-1260. doi:10.1044/jslhr.4006.1245

[10] Conti-Ramsden, G. and Durkin, K. (2011) Specific Language Impairment, in Child Psychology and Psychiatry: Frameworks for Practice, Second Edition (eds D. Skuse, H. Bruce, L. Dowdney and D. Mrazek), John Wiley & Sons, Ltd, Chichester, UK. doi: 10.1002/9781119993971.ch29

[11] Beesems, A.G.: Foundation Developmental Dysphasia. Theory Diagnosis and Treatment. Speech Language Therapist, Developmental Dysphasia Foundation Amsterdam.

[12] Bočková, B.: The Support of Pupils with Specific Language Impairment (in Czech: Podpora žáků se specificky narušeným vývojem řeči). 1ed, Brno: Masarykova univerzita, 2011, ISBN 978-80-210-5609-1.

[13] http://www.logopedonline.cz/vady-reci/vyvojova-dysfazie.html

[14] van der Lely HKJ, Payne E, McClelland A (2011) An Investigation to Validate the Grammar and Phonology Screening (GAPS) Test to Identify Children with Specific Language Impairment. PLoS ONE 6(7): e22432. doi:10.1371/journal.pone.0022432

[15] C.-N. Anagnostopoulos, T. Iliou, and I. Giannoukos, "Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011," Artificial Intelligence Review, pp. 1–23, 2012.

[16] A. Stuhlsatz, C. Meyer, F. Eyben, T. ZieIke, G. Meier, and B. Schuller, "Deep neural networks for acoustic emotion recognition: Raising the benchmarks," in ICASSP, 2011.

[17] T. Vogt and E. Andre, "Comparing feature sets for acted and spontaneous speech in view of automatic emotion recognition," in ICME, 2005.

[18] A. Batliner, K. Fischer, R. Huber, J. Spilker, and E. Noth, "Desperately seeking emotions: Actors, wizards, and human beings," in Proc. ISCA Workshop on Speech and Emotion, 2000.

[19] A. Kazemzadeh, J. Gibson, J. Li, S. Lee, P. Georgiou, and S. Narayanan, "A sequential bayesian dialog agent for computational ethnography," in Interspeech, 2012.

[20] D. Bone, C.-C. Lee, and S. S. Narayanan, "A robust unsupervised arousal rating framework using prosody with cross-corpora evaluation," in Interspeech, 2012.

[21] E. Mower, M. Mataricc, and S. Narayanan, "A framework for automatic human emotion classification using emotion profiles," Audio, Speech, and Language Processing, IEEE Transactions on, vol. 19, pp. 1057–1070, 2011.

[22] J. McCann and S. Peppe, "Prosody in autism spectrum disorders: A critical review," International Journal of Language & Communication Disorders, vol. 38(4), pp. 325–350, 2003.

[23] J. van Santen, E. Prudhommeaux, L. Black, and M. Mitchell, "Computational prosodic markers for autism," Autism, vol. 14, pp. 215–236, 2010.

[24] D. Bone, M. P. Black, C.-C. Lee, M. E. Williams, P. Levitt, S. Lee, and S. S. Narayanan, "Spontaneous-speech acoustic-prosodic features of children with autism and the interacting psychologist," in Interspeech, 2012.

[25] C. Kaland, E. Krahmer, and M. Swerts, "Contrastive intonation in autism: The effect of speaker- and listener-perspective," in Interspeech, 2012.

[26] R. Lunsford, P. A. Heeman, and J. P. H. van Santen, "Interactions between turn-taking gaps, disfluencies and social obligation," in Interspeech, 2012.

[27] M. Swerts and C. de Bie, "On the assessment of audiovisual cues to speaker confidence by preteens with typical development (TD) and atypical development (AD)," in Interspeech, 2012.

[28] G. Kiss, J. P. van Santen, E. Prudhommeaux, and L. M. Black, "Quantitative analysis of pitch in speech of children with neurodevelopmental disorders," in Interspeech, 2012.

[29] B. Schuller, S. Steidl, A. Batliner, A. Vinciarelli, K. Scherer, F. Ringeval, M. Chetouani, F. Weninger, F. Eyben, E. Marchi, M. Mortillaro, H. Salamin, A. Polychroniou, F. Valente, and S. Kim, "The interspeech 2013 computational paralinguistics challenge: Social signals, conflict, emotion, autism," in Interspeech, 2013.

[30] Schuller B., Steidl S., Batliner A., Burkhardt F., Devillers L. , Müller Ch., Narayanan S., Paralinguistics in speech and language-State-of-the-art and the challenge, Computer Speech and Language, v.27 n.1, p.4-39, January, 2013, ISSN 0885-2308.

[31] Bishop DVM, Adams CV, Norbury CF. Distinct genetic influences on grammar and phonological short-term memory deficits: Evidence from 6-year-old twins. Genes, Brain and Behavior. 2006

[32] Kohnert, K., Windsor, J., & Ebert, K.D. (2009). Primary or "Specific"□ Language Impairment and children learning a second language. Brain and Language, 109

[33] Archibald, L. M., & Gathercole, S. E. (2006). Short-term and working memory in Specific Language Impairment. In Alloway, T. P. & Gathercole, S. E. (Eds), Working memory in neurodevelopmental conditions, pp. 139-160. Psychology Press.

[34] Montgomery JW, Magimairaj BM, and Finney MC (2010) Working memory and specific language impairment: An update on the relation and perspectives on assessment and treatment. American Journal of Speech-Language Pathology 19

[35] Bishop, D. V. M., & Snowling, M. J. (2004). Developmental dyslexia and specific language impairment: Same or different? Psychological Bulletin, 130

[36] van Weerdenburg M, Verhoeven L, Bosman A, van Balkom H. (2011). Predicting word decoding and word spelling development in children with Specific Language Impairment. Journal of Communication Disorders 44(3)

[37] Grela, B., Collisson, B., & Arthur, D. (2013). Language Processing in Children With Language Impairment. In J. Guendouzi, F. Loncke & M. J. Williams (Eds.), The Handbook of Psycholinguistic and Cognitive Processes. New York, NY: Psychology Press.

