

CZECH TECHNICAL UNIVERSITY IN PRAGUE

**Faculty of Electrical Engineering**

*Doctoral Thesis*

2018

*Jan Hlavnička*



CZECH TECHNICAL UNIVERSITY IN PRAGUE

Faculty of Electrical Engineering

Department of Circuit Theory



# AUTOMATED ANALYSIS OF SPEECH DISORDERS IN NEURODEGENERATIVE DISEASES

Doctoral Thesis

*by*

**Jan Hlavnička**

Ph.D. Program: Electrical Engineering and Information Technology

Branch of Study: Electrical Engineering Theory

Prague, 2018

Supervisor: Prof. Ing. Roman Čmejla, CSc.

Assistant supervisor: Ing. Jan Ruzs, Ph.D



# ACKNOWLEDGEMENT

I would like to thank my assistant supervisor Jan Ruzs for his guidance and supervisor Roman Čmejla for his forbearance. Both mentors gave me the freedom to pursue my research ideas and supported me with books, advice, and encouragement. I also want to thank speech-language pathologist Hana Růžičková, who was a great source of inspiration and devoted her time and effort to testing the methodology in a clinical setting.

This thesis represents an outgrowth of studies that were supported financially by the Czech Science Foundation under grant No. 6-03322S, grant No. 16-07879S, grant No. 16-03322S, grant No. 16-19975S, and grant No. 102/12/2230, the Czech Ministry of Health under grant No. 15-28038A and grant No. 16-28914A, the Charles University in Prague under grant No. PRVOUK-P26/LF1/4, the Czech Technical University in Prague under grant No. SGS12/185/OHK4/3T/13, and SGS15/199/OHK3/3T/13, and the Czech Ministry of Education, Youth and Sports, OP VVV MEYS under grant No. CZ.02.1.01/0.0/0.0/16\_019/0000765.

# AFFIDAVIT

I hereby declare that my thesis entitled is the result of my own work and includes nothing, which is the outcome of work done in collaboration except where specifically indicated in the text. It has not been previously submitted, in part or whole, to any university of institution for any degree, diploma, or other qualification.

.....

Prague, December 19, 2018

Jan Hlavnička



# ABSTRACT

Automated vocal biomarkers are becoming increasingly desired by speech pathologists and neurologists in order to extend current noninvasive measures of speech motor abnormalities associated with neurodegeneration. Clinical information concerning acoustical features and patterns can be invaluable only if the measures are based on interpretable hypotheses and described with regard to the impact of the disease, sexual dimorphism, and any age dependency. The complexity of interpretation is the main barrier between engineering applications and clinical practice. Despite huge developments in the field, no applicable methodology for complex acoustic analysis have been proposed yet. This thesis aims to design and define the automated acoustic analysis that could provide profound insight into speech disorders caused by neurodegeneration.

The database used in this research is comprised of 42 subjects with idiopathic rapid eye movement sleep behavior disorder; 32 subjects with early, untreated Parkinson's disease; 26 subjects with treated Parkinson's disease; 22 subjects with multiple system atrophy; 15 subjects with progressive supranuclear palsy; 18 subjects with untreated Huntington's disease; 13 subjects with treated Huntington's disease; 17 subjects with cerebellar ataxia; 101 subjects with multiple sclerosis; and 284 subjects with no history of neurological or communication disorders (HC). Each speaker performed the sustained vowels /A/ and /I/, took a rhythm test, read a passage, performed a monologue, and completed a diadochokinetic task. Acoustic signals were recorded using a standardized procedure. Signals were processed by fully automated methods. Normative data were estimated by selecting an HC subgroup to match any speaker in terms of age and sex. All measured values were normalized by corresponding normative data and expressed in terms of probabilities and z-scores. A novel approach for supervised learning based on the weighted fusion of z-scores (SWFS) was employed for recognition of certain tendencies of disordered speech. Finally, the methodology was implemented in a software application and tested extensively in a clinical setting by an experienced speech-language pathologist for more than one year.

Based on a thorough evaluation, the proposed processing methods represent the most precise technology for the extraction of given acoustic features available up to the date of this thesis. The majority of speech features showed abnormalities in at least one disease group compared to the HC. Individual speech features did not exhibit specificity to disease. Nevertheless, clear tendencies with discriminative qualities were observed in combined features. The SWFS showed the ability to decompose any speech pattern and quantify its severity in terms of abnormalities, whereas the recognition accuracy was comparable with conventional classifiers. The clinician rated the methodology as practicable, clinically relevant, interpretable, and of benefit. Two case studies are presented to demonstrate the capacity of the proposed methodology.

This thesis introduces a methodology for the extraction of highly interpretable speech features using a new approach in digital signal processing, machine learning, and the modeling of sexual dimorphism and age dependency; investigates a large database of patients affected by neurodegeneration; and discusses clinical applicability based on the successful experimental use of the implementation in a clinical setting. The methodology was designed to meet the demands of clinical practice with a hope that the presented results will lead, inspire, and bolster the future development of automated methods for the assessment of speech disorders.

**Key words:** *Speech disorders, Neurodegeneration, Parkinson's disease, Rapid eye movement sleep behavior disorder, Multiple system atrophy, Progressive supranuclear palsy, Huntington's disease, Cerebellar ataxia, Multiple sclerosis, Dysarthria, Acoustic analysis, Speech pattern recognition.*

# ABSTRAKT

Biomarkery získané automatickou analýzou hlasu se těší rostoucímu zájmu logopedů i neurologů v souvislosti s možností rozšířit dosud značně limitovaná neinvazivní měření motorických poruch řeči způsobených neurodegenerativními onemocněními. Akustické řečové příznaky mohou být v klinické praxi vskutku neocenitelné, avšak pouze tehdy, jsou-li podloženy vysvětlitelnými hypotézami a popsány z hlediska dopadu onemocnění, pohlavní dvojtvárnosti a vlivu stárnutí. Spletitost interpretace těchto faktorů tvoří hlavní překážku bránící využití hlasových analýz v klinické praxi, která navzdory značnému rozvoji tohoto oboru nebyla dosud překonána. Tato práce zavádí metodologii pro získání srozumitelných akustických příznaků pomocí číslicového zpracování signálů a strojového učení a modelování pohlavní dvojtvárnosti a vlivu stárnutí; vyšetřuje velkou databázi pacientů s neurodegenerativními onemocněními a diskutuje použitelnost metody na základě experimentálního odzkoušení metody v klinické praxi.

Databáze zahrnovala 42 pacientů s idiopatickou poruchou chování v REM spánku (REM = rapid eye movement, česky: rychlé pohyby očí), 32 neléčených pacientů v rané fázi Parkinsonovy nemoci, 26 léčených pacientů Parkinsonovy nemoci, 22 pacientů s multisystémovou atrofií, 15 pacientů s progresivní supranukleární obrnou, 18 neléčených pacientů s Huntingtonovou nemocí, 13 léčených pacientů Huntingtonovy nemoci, 17 pacientů s mozečkovou ataxií, 101 pacientů s roztroušenou sklerózou a 274 zdravých kontrolních subjektů, kteří nevykazují a nikdy neprodělali neurologickou poruchu ani poruchu komunikace. Každý účastník provedl úlohu prodloužené fonace hlásky /A/ a /I/, rytmický test, čtení textu, monolog a diadochokinetický test. Akustické signály byly nahrány standardizovanou procedurou. Signály byly zpracovány automatickým algoritmem. Pro každého možného řečníka byly ze skupiny zdravých kontrolních subjektů vybrány subjekty srovnatelné věkové skupiny a pohlaví a na jejich základě byla odhadnuta normativní data. Všechna měření byla normalizována pomocí normativních dat a vyjádřena jako pravděpodobnost a z-skóre (SWFZ). Nový přístup v rozpoznávání vzorů učených s učitelem založený na vážené fúzi z-skóre byl použit k popisu základních tendencí řečových poruch. Celá metodologie byla nakonec implementována do podoby softwarové aplikace a testována po dobu více než jednoho roku zkušeným logopedem v podmínkách klinické praxe.

Důkladná analýza ukázala, že navržené metody zpracování signálů představují v současnosti nejpřesnější technologie pro měření příslušných akustických příznaků. Jednotlivé příznaky nebyly specifické pro jednotlivá onemocnění, avšak kombinace příznaků ukázala specifické a rozlišitelné tendence řečových poruch. Navržená metoda SWFZ projevila rozpoznávací přesnost těchto tendencí srovnatelnou s běžnými klasifikátory, přičemž umožňuje rozložit tyto tendence na jednotlivé komponenty a odhadnout tíži poruchy. Metoda byla testováním v praxi ohodnocena jako použitelná, prospěšná a poskytující klinicky relevantní a interpretovatelné výsledky. Způsobnost metody byla demonstrována na dvou kauzistikách.

Prezentovaný proces automatické analýzy řeči poskytuje výsledky nezkrácené pohlavní dvojtvárností a vlivem stárnutí a umožňuje získat hluboký vhled do řečové poruchy způsobené neurodegenerací. Metodologie byla navržena pro uspokojení nároků klinické praxe s nadějí, že prezentované výsledky povedou, inspirují a podpoří budoucí vývoj automatických metod pro ohodnocení řečových poruch u neurodegenerativních onemocnění.

**Key words:** *Poruchy řeči, Neurodegenerace, Parkinsonova nemoc, Porucha chování v REM spánku, Multisystémová atrofie, Progresivní supranukleární obrna, Huntingtonova nemoc, Cerebelární ataxie, Roztroušená skleróza, Dysartrie, Akustická analýza, Rozpoznávání řečových vzorů.*

# TABLE OF CONTENTS

|  |             |
|--|-------------|
| <b>ACKNOWLEDGEMENT</b> .....   | <b>iii</b>  |
| <b>AFFIDAVIT</b> .....   | <b>iii</b>  |
| <b>ABSTRACT</b> .....  | <b>v</b>    |
| <b>ABSTRAKT</b> .....  | <b>vii</b>  |
| <b>TABLE OF CONTENTS</b> .....                                       | <b>ix</b>   |
| <b>LIST OF TABLES</b> .....  | <b>xiii</b> |
| <b>LIST OF FIGURES</b> .....   | <b>xiii</b> |
| <b>LIST OF EQUATIONS</b> .....                                       | <b>xv</b>   |
| <b>NOMENCLATURE</b> .....  | <b>xvii</b> |
| <b>1 INTRODUCTION</b> .....  | <b>1</b>    |
| 1.1 MOTOR SPEECH DISORDERS.....                                      | 2           |
| 1.2 SELECTED DISEASES AND PRECURSORS.....                            | 2           |
| 1.2.1 <i>Parkinson's disease</i> .....                               | 2           |
| 1.2.2 <i>Atypical parkinsonian syndromes</i> .....                   | 2           |
| 1.2.3 <i>Rapid eye movement sleep behavior disorder</i> .....        | 3           |
| 1.2.4 <i>Huntington's disease</i> .....                              | 3           |
| 1.2.5 <i>Multiple sclerosis</i> .....                                | 3           |
| 1.2.6 <i>Cerebellar ataxia</i> .....                                 | 4           |
| 1.3 EXAMINATION OF DYSARTHRIA.....                                   | 4           |
| 1.4 ON THE DECOMPOSITION OF SPEECH PROCESSES.....                    | 7           |
| 1.5 AUTOMATED ANALYSIS OF DYSARTHRIA.....                            | 7           |
| 1.5.1 <i>Acoustic analysis</i> .....                                 | 8           |
| Connected speech.....  | 8           |
| Rhythm test.....   | 9           |
| Diadochokinetic test.....  | 10          |
| Sustained vowels.....  | 11          |
| 1.5.2 <i>Modeling of speech patterns</i> .....                       | 12          |
| 1.6 AIMS AND OBJECTIVES.....   | 14          |
| <b>2 METHOD</b> .....  | <b>17</b>   |
| 2.1 DATABASE.....  | 18          |
| 2.2 RECORDING PROCESS.....   | 19          |
| 2.3 ACOUSTIC ANALYSIS.....   | 20          |
| 2.3.1 <i>Sustained vowels</i> .....                                  | 20          |
| Segmentation.....  | 20          |
| Analysis of the modal and subharmonic vibrations of vocal folds..... | 21          |
| Speech features.....   | 25          |
| 2.3.2 <i>Rhythm test</i> .....                                       | 28          |
| Segmentation.....  | 28          |
| Speech features.....   | 31          |
| 2.3.3 <i>Connected speech</i> .....                                  | 32          |
| Segmentation.....  | 32          |
| Speech features.....   | 33          |
| 2.3.4 <i>Diadochokinetic test</i> .....                              | 39          |

