Czech Technical University in Prague
Faculty of Electrical Engineering

Doctoral Thesis

November, 2016

Tomáš Lustyk

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

Analysis of disfluency in pathological speech

Doctoral Thesis

Tomáš Lustyk

Prague, November 24, 2016

PhD Programme: Electrical Engineering and Information Technology
Branch of Study: Electrical Engineering Theory

Supervisor: Doc. Ing. Roman Čmejla, CSc.

Abstract

The thesis focuses on objective evaluation of disfluent speech of people who stutter. Stuttering belongs among the main speech fluency disorders and is characterised by symptoms such as repetitions of sounds, syllables, and words, prolongation of sounds, frequent pauses, revisions, incomplete words. Even though that evaluation of the disorder is mainly performed on the basis of listening to the speech recordings by experienced speech–language pathologist, the nature of this procedure is subjective. Automatic methods relying on speech signal processing could help to current subjective methods by bringing an objective view on speech disfluency. The main objective of the thesis is to analyze disfluent speech by means of objective methods without manual intervention and find out whether these methods can estimate the level of speech disfluency in recordings of different speaking tasks.

The thesis is divided into four studies and the first one concentrates on read recordings of disfluent speech. The study tries to determine whether the proposed measurements are able to describe speech fluency in read recordings. It compares the measurements to the subjective evaluation of speech–language pathologists and attempts to select the most appropriate setting of the algorithms. These measures analyze, for example, the amount of silence or the number of abrupt spectral changes in a speech signal. All the measures were designed to take into account symptoms of speech disfluency. In the study, 118 audio recordings of read speech of Czech native speakers were employed. The results indicate that subjective assessment of disfluency in read speech can be predicted by automatic measurements. The results also imply that there are measures that can describe partial symptoms of speech disfluency (especially fixed postures without audible airflow).

The second study compares automated acoustic measures to behavioral measures of speech fluency in two different speaking conditions (reading and spontaneous monologue) in the speech of people who stutter. The main aim was to investigate the influence of the speaking tasks on participants' fluency levels. Participants were 92 adults (8 control speakers, and 84 stuttering participants). Analysis of read and spontaneous recordings was undertaken by means of two automatic measures selected according to the results on read recordings. The measures were able to estimate the level of speech fluency in both speaking tasks (they correlate to evaluation of speech-language pathologists with -0.83 and -0.52 in the reading and spontaneous task). In particular, the results indicate that speakers with different degrees of disfluency react differently to the speaking conditions. Fluent speakers, and participants with mild and moderate levels of disfluency tend to speak more slowly when performing a spontaneous task compared with reading, while fluency of participants with severe and very severe dysfluency was at the similar or slightly better level in spontaneous condition.

The third study of the thesis describes experiments where automatic acoustic algorithms initially intended to be used on stuttering Czech speakers were applied to recordings of stuttering German speakers. The results suggest that it is basically possible to perform language–independent analysis of disfluent speech. A short (fourth) part outlines an experiment on read recordings with delayed auditory feedback.

All parts of the thesis together form a methodology that could assist to the present methods of speech disfluency assessment by providing an objective instrument to measure the level of speech disfluency in audio recordings of people who stutter.

Abstrakt

Práce zkoumá možnosti objektivního hodnocení neplynulosti řeči mluvčích trpících koktavostí pomocí automatických algoritmů. Koktavost patří mezi poruchy plynulosti řeči, je charakterizována repeticemi (hlásek, slabik, nebo slov), prolongacemi hlásek, četnými pauzami v řeči a dalšími projevy. Hodnocení poruchy neplynulosti je založeno zejména na poslechu nahrávek zkušenými lékaři, i přesto je ale její základ subjektivní. Automatické metody zpracování signálů by mohly pomoci v současnosti používaným metodám tím, že by přinesly objektivní náhled. Proto je hlavním cílem práce analyzovat neplynulou řeč mluvčích s koktavostí pomocí automatických a objektivních metod, a zjistit, zda mohou odhadnout úroveň neplynulosti řeči v různých typech řečových úloh.

Práce je rozdělena na čtyři části. První studie se zaměřuje na neplynulost ve čtených promluvách. Studie zkoumá, zda navržené algoritmy jsou schopny popsat neplynulost řeči ve čtených promluvách, porovnává měření se subjektivním hodnocením lékařů a zkouší nalézt nejvhodnější nastavení algoritmů. Algoritmy analyzují například množství ticha v promluvě nebo počet významných náhlých změn v řečovém signálu. Všechna měření byla navržena tak, aby se zaměřovala na jednotlivé řečové projevy koktavosti. V této studii bylo použito 118 audio nahrávek rodilých mluvčích češtiny. Výsledky ukazují, že automatické metody mohou velmi dobře odhadnout úroveň neplynulosti určenou lékaři ve čtených promluvách. Zároveň se ukazuje, že některá měření velmi dobře popisují jednotlivé symptomy neplynulosti řeči (nejlépe četné pauzy v řeči).

Druhá studie zkoumá efekt řečové úlohy (čtení a spontánní promluva) na neplynulost koktavých. Cílem bylo zjistit možnosti algoritmů ve spontánních promluvách a vliv řečové úlohy na úroveň neplynulosti. V experimentu bylo použito 92 nahrávek (8 kontrolních a 84 koktavých mluvčích), všichni rodilý mluvčí češtiny. Pro analýzu čtených a spontánních nahrávek byla vybrána dvě automatická měření dle jejich výsledků na čtených promluvách a dvě standartní měření rychlosti řeči. Měření jsou schopna odhadnout úroveň neplynulosti v obou řečových úlohách. Korelace s hodnocením lékařů dosahují koeficientu -0.82 pro čtené a -0.52 pro spontánní promluvy. Výsledky ale zejména ukazují, že mluvčí s různým stupněm neplynulosti reagují rozdílně na řečovou úlohu. Plynulí mluvčí, mluvčí s mírnou a střední neplynulostí mluví pomaleji a zhoršují se při spontánní promluvě oproti čtení textu, zatímco mluvčí s vážnou a velmi vážnou neplynulostí zůstávají na stejné úrovni nebo se mírně zlepšují ve spontánní úloze.

Třetí část práce popisuje experimenty, kde byla automatické měření, původně navržená pro české pacienty, použita pro promluvy německy mluvících koktavých. Výsledky naznačují, že je možné provádět analýzu neplynulé řeči nezávisle na jazyku, zde demonstrováno na čtených promluvách německých mluvčích. Krátká čtvrtá část práce nastiňuje experimenty provedené na nahrávkách koktavých při použití zpožděné sluchové vazby.

Všechny části práce tvoří celek, který by mohl pomáhat v současnosti používaným metodám hodnocení neplynulosti tím, že přináší objektivní nástroj pro měření neplynulosti řeči v nahrávkách mluvčích s poruchou koktavosti.

Acknowledgement

I would like to thank my supervisors Doc. Ing. Roman Čmejla, CSc. and Ing. Petr Bergl, PhD, for their guidance during the entire studies, many valuable advices, impulses, and suggestions. Also, I am glad that here I can thank to my colleagues at the room 460. Further, I am really grateful to Dr. Ing. Jan Vokřál from Department of Phoniatrics at the 1st Faculty of Medicine, Charles University and General Faculty Hospital for providing the speech signals and mediation of medical data.

Furthemore, I want to thank Tereza Tykalová, Libor Černy, and Miroslava Hrbková for evaluation of read recordings and to Tereza Hladilová, Jitka Mercelová, Jaroslava Mrkvičková, Barbora Richtrová, and Jan Dezort for assessment of spontaneous recordings.

I also express many thanks to Prof. Dr.-Ing. Elmar Nöth, Dr.-Ing. Tino Haderlein, Dipl.-Inf. Florian Hönig, and other members of Pattern Recognition Lab, Department of Computer Science 5 at the Friedrich–Alexander–University Erlangen–Nürnberg for their supervising, support, hospitality, and company during my abroad studies in Germany.

I would really like to thank Olga Stepankova and Jiri Anyz for their help with the statistical analysis in the study dealing with effect of speaking task.

The research has been supported by the project No. GACR P102/12/2230, and by the Grant Agency of the Czech Technical University in Prague SGS12/185/OHK4/3T/13.

At the end, I would really like to thank my parents for their support and patience and all my friends for their friendship. Thanks to all of you!

Statutory declaration

I hereby declare that my thesis entitled “Analysis of disfluency in pathological speech” is the result of my own work. The thesis directly follows the previous research done by Ing. Petr Bergl, PhD, whose thesis entitled “Objektivizace poruch plynulosti řeči” (Objectification of speech fluency disorders) is the outcome of the mentioned research. Therefore, some similarities can appear in this thesis.

I did not receive any help or support from commercial consultants. All sources and/or materials applied are listed and specified in the thesis. Furthermore, I verify that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form. However, partial results of the research have been submitted and/or published in scientific journals and conferences.

I also agree with the time–unlimited publication of the thesis in electronic form.

Prague, November 24, 2016

signature

Contents

List of abbreviations

ALS	Average Length of Silence
AC	Abrupt Spectral Change
ANOVA	Analysis of Variance
BACD	Bayesian Change–point Detector
DAF	Delayed Auditory Feedback
ESF	Extent of Speech Fluency
F0	Fundamental Frequency
FPWAA	Fixed Postures With Audible Airflow
FPWOAA	Fixed Postures Without Audible Airflow
HMM	Hidden Markov Models
ISR	Incomplete Syllable Repetition
LBDL	Lidcombe Behavioral Data Language of Stuttering
MSUR	Multi—Syllable Unit Repetition
NS	Not Significant
NSI	Number of Spectral Changes in Speech Segments
REV	Regularity of Voicing
ROS	Rate of Speech
RSE	Regularity of Speech Energy
RT	Total Reading Time
SCSI	Average Number of Spectral Changes in Short Intervals
SD	Standard Deviation
SDI11	Standard Deviation of 11 Successive Intervals
SET	Spacing when Exceeding the Threshold
SLP	Speech–Language Pathologist
SNB	Superfluous Nonverbal Behaviors
SR	Syllable Repetition
SVB	Superfluous Verbal Behaviors
VAD	Voice Activity Detection
yr	year

List of symbols

Δ	classification error
$\overline{x}$	mean value
β_0	parametr of general linear model
$\beta_{CONDITION}$	parametr of general linear model
β_{LBDL}	parametr of general linear model
$\beta_{LBDL,\ CONDITION}$	parametr of general linear model
ϵ	residuals in general linear model

List of appendices

Foreword

The thesis presented to obtain the PhD. degree in Electrical Engineering Theory by the Czech Technical University in Prague; the Faculty of Electrical Engineering is largely a result of three studies performed at the Department of Circuit Theory and for one study in close cooperation with Pattern Recognition Lab, Department of Computer Science 5 at the Friedrich–Alexander–University Erlangen–Nürnberg. One article has been published in an international journal, one is currently submitted to an international journal, and one study was presented during the international conference INTERSPEECH 2015.

Articles:

“Evaluation of disfluent speech by means of automatic acoustic measurements”
Lustyk, T., Bergl, P., and Cmejla, R.
Journal of Acoustical Society of America (2014), 135(3), 1457–1468. ISSN 0001-4966,
doi:10.1121/1.4863646

“Comparison between read and spontaneous recordings of disfluent speech by means of objective measures”
Lustyk, T., Bergl, P., and Cmejla, R.
currently submitted

Conference Proceedings:

“Language–independent method for analysis of German stuttering recordings”
Lustyk, T., Bergl, P., Haderlein, T.,Noth, E., and Cmejla, R.
Proceedings of the 16th Annual Conference of the International Speech Communication Association (INTERSPEECH 2015), Bochum: ISCA – International Speech Communication Association, 2015, art. no. 2947, ISSN 2308–457X, 2947–2951.

A complete list of all publications related to the thesis is presented at the end of the manuscript.

Chapter 1

Introduction

Stuttering belongs among the speech fluency disorders. It is characterized by impaired natural fluency of speech production (Conture, 2001). The symptoms predominantly arise in speech: repetition (of sounds, syllables, words, or phrases), prolonged sounds, interjections, revisions, incomplete phrases, and broken words (Bloodstein and Bernstein Ratner, 2008). See Figures 1.1 and 1.2 for examples of stuttering in speech signal. There is also an element of the disorder that influences the psychological and social state of a person who stutters (Kalinowski, 2003; Ezrati-Vinacour and Levin, 2004), and the disease has a notable adverse effect on the quality of life of a person who stutters (Craig *et al.*, 2009).

Although the pathophysiological mechanism of developmental stuttering has not been completely understood, there is agreement that the basal ganglia plays an important role (Alm, 2004; Kubikova *et al.*, 2014).Our incomplete understanding of the mechanism behind stuttering is probably the reason why its definition according to the World Health Organisation is quite general: "Disorders in the rhythm of speech in which the individual knows precisely what he wishes to say but at the time is unable to say because of an involuntary repetition, prolongation, or cessation of a sound" (World Health Organization, 1977).

Stuttering occurs in the whole population regardless of education, economic level, or race. Developmental stuttering typically starts between 2 and 7 years of age with a prevalence of about 5% in preschool children (Yairi and Ambrose, 1999; Mansson, 2000). Symptoms of stuttering in most of the children disappear in the early years (Bloodstein and Bernstein Ratner, 2008). The symptoms persist into adulthood in approximately 1% of the population. The ratio between females and males is estimated to be 1:3 for 2–10 yr olds, and the ratio does not remain stable with age (1:4 for 11–20 yr old, 1:2 between 21–49 yr, and 1:1.4 for the population over age 50) (Craig and Tran, 2005; Bloodstein and Bernstein Ratner, 2008).

Beside developmental stuttering, the neurogenic (or acquired) stuttering can occur later in life as the consequence of a head injury or neurological disease. This type of stuttering exhibits both similarities and differences with developmental stuttering, and some cases seem to be indistinguishable from developmental stuttering (Alm, 2004).

The diagnostic of stuttering is usually based on the judgement of clinical experts. When performing diagnosis, both the speech and also psychological state of a person are taken into account. There exist several stuttering scales regarding the level of speech fluency performance, such as the Stuttering Severity Instrument (Riley, 2009), the Lidcombe Behavioral Data Language of Stutter-

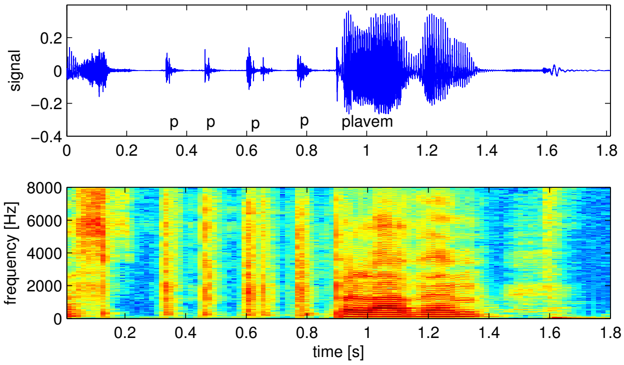


Figure 1.1: Stuttered speech illustration. Repetition of a phoneme /p/ in a Czech word “splavem” (weir).

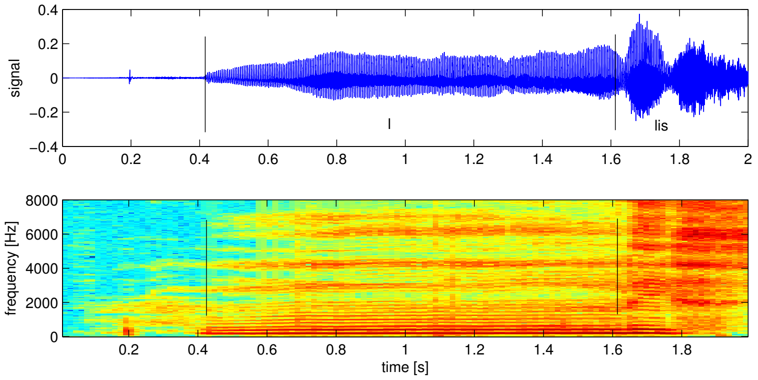


Figure 1.2: Stuttered speech illustration. Prolongation of a phoneme /l/ in a Czech word “listi” (leaves).

ing (Teesson *et al.*, 2003), Kondas scale (Lechta and collective, 2004), or others, for example, scales referred in Manning (2009). However, the evaluation methods are based on subjective assessment.

At the same time, automatic and objective methods can efficiently support the diagnosis of fluency disorders and/or evaluation of therapy outcomes (Van Borsel *et al.*, 2003), which could also save time and effort of the speech–language pathologists (SLPs).

The study follows the previous research done by Ing. Petr Bergl, PhD, at the Faculty of Electrical Engineering, Czech technical University in Prague (Bergl, 2010), who began to develop and design automatic methods for stuttering and applied them to recordings of disfluent speech. The current thesis aims to broaden the previous work by designing additional automatic methods (without manual intervention), testing all the algorithms on read recordings, extending the experiments to spontaneous recordings and evaluating the effect of the speaking tasks. Further, we had an opportunity to try the algorithms initially designed for stuttering Czech speakers on recordings of stuttering German speakers. Therefore, we would like to test the possible language–independence of the algorithms.

Chapter 2

The state of the art

This chapter gives a short state of the art in acoustic methods used in stuttering research and related fields. It mentions the researches that were the most influential on the thesis. As the thesis is divided into three main studies, more articles, thesis, and researches that are related to the subject of the individual study are reported there.

2.1 Objectification of speech disfluencies

The current thesis builds on the previous thesis and research written and carried out by Ing. Petr Bergl, PhD, who defended his doctoral thesis in 2010 under supervision of Doc. Ing. Roman Cmejla, CSc. at the Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague. The outcomes of the research can be found in thesis (Bergl, 2010), and research articles (Bergl, 2006; Bergl and Cmejla, 2007).

The aim of the study was to design algorithms that would be able to estimate automatically and objectively the level of speech fluency disorder. The work is based on the assumption, that there are no manual intervention in the processing of recordings and the recordings are processed as a unit.

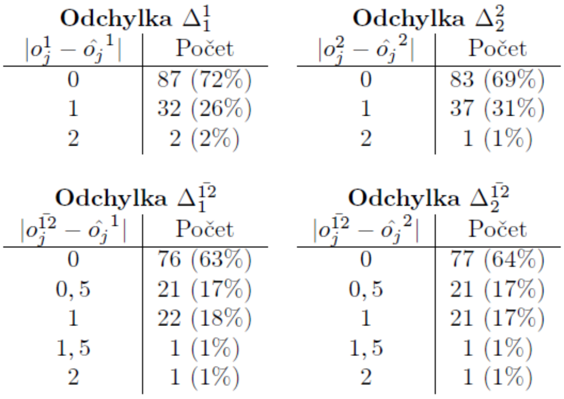
The experiments use 121 recording of read speech. The participants read a 70-word-long text taken from the book Babicka (Grandmother) written by Bozena Nemcova, which is a part of recommended literature for the 6–8th grade of grammar school in Czech Republic (children at the age of 12–14 years). The range of age in the database was from 8 to 49 years. The highest number of participant was at the age between 12 and 15 years. Also, all the levels of speech fluency disorder were represented. Each recording was given an evaluation from two experienced speech fluency experts who assessed the recordings independently, assigning grades from 0 (fluent) to 4 (very severe disfluency).

The thesis presents several automatic and objective measures. They examine the ratio between length of silence and speech parts, number of speech/silence segments, envelope of speech signal energy, regularity of speech energy output, or voicing. Part of measures are based on detection of significant abrupt spectral changes. Two abrupt spectral detectors were used: Bayesian Change–point Detector (BACD), and detector based on GLR (General Likelihood Ratio). The analysis is based on the number of changes made in a recording, and the distances between the changes or their standard deviation. The measurements are independent on the length of the recording. It should guarantee their wide use for different tasks in evaluation of stuttering speech.

The highest correlation achieved by the measures are: the average length of silence with successive removal of short silent/speech parts of signal, 0.793 (evaluation of the first speech fluency expert) and 0.722 (the second clinician), further, the extent of speech fluency (number of significant abrupt spectral changes), -0.782 and -0.783 for the first and second clinical expert, respectively. Also, very good results were reached by the measure derived from speech recognition tool HTK (Hidden Markov Model Toolkit) based on HMM (Hidden Markov Model) with very simple grammar enabling any repetition of sounds. The measure counted the number of transition between sounds (boundaries between phonemes). The correlations were -0.795 and -0.746 for both speech–language pathologists.

The measurements results were also analyzed with the Wilcoxon test to determine whether they are able to distinguish between individual levels of disfluency. The best results were achieved by the measure based spectral changes (HMM, BACD, and GLR) which were able to find significant differences between all classes (significance level 0.05 and 0.01).

At the beginning of the experiment it was hypothesized that different algorithms would correlate to different kinds of stuttering symptoms. This was confirmed and it suggests that the measures, which correlate with different speech symptoms, can be combined to a single and more successful parameter. The parameter composed of a smaller number of measures achieved very good results (correlation up to 0.839), their deviation from subjective evaluation was lower than that of individual measures. For example, one experimental system proposed in the thesis classified about 70% of the participants at the correct fluency level, the classification error higher than two classes was observed only in two participants, see Figure 2.1.



Odchylka Δ_1^1

$\lvert o_j^1 - \hat{o_j}^1 \rvert$	Počet
0	87 (72%)
1	32 (26%)
2	2 (2%)

Odchylka Δ_2^2

$\lvert o_j^2 - \hat{o_j}^2 \rvert$	Počet
0	83 (69%)
1	37 (31%)
2	1 (1%)

Odchylka $\Delta_1^{\bar{1}2}$

$\lvert o_j^{\bar{1}2} - \hat{o_j}^1 \rvert$	Počet
0	76 (63%)
0, 5	21 (17%)
1	22 (18%)
1, 5	1 (1%)
2	1 (1%)

Odchylka $\Delta_2^{\bar{1}2}$

$\lvert o_j^{\bar{1}2} - \hat{o_j}^2 \rvert$	Počet
0	77 (64%)
0, 5	21 (17%)
1	21 (17%)
1, 5	1 (1%)
2	1 (1%)

Figure 2.1: The table shows the number (and percentage) of participant correctly classified by the experimental system composed from individual measures in comparison with subjective evaluations of two speech language pathologists, (Bergl, 2010).

The current thesis and the thesis (Bergl, 2010) were carried out on the part of the database employed in the research (Lastovka *et al.*, 1998). The research applied following parameters to evaluate stuttering: total reading time in seconds, total time of pauses, total speech time, number of speech segments, marginal density of pauses (normalised histograms considering eight intervals), marginal density of speech segments (normalised histograms considering seven intervals), table based on the duration of pause and speech intervals. The best results were achieved when the parameters were classified by discrimination analysis and leave–one–out strategy, 95% of recordings were classified correctly, the rest of the recordings were classified with the error ±1, except for one participant.

2.2 Rate of speech

Rate of speech is a measure which can indicate a speaker's fluency, both in normal and disordered speech. It could be represented in sounds, syllables, or words per minute. Speech rate of stutterers has been extensively studied since the start of modern stuttering research. It has been found that speech rate and pauses are potential perceptual cues for listeners attempting to discriminate the speech of stutterers from that of nonstutterers (Prosek and Runyan, 1982).

2.2.1 Speech rate of stutterers and nonstutterers in different speaking task

One of the most influential research projects is the study made by Johnson (1961). The research examined speech rate and disfluencies of female and male stutterers and nonstutterers. It was carried out to address three issues: 1) to develop procedures of analysis; 2) to obtain normative and comparative data; 3) to compare the achieved result to the other studies.

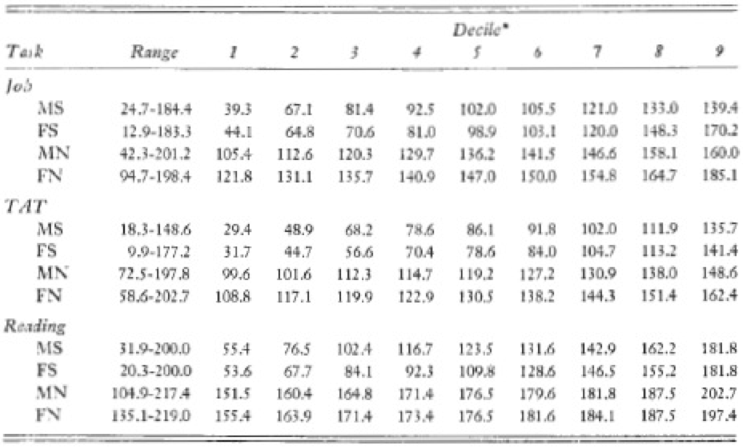
The experiment employed 50 male stutterers, 50 female stutterers, 50 male nonstutterers, and 50 female nonstutterers. Three speaking tasks were considered, spontaneous speech (at least 2 minutes, topic – job), description of a picture (at least 3 minutes), reading a 300–word passage. The female stuttering participants were in the range from 17 to 41, with mean age 21.4, the male stuttering participants were in the mean age of 19.6 years, the range was from 16 to 24 years. The group of normal speakers ranged from 17 to 24 years, mean age of female and male group was 19.3 and 19.2 years, respectively. The majority of speakers were college students.

The following types of disfluencies were considered in the study: interjections of sounds, syllables, words and phrases; part–word repetitions; word repetition; phrase repetitions; revision; incomplete phrases; broken words; and prolonged sounds. Pauses were not considered in the research due to the unsystematic naturel of judging what is a pause or not. The number of words in a recording was also counted and used in combination with the length of the recoding to obtain the rate of speech for each participant and task.

Each recording was assigned a score computed as a number of instances of each type of disfluency. Also, the total score (considering the total number of instances) was calculated. This constitutes the subjective evaluation. Recordings were scored by three investigators with a high level of reliability (> 0.9).

The results of the speech rate study are presented in Figure (table) 2.2. The main findings are: firstly, there are a significant differences in the rate of oral reading and the spontaneous task (monologue, picture description). Secondly, fluent speakers as a group show higher rates, being considerably more fluent, and more consistent than stutterers who reported a higher variability in rate and were less fluent. The study also found an overlap of distribution of disfluency measures (speech rate, number of disfluencies) between both groups in reading and spontaneous speech. That means that some of the speakers who were considered as stutters are more fluent at least in certain aspects than some speakers who are regarded as normal speakers. The research pointed out that there are some types of disfluencies that are considered as stuttering more than others, which are perceived more as normal. The part–word repetition (repetition of sounds and syllables) are perceived by listeners as to be more stuttering.

The overlap of the groups is considered as normal, because fluent speakers usually produce some signs of disfluencies (Roberts *et al.*, 2009), such that the number of normal disfluencies can be even higher than of people with stuttering (Roberts *et al.*, 2009). However, reports by Ingham *et al.* (2012); Pinto *et al.* (2013) confirmed that the fluent speakers generally have a higher speech rate (in syllables/minute) than stuttering participants in both speaking tasks. Simultaneously, they observed that the difference between the speech rate in the read and spontaneous tasks is more obvious in the group of control speakers and the speech rate of stuttering participants is very similar in the oral reading and the monologue.



		Decile*								
Task	*Range*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Job										
MS	24.7-184.4	39.3	67.1	81.4	92.5	102.0	105.5	121.0	133.0	139.4
FS	12.9-183.3	44.1	64.8	70.6	81.0	98.9	103.1	120.0	148.3	170.2
MN	42.3-201.2	105.4	112.6	120.3	129.7	136.2	141.5	146.6	158.1	160.0
FN	94.7-198.4	121.8	131.1	135.7	140.9	147.0	150.0	154.8	164.7	185.1
TAT										
MS	18.3-148.6	29.4	48.9	68.2	78.6	86.1	91.8	102.0	111.9	135.7
FS	9.9-177.2	31.7	44.7	56.6	70.4	78.6	84.0	104.7	113.2	141.4
MN	72.5-197.8	99.6	101.6	112.3	114.7	119.2	127.2	130.9	138.0	148.6
FN	58.6-202.7	108.8	117.1	119.9	122.9	130.5	138.2	144.3	151.4	162.4
Reading										
MS	31.9-200.0	55.4	76.5	102.4	116.7	123.5	131.6	142.9	162.2	181.8
FS	20.3-200.0	53.6	67.7	84.1	92.3	109.8	128.6	146.5	155.2	181.8
MN	104.9-217.4	151.5	160.4	164.8	171.4	176.5	179.6	181.8	187.5	202.7
FN	135.1-219.0	155.4	163.9	171.4	173.4	176.5	181.6	184.1	187.5	197.4

Figure 2.2: The figures in the table show the ranges and deciles of distributions of values for speaking and reading rates in words per minute, for each task for 50 male stutterers (MS), 50 female stutterers (FS), 50 male nonstutterers (MN), and 50 female nonstutterers (FN), (Johnson, 1961).

2.2.2 Relation between speech rate and severity of the fluency disorder

The research (de Andrade *et al.*, 2003) analyzes relation between stuttering severity and speech rate. The study was carried out among native adult Brazilian Portuguese speakers. There was 19 female and 51 male speakers, aged 18 years and older. Each of the recordings was evaluated by Stuttering Severity Instrument (Riley, 1972). Speech samples were obtained during spontaneous speech and contained at least 200 fluent syllables.

Results indicate that the speech rate decreases with the stuttering severity. The more stuttering there was, the smaller number of items per minute was observed. The conclusion applies for both units, syllables and words per minute. The results are presented in Figure 2.3.

Further, the article discusses neurophysiological processing in stuttering and attempts to identify possible subtypes of stuttering. It discusses the hypothesis of an asymmetry in brain activation among stutters and fluent speakers.

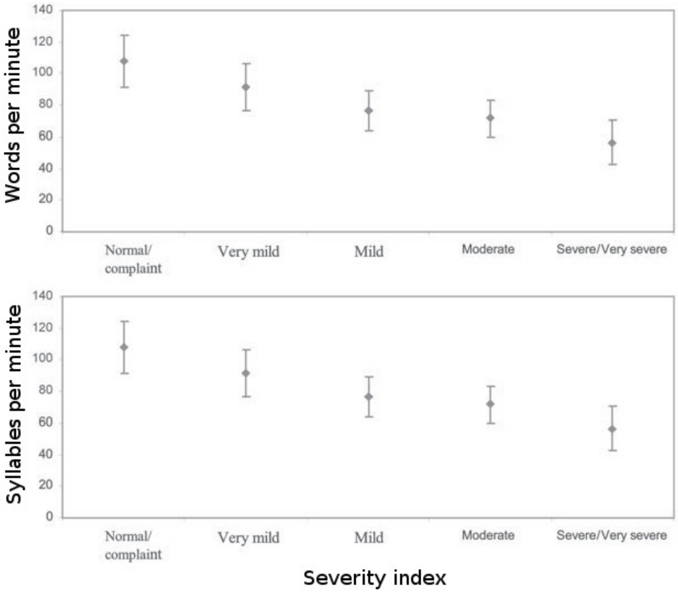


Figure 2.3: Severity index and speech rate (upper part in words per minute, lower part in syllables per minute), (de Andrade *et al.*, 2003).

2.3 Stuttering recognition using Hidden Markov Models

The study (Nöth *et al.*, 2000) is devoted to detection of disfluencies by the HMM. The research employs 16 stutterers and 16 nonstutterers, all German native speakers. The participants read the text "Nordwind und Sonne" (North Wind and Sun). For each recording the number of disfluencies was known.

The speech recognizer (its grammar) was adjusted to disfluent speech. These adjustments were: after each phoneme a silence or filled pause can occur; each phoneme can be repeated; after each phoneme the syllable can be restarted; after each phoneme the word can be restated; and also after each *hot spot* the phrase can be restarted. The *hot spot* is a location where an increased number of disfluencies can arise. These additional rules helps to track stuttered speech.

The results show that the recognizer with adjusted grammar can be successfully applied to stuttered speech. The system reached a correlation of 0.99, see Figure 2.4 in comparison of automatic measure and subjective rating. Also, another parameter, phoneme error rate, was highly correlated to the number of disfluencies (0.95). Another conclusion of the research was that, on average, no significant differences can be found between stutterers and nonstutterers in the duration

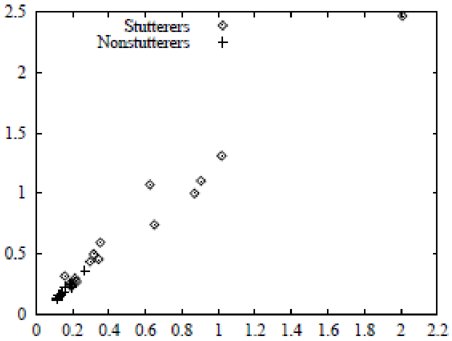


Figure 2.4: Number of disflunecies (counted by therapist, x axis) and hypothesized disfluencies (detected by algorithm, y axis), (Nöth *et al.*, 2000).

of phonemes. The same was found for words. The research also pointed out that pauses in speech play an important role in distinction between stuttered and fluent speech.

Several authors also utilized HMM in stuttered speech research. For example, Wisniewski *et al.* (2007a,b) uses this approach to reveal prolonged fricative phonemes and also blockades with repetition of stop phonemes. The best result for the largest codebook was 80% of successful recognition of prolongations of fricative phonemes. The experiments were performed with a small database and the question remains how the algorithm would perform with a larger database.