[38] Clahsen, H. (1989). The grammatical characterization of developmental dysphasia. Linguistics

[39] Gopnik, M., Dalalakis, J., Fukuda, S. E., Fukuda, S., & Kehayia, E. (1997). Genetic language impairment: Unruly grammars. In W. G. Runciman, & J. Maynard (Eds.), Evolution of social behaviour patterns in primates and man: 88. Proceedings of the British Academy. Oxford, UK: Oxford University Press.

[40] Locke, J. L. (1994). Gradual emergence of developmental language disorders. Journal of Speech and Hearing Research, 37,

[41] Kantardzic, M. Data mining: concepts, models, methods, and algorithms. 2nd ed. Hoboken, N.J.: IEEE Press, c2011, xvii, 534 p. ISBN 978-1-118-02913-8.

[42] http://neuroscience.uth.tmc.edu/s4/chapter08.html

[43] Gick B., Wilson I., Derrick D. (2012) Articulatory Phonetics. Wiley. ISBN 9781118438084.

[44] Thmoas W. P. (1986) Voice and Speech Processing. McGraw-Hill Inc. ISBN

[45] Hardcastle W. J. (1976) Physiology of Speech production. Academic press Inc. ISBN

[46] Psutka J, Müller L, Matoušek J, Radová V (2006) We talk Czech with computer (in Czech), Academia Praha, ISBN 80-200-0203 0.

[47] Uhlíř J, Sovka P, Pollák P, Hanžl V, Čmejla R (2007) Technologie hlasových komunikací. Nakladatelství ČVUT, 2007, ISBN 978-80-01-03888-8

[48] B. P. Bogert, J. R. Healy and J. W. Tukey, “The Que-frency Analysis of Time Series for Echoes: Cepstrum, Pseudo-Autocovariance, Cross-Cepstrum, and Saphe Crack- ing,” Proceedings of the Symposium on Time Series Ana- lysis, 1963, pp. 209-243.

[49] R. Mammone, X. Zhang, R. Ramachandran, "Robust Speaker Recognition", IEEE Signal Processing Magazine, September 1996.

[50] Olson, Harry F. (February 1972). "The Measurement of Loudness".Audio: 18–22.

[51] Guimarães, Isabel. A Ciência e a ÜArte da Voz Humana. Escola Superior de Saúde de Alcoitão, 2007.

[52] Smith W S. The Scientist and Engineer's Guide to Digital Signal Processing. In http://www.dspguide.com/pdfbook.htm

[53] H. Hermansky. Perceptual linear predictive (plp) analysis for speech. The Journal of the Acoustical Society of America, 87:1738–1752, 1990.

[54] Waele S D, Broersen P M T. (2000) The burg algorithm for segments. IEEE Trasncsactions on Signal Processing, 48(10), October 2000, 2876–80.

[55] Palková Z (1994) Phonetics and phonologics of the Czech language (book in the Czech) Prague: Karolinum. ISBN 80-7066-843-1.

[56] Press, W.H. Numerical Recipes 3rd Edition: The Art of Scientific Computing, Cambridge University Press, 2007, ISBN 9780521880688

[57] Sharma, A.K. (2005) Text Book Of Biostatistics I. Discovery Publishing House Pvt. Limited ISBN 9788183560306.

[58] Shapiro, S. S.; Wilk, M. B. (1965). "An analysis of variance test for normality (complete samples)". Biometrika 52 (3-4): 591–611. doi:10.1093/biomet/52.3-4.591. JSTOR 2333709. MR 205384

[59] Montgomery, D.C., Runger, G.C. (2010) Applied Statistics and Probability for Engineers. John Wiley & Sons ISBN 9780470053041

[60] Gibbons, J. D., and S. Chakraborti. Nonparametric Statistical Inference, 5th Ed., Boca Raton, FL: Chapman & Hall/CRC Press, Taylor & Francis Group, 2011.

[61] Meloun, M., Militky, J. Statisticka analyza experimentalnich dat (book in the Czech). Praha: ACADEMIA, 953 s. ISBN 80-200-1254-0.

[62] Panik, M. J. Advanced statistics from an elementary point of view. Boston: Elsevier/Academic Press, xvii, 802 p. ISBN 01-208-8494-1.

[63] http://math.tutorvista.com/statistics/quartiles.html

[64] https://onlinecourses.science.psu.edu/stat500/node/38

[65] https://www.princeton.edu/~achaney/tmve/wiki100k/docs/Statistical_dispersion.html

[66] http://www.sagepub.com/sites/default/files/upm-binaries/40007_Chapter8.pdf

[67] Suzuki, K. (2011). Artificial Neural Networks - Methodological Advances and Biomedical Applications. InTech ISBN 978-953-307-243-2

[68] Kröse B.; Smagt P. (1996). An Introduction to Neural Networks, The University of Amsterdam, Amsterdam.

[69] Gurney, K. (1997). An Introduction to Neural Networks, Routledge, ISBN 1-85728-673-1 London

[70] Steve, F., Steve, T. Neural systems engineering J. R. Soc. Interface:20074 193-206;DOI: 10.1098/rsif.2006.0177.Published 22 April 2007

[71] Tučková J (2009) Selected applications of the artificial neural networks at the signal processing. (book in the Czech). Prague: Czech Technical University in Prague, ISBN 978-80-01-04229-8.

