We are IntechOpen, the world's leading publisher of Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Classification and Detection of Specific Language Impairments in Children Based on their Speech Skills

Pavel Grill and Jana Tučková

Additional information is available at the end of the chapter

Abstract

The ability to use the spoken language is one of the most important characteristics in child development. Speech is difficult to replace in real life, although there are several other options for communication. Inabilities to communicate with speech skills can isolate children from society, especially children with specific language impairments. This research study focused on a specific disorder, known as specific language impairment (SLI); in the Czech language, it is specifically known as developmental dysphasia (DD). One major problem is that this disorder is detected at a relatively late age. Early diagnosis is critical for successful speech therapy in children. The current chapter presents several different approaches to solve this issue, including a simple test for detecting this disorder. One approach involves the use of an original iPad application for detecting SLI based on the number of pronunciation errors in utterances. One advantage of this method is its simplicity; anyone can use it, including parents.

Keywords: specific language impairment, developmental dysphasia, pathological children speech, children speech database, artificial neural networks, formants

1. Introduction

Specific language impairment (SLI) [1–4] is a diagnosis in children with disordered or delayed language development without any reason for the disorder or delay. In children with this disorder, there are specific delays in the mastery of language skills without other developmental delays or hearing loss. Other names for this disorder include developmental dysphasia (DD), which it is frequently referred to in the Czech language, as well as language delay or developmental language disorder. Developmental language disorders are among the most

common learning disorders in children. Approximately 5–7% of all children aged 4–12 years old have these disorders [5]. The impact of these disorders in real life is that a child does not have the same speech skills as other children of the same age because his or her speech skills are delayed. Children with SLI fail to acquire their native language properly or completely despite having normal nonverbal intelligence, a lack of hearing problems, a lack of known neurological dysfunctions and a lack of behavioral, emotional or social problems [4]. These experiences can disrupt children's social lives and separate children from their contemporaries, which can create a specific social barrier. There is a relationship between the development of a child's language skills, age and success with treatment.

The determination that SLI includes a significant genetic component was demonstrated in various studies of heritability, for example, in a study of genetic etiology, a study of twins and a study of family evaluations [6]. SLI affected more boys than girls in another study [5]. The manifestation of the disorder primarily occurs in manipulating the linguistic rules of derivation and inflection, resulting in incorrect syntactic structures in their native tongue. Furthermore, there is reduced development of vocabulary at early ages. Usually, the production of language for those with the disorder is worse than their language comprehension. Various difficulties can be present in children with SLI in nonlinguistic cognitive skills, for example, motor ability, mental rotation or executive functions [7]. Other difficulties can be associated with impairments in reading and problems with working memory [8–11]. Many studies evaluate the problem underlying and causing the observed language difficulties. In these studies, theories of language acquisition as well as language representation and processing have been applied [4, 12]. The most frequently listed hypotheses for the causes of SLI are as follows:

(a) Slower linguistic processing despite relatively normal linguistic representation [4, 12];

(b) Normal linguistic and other cognitive skills with later timing in the onset or triggering of language acquisition processes, leading to developmental delay in language acquisition [13]; and

(c) Problems with grammar and specific subgroups of grammar, while cognitive abilities are relatively intact [14, 15].

The Laboratory of Artificial Neural Network Applications (LANNA) [16] at the Czech Technical University in Prague, with the participation of R&D Laboratory at the Military Technical Institute, 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. The project focuses on children with SLI. A partial aim of this project is to obtain data about SLI and speech disorders using automatic utterance analysis by self-organizing neural networks. The goal of this research is to determine the parameters that correspond to correlations across the results generated from diagnostics (from several different specialists, for example, speech therapists and specialists, psychologists, neurologists, and EEG and MRI tractography) and tests. LANNA uses methods based on computer speech analysis to determine whether children have SLI.

2. Methods

2.1. Ethical statement

The research was performed in compliance with ethical standards and was in accordance with the ethical standards of the Ethics Committee of Motol University Hospital in Prague, Czech Republic. The parents of the participants were informed and provided written informed consent for participation in this research.

2.2. Speech databases and participants

To investigate the effects of speech problems on children with SLI, it was necessary to create a speech database. The LANNA research group created the database [17]. The stimulus for its creation came from cooperation with the Department of Paediatric Neurology in the 2nd Faculty of Medicine of Charles University in Prague and Motol University Hospital, which was supported by grants from IGA MZ CR (Science Foundation of the Ministry of Health of the Czech Republic). The database contained three partial databases of speech recordings of the speech from the following different speaker types: H-CH (children without speech disorders), SLI-CH I (children with SLI), and SLI-CH II (children with SLI with three different degrees of diagnosis severity, which include mild, moderate and severe). This classification of degrees was chosen based on the decisions of speech therapists and specialists from Motol Hospital.

A total of 54 native Czech participants with SLI-CH II (hereafter referred to as *"cases"*) consisted of 35 boys and 19 girls, aged 70–131 months (mean age = 96 ± 16.3 months and median age = 94 months). The participants included in the study had to be examined by a clinical psychologist. The examinations were performed in the Department of Pediatric Neurology of the 2nd Faculty of Medicine of Charles University in Prague. The examination lasted all day, and the parents were present during the exam. The participants (children) were subjected to the following tests over one day: the Stanford-Binet Intelligence Test (Fourth Edition) [18]; Gessel Developmental Diagnosis [19], another standardized and specialized test for the Czech language (world differentiation and sound differentiation tests, auditory analysis and synthesis test); special graphomotor and perceptual skills tests; a test for visuomotor coordination; a test of figure drawing and tracing; and, finally, spontaneous talk evaluations [1, 2, 20, 21]. The inclusion criteria were the following: performance intelligence quotient (PIQ) ≥ 70; disturbed phonemic discrimination; and disturbed language at various levels, which included phonologic, syntactic, lexical, semantic and pragmatic levels [22]. The participants were assessed by other specialists. Neurological examinations showed no abnormalities. Motor milestones were within normal ranges. None of the children had hearing impairments. None were receiving antiepileptic medications. No child was diagnosed with a pervasive developmental disorder or other dominating behavioral problem. None of the children had a history of language or other cognitive regression [22].

A total of 44 native Czech participants from the H-CH subgroup (hereafter referred to as the *"controls"*) with no history of neurological and/or communication disorders were recruited

as a control group. There were 35 boys and 19 girls who were 70–124 months old (mean age = 106 ± 15.4 and median = 110 months). None of the controls underwent voice therapy.

All recordings, data and applications were saved on the server of the LANNA research group, and they are available to authorized users or those who have access to the server of the LINDAT/CLARIN Centre for Language Research Infrastructure. The saved data lack identifying information and are free to use, for scientific purposes, on the server of the LINDAT/CLARIN (http://hdl.handle.net/11372/LRT-1597) [23].

2.3. Procedures and speaking tasks

The selected utterances and first seven tasks, with the English translations of the original Czech utterances used in the current research, are listed in **Table 1**. Only words (a total of 38), no phrases or sentences, were chosen for inclusion from all suitable utterances.

Task code	Description	# Patterns	Language	Utterances
[T1]	Vowels	5	Czech	„a – e – i – o – u"
			English	„a – e – i – o – u"
[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]	Two-syllable words	5	Czech	„kolo - pivo - sokol - papír – trdlo"
			English	„wheel - beer - falcon - paper – boob"
[T5]	Three-syllable words	4	Czech	„dědeček - pohádka - pokémon – květina"
			English	„grandfather - fairy tale - Pokemon – flower"
[T6]	Four-syllable words	3	Czech	„motovidlo - televize – popelnice"
			English	„niddy noddy - television – dustbin"
[T7]	Five-syllable words	2	Czech	„různobarevný – mateřídouška"
			English	„varicoloured – thyme"

Table 1. List of the vocal tasks.

Clinical psychologists and speech therapists collaborated on the selection of suitable utterances, and they formulated the test based on their own experience and acknowledged tests. With this test, the participants repeated spoken utterances, which were necessary to ensure the same conditions for all participants because the younger children could not yet read. The structures of the utterances included a range of words and phrases for a total of 68 different utterances. All utterances were previously described [17].