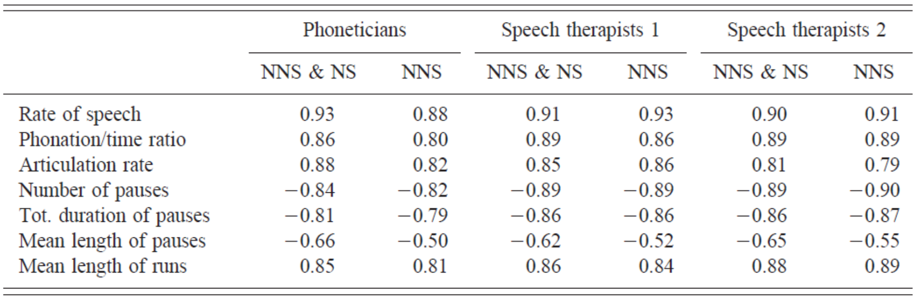
2.4 Studies in other areas of speech research

The studies in different research areas of speech production, which have been influential for this thesis, are introduced in this part of state of the art.

2.4.1 Second language learners' fluency

Cucchiarini *et al.* (2000) investigated the quantitative assessment of second language learners' fluency by means of automatic speech recognition technology in reading. The research was conducted with 20 fluent native and 60 non–native speakers of Dutch. The recordings were subjectively assessed by nine experts. The automatic measures were based on the continuous speech recognizer. The list of measures is specified below:

- Rate of speech (number of phonemes/total duration of speech including sentence–internal pauses)
- Phonation/time ratio (100% x total duration of speech without pauses/total duration of speech including sentence–internal pauses)
- Articulation rate (number of phonemes/total duration of speech without pauses)
- Number of silent pauses (number of sentence–internal pauses of no less than 0.2 s)



	Phoneticians		Speech therapists 1		Speech therapists 2	
	NNS & NS	NNS	NNS & NS	NNS	NNS & NS	NNS
Rate of speech	0.93	0.88	0.91	0.93	0.90	0.91
Phonation/time ratio	0.86	0.80	0.89	0.86	0.89	0.89
Articulation rate	0.88	0.82	0.85	0.86	0.81	0.79
Number of pauses	−0.84	−0.82	−0.89	−0.89	−0.89	−0.90
Tot. duration of pauses	−0.81	−0.79	−0.86	−0.86	−0.86	−0.87
Mean length of pauses	−0.66	−0.50	−0.62	−0.52	−0.65	−0.55
Mean length of runs	0.85	0.81	0.86	0.84	0.88	0.89

Figure 2.5: Correlation among the fluency rating by the three rater groups and the quantitative measures, for the whole group ($n = 80$) and for the non–native only ($n = 60$), NNS – non–native speakers, NS – native speakers, (Cucchiarini *et al.*, 2000).

- Total duration of pauses (total duration of all sentence–internal pauses of no less than 0.2 s)
- Mean length of pauses (mean length of all sentence–internal pauses of no less than 0.2 s)
- Mean length of runs (average number of phonemes occurring between unfilled pauses of no less than 0.2 s)
- Number of filled pauses (number of /uh/, /er/, /mm/, etc.)
- Number of disfluencies (number of repetitions, restarts, repairs)

The results of the research indicate that it is possible to obtain a reliable subjective rating of fluency. Reliability was high for all three groups of experts (Cronbach's alpha varied among 0.9 and 0.96). The automatic measures showed a very high correlations with the subjective rating. Six of nine measures exhibited correlation in range between 0.81 and 0.91, see Figure 2.5 (table) for more details. The rate of speech appeared to be the best predictor of perceived fluency, the correlations with three groups of experts varied among 0.90 and 0.93. Also, all automatic measures are strongly correlated with each other.

The research suggests that there are two important factors in perceiving fluency in read speech, these being the rate at which speakers articulate and the number of pauses they make.

Cucchiarini *et al.* (2002) applied the automatic measures based on continuous speech recognizer (same as in Cucchiarini *et al.* (2000)) for assessment of spontaneous speech in further continuation of research on second language learners' fluency. We can see comparison between results on read and spontaneous recordings in the research. The read part of the database was the same as in the first experiment (Cucchiarini *et al.*, 2000). Recordings of a group of 57 non–native speakers of Dutch constitutes the spontaneous part.

The main findings of the research are: 1) the automatic objective measures can be employed to predict fluency rating, although the predictive power is stronger in read speech than in spontaneous; 2) the objective measures indicate that speakers appear to be less fluent in the spontaneous task than in reading; 3) the nature of the task can influence the performance of the speakers. In addition to point 3, there were two proficiency groups in spontaneous database, beginners and intermediate. Each group carried out different spontaneous tasks. The intermediate group had to perform a longer and more advanced task, requiring a high cognitive load. Conversely, the beginners group performed a shorter task with lower cognitive load. This resulted in the fact that the intermediate group seemed to be less fluent than beginners group.

2.4.2 Speech of patients with Parkinson's disease

Acoustical methods have been also applied as non–invasive biomarkers of Parkinson's disease (Sapir *et al.*, 2010; Rusz *et al.*, 2011). Study (Rusz *et al.*, 2011) examined 46 Czech native speakers, 23 were with untreated early diagnosed Parkinson's disease (PD). Each of the patient underwent eight speech tasks: sustained phonation of /i/, rapid steady /pa/-/ta/-/ka/ syllables repetition, sustained phonation of vowels /a/, /i/, /u/ (all on one breath), reading standard 136–word text, approximately 90 s monologue, reading 8 sentences with varied stress pattern, reading 10 sentences with specific emotions, rhythmical reading of text containing 8 rhymes.

The automatic acoustic measures aimed at one of the speech areas *phonation, articulation*, and *prosody.* The list of measures in various tasks was reduced to 19 measures such as jitter, shimmer, noise to harmonic ratio, harmonic to noise ratio (phonation), rate and regularity in rapid steady repetition, vowel area, relative intensity range variation, robust formant periodicity correlation (articulation), percent pause time, articulation in reading, rhythm measured by dynamical time warping (prosody). The combination of the 19 measures was used as a classification tool of patients with PD.

The main finding is that 78% of subjects with untreated early diagnosed PD show some vocal impairment. The impairment was observed in three subsystems: phonation, articulation, prosody. The impairment occurs not only separately in one subsystem but also frequently in combination with others. Figure 2.6 shows percentage of patients with affected speech subsystems. The most affected subsystem in PD patients was phonation, followed by articulation as the second most affected subsystem.

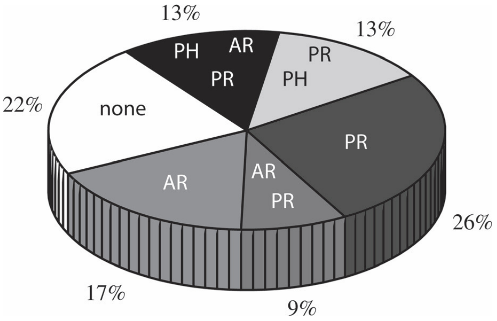


Figure 2.6: Percentage of PD subjects with affected subsystems of speech. PH – phonation, PR – prosody, and AR – articulation, (Rusz *et al.*, 2011).

The effect of the speaking task in patients with PD has been studied on imprecise vowel articulation in Rusz *et al.* (2013). As before, the patients were unmedicated and were cases of early diagnosis. Patients performed four speaking tasks: sustained phonation, sentence repetition, reading passage, and monologue. A group of 20 PD and 15 age–matched healthy control individuals were examined. The vowels /a/, /i/, and /u/ were extracted from recordings of each task for each participant according to strict criteria. The measurements were based upon first (F1) and second (F2) formant frequencies, vowel space area, F2i/F2u, and vowel articulation index (VAI).

The main finding of the study is that the PD subjects manifest abnormalities compared to healthy speakers in vowel articulation through the measurements F2u, vowel space area, F2i/F2u, and VAI. Further, the measurements of imprecise vowel articulation can find differences in various speaking tasks. The vowels extracted from monologue were the most sensitive for detecting abnormalities and difference between controls and PD patients, classification accuracy 80%. Sustained phonation was found to be inappropriate for examining vowel articulation. The measurements were able to find even minor abnormalities in speech. The research also pointed out that a certain type of speaking task can be more useful in the search for difference between healthy and impaired speech. More precisely, complex speech tasks such as monologue can demonstrate speech deficits in the speech of Parkinson's patients with greater frequency than other speaking tasks.

2.5 Objectives and hypothesis

This thesis aims to analyze disfluent speech of patients with stuttering by means of automatic and objective measurements. The base of the thesis is formed as a collage of studies on automatic objective measures of stuttering (Bergl, 2010; Nöth *et al.*, 2000), studies of speech rate in stuttering (Johnson, 1961; de Andrade *et al.*, 2003), research of second language learners' fluency in read and spontaneous recordings (Cucchiarini *et al.*, 2000, 2002), and study of speaking task effect in speech of patients with Parkinson's disease (Rusz *et al.*, 2013).

The main text of the thesis is divided into three studies and one short experiment. The *first* study of the thesis focuses on read recordings of stuttered speech. It serves to introduce all automatic and objective measurements which were designed to describe speeech disfluency and take into account symptoms of stuttering occurring in speech. This part directly follows the previous thesis (Bergl, 2010). In the first study, the main questions was:

- Are the automatic measures able to describe fluency/disfluency in speech of stutterers?

Speech fluency has been studied from different points of view (Johnson, 1961; Nöth *et al.*, 2000; Bergl, 2010) and in different areas of speech research (Cucchiarini *et al.*, 2000). It has been shown that fluency can be predicted by means of automatic measures. Therefore, we suppose that speech fluency of subjects with stuttering can be estimated by automatic and objective measurements.

The *second* study extends the analyses to spontaneous recordings, it examines the effect of the speaking task. It takes the findings of the first study and applies them to spontaneous recordings. The main questions we would like to answer:

- Are the measures able to estimate the level of speech fluency disorder in spontaneous recordings?
- How does stuttering speakers' fluency differ in read and spontaneous speaking tasks?
- Does the level of speech fluency play a role in speaking task effect?

The hypothesis is that the algorithms are able to describe the level of speech disfluency in spontaneous recordings of stutterers, however, we can expect that the results are better in read recordings than in spontaneous (Cucchiarini *et al.*, 2002). The second question refers to the effect of the speaking task in stuttering. There exist several studies of speaking task effect (for example a very extensive study (Johnson, 1961)), but there are only two of them which consider the division of stuttering subjects into levels (Vanryckeghem *et al.*, 1999; Blomgren and Goberman, 2008). Therefore, we would like to broaden information about the speaking task effect with respect to levels of speech fluency. Our assumption is that the control group as well as stuttering speakers will be less fluent and speak at a slower speech rate in the spontaneous task and the difference between individual groups of disfluency will fade. In the second part of the question, we hypothesized that subjects across disfluency levels will be generally influenced to an equal extent. The study could be compared to the research studies as Rusz *et al.* (2013) who studied effect of speaking task in patients with Parkinson's disease or study of subjects with Huntington disease (HD) (Rusz *et al.*, 2014). It also serves as a validation of the results from the read recordings.

The *third* part of the thesis applies previously obtained results and presents results of measures on recordings of German native stuttering speakers. The measurements were designed to take into account symptoms of stuttering and disfluent speech, they were originally proposed to be used for Czech stuttering speakers but they do not take the specifics of the Czech language into account. Therefore, the question we would like to answer is:

- Are some measures able to describe level of speech disfluency in different languages?

Because the majority of measures were designed in a way that does not take the specifics of Czech language into account, they are not based either on dictionary or grammar, we can assume that there would be an agreement between the measures and subjective evaluation of speech disfluency in another language, in this German. However, there remains question how strong the agreements are. The results of the measurements are compared to measurements aimed for German language with stuttering–adapted grammar (Nöth *et al.*, 2000), since we had an opportunity to work on the same databases of read recordings.

The short *fourth* part of the thesis concentrates on read recordings with Delayed Auditory Feedback (DAF). We would like to find out whether it would be possible to use the objective measurements in setting appropriate delay for the DAF device in individual patients.

Chapter 3

Methods

This part of the thesis introduces two subjective evaluation scales used to assess the recordings by speech–language pathologist (SLP) and the automatic measures of speech disfluency. More detailed description of applied methods (participants, recordings, subjective rating) is given in individual studies.

3.1 Subjective rating of recordings

A reliable expert rating is essential to verify whether a measure is suitable for evaluation (Cordes and Ingham, 1994). Therefore we decided to use two evaluation scales to get more insight on participants' disfluency. The first one is the modified Kondas scale (Lechta and collective, 2004). The second set of expert ratings was produced by means of the Lidcombe Behavioral Data Language of Stuttering (LBDL) (Teesson *et al.*, 2003).

3.1.1 Kondas scale

The modified Kondas scale is a standard system used by Czech speech therapists for rating stuttering (Lechta and collective, 2004). The discrete scale consists of 5 stages (from 0 to 4):

- 0 is normal healthy speech (without frequent signs of disfluency),
- 1 is mild disfluency (up to 5% disfluent words),
- 2 is moderate disfluency (6%–20% disfluent words),
- 3 is severe disfluency (20%–60% disfluent words),
- 4 is very severe disfluency (more than 60% disfluent words).

The distribution of Kondas scale is rather logarithmic (Bergl, 2010; Cmejla *et al.*, 2013), see Figure 3.1.

The information for speech specialist who evaluated recordings and also the form are given in section Appendices E and F (in Czech).

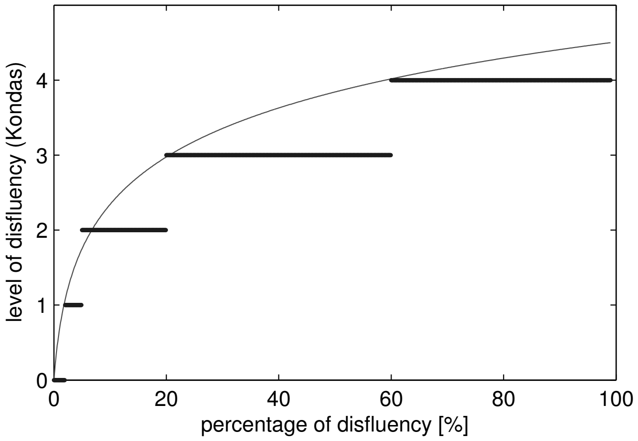


Figure 3.1: Illustration of the Kondas scale, *red* curve represents logarithm of the number of disfluencies, *blue* represents the level of speech disfluency on the Kondas scale, (Bergl, 2010; Cmejla *et al.*, 2013).

3.1.2 The Lidcombe Behavioral Data Language of Stuttering

The second set of expert ratings was produced by means of the Lidcombe Behavioral Data Language of Stuttering (LBDL) (Teesson *et al.*, 2003). We adopted the taxonomy in the experiment because its results are valid and reliable when the system is used by experienced judges (Teesson *et al.*, 2003). The LBDL considers seven descriptors of stuttering symptoms:

- Syllable repetition (SR),
- Incomplete syllable repetition (ISR),
- Multisyllable unit repetition (MSUR),
- Fixed posture with audible airflow (FPWAA),
- Fixed posture without audible airflow (FPWOAA),
- Superfluous verbal behaviors (SVB),
- Superfluous nonverbal behavior (SNB).

We can also define summary descriptors. The individual categories form the *overall* category which is subsequently normalised by the number of words in recording, which gives us a continuous scale from 0 to 100. The number of words was counted by the evaluators. All the descriptors are detectable in a speech signal except the descriptor SNB, which should be looked for in video recordings. The descriptor SNB is not used in these experiments because the video recording was not available. The descriptors *overall* (all descriptors except SNB), *repeated* (SR + ISR + MSUR), and *fixed* (FPWAA + FPWOAA) are considered in this thesis.

The reasons why this taxonomy was adopted in the thesis are: 1) when the system is used by experienced judges, the results are valid and reliable (Teesson *et al.*, 2003), 2) this taxonomy is easy to use, 3) when using the LBDL with all its categories the measures which fit the most for particular descriptors can be found. An example of using the LBDL in a quite similar problem can be seen in the research on stuttering symptoms in Parkinson's disease (Goberman *et al.*, 2010).

The information for SLPs, examples of individual symptoms, and also the form are given in section Appendices E and F (Figure 3) (information are given in Czech).

3.2 Objective measures of speech fluency

This stage is dedicated to automatic and objective measures of speech fluency. Several measures of fluency were introduced in Chapter 2. Many of them are based on the manual extraction of speech parts (Yaruss and Conture, 1993; Robb *et al.*, 1998; Healey and Gutkin, 1984; Hall and Yairi, 1992). Also, several studies introduced measures that can automatically process the entire speech signal without any external interference (Cucchiarini *et al.*, 2000; Nöth *et al.*, 2000; Wisniewski *et al.*, 2007a).

The measures were designed to be independent of any manual intervention and they take stuttering speech symptoms into account. Several of the measurements introduced here were presented in Bergl (2010). Three measures in Bergl (2010) were based on distribution of speech and silence parts of speech (speech/silence rate, number of speech and silence parts, and average length of silence), two on energy of speech signal (number of energy edges, histograms), regularity of speech energy, regularity of voicing. Moreover, the study described measures based on frequency analysis of the speech signal, in particular, analysis based on detection of abrupt spectral changes by Bayesian change–point detector (BACD). These measures were: average number of abrupt spectral changes, analysis of interval between spectral changes, variability of intervals between spectral changes. Furthermore, the the measures were tested using different spectral changes detectors, the first one, mentioned BACD, then GLR (General Likelihood Ratio), and HTK (Hidden Markov Model Toolkit).

In the current study, we take the measures designed by Bergl (2010) as well three novel measures and apply them to different recordings of disfluent speech. The automatic measurements designed to estimate the level of the speech fluency disorder in this thesis are listed below:

- The average length of silence (ALS)
- The regularity of speech energy (RSE)
- The regularity of voicing (REV)
- The extent of speech fluency (ESF)
- The spacing when exceeding the threshold (SET)
- The standard deviation of 11 successive intervals (SDI11)
- The average number of spectral changes in short intervals (SCSI)

- The number of spectral changes in speech segments (NSI)
- The rate of speech using continuous speech recognizer (ROS)

3.2.1 The average length of silence (ALS)

The first algorithm, the *average length of silence* (ALS), uses the voice activity detector (VAD) and assumes that subjects with stuttering have more silence and pauses in speech than healthy subjects do. The study (Prosek and Runyan, 1982) reported that pauses are potential perceptual cues for listeners attempting to discriminate the speech of stutterers from that of non–stutterers. When the placement, number and length of all silences in a speech signal are known, the average duration of silent parts can be computed:

$$ALS = \frac{1}{N_{SIL}} \sum_{i=1}^{N_{SIL}} T_{SILENCE}(i), \tag{3.1}$$

where $T_{SILENCE}(i)$ is the duration in seconds of the ith segment of silence and N_{SIL} is the number of segments of silence. See Figure 3.2, where in part (a), the speech signal, and in (b), the detected voice activity are depicted. The final value of the ALS is modified by summing up with 1 (to avoid the situation when the ALS equals 0) and then using the logarithm.

To make the difference between fluent and disfluent speech greater, an innovative procedure was developed (Bergl, 2010). Short speech parts, such as repetition, superfluous verbal behaviour, and parts of incorrectly pronounced words, could be removed by this method and the amount of silence increased in a speech signal with such disfluencies. The procedure uses successive removals of short segments of speech and silence. First, the speech segments shorter than 125 ms are removed (silent segments shorter than 125/2 ms are removed at the same time), next, there follows 150 ms, and this process continues up to the value 5000 ms. The procedure is demonstrated in Figure 3.2 (b) and (c). In this example, the ALS (without taking the logarithm and adding 1) is approximately 0.2 s (in (b)) and after removing intervals shorter than 150 ms, the ALS rises to 1 s (in (c)). The results presented in the results section are for the time limit values from 125 to 1500 ms.

3.2.2 The regularity of speech energy (RSE)

Changes in the tempo of speech are typical of disfluent speech, where the intervals of an ordinary speech rate alternate with intervals of a low speech rate. This irregularity can be observed in the irregular release of energy.

The following procedure calculates the *regularity of speech energy* (RSE): 1) square root of signal samples *En*; 2) these values are successively added into the accumulator; 3) when the value of the accumulator is lower than a threshold *Th*, then step 2) is repeated, in the opposite case (a threshold has been exceeded, *cumsum(En)= Th*), the accumulator resets and the time of the threshold exceeding is stored. See Figure 3.3 for an example of a speech signal (a), its accumulator curve (b), and the positions of threshold exceeding (c). The process results in the series of the indices indicating a threshold exceeding. Their average number, mean distance, or variation can be examined. The RSE investigates the mean distance between the indices:

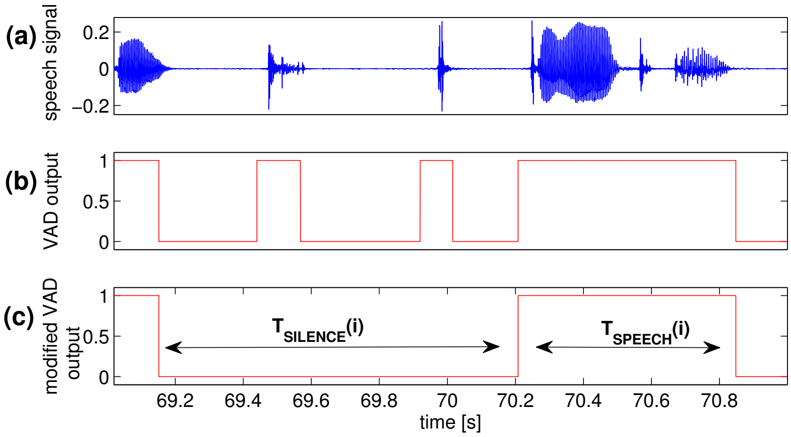


Figure 3.2: Steps of the calculation of the measure ALS. (a) Speech signal, repetition of the Czech word "k k kvitky" (flowers). (b) VAD detection. (c) Modified VAD output after successively removing speech segments shorter than 150 ms and silences shorter than 150/2 ms.

$$RSE = \frac{1}{N_{ACC}-1}\sum_{i=1}^{N_{ACC}-1} T_{ACC}(i), \tag{3.2}$$

where N_{ACC} is the number of indices and T_{ACC} is the time in seconds between two successive indices (see Figure 3.3 (c)). When we take the part of the signal in the figure as an example, the RSE of this fluently pronounced word is 0.024.

Previous experiments showed that one threshold does not work for the whole database. Therefore, an adaptive threshold is computed on the basis of the signal energy for each separate signal. The procedure for its calculation includes: the square root of the signal samples are filtered by a filter (an integrator with coefficient a= [1 -0.995]), then it is smoothed by a lowpass Butterworth filter with cutoff frequency 30 Hz. The threshold value is determined as a fraction of the kth highest maximum of this energy curve. The tested input values were: for the multiplication constant, from 0.25 to 2, and for the kth highest maximum, from 1 to 17. The best setting (1.5 for the multiplication constant and k = 15) were obtained experimentally by comparison to the control data and are presented in the result section.

3.2.3 The regularity of voicing (REV)

In the review section of the thesis several of the researches studied measurements based fundamental frequency (Healey and Gutkin, 1984; Healey and Ramig, 1986; Hall and Yairi, 1992). The measurement *regularity of voicing* (REV) attempt to use information derived from fundamental frequency for assessment of disfluency.

The basic step in the procedure is detection of voiced intervals and computation of the fundamental frequency F0 for these intervals. This part of procedure is calculated in Praat (Boersma,

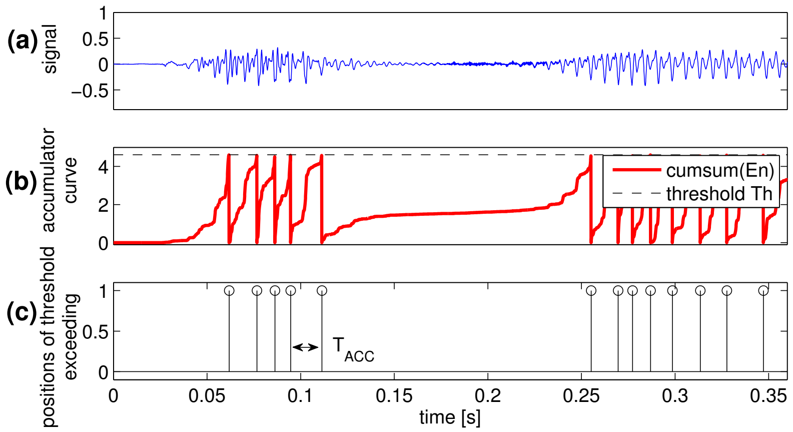


Figure 3.3: The procedure for calculating the RSE for the example of fluently pronounced “podzim” (autumn). (a) Speech signal. (b) The curve of accumulated energy and the threshold. (c) Positions of threshold exceeding.

2002). See the calculated fundamental frequency F0 in the short time window in Figure 3.4 (up). The data from Praat are later processed in Matlab script. Each value of F0 is added to the accumulator (see Figure 3.4 (*bottom*)), when the value of the accumulator is exceeded the time point if the exceeding is stored. These timestamps are then further processed. The procedure is very similar to the measure RSE. The limit of the accumulator (threshold) is set individually for every speech signal as median value of pitch period F0. The monitored variable is the variability of interval length defined by threshold exceeding.

The information about F0 enables calculation of many different measures as jitter or shimmer, however, Bergl (2010) did not find relation between them and speech disfluency.

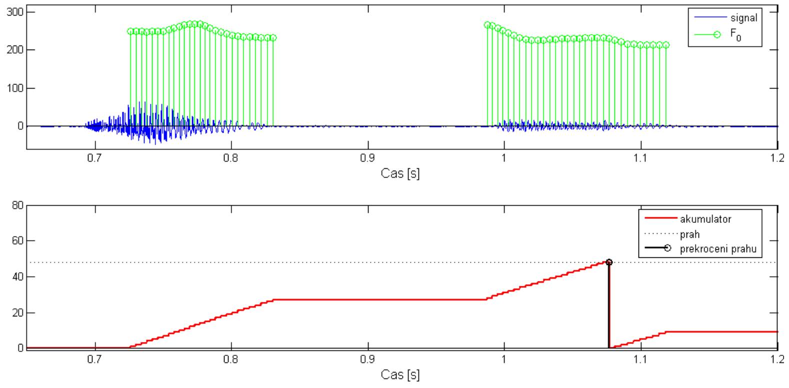


Figure 3.4: An example of calculating the REV (Bergl, 2010). (top) Speech signal and value of fundamental frequency F0 for each segment. (bottom) The curve of accumulated F0 with position of threshold exceeding.

3.2.4 The extent of speech fluency (ESF)

The speech rate has been found to be an important indicator of speech fluency (Johnson, 1961; Ryan, 1992; de Andrade *et al.*, 2003), and there have been experiments to automatically measure the rate of speech connected to speech fluency (Cucchiarini *et al.*, 2000). The measure, *extent of speech fluency*, is very close to the speech rate. The significant abrupt spectral changes (AC) found in a speech signal should correspond to the phoneme boundaries or transitions from speech to silence (and vice versa). Therefore, the number of significant ACs should closely match the number of phonemes. The Bayesian autoregressive changepoint detector (BACD) identified ACs in this experiment. The detector is based on the analytical solution of the changepoint problem between two autoregressive models (Ruanaidh and Fitzgerald, 1996). A detailed description is given in Cmejla *et al.* (2013). The output curve of the detector shows the probability of a spectral change in signal: the higher the probability, the higher the spectral change. The procedure of further analysis is as follows: first, all spectral changes are found in a signal, followed by the selection of those proved to be significant (i.e., which correspond to the phoneme boundaries). To distinguish between significant and less significant ACs, a threshold is used. Following identification, the procedures for the particular measures based on BACD start to differ.

The *extent of speech fluency* (ESF) counts the number of those significant ACs found by spectral changes detector, here BACD, which are higher than the threshold and therefore they should correspond to phoneme boundaries. The number of ACs is divided by the duration of the speech signal, analogous to the speech rate. The ESF assumes that the disfluent speakers speak slower, thus having a lower number of phonemes than fluent speakers (Johnson, 1961). In other words, the more disfluent the signal is, the fewer ACs are found in the signal.

Because the BACD output curve includes both significant and less significant abrupt changes, the following procedure is needed. The output of the BACD is filtered by a low–pass filter with the cut–off frequency at 20 Hz (to smooth the BACD output curve). The local minima are calculated in the smoothed output curve, thereafter the local maxima are found in the appropriate segments (between two local minima). Many of those local maxima do not correspond to significant spectral changes (phoneme boundaries) and they should be excluded. A threshold is utilized to separate these maxima (Figure 3.5 (b)). Then, the significant abrupt changes are obtained (Figure 3.5 (c)), their number is determined, and the ESF is calculated by the formula

$$ESF = \frac{\sum_{i=1}^{N_{AC}} AC(i)}{T_{SIGNAL}}, \tag{3.3}$$

where $AC(i)$ is an abrupt spectral change, N_{AC} is the number of abrupt spectral changes, and T_{SIGNAL} is the length of the speech signal in seconds. For example, in the figure there are 27 abrupt changes and the duration of this part of signal is 3 s, so the ESF of this fluently pronounced speech is 9.

The analysis, carried out on the detector outputs from different participants, showed that we are not able to use one threshold for the entire database. Hence a method of adaptive threshold extraction for each signal was used. The threshold is determined as a fraction of the kth highest maxima (the height of the maximum plays an important role in this selection). Several algorithm settings were tested: from 1 to 9 for k, and from 0.1 to 0.3 for the multiplication constant and their results are shown in results section in comparison to speech specialist evaluation. The best setting ($k = 4$ and 0.15 for the multiple) was established experimentally by comparison with the expert rating. The BACD of a sixth order AR model with a window length 60 ms was used in the whole experiment (Cmejla *et al.*, 2013; Bergl, 2010).

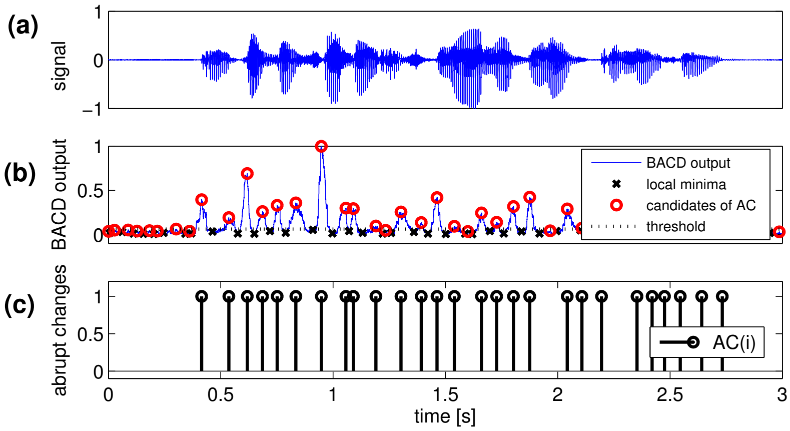


Figure 3.5: Identifying abrupt spectral changes. (a) Speech signal, Czech sentence "ozdoba sadu uschovana byla v komore" (garden adornment was kept in the pantry). (b) BACD output, local minima, and candidates of abrupt spectral changes. (c) Abrupt spectral changes.

3.2.5 The spacing between threshold exceeding (SET)

Two spectral changes define an interval of a certain length, and the difference in spacing of the spectral changes could be examined to obtain additional information about the disfluency.

The start of the analysis is identical to the previous measure (ESF), here after identifying the significant abrupt changes, the intervals between them are calculated. The histograms of the intervals between spectral changes are obtained, and the relative number of occurrences is captured (see Figure 3.6). The histograms are accumulated from left to right (the increasing line in the figure). When the accumulator exceeds a stated threshold (in the figure, the dashed line at 0.8), the value of the length of the intervals is stored: this value is a numerical representation of the SET.

We experimentally examined several settings of the algorithm, including the input values of the BACD and the threshold. The values were: from the second highest abrupt change to the sixth highest abrupt change; multiplication constants 0.1, 0.15, and 0.2; and SET thresholds of 0.6, 0.7, and 0.8. The best results, as determined by a comparison to the control data, were achieved using the fourth highest maximum, a multiple of 0.15, and a SET threshold of 0.8, see the results in the first study of the thesis.

There is an assumption that control (healthy) speakers accumulate faster (they have steeper rises) than speakers with fluency disorder. The reason is that fluent speakers have more shorter intervals, which is the opposite from speakers with the fluency disorder. This can be quantitatively expressed by the interval where the accumulator crosses a threshold value. Figure 3.6 depicts the histogram and accumulator curve for a control participant (LBDL score 0%, Condas's scale score 0, SET value 0.15) and a participant with the speech fluency disorder (LBDL score 51.95%, Condas's scale score 3, SET value 0.43), to see the difference between fluent and disfluent speech.