[72] Tučková J, Bártů M, Zetocha P (2009) Aplikace umělých neuronových sítí při zpracování signálů. (book in the Czech) Prague: Czech Technical University in Prague. ISBN 978-80-01-04400-1.

[73] Haykin, Simon (1999). "9. Self-organizing maps". Neural networks - A comprehensive foundation (2nd ed.). Prentice-Hall. ISBN 0-13-908385-5.

[74] Kohonen, T. (2001). Self-Organizing Maps. Third, extended edition. Springer.

[75] Mehrotra K., Mohan C.K., Ranka S. (1997) Elements of artificial neural networks. MIT Press ISBN 0262133288

[76] Dozono H. (2012) Applications of Self-Organizing Maps.

[77] Hertz J, Krogh A, Palmer R G (1991) Introduction to the theory of neural computation. Lecture Notes Volume I, Santa Fe Institute, Studies in the sciences of complexity. Addison-Wesley Publishing Company, 1991.

[78] http://blogs.scientificamerican.com/expeditions/2014/10/18/mit-neurotech-mapping-the-brain-with-connectomics/

[79] http://webvision.med.utah.edu/2011/05/something-for-the-cortex/

[80] http://www.wiringthebrain.com/2012/08/are-human-brains-especially-fragile.html

[81] http://rsif.royalsocietypublishing.org/content/royinterface/4/13/193.full.pdf

[82] http://www.apple.com/

[83] http://www.sony.com/

[84] Cool Edit Pro 2 Adobe Systems Inc. Available: http://www.adobe.com/special/products/audition/syntrillium.html. Accessed 11 August 2014

[85] PRAAT: doing phonetics by computer. Available: http://www.fon.hum.uva.nl/praat/. Accessed 11 August 2015

[86] Tučková J, Bártů M, Zetocha P (2009) Aplikace umělých neuronových sítí při zpracování signálů. (lecture notes in the Czech) Prague: Czech Technical University in Prague. ISBN 978-80-01-04400-1.

[87] MATLAB MathWorks Inc. Available: http://www.mathworks.com/products/matlab/. Accessed 11 August 2015

[88] Eyben F, Wöllmer M, Schuller B (2010)"openSMILE - The Munich Versatile and Fast Open-Source Audio Feature Extractor", In Proc. ACM Multimedia (MM), ACM, Florence, Italy, ACM, ISBN 978-1-60558-933-6, pp. 1459-1462, October 2010. doi:10.1145/1873951.1874246

[89] The R Project for Statistical Computing Available: http://www.r-project.org/index.html. Accessed 03 November 2014

[90] Zetocha P (2007) SOM Laboratory for Speech Analysis. ECMS 2007 & Doctoral School 8th International workshop on Electronics, Control, Modelling, Measurement and Signals 2007 & Doctoral School (EDSYS, GEET), May 21-23, 2007, Liberec, Czech Republic.

[91] LANNA - Laboratory of Artificial Neural Network Applications, Department of Circuit Theory at the FEE-CTU in Prague. http://ajatubar.feld.cvut.cz/lanna/

[92] Vesanto J, Himberg J, Alhoniemi E, Parhankangas J (2000) SOM Toolbox for Matlab 5, SOM Toolbox Team, Helsinki Univesity of Technology, Finland, Homepage of SOM Toolbox: www.cis.hut._/projects/somtoolbox

[93] http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html#toc

[94] Thorndike R L, Hagen E P, Sattler J M (1995) Stanford-Binet Intelligence Test (Terman-Merrill) IV. revision. Brno. Psychodiagnostika s.r.o., 1995

[95] Swierkoszová J (1990) Špecifické poruchy učenia. In Lechta,V. a kolektív: Logopedické repetitorium. Bratislava: SPN. ISBN 80-08-00447-9.

[96] An Overview of Cryptographic Hash Functions and Their Uses. Available: http://www.sans.org/reading-room/whitepapers/vpns/overview-cryptographic-hash-functions-879. Accessed 30 June 2014

[97] Komárek V, Kynčl M, Šanda J, Vránová M (2011) Diffusion Tensor Imaging: Ventral and Dorsal Connections between Language Areas in Developmental Dysphasia. In 9th EPNS Congress, May 2011, Dubrovnik, Croatia.

[98] Kantardzic, M. Data mining: concepts, models, methods, and algorithms. 2nd ed. Hoboken, N.J.: IEEE Press, c2011, xvii, 534 p. ISBN 978-1-118-02913-8.

[99] L. B. Leonard and S. Ellis Weismer and C. A. Miller and D. J. Francis and J. B. Tomblin and R. V. Kai (2007) Speed of Processing, Working Memory, and Language Impairment in Children. Journal of Speech Language and Hearing Research. vol. 50, no. 2, pp. 408-428, 2007 DOI: 10.1044/1092-4388(2007/029)

[100] Grill P (2006). Vliv věku na rozložení formantů v řečovém signálu. Bachelor thesis.

[101] Grill P (2011). Automatic detection of formants from speech signal with the use to the classification of children with neurologocal disorder. Doctoral Thesis Statement.

[102] Syrovy V (2014) Hudební akustika (third edition) AMU Press. ISBN 9788073312978

[103] Tučková J, Komárek V (2008) Effectiveness of Speech Analysis by Self-Organizing Maps in Children with Developmental Language Disorders. In: Neuroendocrinology Letters. Ed.: Peter G. Fedor-Freybergh. Society of Integrated Sciences, vol. 29, No.6, Nov/Dec 2008, ISSN 0172-780X.