Only the participants and speech therapist were present during the recordings to maintain the participant's attention during the recording. The procedure of recording the participant was as follows. The participant repeated text after the speech therapist. The same conditions were used for both groups of participants (controls and cases). The recording equipment consisted of digital devices, specifically a digital Dictaphone from Sony Corporation (MD SONY MZ-N710) and an iBook laptop computer by Apple Inc., with professional solution software by Avid Technology, Inc. More information about the recordings of the H-CH and SLI-CH II subgroups can be found in a previously published study [17].

2.4. Processing the recordings and the software used

The following programs were used: Cool Edit Pro 2 [24] and Labelling [25, 26]. The Labelling program was used to segment the speech signal. It was written in the MATLAB programming environment as part of the SOMLab [26] programming system, which was developed in the LANNA. Statistics Toolbox in MATLAB [27] and R software were used for statistical computing [28]. The R Project for Statistical Computing is a language and environment for statistical computing and graphics. It is a GNU project that was developed at Bell Laboratories by John Chambers.

3. Error analysis: transcriptional analysis

In this part of the chapter, a new method, called error analysis, is presented to identify cases based on the number of pronunciation errors in the utterances. Pronunciation requires the ability to distinguish the sounds of spoken language via hearing. The cases had a distinctly impaired ability to aurally differentiate phonemes, and they could not distinguish acoustically similar words. These 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. One requirement of pronunciation is the ability to distinguish the sounds of spoken language by hearing. Analysis was performed by comparison of the words pronounced by the cases versus the words pronounced by the controls, and it was focused on the description of errors in individual words. During the research on the cases, their utterances included many more errors than controls. These errors occurred across all age categories (our research included children aged 39–131 months).

3.1. Description

Three matrices, that is, reference matrix *[RM]*, test matrix *[TM]* and confusion matrix *[CM]*, and two parameters of utterance, that is, utterance of speech therapist *[ut_i]* and utterance

of participant $[ut_2]$ (see **Table 1**), comprise the basic input for this method. *RM* is defined as a square reference matrix with *k* parameters. *K* is characterized as the number of phonemes in ut_1 or ut_2 depending on the size, where a larger *K* is more decisive (see the following equations):

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

$$\sum_{i=1}^{k} r\,m_{kk} = k \qquad (2)$$

The number of errors is obtained as a penalty score from comparing the phonemes from ut_1 and ut_2 (see the following equation):

$$PS = wp + up + mp \qquad (3)$$

where *PS* is the penalty score, *wp* is the number of incorrect phonemes, *up* is the number of unspoken phonemes and *mp* is the number of missing phonemes. A detailed description of the error analysis and all algorithms are provided in Ref. [29]. The input data for error analysis were the recorded ut_1 and ut_2, and the output from error analysis was a *PS* of the analyzed ut_2. In simple terms, comparison of the *TM* and *RM* matrices generates the *CM* matrix. The *CM* matrix contains all information about the errors in *ut2*.

3.2. Statistical evaluation and results

Research data were divided into two groups, controls (*p_h*) and cases (*p_sli*). The Shapiro-Wilk test for normality was used to determine that the data were statistically normal. The obtained scores (*p_h:* W = 0.9175, *p-val* = 0.00444; *p_sli:* W = 0.83, *p-val* = 2.28e−06) were too small to confirm the hypothesis that the groups had a normal distribution. The Wilcoxon's rank-sum test is a nonparametric test used as a substitute for the *t*-test. The obtained scores for *p_h* vs. *p_sli* (*w*) were as follows: *p-val* = 1.01e−15, *zval* = -8.3166 and *ranksum* = 963. The *p*-value was less than the significance level of 0.05; therefore, the null hypothesis of equal medians was rejected. There was sufficient evidence in the data to suggest that the controls and cases were not the same at the default 5% significance level, which was sufficient for significant contention. These results could be considered correct, and it could be argued that there was a significant difference in the number of errors in the speech of the cases and controls.

The results of the analyses of utterance errors are displayed in **Figure 1**, which presents all participants included in our current study.

The controls are displayed in red (or at a higher position), and the cases are displayed in blue (or at a lower position) or in grayscale. Pronunciation errors are displayed in the upper graph. A higher value indicates more errors. The cases had a total number of errors in their
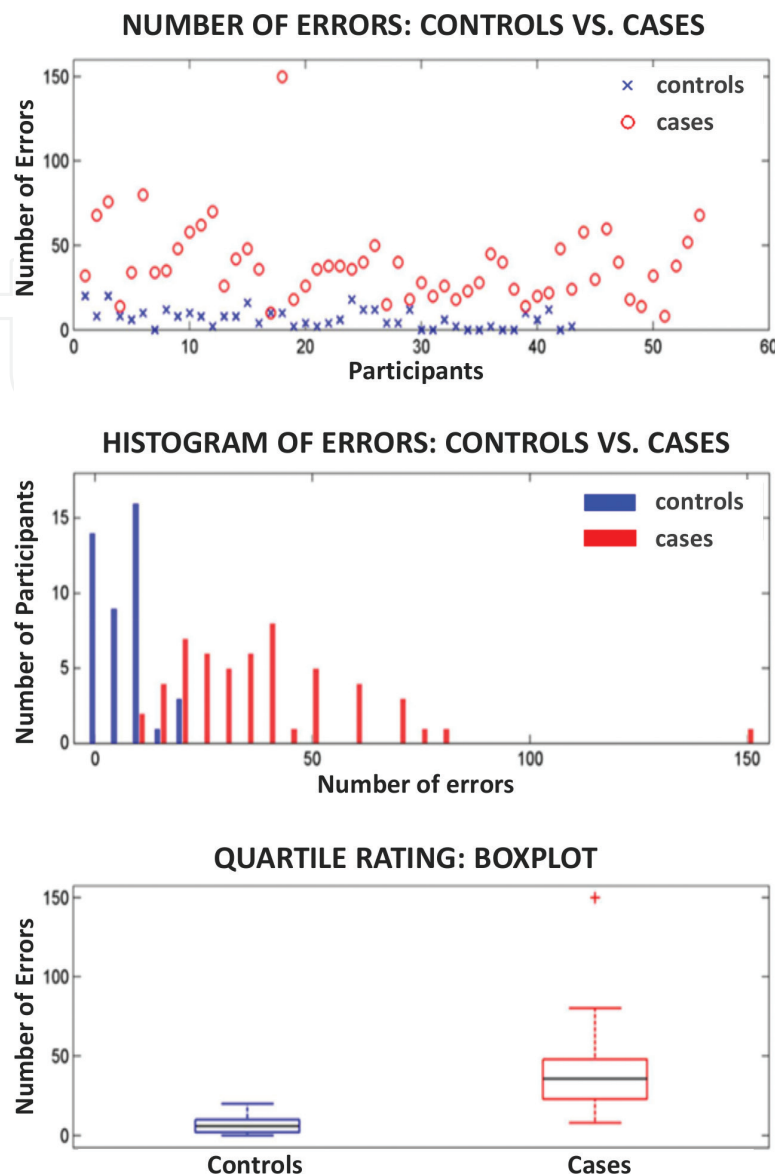
Figure 1. Evaluation of the error analysis. Data from controls are shown in blue (or at a lower position), and data from cases are shown in red (or at a higher position) or in grayscale. Samples with more errors are at a higher position in the upper graph. The histogram of the errors for each participant is shown in the middle graph. Boxplots represent the distributions of the numbers of utterance errors for controls and cases in the bottom chart.

utterances that was much greater than the number of errors for the controls. The distributions of utterance errors of the controls and cases are displayed in the middle graph. The distribution of utterance errors of the controls was clustered around the lower values compared with the distribution of the cases. Box plots representing the distributions of utterance errors of the controls and cases show clear differences between these groups. The cases made more errors in their utterances than controls of the same ages. **Table 2** shows the difference in the average number of errors between controls and cases.