|   |           |
|---|-----------|
| Segmentation.....   | 39        |
| Speech features .....   | 41        |
| 2.4 MODELING OF SPEECH PATTERNS .....                         | 42        |
| 2.4.1 Normalization.....                                      | 42        |
| 2.4.2 Combination of probabilities.....                       | 43        |
| 2.4.3 Pattern analysis.....                                   | 44        |
| 2.4.4 Pattern decomposition.....                              | 45        |
| 2.4.5 Excitatory and inhibitory speech patterns.....          | 45        |
| 2.5 INTERPRETATION AND VISUALIZATION OF RESULTS.....          | 46        |
| 2.6 STATISTICAL ANALYSIS .....                                | 47        |
| 2.7 CLASSIFICATION EXPERIMENT.....                            | 48        |
| 2.8 QUESTIONNAIRE FEEDBACK FROM CLINICIAN.....                | 49        |
| <b>3 RESULTS .....</b>  | <b>51</b> |
| 3.1 TRACKING THE ACCURACY OF THE ANALYSIS .....               | 52        |
| 3.1.1 Connected speech.....                                   | 52        |
| 3.1.2 Rhythm .....  | 53        |
| 3.1.3 Diadochokinetic task .....                              | 53        |
| 3.1.4 Sustained vowels.....                                   | 54        |
| 3.2 STATISTICAL ANALYSIS .....                                | 56        |
| 3.3 CLASSIFICATION EXPERIMENT.....                            | 57        |
| 3.4 QUESTIONNAIRE FEEDBACK .....                              | 61        |
| 3.5 CASE STUDIES.....   | 61        |
| 3.5.1 Case A.....   | 61        |
| Neurological diagnosis .....                                  | 61        |
| Speech-language-swallowing pathology diagnosis .....          | 62        |
| Therapy of speech and swallowing .....                        | 63        |
| Acoustic analysis .....                                       | 64        |
| 3.5.2 Case B.....   | 70        |
| Neurological diagnosis .....                                  | 70        |
| Speech-language-swallowing pathology diagnosis .....          | 70        |
| Therapy of speech and swallowing .....                        | 71        |
| Acoustic analysis .....                                       | 72        |
| <b>4 DISCUSSION .....</b>                                     | <b>79</b> |
| 4.1 ACOUSTIC ANALYSIS.....                                    | 80        |
| 4.1.1 Connected speech.....                                   | 80        |
| Segmentation.....   | 80        |
| Speech features .....   | 80        |
| 4.1.2 Rhythm .....  | 81        |
| Segmentation.....   | 81        |
| Speech features .....   | 82        |
| 4.1.3 Sustained vowels.....                                   | 82        |
| Segmentation.....   | 82        |
| Speech features .....   | 83        |
| 4.1.4 Diadochokinetic test.....                               | 84        |
| Segmentation.....   | 84        |
| Speech features .....   | 85        |
| 4.2 SPEECH PATTERNS .....                                     | 85        |
| 4.3 CLINICAL APPLICABILITY .....                              | 87        |
| 4.4 LIMITATIONS AND FUTURE STEPS .....                        | 89        |
| <b>5 CONCLUDING REMARKS.....</b>                              | <b>93</b> |
| <b>APPENDIX A: NORMATIVE DATA FOR THE CZECH LANGUAGE.....</b> | <b>95</b> |

|   |            |
|---|------------|
| <b>APPENDIX B: NORMALIZED VALUES OF SPEECH FEATURES .....</b> | <b>105</b> |
| <b>APPENDIX C: SOFTWARE APPLICATION.....</b>                  | <b>115</b> |
| <b>APPENDIX D: QUESTIONNAIRE FEEDBACK .....</b>               | <b>121</b> |
| <b>REFERENCES.....</b>  | <b>127</b> |
| <b>LIST OF AUTHOR'S PUBLICATIONS AND RECOGNITION .....</b>    | <b>137</b> |
| PUBLICATIONS RELATED TO THE DOCTORAL THESIS .....             | 137        |
| <i>Articles in journals with impact factor .....</i>          | <i>137</i> |
| <i>Articles in peer-reviewed journals.....</i>                | <i>138</i> |
| <i>Other articles indexed by the SCOPUS.....</i>              | <i>138</i> |
| <i>Other articles and abstracts.....</i>                      | <i>138</i> |
| OTHER PUBLICATIONS .....                                      | 139        |
| <i>Other articles and abstracts.....</i>                      | <i>139</i> |
| <i>Other articles indexed by the SCOPUS.....</i>              | <i>139</i> |
| CITATIONS INDEXED IN THE WEB OF SCIENCE AND SCOPUS .....      | 140        |
| AWARDS .....  | 142        |



# LIST OF TABLES

|  |    |
|--|----|
| TABLE 1: SUMMARY OF DYSARTHRIA CATEGORIES. ....  | 5  |
| TABLE 2: CLINICAL CHARACTERISTICS OF ALL GROUPS IN THE DATABASE. ....  | 18 |
| TABLE 3: CORRELATIONS BETWEEN THE REFERENCE AND AUTOMATED SPEECH FEATURES. ....                              | 55 |
| TABLE 4: SEGMENTATION ACCURACY IN SUSTAINED VOWELS EXPRESSED IN PERCENT. ....                                | 56 |
| TABLE 5: MEDIAN PREDICTION ERRORS MEASURED ON THE DATABASE OF SYNTHETIC PHONATIONS. ....                     | 57 |
| TABLE 6: SUMMARY OF ACOUSTIC FEATURES MEASURED ON DIADOCHOKINETIC TASK, RHYTHM, AND SUSTAINED VOWELS. ....   | 58 |
| TABLE 7: SUMMARY OF ACOUSTIC FEATURES MEASURED ON CONNECTED SPEECH. ....                                     | 59 |
| TABLE 8: INCIDENCES OF SPEECH PATTERNS BY RANDOMIZED STRATIFIED CROSS-VALIDATION. ....                       | 60 |
| TABLE 9: SUMMARY OF MOST SEVERE SPEECH FEATURES OF CASE A MEASURED AT THE FIRST RECORDING SESSION. ....      | 65 |
| TABLE 10: SUMMARY OF THE MOST SEVERE SPEECH FEATURES OF CASE B MEASURED IN THE FIRST RECORDING SESSION. .... | 73 |

# LIST OF FIGURES

|  |    |
|--|----|
| FIGURE 1: ILLUSTRATED OBJECTIVES OF THE THESIS. ....   | 15 |
| FIGURE 2: PROCESS DIAGRAM ILLUSTRATING THE ANALYSIS OF MODAL AND SUBHARMONICS VIBRATIONS. ....       | 24 |
| FIGURE 3: ILLUSTRATION OF PERTURBATION ANALYSIS. ....  | 28 |
| FIGURE 4: PROCESS DIAGRAM OF SYLLABLE IDENTIFICATION. ....   | 30 |
| FIGURE 5: ILLUSTRATION OF DESIGNED RHYTHM FEATURES. ....   | 31 |
| FIGURE 6: AUTOMATED SEGMENTATION OF CONNECTED SPEECH. ....   | 34 |
| FIGURE 7: ILLUSTRATION OF THE NORMALIZATION PROCESS. ....  | 43 |
| FIGURE 8: DETECTION EFFICIENCY OF PAUSE AND RESPIRATORY INTERVALS IN CONNECTED SPEECH. ....          | 52 |
| FIGURE 9: CUMULATIVE DISTRIBUTION OF SEGMENTATION ERRORS IN THE DIADOCHOKINETIC TASK. ....           | 54 |
| FIGURE 10: ACCURACY OF $F_0$ DETECTION BY THE PROPOSED METHOD AND PUBLICLY AVAILABLE DETECTORS. .... | 57 |
| FIGURE 11: INCIDENCES ESTIMATED BY THE LEAVE-ONE-OUT CROSS-VALIDATION EXPERIMENT. ....               | 61 |
| FIGURE 12: ILLUSTRATED RESULTS OF CASE A MEASURED AT THE FIRST RECORDING SESSION. ....               | 66 |
| FIGURE 13: ILLUSTRATED RESULTS OF THE CASE A MEASURED AT THE LAST RECORDING SESSION. ....            | 67 |
| FIGURE 14: SPEECH PATTERNS OF THE CASE A MEASURED AT THE FIRST RECORDING SESSION. ....               | 68 |
| FIGURE 15: LONGITUDINAL DATA OF SELECTED SPEECH FEATURES MEASURED ON CASE A. ....                    | 69 |
| FIGURE 16: ILLUSTRATED RESULTS OF CASE B AS MEASURED IN THE FIRST RECORDING SESSION. ....            | 74 |
| FIGURE 17: ILLUSTRATED RESULTS OF CASE B MEASURED IN THE LAST RECORDING SESSION. ....                | 75 |
| FIGURE 18: SPEECH PATTERNS FOR CASE B MEASURED IN THE FIRST RECORDING SESSION. ....                  | 76 |
| FIGURE 19: LONGITUDINAL DATA OF SELECTED SPEECH FEATURES MEASURED ON CASE B. ....                    | 77 |



# LIST OF EQUATIONS

|                  |    |
|------------------|----|
| EQUATION 1.....  | 10 |
| EQUATION 2.....  | 10 |
| EQUATION 3.....  | 22 |
| EQUATION 4.....  | 22 |
| EQUATION 5.....  | 22 |
| EQUATION 6.....  | 23 |
| EQUATION 7.....  | 23 |
| EQUATION 8.....  | 23 |
| EQUATION 9.....  | 23 |
| EQUATION 10..... | 23 |
| EQUATION 11..... | 23 |
| EQUATION 12..... | 23 |
| EQUATION 13..... | 23 |
| EQUATION 14..... | 24 |
| EQUATION 15..... | 24 |
| EQUATION 16..... | 24 |
| EQUATION 17..... | 27 |
| EQUATION 18..... | 27 |
| EQUATION 19..... | 27 |
| EQUATION 20..... | 29 |
| EQUATION 21..... | 31 |
| EQUATION 22..... | 32 |
| EQUATION 23..... | 32 |
| EQUATION 24..... | 32 |
| EQUATION 25..... | 32 |
| EQUATION 26..... | 32 |
| EQUATION 27..... | 36 |
| EQUATION 28..... | 40 |
| EQUATION 29..... | 40 |
| EQUATION 30..... | 40 |
| EQUATION 31..... | 43 |
| EQUATION 32..... | 44 |
| EQUATION 33..... | 45 |
| EQUATION 34..... | 45 |
| EQUATION 35..... | 55 |
| EQUATION 36..... | 55 |
| EQUATION 37..... | 55 |
| EQUATION 38..... | 55 |
| EQUATION 39..... | 55 |