10.1 List of author’s publications related to the thesis

[104] Grill P, Tuckova J (2014) A database of normal and dysphatic children's speech. In: PLoS One Manuscript draft with manuscript number: PONE-D-14-47426 “Send to press”

[105] Vavřina J, Grill P, Olšanský V, Tučková J (2012) A phoneme classification using PCA and SSOM methods for a children disorder speech analysis. Lékař a technika, 2012, vol.42, no.2, pp.85-88.

[106] Tučková J, Bártů M, Zetocha P, Grill P (2011) Self-Organizing Maps as Data Classifier in Medical Applications. Proceedings of the 3rd International Joint Conference on Computational Intelligence - International Conference on Neural Computation Theory and Applications, Paris, France, 2011. SciTePress , Madeira.

[107] Grill P, Vavřina J, Tučková J (2013) Databases and their applications for diagnosis of developmental dysphasia. Proceedings of the 11th international workshop IEEE, ECMSM 2013, Toulouse, France, 2013

[108] Grill P, Tučková J (2011) Identification the speaker from continual signal by formants analyses. In Proceedings of the 8th International Conference on Digital Technologies 2011 [CD-ROM]. Žilina: Slovenská elektrotechnická společnost, 2011, p. 127-131. ISBN 978-80-554-0437-0.

[109] Grill P, Tučková J (2010) Formant analysisis - vowel detection of children with developmental dysphasia. In Digital Technologies 2010. Žilina: TU v Žilině, 2010, ISBN 978-80-554-0304-5.

[110] Grill P, Tučková J (2008) Formant Analysis - The Appropriate Frequency of Formant Analysis. Proceedings of the 5th International Conference on Digital Technologies 2008, Žilina, Slovakia

[111] Tučková J, Grill P, Zavadil O, Wasserbauer V (2007) Formant analysis – FORANA. In Proceedings of the 4th International Conference on Digital Technologies 2007, Žilina, Slovakia, 2007.

[112] Tučková J, Grill P (2009) Určování formantů pro analýzu emotivně zabarvené řeči . Proceedings of the 8th Czech-Slovak Conference Trends in Biomedical Engineering, Bratislava, Slovensko.

[113] Grill P (2012) Odhad formantů v patologické řeči. In LETNÍ DOKTORANDSKÉ DNY 2012. Praha: ČVUT, 2012, s. 82-85. ISBN 978-80-01-05050-7.

[114] Grill P, Tučková J (2009) FORANA. In Technical Computing Prague 2009. Praha: Humusoft, 2009, s. 32-39. ISBN 978-80-7080-733-0.

[115] Grill P (2012) The database of the children with developmental dysphasia. In BioDat 2012 - Conference on Advanced Methods of Biological Data and Signal Processing. Prague: Czech Technical University in Prague, 2012, vol. 1, p. 32-34. ISBN 978-80-01-05153-5.

[116] Vránová M, Tučková J, Kynčl M, Grill P, Komárek V, et al. (2011) MRI abnormality řečových drah. In Celostátní mezinárodní logopedická konference Tábor, 4. 10. 2011

[117] Vránová M, Grill P (2011) MRI abnormality řečových drah a počítačová analýza promluv dětí s vývojovou dysfázií. Celostátní konference LSMS , Praha, 20. 5. 2011

[118] Grill, P. - Tučková, J.: Formants application to diagnose of children developmental dysphasia. 9th Czech-Slovak Conference Trends in Biomedical Engineering. Ostrava: VŠB, 2011, s. 98-101. ISBN 978-80-248-2475-8.

11 Appendix

11.1 Appendix 1: The Speech Chain

Speech is a natural form of communication for human. Development of speech as of a complex system is a very lengthy process. By age four, most children have developed an ability to communicate through oral language. By age six or seven, are speech organs completely developed and most children can comprehend, as well as express, written thoughts. Major role in speech production have a vocal tract and vocal cords. Articulatory movements, such as the movement of the vocal cords, lips, tongue, and other organs, are among the subtlest and most adept of any actions performed by human beings. The whole mechanism of speech creation is not composed only from articulatory movement, but it is combining the processes that occur in the brain, ears, mouth, lungs, and finally in articulatory organs. The mechanism of speech is described in so called „The Speech Production Chain“ (see Figure 11.1).

Speech is not only a mean for communication, but it is special tool that can be used for a wider understanding of the speaker, e.g. speech conveys not only the meaning but it also expresses the emotion of the speaker and individual information about the speaker. In the following text I focused precisely on this last part. I compiled the analysis of the speech from the perspective of children's speech pathology neurological needs.

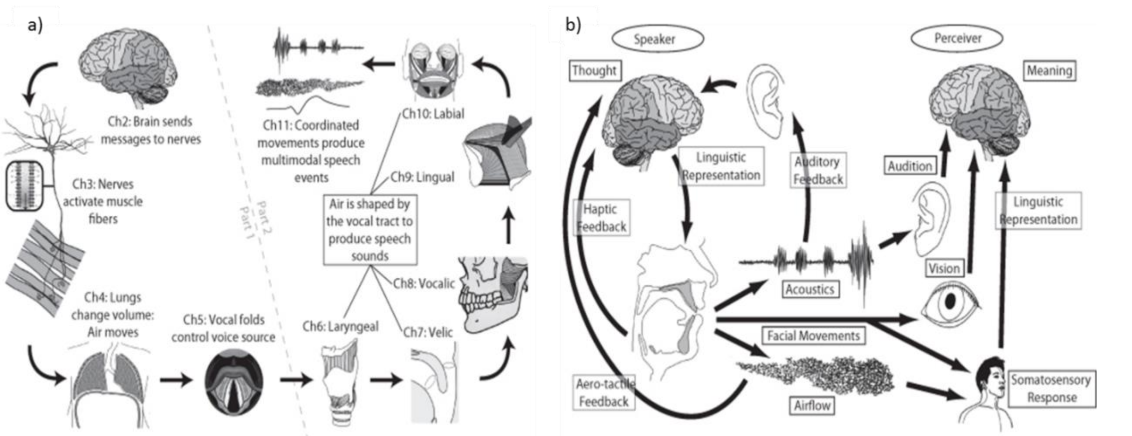


Figure 11.1 Speech chain models (a) Speech production chain, (b) Multimodal speech chain with feedback loops it consists of several states, i.e. linguistic and physiological level (on the speaker side), acoustic level and physiological and linguistic level (according to [43]).