The final evaluation of the error analysis results is shown in **Table 3**. The percentage success rate for the best method, *HM* (*"hand-made"*), was 93.81%. It was necessary to set the limits for each group as the thresholds using the maximum and minimum values from both groups.

Error analysis: controls vs. cases: participants					
Age category	Average error		Difference 2 vs. 1	Comparison	Difference [%]
	Cases (2)	Controls (1)			
All	38.89	4.93	33.96	2 vs. 1	688.84

Table 2. Error analysis: comparison of both groups: average number of errors for controls and for cases.

Error analysis: the success of classification				
Methods	HM (%)	ANN1 (%)	ANN2 (%)	ANN3 (%)
Classification for controls (p_h)	**95.35**	81.40	97.67	97.67
Classification for cases (p_sli)	**92.60**	88.89	83.33	87.04
Final classification	**93.81**	85.14	90.50	92.36

Table 3. Final results for classifiers based on the HM and ANN methods. The method with the highest rate of success in bold.

Classification labeled as misclassification ("*misclass*" in **Figure 2**) indicates the values located outside these limits. As the final criterion for the classification of several words containing an error, the group of cases comprised all children who had more than six words with any error during testing. Self-Organizing Maps (SOMs) [30], a subgroup of an Artificial Neural Network (ANN) [31], was the basis for the other three methods. Parameters for the ANN were set with a standard approach, that is, ratios were 0.7 for training, 0.15 for testing and 0.15 for validation. Differences were observed in the values for weights. ANN1 comprised original default values, ANN2 comprised minimum and maximum values from both groups and ANN3 comprised the weights set to the mean values of these groups. **Figure 2** provides a process diagram for the classification of the error analysis method.

Approximately the same final results are observed for both classifiers, but the HM classifier is easier to implement and use. The results indicate that children with SLI had a greater number of errors in their utterances than typical children.
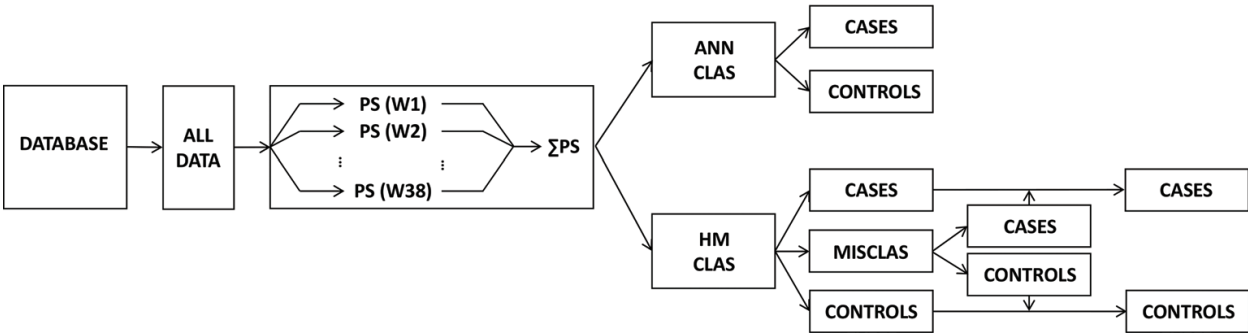
ERROR ANALYSIS: PROCESS DIAGRAM

Figure 2. Process diagram illustrating the principle of error analysis. Overview and comparison of the classification through ANN and HM method.

4. Feature analysis: acoustic analysis

This part of the chapter presents a method, called feature analysis, based on the analysis of acoustic speech features. Children with SLI have a specific problem with the production and perception of spoken language as well as show signs of motor, auditory and phonology difficulties.

Achievements in the recognition of emotions were the inspiration for the use of this method of speech processing. An analogy between the research of emotions and the research of pathological speech, that is, speech of children with SLI, can be made. Typical children, who lack pathological changes in speech and a diagnosis of any disease, can be compared with a neutral emotion. Children with SLI can be compared with some unspecified emotion, for example, anger or fear.

This method is focused on acoustic speech parameters from individual words. Analysis was performed by comparing the words pronounced by the cases versus the words pronounced by the controls. The aim is to identify features that can be used to uniquely identify cases.

4.1. Description

The examined issue, the classification of children with SLI, has the following implementation structure. The implementation can be divided into four parts whose respective key components can be described as follows:

(a) Input data: The data used to identify children with SLI were selected from our speech database, particularly from the H-CH (44 participants from controls) and SLI-CH II (54 participants from cases) subgroups.

(b) Feature extraction: The OpenSMILE toolkit [32] was used to extract acoustic speech parameters. This software pack can produce a wide range of acoustic speech features. The obtained feature set for description of the speech signal contains a total of 1582 acoustic features, that is, 21 statistical functionals used for 34 low-level descriptors (LLDs) and their deltas, which were calculated every 25 ms from the speech signal. The names and numbers of the 34 low-level descriptors as they appear in the output file are as follows: *pcm_loudness (1), mfcc (15), logmelfreqband (8), lspfreq (8), f0finenv (1) and voicingfinalunclipped (1)*. The names of the 21 functionals are as follows: *maxpos, minpos, amean, linregc1, linregc2, linregerra, linregerrq, stddev, skewness, kurtosis, quartile1, quartile2, quartile3, iqr1-2, iqr2-3, iqr1-3, percentile1.0, percentile99.0, pctlrange0-1, upleveltime75 and upleveltime90*. A more detailed description is given in the openSMILE toolkit tutorial [32].

(c) Many features increase the probability of successful classification as well as increase the possibility of calculating redundant or irrelevant data. The procedure of feature selection was as follows: a value of 1 indicates correct classification, and a value of 2 indicates incorrect classification or so-called *"misclassification."* The whole process is provided in **Figure 3**, and more information is provided in a previous study [17]. Finally, those features that have the best classification rate are selected.

(d) Classification: The classification of the children into two groups, controls and cases, was relatively simple. The evaluation of the selected speech features of all words (selected words listed in **Table 1**) in this study was performed. The resulting classification for each participant was evaluated from the winning class based on the number of classifications (i.e., a value of one corresponded to the correct classification, and a value of two corresponded to a misclassification). Two different approaches of feature selection were used (FS: constant and FS: variable) for classification. For the variable type, properties with the highest accuracy rate (with more than a 90% success of classification for each word) were selected. For the constant type of feature selection, participants were classified using 268 features obtained from 38 words. If a constant number of features (with the 30 best parameters for each word) was used, each participant was classified using 760 features. The entire process of identifying the children with SLI is shown in **Figure 4**. Comparison of the approaches of feature selection is shown in **Figure 5**. The success rate of classification of the FS variable is shown in red (or is presented as the top values), and that of the FS constant is shown in blue (or is presented as the bottom values) or by grayscale. The horizontal line in this chart is the critical line for classification success. The x-axis represents all words, and the y-axis represents the success rate as a percentage.

4.2. Statistical evaluation and results

The data were divided into four groups depending on the classification, that is, correct or incorrect classification for controls (p_h) and correct or incorrect classification for cases (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 as follows: for p_h, correct: $W = 0.5969$ and p-$val = 8.965e{-}10$ and wrong: $W = 0.5678$ and p-$val = 3.567e{-}10$; for p_sli, correct: $W = 0.7825$ and p-$val = 1.598e{-}07$ and wrong: $W = 0.7898$ and p-$val = 2.344e{-}07$. These values were too small (p-$val < 0.05$) to use to confirm the hypothesis that the groups have a normal distribution. The