# NOMENCLATURE

| Abbreviation | Meaning   | Abbreviation       | Meaning   |
|--------------|---|--------------------|---|
| AMR          | Alternating motion rate   | RMSE               | Root mean square error of fundamental frequency in semitones              |
| APS          | Atypical parkinsonian syndromes                                       | RFA                | Resonant frequency attenuation  |
| AST          | Acceleration of speech timing   | RSR                | Rate of speech respiration  |
| BACD         | Bayesian autoregressive change-point detector                         | RST                | Acceleration of speech timing   |
| BSCD         | Bayesian step change-point detector                                   | SARA               | Scale for the Assessment and Rating of Ataxia                             |
| CA           | Cerebellar ataxia   | SD                 | Standard deviation  |
| CPSD         | Cepstrum of power spectral density                                    | SDE                | Standard deviation of error of fundamental frequency in semitones         |
| DAB          | Diagnostic system introduced in studies by Darley, Aronson, and Brown | stdFO              | Standard deviation of fundamental frequency                               |
| DDKI         | Diadochokinetic irregularity  | stdPSD             | Standard deviation of power spectral density                              |
| DDKR         | Diadochokinetic rate  | stdPWR             | Standard deviation of power   |
| DFA          | Detrended fluctuation analysis  | SHR                | Subharmonic-to-harmonic ratio   |
| DPI          | Duration of pause intervals   | SVG                | Scalable Vector Graphics  |
| DUF          | Decay of unvoiced fricatives  | SVM                | Support vector machine  |
| DUS          | Duration of unvoiced stops  | SWFZ               | Supervised weighted fusion of z-scores                                    |
| DVA          | Degree of vocal arrests   | TEO                | Teager energy operator  |
| DVI          | Duration of voiced intervals  | TH                 | Threshold   |
| EDSS         | Expanded Disability Status Scale                                      | UHDRS              | Unified Huntington's Disease Rating Scale                                 |
| Efn_M        | Degree of hypernasality   | UPDRS III          | Unified Parkinson's Disease Rating Scale motor score                      |
| Efn_SD       | Intermittend hypernasality  | VD                 | Vowel duration  |
| EM           | Expectation-maximization algorithm                                    | VOT                | Voice onset time  |
| EST          | Entropy of speech timing  | ZCR                | Zero-crossing rate  |
| FO           | Fundamental frequency   |                    |   |
| FFT          | Fast Fourier transformation   | Symbol             | Meaning   |
| GMM          | Gaussian mixture model  | $A$                | Amplitude   |
| GUI          | Graphical user interface  | $avIntDur_{1-4}$   | Average intervals between syllables of the sequence 1-4 in rhythm task    |
| GVI          | Gaping in-between voiced intervals                                    | $avIntDur_{5-12}$  | Average intervals between syllables of the sequence 5-12 in rhythm task   |
| HC           | Healthy control   | $avIntDur_{13-20}$ | Average intervals between syllables of the sequence 13-20 in rhythm task  |
| HD           | Huntington's disease  | $C_i$              | Relative contribution of the speech feature                               |
| HNR          | Harmonics-to-noise ratio  | $COV_{5-20}$       | Coefficient of variation of syllables of the sequence 5-20 in rhythm task |
| HTML         | HyperText Markup Language   | dB                 | Decibel   |
| LFCC         | Linear-frequency cepstral coefficients                                | $D_M$              | Mahalanobis distance  |
| LPSD         | Cepstrally lifted power spectral density                              | $e$                | Natural exponential function  |
| LRE          | Latency in respiratory exchange                                       | $e_n$              | Error of the estimated value of fundamental frequency                     |
| LSI          | Location of subharmonic intervals                                     | $f$                | Logistic function   |
| MAE          | Median absolute error   | $f_0$              | Value of modal fundamental frequency                                      |
| ME           | Mean semitone error of fundamental frequency                          | $\dot{f}_0$        | Derivation of modal fundamental frequency                                 |
| MFCC         | Mel-frequency cepstral coefficients                                   | $\hat{f}_x$        | Value of a feature  |
| MPAF         | Maximal peak in the autocorrelation function                          | $F$                | State transition model of Kalman filter                                   |
| MPT          | Maximum phonation time  | $F_x$              | A feature   |
| MS           | Multiple sclerosis  | $g$                | Template of the cross-correlation function                                |
| MSA          | Multiple system atrophy   | $\bar{g}$          | Average of template of cross-correlation function                         |
| N/A          | Not available   | $\hat{g}$          | Normalized template of the cross-correlation function                     |
| NSR          | Net speech rate   | $G$                | Total number of hypotheses  |
| NNIPPS       | Natural history and neuroprotection on Parkinson Scale                | $h$                | Hamming window  |
| PD           | Parkinson's disease   | $H$                | Matrix mapping input measurement to space observed in Kalman filter       |
| PDU          | Early untreated Parkinson's disease                                   | Hz                 | Herz  |
| PDT          | Treated Parkinson's disease   | $i$                | Index in series   |
| PIR          | Pause intervals per respiration                                       | $j$                | Imaginary unit  |
| PSD          | Power spectral density  | $J$                | Cost function   |
| PSI          | Proportion of subharmonic intervals                                   | $k$                | Index in series   |
| PSP          | Progressive supranuclear palsy  |                    |   |
| PWR          | Power of the signal   |                    |   |
| RA           | Rhythm acceleration   |                    |   |
| RBD          | Rapid eye movement sleep behavior disorder                            |                    |   |
| RI           | Rhythm instability  |                    |   |
| RLR          | Relative loudness of respiration                                      |                    |   |

## Nomenclature

---

| Symbol            | Meaning   |
|-------------------|---|
| $K$               | Kalman gain   |
| $L$               | Degrees of freedom of the Chi-square distribution                                     |
| ms                | Millisecond   |
| $M$               | Length of series  |
| $n$               | Index in series   |
| $n_p$             | Number of pause intervals   |
| $n_r$             | Number of respiratory intervals   |
| $n_t$             | Total number of intervals   |
| $n_u$             | Number of unvoiced intervals  |
| $n_v$             | Number of voiced intervals  |
| $N$               | Length of series  |
| $p$               | Probability   |
| $PA$              | Pace acceleration   |
| $P_t$             | Error covariance matrix   |
| $P_X$             | Power of inlier   |
| $P_Y$             | Power of outlier  |
| $q$               | Percentile  |
| $Q$               | Covariance matrix of the process noise  |
| $r$               | Correlation coefficient   |
| $R_t$             | Covariance of observation noise   |
| $R_x$             | Normalized autocorrelation function   |
| $s$               | Steepness of the logistic function  |
| $sdIntDur_{5-20}$ | Standard deviation of intervals between syllables of the sequence 5-20 in rhythm task |
| $S$               | Covariance matrix   |
| $t$               | Time  |
| $T$               | Period  |
| $u_n$             | Reference value of fundamental frequency  |
| $\hat{u}_n$       | Estimated value of fundamental frequency  |
| $v_t$             | Normally distributed process noise  |
| $w_i$             | Weight assigned to the hypothesis   |
| $w_t$             | Normally distributed observation noise  |
| $W$               | Set of optimized weights of hypotheses  |
| $x$               | Signal  |
| $\bar{x}$         | Average of signal   |
| $\hat{x}$         | Prediction of modal fundamental frequency   |
| $x_n$             | Sample of the signal  |
| $x_s$             | Observation of parameterized syllable   |
| $x_t$             | Value of the modal Fundamental frequency  |
| $X$               | Samples of Fourier transform of the signal  |
| $X_s$             | Distribution of observed parameterized syllables                                      |
| $y$               | Signal reconstructed from phase   |
| $y_{cc}$          | Normalized cross-correlation  |
| $y_n$             | Sample of the signal reconstructed from the phase                                     |
| $Y_k$             | Reference label of the speaker  |
| $\hat{Y}_k$       | Predicted label of the speaker  |
| $z_t$             | Measurement in Kalman filtering   |
| $Z$               | Z-score   |
| $Z_0$             | One-tailed z-score corresponding to the level of significance                         |
| $Z_i$             | Z-score of the hypothesis   |
| $Z_k$             | Z-score of the hypothesis for the speaker   |
| $\Delta t$        | Interval between consecutive syllables  |
| $\varepsilon$     | Residuals of the regression model   |
| $\theta$          | Phase of the Fourier transform  |
| $\mu$             | Mean  |
| $\mu_x$           | Mean of the signal  |
| $\pi$             | Archimedes' constant  |
| $\sigma$          | Standard deviation  |
| $\sigma_x$        | Standard deviation of the signal  |
| $\sigma_{f_0}^2$  | Variance of the initial model of modal fundamental frequency                          |
| $\Phi$            | Cumulative distribution function  |
| $\chi^2$          | Chi-square distribution   |

# 1

## INTRODUCTION

*She would be all right for a while and treat us kids as good as any mother, and all at once it would start in—something bad and awful—something would come over her, and it came by slow degrees. Her face would twitch and her lips would snarl and her teeth would show. Spit would run out of her mouth and she would start out in a low grumbling voice and gradually get to talking as loud as her throat could stand it; and her arms would draw up at her sides, then behind her back, and swing in all kinds of curves...and she would double over into a terrible-looking hunch—and turn into another person.*

—Woody Guthrie, *Bound for Glory*, 1943

Speech represents one of the most complex human activities, as it involves cognitive-linguistic processes, motor speech planning, programming, control, and neuromuscular execution. The disordered nervous system may manifest in predictable and clinically recognizable speech changes. Studying patterns of speech changes with regard to the underlying neuropathology is beneficial for an understanding of the anatomical and functional organization of speech production, differential diagnosis and localization of a neurological disease, management of a speech disability, and tracking responses to therapy.

Speech analysis has been limited to subjective auditory perceptual assessment or laborious manual analysis of recordings for many generations. With the current astounding availability of data acquisition tools and computational power, digital signal processing stands at the forefront of research in speech pathology. This thesis tackles the main problems of the acoustic analysis of speech, which revolve around the applicability of methods on various speech pathologies, interpretability of speech features, and modelling of complex speech patterns. The method herein described represents one of the first and fundamental steps towards the development of a clinical tool for the complex assessment of speech disorders in neurodegenerative diseases.

## 1.1 MOTOR SPEECH DISORDERS

Speech disorders resulting from impaired motor speech planning, programming, control, or neuromuscular execution are called motor speech disorders (Duffy **2013**). Motor speech disorders can be classified into apraxia and dysarthria.

Apraxia of speech is characterized by the impaired capacity to plan or program sensorimotor commands for directing speech movements (Duffy **2013**). Apraxia of speech is caused mostly by non-hemorrhagic stroke and less frequently by trauma, neurosurgery, or tumors with a lesion in the dominant hemisphere. Although apraxia can result from neurodegeneration, the majority of neurodegenerative diseases are rarely or never associated with apraxia of speech (Duffy **2013**).

Dysarthria is an umbrella term for speech disorders resulting from poor control and coordination of the speech motor system. Speech movements in dysarthria are abnormal in the strength, steadiness, range, tone, or accuracy. Dysarthria can be categorized into several types based on common perceptual characteristics, yielding implications for the localization of a lesion. A variety of causes can lead to dysarthria, including a neurodegenerative disease or brain injury with traumatic, metabolic, or toxic origin. Table 1 provides a brief overview of dysarthria categories, their lesions, distinguishing speech characteristics, and associative neurodegenerative disorder.