11.2 Appendix 2: Statistical Methods

Descriptive statistics [56], [57], [58], [59], [61], [62], [63], [64], [65], [66] are used to describe the basic features of the data in a study. They provide simple summaries about the sample and the measures. Together with simple graphics analysis, they form the basis of virtually every quantitative analysis of data. Descriptive statistics you are simply describing what is or what the data shows.

Statistical Dispersion

Statistical dispersion is variability or spread in a variable or a probability distribution. Common examples of measures of statistical dispersion are the range, variance, standard deviation and interquartile range.

$$R = x_{\max} - x_{\min} \tag{0.39}$$

Central Tendency

A measure of central tendency is a one single or central value for a probability distribution. It attempts to describe a set of data by identifying the central position within that set of data. The mean, median and mode are all valid measures of central tendency.

Moment

Moment can be characterized as a numerical characteristic of a probability distribution. The first four parameters for a probability model are the mean, the variance, the skewness, and the kurtosis.

- **variance** is denoted as σ^2 and the square root of the variance, σ, is known as the **standard deviation** (both are second moment).

$$\mathrm{Var}(x_1 ... x_N) = \sigma^2 = \frac{1}{N-1}\sum_{j=1}^{N}\left(x_j - \bar{x}\right)^2 \tag{0.40}$$

$$\sigma = \sqrt{\mathrm{Var}(x_1 ... x_N)} \tag{0.41}$$

- **Skewness** (third moment) is characterized as the degree of asymmetry of a distribution around its mean. The number characterizes only the shape of the distribution.

$$Skew(x_1 ... x_N) = \gamma_1 = \frac{1}{N}\sum_{j=1}^{N}\left[\frac{x_j - \bar{x}}{\sigma}\right]^3 \tag{0.42}$$

  $\sigma = \sigma\left(x_1 \ldots x_N\right)$ → is the distribution's standard deviation. A distribution that is skewed to the left/right will have a negative/positive skewness.

- **Kurtosis** (fourth moment) is characterized as the measure of whether the data are peaked or flat relative to a normal distribution. Positive kurtosis indicates a relatively peaked distribution and negative kurtosis indicates a relatively flat distribution.

$$\text{Kurt}(x_1 \ldots x_N) = \gamma_2 = \left\{ \frac{1}{N} \sum_{j=1}^{N} \left[\frac{x_j - \bar{x}}{\sigma} \right]^4 \right\} \tag{0.43}$$

Very often kurtosis is quoted in the form of excess kurtosis (kurtosis relative to normal distribution kurtosis). Excess kurtosis is simply kurtosis less 3.

Quantiles

The quantiles can be defined as the values taken at regular intervals from the inverse of the cumulative distribution function of a random variable. Some *q*-quantiles have special names (2-quantile is median, 4-quantiles are quartiles, 100-quantiles are percentiles, etc.)

- **Percentiles** (P) can be defined as the values that divide the whole series into 100 equal parts.

$$P_k = k\frac{n+1^{th}}{100} \text{ item} \tag{0.44}$$

- **Quartiles** (Q) Just like the percentiles divide the set of observation into 100 equal parts when arranged in the numerical order, in the same way quartile divides the set of observation into 4 equal parts.
    - **first quartile** (Q_1) also called the lower quartile or the 25th percentile:

$$Q_1 = \frac{n+1^{th}}{4} \text{ item} \tag{0.45}$$

    - **second quartile** (Q_2) also called the median or the 50th percentile:

$$Q_2 = \frac{n+1^{th}}{2} \text{ item} \tag{0.46}$$

    - **third quartile** (Q_3) also called the upper quartile or the 75th percentile:

$$Q_3 = 3\frac{n+1^{th}}{4} \tag{0.47}$$

    - **interquartile range** (IQR) is the difference between the upper and lower quartiles:

$$IQR = Q_3 - Q_1 \tag{0.48}$$

Shapiro-Wilk normality test

Hypothesis:

H_0 → The population has a normal distribution

$H_1 \rightarrow$ The population does not have a normal distribution

The formula for the statistic is:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2} \tag{0.49}$$

$x_i \rightarrow$ the ordered random sample values

$a_i \rightarrow$ the Shapiro-Wilk coefficients

Decision:

$W < W_{n\alpha} \rightarrow$ reject the null hypothesis (H_0)

If the value of the test statistic exceeds the critical value $W_{n\alpha}$ tabulated by the *SW* test, the null hypothesis, which is that the population is normally distributed, is confirmed at that significance level. The alternative hypothesis (null hypothesis is rejected) is accepted for data that are not from a normally distributed population. The maximum value of the statistic test is 1 (the random sample is not from a normal distribution function). A *p*-value of $p < 0.05$ was considered statistically significant for the *SW* test. The results obtained at these levels have significant interpretive value.