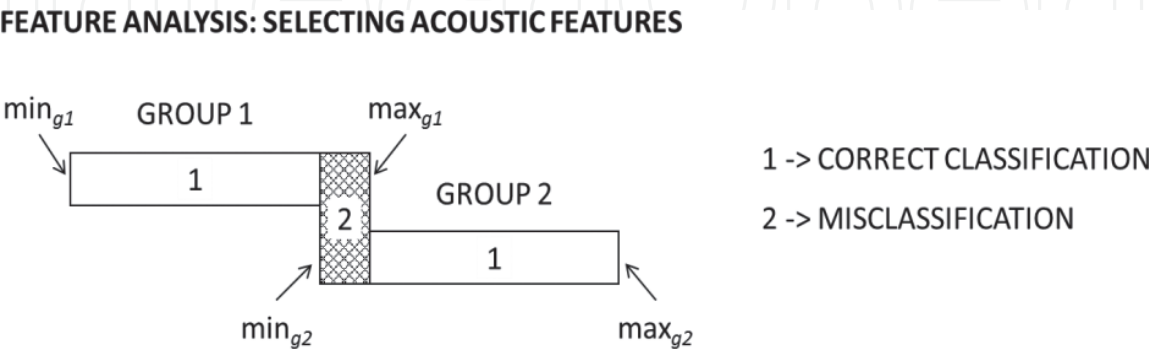**FEATURE ANALYSIS: SELECTING ACOUSTIC FEATURES**

Figure 3. Procedure for selecting appropriate acoustic features. The value of 1 indicates correct classification, and the value of 2 indicates incorrect classification. Parameters "max_{g1}" and "min_{g2}" indicate maximum and minimum threshold limits for incorrect classification.
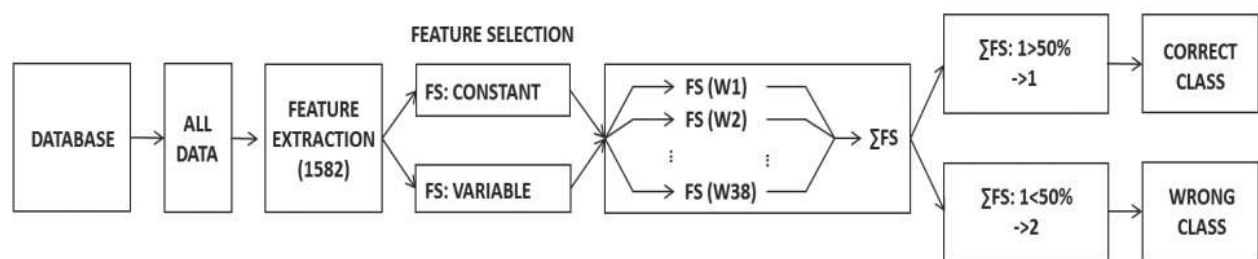
FEATURE ANALYSIS: PROCESS DIAGRAM

Figure 4. Process diagram illustrating the principle of feature analysis. Overview of the classification individual groups.
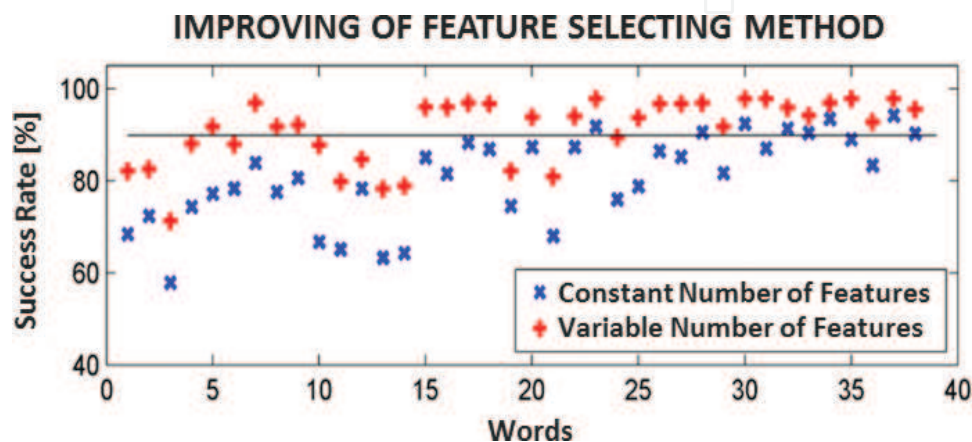

Figure 5. Feature analysis: improving of feature selection method. The success rate of classification of the FS variable is shown in red (or is presented as the top values), and that of the FS constant is shown in blue (or is presented as the bottom values) or by grayscale.

scores of the Wilcoxon rank-sum test, which was used as a substitute for the t-test, are as follows: for p_h, correct vs. wrong: $p\text{-}val = 1.7510e{-}15$, $zval = 7.9578$ and $ranksum = 2911$; for p_sli, correct vs. wrong: $p\text{-}val = 3.3145e{-}19$, $zval = -8.9577$ and $ranksum = 1485$. The null hypothesis of equal medians was rejected because the p-values were too small, that is, a smaller one than the significance level was set, and the values for the group were not the same at this significance level. These results indicate significant differences in the number of classifications between wrong and correct evaluations for controls and cases.

Table 4 presents the final evaluation used to distinguish the two groups, that is, controls vs. cases. The success rate was almost 97%, exactly 96.94%. Three participants (from controls) out of 98 were classified incorrect. Obtained results proved that it is possible to find method based on the acoustic features that can distinguish typically children from children with SLI with high accuracy.

The results of the feature analyses for all participants are displayed in **Figure 6**. Correct classifications of the control group are displayed in blue (or at a higher position), and incorrect classifications of the cases are displayed in red (or at a lower position) or by grayscale. The upper graph showed the total number of classifications where the values in the higher positions indicate a more successful classification. The middle histogram represents the

Feature analysis: evaluation of percent success rate

Age category	Classification of participants			Success rate [%]
	Group	**Correct**	**Wrong**	
All	P_SLI (2)	54	0	100.00
	H-CH (1)	44	41	93.18
	∑(1 + 2)	98	95	**96.94**

Table 4. Evaluation of percent success rate of method based on the acoustic features. The final success rate is in bold.
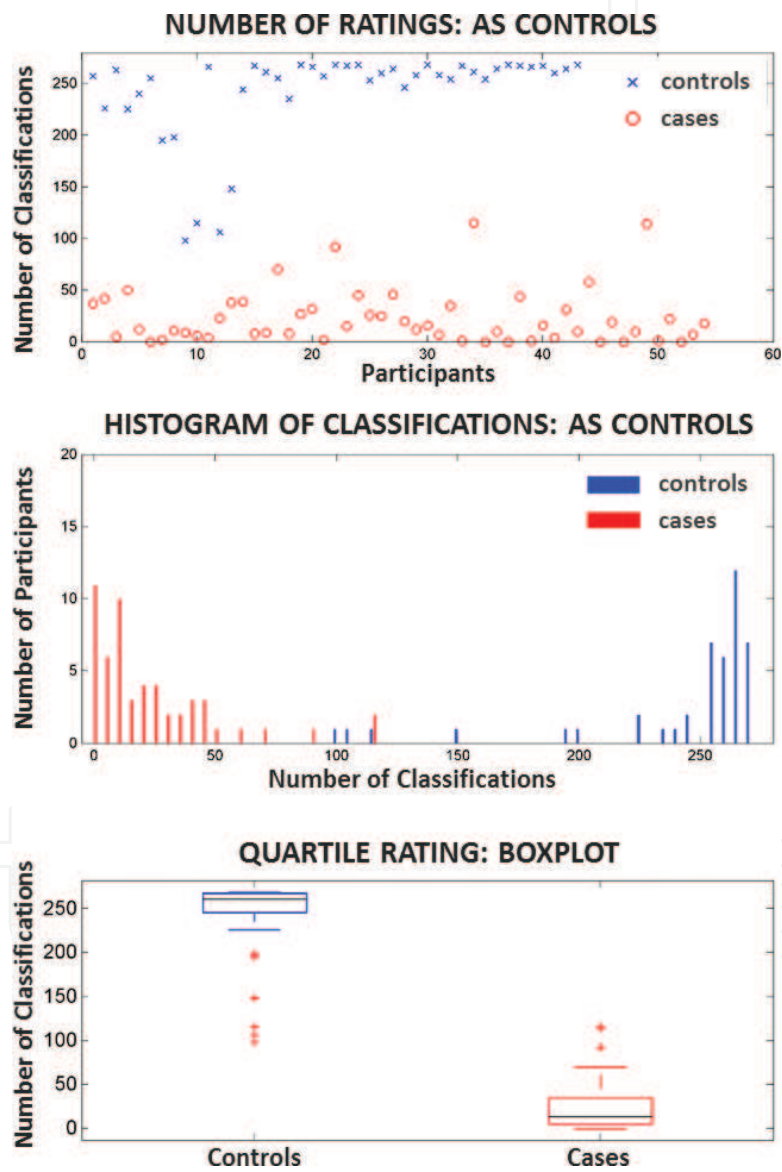

Figure 6. Evaluation of the feature analysis for cases. The correct classification is shown in blue (or at a higher position), and incorrect classification is shown in red (or at a lower position) or in grayscale. Samples with a more successful classification are at a higher position in the upper graph. The histogram represents the distributions of the correct classifications of controls and incorrect classifications of cases (more classifications in the right) in the middle graph. The boxplots show significant differences between the correct (the left boxplot) and incorrect classifications (the right boxplot) in the bottom chart.

distributions of the correct classifications of controls and incorrect classifications of cases. Participants in the higher positions (in the right part of the chart) have more successful classifications. The bottom boxplots show significant differences between the correct (blue or the left boxplot) and incorrect classifications (red or the right boxplot). There was an analogous situation for cases.