## 1.2 SELECTED DISEASES AND PRECURSORS

### 1.2.1 Parkinson's disease

Idiopathic Parkinson's disease (PD) is characterized by a progressive loss of dopaminergic neurons in the substantia nigra pars compacta. The resulting imbalance of dopamine and acetylcholine disturbs the function of the basal ganglia, which participates in the planning, regulation, and execution of movements. Clinical symptoms, include tremors, rigidity, bradykinesia, and postural instability, manifest when more than 40-60% of the dopaminergic neurons have died (Fearnley and Lees **1991**). Approximately 70-90% of PD patients develop a multidimensional speech impairment called hypokinetic dysarthria (Logemann et al. **1978**, Ho et al. **1998**). Hypokinetic dysarthria manifests typically in the imprecise articulation of consonants and vowels, monoloudness, monopitch, inappropriate silences and rushes of speech, dysrhythmia, reduced vocal loudness, and harsh or breathy vocal quality.

### 1.2.2 Atypical parkinsonian syndromes

Atypical parkinsonian syndromes (APSs) are progressive neurodegenerative disorders that involve various neural systems in addition to the substantia nigra. Their manifestations include parkinsonian symptoms plus characteristic clinical signs; hence, APS is also called Parkinson's plus syndrome. The characteristic representatives of APS are multiple system atrophy (MSA) and progressive supranuclear palsy (PSP). MSA causes degeneration in the substantia nigra, striatum, inferior olivary nucleus, and cerebellum. Common symptoms of MSA include difficulties in coordinating movement and balance, postural or orthostatic hypotension, incontinence, impotence, loss of sweating, dry mouth, and vocal cord paralysis. PSP affects neurons and glial

cells in the basal ganglia, brainstem, cerebral cortex, spinal cord, and dentate nucleus. Patients with PSP suffer from a loss of balance while walking; an inability to aim their eyes properly; stiffness; sleep disturbances; depression and anxiety; loss of interest in pleasurable activities; impulsive behaviors, including laughing or crying for no reason; and problems with speech and swallowing. The pattern of symptoms may vary between individuals, making a diagnosis of APS difficult. APS has a generally reduced response to dopaminergic therapy and a more rapid progression, with early development of early-onset postural instability. Speech in PSP and MSA is affected by mixed dysarthria with various combinations of hypokinetic, spastic, and ataxic components (Kluin et al. **1993, 1996**). Excess pitch, reduced intonation variability, reduced maximum phonation time, reduced speech rate, and substantial prolongation of pauses are evidenced in speech affected by PSP (Skodda et al. **2011**, Sachin et al. **2008**, Saxena et al. **2014**). Kim et al. (**2010**) described speech in MSA as slow and effortful with a strained-strangled vocal quality.

### 1.2.3 Rapid eye movement sleep behavior disorder

Idiopathic rapid eye movement sleep behavior disorder (RBD) is parasomnia characterized by motor behavior in response to dream content due to loss of muscle atonia during REM sleep. In recent years, clinical researchers have developed a consensus on the association of RBD and a high risk of alpha-synucleinopathy, particularly PD or dementia with Lewy bodies, and less frequently with MSA (Schenck et al. **1996**, Iranzo et al. **2006**, Postuma et al. **2009**). Iranzo et al. (**2014**) estimated the risk of developing a neurodegenerative disorder at 33.1% at five years, 75.7% at 10 years, and 90.9% at 14 years after diagnosis of RBD. Subtle markers of neurodegeneration, such as reduced color discrimination and olfactory impairment, can be observed in RBD before clinical symptoms of neurodegeneration emerge (Postuma et al. **2009**). A survey of the speech abnormalities in RBD may yield early speech markers of neurodegeneration.

### 1.2.4 Huntington's disease

Huntington's disease (HD) is a predominantly inherited neurodegenerative disorder with a widespread neural loss of both white and grey matter. The broad impact of HD leads to mobility, cognitive, and psychiatric disorders. Symptoms may vary from person to person and stages of the disease. Patients with HD suffer from involuntary, random, jerky movements called chorea; diminished coordination; difficulty in walking and swallowing; speech disorders; problems with concentration, planning, making decisions, and recall; depression; apathy; irritability; anxiety; and obsessive behavior. Symptoms typically develop in middle age, but the disease may onset in a juvenile form with rapid progression or late with slower progression. Involuntary, unpredictable movements may affect any speech dimension, causing the typical characteristics of hyperkinetic dysarthria represented by intermittent hypernasality and nasal emissions, brief speech arrests, irregular articulatory breakdowns, articulatory imprecision, excess loudness variation, sudden forced respiration, constant or varying strained-harsh voice quality, voice stoppages, and abnormal flows of speech (Duffy **2013**).

### 1.2.5 Multiple sclerosis

Multiple sclerosis (MS) a chronic immune-mediated disease of the central nervous system. The pathogenesis of MS is not well understood. Although immune-mediated inflammation is assumed to be the primary cause of damage in relapsing-remitting multiple sclerosis, neurodegeneration

seems to be major contributor to irreversible neurological disability in progressive multiple sclerosis (Trap and Nave **2008**, Ontaneda et al. **2017**). Various motor, sensory, visual, and autonomic systems can be disturbed and any symptoms and signs of central nervous system issues can be present in MS, including numbness, weakness, vertigo, clumsiness and poor balance, cognitive impairment, emotional lability, paroxysmal symptoms, stiffness, painful spasms, impaired swallowing, speech disorder, diplopia, oscillopsia, painful loss of vision, constipation, and erectile dysfunction (Compston and Coles **2008**). Speech disorder in MS can resemble almost any single dysarthria or a combination of the various types (Duffy **2013**). Therefore, dysarthria in MS is not specified, despite speech disorder in MS manifesting most commonly as mixed dysarthria with spastic and ataxic components.

### 1.2.6 Cerebellar ataxia

Cerebellar ataxia (CA) is a term for ataxia caused by a dysfunctional cerebellum. Stroke; tumor; intoxication; poisoning, typically by ethanol; degeneration; and many other causes may lead to CA. Degeneration of the cerebellum can be idiopathic or hereditary. Multiple types of CA can be categorized based on specific symptoms and genetic markers. Hereditary CAs are classified based on the mode of inheritance (autosomal dominant, autosomal recessive, X-linked, or mitochondrial) and gene. The majority of autosomal dominant CAs are referred to as spinocerebellar ataxias (SCAs), as they also involve afferent pathways. Patients with CA suffer from a lack of voluntary coordination of muscle movements, which is called ataxia. The most common clinical symptom is an uncoordinated gait or gait ataxia (Rossi et al. **2013**). Less frequent symptoms represented by unspecified ataxia are dysarthria dizziness, diplopia, visual impairment, vomiting, chorea-dyskinesia, seizures, limb ataxia, intention or postural tremor, and Parkinsonism may be observed in various types of CA (Rossi et al. **2013**). Inaccurate articulation, excess and equal stress, prolonged phonemes and intervals, harsh voice, alteration in speech rhythm, reduced speech rate, increased duration and variability of speech intervals, and increased vocal instability have been reported in CA (Darley et al. **1969B**, Brendel et al. **2015**, Skodda et al. **2013**, Schalling et al. **2007**, Schalling and Hartelius **2013**). Speech disorder in CA gives the impression of slow and imprecise speech with a “drunken” character (Duffy **2013**). Although various speech abnormalities present in other dysarthrias may be present in CA due to neurological impairment extending beyond the cerebellum, speech symptoms in CA resemble predominantly ataxic dysarthria (Duffy **2013**).

## 1.3 EXAMINATION OF DYSARTHRIA

The clinical assessment of dysarthria is described briefly here in order to explain the purpose of an acoustic analysis in a clinical context. Generally, the examination procedure aims to describe the speech disorder, establish the diagnostic possibilities and final diagnosis, establish implications for localization, make a disease diagnosis, recommend management, and specify the severity of the speech disorder (Duffy **2013**).

First, the examiner characterizes the features of the speech disorder. Non-speech oral function is examined in terms of strength, symmetry, range, tone, steadiness, and accuracy of movements. Size and shape of articulators are also observed. The face, jaw, tongue, velopharynx, and larynx, plus respiration, reflexes, and volitional vs. automatic / overlearned responses of non-speech muscles are all subject to analysis. Subsequently, the examiner instructs the patient to

perform various speech tasks and describes the speech disorder by using defined auditory-perceptual characteristics. The most widely used system for auditory-perceptual characterization of dysarthria was established by Darley, Aronson, and Brown (1969A, 1969B), hence it is referred to as the DAB system. The DAB uses 38 speech dimensions grouped into pitch, loudness, voice quality, resonance, respiration, prosody, and articulation and rated on a 7-point scale. The DAB

| Dysarthria                    | Lesion (deficit)   | Speech characteristics  | Associative neurodegenerative disorder  |
|-------------------------------|--|---|---|
| Ataxic                        | Cerebellum or its pathways (incoordination)  | Excess and equal stress, irregular articulatory breakdowns, irregular AMRs, distorted vowels, excess loudness variation, prolonged phonemes, telescoping of syllables, <i>slow rate, slow and irregular AMRs</i>  | Cerebellar ataxia, a component of mixed dysarthria in Friedreich's ataxia, multiple system atrophy, and progressive supranuclear palsy.         |
| Flaccid                       | Cranial or spinal nerves or lower motor neuron system (weakness)                             | Hypernasality, breathiness, diplophonia, nasal emission (audible), audible inspiration (stridor), short phrases, rapid deterioration and recovery with rest, speaking on inhalation, <i>pitch breaks, monopitch, monoloudness, reduced loudness</i>   | Typically as a component of mixed dysarthria in amyotrophic lateral sclerosis   |
| Spastic                       | Upper motor neuron (spasticity)  | Harshness, low pitch, slow rate, strained-strangled quality, pitch breaks, slow and irregular AMRs, <i>hypernasality, short phrases, excess and equal stress, monopitch, monoloudness, intermittent breathy/aphonic segments</i>  | Primary lateral sclerosis, a component of mixed dysarthria in multiple sclerosis, progressive supranuclear palsy, amyotrophic lateral sclerosis |
| Hypokinetic                   | Basal ganglia circuit: substantia nigra pars compacta (rigidity, reduced range of movements) | Monopitch, reduced stress, monoloudness, reduced loudness, inappropriate silences, short rushes of speech, variable rate, increased rate in segments, increased overall rate, rapid, "blurred" AMRs, repeated phonemes, palilalia, <i>hypernasality, breathiness, echolalia</i>   | Parkinson's disease, component of dysarthria in multiple system atrophy, progressive supranuclear palsy   |
| Hyperkinetic                  | Basal ganglia circuit: putamen or caudate nucleus (involuntary movements)                    | Irregular AMRS, distorted vowels, excess loudness variation, prolonged intervals, sudden forced inspiration/expiration, voice stoppages/arrests, transient breathiness, voice tremor, myoclonic vowel prolongation, intermittent hypernasality, slow and irregular AMRs, marked deterioration with increased rate, inappropriate vocal noises, coprolalia, intermittent strained voice/arrests, intermittent breathy/aphonic segments, <i>hypernasality, audible inspirations (stridor), short phrases, harshness, low pitch, slow rate, strained-strangled voice quality, irregular articulatory breakdowns, prolonged phonemes, monopitch, inappropriate silences, variable rate, echolalia, inconsistent articulatory errors</i> | Huntington's disease, dystonia musculorum deformans   |
| Unilateral upper motor neuron | Unilateral upper motor neuron system (weakness, incoordination, spasticity)                  | <i>Slow rate, irregular articulatory breakdowns, irregular AMRs, reduced loudness</i>   | N/A   |
| Mixed                         | Combination of the above   | Combination of the above  | Multiple sclerosis, Friedreich's ataxia, progressive supranuclear palsy, multiple system atrophy, amyotrophic lateral sclerosis                 |
| Undetermined                  | N/A  | Ambiguous pattern of speech characteristics   | N/A   |

Table 1: Summary of dysarthria categories.