Student's t-test: Comparison of two means

The requirements for student's test are:

- two independent samples
- data should be normally distributed
- both population standard deviation, σ_1, σ_2, are unknown, but are assumed to be equal

Hypothesis:

$H_0 : \mu_1 = \mu_2 \rightarrow$ no significant difference between the means of the two data sets

$H_0 : \mu_1 \neq \mu_2 \rightarrow$ significant difference between the means of the two data sets

The formula for the statistic is:

$$t = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}} \tag{0.50}$$

$\overline{x}, \overline{y}$ → means

s_x, s_y → standard deviations

n, m → the sample sizes

$$s = \sqrt{\frac{(n-1)s_x^2 + (n-1)s_y^2}{n+m-2}} \tag{0.51}$$

s → the poled standard deviation

The *p*-value is the probability or the area in the tail of the t distribution with the degrees of freedom:

$$df = n + m - 2 \tag{0.52}$$

Decision:

$t > t_{tab}\left(p - value < \alpha\right)$ → reject the null hypothesis (H_0)

$t < t_{tab}\left(p - value > \alpha\right)$ → retain the null hypothesis (H_0)

The t value is compared with the critical value t_{tab} corresponding to the given degrees of freedom $n+m-2$ and a significance level of α. A *p*-value of $p < 0.01$ was considered statistically significant for the Student's *t*-test. The results obtained at these levels have highly significant interpretive value.

Wilcoxon Rank-Sum test

The requirements for test are:

- two independent random samples
- not require normally distributed population
- large sample $n_i \geq 10$η

Hypothesis:

$H_0 : \mu_1 = \mu_2$ → no significant difference between the median of the two data sets

$H_0 : \mu_1 \neq \mu_2$ → significant difference between the median of the two data sets

The formula for the statistic is:

$$z = \frac{R - \mu_R}{\sigma_R} \tag{0.53}$$

$$\mu_R = \frac{n_1\left(n_1 + n_2 + 1\right)}{2} \tag{0.54}$$

$$\sigma_R = \sqrt{\frac{n_1 n_2\left(n_1 + n_2 + 1\right)}{12}} \tag{0.55}$$

R → sum of ranks for smaller sample size (n_1)

n_1 → smaller of sample sizes

n_2 → larger of sample sizes

Decision:

$|z| > z_{tab}$ → reject the null hypothesis (H_0)

The value of z_{tab} represents the corresponding upper critical value of a standard normal distribution. A *p*-value of $p < 0.05$ was considered statistically significant for the *Wilcoxon* Rank-Sum test. The results obtained at these levels have significant interpretive value.

Covariance

The covariance of two variables *x* and *y* in a data sample measures how the two are linearly related. A positive covariance would indicates a positive linear relationship between the variables, and a negative covariance would indicate the opposite.

$$\sigma_{xy} = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \mu_x\right)\left(y_i - \mu_y\right) \tag{0.56}$$

Correlation

The relationship between more than one variable is considered as correlation. It is a statistical method to finding whether two variables are related and to what extent they are related. Correlation is considered as a number which can be used to describe the relationship between two variables. There are several correlation coefficients, often denoted ρ or r, measuring the degree of correlation. Every correlation has two qualities: strength and direction. The direction of a correlation is either positive or negative. In a negative correlation, the variables move in inverse or opposite directions (one increases, the other decreases). In a positive correlation, the variables move in the same direction (both increases). We determine the strength of a relationship between two correlated variables by looking at the correlation coefficient from following formula:

$$-1 \le r \le 1 \tag{0.57}$$

The value of $r = 1$ is for perfect positive correlation and $r = -1$ is for perfect negative correlation and $r = 0$ is for no correlation relationship (see Figure 11.2).

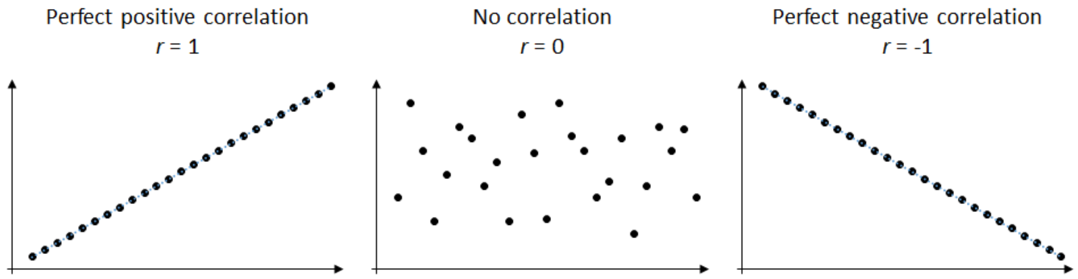


Figure 11.2 Right graph shows a perfect positive correlation, left graph shows perfect negative correlation and middle graph is for situation, when between the data is no apparent relationship between variables.

The strength of correlation is possible generally describe using the following guide for the absolute value of:

$$r \le |0.19| \rightarrow \text{“very weak”}$$

$$r \le |0.39| \rightarrow \text{“weak”}$$

$$r \le |0.59| \rightarrow \text{“moderate”}$$

$$r \le |0.79| \rightarrow \text{“strong”}$$

$$r \le |1.00| \rightarrow \text{“very strong”}$$

The most common of these are the Pearson correlation coefficient and Spearman rank correlation coefficient.

- **Pearson Product-Moment Correlation Coefficient**

(commonly referred to as Pearson's r) is a measure of the strength of a linear association between two variables and is denoted by *r*. If the relationship between the variables is not linear, then the correlation coefficient does not adequately represent the strength of the relationship between the variables. Standard formula is:

$$r = \frac{n\sum xy - \sum(x)(y)}{\sqrt{\left[n\left(\sum x^2\right) - \left(\sum x\right)^2\right]\left[n\left(\sum y^2\right) - \left(\sum y\right)^2\right]}} \tag{0.57}$$

The Pearson correlation coefficient *r* can take a three different states. When r is greater than 0 indicates a positive association, when *r* is less than 0 indicates a negative association and finally when *r* is 0 indicates no association between the two variables. This situation is in following graphs (see Figure 11.3).