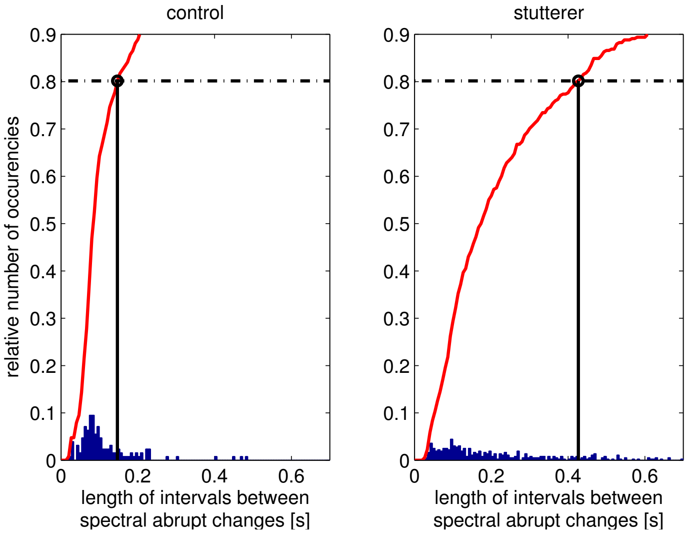


Figure 3.6: Histograms of spacings, the progress of the accumulator (steep line), threshold value (dashed line), and the place where the threshold has been exceeded, for two speakers (control speaker on the left, speaker with severe level of speech fluency disorder on the right).

3.2.6 The standard deviation of 11 successive intervals (SDI11)

Another approach to analyzing the intervals defined by abrupt spectral changes is to examine their standard deviation. The measure called the *speech fluency variability* (SFV) was described in Cmejla *et al.* (2013), and is defined as the logarithm of the standard deviation of the distances between two successive spectral changes. The SFV computes the standard deviation at once, it does not capture the order followed by the intervals. In other words, from the point of view of SFV, it does not matter whether very short intervals follow very long ones, or vice versa, as we can expect in disfluent speech. Their mutual and overall position does not play a role. The measure standard deviation of 11 successive intervals (SDI11) take this into account.

The calculation of the measure is as follows: the first step is the identification of the relevant spectral changes, then an 11-point moving average is used on the intervals between the identified spectral changes (T_{AC}). In more detail: the average of the first eleven intervals (T_{AC}(1),..., T_{AC}(11)) is calculated; then one moves one interval ahead and the average of the intervals T_{AC}(2),..., T_{AC}(12) is computed. The standard deviation is calculated for this set of averages. This can be expressed by the equation

$$SDI11 = std\{MA_{11}(T_{AC})\}, \tag{3.4}$$

where *std* means the standard deviation and MA_{11}, the 11–point moving average. The logarithm is used for its final value. The tested settings of the algorithm were the same as for the ESF. The highest correlation with the control data was achieved by using the fourth highest maximum and the multiplication constant 0.15, more in results of the first study of the thesis.

3.2.7 The average number of spectral changes in short intervals (SCSI)

In Chapter 1, it was mentioned that disfluent speech consists of many prolongations, frequent pauses, and broken words. The *average number of spectral changes in short intervals* (SCSI) tries to capture these phenomena by processing the BACD output in short windows. If the output of the BACD is processed in short segments, the difference in the number of abrupt changes could be significant for segments with speech activity as opposed to segments with silence, taking into account the comparison of disfluent speech to healthy speech. For participants with disfluencies, it is expected that more silence appears in stuttered than in fluent speech (the average number of changes in the window is smaller). The number varies and in many cases is zero. Conversely, the number of changes for healthy speakers is more stable and the average should be higher.

The procedure of analysis using the average number of ACs in a short interval begins with identification of significant abrupt spectral changes. It is followed by the processing of the detector output in a short window. The number of spectral changes is found in each window and the average number of abrupt spectral changes in the windows is quantified. The logarithm is used for the final value. An example of this calculation can be seen in Figure 3.7, where the value of the logarithm of the SCSI is 0.84 (it is a part of a disfluent speech signal with severe disfluency) and the window length is 2 s.

The tested window lengths were 1, 2, and 4 s, with half–overlap and all used window lengths reached very similar results. The studied BACD settings were as for the ESF. The window length 2 s and all settings of BACD are presented in the results section.

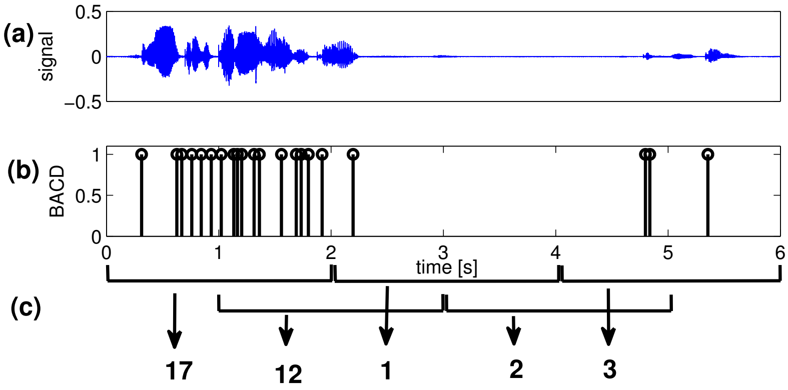


Figure 3.7: Procedure for calculating the average number of BACD changes in a short interval. (a) Speech signal, a part of the Czech sentence “chomáče starého listí bůh ví kam” (bunch of old leaves God knows where). (b) Modified output signal of Bayesian detector. (c) Processing by means of a window with marked number of abrupt spectral changes.

3.2.8 The number of spectral changes in speech segments (NSI)

This measure, the *number of spectral changes in speech segments* (NSI), makes the same assumption as the ESF, and combines the BACD and VAD. The algorithm looks for significant spectral changes which are higher than the threshold and within speech segments, then the number of these ACs is divided by the length of the recording (analogous with the speech rate). A very important part of the NSI algorithm is the successive removal of short speech segments, a procedure similar to the ALS measure. The time limits used in this experiment varied from 100 up to 5000 ms in regular intervals.

The beginning of the procedure is the same as in the previous BACD algorithms up to the point of identifying the relevant spectral changes, at which point one then implements the following step: applying the VAD and spectral changes in speech segments are identified. The number of spectral changes in each of the speech segments is determined (the number of spectral changes in the ith speech segments is labelled as $N_{AC\,speech}(i)$) and finally the number of spectral changes in all speech segments is summed up and divided by the length of the speech signal T_{SIGNAL} in seconds. See Figure 3.8 for a short demonstration of the method. When N_{speech} is the number of speech segments, the measure NSI can be written as follows:

$$NSI = log_{10}\frac{\sum_{i}^{N_{speech}} N_{AC\,speech}(i)}{T_{SIGNAL}}. \tag{3.5}$$

There is also an additional step to this procedure, as at the measurement of the ALS, it is the successive removal of short speech segments which increases the difference between fluent and disfluent speech. The example in the figure shows a short part of the speech signal "Podzim na starem belidle" (autumn at the old bleachery) where the value of the NSI would be 0.43. All tested settings of the BACD are shown in the results section, the time limit for removing short speech and silence segments is 1000 ms based on the ALS algorithm results.

3.2.9 The rate of speech (ROS)

The following measure, the *rate of speech* (ROS), uses the automatic recognizer of Czech phonemes based on long temporal context (Schwarz, 2009). The measure is defined as the number of phonemes found by the recognizer and divided by the duration of speech including utterance internal silences (Cucchiarini *et al.*, 2000). This measure as well as NSI and ESF should highly coincide with the rate of speech measured as phonemes/time. The example of a recognizer output file is as follows:

```
0 1100000 pau
1100000 4100000 spk
4100000 5000000 p
5000000 5600000 o
5600000 6800000 d_z
6800000 7200000 i
7200000 8400000 n
8400000 8700000 a
```

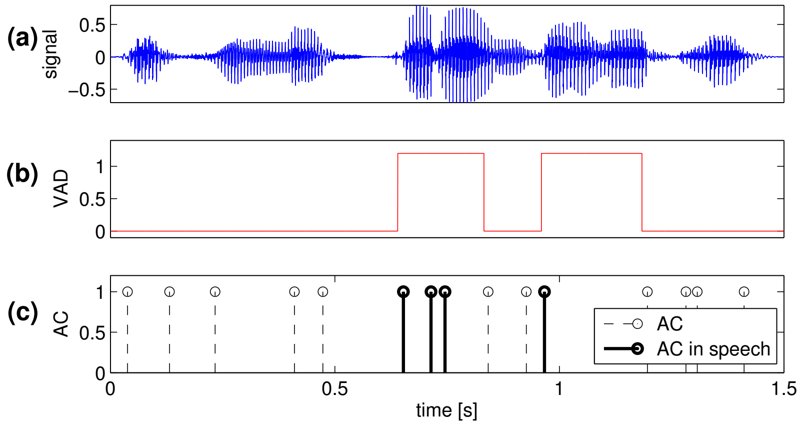


Figure 3.8: Calculation of the number of spectral changes in speech segments (NSI). (a) Speech signal. (b) Speech activity output with successive removed speech and silence parts (1 – speech activity, 0 – silence/pause). (c) Abrupt changes (dashed line) and ACs included in speech segments (thick line).

where the first number represents the time of the beginning of the phoneme, the second number the time of the end, and finally the phoneme (or event in the speech signal).

The very same measure was used in Bergl (2010), the difference between the previous measure and the ROS is that a different speech recognizer is applied. The automatic recognizer of Czech phonemes based on a long temporal context (Schwarz, 2009) is trained on larger database and is tested in real situations; therefore the detection rate should be higher.

Chapter 4

Analysis of speech fluency in read recordings

The first study deals with the recordings of read speech of subjects with no speech fluency impairment and participants with stuttering. Basically, it describes the first experiments on recordings of disfluent speech.

The diagnosis and evaluation of the severity of a speech disorder are traditionally performed by clinical experts. Several stuttering scales have been introduced, such as the Lidcombe Behavioral Language of Stuttering (Teesson *et al.*, 2003), Stutering Severity Instrument (Riley, 1972), but there has been a need for automatic and objective methods. Such a method would be helpful in diagnosis, the choice of treatment approach, and the evaluation of treatment progress and results (Metz *et al.*, 1983; Van Borsel *et al.*, 2003).

The application of acoustical analysis could provide an objective and quantitative instrument to mark the presence of stuttering symptoms and/or describe the severity, characteristics, and progress of the disorder and its treatment (Kent *et al.*, 1999). Studies (Di Simony, 1974; Metz *et al.*, 1983; Adams, 1987) have focused on the temporal characteristics of stuttered speech, investigating, for example, vowel duration and voiced stop consonant intervocalic intervals. The rate of speech (manually measured) has been also recognized as a helpful tool for the evaluation of stuttering (Johnson, 1961; Ryan, 1992; de Andrade *et al.*, 2003).

Methods based on digital signal processing may offer insight into stuttered speech. A great effort has been devoted to studying the behaviour of formant frequencies, the fundamental frequency, and the voice onset time (VOT). The transition of the second formant frequency has been studied in Yaruss and Conture (1993), formant frequency fluctuation in Robb *et al.* (1998), fundamental frequency and fluent VOT in Healey and Gutkin (1984), fluent VOT and phrase duration in Healey and Ramig (1986), and fundamental frequency, jitter, and shimmer in Hall and Yairi (1992). Computer programs can be efficiently applied to the objective analysis of pathological speech. The computer system Multi–Dimensional Voice Program developed by Kay Elemetrics Corp. (Kay Elemetrics Corp., 2003), and the freely available PRAAT (Boersma, 2002), are among these programs and provide several measures for speech evaluation. However, the disadvantage of these programs is mostly the need for user control of the analysis. This can be avoided by using methods that process the entire signal without user control. An approach simply using temporal characteristics to find repetition and prolongation can be seen in Howell *et al.* (1986). Advanced digital signal processing methods have been employed for identifying manually selected stuttered parts of speech: Mel Frequency Cepstral Coefficients in Ravikumar *et al.* (2009), and Linear Predictive Cepstral Coefficients in Hariharan *et al.* (2012). Hidden Markov Models (HMM) have been utilized in Nöth *et al.* (2000); Wisniewski *et al.* (2007a,b) to reveal repeated or prolonged parts of disfluent speech. A method does not have to look for symptoms of speech disorder: it could process the signal in another way. Such a method could investigate the energy of the speech signal (speech envelopes) (Kuniszyk-Jozkowiak, 1995, 1996) or could utilize Kohonen networks for the detection of speech nonfluency (Szczurowska *et al.*, 2009).

Research on other speech disorders and in different areas of acoustics could supply interesting results and ideas. Maier *et al.* (2011) used automatic methods for evaluation of reading disorder in children's speech where the total reading time is one of the most useful feature. Articulation disorder in children with a cleft lip or palate were investigated in Maier *et al.* (2009b), and patients who have had their larynx removed due to cancer and children with a cleft lip or palate in Maier *et al.* (2009a). Study (Godino-Llorente and Gomez-Vilda, 2004) has used short–term cepstral parameters to identify vocal fold impairment due to cancer. Acoustical methods have been applied to non–invasive biomarkers of patients with Parkinson's disease (Sapir *et al.*, 2010; Rusz *et al.*, 2011). Cucchiarini *et al.* (2000) applied a continuous speech recognizer to the quantitative assessment of second language learners' fluency. Nine automatic measurements based on temporal features of speech, such as the rate of speech, articulation rate, or the total duration of the pauses, have been employed.

The aim of the study was to determine whether the level of the speech fluency disorder can be estimated by means of automatic acoustic measurements. Investigating recordings of read text could be a step towards spontaneous speech which is more common in clinical practice, and clinical experts would appreciate a method that could help with evaluating.

4.1 Method

4.1.1 Participants and recordings

The speech signal database, which the read recordings are part of, was formed in past several years at the Department of Phoniatrics of the 1st Faculty of Medicine at the Charles University and the General Faculty Hospital in Prague. The database is very large and not all the recordings could have been included in the experiments because of their technical quality.

The read part contains recordings of 118 Czech native speakers (28 women and 90 men) with different ages and levels of speech fluency disorder. The age structure of the database is as follows: Mean age 18.1 yr [±standard deviation (SD), 9.9 yr], the youngest participant was 8 yr old, the oldest was 50 yr old. Fifteen recordings (5 women and 10 men) are utterances of speakers without speech fluency disorder [mean age 27.37 yr (±SD, 7.4 yr)] and speakers with speech fluency disorder were in age, mean 16.73 yr (±SD, 9.4 yr). See the distribution of participants' age in Figure 4.1. All participants read the standard text used by Czech speech therapists (*Podzim na Starém bělidle*), the text is about 70 word–long, it is phonetically non–balanced, and it does not include tongue twisters. The book is a part of recommended literature for the 6–8th grade of grammar school in Czech Republic (children at the age of 12–14 years). The entire text can be seen at the end of the document in section Appendices A. The average length of a recording is 66.1 s (±SD, 33.3 s).

The utterances were recorded with a sampling frequency of 44 kHz. The signals were down–sampled to 16 kHz for the subsequent analysis.

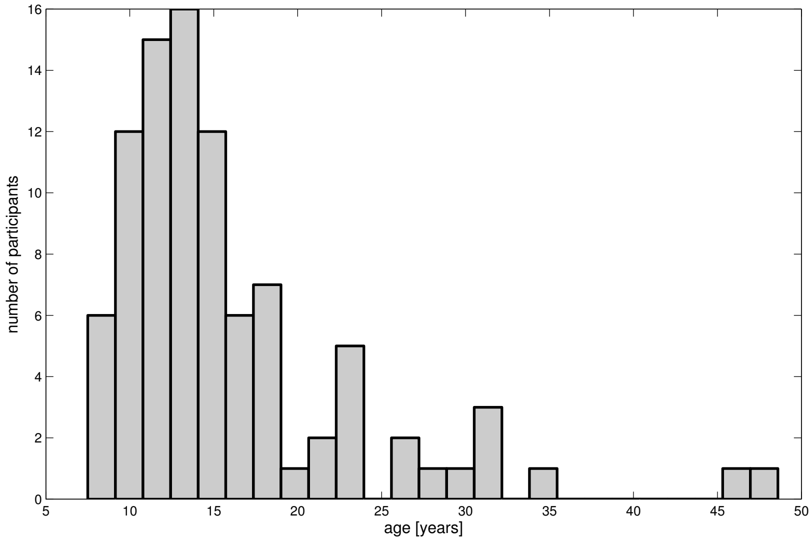


Figure 4.1: Histogram, distribution of participants' age in the study on read recordings (118 participants).

4.1.2 Subjective rating of read recordings

The evaluation of read recordings was performed by two professional speech pathologists using the Kondas scale. The evaluators assessed recordings independently and according to the performance of subject and the best knowledge of the evaluator. The two judgements of read recordings were merged for further procedures. In the case that the ratings differed (for example, the first assigned the level 2 and the second 3), the higher level was adopted. The structure of speech fluency disorder according to the modified Kondas scale (merged judgement of two therapists) is as follows: the groups of 0, 1, 2, 3, and 4 include 15, 24, 41, 31, and 7 recordings.

One evaluator listened to the read recordings and wrote down the number of occurrences of all LBDL categories in each recording. The distribution of evaluation on read recordings is depicted in Figure 4.2.

In addition to the two evaluation scales, we also consider the rate of speech. The rate of speech [words/time] was obtained from the number of words counted during the LBDL evaluation procedure and the length of recording.

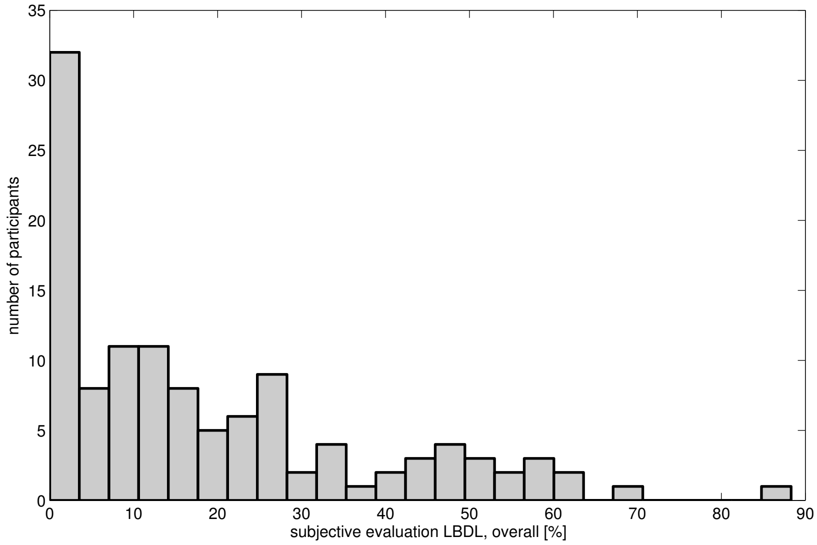


Figure 4.2: Histogram of the values of subjective evaluation made by means of the LBDL on read recordings (118 participants).

4.1.3 Measures of speech fluency

Study 1 uses all measurement of speech fluency introduced in the Chapter 3 – Method. In addition, the comparative measure the total reading time is also used. The measures are listed below:

- The average length of silence (ALS)
- The regularity of speech energy (RSE)
- The regularity of voicing (REV)
- The extent of speech fluency (ESF)
- The spacing when exceeding the threshold (SET)
- The standard deviation of 11 successive intervals (SDI11)
- The average number of spectral changes in short intervals (SCSI)
- The number of spectral changes in speech segments (NSI)
- The rate of speech using continuous speech recognizer (ROS)
- The total reading time (RT)

4.2 Statistics

The ability to recognize levels of speech disfluency was examined using the Pearson product–moment correlation, the classification with the Linear Discriminant Analysis (LDA), and the statistical method ANOVA with *post hoc* Bonferroni adjustment. Firstly, all settings of each algorithm are examined by means of correlations and deviations with respect to the expert ratings. Secondly, the ANOVA analysis is performed for one selected setting of each of the four acoustic measures to find significant differences between fluency levels. Then, the relationship between the acoustic measures and all categories of the LBDL is evaluated by the Pearson product–moment correlation. The Kolmogorov–Smirnov test was used to examine the normality of the distribution of the data.

To demonstrate how the algorithms are able to separate all subjects into disfluency levels, the LDA is used. The LDA (Harrington and Cassidy, 1999), a statistical technique, takes the knowledge that an element from training data set belongs to a certain group/level. On the basis of the elements' mean and standard deviation, the discriminant function is determined for each group from training data set. These discriminant functions could then be used for classification of a new element. Because the number of participants in the experiment is rather lower, especially in peripheral levels 0 and 4, we decided to perform the leave–one–out cross–validation instead of dividing the database into test and validation group. When using this method, all elements of the data set except one serve as the training set and the one element is used as the validation data. This is repeated for each element of the data set, thus each element is used as the validation data.

The deviation Δ is defined to assess the success of classification

$$\Delta = \sum_{i=1}^{N} (|o_i - \widehat{o}_i|). \tag{4.1}$$

where o_i is the merged evaluation of SLPs for the ith subject in the database, $\widehat{o}_i$ represents the estimated level for the same subject, the difference $o_i - \widehat{o}_i$ represents the classification error for one subject, and N is the number of subject in the database. When inspecting the results, we can follow the theorem: the smaller deviation Δ, the better result of classification achieved.

4.3 Result

4.3.1 Reliability of subjective evaluation on read recordings

All the read speech recordings were evaluated by two evaluators using the modified Kondas's scale and by one evaluator using the LBDL. As appears from the Pearson correlation and Cronbach's alpha (standardised), the expert rating (the Kondas's scale) shows a very high relationship between both therapists, a correlation of 0.91 ($p < 0.001$) and Cronbach's alpha 0.95. The evaluation made by the speech therapists was also compared to the subjective evaluation made by means of the LBDL (overall), and here the Pearson correlation coefficient was 0.93 ($p < 0.001$) for the first expert, 0.92 ($p < 0.001$) for the second, and for the merged evaluation: 0.93 ($p < 0.001$). The logarithm of the LBDL evaluation was used because the Kondas's scale is rather logarithmic (Cmejla *et al.*, 2013).

The LBDL reports a high level of agreement in describing stuttering events and shows consistent results for intra– and inter–judge agreement (Teesson *et al.*, 2003). Thirty recordings (20% of the 118 recordings) were assessed twice to obtain the intra–judge agreement; the same 30 recordings were assessed by the second evaluator to obtain the interjudge agreement, similar to Goberman *et al.* (2010). The lowest correlation coefficients across all descriptors for intrajudge agreement were 0.87 ($p < 0.001$) and 0.89 ($p < 0.001$), the others exceeded 0.94 ($p < 0.001$). The results for interjudge agreement also seem to show good results (> 0.76) except for descriptor superfluous verbal behaviors (SVB). The agreement for descriptor SVB was 0.32 ($p = 0.08$). These events are important but not as much so as the other events (repetition, prolongation, pauses), therefore we decided to use this evaluation but the results related to the descriptor SVB are viewed carefully.

The results of the intra– and inter–judge reliability in the read experiment achieved a high level of agreement, and we can conclude that the evaluation is reliable and applicable for the purpose of this experiment.

4.3.2 Results of the ALS

Table I presents the correlation of the objective measurement ALS with the merged subjective evaluation (Kondas scale, levels 0–4), all tested settings of the algorithm are included in the table. The ALS measure yielded the best correlation 0.64, the correlation rises with increasing time limit (the scope of setting) achieving its top at 1000 ms followed by a slow decline. The trend of classification deviation is the opposite; the deviation decreases with rising time limit (the classification is more successful), one of the local minima is reached at the time limit 1000 ms $\Delta = 89$, but the smallest deviation is 82 at the time limit 1500 ms (for details, see Table I). According to the correlation coefficients and classification results, we selected the setting with threshold 1000 ms, which is used in the following experiments.

The typical values of the measurement and correlation with the second subjective rating are presented in Table IX. We hypothesised that the ALS measurement would rise with level of speech disfluency and Figure 4.3 supports our assumption. Correlations with the second set of subjective evaluation according to the LBDL are presented in Table XXIII. The highest correlation can be found between objective measures and individual categories FPWOAA (0.73), and summary categories *fixed* (0.72) and *overall* (0.64). Agreement with other categories did not overcome coefficient 0.5.

The ANOVA analysis of the ALS measure found significant differences between levels moderate and severe disfluency (2 vs. 3), severe and very severe (3 vs. 4) at the significance level 0.001 (details in Table XI). The measure classified 47 subjects into correct disfluency level, 53 subjects were classified with 1–level error, and 18 subject were classified with 2–level difference from correct class.

4.3.3 Results of the RSE

Correlation coefficients with Kondas scale and results of classification for all tested algorithm settings are presented in Table II. The results show that the algorithm settings are robust, the correlation for more than half of settings is similar (0.64 – 0.65, varying at the third decimal place). The highest correlation reaches 0.65. Also, the classification results exhibit robustness, the classification deviation vary around 90 and the smallest is 87. According to the results, the setting with $k = 15$ and ,multiplication constant 1.5 was selected for further analysis.

Typical values of the RSE measure and the measure's dependency on the level of the disfluency are shown in Table IX. See also the visual form of these values in Figure 4.4. The RSE grows with

TABLE I: The Pearson correlation and results of classification using the LDA (the deviation Δ from subjective evaluation) for all algorithm settings of the ALS in comparison to the merged subjective evaluation, the time limit for successive removing of short speech and silence segments is the subject of setting.

settings [ms]	correlation (deviation Δ)
125	0.35 (172)
150	0.34 (153)
200	0.36 (146)
300	0.40 (123)
400	0.46 (124)
500	0.50 (121)
700	0.56 (109)
800	0.59 (105)
900	0.62 (94)
1000	**0.64 (89)**
1100	0.64 (91)
1200	0.62 (98)
1300	0.62 (102)
1400	0.62 (112)
1500	0.62 (82)

the rising level of speech disfluency. Table XXIII includes Pearson correlation of the RSE measure with categories of LBDL taxonomy. High correlation can be observed for ISR (0.66), FPWOAA (0.65), and summary categories *repeated* (0.67), *fixed* (0.69), and overall (0.76).

The ANOVA, see Table XI, found significant differences for the RSE measure between levels 2 vs. 3 vs. 4 (moderate vs. severe vs. very severe disfluency) at the significance level 0.001. In the classification task, the measure was able to correctly classify 44 subjects, 59 with error of 1 level, and 12 with 2–level classification deviation.

TABLE II: The Pearson correlation and results of classification using the LDA (the deviation Δ from subjective evaluation) for all algorithm settings of the RSE in comparison to the merged subjective evaluation, to set the algorithm the kth highest maximum and multiplication constant are used.

	multiplication constant, correlation (deviation Δ)						
k	25	50	75	100	125	150	200
0	0.41 (150)	0.41 (149)	0.41 (149)	0.41 (149)	0.41 (149)	0.41 (149)	0.41 (149)
4	0.53 (127)	0.53 (127)	0.53 (127)	0.53 (127)	0.53 (127)	0.53 (127)	0.53 (127)
9	0.59 (100)	0.59 (101)	0.60 (101)	0.60 (101)	0.60 (101)	0.60 (101)	0.59 (102)
12	0.64 (89)	0.64 (90)	0.64 (89)	0.64 (89)	0.64 (90)	0.64 (87)	0.64 (89)
14	0.65 (94)	0.65 (93)	0.65 (93)	0.65 (93)	0.65 (92)	**0.65 (92)**	0.65 (92)
16	0.64 (92)	0.65 (93)	0.65 (91)	0.65 (89)	0.65 (89)	0.65 (90)	0.65 (90)

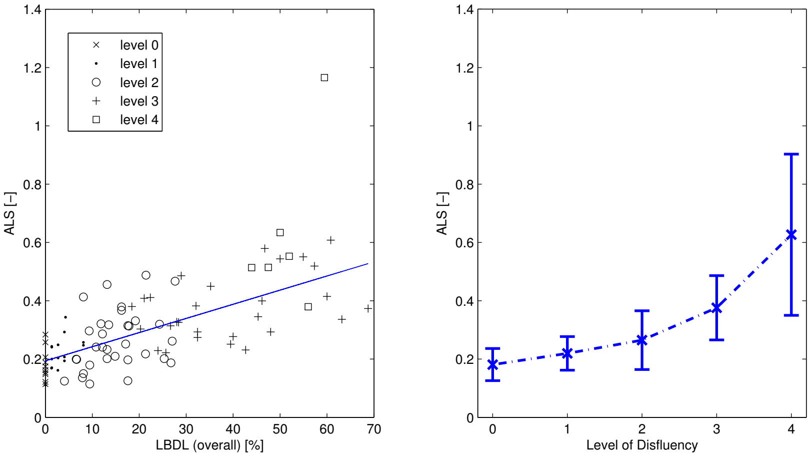


Figure 4.3: Comparison of the measurement ALS to subjective evaluation made with LDBL (left) and Kondas scale (right) in read recordings. The individual levels of disfluency are labelled by marks in the left figure. In the right figure, a marker indicates the mean value for the given level of fluency, error bars represent SD.

4.3.4 Results of the REV

The comparison of all settings of measure REV and subjective evaluation of speech disfluency to find the most suitable setting is given in Table III. The highest correlations ranges from 0.61 to 0.64 and classification deviations from 112 to 117. The setting marked as *0.50* seem to be the best combination, correlation 0.64 and classification deviation 112. This setting was selected for further analyses.

Table IX and Figure 4.5 show the typical values of the measure in read recordings and demonstrate the trend of the measure with progressing disfluency. The REV increases with rising level of the disfluency. A problem is that fluent speakers are, according to the measure, less fluent than the mild disfluency group. Pearson correlation coefficients of the measure and second set of subjective evaluation (LBDL) are presented in Table XXIII. We observed good correlations with ISR (0.55), FPWOAA (0.66) and summary descriptors *repeated* (0.57), *fixed* (0.65) and *overall* (0.68).

Table XI includes the ANOVA analysis of the REV measurement. The analysis found significant differences between levels 1 vs. 2 (significance level $p = 0.05$), 2 vs. 3 (0.01), and 3 vs. 4 (0.01). The REV measure classified 45, 41, 25, and 7 with classification error 0, 1, 2, and 3, respectively.

4.3.5 Results of the ESF

To select the most suitable setting of the ESF algorithm, the correlation coefficients of the measure and subjective evaluation (Kondas) are presented in Table IV. Several settings overcome coefficient -0.7. The highest (-0.77) was found for $k = 4$ and multiplication constant 0.15. This setting is used for further analyses. Typical values for different levels of speech disfluency are presented

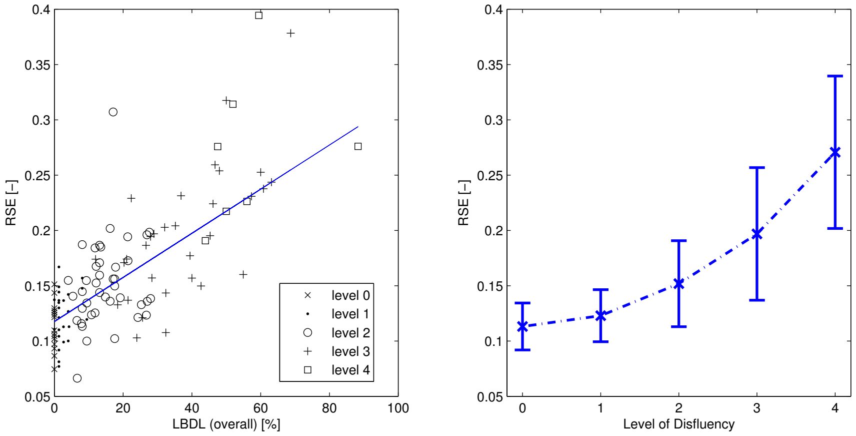


Figure 4.4: Comparison of the measurement RSE to subjective evaluation made with LDBL (left) and Kondas scale (right) in read recordings. The individual levels of disfluency are labelled by marks in the left figure. In the right figure, a marker indicates the mean value for the given level of fluency, error bars represent SD.

in Table IX and Figure 4.6. The measure decreases with the growing level of disfluency. Pearson correlation coefficients of the measure and second set of subjective evaluation (LBDL) are presented in Table XXIII. We observe good correlation with ISR (-0.51), MSUR (-0.4), FPWOAA (-0.67) and summary descriptors *repeated* (-0.63), *fixed* (-0.73) and *overall* (-0.76).

Table XI shows the ANOVA analysis of the ESF measurements. Significant differences were found between levels 1 vs. 2 (significance level $p = 0.001$), 2 vs. 3 (0.001). Differences between other levels were not significant. In classification task, the ESF classified correctly 57 subjects, 48 subjects were classified incorrectly of 1 level, and 13 participants missed their level of 2.

4.3.6 Results of the SET

Table V presents correlation coefficients of the all SET measure settings with subjective evaluation on read recordings. According to the table, we selected an appropriate setting. The measure exhibits robustness as several settings exceeded correlation of 0.7. The highest was 0.78 for the threshold 80, k = 4, and multiplication constant 0.15, this setting is used for further analyses. Table IX includes typical values of SET at various levels of speech disfluency for the selected setting; a graphical interpretation of these results is in Figure 4.7. The measurement rises with growing level of speech disfluency, the correlation coefficients are positive. The correlation of the selected setting with LBDL scale are shown in Table XXIII. The highest correlation can be seen for individual LBDL descriptors MSUR (0.56) and FPWOAA (0.74), while for summary descriptors *repeated* (0.61), *fixed* (0.79) and *overall* (0.77).