Speech characteristics were adopted from Duffy (2013). Distinguishing speech characteristics are typed using normal font. Non-distinguishing speech characteristics are emphasized in italics. The association between dysarthria and neurodegenerative disease was generalized and restricted to common clinical findings.

**Abbreviations:** AMRs = alternating motion rates, N/A = not applicable.

has shown that patterns in auditory-perceptual speech dimensions differ depending on the underlying neuropathology and introduced categorisation of dysarthria based on auditory-perceptual speech characteristics. Auditory-perceptual characteristics play a prominent role in the clinical assessment of dysarthria. Perceived intelligibility of speech may serve as an index of the speaker's ability to communicate. In addition to auditory perceptual characteristics, speech pathologists may employ visual imaging, physiologic, or acoustic methods. Visual imaging methods, such as videofluoroscopy, nasoendoscopy, laryngoscopy, videostroboscopy, and videokymography, are the most commonly used instrumentation techniques. Visual imaging can contribute to an evaluation of swallowing and velopharyngeal and laryngeal function. Results of imaging techniques can be interpreted visually in the context of auditory perception. Physiologic methods, such as electromyography, aerodynamic measures, and electroglottography and acoustic measures, provide mainly quantitative data, which may cause some inconveniences in interpretation. Speech pathologists use instrumentation methods rather exceptionally due to the lack of widely accepted standards, methods and their parameters, and normative data (Till **1995**, Duffy **2013**). Moreover, the majority of speech pathologists may not be armed with the complex knowledge required for analysis and interpretation or may not be convinced about possible benefits (Gerratt et al. **1991**). Although acoustic analysis involves the most convenient instrumentation for the automated assessment of speech disorders, its extensive application in clinical practice is hindered by the frequent correlation of acoustic characteristics with age, sexual dimorphism, and language. Interpreting a large set of raw acoustic features can be an unbearable problem for experts in digital signal processing and even more so for speech pathologists. No such application for acoustic speech analysis which respects the educational background of speech pathologists and the complexity of speech patterns has been provided as of the writing of this thesis. In summary, the existing instrumentation techniques only serve to complement the use of auditory-perceptual characteristics.

Diagnostic possibilities are inferred from a comprehensive description of speech. The clinician can establish the most reasonable diagnosis by considering if the problem is neurologic, organic, psychogenic, or even abnormal at all. Lesion loci can be derived from diagnosed dysarthria only when speech characteristics support the association unambiguously. Classification of dysarthria in the context of other neurological symptoms is common at least in the clinical practice of neurologists (Fonville et al. **2008**). Clinicians may also consider the possible incompatibility of the dysarthria category with the neurologic diagnosis. Generally, a diagnosis of dysarthria requires a holistic approach and cannot rely solely on auditory perceptions. For illustration, Zyski and Weisinger (**1987**) asked experienced clinicians as well as graduate students to classify dysarthria from 28 speech recordings representing all of the categories in the DAB. The reported accuracy of 56% in discriminating dysarthria types was not significantly different between experienced clinicians and students (Zyski and Weisinger **1987**). Another study by Fonville et al. (**2008**) focused on neurologists' ability to discriminate dysarthrias demonstrated an even lower accuracy of 35%, with no significant difference between experienced clinicians and students reported. Van der Graaff et al. (**2009**) asked eight neurologists, eight residents, and eight speech therapists to rate speech samples from 18 patients with flaccid, spastic, ataxic, hypokinetic, hyperkinetic, and mixed dysarthria and four healthy controls (HC). Neurologists showed a 40%, residents a 41%, and speech therapists a 37% accuracy in the identification of dysarthria (Van der Graaff et al. **2009**). Listeners' abilities to discriminate dysarthrias were very low (71%), even when the possible diagnoses were restricted to hypokinetic, spastic, and ataxic dysarthria (Auzou et al. **2000**). When the auditory-perception characteristics from which the dysarthria categories were derived are not sufficient for a diagnosis (Zyski and Weisinger **1987**), no other single approach may work alone.

Finally, instrumentation techniques, such as the acoustic analysis presented in this thesis, are meant to extend the diagnostic capabilities of the clinician, not as a substitute for his or her experience and common sense.

## 1.4 ON THE DECOMPOSITION OF SPEECH PROCESSES

Speech is produced by the interaction of various speech subsystems, including timing, articulation, resonance, phonation, and respiration. For illustration, respiratory flow modulated by glottal pulses convolutes with resonances of the vocal tract. The interaction of subsystems makes localization of the breakdown in the production of speech difficult. The trained ear of a speech-language pathologist can identify a broad spectrum of speech characteristics that can be linked to certain speech movements or dimensions. Acoustic analysis of speech aims to do the same thing via the segmentation of a digital speech signal, followed by the computation of interpretable speech features. Segmentation determines the temporal position of a speech event, and speech features describe its quality. In summary, both auditory-perceptual assessment and acoustic speech analysis decompose speech processes into features that describe elementary tendencies of speech movements in an understandable way.

Despite the incredible abilities of humans in processing acoustic and visual information, speech pathologists commonly employ various speech tasks that endeavor to isolate specific speech movements. Indeed, specific aspects of speech can be inspected in more detail by using specific speech tasks because speech tasks can diminish the possible influence of other processes of speech production and cognitive deficits. Speech tasks also allow speech pathologists to observe specific aspects of speech for longer periods of time or through multiple repetitions.

Connected speech highlights the most natural and challenging cooperation between all of the subsystems of speech. A monologue on a given topic or the reading of a standardized text is used frequently for the assessment of connected speech. Basic timing aspects, such as rhythm stability and rhythm acceleration, can be examined using a rhythm task that requires the syllable /Pa/ to be articulate in a steady rhythm. Articulatory performance is commonly evaluated via antagonistic movements, such as the use of the syllables /Pa/ /Ta/ /Ka/ in quick succession, which is called the diadochokinetic test. The quality of articulation can be rated via individual words or sentences. Phonatory characteristics are usually measured via sustained vowels. Several other aspects, such as lexical and prosodic stress, are assessed via phonetically-balanced texts or rhymes. Many other tasks that are beyond the scope of this study can be exploited in the examination of specific aspects of dysarthria. The list of tasks used in this thesis is limited to an examination of connected speech via the performance of a monologue and reading of a text, the rhythm test, the diadochokinetic test, and inspection of sustained vowels, representing a tradeoff between the number of tasks covered and the complexity of the assessment for the analyzed set of dysarthrias.

## 1.5 AUTOMATED ANALYSIS OF DYSARTHRIA

The term “automated analysis of dysarthria” denotes a methodology for the enumeration of interpretable speech symptoms and/or speech patterns which does not require manual intervention. Theoretically, any instrumentation method could provide a foundation for an automated analysis, including acoustic measures, physiologic measures, and visual imaging.

Nevertheless, acoustic measurement is the primary method used for automation due to the following reasons: All of the subsystems of speech can be captured by one non-invasive and cost-effective measurement of acoustic waves. Unlike physiological measures and visual imaging that monitor the process of speech production via, for example, mechanical, biomechanical, or neural activity; acoustic data describe the final product of speech movements that matter most for speech therapy. Finally, differential speech patterns could be hypothetically detected by acoustic measures, since they are defined by auditory perceptual features. It should also be noted that acoustic analysis involves time series analyses and a very complicated analysis of speech patterns, both of which point to automation because manual analysis is typically laborious or may be principally unbearable.

Despite the long history of the acoustic analysis of speech, which began in 1902 with *The Elements of Experimental Phonetics* by Scripture and took on a new dimension with the technological achievements of the '90s, the acoustic analysis of dysarthria is still subject to research and clinical applications of acoustic analysis are very limited. The most vital developments in acoustic analysis are recent, having been facilitated by easy access to data collection technologies, increased computational power, and increased interest on the part of engineers in speech analysis. Current state-of-the-art acoustic analysis of dysarthria represents a multidisciplinary approach that bridges the disciplines of digital signal processing, machine learning, speech pathology, and neurology. Although acoustic methods are increasingly popular among researchers, the gap between the disciplines has prevented the implementation of results in clinical practice. Clinicians demand knowledge-driven models with universal application, but engineers offer mostly data-driven models that have rarely been validated for more than a single category of dysarthria. Analytical methods are usually specific not only to speech task, but also to dysarthria category. For these reasons, the state-of-the-art acoustic methods documented here focus only on the speech tasks and categories of dysarthria surveyed in the previous sections.

### 1.5.1 Acoustic analysis

#### CONNECTED SPEECH

##### *Segmentation*

Although the segmentation of connected speech has been subject of study by signal processing engineers, the assessment of disordered speech requires more precise segmentation than state-of-the-art voice activity detectors currently provide. The segmentation of connected speech in dysarthria is difficult due to the increased perturbation of voiced intervals and pauses, non-speech sounds, decreased energy in unvoiced speech, loud respirations, and imprecise articulation. The only method in use for the segmentation of connected speech is limited to the detection of pauses and speech intervals (Rosen et al. **2010**).

##### *Speech features*

Connected speech represents a natural task for the examination of prosody. Not surprisingly, intonation variability as well as rate and pause characteristics are commonly measured using connected speech. Regarding the complexity of connected speech, automated prosodic measures are limited by the lack of technologies available for sophisticated segmentation and subsequent qualitative analysis, such as detection of pitch or spectral analysis. Evaluation of pitch still relies on pitch detectors developed for healthy speech or ensembles of detectors (Tsanas et al. **2014**, Berisha et al. **2017**). The assessments of speech rate (Martens et al. **2015**, Jiao et al. **2015**) and the

basic temporal characteristics of short, standardized sentences (Bandini et al. 2015) have been automated.

In addition to prosody, any other speech dimension can be examined in connected speech. Assessment of articulatory features is limited currently to the automated assessment of vowel space area for connected speech proposed by Sandoval et al. (2013). Unfortunately, the method has been evaluated only on healthy speakers. Cross-linguistic metrics based on manual segmentation were analyzed in studies by Liss (2009) and Lowit (2014), but both studies discovered no difference between the HC and dysarthria groups, possibly due to the small sample of patients.

Analysis of phonation is rarely performed because, currently, phonatory features are represented in the quantitative analysis of such items as perturbation characteristics, which can be measured on sustained vowels more conveniently. Likewise, analysis of resonance is measured on sustained vowels since the analysis of resonance is technically challenging in situations in which articulation varies.

Although respiration is placed first in the hierarchy of items requiring clinical attention and treatment (Dworkin 1991), no comprehensive automated acoustic analysis of respiratory patterns in the connected speech has been published. Moreover, respiratory features are analyzed in connected speech, which is one of the tasks used in therapy for and the tracking of the respiratory subsystem (Dworkin 1991). Respiration in connected speech can hardly be analyzed by common measurements, such as a spirometer, because the physical measurement of respiration impedes the ability to speak. However, a microphone located close to the patient's mouth can capture respiration patterns very well without any additional discomfort to the patient.

In summary, the complex assessment of connected speech with regard to prosody, phonation, articulation, and respiration represents a fundamental source of information, as it represents the most natural speech task that can be possibly captured in the form of daily conversations by something as simple as a background app on a smartphone. Regrettably, no methodology for the complex assessment of connected speech had yet been made available at the time of this writing (Hlavnička et al. 2017A).