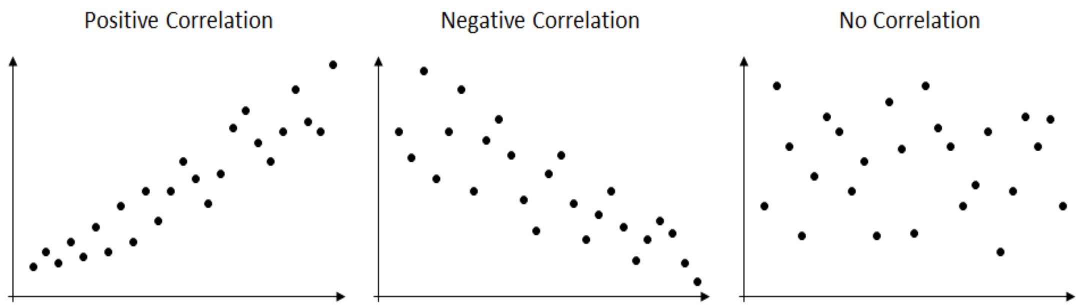


Figure 11.3 The Pearson correlation

- **Spearman Rank Correlation Coefficient**

is the nonparametric version of the Pearson Product-Moment Correlation. Spearman's correlation coefficient measures the strength of association between two ranked variables. The value of each variable are rank-ordered from 1 to *n*, where *n* is the number of pairs of values. Standard formula is:

$$r_s = 1 - \frac{6\sum d_i^2}{n\left(n^2 - 1\right)} \tag{0.58}$$

And its interpretation is similar to that of Pearson. It is necessary to understand a meaning of a monotonic function. It is a function between ordered sets that preserves the given order. Function is always increasing or decreasing, and never changes its direction. The following graphs (Figure 11.4) illustrate difference between monotonic and not monotonic relationships:

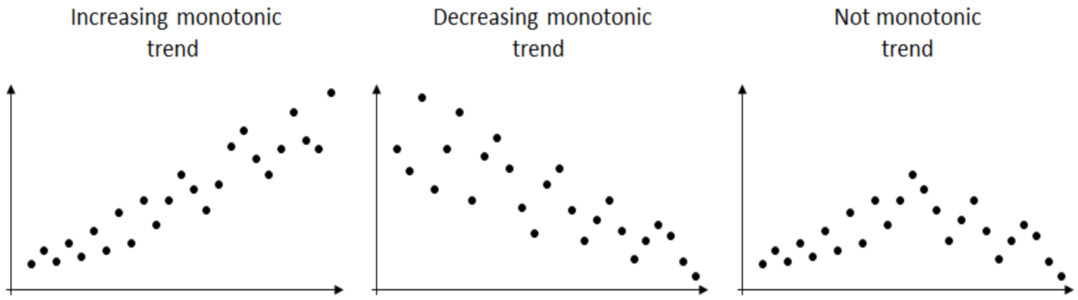


Figure 11.4 Monotonic and not-monotonic relationships. First is for decreasing, second is for increasing and last is for non-monotonic.

Regression analysis

A regression is a statistical analysis that utilizes the relation between two or more quantitative variables so that one variable can predicted from another. Simple linear regression is used to examine the relationship between one dependent and one independent variable. After performing an analysis, the regression statistics can be used to predict the dependent variable when the independent variable is known. In a cause and effect relationship, the independent variable is the cause, and the dependent variable is the effect. Least squares linear regression is a method for predicting the value of a dependent variable *y*, based on the value of an independent variable *x*. Regression goes beyond correlation by adding prediction capabilities.

The simple linear regression function is:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \tag{0.59}$$

Regression coefficient β_1 indicates the change in the mean of the response variable per unit increase in the predictor variable. β_0 and β_1 are unknown population parameters, therefore are estimated from the data (see Figure 11.5).

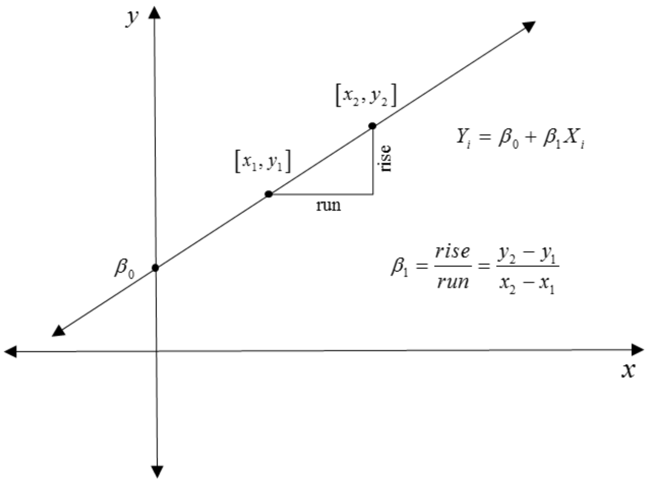


Figure 11.5 Estimation of regression coefficient β_0 and β_1.

Parametres β_0 and β_1 are obtained by finding the values of β_0 and β_1 that minimize the sum of the squared residuals (method of ordinary least squares - OLS):

$$Q = \sum_{i=1}^{n} \left(Y_i - \hat{Y}\right)^2 = \sum_{i=1}^{n} \left(Y_i - a - bX_i\right)^2 \tag{0.60}$$

On finding out the second differential we get the following formulas:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \tag{0.61}$$

$$\hat{\beta}_1 = \frac{\sum\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum\left(X_i - \bar{X}\right)^2} = \frac{Cov(X,Y)}{Var\left(X\right)} \tag{0.62}$$

To measure the strength of the linear relationship we use the coefficient of determination (also called R-squared and is denoted as R^2) by formula:

$$TSS = \sum\left(Y - \bar{Y}\right)^2 \tag{0.63}$$

$$RSS = \sum\left(Y - \hat{Y}\right)^2 \tag{0.64}$$

$$ESS = \sum\left(\hat{Y} - \bar{Y}\right)^2 \tag{0.65}$$

$$TSS = RSS + ESS \tag{0.66}$$

$$R^2 = 1 - \frac{RSS}{TSS} \tag{0.67}$$

TSS $\rightarrow$ total sum of square

RSS $\rightarrow$ residual sum of square

ESS $\rightarrow$ explained sum of square

Value of the strength of the linear relationship is between $0 \leq R^2 \leq 1$.