5. Time duration analysis

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

The procedure of the experiment was as follows. The average duration was calculated for all words and all participants. Obtained values were divided into two groups. The first group contained the values from controls, and the second group contained the values from cases. Both groups were compared with an average duration of each word.

The evaluation of the time duration is displayed in **Figure 7**. The x-axis represents all words, and the y-axis represents time (s). The time values for cases are displayed in blue (the lower curve), and the time values for controls are displayed in red (the upper curve) or grayscale. **Table 5** illustrates the average duration of all words for both groups. The result is an average duration for cases that is approximately 27.56% higher than that of controls.

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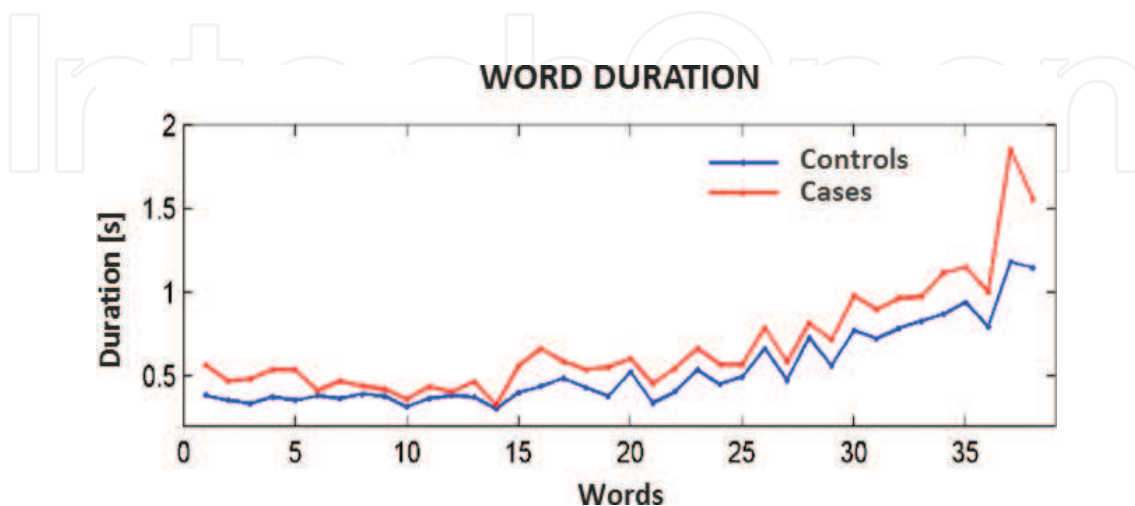

Figure 7. Average duration of words at controls and cases. The time values for cases are displayed in blue (the lower curve), and the time values for controls are displayed in red (the upper curve) or grayscale.

Word duration: controls vs. cases				
ID	Group	Average duration [s]	Comparison	Difference [%]
1	Controls	0.54		
2	Cases	0.69	2 vs. 1	**27.51**

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

Table 5. The success of classification. The final percent rate of difference of the time duration between controls and cases in bold.

6. Formant analysis

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 [20–22]. Formants are normally defined as the spectral peaks of the sound spectrum of the voice (or the concentration 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. A physical dimension of the formants as a classification parameter is based on the presence of an acoustic energy across the speech spectrum, that is, the formants are affected by the movement of the articulatory system based on the human brain activity. This hidden relationship of formants can be used for classifying children with SLI. One of the conditions for using formants as classification parameters is the ability to calculate formants with a minimal error rate. Originally, the extraction of formant frequencies from speech signals was done by using PRAAT [33] acoustic analysis software. However, since the use of the PRAAT software produced formant classification errors in the course of the analysis, the results obtained using this approach could not be treated as relevant (specifically the use of Burg's algorithm to compute formants with method: "*To Formants (burg)…*"). To acquire suitable formants (formants with a minimal error rate), FORANA, a software tool, was developed [34].

Formants provide information about the vowels in the frequency spectrum when the two conditions are fulfilled, that is, the formants must be correctly classified and the utterance must be properly spoken. Especially if we put the first two formants (F1 and F2) into context with each other, we get what we refer to as vocalic triangle. The triangle divides individual vowels into three different categories, depending on the position of the given formant. The first category is represented by the vowel "*a*", the second category is represented by "*e*" and "*i*" and the third category is represented by "*o*" and "*u*" for the Czech speech. The main idea of using formants and vocalic triangle is to verify the correctness of the spoken utterances by using precisely defined vowel locations in the vocalic triangle. Participants from cases (children with SLI) have problems with correctly speaking difficult utterances or words compared with participants from controls (typically children). Formant analysis clearly verifies whether the vowels are correctly pronounced. Otherwise, if there are any errors in the analyzed vowel, there is a shift in the frequency spectrum. This observation means that the speakers have articulatory organs in a bad position and the distribution of articulatory cavities is the wrong shape for forming vowels. This positioning leads to the malfunction of speech control in the brain, which can be used to classify and identify children with SLI. More about this issue can be found in prior studies [35–38].
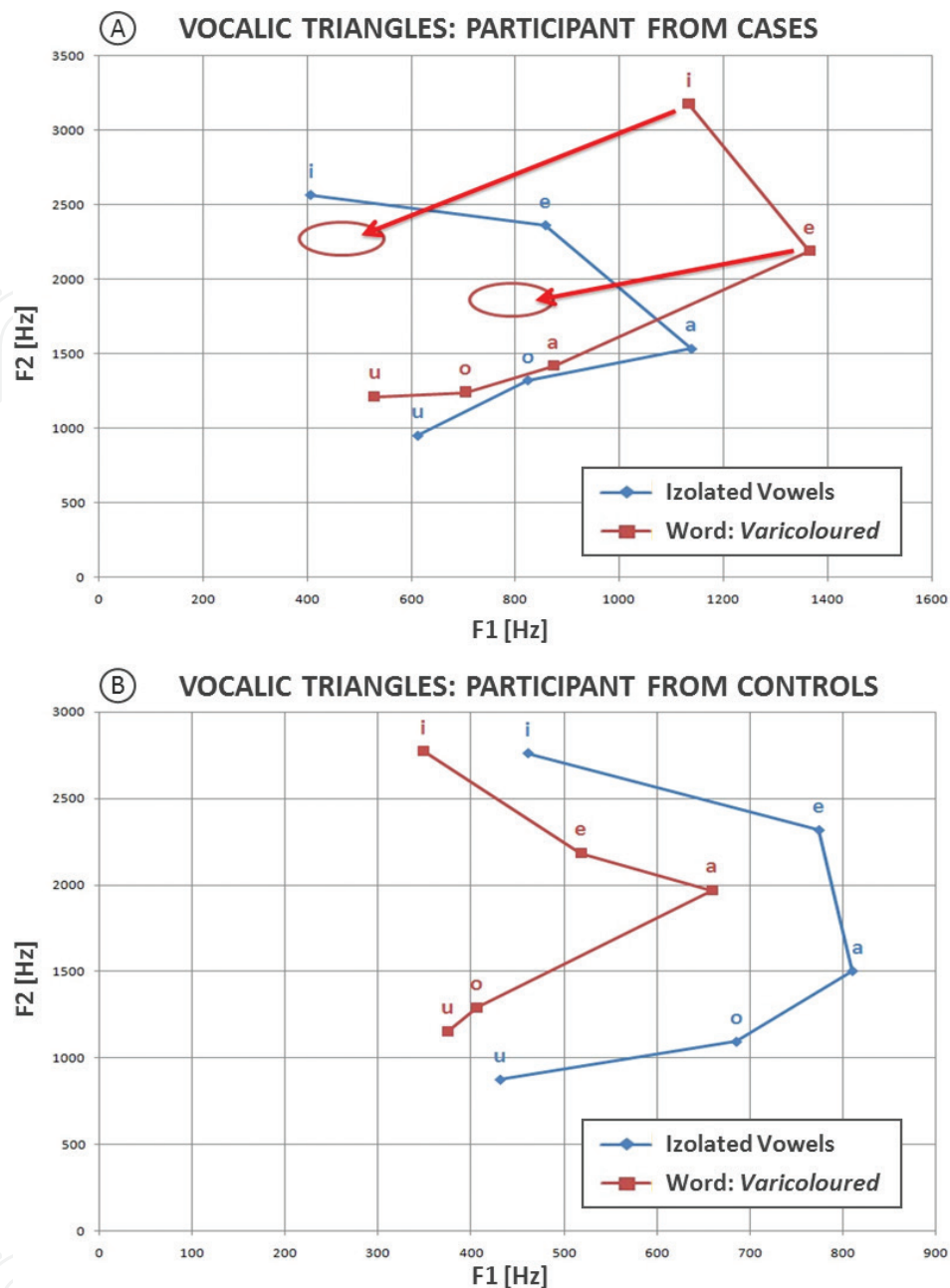
Figure 8. Vocalic triangles, upper chart, obtained from child with SLI and vocalic triangles, lower chart, obtained from typically child. Both children were at age 10 years.