## **RHYTHM TEST**

### ***Segmentation***

Identification of individual syllables can be very tricky despite the simplicity of the task. Non-speech noises, such as incomplete occlusion, tongue clicks, and excessive inspirations, can occur frequently in dysarthric speech. Additionally, voicing may continue between syllables, and syllables themselves may vary in loudness as well as spectrum. Segmentation of the rhythm task has been based solely on laborious, manual hand labeling up to the release date of the algorithm developed by the author of this thesis (Rusz et al. 2015A).

### ***Speech features***

Pace, as represented by the location of detected syllables, can be evaluated in terms of rate, acceleration, and instability. The clinical relevance of the pace rate calculated as the number of syllables per second is limited when considering the facts that the speaker is allowed to choose his own pace and no dysarthria influences specifically a preference for a fast or slow rhythm (Duffy 2013). Although pace acceleration can be observed exclusively in hypokinetic dysarthria and may influence the measured pace rate, increased self-pacing cannot be inferred. Acceleration of pace (*PA*) can be measured as the difference between the average intervals between the syllables of the

sequences 5-12 ( $avIntDur_{5-12}$ ) and 13-20 ( $avIntDur_{13-20}$ ), normalized by the reference determined as the average intervals between syllables of the sequence 1-4 ( $avIntDur_{1-4}$ ) (Skodda et al. **2010**):

$$PA = 100 \cdot \frac{avIntDur_{5-12} - avIntDur_{13-20}}{avIntDur_{1-4}} \quad \text{Equation 1}$$

Skodda et al. (**2010**) associated values of  $PA$  higher than 1 with acceleration of speech. The authors proposed to measure pace stability using the coefficient of variation ( $COV_{5-20}$ ) based on a similar principle:

$$COV_{5-20} = 100 \cdot \frac{sdIntDur_{5-20}}{\sqrt{16} \cdot avIntDur_{1-4}}, \quad \text{Equation 2}$$

where  $sdIntDur_{5-20}$  denotes the standard deviation of intervals between syllables of the sequence 5-20 (Skodda et al. **2010**). Evidently, the resulting values of  $COV_{5-20}$  and  $PA$  are inversely proportional to the speaker's performance of the first four syllables, quantified as  $avIntDur_{1-4}$ . Moreover, the evaluation requires the speaker to perform a sequence of at least 20 syllables, which may prove to be an overwhelming task for speakers with severe dysarthria. Speech feature assessments for rhythm acceleration and instability need to be redefined to increase reliability and applicability.

## DIADOCHOKINETIC TEST

### *Segmentation*

Assessment of articulatory movements rises and falls on precise detection of burst, voice onset, and occlusion of each articulated syllable since speech features are calculated from the location of detected articulatory events. Novotný et al. (**2014**) developed an automated assessment of the diadochokinetic test incorporating robust segmentation and a set of speech features. The algorithm showed superior accuracy on the datasets for HC and PD. Novotný et al. (**2015**) continued the research and improved the detection accuracy for HD subjects. Unfortunately, further evaluation of the recordings of speakers with CA and APS uncovered limitations of the algorithm in terms of the detection of poorly articulated bursts, silent syllables, and imperfectly separated syllables. More robust detection of voiced intervals and improved detection of bursts is required for the applicability of these speech measurements for any clinical population.

### *Speech features*

The essential measures of diadochokinesis have a very long history. For illustration, the concept of voice onset time, i.e., the time interval between burst and voice onset, came about in the late 19<sup>th</sup> century and was fully defined in the '60s for the categorization of phonemes (Lisker and Abramson **1964**). In addition to voice onset time, the syllabic rate, regularity, vowel duration, and other locational measures used to describe the performance of the diadochokinetic task can be calculated easily using descriptive statistics, such as the mean and standard deviation, when the positions of burst, voice onset, and occlusion are known. The precision of these measures is tied strongly to the precision of segmentation but can be improved by application of more robust estimators, such as the median. Unfortunately, the majority of authors prefer easily defined formulas that can be greatly influenced by outliers resulting from misdetections.

## SUSTAINED VOWELS

### *Segmentation*

The decision as to whether or not the analyzed interval is voiced precedes any analysis of the phonation subsystem, especially in the case of sustained vowels. A variety of voicing determination algorithms have been developed throughout history, but the majority of them fall into one of three main categories indicated by Hess (1983): threshold analyzing, pattern recognition, or voicing determination algorithms combined with pitch determination. Threshold analyzers operate simply by testing a level of the parameter that describes voicing. Common parameters used in voicing determination are energy, the coefficient of the normalized autocorrelation function, the number of zero-crossings, the error in linear prediction, and so forth. The decision as to an interval being voiced can be inferred not only from individual independent parameters but also from their combination using unsupervised or supervised learning. The decision can also be made by testing the periodicity of the estimated pitch, which prevents measurement errors due to the failure of voice detection. Additionally, decisions of multiple detectors can be combined to increase the reliability of decision further.

The idea to decide voicing based on the periodicity of the pitch can be traced back to the '60s, when pioneers of digital signal processing recognized the potential of digital vocoders for voice transmission (Noll 1967, Sondhi 1968). The purpose of a pitch detector is to identify whether the pitch is measurable or not in the analyzed interval. As the quality of the estimate of the pitch depends strongly on the ability to distinguish voiced and unvoiced segments, thresholds for the voiced decision can be expected to be set high enough for precise measurement of the pitch. Accordingly, aberrant vibrations that represent a challenging condition for detection of pitch may be considered as unvoiced, although they represent clinically important intervals of phonation. An algorithm that would detect intervals of increased perturbation and subharmonics as voiced is crucial for subsequent qualitative analysis. Unfortunately, the accuracy of voiced detection in pathological speech is marginalized, and the majority of speech analyzers do not provide a detailed description of the procedures or enumerated accuracy in such cases. No transparent methodology for the voiced/unvoiced decision to be made on the sustained vowels of dysarthric speakers was yet available at the time of writing.

### *Speech features*

Numerous technologies have been proposed for the acoustic analysis of sustained vowels. Two major categories of measurements can be analyzed for this task. The first category consists of measurements related to the function of the larynx, including the characteristics of fundamental frequency, such as variability or range; perturbation measurements; and measurements of the glottal pulse shape. Phonatory measurements are very popular. As an illustration, methods focused only on the assessment of voice quality account for more than 500 studies (Buder 2000). Jitter, shimmer, and the harmonics-to-noise ratio are the most common perturbation measurements with straightforward interpretations. However, these measurements require precise detection of the fundamental frequency, which is not an easy task in conditions of severe perturbation, an abnormal variation in melody, sudden shifts in pitch, and vocal arrests caused by dysarthria. Additionally, the majority of technologies measuring the harmonics-to-noise ratio are based on an autocorrelation function that can be influenced strongly by increased jitter and shimmer. Available technologies, such as CSpeech (Paul Milenkovic, Madison, Wisconsin, USA), the Computerized Speech Laboratory (Kay Elemetrics, Pine Brook, New York, USA), MDVP (Kay Elemetrics, Pine Brook, New York, USA), Dr. Speech (TigerDRS, Seattle, Washington, USA), TF32 (Paul Milenkovic, Madison, Wisconsin, USA), and PRAAT (Boersma, P. & Weenink 2018), were not designed

explicitly for dysarthria or evaluated on a large corpus of dysarthric speakers. Moreover, values measured by different methodologies are not comparable even if measured on identical data (Bielamowicz et al. **1996**). In addition to standard perturbation measurements, the randomness of the signal can be quantified in terms of fractal dimensionality (Baken **1990**, Little et al. **2007**). Fractal-based measurements are interesting research instruments that do not require detection of the fundamental frequency, but the lack of interpretability impedes their broader application. Generally, all well-established phonatory measurements require detection of the fundamental frequency. One exception is represented by cepstral peak prominence, which is measured as a maximal peak in cepstrum against the trend line and thus, by definition, does not require estimation of the fundamental frequency. However, detection of maximal peaks in cepstrum is related strongly to the detection of the fundamental frequency (Noll **1967**, Hillenbrand and Houde **1996**), which raises the question as to whether the original definition is too loose.

Although methods related to the detection of the fundamental frequency are referred to in more than 3000 publications (Benesty et al. **2007**), no ultimate solution suitable for any situation has been discovered yet. Detectors are usually designed to perform well under a specific set of circumstance, such as degraded signal quality, increased environmental noise, and the presence of multiple speakers. Detectors for dysphonia, including an ensemble of detectors (Tsanas et al. **2014**), were never validated with regard to abnormal vibration regimes. Moreover, the available detectors and terminology for a description of dysphonia are related to the fundamental frequency ( $F_0$ ) regarding perceived pitch. Perception of  $F_0$  can be tricky and subjective when vocal folds vibrate with an alternating period, amplitude or both. These aberrant vibrations are called subharmonic vibrations. Although subharmonics represent unique phenomena, clinicians can categorize this phenomenon into one of three different categories, namely diplophonia, harsh voice, or a sudden shift in pitch called pitch break (Weismer **2006**). Subharmonics manifest in the spectrum as local extremes at an integer fraction of the  $F_0$ , most frequently  $F_0/2$ . When alternation exceeds a subjective threshold, subharmonics start to mask  $F_0$ , causing a change of pitch perceived as a jump to an octave below (Bergan and Titze **2001**). Although speech processing engineers are aware of the difference between true fundamental frequency and perceived fundamental frequency, no answer to the problem that would respect both speech pathology and signal processing has been provided yet. Given the above information, a term (modal  $F_0$ ) that describes the  $F_0$  corresponding to the modal register of the voice will be introduced in this thesis. Modal  $F_0$  is meant to represent the frequency at which vocal folds would vibrate without alternation. In other words, modal  $F_0$  reflects vibrations respective to the set-up of laryngeal muscles apart from factors that cause subharmonics. Unfortunately, no method for the tracking of modal  $F_0$  and the detection of subharmonic intervals in dysarthria has been published to date.

The second category of measurements concern the velopharynx and are related to resonance characteristics, such as the measurement of the degree and variability of hypernasality. Automated measurements of hypernasality were designed by Novotný et al. (**2016**).

## 1.5.2 Modeling of speech patterns

Acoustic analysis returns a set of raw values that enumerate speech manifestations. What the values of acoustic features measured on one particular speaker mean cannot be inferred from the values themselves. Analogous to any other measurement, a model is required to elucidate the trend in and severity of a speech feature. When a value is comparable to a statistical model for healthy speakers, then the value can be considered to be normal, i.e., no malfunctions or abnormalities

indicating underlying pathophysiology are present. Models of single speech features can be beneficial for describing individual speech abnormalities, whereas models combining multiple speech features can predict overall tendencies, categorization, or a defined rating scale.

Modeling of a single speech feature is generally marginalized, and authors focus mostly on multivariate models for the categorization of a disorder. Although various studies have shown how incredibly accurate various technologies can be (Little et al. **2009**, Tsanas et al. **2012**, Hariharan et al. **2014**, Orozco-Arroyave et al. **2016**, Vaiciukynas **2017**), the majority of recently developed technologies seem to have limited impact on clinical practice. The reasons for this failure lie not only in limited databases but also in the following technical issues.

First, the majority of authors have proposed algorithms that provide binary classifications or, as the authors frequently state, the “diagnosis” or “detection” of neurodegenerative disease vs. healthy controls. Such algorithms are predominately based on features selected by brute force from an enormous set of descriptors with no hypothetic relation to pathophysiology. An extreme example can be found in the work of Vaiciukynas (**2017**), which analyses 99 speakers using subsets of 47,229 descriptors in total. Such a procrustean solution can hardly prove beneficial to any speech pathologist interested in reliable evidence for his or her own responsible diagnosis.