The ANOVA found significant differences between levels 1 vs. 2 (significance level $p = 0.05$) and 2 vs. 3 (0.001). Differences between other levels were not significant. See Table XI for details.

TABLE III: The Pearson correlation and results of classification using the LDA (the deviation Δ from subjective evaluation) for all algorithm settings of the REV in comparison to the merged subjective evaluation, the time limit for successive removing of short speech and silence segments is the subject of setting.

settings [-]	correlation (deviation Δ)
0.01	0.32 (123)
0.05	0.61 (114)
0.10	0.63 (116)
0.20	0.64 (114)
0.30	0.63 (117)
0.40	0.63 (114)
0.50	**0.64 (112)**

TABLE IV: The Pearson correlation and results of classification using the LDA (the deviation Δ from subjective evaluation) for all algorithm settings of the ESF in comparison to the merged subjective evaluation, to set the algorithm the *k*th highest maximum and multiplication constant are used.

	multiplication constant, correlation (deviation Δ)				
k	0.10	0.15	0.20	0.25	0.30
1	-0.75 (85)	-0.73 (95)	-0.67 (98)	-0.61 (105)	-0.53 (126)
2	-0.76 (80)	-0.76 (82)	-0.72 (96)	-0.65 (102)	-0.60 (105)
3	-0.75 (74)	-0.76 (79)	-0.72 (102)	-0.66 (120)	-0.61 (135)
4	-0.75 (79)	**-0.77 (74)**	-0.73 (95)	-0.68 (112)	-0.62 (127)
5	-0.74 (74)	-0.76 (72)	-0.74 (92)	-0.69 (111)	-0.65 (121)
6	-0.74 (87)	-0.77 (74)	-0.75 (87)	-0.70 (112)	-0.66 (125)
7	-0.72 (97)	-0.77 (73)	-0.75 (86)	-0.71 (106)	-0.67 (121)
8	-0.72 (93)	-0.76 (75)	-0.76 (82)	-0.73 (106)	-0.68 (121)
9	-0.72 (83)	-0.76 (75)	-0.76 (82)	-0.74 (101)	-0.70 (108)

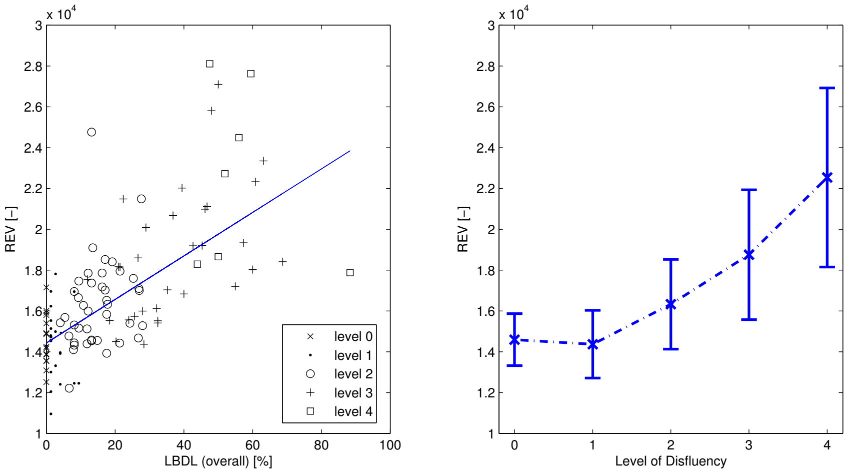


Figure 4.5: Comparison of the measurement REV to subjective evaluation made with LDBL (left) and Kondas scale (right) in read recordings. The individual levels of disfluency are labelled by marks in the left figure. In the right figure, a marker indicates the mean value for the given level of fluency, error bars represent SD.

In the classification task, the measure classified correctly 38 subjects, 60 with error 1 level, 20 subjects with error higher or equal 2.

4.3.7 Results of the SDI11

According to correlation coefficients of the measure and subjective evaluation (Kondas), which is presented in Table VI, we selected the most suitable setting of the algorithm, $k = 4$, multiplication constant 0.15. The mean and SD of the measure (selected setting) with respect to the level of disfluency can be seen in Table IX or in Figure 4.8. The mean values increase with rising level of speech disfluency. Table XXIII presents correlation of selected measure setting with continuous LBDL scale. We observed good correlations in individual LBDL categories: ISR (0.52), MSUR (0.52), and FPWOAA (0.65); while in summary descriptors *repeated* (0.61), *fixed* (0.71) and *overall* (0.74).

Significant differences between levels 1 vs. 2 (significance level $p = 0.001$) and 2 vs. 3 (0.001) were found by means of ANOVA analysis (Table XI). Significant differences between other levels were not found. The measurements was able to classify 43 subject into correct level, 63 missing the correct level of one level, and 12 subjects was classified with error equal 2 levels.

4.3.8 Results of the SCSI

Table VII displays correlations of the SCSI measure and subjective evaluation by means of the Kondas scale. According to the table, we chose one suitable setting which is presented in further analyses. The setting uses a short–time window of the length 2 s, $k = 6$ (sixth highest maxima in

TABLE V: The Pearson correlation and results of classification using the LDA (the deviation Δ from subjective evaluation) for all algorithm settings of the SET in comparison to the merged subjective evaluation, to set the algorithm the kth highest maximum and multiplication constant are used.

	threshold 60, multiplication constant, correlation (deviation Δ)		
k	0.10	0.15	0.20
1	0.72 (94)	0.72 (77)	0.66 (91)
2	0.70 (110)	0.71 (89)	0.66 (111)
3	0.70 (117)	0.73 (92)	0.69 (106)
4	0.69 (117)	0.73 (94)	0.69 (101)
5	0.69 (110)	0.73 (107)	0.70 (102)
	threshold 70		
k	0.10	0.15	0.20
1	0.74 (89)	0.72 (93)	0.69 (104)
2	0.70 (106)	0.72 (102)	0.67 (106)
3	0.71 (105)	0.74 (98)	0.71 (107)
4	0.71 (107)	0.75 (103)	0.69 (106)
5	0.70 (113)	0.75 (104)	0.69 (107)
	threshold 80		
k	0.10	0.15	0.20
1	0.77 (98)	0.77 (90)	0.73 (103)
2	0.74 (114)	0.74 (104)	0.69 (115)
3	0.74 (114)	**0.78 (103)**	0.72 (118)
4	0.74 (107)	0.77 (105)	0.72 (116)
5	0.75 (104)	0.77 (106)	0.73 (114)

TABLE VI: The Pearson correlation and results of classification using the LDA (the deviation Δ from subjective evaluation) for all algorithm settings of the SDI11 in comparison to the merged subjective evaluation, to set the algorithm the kth highest maximum and multiplication constant are used.

	multiplication constant, correlation (deviation Δ)				
k	0.10	0.15	0.20	0.25	0.30
1	0.73 (77)	0.74 (83)	0.73 (96)	0.72 (98)	0.70 (98)
2	0.73 (77)	0.75 (82)	0.76 (84)	0.74 (96)	0.73 (110)
3	0.71 (82)	0.74 (89)	0.75 (92)	0.73 (103)	0.73 (101)
4	0.70 (83)	0.74 (87)	0.75 (88)	0.74 (95)	0.73 (102)
5	0.70 (84)	**0.74 (79)**	0.74 (89)	0.73 (98)	0.73 (97)
6	0.69 (84)	0.74 (78)	0.74 (90)	0.73 (93)	0.73 (97)
7	0.67 (79)	0.73 (78)	0.74 (92)	0.73 (95)	0.73 (104)
8	0.66 (80)	0.74 (79)	0.74 (88)	0.73 (95)	0.72 (101)
9	0.65 (78)	0.73 (78)	0.74 (88)	0.73 (94)	0.72 (100)

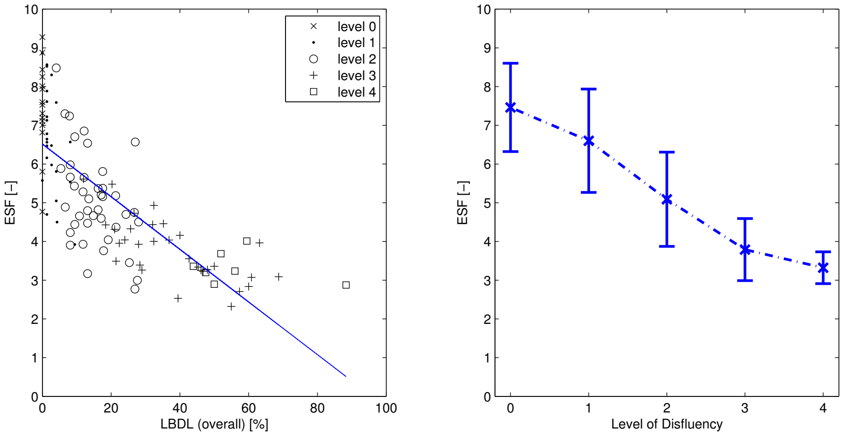


Figure 4.6: Comparison of the measurement ESF to subjective evaluation made with LDBL (left) and Kondas scale (right) in read recordings. The individual levels of disfluency are labelled by marks in the left figure. In the right figure, a marker indicates the mean value for the given level of fluency, error bars represent SD.

the BACD output), and multiplication constant 0.15. This setting reached correlation -0.77 and classification deviation $\Delta = 75$. The mean and SD of the measure in dependence on the speech difluency level is presented in Table IX and Figure 4.9. The measure values decrease with increasing level of speech disfluency. Table XXIII introduces the correlations of the selected measure setting with continuous LBDL scale. Correlation coefficients higher than 0.5 were observed in: individual LBDL categories -0.53 (ISR), -0.59 (MSUR), and -0.72 (FPWOAA); summary categories *repeated* (0.65), *fixed* (0.78) and *overall* (0.79).

Significant differences between levels mild, moderate, severe, and very severe disfluency were found by ANOVA analysis, significance level $p = 0.001$), see Table XI). The only groups fluent vs. mild disfluency were not distinguished. Fifty–five subjects were correctly classified in classification task, 55 participants with error 1 level, and 10 subjects with error of 2 levels.

4.3.9 Results of the NSI

Table VIII presents the correlation of the NSI measure and subjective evaluation (Kondas scale). We selected the most suitable setting of the measure according to this table. The results of the algorithm are presented further with this setting. The ALS algorithm uses removal of short segments of speech and silence, which was also subject of the setting. Because there exist a lot of possible combinations and the tables would become very large, we present only results for threshold 1000 ms. Further parameters of the algorithm setting are: $k = 6$ and multiplication constant 0.15. This setting reached correlation -0.78 and classification deviation $\Delta = 64$. Table IX and Figure 4.10 present mean and SD of the measure in dependence on the speech difluency level. The measure values decrease with increasing level of speech disfluency. Table XXIII introduces correlation of

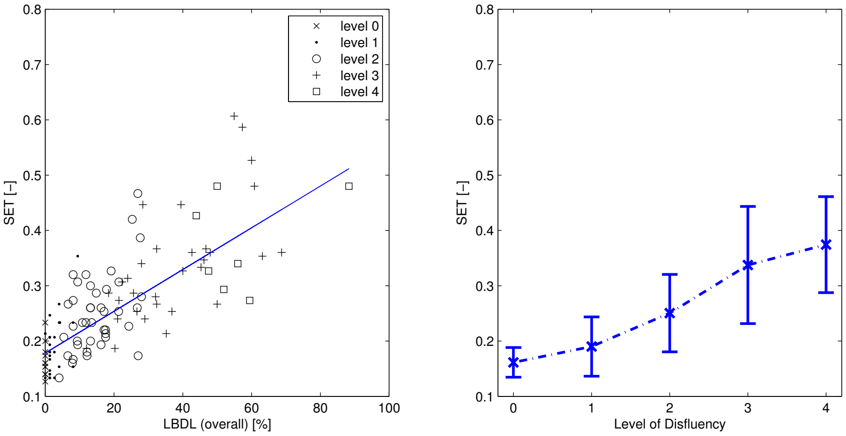


Figure 4.7: Comparison of the measurement SET to subjective evaluation made with LDBL (left) and Kondas scale (right) in read recordings. The individual levels of disfluency are labelled by marks in the left figure. In the right figure, a marker indicates the mean value for the given level of fluency, error bars represent SD.

selected measure setting with continuous LBDL scale. We observed correlations higher than 0.5 in: individual LBDL categories -0.52 (ISR), -0.50 (MSUR), and -0.84 (FPWOAA); summary categories *repeated* (0.63), *fixed* (0.85) and *overall* (0.82).

The ANOVA analysis found significant differences between levels mild, moderate, severe, and very severe disfluency at significance level $p = 0.001$), see Table XI). The ANOVA was not able to find significant difference between levels normal speech and mild disfluency. In the classification task, the NSI was able to classify 61 subjects correctly, 50 subjects with classification error 1 level, and 7 subjects differing from the correct level of 2 levels.

4.3.10 Results of the ROS

The measure ROS does not have various setting of the algorithm. The correlation coefficient of the measure and subjective rating by means of Kondas scale is -0.77 in read recordings. The value of ROS decreases with a higher level of disfluency, see Table IX and Figure 4.11. The measurement reached correlation -0.77 with merged evaluation of speech–language pathologists, the overall classification error $\Delta = 70$. The correlations of measure and all LBDL categories can be seen in Table XXIII. Correlation higher than 0.5 were observed for individual categories SR (-0.51), ISR (-0.56), MSUR (-0.56), and FPWOAA(-0.71) and summary categories *repeated* (-0.67), *fixed* (-0.75) and *overall* (-0.79).

The ANOVA analysis found significant differences between levels 1 vs. 2 vs. 3 vs. 4 at significance level $p = 0.001$), Table XI) shows all the figures. The ANOVA was not able to find significant difference between levels 0 (normal speech) and 1 (mild disfluency). The ROS measure classified correctly 54 subjects, 58 subjects with error of 1 level, and 6 participants were classified incorrectly

TABLE VII: The Pearson correlation and results of classification using the LDA (the deviation Δ from subjective evaluation) for all algorithm settings of the SCSI in comparison to the merged subjective evaluation, to set the algorithm the kth highest maximum and multiplication constant are used, the window for processing is 2 s.

	multiplication constant, correlation (deviation Δ)				
k	0.10	0.15	0.20	0.25	0.30
1	-0.75 (82)	-0.73 (88)	-0.68 (112)	-0.63 (120)	-0.58 (120)
2	-0.76 (79)	-0.76 (80)	-0.72 (93)	-0.67 (99)	-0.62 (99)
3	-0.75 (80)	-0.75 (83)	-0.72 (105)	-0.67 (120)	-0.62 (134)
4	-0.75 (76)	-0.77 (77)	-0.73 (101)	-0.69 (119)	-0.64 (123)
5	-0.74 (75)	-0.77 (76)	-0.74 (92)	-0.70 (111)	-0.66 (122)
6	-0.74 (87)	**-0.77 (75)**	-0.75 (92)	-0.71 (110)	-0.67 (123)
7	-0.72 (100)	-0.76 (77)	-0.75 (89)	-0.71 (114)	-0.67 (118)
8	-0.72 (97)	-0.77 (75)	-0.76 (89)	-0.73 (106)	-0.68 (118)
9	-0.71 (104)	-0.77 (78)	-0.76 (82)	-0.73 (105)	-0.69 (111)

TABLE VIII: The Pearson correlation and results of classification using the LDA (the deviation Δ from subjective evaluation) for all algorithm settings of the NSI in comparison to the merged subjective evaluation, to set the algorithm the kth highest maximum and multiplication constant are used, the time limit for successive removing of short speech and silence segments is set to 1000 ms.

	multiplication constant, correlation (deviation Δ)				
k	0.10	0.15	0.20	0.25	0.30
1	-0.77 (64)	-0.77 (72)	-0.74 (82)	-0.70 (89)	-0.66 (95)
2	-0.77 (65)	-0.78 (68)	-0.76 (69)	-0.73 (82)	-0.69 (93)
3	-0.76 (70)	-0.77 (68)	-0.76 (76)	-0.73 (87)	-0.69 (94)
4	-0.77 (72)	-0.78 (70)	-0.77 (70)	-0.74 (80)	-0.71 (93)
5	-0.76 (68)	-0.77 (70)	-0.77 (73)	-0.74 (80)	-0.72 (91)
6	-0.76 (69)	**-0.78 (64)**	-0.77 (75)	-0.75 (78)	-0.73 (93)
7	-0.76 (71)	-0.77 (66)	-0.77 (75)	-0.75 (82)	-0.73 (95)
8	-0.76 (70)	-0.77 (66)	-0.78 (74)	-0.76 (76)	-0.74 (87)
9	-0.76 (68)	-0.77 (64)	-0.78 (73)	-0.77 (77)	-0.74 (84)

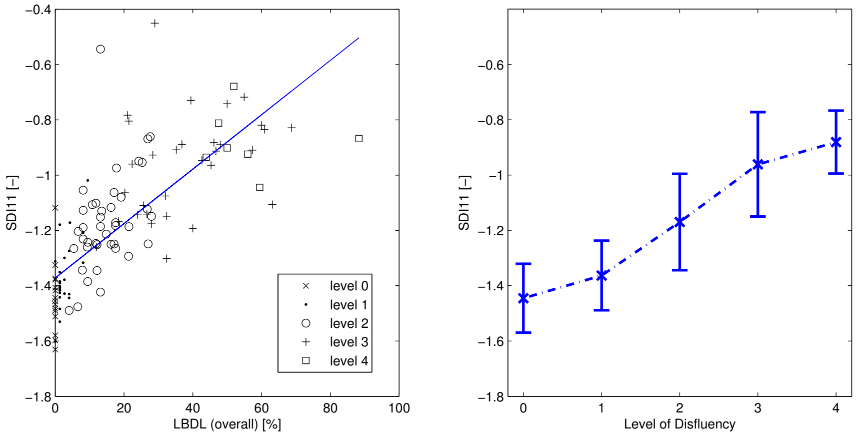


Figure 4.8: Comparison of the measurement SDI11 to subjective evaluation made with LDBL (left) and Kondas scale (right) in read recordings. The individual levels of disfluency are labelled by marks in the left figure. In the right figure, a marker indicates the mean value for the given level of fluency, error bars represent SD.

with error 2 levels.

4.3.11 Results of the RT

Results of the measure total reading time are presented to have a comparison with the standard measure used in clinical evaluation. There are no parameters to set. The correlation coefficient with subjective evaluation is 0.77. The mean and SD of the the measure are presented in Table IX and Figure 4.12, we can see that the measure values grow with rising level of disfluency. The correlations among the LBDL and the objective algorithm are in Table XXIII. The measurement correlates with merged evaluation of SLPs with coefficient 0.77, the overall classification error $\Delta = 64$. The correlation of measure and all LBDL categories can be seen in Table XXIII. Correlation higher than 0.5 were observed for individual categories SR (-0.54), ISR (-0.65), MSUR (-0.60), FPWOAA(-0.68) and SVB (0.60) and summary categories *repeated* (-0.75), *fixed* (-0.78) and *overall* (-0.86).

The analysis ANOVA found significant differences between levels mild vs. moderate ($p = 0.05$), moderate vs. severe vs. very severe disfluency at significance level 0.001, see Table XI). The ANOVA was not able to find significant difference between levels normal speech and mild disfluency. The total reading time classified 59 subject into correct level, 54 subjects were classified with error 1 level, and 5 subjects with 2–levels error.

4.4 Discussion, read recordings

The study of read recordings presents automatic and objective measures applied to the analysis of read audio recordings of stutterers. The main goal of this study is to find out whether the

TABLE IX: The mean $\overline{x}$ and standard deviation SD of fluency measures.

	ALS		RSE		REV		ESF		SET	
	$\overline{x}$	SD	$\overline{x}$	SD	$\overline{x}$	SD	$\overline{x}$	SD	$\overline{x}$	SD
Normal healthy speech (0)	0.19	0.06	0.11	0.02	14959	1274	7.46	1.4	0.16	0.03
Mild disfluency (1)	0.21	0.05	0.12	0.02	14371	1660	6.60	1.34	0.19	0.05
Moderate disfluency (2)	0.26	0.10	0.15	0.04	16330	2198	5.09	1.22	0.25	0.07
Severe disfluency (3)	0.38	0.11	0.20	0.06	18756	3180	3.79	0.80	0.34	0.11
Very severe disfluency (4)	0.60	0.26	0.27	0.07	22541	4388	3.32	0.41	0.37	0.09

	SDI11		SCSI		NSI		ROS		RT	
	$\overline{x}$	SD	$\overline{x}$	SD	$\overline{x}$	SD	$\overline{x}$	SD	$\overline{x}$	SD
Normal healthy speech (0)	-1.44	0.12	1.23	0.08	0.87	0.06	8.71	1.09	34.8	8.0
Mild disfluency (1)	-1.36	0.13	1.17	0.08	0.76	0.11	7.83	1.26	43.9	11.2
Moderate disfluency (2)	-1.17	0.17	1.06	0.09	0.59	0.14	6.47	1.15	58.2	14.2
Severe disfluency (3)	-0.96	0.19	0.95	0.09	0.42	0.15	5.16	1.08	92.0	25.7
Very severe disfluency (4)	-0.88	0.11	0.90	0.04	0.14	0.35	3.84	0.34	140.0	40.3

TABLE X: The Pearson correlation coefficients and the levels of significance (in parentheses when $p > 0.001$) for one selected setting of each measure in comparison to the LBDL descriptors and the merged subjective evaluation of speech pathologists.

	measure									
descriptor	ALS	RSE	REV	ESF	SET	SDI11	SCSI	NSI	ROS	RT
SR	0.38	0.43	0.37	-0.49	0.45	0.42	-0.47	-0.48	-0.51	0.54
ISR	0.44	0.66	0.55	-0.51	0.49	0.52	-0.53	-0.52	-0.56	0.65
MSUR	0.28	0.42	0.39	-0.54	0.56	0.52	-0.59	-0.50	-0.56	0.60
FPWAA	0.25	0.39	0.22	-0.46	0.47	0.46	-0.47	-0.38	-0.40	0.49
FPWOAA	0.73	0.65	0.66	-0.67	0.74	0.65	-0.72	-0.84	-0.71	0.68
SVB	0.28	0.37	0.32	-0.31	0.28	0.39	-0.34	-0.29	-0.36	0.60
repeated	0.49	0.67	0.57	-0.63	0.61	0.61	-0.65	-0.63	-0.67	0.75
fixed	0.72	0.69	0.65	-0.73	0.79	0.71	-0.78	-0.85	-0.75	0.78
overall	0.68	0.76	0.68	-0.76	0.77	0.74	-0.79	-0.82	-0.79	0.86
specialists (merged)	0.64	0.65	0.62	-0.77	0.66	0.75	-0.77	-0.78	-0.77	0.77

TABLE XI: Statistical significance by means of the ANOVA analysis with comparison between levels by the *post hoc* Bonferroni adjustment for read and spontaneous recordings.

	measure									
	ALS	RSE	REV	ESF	SET	SDI11	SCSI	NSI	ROS	RT
ANOVA F(4, 117)	27.97^{*}	25.72^{*}	23.14^{*}	42.84^{*}	23.48^{*}	38.07^{*}	43.33^{*}	48.89^{*}	44.07^{*}	60.08^{*}
0 vs. 1	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
1 vs. 2	NS	NS	p^{***}	p^{*}	p^{***}	p^{*}	p^{*}	p^{*}	p^{*}	p^{***}
2 vs. 3	p^{*}	p^{*}	p^{**}	p^{*}	p^{*}	p^{*}	p^{*}	p^{*}	p^{*}	p^{*}
3 vs. 4	p^{*}	p^{*}	p^{**}	NS	NS	NS	NS	p^{*}	NS	p^{*}

NS = not significant

$^{*}p < 0.001$, $^{**}p < 0.01$, $^{***}p < 0.05$

TABLE XII: Cross–correlation between all objective measurements on read recordings.

	measure								
measure	RSE	REV	ESF	SET	SDI11	SCSI	NSI	ROS	RT
ALS	0.75	0.69	-0.58	0.50	0.64	-0.61	-0.85	-0.70	0.69
RSE		0.77	-0.56	0.48	0.61	-0.59	-0.73	-0.69	0.75
REV			-0.55	0.42	0.65	-0.60	-0.69	-0.65	0.68
ESF				-0.86	-0.90	0.98	0.88	0.89	-0.77
SET					0.73	-0.88	-0.82	-0.77	-0.74
SDI11						-0.93	-0.82	-0.85	0.81
SCSI							0.90	0.88	-0.81
NSI								0.88	-0.81
ROS									-0.87

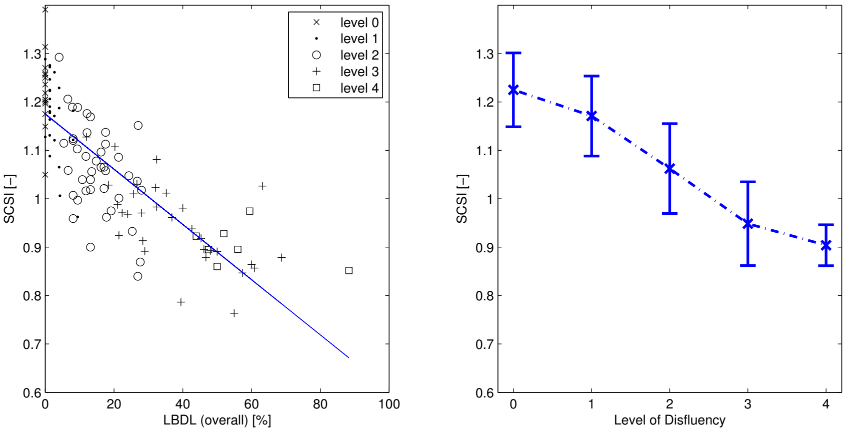


Figure 4.9: Comparison of the measurement SCSI to subjective evaluation made with LDBL (left) and Kondas scale (right) in read recordings. The individual levels of disfluency are labelled by marks in the left figure. In the right figure, a marker indicates the mean value for the given level of fluency, error bars represent SD.

automatic measurements are able to estimate the level of the speech fluency disorder in read speech. Eventually, to find the most appropriate settings of algorithms which would correspond to subjective rating in read recordings. A large part of the results presented here has been introduced in Lustyk *et al.* (2014).

The expert ratings are very important when comparing automatic measurements to subjective assessments. To have more information about the extent of the speech fluency disorder, two different evaluation scales were applied: the first is the modified Kondas's scale (Lechta and collective, 2004) and the second is the LBDL taxonomy (Teesson *et al.*, 2003) (see section Chapter 3 – Method). All 118 audio recordings of read speech were evaluated by two experienced phoniatric experts using the Kondas's scale. The Pearson correlation coefficient and Cronbach's alpha showed a very high relationship between both speech therapists. The second subjective evaluation was made by one evaluator who assessed all recordings by means of the LBDL taxonomy. The evaluation of 30 recordings for the second time and by another judge was used for intra– and inter–judge reliability. The same procedure was used in Goberman *et al.* (2010). The Pearson correlation coefficient showed a strong agreement between the original and the repeated evaluation using the LBDL, which is consistent with Teesson *et al.* (2003); Goberman *et al.* (2010), where very high intra–judge agreement was achieved. When we consult the inter–judge agreement, the lowest correlation (0.32) was found for superfluous verbal behaviors, the other categories of the LBDL report significant positive correlations. Because of the low correlation of the characteristic superfluous verbal behaviors, the results dealing with this characteristic are viewed carefully. When comparing the individual or merged evaluations by experts (Kondas's scale) and the descriptor *overall* of the LBDL, the conclusion can be adopted that these two evaluations report very strong relationships (the Pearson correlations for the individual experts and the merged evaluation with the LBDL

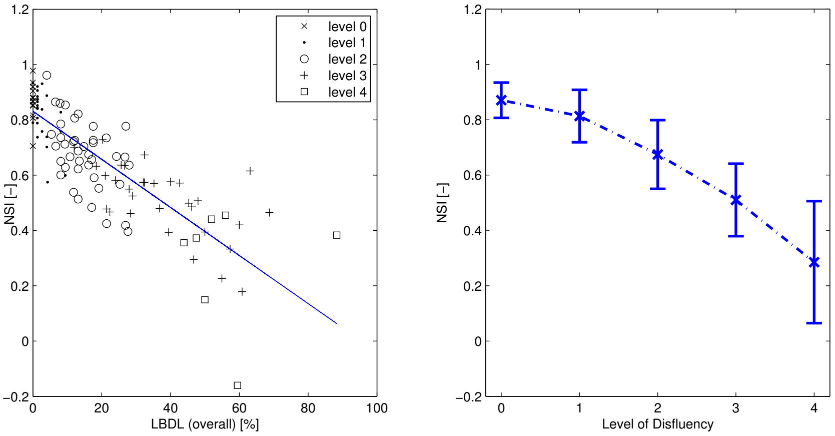


Figure 4.10: Comparison of the measurement NSI to subjective evaluation made with LDBL (left) and Kondas scale (right) in read recordings. The individual levels of disfluency are labelled by marks in the left figure. In the right figure, a marker indicates the mean value for the given level of fluency, error bars represent SD.

surpasses 0.9), these results of assessment suggests that the expert ratings are reliable and useful for the purposes of this experiment.

The main findings dealing with automatic measurements of audio recordings for the evaluation of speech disfluency can be expressed as follows. First, the measures are able to indicate the overall level of the speech fluency disorder. This finding is supported by the results where all measures have magnitudes of the correlation coefficient with two experienced speech pathologists (Kondas scale) higher than 0.62 and with the LBDL evaluation *overall* score exceeding 0.68. The highest correlation with Kondas scale were achieved by NSI (-0.78), followed by ESF, SCSI, and ROS (all -0.77). The comparative measure total reading time achieved very similar correlation (0.77 for speech experts, Kondas). Looking at the second set of subjective rating, the LBDL scale, the highest correlation we reached by NSI (-0.82), followed by ROS and SCSI (-0.79). The comparative measure total reading time (RT) reached correlation of -0.86.

The correlation are supported by results of classification using the Linear Discriminant Analysis with the leave–one–out cross–validation, when the selected setting of the NSI algorithm classified 61 subjects (52%) into the correct level of the Kondas's scale, 50 subjects (42%) with the classification error 1 (the estimated level by algorithm differs by one level from the subjective evaluation), and seven participant (6%) with classification error 2, the total deviation from the speech therapists evaluation is 64 (the lowest of all measures). For comparison, the total reading time classified 59 subjects correctly (50%), 54 subjects with the classification error 1 (46%), and five subjects (4%) with the classification error 2 (the total deviation from subjective evaluation is 64). Assessment of group differences confirms that the measure NSI is able to find statistically significant differences ($p < 0.001$) between the groups mild and moderate, moderate and severe, and severe and very severe. The measures ALS, RSE, ESF, SET, SCSI, and ROS can separate one group less, they

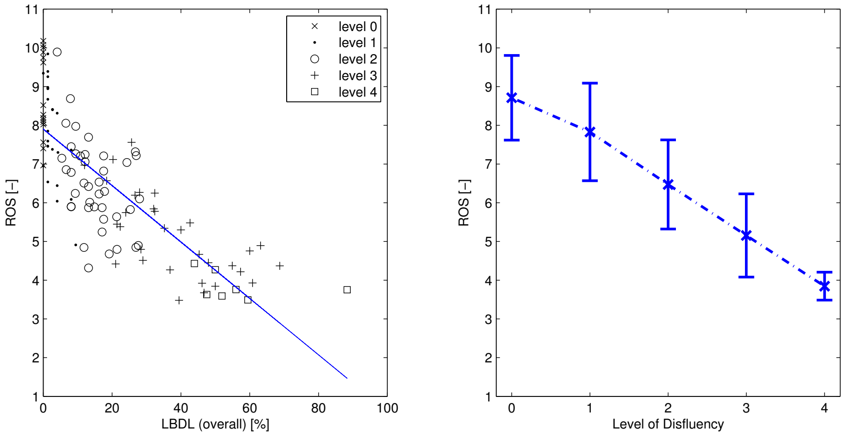


Figure 4.11: Comparison of the measurement ROS to subjective evaluation made with LDBL (left) and Kondas scale (right) in read recordings. The individual levels of disfluency are labelled by marks in the left figure. In the right figure, a marker indicates the mean value for the given level of fluency, error bars represent SD.