This experiment is based on the comparison of two different vocalic triangles for all tested individuals. Speech signal analysis was performed for the following two types of participants. Participants were chosen randomly, and both were at the same age. One was from cases (from the SLI-CH II group), and the other was from controls (from group H-CH). Both participants were analyzed by using the same utterances, namely isolated vowels and word "*různobarevný*" (in en: "*varicoloured*"). This particular word contains all vowels, and it therefore makes it possible for us to make a comparison between the different vowels. The upper chart (part A in **Figure 8**) represents participant from cases, and the bottom chart (part B in **Figure 8**) represents participant from controls. Both charts show two vocalic triangles, a blue (or the one on the left) one for the isolated vowels and a red (or the one on the right) one for the vowels

in *"různobarevný"* (in en: *"varicoloured"*). The vocalic triangle is presented for simple speech, that is, for isolated vowels; on the other side, the vocalic triangle is absent for more complex speech, that is, for word "různobarevný" (in en: "varicoloured"). The arrows point to the positions where the vowels should be located under ideal circumstances. This corresponds to the situation in the upper chart (part A in **Figure 8**). 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 in children with specific language impairments. Bottom chart (part B in **Figure 8**) shows the vocalic triangles obtained from participant from controls; the triangle is present for both situations (simple speech and speech that is more complex).

Formant analysis: vocalic triangle classification			
Participants	Classification		
	Correct	**Wrong**	**Not**
24	21	1	2
Success rate [%]	87.50	4.16	8.34

Twenty-one participants were classified into the correct class, controls vs. cases, one participant was classified into the wrong class and two participants were not classified (based on real example, two recordings were analyzed for one participant; the result obtained from the first recording was determined as controls, and the result obtained from the second recording was determined as controls).

Table 6. Success rate of method is based on the vocalic triangle classification.

The experiment only involved participants from cases. A total of 24 participants were randomly selected with 54 recordings. Some participants had one speech recording on record, and some had several. The whole experiment was based on the comparison of the two different vocalic triangles, namely isolated vowels (*"a"*, *"e"*, *"i"*, *"o"*, *"u"*) and multisyllabic word *"různobarevný"* (in en: *"varicoloured"*). A prerequisite of this method is the difference in the shape of the vocalic triangles, that is, for isolated vowels, it has the correct shape, while the shape for multisyllabic word is misshapen. The three possible classifications were obtained, that is, correct, wrong and not classified. The results obtained from the vocalic triangle classification method are shown in **Table 6**.

7. SLIt tool

The test of Specific Language Impairments (SLIt tool) is a tablet application that uses a very simple test for identifying children with SLI on an iOS platform (Apple, Inc.), specifically for use on an iPad (iPad third generation or newer) that is based on the procedures used in error analysis. The aim was to create a simple tool that is user-friendly and is easy to use for anyone, for example, parents. Devices such as tablets are light and portable. The test is possible to perform anywhere, for example, at home, instead of only in a specialized clinic.

Figure 9 shows an application SLIt Tool. An application is divided onto four parts. Part 1 contains text from the research for testing children (see **Table 1**). Part 2 contains tools for recording speech. Part 3 contains the corrective mechanisms from error analysis. In the last part (4), a final evaluation of the test is performed. It is possible to view general information about our research on children with SLI and about this application, for example, description of SLI, specification of users, advantages of application and information about the supporting grant.

The test for a child is a very simple, and the course of recordings proceeds the same way as in our research. The procedure is as follows:
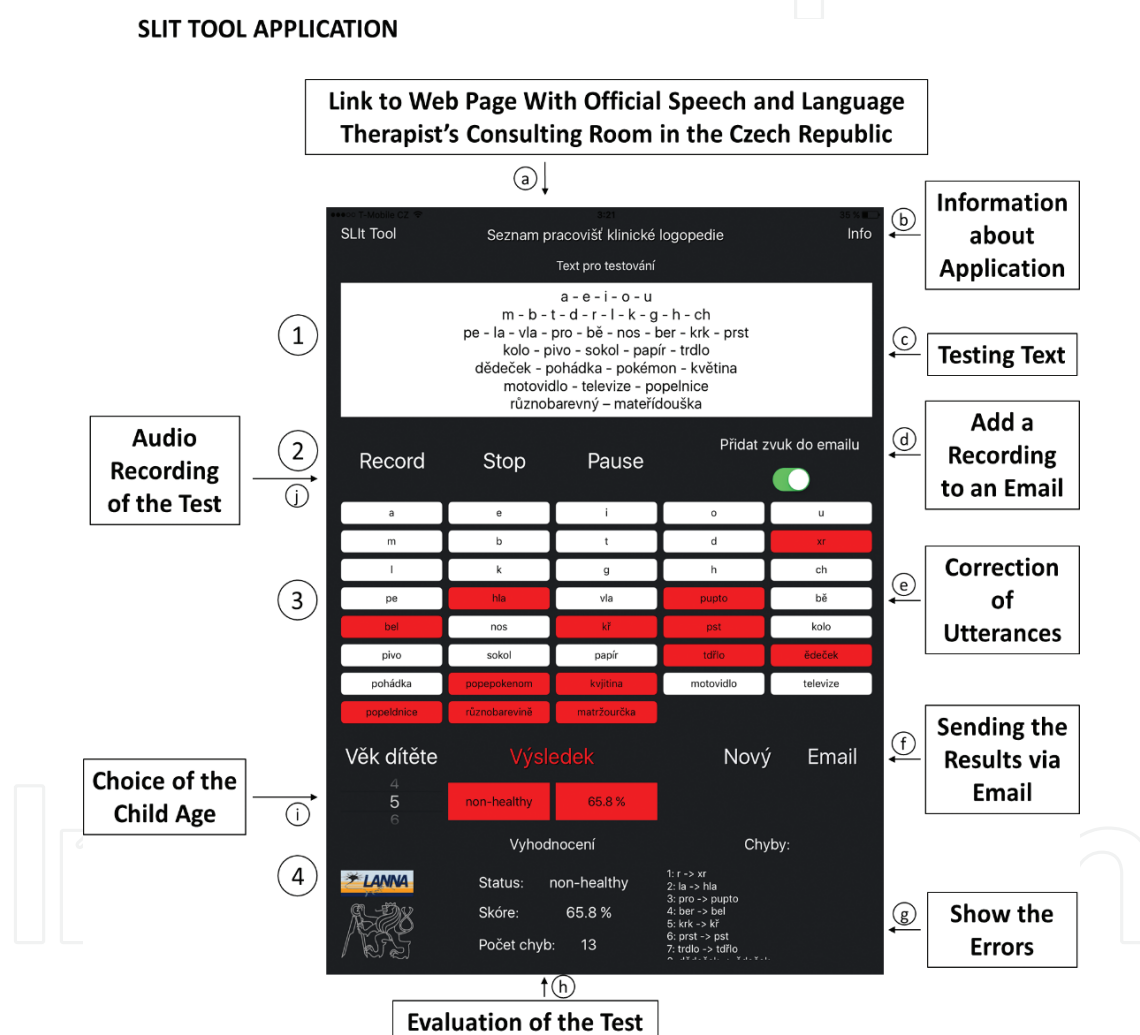