Second, the majority of authors seek complex models, such as a support vector machine (SVM) with radial basis function or deep neural networks, in order to reach the highest possible accuracies of classification, which raises serious doubts as to whether the classifier aims to describes interpretable principles behind speech patterns instead of just constellations of data. For illustration, see the first subfigure on the left top of Figure 6 in the publication by Little et al. (**2009**). This classification boundary is placed around just five healthy subjects with considerably increased values of a feature called detrended fluctuation analysis (DFA), although increased DFA is associated with voice disorder. Generally, the possibility of fitting a model that contradicts hypotheses can be expected anytime training is executed by minimizing the error function. Some classifiers, such as decision trees, allow one to inspect their consistencies with hypotheses in a decision structure, but the majority of classifiers require thorough evaluation via simulated data.

Furthermore, there is a rising trend in regulating machine learning with the “right to an explanation” (Edwards and Veale **2018**). A nice example can be found in the General Data Protection Regulation act of the European Union, which requires any automated decision-making in the European Union to provide “meaningful information about the logic involved” (Parliament and Council of the European Union **2016**). The implication of the law on machine learning is still the subject of ongoing debate, and some authors question its legal status (Goodman and Flaxman **2016**, Wachter et al. **2017**). Nevertheless, stricter legislation can be expected in the near future after machine learning penetrates society on a deeper level. Regardless of current legislation, clinicians may require and demand a more explanatory approach that can be juxtaposed with other outcomes of their examinations, such as perceptual findings and the patient’s history and socioeconomic status.

Binary classification of speech patterns can yield valuable information about the complex interactions between speech features. However, it should be questioned whether any binary decision about speech patterns could benefit a clinician when acoustic analysis is only one of the many other descriptors which could be considered for the formulation of a diagnosis (see section 1.3 EXAMINATION OF DYSARTHRIA, page 4). A binary decision is more likely to induce bias into a diagnosis than yield new insight into a speech disorder. Estimation of speech pattern severity seems to be more desirable than a diagnostic shortcut.

The efforts of the community of speech processing engineers have been confined mostly to mapping speech patterns to intelligibility (Pathological speech sub-challenge, Interspeech **2012**) or clinical scales, such as the Unified Parkinson's Disease Rating Scale motor score (UPDRS III; Parkinson's condition sub-challenge, Interspeech **2015**) and Frenchay Dysarthria Assessment scales (Orozco-Arroyave et al. **2018**). Despite the considerable attractiveness of these targets for increased objectivity or remote monitoring, assessment of individual speech patterns with regard to the lesion or defined category is more relevant clinically, as was demonstrated by DAB. In summary, simple and explanatory models that respect sexual dimorphism and age dependency are in demand. Currently, no such methodology for modeling the severity of speech deficits described by individual speech features and combinations of speech features is available.

## 1.6 AIMS AND OBJECTIVES

This thesis is centered on the automation of the analysis of dysarthria using acoustic signals. Automated segmentation and the calculation of descriptive speech features are the paramount aims of the thesis. This thesis is framed as a coherent application that addresses general issues of clinical practice, namely, interpretability of values measured by acoustic analysis and modeling of speech patterns. The following objectives summarize the scope of the thesis (see Figure 1):

- **Limitations:** The thesis specifies the recording process and discusses limitations of the proposed methodology with regard to a recording device, the recording process, and applicability of the proposed methodology.
- **Acoustic analysis:** Development of novel methods for automated analysis of connected speech, sustained vowels, rhythm, and the diadochokinetic task regarding segmentation and acoustic features represents the essential goal of the thesis since state-of-the-art methods do not currently provide a clinically applicable solution with acceptable accuracy of detection.
- **Modeling of speech patterns:** Modeling of individual acoustic features regarding sexual dimorphism and age as well as a new approach for pattern recognition which was designed to answer the fundamental limitations of acoustic analysis in clinical settings.
- **Statistical comparison:** Comprehensive analysis of individual acoustic features across selected groups of diseases, covering patients from the subclinical (RBD) to clinical (PD, HD, CA, MS) stages of diseases.
- **Classification experiment:** A novel methodology for pattern classification was compared with selected state-of-the-art classifiers. Subsequently, incidences of proposed speech patterns were estimated.
- **Visualization:** The comprehensive report was designed to convey the results of the analysis in an intelligible way by exploiting all properties of the proposed modeling of speech patterns. The whole methodology, including acoustic analysis and modeling of speech patterns, was implemented into a software application that automatically generates the report. Although the software was not the key goal of the thesis and no code is provided by the thesis, it was crucial for the evaluation of clinical applicability.
- **Experimental use:** The methodology was cultivated in cooperation with an experienced speech pathologist who tested intensively the software implementation of the proposed methodology in clinical practice.

- **Survey:** Feedback was gathered from the clinician in order to evaluate the proposed methodology in terms of customer satisfaction, clinical relevance, interpretability of provided results, benefits, and limitations.
- **Case study:** The application of the proposed methodology was demonstrated using two illustrative case studies. Examinations utilizing the proposed methodology were accompanied by detailed anamnesis and commentary.

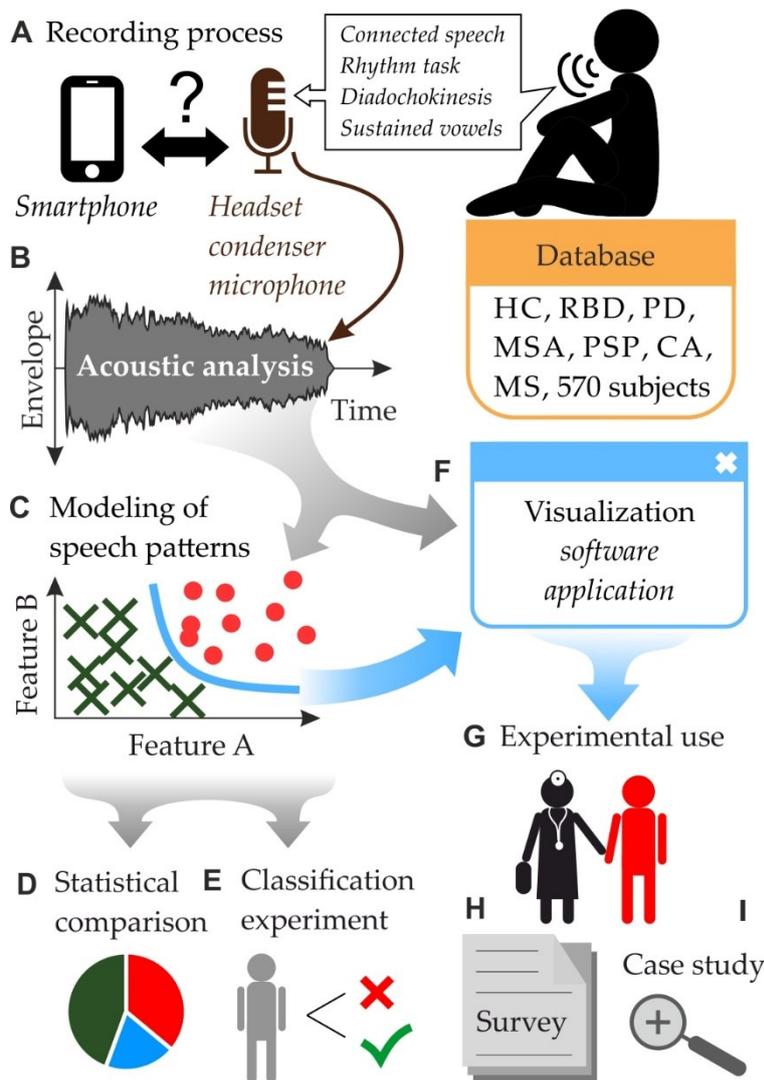


Figure 1: Illustrated objectives of the thesis.

(A) Definition of the recording process and addressing limitations of the proposed methodology. (B) Development of automated methods for the assessment of selected speech tasks. (C) Development of a method for analysis and decomposition of speech patterns. (D) Statistical evaluation of resulting speech features obtained from a large database of subjects. (E) Comparison of various classification methods and proposed modeling of speech patterns. (F) Development of a software application that executes the analysis and generates an interpretable report. (G) Optimization of the methodology through experimental use. (H) Evaluation of the application by a clinician. (I) Demonstration of the final application on selected speakers with dysarthria annotated with the patient’s history and interpretation of the analysis.

**Abbreviations:** HC = healthy control, PD = Parkinson’s disease, MSA = multiple system atrophy, PSP = progressive supranuclear palsy, HD = Huntington’s disease, RBD = rapid eye movement sleep behavior disorder, MS = multiple sclerosis, CA = spinocerebellar ataxia.



# 2

## METHOD

*By the time Bob met him, Guthrie's voice was slurred by Huntington's chorea. The sound of his breath preceded his words. Instead of clearly singing, for example 'I'm ramblin' around,' he would buff, 'bb-I'm ramblin' a-bb-round.' Marjorie Guthrie, who later founded the Huntington's Disease Society and became an authority on the illness, believed Bob, as well as the other young musicians who visited, mistakenly copied these vocal eccentricities as the authentic Guthrie voice. Her daughter Nora says that 'She was convinced that these young guys were picking up these early Huntington's symptoms ... holding a note and then kind of trailing off, which was really a lack of control. That became the style and the jumping off point for Dylan.'*

*—Howard Sounes, Down the Highway: The Life of Bob Dylan, 2001*

The methodology for acoustical assessment is entirely described in this chapter. The clinical characteristics of a large database of speakers as well as a detailed definition of the recording process, including technical details and instructions, are presented here in a comprehensive fashion as a product of the fruitful cooperation between the Department of Neurology and Centrum of Clinical Neuroscience, First Faculty of Medicine of Charles University and Faculty of Electrical Engineering of Czech Technical University in Prague. The recording standard used was developed by Jan Ruzs and Tereza Tykalová. The database was recorded over the course of many years by Jan Ruzs, Tereza Tykalová, Michal Novotný, Hana Ružičková, and other collaborators, as well as marginally by the author of this thesis. All recorded signals were processed digitally via a methodology designed by the author of the thesis, with one exception: the hypernasality measures originated in the research by Michal Novotný et al. (2016). Each methodology is described in terms of the segmentation and acoustic features accompanied by the underlying pathophysiology. A novel approach for the treatment of age dependency and sexual dimorphism is then applied to the measured values of the acoustic features. Central tendencies of speech were categorized from a new perspective, which was deduced from the functional pathways of a speech motor control circuit and classified using a novel approach to information fusion. The classification experiment and statistical methods used are also specified in this chapter. Finally, a special note on the interpretation and application of acoustic analysis is given here to frame the proposed method in a wider clinical context.

## 2.1 DATABASE

The majority of the subjects were recruited originally for previous studies by the Signal Analysis, Modeling, and Interpretation group of the Faculty of Electrical Engineering, Czech Technical University in Prague. The database was extended by speakers that were not included in previous studies or were recorded after a project terminated. Previous studies explored the database in the context of a disease or a syndrome, but a comparison of all speakers was never previously published due to the incomparable age of onset for various diseases.

A sample of 570 Czech native speakers was comprised of 42 subjects with idiopathic RBD (37 males, 5 females), 32 subjects with early untreated PD (PDU; 22 males, 10 females), 26 subjects with treated PD (PDT; 13 males, 13 females), 22 subjects with MSA (10 males, 12 females), 15 subjects with PSP (9 males, 6 females), 18 subjects with untreated HD (HDU; 6 males, 12 females), 13 subjects with treated HD (HDT; 8 males, 5 females), 17 subjects with CA (10 males, 7 females), and 101 subjects with MS (24 males, 77 females). Additionally, 284 subjects (141 males, 143 females) with no history of a neurological or communication disorder were included as HC. Clinical characteristics are summarized in Table 2.