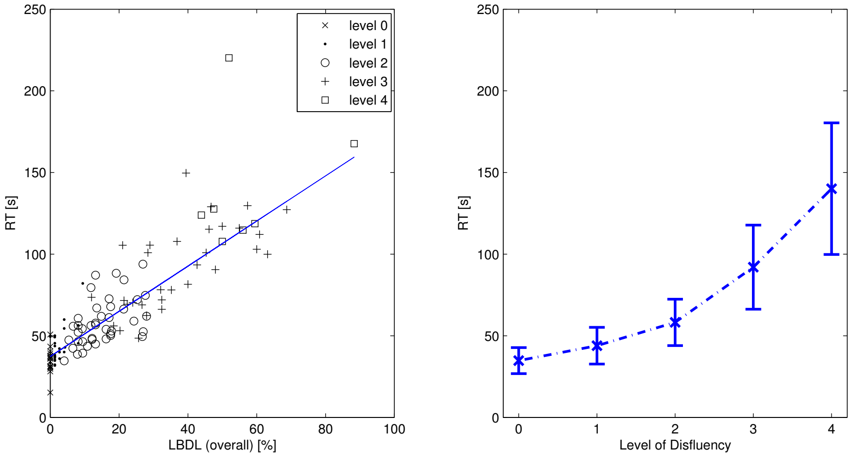


Figure 4.12: Comparison of the measurement RT to subjective evaluation made with LDBL (left) and Kondas scale (right) in read recordings. The individual levels of disfluency are labelled by marks in the left figure. In the right figure, a marker indicates the mean value for the given level of fluency, error bars represent SD.

are usually able to find significant differences between groups mild and moderate (1 vs. 2, $p < 0.001$) and moderate vs. severe (2 vs. 3). In comparison, the total reading time can differentiate levels moderate, severe, very severe ($p < 0.001$), and mild and moderate ($p < 0.05$). A major problem is distinguishing between normal fluent speech and mild disfluencies: no measure is able to recognize a statistically significant difference here (the similar phenomenon can be observed in classification). This is probably caused by the definition of the levels of the modified Kondas's scale, where the level 0 (normal healthy speech—without frequent signs of disfluency) and the level 1 (mild disfluency, up to 5% disfluent words) are very close. These two groups often overlap, because normal fluent speakers usually exhibit some signs of disfluencies (Johnson, 1961; Yairi and Clifton, 1972; Goberman *et al.*, 2010) and it is difficult to recognize the difference (Onslow *et al.*, 1992).

Second, some measures are able to describe individual or summary characteristics of the LBDL. The best results can be found for the fixed postures without audible airflow: five measures achieved a Pearson product–moment correlation higher than 0.7 in magnitude (the highest was 0.84 for the measure NSI). This finding suggests that a large part of the fluency evaluation in read speech may lie in the pauses, which is in line with Cucchiarini *et al.* (2000). Also Nöth *et al.* (2000) found pauses very important for automatic evaluation of stuttered speech. This finding led us to examine the cross–correlations between all characteristics of the LBDL and a strong relationship between *overall* and fixed postures without audible airflow was found (Pearson correlation of 0.81), which means that pauses constitute a large part of the subjective evaluation of read speech at least in this case. Thus, the measures which obtained a good agreement with the fixed postures without audible airflow have a strong relation with the *overall* subjective evaluation based on the LBDL. On the contrary, the total reading time has balanced results for all individual categories and manages to achieve a very good results for the *overall* score. The results for the other individual categories of LBDL do not reach those for pauses.

The total reading time was found distinctive for evaluation of disfluencies in read speech (Maier *et al.*, 2011). This measure was added to the experiment to have a comparison to other possibility of how to measure stuttering severity. It turned out to be a very good instrument for the evaluation even though it is very simple. The results are comparable and in some cases better than those of introduced algorithms and it could be possible to replace the algorithms with the total reading time. However, we would like to use these algorithms for evaluation of spontaneous speech where the utterances are mostly limited by time and the total time of a recording will not be as influential as in recordings of read speech.

Because of the basic method used for the larger part of the measures (the Bayesian abrupt spectral changes detector), it is appropriate to investigate the relationships between these measures, and a strong relationship can be expected, as in Cucchiarini *et al.* (2000). Examining these results (Table. XII), we can see that all the measures based on the BACD are strongly correlated (some of the coefficients exceed 0.9). In case of lesser correlation, there exists a high probability that a combined measure created from less correlated measures will be more successful. An experiment was carried out to see whether this is so, by a simple combination (summing up the normalised values of measures ALS, ESF, SCSI, and NSI), and a correlation coefficient of 0.8 with speech pathologists and 0.82 with the overall LBDL characteristic was achieved, which is higher than that for any single measure. A suitable combination and selection of measures could be a future focus of research.

A possible limitation of the algorithms is that they are able to describe fixed postures without audible airflow with good agreement and the other individual characteristics of the subjective evaluation, such as syllable and incomplete syllable repetitions or prolongations, to a limited extent. The results of this study for these symptoms do not reach the results of Nöth *et al.* (2000); Wisniewski *et al.* (2007a,b), but on the other hand, we are not aware of other studies concentrating on automatically measured temporal speech characteristics in stuttered speech which do not use Hidden Markov Models. The database could be considered a weak point of the present study, and especially its gender imbalance and its distribution of participants across the levels of the disorder. There were only a few participants at the very severe level, and most participants were located at the mild, moderate, or severe levels. However, the database reflects the situation in common practice (Yairi and Ambrose, 1999; Bloodstein and Bernstein Ratner, 2008).

An advantage of our methods could be the possibility to exchange one instrument for another. In other words, it provides the opportunity to apply other reliable abrupt spectral changes detectors or voice activity detectors. The BACD (Cmejla *et al.*, 2013) applied in this study was tested using synthetic and real speech signals (Bergl and Cmejla, 2007) or for stuttered speech (Bergl, 2010) in comparison to other divergence metrics with very good results. Algorithms, from simpler ones such as spectral or cepstral distance, to more complex ones, such as General Likelihood Ratio (Appel and Brandt, 1983) and Kullback–Leiber divergence, could be employed. A great advantage of BACD– and VAD–based measures could be that they are language independent, and there is no need for a training database as in the case of systems based on Hidden Markov Models. A possible language–independence of the algorithms is discussed in the section where the measures were applied on the recordings of German stuttering speaker. The measures could be considered for use in experiments with second language learning as in Cucchiarini *et al.* (2000, 2002); Maier *et al.* (2009c). Another VAD was also tested, one based on parameters (Atal and Rabiner, 1976) in cooperation with the Support Vector Machine making the decision about speech vs. silence. When this VAD was applied, very similar results were obtained.

Chapter 5

Effect of speaking task in disfluent speech

The study compares automated acoustic measures to behavioral measures of speech fluency in two different speaking conditions (reading and spontaneous monologue) in the speech of people who stutter. The main aim was to investigate the influence of the speaking tasks on participants' fluency levels. Participants were 92 adults (8 control speakers, and 84 stuttering participants). Analysis of recordings of their reading and spontaneous speech was undertaken. The analysis was carried out with two measurements (NSI and ROS), they were selected according to the results on read recordings.

Automatic and objective methods can efficiently support the diagnosis of fluency disorders and/or evaluation of therapy outcomes (Van Borsel *et al.*, 2003). The methods should be versatile and precise. Versatile methods could be applied in different speaking tasks that are typically used (reading, narration, or monologue). A precise method would accurately reflect a speaker's fluency in different speaking conditions.

Acoustic analysis (Healey and Ramig, 1986; Yaruss and Conture, 1993) and/or advanced methods of signal processing (Nöth *et al.*, 2000; Wisniewski *et al.*, 2007b; Esmaili *et al.*, 2016) may provide an objective and quantitative instrument for assessing speech fluency. In addition to traditional characteristics, Prosek and Runyan (1982) stated speaking rate and pauses as potential perceptual cues for listeners attempting to differentiate the speech of people who stutter from that of non–stutterers.

In a non–spontaneous task such as reading a passage, speakers produce ready–made text and therefore can pay more attention to articulatory planning (Levelt, 1989). The spontaneous monologue requires linguistic formulation and thus places greater demand on neural resources, which might suggest that stuttering frequency would be greater in spontaneous tasks than in reading (Armson and Stuart, 1998).

The effect of speaking task on the speech of people who stutter has been investigated in several studies. Johnson (1961) studied speaking rate and disfluencies in people who stutter and those with typical speech. Significant differences were noted between the two groups for rate of oral reading, spontaneous monologue, and picture description. Fluent speakers, as a group, showed higher rates and were more fluent and consistent than people who stutter. The study also found an overlap in distribution of disfluency measures for both groups in reading and spontaneous speech.

Reports by Ingham *et al.* (2012); Pinto *et al.* (2013) confirmed that fluent speakers generally have a higher speech rate (in syllables/minute) than stuttering participants for both speaking tasks. At the same time, they observed that the difference between the speech rate in the two tasks was more obvious in the control group and the speech rate of stuttering participants was very similar in both tasks. However, the literature is equivocal regarding the effect of speaking task on disfluency (Armson and Stuart, 1998). Some report that the frequency of stuttering is greater in spontaneous monologues (Johnson, 1961; Silverman, 1974), yet others Blood and Hood (1978) note no differences in stuttering frequency during oral reading and spontaneous speech.

Studies of speaking task effect have compared the speech of fluent speakers with the speech of people who stutter. Usually, the people who stutter have been treated as a homogeneous group with no further subdivision, except in a few cases (Blomgren and Goberman, 2008; Vanryckeghem *et al.*, 1999).

Automatically measured temporal characteristics have been used to evaluate second language learners' fluency in reading (Cucchiarini *et al.*, 2000), and further in reading and spontaneous speech tasks (Cucchiarini *et al.*, 2002). Spontaneous monologue was found to be the most sensitive of four tasks for eliciting an articulatory deficit in patients with early diagnosed Parkinson's disease (Rusz *et al.*, 2013).

This study aims to investigate whether automatic acoustic measures can estimate the level of speech fluency in oral reading and spontaneous monologue. We hypothesize that the measures will be more accurate for reading samples, and that it will be more difficult to determine the level of fluency in spontaneous speech samples. Second, we investigated whether speakers' performances differ under different speaking condition, and if severity subgroups (mild, moderate, severe) are equally affected. We assume that both control speakers and speakers who stutter will be less fluent and speak at a slower rate in the spontaneous task and the difference between the severity subgroups will fade. We hypothesized that speakers at all levels of fluency will be affected equally by the speaking conditions.

5.1 Method

5.1.1 Participants and recordings

A group of 92 (16 women and 76 men) Czech native speakers with different ages and levels of speech fluency disorder were recruited for the study of speaking task effect. For each of the 92 participants both the read and spontaneous recordings are available.

The mean age 15.8 yr [±standard deviation (SD), 7.6 yr], ranging from 8 to 49 years old. Eight speakers (one female and seven males) were subjects without any speech fluency disorder (mean age 25.7 yr [±SD, 5.5 yr]) and 84 speakers with speech fluency disorder (15 women and 69 men) were at a mean age of 14.8 yr [±SD, 7.1 yr]. See the distribution of participants' age in Figure 5.1.

The reading task is the same as described in previous study in the thesis. Spontaneous monologues were elicited through a picture description task. Two pictures were shown of a small town in two different situations, the first one is positive (clean streets, calm people), the second one is negative (people and streets in a mess). The duration of each recording was intended to last 90 s, with half of the allotted time for description of one picture. If the subject spoke longer, the recording was shortened as close as possible to 90 s. The mean duration of recordings is 91.7 s [±SD, 4.9 s].The pictures can be seen in section Appendices B and C.

The recording took place in the clinic consulting room. All speech signals were recorded with the sampling frequency 44 kHz, and the signals were subsequently down–sampled to 16 kHz for analysis.

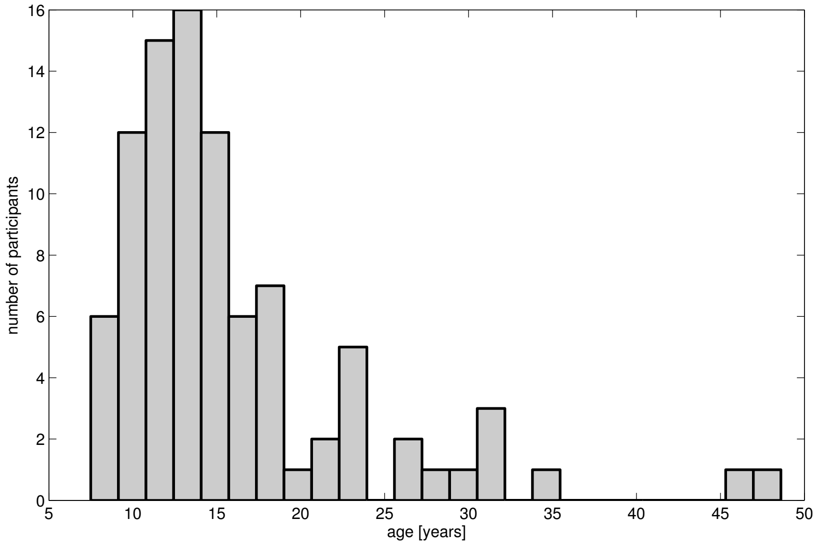


Figure 5.1: Histogram, distribution of participants' age in the study of speaking task effect (92 participants).

5.1.2 Evaluation of recordings performed by speech language pathologists

The evaluation was obtained by means of the Lidcombe Behavioral Data Language of Stuttering (LBDL) Teesson *et al.* (2003). The LBDL considers seven descriptors of stuttered speech characteristics: syllable repetition, incomplete syllable repetition, multisyllable unit repetition, fixed posture with audible airflow, fixed posture without audible airflow, superfluous verbal behaviors, and superfluous non–verbal behavior. The superfluous non–verbal behavior was not counted because the behavior is only detectable through video recordings, which were not available. For the evaluation procedure, SLPs listened to the recordings and wrote down the number of disfluencies in individual categories followed by the total number of words. The sum of individual categories together formed the overall category which was subsequently normalized by the number of words in the recording, yielding a continuous scale (0 – 100%).

SLPs assessed recordings independently. They focused only on speech disfluency and did not report other behaviors associated with a comprehensive assessment of stuttering. For the assessment using the LBDL scale, one evaluator assessed the reading recordings. Five evaluators used the LBDL to assess each spontaneous recording. After the analysis of inter-rater reliability, the evaluations of spontaneous speech were merged for each recording to give a mean value.

Apart from the LBDL evaluation scale, we considered speech rates in [words/min] and [phonemes/min] for both speaking tasks. The first was obtained from the number of words manually counted during the LBDL evaluation procedure (mean number of words from evaluators) and divided by the duration of the recording. The second comes from the number of phonemes in verbatim transcriptions of recordings.

The distribution of subjective evaluation on read recordings is depicted in Figure 5.2, on spontaneous recordings in Figure 5.3. See also Figure 4 in section Appendices G where the histograms for each of the five judges and the merged evaluation on spontaneous recordings is shown.

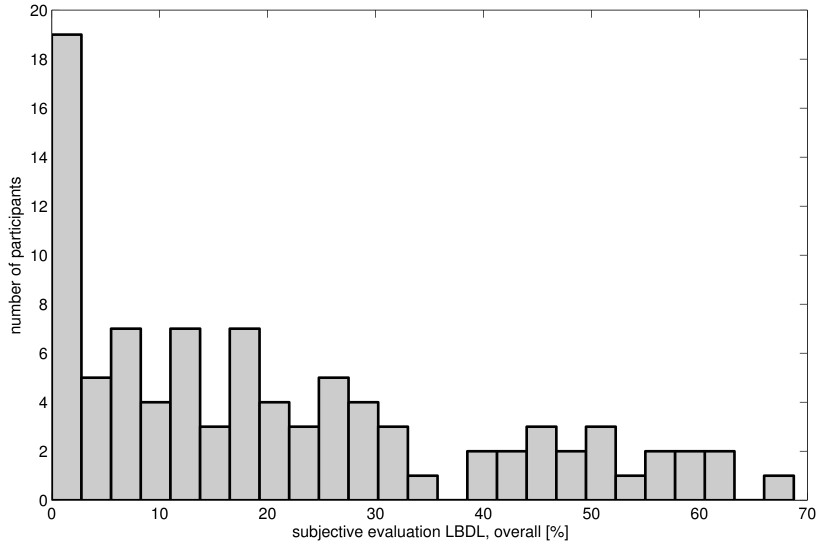


Figure 5.2: Histogram, subjective evaluation made by means of the LBDL on read recordings (92 participants) for study of speaking task effect.

5.1.3 Measures of speech fluency

The experiment closely follows previously published study Lustyk *et al.* (2014) that focused on read recordings. The function of measurements was describe earlier in the thesis, therefore only the list of measures follows:

- The number of spectral changes in speech segments (NSI)
- The rate of speech (ROS)

5.1.4 Statistics

The Pearson product-moment correlation reveals agreement between automatic measures and evaluation of SLPs. The difference between read and spontaneous recordings is evaluated by means

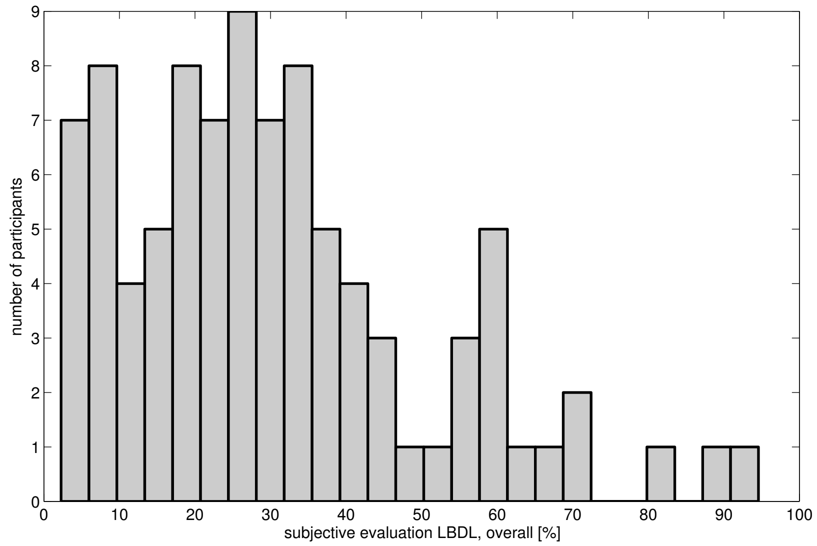


Figure 5.3: Histogram, merged subjective evaluation made by using the LBDL on spontaneous recordings (92 participants) for study of speaking task effect.

of non-parametric pair comparison (the Wilcoxon signed rank test (Wilcoxon, 1947)). Inspired by Spaniel *et al.* (2016), we applied generalized linear model (McCullagh and Nelder, 1989) to answer the question, if there is a different effect of the speaking task on speech fluency. The automatic measurements were separately treated as dependent variables, while assessments of SLPs were modeled as covariates and the tasks represented fixed factor. The linear model included intercept, and factors of task, SLPs assessment, and interaction between speaking task and assessment. The handling and statistical data analysis was performed in Matlab (Matlab R2011b, The Mathworks, Inc. Natick, Massachusetts, USA) and SPSS software (IBM SPSS Statistics 22, IBM Corp., Armonk, New York, USA).

5.1.5 Details of the generalized linear model

To analyze the effects of speaking condition (task) and evaluation based on judgments of speech-language pathologists on automatic measurement we applied the generalized linear model. The choice of Gamma turned out to be appropriate (data are continuous and nonnegative). The characteristics of the measurements suggested that the relation between them and SLPs evaluation is linear (link function). The model equation is:

$$\begin{aligned} measurement = {} & \beta_0 + \beta_{CONDITION} CONDITION + \beta_{LBDL} LBDL \\ & + \beta_{CONDITION,\ LBDL} CONDTION\ \ LBDL + \epsilon \end{aligned} \tag{5.1}$$

where *measurement* in the response variable (in our case automatic measurements NSI, ROS and manually measured speech rates), β_0 represents the model intercept, the *CONDITION* are

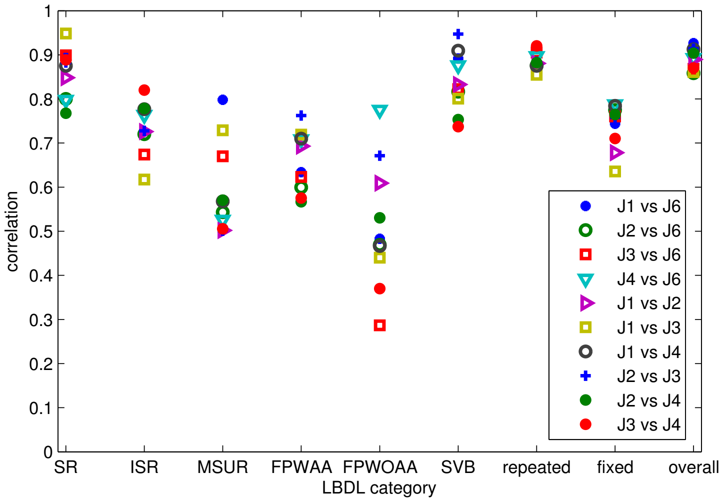


Figure 5.4: Interjudge agreement of 92 spontaneous recordings in the experiment with speaking task effect, evaluation scale – LBDL, 5 judges (J1, J2, J3, J4, and J6). Correlation coefficient between individual judges.

predictor variables for the speaking conditions (read and spontaneous), $\beta_{CONDITION}$ are coefficients for those predictors (dependence of the response on speaking conditions), $LBDL$ are predictor variables for the covariate (i.e. evaluation of speech-language pathologists, LBDL), β_{LBDL} are coefficients for those predictors (effect of the SLPs evaluation on dependent variable), the equation element $\beta_{CONDITON,\ LBDL} CONDITION\ LBDL$ represents interaction between speaking condition and evaluation of speech-language pathologists, $\beta_{CONDITON,\ LBDL}$ are coefficients for interaction. The residuals are described by ϵ. The applied model handled the measurements separately as dependent variables. The linear model included intercept, and factors of task, SLPs assessment, and interaction between speaking task and assessment.

When modeling, we proceeded from a simple model with only main effect of evaluation and speaking conditions, adding interaction between model parameters, while looking for an improvement in model description of the data.

5.2 Results

5.2.1 Reliability of recording evaluation made by speech-language pathologists

The group of 92 read recordings were assessed by one evaluator. To obtain the intra- and inter-judge agreement a set of 27 recordings was selected. These recordings were evaluated for the second time by the first evaluator and also once by the second evaluator. Similar procedure was used in Goberman et al. 2010. The participants selected for the reliability assessment covered the entire spectrum of disfluency. The correlation coefficients for intra- and inter-judge agreement in the

TABLE XIII: The Pearson correlation coefficient and the level of significance (*when $p > 0.001$) for automatic measures compared with the evalution of speech pathologists (overall LBDL descriptor) and manually measured speech rate [words/min] and [phonemes/min] for both speaking conditions. The Pearson correlation coefficients for manually measured speech rates in comparison to speech language pathologist assessment.

	measure							
	NSI		ROS		speech rate [words/min]		speech rate [phonemes/min]	
evaluation/task	read	spont	read	spont	read	spont	read	spont
LBDL	−0.83	−0.52	−0.80	−0.52	−0.83	−0.72	−0.82	−0.58
speech rate [words/min]	0.87	0.61	0.98	0.73	-	-	0.98	0.92
speech rate [phonemes/min]	0.87	0.62	0.98	0.71	-	-	-	-

overall LBDL category reached 0.98 and 0.94 (p< 0.001), respectively.

Five judges independently evaluated all 92 recordings. The range of inter-judge agreement in the overall LBDL among the judges varied between 0.86 and 0.93 (p< 0.001), see Figure 5.4. Twenty-three recordings (25% of the 92 spontaneous recordings) were assessed twice by one evaluator to determine intra-rater agreements, with a correlation of 0.98 (p< 0.001).

The results of intra- and inter-rater reliability showed a good level of agreement. Therefore, we decided the evaluation made by SLPs were appropriate for the purpose of this experiment.

5.2.2 Analysis of agreement between measurements and evaluation of speech–language pathologists

Because the results from read recordings were presented in the previous study (Lustyk *et al.*, 2014), the findings from spontaneous recordings are emphasized here. It is relevant to point out that the results from read recordings in the previous study were obtained for 121 participants (92 recordings are used in the current study), and hence the results may slightly differ.

The measures NSI and ROS decrease as the level of disfluency grows, see the Figure 5.5 that shows the automated measures as a function of SLPs' evaluation. The correlations of the automatic measurements are given in the Table XIII. The correlations in spontaneous task reached -0.52 (ROS), -0.52 (NSI), p< 0.001. The higher correlations were found in read than in spontaneous recordings (-0.80 for ROS, -0.83 for NSI).

Manually measured speech rates also decrease with increasing level of speech disfluency. The speech rates highly correlate with assessment of SLPs, the coefficients for manually measured speech rate in [word/min] are -0.83 and -0.72 in read and spontaneous conditions, respectively. Speech rate measured in [phonemes/min] yielded correlation -0.82 and -0.58 for read and spontaneous recordings.

Because this study concentrates only on two automatic measures (NSI and ROS), it is interesting to see the results of other measurements in spontaneous spech. Therefore, the results of comparison of all measeurements and all LBDL descriptors (merged evaluation of speech-language pathologists) in spontaneous recordings are presented in Appendices H.

5.2.3 Analysis of the speaking task effect, pair test

A Wilcoxon signed-rank test showed that the different speaking condition (read and spontaneous) elicit a statistically significant change in both automatic measurements of speech disfluency in individuals with and without stuttering (for ROS, Z= -7.476 and NSI, Z= -5.681; p< 0.001). See Table II. for detailed description of the comparison by the Wilcoxon signed rank test.

Figure 5.6 completes the results of pair test with the graphical representation of the difference between fluency in read and spontaneous recordings. Even though that the difference was statistically significant we observe a trend in the data. The difference between fluency in read and spontaneous conditions decreases as the level of speech disfluency increases.

Also, Wilcoxon signed-rank test revealed statistically significant difference for manually measured rate of speech in [words/min] (Z= -5.685, p< 0.001) and in [phonemes/min] (Z= -6.156, p< 0.001) in the same group of individuals as automatic measurements. The Figure 5.6 shows results for the measure NSI, the figure supporting the results of pair test for ROS measurement and both speaking rates can be seen in Figure 5.7 .

TABLE XIV: The mean $\overline{x}$ and standard deviation SD and median values for automatic measures in read and spontaneous recordings. Wilcoxon signed rank test for measurements of speech fluency to looks for difference read and spontaneous, descriptive statistics for 92 participants, automatic measures and manually measured speech rates.

		measurement			
		NSI	ROS	speech rate, manually [words/min]	speech rate, manually [phonemes/min]
mean	read	0.74	6.32	77.8	379.6
	spont	0.66	5.10	62.8	310.6
SD	read	0.15	1.71	21.0	143.4
	spont	0.11	1.33	22.8	115.3
median	read	0.74	6.23	78.9	377.58
	spont	0.66	4.96	58.6	285.0
Z		$-5.681^{*,a}$	$-7.476^{*,a}$	$-5.685^{*,a}$	$-6.156^{*,a}$

*Asymt. Sig. (2-tailed),$p < 0.001$
a - based on positive ranks

5.2.4 Analysis of factors influencing disfluency: model results

The linear model revealed that there is a significant effect of the SLPs' assessment (LBDL) on automatic measurements NSI and ROS (p< 0.001). Also, the effect of the task on fluency measured by means of automatic measurements was found to be significant (p<0.001). At the same time, the coefficient for interaction between LBDL evaluation and speaking task was statistically significant (p< 0.001), showing that the trend in both speaking conditions is different. This also confirms the trends observed in the figures.

The models build for both speech rates convey similar information as the models for automatic measurements (significant effect of speaking condition, SLPs evaluation and interaction between them).

Detailed analysis of the model parameters follows on the example of the NSI measurements. The model intercept for spontaneous speaking condition is β_0= 0.738 (when a participant has the evaluation from speech-language pathologists equal to 0 – fluent speech, her/his NSI value will be in average 0.738). The NSI value in reading task is shifted of 0.130 (coefficient $\beta_{CONDITION}$) compared to spontaneous conditions for fluent participants [LBDL= 0] (in overall, model estimate for fluent participant of the NSI measure value in reading is 0.868).

The trend in the spontaneous speaking condition (coefficient β_{LBDL}) is estimated to be -0.003. The sign represents the direction of the change, here, with increasing level of speech disfluency in LBDL the value of NSI decreases. A unit change in LBDL evaluation (fluency measured in LBDL better of 1%) cause a drop of -0.003 in NSI measurement value.

The model parameter for interaction between task and speech-language pathologists' evaluation indicates that the trends in reading and spontaneous tasks are different. The coefficient $\beta_{CONDITION,\ LBDL}$ is equal to -0.004, therefore, the slope in reading conditions is steeper than in spontaneous task (in overall, slope in reading task is -0.007). Suggesting that the fluency of participants at different levels is affected differently by speaking task, participants' ability to speak in spontaneous task is affected less by their level of speech disfluency than in reading. According to model, for example, the NSI estimate for a fluent participant (LBDL = 0) in read conditions is 0.868, while in spontaneous conditions it is 0.738 (read > spontaneous). As the disfluency decreases the relations changes. The NSI estimate for a participant with LBDL = 60% (severe disfluency) in read would be 0.448, whilst in spontaneous 0.558 (read < spontaneous).

The slopes in both speaking conditions support conclusion obtained with correlations, that the measurements are able to estimate level of speech fluency in both speaking tasks. The same conclusion as for NSI measurements can be drawn for other measurements (ROS, speech rate manually in [words/min] and [phonemes/min]), the change would be in parameter values (see Table XV), the interpretations remain the same.

TABLE XV: Estimates of model parameters for automatic and manual measurements.

		parameter value/measurement			
model parameter	p	NSI	ROS	speech rate, manually [words/min]	speech rate, manually [phonemes/min]
intercept [β_0]	<0.001	0.738	6.105	83.26	391.2
task [β_1]	<0.001	0.130	1.546	16.73	93.17
evaluation LBDL [β_2]	<0.001	-0.003	-0.033	-0.678	-2.568
interaction task*LBDL [β_3]	<0.001	-0.004	-0.031	-0.415	-2.451

5.3 Discussion

In the current study, we analyzed the effect of oral reading and spontaneous monologue in a sample of typical and disfluent speakers. The analysis was carried out using automatic objective measurements. Several previous studies have considered stuttering individuals to constitute a homogenous group. The current study broadens information about the speaking task effect with respect to the level of speech fluency. The objective measures and their comparison with evaluation made by experienced SLPs show that disfluency level has a significant impact on how speakers perform on different speaking tasks.

5.3.1 Speech-language pathologists' evaluation of recordings

A reliable expert rating is essential to verify if a measure is suitable to effectively support the SLPs evaluation (Cordes and Ingham, 1994). The LBDL taxonomy (Teesson *et al.*, 2003) was employed to examine how the automatic measures are able to describe fluency in disfluent recordings. The set of evaluations in both speaking conditions showed very high inter- and intra-rater reliability, which fits with Teesson *et al.* (2003); Goberman *et al.* (2010). Correlations did not drop below 0.86. The evaluation was considered to be adequate for the purpose of the study.

5.3.2 Measures as indicators of speech fluency

The comparison of automatic measurements with evaluation made by speech-language pathologists showed that the measurements are able to detect fluency disorder and describe level of speech fluency of people who stutter. Measurements indicated that the control participants are more fluent and speak with faster rate than the participants with speech disfluency. These findings agree with Johnson (1961); Ingham *et al.* (2012); Pinto *et al.* (2013). Fluency level of control and stuttering participants overlap to some extent. The situation is affected by the fact that several of the control speakers in this study exhibited disfluencies, it is common that fluent speakers produce disfluencies (Roberts *et al.*, 2009).

The findings are supported by correlation coefficients for the overall LBDL category in comparison to automatic measurements, they were -0.83 for oral reading and -0.52 for spontaneous task. Also, the applied generalized model confirmed that the measurements reflect speakers' level of fluency.

To compare the automatic measurements with standard measure, the speech rate in [words/min] and [phonemes/min] were applied. Both speech rates highly correlate with rating of speech-language pathologists, confirming results of Prosek and Runyan (1982). Strong positive correlations were also found between automatic measures and manually measured speech rates. The comparison of automatic measurements results and inter-rater reliability of speech-language pathologists' evaluation suggests that the automatic measurements do not reach the correlations of inter-rater reliability (mainly in spontaneous speaking task). However, the standard measurements (manually measured speech rates) are as well at the similar level as the automatic measurements.

Cucchiarini *et al.* (2002) assessed the fluency of second language learners by means of automatic measurements and noted lower correlations in a spontaneous speech task compared with reading. Our results show a similarly reduced correlation. Possible reasons are that the SD of the entire participants group in the spontaneous task is smaller compared with the reading task (values are more compressed, trend in the spontaneous data is less distinct, Figure 5.5, Table XIV). Consequently, the correlations decrease and differences between individual levels of disfluency fade.

The measurements are not only ableto describe general conclusions for the entire group of speakers (especially with hlp of Wilcoxon signed rank test and generalized linear model), but they can capture individal differences of participants. It can be seen mainly in the fires with depicted differences for all participants. For example, we can see that there are participants who have greater difference between read and spontaneous recording and aslo those who have small difference. Capturing these characteristic features of fluency for individual participants could be benefitial for example in treatment.

5.3.3 Effect of speaking task

The pairwise comparison demonstrated statistically significant change in automatic measurements induced by speaking conditions (read and spontaneous task). However, the data exhibit a trend (see Figure 5.6) and it suggests that the speaking tasks has a different effect on participants at different levels of speech fluency. The fluent participants and participants at lower levels of speech disfluency in general display that their fluency is better in reading than in spontaneous task. While most of the participants with severe disfluency showed the opposite. Additionally, the distinct effect of speaking conditions on participants' fluency was confirmed by linear model, the participants' ability to speak in spontaneous task is affected less by their level of speech disfluency than in oral reading.