Figure 9. Screen shot of SLIt Tool application on an iOS platform. This figure shows a basic window for testing of the children. Labels inside the rectangular black boxes describe the main functions and possibilities of using this application (description using the letters beside the arrows): (a) link to web pages of the professional association of clinical speech pathologists; (b) complete description and instructions for use an application; (c) text for testing children; (d) the possibility of choosing to send an audio recording of the test via email; (e) the possibility to make corrections in spoken words; (f) the possibility of choosing to send the test via email to speech and language pathologist; (g) all errors of the test are displayed in this part; (h) final evaluation of the test is displayed in this part; (i) choice of child's age and (j) possibility to make an audio recording of the test.

(a) A parent or someone else reads the text, and the child repeats the same text (see **Table 1**).

(b) The child's speech can be recorded for later replay and evaluation.

(c) The text box for the correction of spoken words is pre-filled. The wrong form of a spoken word needs to be replaced, for example, changing the wrong form of the word *"nos"* (in en: *"nose"*) to "*los*" (based on real example).

(d) The final evaluation, test results and recording of the child's speech can be sent to a speech therapist for a more detailed classification.

The application allows for viewing of a list of therapeutic consulting rooms, which are associated with the professional association of clinical speech pathologists in the Czech Republic (AKL CR). Here, it is possible to identify concrete speech and find a language pathologist who can evaluate the results via email. This email with the test evaluation also contains information about the test, the obtained errors, a recommendation based on the final score and an audio recording of test. Audio recordings can be especially beneficial in a comprehensive report on the possible language and speech difficulties of a child. The SLIt Tool is free to use and is available from iTunes.

8. Artificial neural networks analysis

The Supervised Self-Organizing Map (supervised SOM, or SSOM) is based on clustering. These maps and their subsequent visualization help to monitor the progress of trends and magnitude of the degree of impairment. The algorithm of the SSOM represents a very effective classification approach, but it is only effective for well-known input data or for well-known classes of input data. ANNs were selected because of their notable robustness and strong ability to perform data visualization; hence, they can also process less qualitative signals.

8.1. Description of method

This study included 72 controls and 14 cases. The goal was to categorize the subjects into two classes, controls vs. cases. We obtained the results from the speech analysis via vowel mapping with speech from cases by SSOM.

SSOM classification: SSOM was formed by a two-dimensional map with 24 × 24 units. The type of map had a hexagonal grid with a random initialization of the vectors. The following two stages of training were used:

(a) The first stage (rough): The Batch Map algorithm was used with the Gaussian neighborhood function, which decreased monotonically from 24 to 1. The training steps were set to 5000.

(b) The second stage (fine): The Batch Map algorithm was used with the Gaussian neighborhood function, which decreased monotonically from 2 to 0. The training steps were set to 1000.
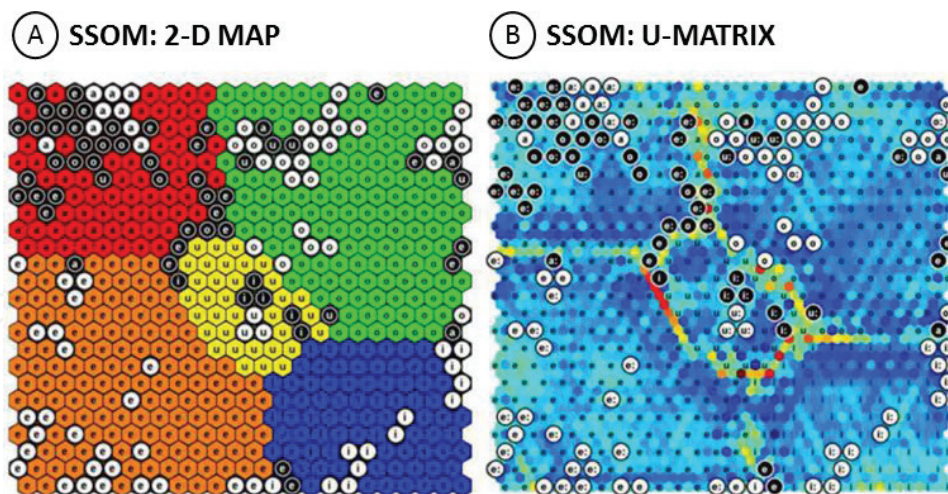
Figure 10. The results of the vowel classification by using SSOM maps. Part A represents a 2-D map, and part B represents a U-matrix.

8.2. Evaluation and results

The training data were set to the dimension of 31,475 × N, where N represents several speech coefficients. The number of wav-files was 1495, and the number of phonemes was 2299. **Figure 10** shows the classification via SSOM trained for vowels for cases. The left panel or part A of the chart represents a 2-D map, and the right panel or part B of the chart represents a U-matrix. These colors or parts of the map represent the vowels in the map; a red color (or the upper left part) represents *"a"*, an orange color (or the lower left part) represents *"e"*, a blue color (or the lower right part) represents *"i"*, a green color (or the upper right part) represents "o" and a yellow color (or the middle part) represents "u".

For the training set, the utterances of all controls and cases were classified with these maps. A white color indicates a successful classification, while a black color indicates a failed classification. For cases, there are characteristic replacements for these vowels, that is, *"o"* behind *"e"* and *"u"* behind *"i"*. These replacements are specific for cases and is not observed in controls. This method obtained a success rate for detecting children with SLI of more than 85%.

9. Conclusion

The methods described in this chapter were developed to analyze disordered speech in children, specifically in children with language impairments. The research was conducted over 10 years. The description is focused on the classification, data collection and data analysis of these children. For analysis, only speech skills of children with SLI were used and compared with typical children. The main benefit of this study includes the methods that were developed to classify children with SLI based on direct database processing. The implementation of these approaches in clinical practice could elucidate the progression and treatment of the disease and facilitate efficient disease treatment.

The first method, called error analysis, is based on the number of pronunciation errors in the utterances. A significant advantage is that its function does not require complex computational methods and can be performed by anyone. The success rate in distinguishing between children with SLI and typical children was 93.81%. The second method, called feature analysis, is based on the auditory signal features that are specific to the acoustic features of speech. These features can easily be obtained and calculated without complicated modifications of the speech signal. The success rate was 96.94%, and only three out of 98 participants were classified as incorrect. The third approach, based on the time duration of utterances, verified the hypotheses on the speed of processing and response for a range of tasks. Children with SLI have a longer duration of words than typical children, that is, the difference was 27.51%. In formant analysis, each vowel has a clearly defined location in the vocalic triangle. The difference between children with SLI and typical children is in the possibility (for typical children) or inability (for children with SLI) to create two vocalic triangles. The vocalic triangle for vowels from a multisyllabic word is misshapen in 87.5% of the analyses of children with SLI. The tablet application SLIt Tool uses an algorithm derived from error analysis, which facilitates the testing of children. The output verifies speech skills with possible consultation about the results via email with a speech and language pathologist.