11.3 Appendix 3: Psychological Tests

The Stanford-Binet test is one of the standard intelligence tests used in clinical pediatric psychology. The fourth revision of the test, published in the Czech Republic in 1995, covers ages 2 to 14 years. In this revision, the test focuses on four areas of evaluation: verbal reasoning, abstract-visual reasoning, quantitative reasoning and short-term memory. The advantage of the fourth revision is the possibility of a relatively reliable evaluation profile and comparison of the levels of the individual skills of the child. Furthermore, the different parts of the test can be used separately. The primary goal of the test is to diagnose children with mental retardation and distinguish mental retardation from specific learning disabilities [94][95].

The terms “special test of perceptual skills”, “graphomotor skills” and “visuomotor coordination” [116] cover the set of psychomotor activities that children exercise when writing and drawing. Those activities focus not only on the coordination and movements of the arms, hands and fingers while grasping writing tools (and their manipulation), but also on constructive skills such as various puzzles. Writing and drawing are not associated with the motoric cortex alone; they are influenced by other centers located in different areas of brain, and the reverse is true, as well. Thanks to these complex influences, the graphomotor skills can help with the diagnosis of psychic conditions and neurological disorders.

11.4 Appendix 4: Formants - Evaluation of FF Algorithm

Results from comparison of the new FF algorithm and classical approach are in the Tables 3 (vowel /a/), 4 (vowel /e/), 5 (vowel /i/), 6 (vowel /o/) and 7 (vowel /u/). Each table contains average values of the formants; i.e. for reference, FF and CA methods and percent differences for all formants of all vowels in comparison FF algorithm and CA method with reference data set.

- **evaluation for vowel /a/**

Vowel /a/						
ID	Method	F1	F2	F3	F4	F5
1	reference	1089	1594	4077	5371	7274
2	FF	1112	1587	4093	5529	7704
3	CA	1100	1594	4065	5401	7308
Evaluation Method of Analysis						
Comparison		F1	F2	F3	F4	F5
1 vs. 2	absolut	23,68	6,76	16,51	158,33	430,17
	percent	2,18%	0,42%	0,40%	2,95%	5,91%
1 vs. 3	absolut	11,06	0,2	12,11	29,64	33,49
	percent	1,02%	0,01%	0,30%	0,55%	0,46%

Table 11.1. Formants for vowel /a/.

- **evaluation for vowel /e/**

Vowel /e/						
ID	Method	F1	F2	F3	F4	F5
1	reference	918	2045	3697	5100	6769
2	FF	922	2066	3887	5254	6934
3	CA	927	2111	3961	5369	7056
Evaluation Method of Analysis						
Comparison		F1	F2	F3	F4	F5
1 vs. 2	absolut	4,08	21,58	189,64	153,1	164,73
	percent	0,44%	1,06%	5,13%	3,00%	2,43%
1 vs. 3	absolut	8,42	65,89	263,43	268,21	286,9
	percent	0,92%	3,22%	7,13%	5,26%	4,24%

Table 11.2 Formants for vowel /e/.

- **evaluation for vowel /i/**

Vowel /i/						
ID	Method	F1	F2	F3	F4	F5
1	reference	462	3931	5450	6585	9876
2	FF	461	3859	5234	6268	9589
3	CA	605	3548	4738	6139	7228
Evaluation Method of Analysis						
Comparison		F1	F2	F3	F4	F5
1 vs. 2	absolut	1,13	72,36	216,06	316,8	287
	percent	0,25%	1,84%	3,96%	4,81%	2,91%
1 vs. 3	absolut	143,27	383,27	711,56	446,51	2647,47
	percent	31,04%	9,75%	13,06%	6,78%	26,81%

Table 11.3 Formants for vowel /i/.

- **evaluation for vowel /o/**

Vowel /o/						
ID	Method	F1	F2	F3	F4	F5
1	reference	651	965	3736	4525	6097
2	FF	652	970	3855	4636	6081
3	CA	746	902	4383	6141	7827
Evaluation Method of Analysis						
Comparison		F1	F2	F3	F4	F5
1 vs. 2	absolut	0,83	5,02	118,9	111,22	16,39
	percent	0,13%	0,52%	3,18%	2,46%	0,27%
1 vs. 3	absolut	94,12	62,8	647,54	1616,48	1730,44
	percent	14,45%	6,51%	17,33%	35,72%	28,38%

Table 11.4 Formants for vowel /o/.

- **evaluation for vowel /u/**

Vowel /u/						
ID	Method	F1	F2	F3	F4	F5
1	reference	474	732	2327	4071	4855
2	FF	471	737	2366	4097	4905
3	CA	538	2026	4822	5885	7979
Evaluation Method of Analysis						
Comparison		F1	F2	F3	F4	F5
1 vs. 2	absolut	3,09	5,2	39,69	26,71	50,18
	percent	0,65%	0,71%	1,71%	0,66%	1,03%
1 vs. 3	absolut	64,34	1294,46	2495,43	1814,45	3123,87
	percent	13,58%	176,93%	107,25%	44,57%	64,35%

Table 11.5 Formants for vowel /u/.