The results do not confirm the hypothesis formulated at the outset of the project, i.e. that participants of all levels of fluency would be affected to the same extent. With help of findings of Johnson (1961); Pinto *et al.* (2013) we observed a disproportionate speaking task effect on participants at different fluency levels. Two studies relate effect of speaking task and speech rate to level of stuttering, but are in contradiction. One report (Blomgren and Goberman, 2008) found that changes in speech rate affect speakers with less severe stuttering to a greater extent, and a smaller effect was observed among severe stutterers. However, Vanryckeghem *et al.* (1999) describes an opposite, smaller effect for those who stuttered the least and a higher effect for people with severe stuttering. Studies Vanryckeghem *et al.* (1999); Blomgren and Goberman (2008) have manipulated speech rate. However, speech rate is habitual and the complexity of the task (reading vs. spontaneous) is changed in this study. The question arises as to whether their conclusions are applicable to our case, they are likely caused by differences in applied methods. Nevertheless, our results tend to the findings of Blomgren and Goberman (2008), i.e., that aspects related to speech rate influence severely disfluent participants to a lesser extent than for more fluent speakers.

Accordingly, we have tried to confirm the reported phenomenon for automatic measures and conformity with (Johnson, 1961; Prosek and Runyan, 1982) via manually measured speech rates. The results were very similar, thus the analysis acknowledged the trends and conclusions found for automatic measurements.

A similar phenomenon suggesting that performance of patients with severe speech impairments is not substantially influenced by the nature of the speaking task was observed in patients with Huntington's disease (HD) (Rusz *et al.*, 2014). In particular, the speech performance of patients with HD tends to be similar in reading and monologue(Rusz *et al.*, 2014), allowing reading-based metrics to separate HD participants precisely from control individuals.

5.3.4 Limitation of the study

We are aware of limitations of the study. In the first place, there is different numbers of participants at different levels of speech disfluency. A great part of the participants (about 80 participants in both tasks) has the overall LBDL score up to 50%, the rest of the disfluency spectra is covered by the smaller number of participants. It reflects the situation in common practice (Yairi and Ambrose, 1999; Bloodstein and Bernstein Ratner, 2008): there are fewer people with very severe stuttering.

The unbalanced number of speech-language pathologists that evaluated recordings in reading (1) and spontaneous (5) conditions was used. As we perceive the spontaneous task as more difficult to evaluate, due to the study design and inappropriate timing of the project we recruited more judges for spontaneous task. We understand that this could confine the study. Therefore, a very reliable evaluation LBDL scale (Teesson *et al.*, 2003; Goberman *et al.*, 2010) was employed. The analysis of inter- and intra-judge agreement confirmed a high reliability of the evaluation.

One limitation is linked to the capability of our automatic measurements to describe fluency related to speech rate. If they were applied in clinical evaluation, they would have to be used in conjunction with automatic measurements considering other symptoms of stuttering, for example repetition (Nöth *et al.*, 2000) and prolongation (Wisniewski *et al.*, 2007b) to obtain a comprehensive picture of disfluency. This may suggest future research that could focus on individual symptoms of stuttering in both speaking tasks with respect to the level of speech fluency.

5.4 Conclusion

The study described that the level of speech fluency in stuttering can be estimated by means of automatic acoustic measures in oral reading and spontaneous speech, though better conformity with subjective evaluation was achieved in the reading task. The measurements indicated that participants with varying degrees of disfluency are differently affected by the speaking tasks. Specifically, typical speakers (with no fluency problems) and those with mild and moderate disfluency speak slower in spontaneous monologue compared to reading. While most of the severe stuttering participants display the opposite, meaning that these participants reach on average similar or slightly better fluency. The findings suggest that the measurements are able to reflect speech fluency and hence may efficiently assist the existing methods by offering an automatic and objective evaluation of speech disfluency.

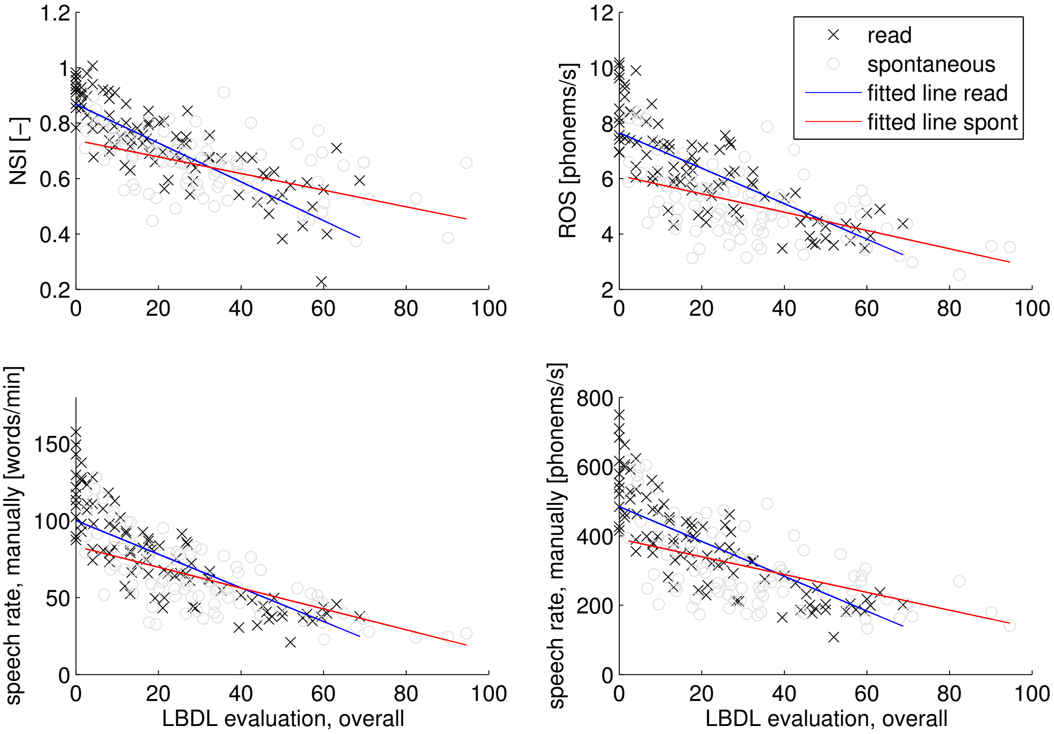


Figure 5.5: Comparison of fluency measures and evalution of SLPs – the overall LBDL category. The upper two graphs show automatic algorithms values for each participants are marked with x (read) and o (spontaneous), fitted linear lines show the trend the speaking tasks (blue and red color for read and spontaneous speaking conditions, respectively). The lines are in accordance with the applied linear model. The lower part presents the comparison of manually measured speech rate in [phonemes/sec] and [words/sec].

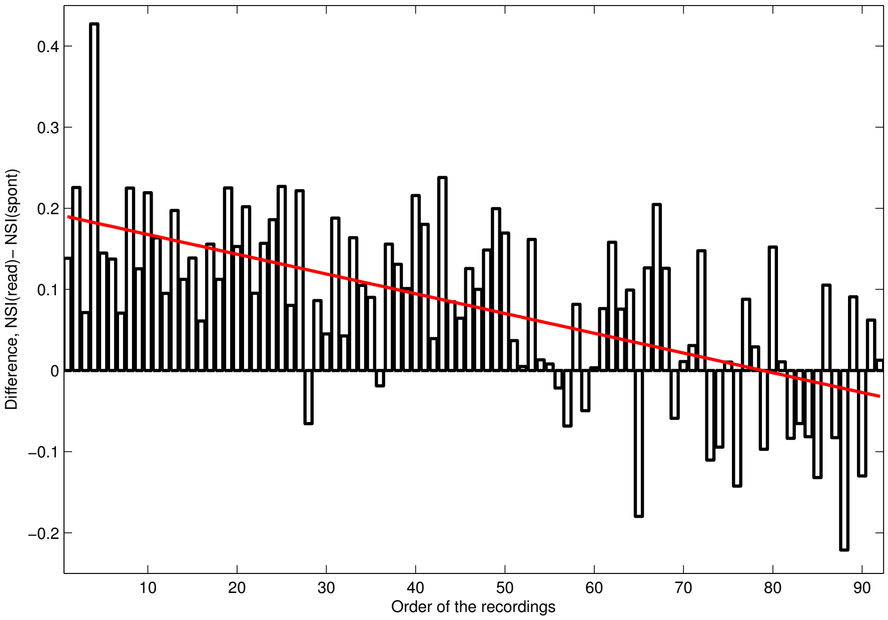


Figure 5.6: The difference between values of automatic measurement NSI in read and spontaneous recording for all 92 participants. The participants are sorted according to the level of speech fluency in LBDL from the fluent speakers (left side) to the most disfluency participants on the right side. Positive value in NSI difference means that a participant was more fluent in read recording than in spontaneous. Negative difference indicates that a person was more fluent in spontaneous speaking condition than in reading.

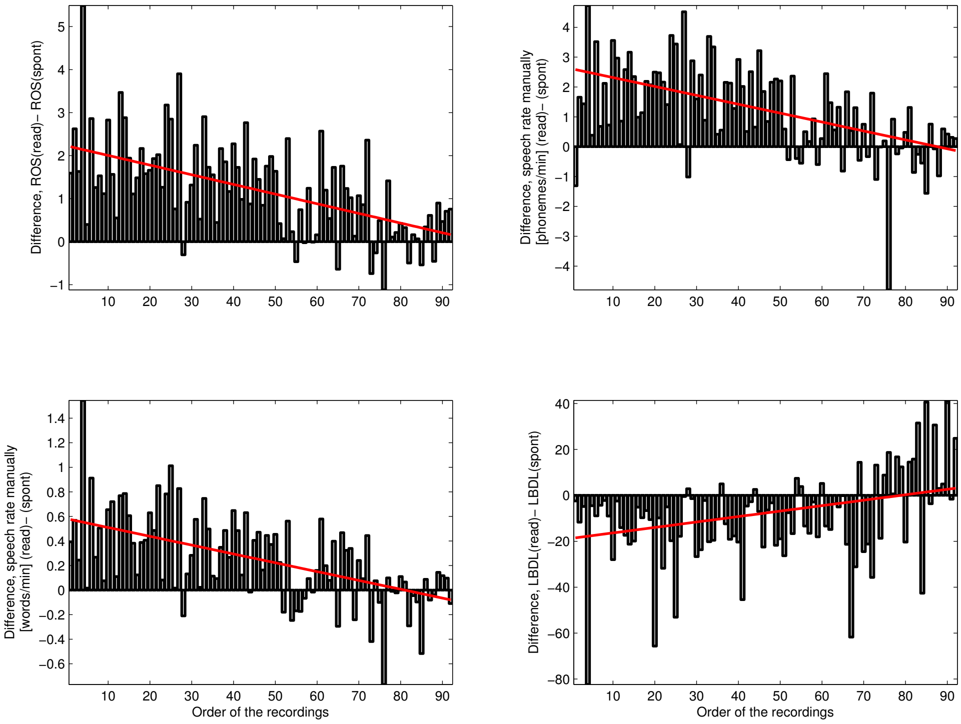


Figure 5.7: The difference between values of automatic measurement ROS, manually measured speech rates, and LBDL in read and spontaneous recording for all 92 participants. The participants are sorted according to the level of speech fluency in LBDL from the fluent speakers (left side) to the most disfluency participants on the right side. Positive value in difference of ROS and manually measured speech rates means that a participant was more fluent in read recording than in spontaneous. Negative difference indicates that a person was more fluent in spontaneous speaking condition than in reading. For LBDL the meaning is opposite.

Chapter 6

Analysis of read recordings of German stuttering subjects

The third study describes experiments where automatic acoustic algorithms initially intended to be used on Czech stuttering speakers were applied to recordings of German stuttering speakers. Our measurements mainly use prosodic characteristics to describe speech fluency. In some aspects they are very similar to methods used for unsupervised speech segmentation (Scharenborg *et al.*, 2010) or methods looking for differences in speech of native speakers and people who are non–native speakers (second language learners) (Cucchiarini *et al.*, 2000, 2002; Scharenborg *et al.*, 2012). Therefore, the basis of the measurements could enable the evaluation of fluency in different languages.

There are differences in speech characteristic in different languages. Prosodic characteristics such as rhythm, stress and intonation in speech convey important information regarding the identity of the spoken language (Mary and Yegnanarayana, 2008). Therefore we could not expect as good results as the language–dependent algorithms.

The nature of the two languages is different: Czech is a Slavic language, and German is a Germanic one. In particular, the differences between Czech and German languages are lie in: German has larger phonetic inventory than Czech, it also has more vowels (Matoušek *et al.*, 2002). For consonants, there are more consonantal phonemes in Czech language, and Czech consonants are influenced by the characteristic voiced/unvoiced much more than German. Some differences could be found in prosodic features of both languages. The main difference lies in stress. However, there exist some similarities between languages in general, therefore it is possible to find language–independent features, for example: pauses, fundamental frequency, and others (Vaissiere, 1983). For example, Czech and German languages distinguish short and long vowels.

The aim of this experiment is to find if it could be possible to perform analysis of speech disfluency independently on language, here demonstrated on read recordings of German speakers. One standard measure and four automatic measures of disfluency are applied to read recordings and their results are discussed. The questions, we hope to answer, were: Are the measures able to describe the level of speech fluency in different languages? If yes, how good they are? Is it possible to classify fluent and disfluent recordings? How would the comparison between language–dependent and independent measurements turn out?

6.1 Method

6.1.1 Recordings and participants

The recordings of German native speakers were collected at the Pattern Recognition Lab, Department of Computer Science 5 at the Friedrich–Alexander–University Erlangen–Nürnberg.

The database contains 34 recordings – 16 signals of stuttering speakers, 18 recordings of speakers who do not have problems with fluency. All the subjects are males and German native speakers. All participants read the same phonetically rich text *Nordwind und Sonne* (North Wind and Sun). The entire passage is presented in the section Appendices D. The text is 108 word long and the average duration of reading is 62.5 s (±SD, 26.1 s). The stuttering part of the database was used in the research experiment published in Nöth *et al.* (2000).

A great disadvantage is that the information about the age of participants in this experiment is not known.

6.1.2 Subjective evaluation of German recordings

The evaluation of 16 read recordings of German speakers is composed of the number of disfluencies found in individual recordings. These recordings and their score were used in the experiments (Nöth *et al.*, 2000). One professional speech specialist assessed/counted the disfluencies. The recordings of 18 fluent speakers did not contain any disfluencies. Figure 6.1 shows the histogram of subjective evaluation (number of disfluencies).

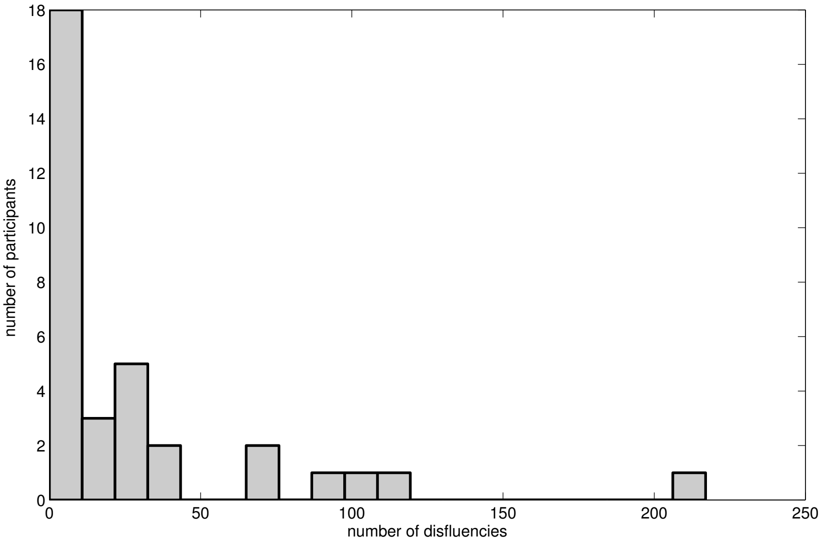


Figure 6.1: Histogram, distribution of subjective evaluation of German read recordings, 34 participants.

6.1.3 Measurements of speech fluency

The measurements were designed to take into account symptoms of stuttering and disfluent speech. They were originally proposed to be used for Czech stuttering speakers but they do not take the specifics of the Czech language into account. The list of all measures is given below:

- The average length of silence (ALS)
- The extent of speech fluency (ESF)
- Spectral changes in short interval (SCSI)
- The number of spectral changes in speech segments (NSI)

6.2 Results

To evaluate the performance of the algorithms with the German recordings, Pearson's correlation and statistical analysis ANOVA are utilized. The measures are compared to the number of disfluencies in utterance (subjective evaluation). The total reading time (RT) – the duration of recording in second, a standard measure used to evaluate pathological speech Foundas *et al.* (2004); Maier *et al.* (2011), is added and its results are displayed along with others.

First, the correlations between settings of the measures and the subjective evaluation are presented. These are ALS, ESF, SCSI, and NSI in Tables XVI, XVII, XVIII, and XIX, respectively. The measures reveal good agreement with the subjective evaluation. The ALS (settings 900 and 1000 ms) reached the correlations of 0.81. The ESF has the worst performance of all measures, the highest coefficient was 0.6, achieved by several settings of the algorithm. The moderate correlations are recognized with the SCSI. Many settings have a correlation coefficient of about 0.65, with the best performance at 0.72. The combination of spectral changes and voice activity detection brings the best results. The measure NSI exceeded with several settings a correlation of 0.8, with the highest at 0.85. The standard measure RT reached correlation 0.89, the highest in this study.

The range of values (mean and standard deviation) of each algorithm is given in Table XX for two groups (0 - without any sign of stuttering or fluent, 1 - subjects with stuttering). The typical values are followed by the results of the ANOVA analysis.

The ALS and comparative RT increase with the level of fluency disorder, on the other hand all the BACD-based measures decrease with the level of speech fluency disorder.

According to the best performance (the highest correlation), the algorithm settings were selected to be presented in ANOVA analysis.

1. ALS, time threshold for successive removal 900 ms;
2. ESF, multiplication constant 0.1, $k = 2$;
3. SCSI, multiplication constant 0.1, $k = 2$, window length 2 s;
4. NSI, multiplication constant 0.1, $k = 2$, threshold for successive removal 900 ms.

TABLE XVI: Pearson's correlation for all settings of the ALS in comparison to the number of disfluencies (subjective evaluation), setting of the measure: the time threshold for successive removal.

setting [ms]	correlation
100	0.51
150	0.58
200	0.65
300	0.69
400	0.71
500	0.75
700	0.78
800	0.80
900	**0.81**
1000	0.81
1100	0.75
1200	0.74
1300	0.76
1400	0.77
1500	0.76

The measures ALS, ESF, and SCSI are able to find statistically significant differences between fluent and disfluent speech at the level 0.01, while the measure NSI at the level 0.001, and the comparative measure RT at the level 0.05.

Another view on the results is shown by Figure 6.2, where the NSI is compared to the subjective evaluation. The range and dependency of the measure on the level of the disorder can be seen.

The classification using linear discriminant analysis and leave-one-out cross-validation was done to validate the results of correlations and ANOVA analysis. The NSI algorithm classified 26 (77%) of 34 recordings correctly, incorrectly 8 (23%, 5 disfluent, 3 fluent). The ALS, ESF, and SCSI placed 24 (71%) recordings correctly, 10 (29%) incorrectly. The comparative measure assigned 24 (71%) speakers to the correct class.

TABLE XVII: Pearson's correlation for all settings of the ESF in comparison to the number of disfluencies (subjective evaluation), setting of the measure: k-th highest maximum and multiplication constant.

	multiplication constant, correlation				
k	0.10	0.15	0.20	0.25	0.30
1	-0.60	-0.59	-0.54	-0.50	-0.50
2	**-0.65**	-0.60	-0.58	-0.55	-0.53
3	-0.60	-0.54	-0.52	-0.50	-0.46
4	-0.60	-0.54	-0.53	-0.51	-0.47
5	-0.60	-0.54	-0.53	-0.52	-0.47
6	-0.59	-0.52	-0.52	-0.50	-0.47
7	-0.59	-0.54	-0.52	-0.51	-0.48
8	-0.60	-0.55	-0.54	-0.53	-0.49
9	-0.59	-0.54	-0.53	-0.51	-0.50

TABLE XVIII: Pearson's correlation of the SCSI in comparison to the number of disfluencies (subjective evaluation), setting of the measure: k-th highest maximum, multiplication constant, and window length 2 s.

	multiplication constant, correlation				
k	0.10	0.15	0.20	0.25	0.30
1	-0.68	-0.66	-0.63	-0.59	-0.58
2	**-0.72**	-0.68	-0.57	-0.64	-0.61
3	-0.66	-0.61	-0.61	-0.58	-0.54
4	-0.67	-0.62	-0.61	-0.60	-0.55
5	-0.66	-0.62	-0.60	-0.60	-0.54
6	-0.65	-0.60	-0.60	-0.58	-0.55
7	-0.65	-0.59	-0.60	-0.59	-0.56
8	-0.67	-0.61	-0.62	-0.61	-0.58
9	-0.66	-0.61	-0.60	-0.59	-0.57

TABLE XIX: Pearson's correlation of the NSI in comparison to the number of disfluencies (subjective evaluation), setting of the measure: k-th highest maximum, multiplication constant, and time threshold 900 ms.

	multiplication constant, correlation				
k	0.10	0.15	0.20	0.25	0.30
1	-0.84	-0.81	-0.77	-0.71	-0.68
2	**-0.85**	-0.81	-0.78	-0.74	-0.70
3	-0.83	-0.79	-0.76	-0.72	-0.67
4	-0.84	-0.80	-0.77	-0.73	-0.67
5	-0.84	-0.80	-0.78	-0.75	-0.69
6	-0.83	-0.80	-0.78	-0.74	-0.70
7	-0.83	-0.80	-0.78	-0.75	-0.71
8	-0.84	-0.81	-0.78	-0.76	-0.72
9	-0.84	-0.81	-0.78	-0.76	-0.72

TABLE XX: The mean $\overline{x}$ and standard deviation SD of fluency measures and statistical significance by means of the ANOVA analysis with comparison between levels by the *post hoc* Bonferroni adjustment.

	ALS		ESF		SCSI		NSI		RT	
	$\overline{x}$	SD	$\overline{x}$	SD	$\overline{x}$	SD	$\overline{x}$	SD	$\overline{x}$	SD
Fluent (0)	0.21	0.05	7.89	0.98	1.23	0.05	0.94	0.05	51.8	8.1
Stuttering (1)	0.36	0.17	6.32	1.45	1.13	0.1	0.78	0.14	74.1	34.1
	Comparison between the classes									
ANOVA F(1, 33)	12.92*		11.29*		12.1*		17.41*		7.05*	
0 vs. 1	p <0.01		p <0.01		p <0.01		p <0.001		p <0.05	

NS = not significant

$^*p < 0.001$

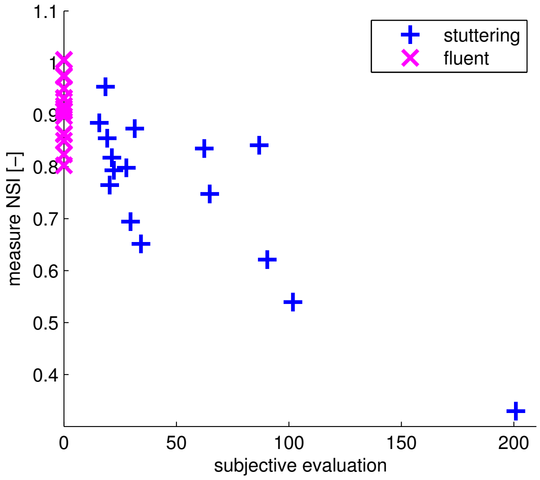


Figure 6.2: The comparison of the NSI with the number of all disfluencies (subjective evaluation).

6.3 Discussion

The study describes an experiment where four automatic and objective measures initially intended to be used on Czech stuttering recordings were applied to German stuttering recordings of read speech. The symptoms of stuttering were considered in the design of the measures. Moreover, the algorithms which were used do not take specifics of any language into account. Therefore, the main goal was to find out if it is possible to utilize these algorithms for evaluation of disfluency in different languages, in other words to confirm their language–independence.

The experiment analyzes read recordings of 16 stuttering and 18 fluent native German speakers. The number of disfluencies was counted in each recording and it constitutes the subjective evaluation. These recordings were used in the study where speech recognition technology was applied to look for stuttering events in recordings of read text Nöth *et al.* (2000).

The main finding of this experiment is that the measures are able to describe the level of the speech fluency disorder for read German speech. This finding is supported by the results where two of four measures exceeded correlation coefficients with the reference evaluation of 0.8, the highest is 0.85 (the number of spectral changes in speech intervals). The standard measure, the total reading time, achieved a correlation of 0.89. The ANOVA analysis confirms the results: three of the measures found statistically significant difference between fluent and stuttering recordings at the level 0.01, for the number of spectral changes in speech segments even 0.001, while for the standard measure, the total reading time, at the level 0.05.

To validate the results of correlation and ANOVA analysis, the classification using linear discriminant analysis and leave–one–out cross–validation was carried out. The NSI algorithm classified 26 (77%) of 34 recordings correctly, 8 incorrectly (23%). The other measures as well as the total reading time classified 71% into the correct class. The overlapping of the groups of normal fluent speakers and speaker at lower levels of disfluency (see Figure 6.2) should not be interpreted as a failing of the algorithms; it is a common finding that speech of normal fluent speaker contains disfluencies (Roberts *et al.*, 2009), also, Johnson (1961) found a high overlap in the speech rate of fluent speakers and participants who were classified as stutterers.

The standard measure, total reading time, was found distinctive for evaluation of pathology in read speech Foundas *et al.* (2004); Maier *et al.* (2011). Also experiments with Czech stuttering speakers acknowledged this statement Lustyk *et al.* (2014). The measure was added to the experiment to have a comparison with the introduced measures. The results of the total reading time turned out to be very good and in some cases better than those of the described methods (correlation coefficient of 0.89). However, there is an advantage of the automatic algorithms over the total reading time. The standard tasks in evaluation of stuttering (as clinical evaluation of spontaneous speech; monologue, picture description) are usually limited by time Johnson (1961), therefore the total duration of recordings would not influence the evaluation but the automatic measures would be able to do the assessment.

We have found some similarities for experiments on current German recordings and previous experiments on Czech recordings Lustyk *et al.* (2014). Firstly, the range of all measures (mean and standard deviation) are very similar for both languages. Secondly, similar settings of the algorithms performed well for both Czech and German recordings. The measures ESF and SCSI have got little worse correlation for German recordings (0.65, 0.72 vs. 0.76, 0.80 on Czech recordings). The ALS and NSI achieved higher correlations 0.81, 0.85 on German vs. 0.68, 0.82 on Czech recordings. We can also conclude that the measures can be considered as robust, especially those based on Bayesian change–point detection, because there are several settings which reached consistent results. The same was observed for Czech recordings.

The stuttering part of the database in this experiment was used in Nöth *et al.* (2000). One can see that the results of current the experiment are not as good as in that research. The highest correlation coefficient here is 0.85 (number of spectral changes in speech intervals) on the contrary the previous study reported correlation of 0.99. The reasons for it could be, firstly, that the measures are considered to be language–independent, they do not take specifics of German language into account. Moreover, algorithms based on hidden Markov models were specially trained for the German language, using the German–stuttering adapted grammar. The language–independence is an advantage of introduced algorithms, although their performance on different languages could be either better or worse in comparison to the language–dependent, because there are differences in languages. Prosodic characteristics such as rhythm, stress and intonation in speech conveys some important information regarding the identity of the spoken language (Mary and Yegnanarayana, 2008). However, there exist some similarities between prosodic features of languages, therefore it is possible to find language–independent features, for example: pauses, fundamental frequency, and others (Vaissiere, 1983). This would influence performance of the measurements. Probably, different measurements settings would have to be used for different languages.

A weak point of this experiment is the database, which consists of only 34 speakers. Although the entire spectrum of speech fluency disorder is represented, it would be preferable to have more subjects on different levels.

Chapter 7

Experiments on recordings with delayed auditory feedback

The experiment on recordings with delayed auditory feedback (DAF) was carried out to find whether it is possible to use automatic and objective measurements for setting appropriate delay of the DAF device for individual patients.

The experiment was performed on a small group of participants, as was not extended to a larger study, because there is a serious problem with the design of the study and character of the recordings. The problem with the study design lies in the nature of the task which the participants underwent. The DAF task was performed as reading of *one* passage with eight different settings of DAF. There exists effect on stuttering patients called adaptation effect, a reduction in stuttering in successive reading of the same material by people with stuttering (Johnson and Knott, 1937, 1939; Bloodstein and Bernstein Ratner, 2008). The effect tends to be very marked during the first few readings and becomes progressively less so, generally reaching its limit where there is little or no further effect. Most of the reduction that is to take place is evident in most cases by the fifth reading. Therefore, in the study of successive reading of the same material and applied DAF, we are not able to distinguish what is the effect of adaptation or DAF.

Nevertheless, we manually evaluated recordings of 11 participants by means of subjective stuttering scale and then compared the results of automatic objective measures to subjective evaluation. The experiment attempted to demonstrate that the measurements are able to track individual stuttering patient's performance.

7.1 Method

7.1.1 Read recordings with delayed auditory feedback

The experiments with delayed auditory feedback were carried out on recordings of 11 speakers. The recordings with DAF comes from the same database which was recorded at Department of Phoniatrics, 1st Faculty of Medicine, Charles University and General Faculty hospital in Prague. The delays chosen for the experiments were: none, 10, 20, 30, 40, 55, 80, and 110 ms. Two of the speakers in this experiment were fluent, the rest of the participants were stutterers at different levels of speech disfluency. The age of the two fluent participants was 31.1 and 29 years. The age

TABLE XXI: The Kondas score (second column) and *overall* LBDL score (third to tenth column) for all participants and recorded settings of DAF (delayed auditory feedback).

		delay setting [ms]/overall LBDL							
signal number	Kondas scale	0	10	20	30	40	55	80	110
74	0	0	0	0	0	0	0	1.35	6.67
75	4	125.81	65.79	36.84	36	32.43	4.05	5.41	4.05
76	3	29.27	2.7	0	1.35	0	2.74	0	1.35
77	1	2.7	0	0	0	0	0	0	2.7
78	3	31.17	8.11	2.74	1.35	5.41	13.7	13.51	0
80	4	55.56	32.39	32.39	22.22	16.42	8.57	4.62	5.8
81	2	17.11	10.81	9.46	15.15	4.05	4.05	4.05	12.16
82	3	35.21	8.57	8.57	4.29	2.86	14.29	2.86	8.57
83	0	0	0	0	1.35	0	1.35	1.35	5.41
84	2	10.53	1.35	1.35	8.11	4.05	0	0	2.7
85	3	57.33	-	25.68	-	-	36	-	36.49

of stuttering participants was from 9.1 to 19.4 yr, the mean age is 13.8 yr, ±SD 3.5 yr.

The participants repeatedly and successively read the same text *Podzim na Starém bělidle* with varied DAF. The passage was the same as in the study of read recordings. Patients had an opportunity to rest for a short time between individual readings.

7.1.2 Subjective evaluation of read recordings with delayed auditory feedback

The group of 11 speakers is a part of larger database described in previous studies. All 11 participants recorded the read recording without DAF, therefore subjective evaluation according to the Kondas scale is available (the merged evaluation of two judges). The number of participants in individual levels 0, 1, 2, 3, and 4 is 2, 1, 2, 4, and 2, respectively. For further analysis, recordings with all settings of DAF for all participants were evaluated by one judge. Moreover, the rate of speech [words/time] was obtained for each recording.

The *overall* score using the LBDL scale for all participants and recorded settings of DAF are given in Table XXI together with the score assigned in the Kondas scale (second column).

7.1.3 Objective measurements

The experiments with delayed auditory feedback are presented only for one automatic measurement (NSI). We carried out the experiments also with another measurements but because of the clarity and simplicity of presentation, the results are described for the measurement NSI.

7.2 Results

The section presents results obtained for recordings with DAF. We compared the measurements to subjective evaluation of each recordings with varied DAF. Figures 7.1, 7.2, and 7.3 show comparison of the NSI measurement and subjective evaluation (*overall* LBDL category – upper part, and manually measured speech rate [words/min] – lower part). In the best possible case the curves of the measurement and subjective evaluation by means of Kondas scale would be identical. The course of the curves of automatic measure and manually measured speech rate are opposite. We calculated the correlation coefficient of the objective measurement and subjective evaluation for individual signal, see Table. XXII. Figures 7.1, 7.2 are examples where the agreement between automatic and manual methods is high. Correlation for the recording with identification number 80 reached 0.90 (speech rate manually) and -0.94 (overall LBDL), the participant is from speech disfluency level 4 (very severe disfluency). Another example is the recording 83, level 0 (no frequent signs of disfluency), the agreement between automatic and manual methods is 0.97 and -0.89 for speech rate (manually measured) and overall LBDL evaluation, respectively. Figure 7.3 displays the course of both methods for recording with identification number 84 is an example where the agreement is not that high. The correlation reached coefficient of 0.40 (speech rate manually) and -0.43 (overall LBDL).