The obtained results demonstrate that it is possible for children with SLI to be clearly identified and distinguished from typical children. The approach combined traditional and alternative procedures to address this issue and generated a resistance tool that is not dependent on the weaknesses of individual methods.

Acknowledgements

Author details

Pavel Grill[1]* and Jana Tučková[2]

*Address all correspondence to: grillpavel@gmail.com

1 R&D Laboratory, Military Technical Institute, Prague, Czech Republic

2 Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, Czech Republic

References

[1] Bishop DVM. Uncommon Understanding (Classic Edition): Development and Disorders of Language Comprehension in Children. England: Psychological Press; 2014. ISBN: 9780203381472

[2] Leonard LB. Children with Specific Language Impairment. MA: MIT Press; 2014. ISBN: 9780262027069

[3] Hulme C, Snowling MJ. Developmental Disorders of Language Learning and Cognition. Chichester, U.K.; Wiley; 2009. ISBN: 9780631206118

[4] Specific Language Impairment. (2017) Available from: http://www.nidcd.nih.gov/health/voice/pages/specific-language-impairment.aspx [Accessed: February 2, 2017]

[5] Tomblin JB, Records NL, Buckwalter P, Zhang X, Smith E, O'Brien M. Prevalence of specific language impairment in kindergarten children. Journal of Speech, Language, and Hearing Research. 1997;**40**:1245-1260. DOI: 10.1044/jslhr.4006.1245

[6] 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. 2013;**27**(1):4-39. ISSN: 0885-2308

[7] 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. PMID: 16507007. DOI: 10.1111/j.1601-183x.2005.00148.x

[8] Kohnert K, Windsor J, Ebert KD. Primary or/specific/language impairment and children learning a second language. Brain and Language. 2009;**109**(2-3);101-111. DOI: 10.1016/j.bandl.2008.01.009

[9] Archibald LM, Gathercole SE. Short-term and working memory in specific language impairment. International Journal of Language & Communication Disorders. 2006;**41**:675-693. Blackwell Publishing Ltd, U.K.; DOI: 10.1080/13682820500442602

[10] Montgomery JW, Magimairaj BM, Finney MC. Working memory and specific language impairment: An update on the relation and perspectives on assessment and treatment. American Journal of Speech-Language Pathology. 2010;**19**:78-94. DOI: 10.1044/1058-0360

[11] Bishop DVM, Snowling MJ. Developmental dyslexia and specific language impairment: Same or different? Psychological Bulletin. 2004;**130**(6):858-886. DOI: 10.1037/0033-2909.130.6.858

[12] van Weerdenburg M, Verhoeven L, Bosman A, van Balkom H. Predicting word decoding and word spelling development in children with specific language impairment. Journal of Communication Disorders. 2011;**44**(3):392-411. DOI: 10.1016/j.jcomdis.2010.12.002

[13] Locke JL. Gradual emergence of developmental language disorders. Journal of Speech Language and Hearing Research. 1994;**37**(3):608. DOI: 10.1044/jshr.3703.608

[14] Clahsen H. The grammatical characterization of developmental dysphasia. Linguistics. 1989; **27**(5), pp. 897-920. DOI: 10.1515/ling.1989.27.5.897

[15] Gopnik M, Dalalakis J, Fukuda SE, Fukuda S, Kehayia E. Genetic language impairment: Unruly grammars. In: Runciman WG, Maynard J, editors. Evolution of Social Behaviour Patterns in Primates and Man. Proceedings of the British Academy. Vol. 88. Oxford, UK: Oxford University Press; 1996

[16] LANNA. Laboratory of Artificial Neural Network Applications. (2016) Prague: Department of Circuit Theory at the FEE-CTU. Available at: http://amber.feld.cvut.cz/lanna/en/index.asp?page=lanna

[17] Grill P, Tučková J. Speech databases of typical children and children with SLI. PLoS One. 2016;**11**(3):e0150365. DOI: 10.1371/journal.pone.0150365

[18] Knobloch H, Stevens F, Malone AF. Manual for Developmental Diagnosis. The Administration and Interpretation of the Revised Gesell and Amatruda Developmental and Neurologic Examination. Hagerstown: Harper and Row; 1980

[19] Thorndike RL, Hagen EP, Sattler JM. Stanford-Binet Intelligence Test (Terman-Merrill): Fourth Revision. Itasca, IL: Riverside Publishing; 1995

[20] Swierkoszova J. Specific Learning Disabilities (in Slovak). Bratislava: SPN; 1990. ISBN: 80-08-00447-9

[21] Vranova M, Tuckova J, Kyncl M, Grill P, Komarek V, et al. MRI abnormalities of speech and computerised processing of speech of children with developmental dysphasia (in Czech). In: AKL Congress; Tabor, Czech Republic. 2011

[22] Tučková J, Komarek V. Effectiveness of speech analysis by self-organizing maps in children with developmental language disorders. Neuroendocrinology Letters. 2008;**29**(6), pp. 939-948. ISSN: 0172-780X

[23] Tučková J, Grill P, Vavřina J, Bártů M. Speech Databases of Typical Children and Children with SLI. LINDAT/CLARIN Digital Library at Institute of Formal and Applied Linguistics, Charles University in Prague. 2013. Available from: http://hdl.handle.net/11372/LRT-1597

[24] Cool Edit Pro 2 Adobe Systems Inc. (2003). Available at: http://www.adobe.com/special/products/audition/syntrillium.html [Accessed: February 11, 2016]

[25] Tuckova J, Bartu M, Zetocha P. Applications of Artificial Neural Networks for Signal Processing (in Czech). Prague: Czech Technical University in Prague; 2009. ISBN: 978-80-01-04400-1

[26] Zetocha P. SOM laboratory for speech analysis. In: ECMS 2007 & Doctoral School 8th International Workshop on Electronics, Control, Modelling, Measurement and Signals 2007 & Doctoral School (EDSYS, GEET); 21-23 May 2007; Liberec, Czech Republic. 2007

[27] MATLAB MathWorks Inc. (2016). Available at: http://www.mathworks.com/products/matlab/ [Accessed: February 11, 2016]

[28] The R Project for Statistical Computing. (2016). Available at: http://www.r-project.org/index. html [Accessed: February 11, 2016]

[29] Grill P. Specific language impairments and possibilities of classification and detection from children's speech. Prague: Czech Technical University in Prague; 2015. Available from: https://dspace.cvut.cz/bitstream/handle/10467/62582/Disertace_Grill_2015.pdf?sequence=1

[30] Kohonen T. Self-Organizing Maps. Springer Series in Information Sciences. 3rd ed. Berlin, Heidelberg: Springer-Verlag; 2001. ISBN: 3-540-67921-9

[31] Tučková J. Selected Applications of the Artificial Neural Networks at the Signal Processing (in Czech). Prague: Czech Technical University in Prague; 2009. ISBN: 978-80-01-04229-8

[32] Eyben F, Wöllmer M, Schuller B. openSMILE—The Munich versatile and fast open-source audio feature extractor. In: Proceedings of the ACM Multimedia (MM). Florence, Italy: ACM; 2010. pp. 1459-1462. ISBN: 978-1-60558-933-6. DOI: 10.1145/1873951.1874246

[33] PRAAT: Doing Phonetics by Computer. (2016) Available at: http://www.fon.hum.uva.nl/praat/ [Accessed: December 11, 2016]

[34] Grill P, Tučková J. FORANA. In: Technical Computing Prague 2009. Praha: Humusoft; 2009. pp. 32-39. ISBN: 978-80-7080-733-0

[35] Grill P. Vliv věku na rozložení formantů v řečovém signálu [bachelor thesis]. 2006

[36] Grill P. Automatic detection of formants from speech signal with the use to the classification of children with neurological disorder [doctoral thesis statement]. Prague: Czech Technical University in Prague; 2011

[37] Grill P, Tučková J. Formant analysis—Vowel detection of children with developmental dysphasia. In: Digital Technologies 2010. Žilina: TU v Žilině; 2010. ISBN: 978-80-554-0304-5

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