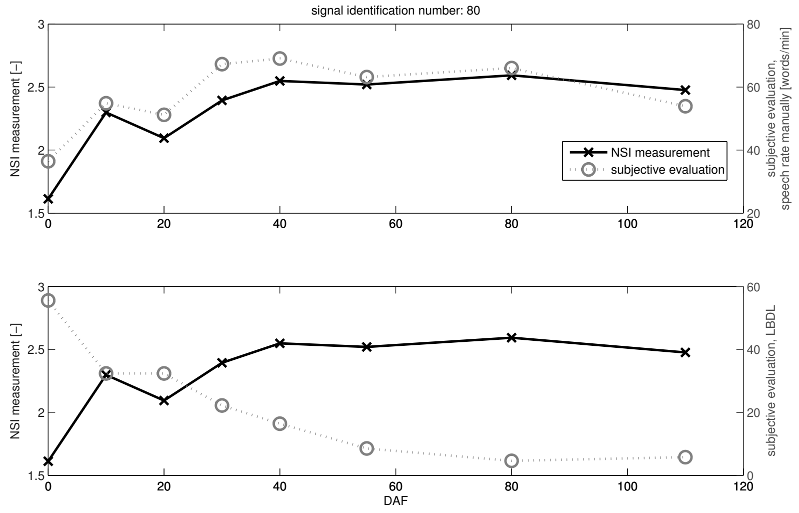


Figure 7.1: The course of the objective measurement NSI and subjective evaluation (LBDL and manually measured speech rate) for the participant with identification number 80 (0 – fluent speaker). It illustrates a good agreement between objective and subjective methods.

7.3 Discussion

The original purpose of the experiment with delayed auditory feedback (DAF) was to find out whether the automatic objective measurements could be used to set up a correct delay for individual patients. However, at the beginning a serious problem with the design of the study was found. The task in the DAF experiment was designed as reading of *one* passage with varied delay. The major problem is that it was only one passage. In stuttering there exists an effect called *adaptation effect* (Johnson and Knott, 1937, 1939; Bloodstein and Bernstein Ratner, 2008), in other words, an effect when reader learns a text and it helps her/him to read the text with less problems (Bloodstein and

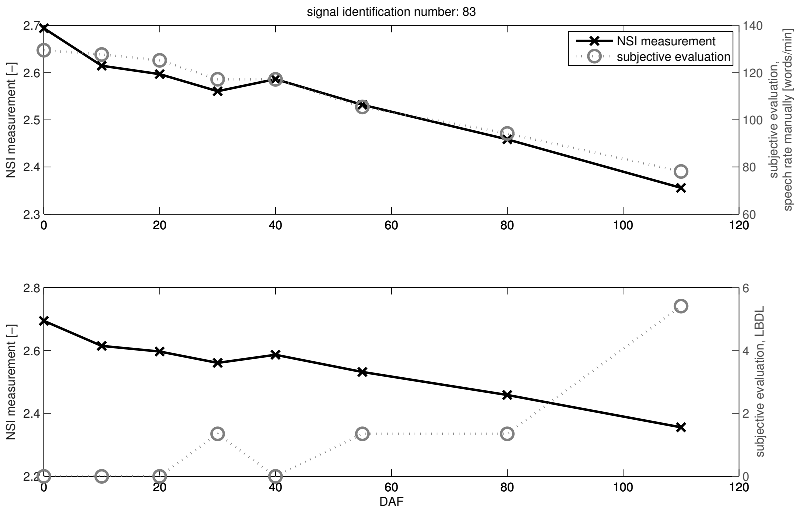


Figure 7.2: The course of the objective measurement NSI and subjective evaluation (LBDL and manually measured speech rate) for the participant with identification number 83 (4 – very severe level of disfluency). It illustrates a good agreement between objective and subjective methods.

Bernstein Ratner, 2008) therefore more fluent. The DAF is a device to enhance fluency in people who stutter. Thus, when there are two possible influences on speakers fluency we would not be able to distinguish what impact is from DAF and what is from adaptation. The study would be possible with different design (to cover all the possible situation with speakers). The groups would be: participants who read the text repeatedly, repeatedly with DAF, with DAF but every time a different text. Even though there was no other group of the subjects and the study design is wrong, the signals were used, but only to demonstrate how the automatic algorithms are able to track individual patients' disfluency in recordings with DAF.

Each of the participants with all settings of DAF was compared to the subjective evaluation made by means of LBDL. The scale is reliable when used by experienced judges (Teesson *et al.*, 2003; Goberman *et al.*, 2010). The assessment was carried out by one evaluator who also assessed read and spontaneous recordings (the reliability of the evaluation in read and spontaneous recordings was considered as good for the study). Therefore, the second round of evaluation to get intra– or inter–judge agreement was not carried out.

The results of the comparison between objective measurements and subjective evaluation show that the measures are able to describe the level of speech fluency in recordings with DAF. For most of the participants, the measurements mimic the results of subjective evaluation (Figures 7.1 and 7.2). Also, the table gives the overview of the agreement (correlations) between objective measures and subjective evaluation for all participants and all DAF recordings. Results for six of eleven participants reached a correlation higher than 0.9 either with manually measured speech rate or subjective evaluation by means of LBDL. Correlations between subjective evaluation and objective

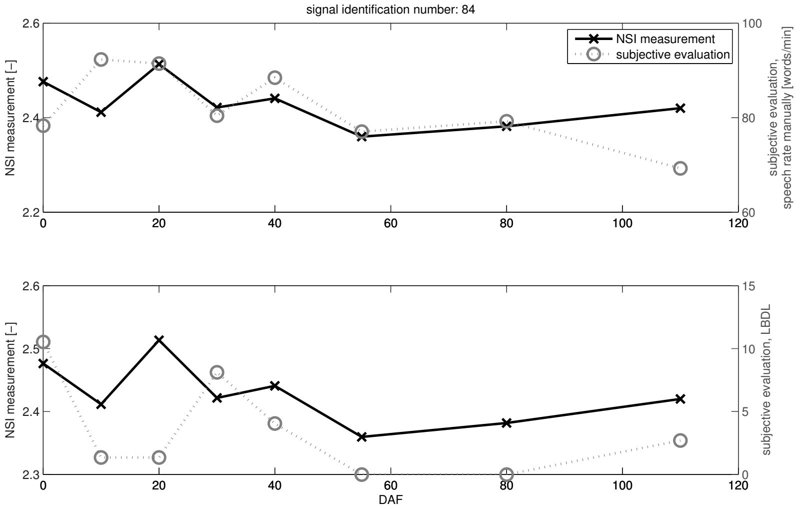


Figure 7.3: The course of the objective measurement NSI and subjective evaluation (LBDL and manually measured speech rate) for the participant with identification number 84 (2 – moderate disfluency). It illustrates worse agreement between objective and subjective methods.

measures for another three participants were higher than 0.68. Agreement between subjective evaluation and objective measures for recordings of two remaining participants did not overcome absolute value of correlation coefficient of 0.46.

The limitation of the study has been mentioned at the beginning of the discussion (the design of the study). Because of it we are not able to differentiate between the effect of adaptation and DAF. However, it suggests one of the future directions of the research.

TABLE XXII: Correlation coefficient between measurements NSI and (delayed auditory feedback).

	correlation		
signal identification number	speech rate	LBDL	disfluency level
74	0.95	-0.75	0
75	0.97	-0.95	4
76	0.93	-0.77	3
77	0.95	-0.14	1
78	0.70	-0.51	3
80	0.90	-0.94	4
81	0.76	-0.46	2
82	0.68	-0.80	3
83	0.97	-0.89	0
84	0.40	-0.43	2
85	0.05	-0.46	3

Chapter 8

Conclusions

The thesis focuses on objective evaluation of disfluent speech of people who stutter. The main objective of the thesis has been to analyze disfluent speech by means of automatic and objective methods and find out whether these methods can estimate the level of speech disfluency in recordings of different speaking tasks. The results confirmed that the automatic and objective measurement can evaluate disfluenct speech of people who stutter. The thesis is divided into three main studies and one short experiment and the following paragraphs provide a summary of the thesis.

The *first* part of the thesis concentrates on read recordings of disfluent speech and the goal is formulated in the question:

- Are the automatic measures able to describe fluency/disfluency in speech of stutterers?

The measurements showed a very good agreement with overall subjective evaluations made by experienced speech–language pathologists (the highest correlation, 0.78). Strong correlations were also found between measurements and individual symptoms of speech disfluency, especially fixed postures without audible airflow (0.84). This finding suggests that it is possible to aim measurements to different stuttering symptoms. In the case that there are measurements which correlate with different stuttering symptoms, combination of these measurement could led to better results. A simple combination of four measurements obtained correlations higher than that for any single measure (0.80). Also, the ANOVA analysis revealed that the measurements are able to separate different levels of speech disfluency, significant differences were found among all levels except fluent and mild disfluency. These results were confirmed by the classification task, when the selected setting of the NSI algorithm classified 61 subjects (52%) into the correct level of the discrete Kondas's scale, 50 subjects (42%) with the classification error 1, and seven participant (6%) with classification error 2.

The *second* study extends the experiments made on read recordings to spontaneous recordings. The aims of the second part are defined in the following questions:

- Are the measures able to estimate the level of speech fluency disorder in spontaneous recordings?
- How does stuttering speakers' fluency differ in read and spontaneous speaking tasks?
- Does the level of speech fluency play a role in the speaking task effect?

The results indicate that the evaluation of speech–language pathologists on speech disfluency can be estimated by means of objective measurements in spontaneous recordings. However, the agreement with the subjective assessment is lower than in read recordings (-0.82 and -0.52 in read and spontaneous recordings, respectively). In particular, this part of the theisis compared read with spontaneous recordings of people who stutter and thereby it broadens information about the speaking task effect on persons with speech disfluency. We have found that the speech of participants at different levels of disfluency is affected to various extent by the speaking tasks. Specifically, the fluent participants and participants at lower levels of speech disfluency in general display that their fluency is better in reading than in spontaneous task. While most of the participants with severe disfluency showed the opposite. The participants' ability to speak in spontaneous task is affected less by their level of speech disfluency than in oral reading. The contribution of our research to this issue consists in that the stuttering group was not taken as a homogeneous group but instead we took various levels of speech fluency into consideration. The conclusion could have an implication on the stuttering assessment procedure, since it points out that a higher distinctiveness is obtained for read recordings (it applies for both subjective and objective methods). However, this does not mean that the spontaneous speech should be left out of the evaluation procedure, it is an integral part of the procedure which brings another perspective for speech–language pathologist observation.

The *third* part of the thesis analyzes a possible use of the objective disfluency methods in different languages. The measurements were originally designed for Czech stuttering speakers but their nature implied possible language–independence. The following question state the aim of the third part:

- Are the measures able to describe the level of speech disfluency in different languages?

The experiment was carried out on read recordings of German stuttering and normal speakers. The results suggest that the language–independent evaluation by means of automatic and objective measures is realistic. The measures showed impaired speech fluency in stuttering participants and were able to find significant differences between typical speakers and people who stutter. Although we observed a high correlation between subjective assessment and objective measurements, the measurements did not reach the correlations obtained for language–dependent methods. This means that the language–independent evaluation of speech disfluency is possible but better results are reached by methods adapted to the selected language.

The *fourth* part of the thesis focuses on the recordings with delayed auditory feedback (DAF). We analyzed eleven read recordings of stuttering participants with different settings of DAF. The goal was to find out whether the objective measurements would be a useful method in evaluation of appropriate delay setting in the DAF device. The results indicate that the measurements are able to track the individual settings of delay and evaluate correct delay setting for individual participants. However, the results of this part have to be considered carefully because of the incorrect design of the study which suggest one of possible directions in future research.

8.1 Future directions

Since the true nature of stuttering has yet to be found there is a space where to focus further research. The automatic objective measures may be helpful in some aspects. Further research dealing with automatic and objective measures could explore several topics:

- Although the study was carried out on a large sample of participants, it would be preferable to confirm the results with a larger and more balanced sample of participants, i.e. more balanced in terms of the number of participants at individual levels of disfluency.
- Tracking individual speakers for a longer time, evaluation of results and outcomes of speech therapy (before, during, and after a therapy, after a longer time period). Individualization of results, refraining from drawing general conclusion. Tracking time evolution of the speech fluency disorder.
- One limitation of presented measurements is that they are primarily capable of describing fluency related to speech rate. If they were applied in clinical evaluation, they would have to be used in conjunction with automatic measurements considering other symptoms of stuttering, for example repetition (Nöth *et al.*, 2000) and prolongation (Wisniewski *et al.*, 2007b) to obtain a comprehensive picture of disfluency. This may suggest that one direction of future research could deal with individual symptoms of stuttering in both speaking tasks with respect to the level of speech fluency.
- A correctly designed study for signals with delayed auditory feedback (DAF). There would have to be more groups of participants to differentiate the effect of adaptation from the effect of DAF, for example: Stuttering group – recordings with DAF, one text; stuttering group – recordings with DAF, different texts; stuttering group – recordings without DAF, one text; stuttering group – recordings without DAF, different texts; fluent participants – one reading passage; fluent participants – different reading passages; fluent participants – one reading passage with DAF; fluent participants – different reading passage with DAF. The study would try to distinguish the influence of DAF and adaptation effect.
- Would it be possible to use the algorithms for differential diagnosis? Even though the results and conclusions are made as general and are applied on people who stutter as a group or divided according to the level of the disfluency), the measures indicate that there are participants who speak differently than the rest of the stuttering group. We observed this especially for the rate of speech in spontaneous recordings. There were participants who had a very severe level of disfluency but their speech rate was at a similar level as fluent participant or participants at mild disfluency. This could suggest a different diagnosis/problem of these participants.

References

Adams, M. R. (**1987**). "Voice onsets and segment durations of normal speakers and beginning stutterers", J. Fluency Disord. **12**, 133–139.

Alm, P. A. (**2004**). "Stuttering and the basal ganglia circuits: a critical review of possible relations", J. Commun. Disord. **37**, 325–369.

Appel, U. and Brandt, V. A. (**1983**). "Adaptive segmentation of piecewise stationary time series", Information sciences **29**, 27–56.

Armson, J. and Stuart, A. (**1998**). "Effect of extended exposure to frequency–altered feedback on stuttering during reading and monologue", J. Speech Lang. Hear. Res. **41 (3)**, 479–490.

Atal, B. and Rabiner, L. (**1976**). "A pattern recognition approach to voiced-unvoiced-silence classification with applications to speech recognition", Acoustics, Speech and Signal Processing, IEEE Transactions on **24(3)**, 201–212.

Bergl, P. (**2006**). "Accuracy of divergence measures for detection of abrupt changes.", in *World academy of science, engineering and technology*, volume 18, 33–36.

Bergl, P. (**2010**). "Objektivizace poruch plynulosti reci (Objectification of speech disfluencies)", Ph.D. thesis, Czech Technical University in Prague, Faculty of Electrical Engineering, Prague, 135 pages, (in Czech).

Bergl, P. and Cmejla, R. (**2007**). "Improved detection of boundaries of phonemes in speech databases", in *Proceedings of the fifth IASTED International Conference: biomedical engineering (BIEN '07)*, 171–174 (ACTA Press, Anaheim, CA, USA).

Blomgren, M. and Goberman, A. M. (**2008**). "Revisiting speech rate and utterance length manipulations in stuttering speakers", J. Commun. Disord. **41 (2)**, 159–178.

Blood, G. W. and Hood, S. B. (**1978**). "Elementary shool–aged stutterers' disfluencies during oral reading and spontaneous speech", J. Fluency Disord. **3 (3)**, 155–165.

Bloodstein, O. and Bernstein Ratner, N. (**2008**). *A handbook on Stuttering*, sixth edition (Delmar, Cengage Learning, Clifton Park, NY), 1–552.

Boersma, P. (**2002**). "Praat, a system for doing phonetics by computer", Glot international **5(9/10)**, 341–345.

Cmejla, R., Rusz, J., Bergl, P., and Vokral, J. (**2013**). "Bayesian changepoint detection for the automatic assessment of fluency and articulatory disorders", Speech Commun. **55**, 178–189.

Conture, E. (**2001**). *Stuttering: Its nature, diagnosis, and treatment*, 1 edition (MA: Allyn & Bacon, Boston), chap. 1.

Cordes, A. K. and Ingham, R. J. (**1994**). "The reliability of observational data ii. issues in the identification and measurement of stuttering events", J. Speech Lang. Hear. Res. **37**, 279–294.

Craig, A., Blumgart, E., and Tran, Y. (**2009**). "The impact of stuttering on the quality of life in adults who stutter", J. Fluency Disord. **34 (2)**, 61–71.

Craig, A. and Tran, Y. (**2005**). "The epidemiology of stuttering: The need for reliable estimates of prevalence and anxiety levels over the lifespan", Int. J. Speech-Language Pathology **7(1)**, 41–46.

Cucchiarini, C., Strik, H., and Boves, L. (**2000**). "Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology", J. Acoust. Soc. Am. **107**, 989–999.

Cucchiarini, C., Strik, H., and Boves, L. (**2002**). "Quantitative assessment of second language learners' fluency: Comparison between read and spontaneous speech", J. Acoust. Soc. Am. **111**, 2862–2873.

de Andrade, C. R. F., Cervone, L. M., and Sassi, F. C. (**2003**). "Relationship between the stuttering severity index and speech rate", Sao Paulo Med. J. **121(2)**, 81–84.

Di Simony, F. G. (**1974**). "Some preliminary observations on temporal compensation in the speech of children", J. Acoust. Soc. Am. **56**, 697–699.

Esmaili, I., Dabanloo, N. J., and Vali, M. (**2016**). "Automatic classification of speech dysfluencies in continuous speech based on similarity measures and morphological image processing tools", Biomed. Signal Process. Control **23**, 104–114.

Ezrati-Vinacour, R. and Levin, I. (**2004**). "The relationship between anxiety and stuttering: a multidimensional approach", J. Fluency Dis. **29**, 135–148.

Foundas, A. L., Bollich, A. M., Feldman, J., Corey, D. M., Hurley, M., Lemen, L. C., and Heilman, K. M. (**2004**). "Aberrant auditory processing and atypical planum temporale in developmental stuttering", Neurology **64 (9)**, 1640–1646.

Goberman, A. M., Blomgren, M., and Metzger, E. (**2010**). "Characteristics of speech disfluency in Parkinson's disease", J. Neurol. **23**, 470–478.

Godino-Llorente, J. and Gomez-Vilda, P. (**2004**). "Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors", Biomedical Engineering, IEEE Transactions on **51(2)**, 380–384.

Hall, K. D. and Yairi, E. (**1992**). "Fundamental frequency, jitter, and shimmer in preschoolers who stutter", J. Speech Hear. Res. **35**, 1002–1008.

Hariharan, M., Chee, L. S., Ai, O. C., and Yaacob, S. (**2012**). "Classification of speech dysfluencies using LPC based parameterization techniques", J. Med. Systems **36**, 1821–1830.

Harrington, J. and Cassidy, S. (**1999**). *Techniques in Speech Acoustics* (Kluwer Academic Press), chap. 9, pp. 239–277.

Healey, E. C. and Gutkin, B. (**1984**). "Analysis of stutterers' voice onset times and fundamental frequency contours during fluency", J. Speech Hear. Res. **27**, 219–225.

Healey, E. C. and Ramig, P. R. (**1986**). "Acoustic measures of stutterers' and nonstutterers' fluency in two speech contexts", J. Speech Hear. Res. **29(3)**, 325–331.

Howell, P., Hamilton, A., and Kyriacopoulos, A. (**1986**). "Automatic detection of repetitions and prolongations in stuttered speech", Speech Input/Output: Techniques and Applications, IEE Publications 252–256.

Ingham, J. R., Grafon, S. T., Bothe, A. K., and Ingham, J. C. (**2012**). "Brain activity in adults who stutter: Similarities across speaking tasks and correlations with stuttering frequency and speaking rate", Brain Lang. **122 (1)**, 11–24.

Johnson, W. (**1961**). "Measurements of oral reading and speaking rate and disfluency of adult male and female stutterers and nonstutterers", J. Speech Hear. Disord. **7**, 1–20.

Johnson, W. and Knott, J. R. (**1937**). "Studies in the psychology of stuttering: I. the distribution of moments of stuttering in successive readings of the same material", J. Speech Disord. **2**, 17–19.

Johnson, W. and Knott, J. R. (**1939**). "Studies in the psychology of stuttering: XIII. a statistical analysis of the adaptation and consistency effect in relation to stuttering", J. Speech Disord. **4**, 79–86.

Kalinowski, J. (**2003**). "Self-reported efficacy of an all in-the-ear-canal prosthetic device to inhibit stuttering during one hundred hours of university teaching: an autobiographical clinical commentary", Disability and Rehabilitation **25(2)**, 107–111.

Kay Elemetrics Corp. (**2003**). *Multi-Dimensional Voice Program (MDVP): Software Instruction manual*, Kay Elemetrics, Lincoln Park.

Kent, R., Weismer, G., Kent, J., Vorperian, H., and Duffy, J. (**1999**). "Acoustic studies of dysarthric speech: methods, progress, and potential.", J. Commun. Dis. **32**, 141–186.

Kubikova, L., Bosikova, E., Cvikova, M., Lukacova, K., Scharff, C., and Jarvis, E. D. (**2014**). "Basal ganglia function, stuttering, sequencing, and repair in adult songbirds", Sci. Rep. **4**, 1–16, 6590; DOI:10.1038/srep06590.

Kuniszyk-Jozkowiak, W. (**1995**). "The statistical analysis of speech envelopes in stutterers and non-stutterers", J. Fluency Disord. **20(1)**, 11–23.

Kuniszyk-Jozkowiak, W. (**1996**). "A comparison of speech envelopes of stutterers and nonstutterers", J. Acoust. Soc. Am. **100(2)**, 1105–1110.

Lastovka, M., Vokral, J., Cerny, L., Radilova, K., and Hrdlickova, M. (**1998**). "Hodnoceni plynulosti reci pomoci neuronovych siti (Speech disfluency assessment using neural networks)", Research report 237/1998/C/1.LF (in Czech).

Lechta, V. and collective (**2004**). *Diagnoza narusene komunikacni schopnosti (Diagnostics of impaired communication ability)* (Portal, Prague), (in Czech), pp. 317–332.

Levelt, W. J. M. (**1989**). *Speaking: From Intention to Articulation. A Bradford book* (The MIT Press, Cambrige, MA), 1–545.

Lustyk, T., Bergl, P., and Cmejla, R. (**2014**). "Evaluation of disfluent speech by means of automatic acoustic measurements", J. Acoust. Soc. Am. **135(3)**, 1457–1468.

Maier, A., Haderlein, T., Eysholdt, U., Rosanowski, F., Batliner, A., Schuster, M., and Nöth, E. (**2009**a). "PEAKS - A system for the automatic evaluation of voice and speech disorders", Speech Communication **51**, 425–437.

Maier, A., Hönig, F., Bocklet, T., Nöth, E., Stelzle, F., Nkenke, E., and Schuster, M. (**2009**b). "Automatic detection of articulation disorders in children with cleft lip and palate", J. Acoust. Soc. Am. **126(5)**, 2589–2602.

Maier, A., Hönig, F., Steidl, S., Nöth, E., Horndach, S., Saurerhöfer, E., Kratz, O., and Moll, G. (**2011**). "An automatic version of a reading disorder test", ACM Trans. Speech Lang. Process. (TSLP) **7(4)**, 17:1–17:15.

Maier, A., Hönig, F., Zeissler, V., Batliner, A., Körner, E., Yamanaka, N., Ackermann, P., Peter, D., and Nöth, E. (**2009**c). "A language-independent feature set for the automatic evaluation of prosody", in *Proceedings of the 10th Annual Conference of the International Speech Communication Association (Interspeech 2009)*, 600–603 (10th Annual conference of the International Speech Communication Association (Interspeech 2009), Brighton, England).

Manning, W. H. (**2009**). *Clinical Decision Making in Fluency Disorders*, third edition (Delmar, Cengage Learning, New York), pp. 1–744.

Mansson, H. (**2000**). "Childhood stuttering: incidence and development", J. Fluency Disord. **25**, 47–57.

Mary, L. and Yegnanarayana, B. (**2008**). "Extraction and representation of prosodic features for language and speaker recognition", Speech Communication **50**, 782–796.

Matoušek, J., Tihelka, D., Psutka, J., and Hesová, J. (**2002**). "German and czech speech synthesis using hmm-based speech segment database", in *Text, Speech and Dialogue*, (pp. 173–180) (Springer Berlin Heidelberg).

McCullagh, P. and Nelder, J. (**1989**). "Generalized linear models", CRC press **37**.

Metz, D. E., Samar, V. J., and Sacco, P. R. (**1983**). "Acoustic analysis of stutterers' fluent speech before and after therapy", J. Speech Hear. Res. **26**, 531–536.

Nöth, E., Niemann, H., Haderlein, T., Decher, M., Eysholdt, U., Rosanowski, F., and Wittenberg, T. (**2000**). "Automatic stuttering recognition using hidden Markov models", in *Sixth International Conference on Spoken Language Processing*, volume 4, 65–68 (Beijing, China).

Onslow, M., Gardner, K., Bryant, K., C.M.Stuckings, and Knight, T. (**1992**). "Stuttered and normal speech events in early childhood: The validity of a behavioral data language", J. Speech Hear. Res. **35**, 79–87.

Pinto, J. C. B. R., Schifer, A. M., and de Alvia, C. R. B. (**2013**). "Disfluencies and speech rate in spontaneous production and in oral reading in people who stutter and who do not stutter", Audiol. Commun. Res. **18 (2)**, 63–70.

Prosek, R. A. and Runyan, C. M. (**1982**). “Temporal characteristics related to the discrimination of stutterers’ and nonstutterers’ speech samples”, J. Speech Lang. Hear. Res. **25 (1)**, 29–33.

Ravikumar, K. M., Rajagopal, R., and Nagaraj, H. C. (**2009**). “An approach for objective assessment of stuttered speech using mfcc features”, ICGST Int. J. on Digital Signal Processing, DSP **9 (1)**, 19–24.

Riley, G. D. (**1972**). “A stuttering severity instrument for children and adults”, J. Speech Hear. Disord. **37**, 314–322.

Riley, G. D. (**2009**). *Stuttering Severity Instrument for children and adults (SSI-4) (4th ed.)*, TX: PRO-ED, Austin.

Robb, M., Blomgren, M., and Chen, Y. (**1998**). “Formant frequency fluctuation in stuttering and nonstuttering adults”, J. Fluency Disord. **23(1)**, 73–84.

Roberts, P. M., Meltzer, A., and Wilding, J. (**2009**). “Disfluencies in non–stuttering adults across sample lengths and topics”, J. Commun. Disord. **42 (6)**, 414–427.

Ruanaidh, J. and Fitzgerald, W. (**1996**). *Numerical Bayesian Methods Applied to Signal Processing* (Springer-Verlag, New York, NY), chap. 5, pp. 96–101.

Rusz, J., Cmejla, R., Ruzickova, H., and Ruzicka, E. (**2011**). “Quantitative acoutic measurements for characterization of speech and voice disorders in early untreated Parkinson’s disease”, J. Acoust. Soc. Am. **129**, 350–367.

Rusz, J., Cmejla, R., and Tykalova, T. (**2013**). “Imprecise vowel articulation as a potencial early marker of parkinson’s disease: Effect of speaking task”, J. Acoust. Soc. Am. **134(3)**, 2171–2181.

Rusz, J., Klempir, J., Tykalova, T., Baborova, E., Cmejla, R., Ruzicka, E., and Roth, J. (**2014**). “Characteristics and occurrence of speech impairment in Huntington’s disease: possible influence of antipsychotic medication”, J. Neural. Transm. **121**, 1529–1539.

Ryan, B. P. (**1992**). “Articulation, language, rate and fluency characteristics of stuttering and nonstuttering preschool children”, J. Speech Hear. Res. **35**, 333–342.

Sapir, S., Ramig, L. O., Spielman, J. L., and Fox, C. (**2010**). “Formant centralization ratio: a proposal for a new acoustic measure of dysarthric speech”, J. Speech Lang. Hear. Res. **53**, 114–125.

Scharenborg, O., Wan, V., and Ernestus, M. (**2010**). “Unsupervised speech segmentation: An analysis of the hypothesized phone boundaries”, J. Acoust. Soc. Am. **127 (2)**, 1084–1095, DOI: 10.1121/1.3277194.

Scharenborg, O., Witteman, M., and Weber, A. (**2012**). “Computational modelling of the recognition of foreign–accented speech”, in *the 13th Annual Conference of the International Speech Communication Association (Interspeech 2012)*, 882–885.

Schwarz, P. (**2009**). “Phoneme recognition based on long temporal context”, Ph.D. thesis, Brno University of Technology, pp. 1–95.

Silverman, S. H. (**1974**). “Disfluency bahavior of elementary–school stutterers and nonstutterers”, Lang. Speech Hear. Ser. 32–37.

Spaniel, F., Bakstein, E., Anyz, J., Hlinka, J., Sieger, T., Hrdlicka, J., Gornerova, N., and Hoschl, C. (**2016**). “Relapse in schizophrenia: Definitively not a bolt from the blue”, Neurosci. Lett. .

Szczurowska, I., Kuniszyk-Jozkowiak, W., and Smolka, E. (**2009**). “Speech nonfluency detection using Kohonen networks”, Neural. Comput. & Applic. **18**, 677–687.

Teesson, K., Packman, A., and Onslow, M. (**2003**). “The Lidcombe behavioral data language of stuttering”, J. Speech Lang. Hear. Res. **46**, 1009–1015.

Vaissiere, J. (**1983**). “Language-independent prosodic features”, in *Prosody: Models and Measurements*, edited by A. Cutler and R. L. (Eds.), 53–65 (Springer Verlag).

Van Borsel, J., Reunes, G., and Van den Bergh, N. (**2003**). “Delayed auditory feedback in the treatment of stuttering: clients as consumers”, Int. J. Lang. Commun. Disord. **38(2)**, 119–129.

Vanryckeghem, M., Glessing, J. J., Brutten, G. J., and McAlindon, P. (**1999**). “The main and interactive effect of oral reading rate on the frequency of stuttering”, Am. J. Speech-Lang. Pat. **8 (2)**, 164–170.

Wilcoxon, F. (**1947**). “Individual comparisons by ranking methods”, Biometrics **3**, 119–122.

Wisniewski, M., Kuniszyk-Jozkowiak, W., Smolka, E., and Suszynski, W. (**2007**a). “Automatic detection of disorders in a continuous speech with the Hidden Markov Models approach”, in *Comp. Recognition System 2, 45 of Advances in Soft Computing*, 445–453 (Springer, Berlin, Germany).

Wisniewski, M., Niewski, M., Kuniszyk-Jozkowiak, W., Smolka, E., and Suszynski, W. (**2007**b). “Automatic detection of prolonged fricative phonemes with the Hidden Markov models approach”, J. Med. Inform. Technol. **11**, 293–297.

World Health Organization (**1977**). *Manual of the international statistical classification of diseases, injuries, and causes of death*, World Health Organization, Geneva, vol. 1.

Yairi, E. and Ambrose, N. (**1999**). “Early childhood stuttering I: persistency and recovery rates”, J. Speech Lang. Hear. Res. **42**, 1098–1112.

Yairi, E. and Clifton, Jr., N. F. (**1972**). “Disfluent speech behavior of preschool children, high school seniors, and geriatric persons”, J. Speech Hear. Res. **15**, 714–719.

Yaruss, J. S. and Conture, E. G. (**1993**). “F2 transitions during sound/syllable repetitions of children who stutter and predictions of stuttering chronicity”, J. Speech Hear. Res. **36**, 883–896.

Publications related to the dissertation thesis

Articles in impacted journals:

"Evaluation of disfluent speech by means of automatic acoustic measurements"
Lustyk, T., Bergl, P., and Cmejla, R. [**33%**]
Journal of Acoustical Society of America (2014), 135(3), 1457–1468. ISSN 0001-4966, doi:10.1121/1.4863646

Articles excerpted in Web of Science:

"Change evaluation of Bayesian detector for dysfluent speech assessment"
Lustyk, T., Bergl, P., Čmejla, R., and Vokřál, J. [**30%**]
In: *Applied Electronics 2011.* Pilsen: University of West Bohemia, 2011, p. 231–234. ISSN 1803–7232. ISBN 978-80-7043-987-6.

Other Articles:

"Porovnání výsledků měření neplynulosti pro čtené a spontánní promluvy koktavých [online]."
Lustyk, T., Bergl, P., and Cmejla, R. [**33%**]
In: RUSZ, J., ČMEJLA, R., a SEDLÁK, J., eds. *V. Letní doktorandské dny 2015.* Letní doktorandské dny 2015. Praha, 28.05.2015 - 29.05.2015. Praha: ČVUT FEL, Katedra teorie obvodů. 2015, ISBN 978-80-01-05749-0. Available z: http://sami.fel.cvut.cz/publication.htm

"Language–independent method for analysis of German stuttering recordings"
Lustyk, T., Bergl, P., Haderlein, T.,Noth, E., and Cmejla, R. [**20%**]
In: *INTERSPEECH 2015.* INTERSPEECH 2015, 16th Annual Conference of the International Speech Communication Association. Dresden, 06.09.2015 - 10.09.2015. Bochum: ISCA - International Speech Communication Association. 2015, ISSN 2308-457X. http://sami.fel.cvut.cz/publication.htm

"Analysis of Stuttering [online]"
Lustyk, T. [**100%**]
In: LUSTYK, T., ČMEJLA, R., a RUSZ, J., eds. *PROCEEDINGS ABSTRACTS on II. CZECH-GERMAN WORKSHOP ON SPEECH PATHOLOGY AND BIOLOGICAL SIGNALS.* II. CZECH-GERMAN WORKSHOP ON SPEECH PATHOLOGY AND BIOLOGICAL SIGNALS. Erlangen, 02.12.2014 - 03.12.2014. Praha: ČVUT FEL, Katedra teorie obvodů. 2014, s. 7. ISBN 978-80-01-05670-7. http://sami.fel.cvut.cz/cgw14/cgw14.pdf

“Proceedings abstracts on II. Czech–German Workshop on speech pathology and biological signals”
Čmejla, R., Rusz, J., Lustyk, T. (ed.) [**30%**]
Prague: CTU, Faculty of Electrical Engineering, Department of Circuit Theory, 2014. ISBN 978-80-01-05670-7.

“Hodnocení koktavosti pomocí automatických algoritmů ve čtených promluvách”
Lustyk, T., Bergl, P., Čmejla, R. [**85%**]
In: III. LETNÍ DOKTORANDSKÉ DNY 2013. Praha: ČVUT, Fakulta elektrotechnická, 2013, díl 3, s. 118–121. ISBN 978-80-01-05251-8.

“Relation Between Stuttering Severity and Automatic Algorithms”
Lustyk, T., Bergl, P., Čmejla, R. [**33%**]
In: Czech–German Workshop on Speech Pathology and Biological Signals - Proceedings. Prague: CTU, Faculty of Electrical Engineering, Department of Circuit Theory, 2012, p. 52–53. ISBN 978-80-01-05164-1.

“The relation between spectral changes distance and prolongation in speech of stutterers”
Lustyk, T., Bergl, P., Čmejla, R. [**33%**]
In: 20th Annual Conference Proceeding’s Technical Computing Bratislava 2012. Prague: HUMUSOFT, 2012, p. 1–5. ISBN 978-80-970519-4-5.

“Akusticke parametry pro analyzu neplynule reci”
Lustyk, T., Bergl, P., Čmejla, R. [**33%**]
In: Sborník 85. akustického semináře. Praha: Nakladatelství ČVUT, 2012, čl. č. 4, s. 25–30. ISBN 978-80-01-05133-7.

“Použití systému hodnocení LBDL jako kontrolních dat pro automatické algoritmy”
Lustyk, T., Bergl, P., Čmejla, R. [**33%**]
In: Novinky ve foniatrii. Praha 5, Na bělidle 34, 150 00: Nakladatelství Galén, 2012, s. 112–114. ISBN 978-80-7262-940-4.

“Analýza patologického hlasu a řeči v laboratoři SAMI ČVUT”
Čmejla, R., Rusz, J., Bauer, L., Lustyk, T., Nejepsová, M., et al. [**11%**]
In: Novinky ve foniatrii. Praha 5, Na bělidle 34, 150 00: Nakladatelství Galén, 2012, s. 28–30. ISBN 978-80-7262-940-4.

“Hodnoceni koktavosti”
Lustyk, T., Čmejla, R., Bergl, P. [**33%**]
In: LETNÍ DOKTORANDSKÉ DNY 2012. Praha: ČVUT, 2012, s. 108–112. ISBN 978-80-01-05050-7.

“Detection of Repetitions in Stuttered Speech by Means of VAD and Time Thresholds”
Lustyk, T. [**100%**]
In: POSTER 2012 – 16th International Student Conference on Electrical Engineering. Praha: Czech Technical University in Prague, 2012, p. 1–4. ISBN 978-80-01-05043-9.

“The Number of Spectral Changes in Speech Segments for Evaluation of Dysfluent Speech”
Lustyk, T., Bergl, P., Čmejla, R., Vokřál, J. [**30%**]
In: 19th Annual Conference Proceedings Technical Computing Prague 2011. Technická 5, 16628 Praha: Vydavatelství VŠCHT Praha, 2011, p. 1–6. ISBN 978-80-7080-794-1.

“Assessment of voice and speech impairment”
Rusz, J., Čmejla, R., Bartošek, J., Janda, J., Lustyk, T., et al. [**12.5%**]
In: Workshop 2011, CTU Student Grant Competition in 2010 (SGS 2010). Praha: ČVTVS, 2011, p. 1–6.

“Hodnocení koktavosti a experimenty s adaptivním prahem”
Lustyk, T. [**100%**]
In: Analýza a zpracování řečových a biologických signálů – sborník prací 2010. Praha: České vysoké učení technické v Praze, 2010, s. 57–62. ISBN 978-80-01-04680-7.

“Assessment of Dysfluency in Stuttered Speech”
Bergl, P., Lustyk, T., Čmejla, R., Černý, L., Hrbková, M. [**30%**]
In: Technical Computing Bratislava 2010. Bratislava: RT systems, s.r.o, 2010, p. 1–3. ISBN 978-80-970519-0-7.

“Hodnocení neplynulosti promluv”
Bergl, P., Lustyk, T., Čmejla, R., Černý, L., Hrbková, M. [**30%**]
In: 8. ČESKO–SLOVENSKÝ FONIATRICKÝ KONGRES. Bratislava: Samedi s.r.o., 2010, s. 25. ISSN 1337–2181.

Appendices

A Passage used in read recordings (Czech)

The text used in recordings of the read speech consists of a passage from the book Babička (Grandmother), Božena Němcová. The text is composed of 74 words, it is phonetically non–balanced, and it does not include tongue twisters.

Podzim na Starém bělidle

V okolí Starého bělidla začínalo být smutno a ticho. Les byl světlejší, stráň žloutla, vítr a vlny odnášely chomáče starého listí buh ví kam. Ozdoba sadu uschována byla v komoře. V zahrádce kvetla astra, měsíčky a umrlčí kvítky. Na louce za splavem růžověly se naháčky a v noci prováděla tam světélka svoje rejdy. Když babička šla s dětmi na procházku, nezapomněli chlapci na papírové draky, které pak na vrchu pouštěli.

B The first picture used in spontaneous speaking task

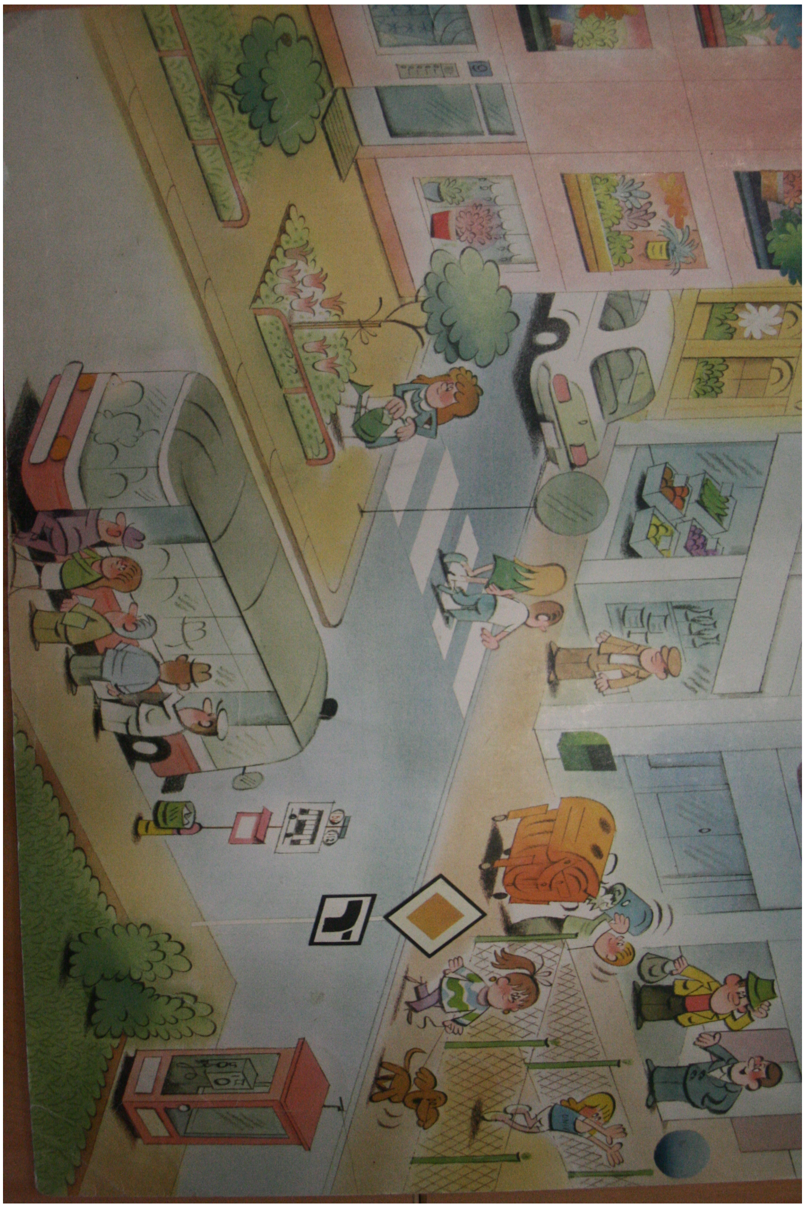


Figure 1: The first of the pictures that were described in the spontaneous speaking task. Positive situation.

C The second picture used in spontaneous speaking task

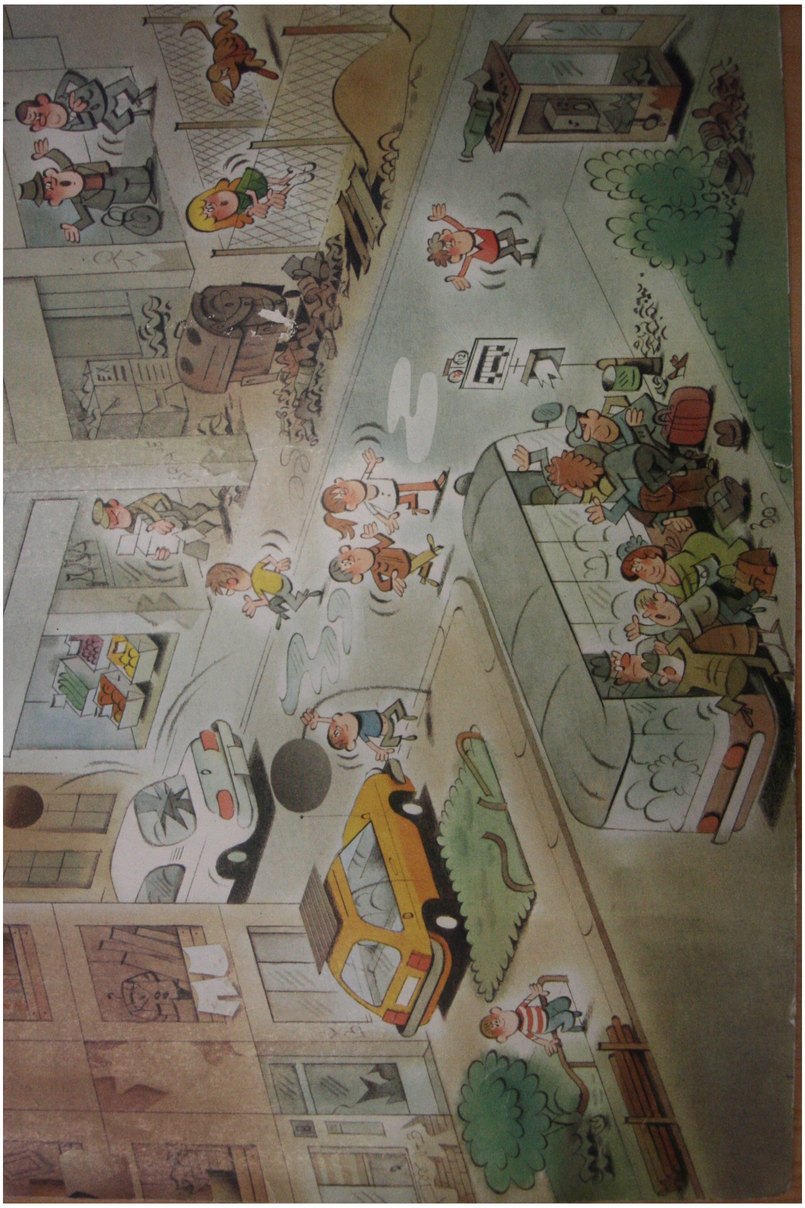


Figure 2: The second of the pictures that were described in the spontaneous speaking task. Negative situation.

D Passage used in read recordings (German)

A passage Nordwind und Sonne (North Wind and Sun) which was used when recording the read speech of German speakers. The text is 108 words long and it is a standard text used in Germany.

Nordwind und Sonne

Einst stritten sich Nordwind und Sonne, wer von ihnen beiden wohl der Stärkere wäre, als ein Wanderer, der in einen warmen Mantel gehüllt war, des Weges daherkam. Sie wurden einig, dass derjenige für den Stärkeren gelten sollte, der den Wanderer zwingen würde, seinen Mantel auszuziehen. Der Nordwind blies mit aller Macht, aber je mehr er blies, desto fester hüllte sich der Wanderer in seinen Mantel ein. Endlich gab der Nordwind den Kampf auf. Nun wärmte die Sonne die Luft mit ihren freundlichen Strahlen, und schon nach wenigen Augenblicken zog der Wanderer seinen Mantel aus. Da musste der Nordwind zugeben, dass die Sonne von ihnen beiden der Stärkere war.

E Instruction for speech specialist for evaluation of Czech recordings (in Czech)

Subjektivní hodnocení neplynulých promluv

Pro hodnocení neplynulých promluv je použit systém hodnocení LBDL (the Lidcombe Behavioral Data Language of Stuttering) a modifikovaná Kondášova stupnice. Oba použité systémy zde budou popsány krátce, ale snad dostatečně, více o hodnocení pomocí LBLD lze najít v článcích [1, 2] a o Kondášově stupnici v [3].

Nejdříve o LBDL. Důvody proč použít toto hodnocení jsou: 1) relativně snadno se dá použít (při poslechu se vlastně dělají čárky za dané projevy); 2) dává informaci o tom, kde má mluvčí problém (není to pouze celková známka, víme jestli má hodně repetic/prolongací/... a můžeme zjistit, která automatická měření fungují, na které projevy); 3) také je spolehlivá (používají-li hodnotitelé s praxí) a s celkem jasně danými projevy (za co dělat čárky).

Kategorie a sedm charakteristik (descriptors), které se hodnotí (je to překlad, proto je pro jistotu uveden i anglický originál s příklady popsanými níže v části příklady a v nahrávkách ve složce se stejným jménem):

koktavost

- Opakované projevy (repeated movements)
  1. repetice (opakování) slabik (syllable repetition)
  2. repetice hlásek (necelých slabik) (incomplete syllable repetition)
  3. repetice víceslabičných výrazů (multisyllable unit repetition)
- Fixované (strnulé) projevy (fixed postures)
  4. fixované projevy se slyšitelným zvukem (prolongace) (fixed postures with audible airflow)
  5. fixované projevy bez slyšitelného zvuku (pauzy) (fixed postures without audible airflow)
- Nadbytečné projevy (superfluous behaviors)
  6. verbální (ústní, mluvené) nadbytečné projevy (superfluous verbal behaviors)
  7. neverbální (neřečové) nadbytečné projevy (superfluous nonverbal behaviors)

Protože pro určení projevu 7 (nadbytečné neverbální projevy) je nutná videonahrávka, je tento projev vypuštěn a používají se zbylé, které je možné určit z audionahrávek. Ve formuláři má každý projev svoji kolonku, do které se dělají čárky (uslyší-li hodnotitel repetici "p p p podzim" udělá se čárka u repetice hlásek (2)). Pro některé druhy událostí, které se objeví v nahrávkách je možné přiřadit více druhů charakteristik z LBDL (věta z článku: Consequently, observers in the present study were given the option of assigning multiple descriptors to stuttering events). Tedy pro některé události je možné udělat více čárek (do různých kolonek, např. u tonoklonu "llll...lllll...lllistí" přijde čárka k repeticím hlásek ale i k prolongacím). Také se může objevit situace, kdy v jednom slovu bude přítomno více druhů událostí "p p po po pod ...(ticho) zim" (čárky za repetici hlásek, slabik a pauzu například).

Pro další zpracování je třeba určit celkový počet slov v promluvě. Počet slov pro všechny nahrávky je znám z dřívějšího zpracování nahrávek, ale pro kontrolu by bylo lepší pokud by každý hodnotitel určil počet slov, který zaznamenal.

Kondášova stupnice je běžně používána v ČR a tak zde bude uveden pouze přepis emailu (komunikace mezi Doc. Čmejlou a MUDr. Hrbkovou), který vše krátce vystihuje: Pro hodnocení jsme zvolili 5–ti stupňovou klasifikaci (0 – 4), modifikaci Kondášovy stupnice koktavosti. V našem pojetí jde o hodnocení pouze podle % neplynulých slov, nikoliv celkové hodnocení narušené komunikační schopnosti (viz [4] str. 331).

- 0 – žádné projevy neplynulých slov,
- 1 – lehká koktavost (balbuties levis), 1–5% neplynulých slov,
- 2 – střední koktavost (balbuties gradus medius), 6–20% neplynulých slov,
- 3 – těžká koktavost (balbuties gravis), 21–60% neplynulých slov,
- 4 – velmi těžká koktavost (balbuties gravis inaptus), nad 60% neplynulých slov či dlouhé prefonační spazmy (bloky) v délce trvání 2 s a více.

Vzhledem k tomu, že velké procento pacientů mělo problémy se čtením ať už ve smyslu pouhého přeřeknutí či dislexie, rozhodli jsme se po dohodě s Vámi o hodnocení dvojí: 1) pod označením *koktavost* je hodnocení podle % zakoktaných slov – tedy pouze projevů koktavosti ve smyslu tonů (např. slovo “máma” – “mmmáma”), klonů (“mma–mma–mmmáma”) či prefonačních spazmů (neschopnost slovo vyslovit i přes viditelnou snahu). 2) pod označením *neplynulost* je hodnocení podle % všech neplynulých slov včetně technických obtíží při čtení, prostého přeřeknutí či projevů specif. proruchy čtení (dislexie). Stupeň ohodnocení v kolonce *koktavost* a *neplynulost* tedy často neodpovídá. Kolonka *poznámka* je asi spíše pro nás (označení ”dlouhé” znamená dlouhé prefonační spazmy).

Snad email vystihuje použitou stupnici dostatečně, pokud ne, bylo by nejlepší se s dotazy obrátit přímo na MUDr. Hrbkovou nebo MUDr. Černého, kteří jako první hodnotili nahrávky (čtené). V tomto experimentu stačí vyplnit pouze známku za neplynulost, dle definice, jak je uvedeno výše ve zprávě MUDr. Hrbkové. Ve formuláři kolonka pro tuto známku je poslední sloupec s označením známka.

Vyplnění formuláře

První sloupec formuláře obsahuje čísla nahrávek, čísla odpovídají signálům umístěným ve složce signaly. Dále následují kolonky pro hodnocení pomocí LBDL, označeny čísly od 1 (opakování hlásek) až po 6 (nadbytečné verbální projevy), kde lze zaznamenávat čárky za jednotlivé projevy neplynulostí. Následuje kolonka počet slov, kde hodnotitel vyplní počet slov. Tyto údaje slouží pro určení celkového hodnocení pomocí LBDL.

Poslední sloupec je kolonka známka, zde hodnotitel vyplní známku dle modifikované Kondášovi stupnice (0, 1, 2, 3 nebo 4). Zde by bylo dobře upozornit, že počet neplynulostí (počet čárek v kolonkách 1 až 6 u LBDL) nemusí odpovídat počtu neplynulých slov, ze kterého se určuje známka pomocí Kondášovi stupnice, protože se v některých slovech může vyskytnout více neplynulostí než pouze jedna.

Ukázka vyplněného formuláře je umístěna v hlavní složce na přiloženém DVD, soubor – ukazkaVyplnenyFormular.jpg.

Přiložené DVD, databáze signálů a příklady

V nahrávkách jsou spontánní promluvy, které vznikaly jako popis dvou obrázků (obrázky používané na Fon. Klinice, oba si jsou velmi podobné, v jednom případě jde ale o obrázek města – pozitivní (lidé s dávají přednost, chodí po chodníku, chovají se slušně), druhý — negativní (lidé se hádají, rozbíjí věci)). Délka promluv se pohybuje přibližně okolo 90 s.

Všech 150 signálů určených pro ohodnocení je uloženo ve složce signaly. Ve složce priklady lze nalézt audio nahrávky k příkladům uvedeným u hodnocení LBDL a souvisejících s ukázkami uvedenými níže. V hlavní složce je kopie tohoto dokumentu (hodnoceni.pdf) a formulář pro vyplnění, které mohou být v případě potřeby vytisknuty (formular.pdf).

Příklady:

1. Repetice slabik - "když ba... ba... babička šla" ("where... where... where's the ball?"); ukázka je v souboru *priklad1.wav* ("vi...vítr"),

2. Repetice necelých slabik (hlásek) - "p... p... p... podzim" nebo "na S... S... S... Starém Bělidle" ("I went to S...S...Sydney..."); nahrávka *priklad2.wav* ("s p... p... p... plavem"),

3. Opakování víceslabičných výrazů - "začí... začí.... začínalo být smutno ..." ("it's a...it's a it's a great...", "what a great oper... oper... opertunity"); nahrávka *priklad3.wav* ("les byl... les byl"),

4. Fixované projevy se slyšitelným zvukem (prolongace) - "smmmmmutno" nebo "SSSSSSStarém" ("mmmmmmy one" "fffffishy gone!"); nahrávka *priklad4.wav* ("byyyyla v komoře"),

5. Fixované projevy bez slyšitelného zvuku - "Podzim na ... (bez zvuku) Starém Bělidle..." ("I ....(no sound) bought...") ukázka v nahrávce *priklad5a.wav* (pauza uvnitř slova "s...(ticho) tráň") a *priklad5b.wav* (pauza mezi dvěma slovy, tam kde by neměla být "kvetla ...(ticho) astra"),

6. Nadbytečné řečové/mluvené projevy - "Podzim na–eh na–eh–oh–o na–eh na Starém Bělidle..." zabručení, mručení, zachrochtání, ... ("I went–oh well–ah–oh well–I–well I went over...", grunting) ukázka v *priklad6.wav* ("(odkašlání)... e... eh.. eh...ee"),

7. Nadbytečné neverbální/nemluvené projevy – grimasy, záškuby těla, tiky, a-další (nutná video nahrávka) (tics, grimacing) bez ukázky (v tomto případě se nepoužívá).

V mnoha případech se nejedná pouze o jednotlivé projevy, ale o kombinaci více, a jak bylo popsáno výše hodnotitelé mají možnost tyto kombinované projevy zaznamenat do více kolonek (za jedno slovo, ve kterém se objeví více projevů koktavosti např. repetice a prolongace zároveň, udělat dvě čárky). Zde jsou uvedeny možné kombinace (nejsou zdaleka všechny):

priklad7.wav ("bje–e bje–e bje–ee bělidla") kombinace repetice slabik a nadbytečných řečových projevů (je tedy možné dát dvě čárky, jedna za repetice slabik a jedna za nadbytečné řečové projevy),

priklad8.wav ("ssss... sssstarého") kombinace repetic hlásek a prolongace (tonoklon), tedy dvě čárky, jedna pro repetice hlásek a jedna pro prolongace,

priklad9.wav ("l...l...les byl...les byl") kombinace repetic hlásek a více slabičných výrazů, je tedy možné připsat čárku za repetice hlásek a opakování více slabičných výrazů,

priklad10.wav ("na lll... na lll... na lllouce") kombinace repetic víceslabičných výrazů a prolongace, můžeme udělat čárku pro opakování víceslabičných výrazů a čárku pro prolongace,

priklad11.wav ("e... vzahrádce... (ticho) ...ee... kvetla") spojení pauzy a nadbytečných řečových událostí, možné dát jednu čárku za pauzu a jednu za nadbytečné projevy,

priklad12.wav ("u... u... schoooovááááána býýýýla v komoře") toto je spíše ukázka, jak může vypadat část hodnoceného signálu, kde se objeví více projevů zároveň, repetice, prolongace,...

U uvedených příkladů a jim podobným bych se spíše klonil k tomu, aby byly zaznamenávány takto vícenásobně, ale je na hodnotiteli, zda to bude dělat (jak je uvedeno, je to pouze možnost, není to požadováno). V některých případech se může stát, že daná část signálu nebude úplně jednoznačná a v tom případě je na hodnotiteli, aby se rozhodl, co lze zaznamenat a co ne.

Literatura

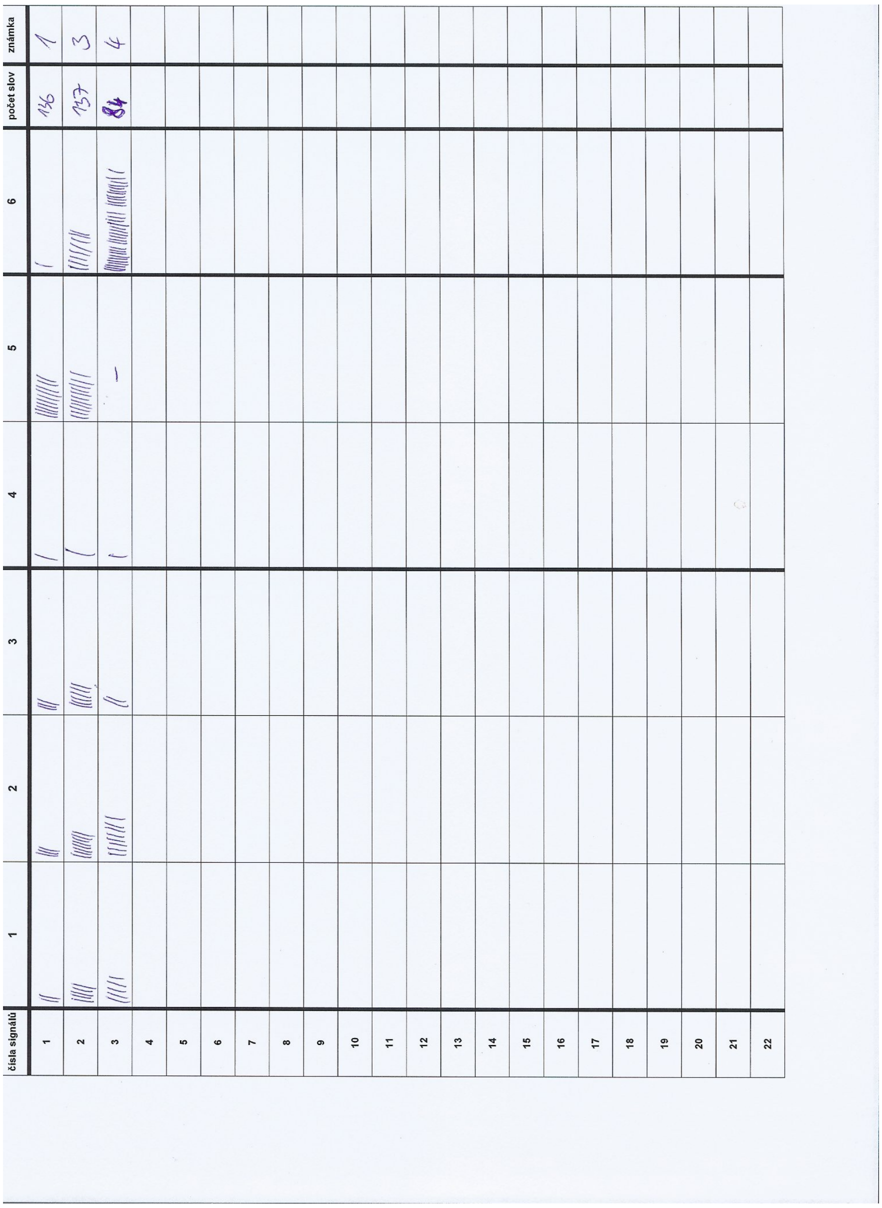
[1] Teesson, K., Packman, A., Onslow, M., (2003) "The Lidcombe Behavioral Data Language of stuttering", J. Speech Lang. Hear. Res. 46, 1009–1015.

[2] Packman, A., Onslow, M. (1998) "The behavioral data language of stuttering", In A.K. Cordes and R.J. Ingham (Eds.), *Treatment efficacy for stuttering: Search fro empirical bases*, (pp. 27–50). San Diego, CA:Singular.

[3] Škodová, E., Jedlička, I., a kolektiv (2003) *Klinická logopedie* (Portál).

[4] Lechta V. a kol. (2003) *Diagnostika narušené komunikační schopnosti* (Portál), Praha.

F Form for subjective evaluation



čísla signálů	1	2	3	4	5	6	počet slov	známka
1	//	///	///	/	/////////	/	136	1
2	/////	//////	/////	/	//////////	////////	137	3
3	/////	////////	//	/	/	/////////////// ////////// //////////	84	4
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								
22								

Figure 3: Partly filled form for subjective evaluation of recordings.

G Histograms of subjective evaluation of spontaneous recordings, effect of speaking task

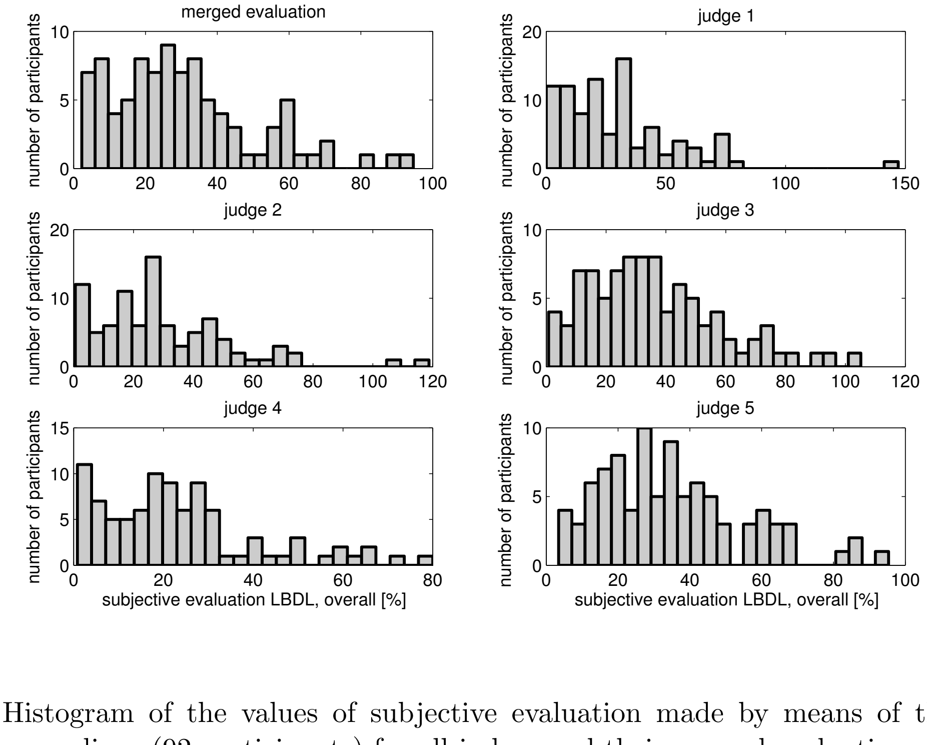


Figure 4: Histogram of the values of subjective evaluation made by means of the LBDL on spontaneous recordings (92 participants) for all judges and their merged evaluation.

H Comparison of all measurements with evaluation of speech-language pathologists (LBDL) in spontaneous recordings

TABLE XXIII: The Pearson correlation coefficients and the levels of significance (in parentheses when $p > 0.001$) for one selected setting of each measure in comparison to the LBDL descriptors in evaluation of speech-language pathologists.

	measure								
descriptor	ALS	RSE	REV	ESF	SET	SDI11	SCSI	NSI	ROS
SR	0.18	0.08	-0.08	-0.03	0.10	-0.07	-0.01	-0.22	-0.18
ISR	0.35	0.24	0.24	-0.16	0.25	0.12	-0.13	-0.33	-0.34
MSUR	0.29	0.19	0.13	-0.15	0.12	0.18	-0.14	-0.23	-0.23
FPWAA	0.36	0.34	0.25	-0.38	0.48	0.27	-0.36	-0.47	-0.41
FPWOAA	0.72	0.42	0.72	-0.57	0.38	0.55	-0.57	-0.72	-0.73
SVB	0.19	0.06	0.21	0.03	0.00	-0.01	0.04	-0.11	-0.26
repeated	0.33	0.20	0.10	-0.12	0.20	0.05	-0.10	-0.33	-0.31
fixed	0.57	0.43	0.50	-0.53	0.51	0.44	-0.51	-0.65	-0.61
overall	0.52	0.33	0.38	-0.30	0.35	0.23	-0.28	-0.52	-0.52