All RBD patients were diagnosed by polysomnography according to the International Classification of Sleep Disorders diagnostic criteria (American Academy of Sleep Medicine **2014**). Diagnosis of PD followed the UK Parkinson's Disease Society Bank Criteria (Hughes et al. **1992**).

| Group<br>(dominant type of dysarthria) | Age (years)<br>Mean / SD<br>(range) | Disease duration (years)<br>Mean / SD<br>(range) | Disease severity #<br>Mean / SD<br>(range) | Speech severity $\chi$<br>Mean / SD<br>(range) |
|--|-------------------------------------|--|--|--|
| HC<br>(none)                           | 54.5 / 17.7<br>(18-89)              | -<br>-   | -<br>-                                     | -<br>-   |
| RBD<br>(none)                          | 66.0 / 8.9<br>(40-83)               | 5.2 / 3.9<br>(1-16)                              | 5.1 / 3.4<br>(0-13)                        | 0.0 / 0.2<br>(0-1)                             |
| PDU<br>(hypokinetic)                   | 65.7 / 8.9<br>(42-79)               | 1.5 / 1.0<br>(0.5-5)                             | 23.6 / 14.0<br>(6-56)                      | 0.5 / 0.5<br>(0-1)                             |
| PDT<br>(hypokinetic)                   | 65.4 / 9.1<br>(48-82)               | 7.8 / 3.8<br>(1-15)                              | 17.8 / 9.2<br>(4-38)                       | 0.8 / 0.7<br>(0-2)                             |
| MSA<br>(ataxic-hypokinetic)            | 61.2 / 6.5<br>(45-71)               | 3.9 / 1.4<br>(2-7)                               | 75.2 / 23.9<br>(35-123)                    | 3.3 / 1.2<br>(1-6)                             |
| PSP<br>(hypokinetic-spastic)           | 66.9 / 6.8<br>(54-84)               | 3.7 / 1.5<br>(2-7)                               | 71.5 / 27.3<br>(19-116)                    | 4.0 / 1.4<br>(2-6)                             |
| HDU<br>(hyperkinetic)                  | 46.3 / 13.8<br>(23-67)              | 5.2 / 3.1<br>(1-13)                              | 19.8 / 11.0<br>(3-51)                      | 0.6 / 0.5<br>(0-1)                             |
| HDT<br>(hyperkinetic)                  | 50.2 / 14.2<br>(30-69)              | 7.0 / 3.4<br>(2-12)                              | 35.0 / 10.6<br>(12-54)                     | 0.9 / 0.5<br>(0-2)                             |
| CA<br>(ataxic)                         | 56.6 / 12.5<br>(34-75)              | 10.0 / 6.6<br>(2-21)                             | 13.4 / 4.3<br>(4-22)                       | 1.9 / 1.3<br>(0-3)                             |
| MS<br>(ataxic-spastic)                 | 43.9 / 11.2<br>(19-74)              | 14.1 / 7.7<br>(2-33)                             | 3.7 / 1.5<br>(1-6.5)                       | 0.3 / 0.6<br>(0-3)                             |

Table 2: Clinical characteristics of all groups in the database.

# Scores on the Unified Parkinson's Disease Rating Scale III (UPDRS III) for RBD, PDU, and PDT (ranging from 0 to 108), Natural history and neuroprotection on Parkinson (NNIPPS) for APS (ranging from 0 to 332), Unified Huntington's Disease Rating Scale (UHDRS) motor sub-score (ranging from 0 to 124) for HDU and HDT, Scale for the Assessment and Rating of Ataxia (SARA) for CA (ranging from 0 to 40), and Expanded Disability Status Scale (EDSS) for MS (ranging from 0 to 10). Higher scores indicate more severe disabilities.

$\chi$  Scores on the UPDRS III item 18 for PD, NNIPPS Bulbar-pseudobulbar signs subscale item 3 for APS, UHDRS dysarthria item for HD, and scores examined by speech specialist for CA and MS.

All scores represent speech motor examination and range from 0 to 4, where 0 represents normal speech, 1 mildly affected speech, 2 moderately impaired speech (still intelligible), 3 markedly impaired speech (difficult to understand), and 4 unintelligible speech.

**Abbreviations:** SD = standard deviation, HC = healthy control, RBD = rapid eye movement sleep behavior disorder, PDU = untreated Parkinson's disease, PDT = treated Parkinson's disease, MSA = multiple system atrophy, PSP = progressive supranuclear palsy, HDU = untreated Huntington's disease, HDT = treated Huntington's disease, CA = cerebellar ataxia, MS = multiple sclerosis.

MSA was diagnosed by the consensus diagnostic criteria for MSA (Gilman et al. **2008**). Cerebellar subtype was identified in just two patients with MSA. PSP was diagnosed using the National Institute of Neurological Disorders and Stroke and the Society for PSP clinical diagnosis criteria (Litvan et al. **1996**). Only two patients manifested PSP-parkinsonism, the rest were diagnosed with PSP-Richardson syndrome. Diagnosis of HD was confirmed by genetic testing (Huntington Study Group **1996**). All CA patients were diagnosed based on molecular testing or clinical findings. Genetic testing identified SCA in 7 subjects. Other CA subjects were diagnosed with idiopathic late-onset cerebellar ataxia based on neuropsychological testing and magnetic resonance imaging. Neuropsychological testing included electromyography, electronystagmography, and genetic analyses of the various SCA mutations (SCA 1, 2, 3, 6, 7, 14, and 17) and the Friedreich's ataxia gene. MS patients were diagnosed with the revised McDonald Criteria (Polman et al. **2011**). Eighty-two patients were diagnosed with relapsing-remitting MS, 4 patients with the clinically isolated syndrome, 7 with secondary progressive MS, and 8 primary progressive MS. Only MS patients in at least a 30-day relapse-free period were accepted for entry into the database.

Patients with RBD or PDU had no history of therapy with antiparkinsonian medication. Patients with PDT had been medicated for at least 4 weeks with levodopa and a different dopamine agonist and were investigated in the ON state<sup>1</sup>. Patients with APS received various doses of levodopa alone or combined with a different dopamine agonist and/or amantadine. Patients with HDU had no history of antipsychotic medication. Patients with HDT were treated with antipsychotic medication alone or combined with antidepressants. None of the patients reported a history of neurological or communication disorders unrelated to their clinical diagnosis or underwent speech therapy while participating in the study.

Disease duration was estimated from the self-reported occurrence of the first motor symptoms. Motor function in RBD and PD patients was scored using the UPDRS III (Stebbins and Goetz **1998**), APS by the Natural History and Neuroprotection in Parkinson Plus Syndromes–Parkinson Plus Scale (NNIPPS; Payean et al. **2011**), HD by the Unified Huntington's Disease Rating Scale (UHDRS; Huntington Study Group **1996**), SCA by the Scale for the Assessment and Rating of Ataxia (SARA; Schmitz-Hübisch et al. **2006**), and MS by the Expanded Disability Status Scale (EDSS; Kurtzke **1983**).

The diagnosis for and scoring of motor function was done by a well-trained professional neurologist with experience in movement disorders. The perceptual severity of speech in RBD, PD, APS, and HD was determined by the speech item on the corresponding clinical scale. The severity of the speech problems in CA and MS was rated perceptually by the speech-language pathologist on a coarse scale ranging from none, mild, moderate to severe, according to Yorkston (**1995**).

All subjects provided informed consent. All studies associated with the database were approved by the Ethics Committee of the General University Hospital in Prague, Czech Republic.

## 2.2 RECORDING PROCESS

Acoustic signals were recorded in a quiet room with low ambient noise using a headset condenser microphone with linear frequency characteristics (Beyerdynamic Opus 55, Heilbronn, Germany). The microphone was placed approximately 5 cm from the mouth. The signal of the microphone

---

<sup>1</sup> The ON state refers to a period when medication suppress the symptoms of PD effectively.

was sampled at 48 kHz with a 16-bit resolution and stored in waveform audio file format via recorder (9Edirol R-09HR, Roland, Shizuoka, Japan). Each participant was recorded in a single session. Each speaker was instructed to perform the following speech tasks:

- **Rhythm:** repeat the syllable /Pa/ at least 20 times at a comfortable, self-determined, and steady pace without acceleration or deceleration.
- **Diadochokinetic task:** repeat the syllables /Pa/ /Ta/ /Ka/ in one breath. Repetition should be performed at least seven times as rapidly, steadily, and accurately as possible.
- **Sustained phonation of vowel /A/:** perform the vowel /A/ for as long and steadily as possible per one breath using the modal register.
- **Sustained phonation of vowel /I/:** perform the vowel /I/ for as long and steadily as possible per one breath using the modal register.
- **Reading passage:** read the standardized text of 80 words.
- **Monologue:** speak about his or her speech interests, job, family, or current activities for approximately 90 seconds in duration.

All tasks except the monologue were performed twice. The values of speech features were averaged across all repetitions of the task in order to reduce the error of measurement. Each task was thus described by a single value for each individual speech feature.

## 2.3 ACOUSTIC ANALYSIS

### 2.3.1 Sustained vowels

#### SEGMENTATION

The signal was decimated to 8 kHz and analyzed in a sliding window 75 milliseconds in length and a 5 millisecond step with hamming weighting. Each position of the window was described by the power of the signal (PWR), maximal peak in the autocorrelation function (MPAF), and the zero-crossings rate of the autocorrelation function (ZCR; see section 2.3.3 CONNECTED SPEECH, SEQUENTIAL SEPARATION, page 32). All values of PWR were normalized by the maximal PWR of the signal and expressed using a logarithmic scale. The first 10% of the signal was not included in the calculation of the maximal PWR to prevent bias due to highly individual PWRs at the onset of phonation. The autocorrelation function was corrected by the autocorrelation function of the hamming window (Boersma 1993) and normalized. Only positive lags of the autocorrelation function corresponding to a frequency range from 50 Hz to 500 Hz were processed further. The MPAF was determined as the maximal value of the autocorrelation function. The ZCR was calculated as the zero-crossing rate of the autocorrelation function and expressed as a frequency.

The signal was labeled as voiced for every position of the window for which the PWR was higher than -50 dB, the MPAF was higher than 0.24, or the value of the ZCR was within a range from 50 to 500. The heuristic was derived from the typical dynamic range described by the PWR, the minimal harmonics-to-noise ratio of -10 dB, and the  $F_0$  ranging from 50 to 500 Hz. Logical disjunction was preferred in the decision process to compensate for the variable quality of pathological voices. All voiced intervals with a PWR lower than -80dB were rejected. Voiced intervals shorter than 100 milliseconds were rejected, as well, because the transient action of vocal folds is not suitable for further analysis. The algorithm described above was developed by the author of this thesis (Hlavnička et al. 2019).

## ANALYSIS OF THE MODAL AND SUBHARMONIC VIBRATIONS OF VOCAL FOLDS

The vibrations of the vocal folds were examined in voiced intervals of speech using the tracking of modal  $F_0$  and the analysis of the harmonic series for modal  $F_0$  and subharmonics devised by Hlavnička et al. (2019). The process of analysis is illustrated in Figure 2.

### *Statistical modelling of modal $F_0$*

A statistical model of modal  $F_0$  was introduced in order to provide support for the tracking of modal  $F_0$  when both modal  $F_0$  and its fractions are present during subharmonic vibrations. The model assumes a normal distribution for modal  $F_0$  as described by its mean and standard deviation. The model was updated for varying fundamental frequencies using a Kalman filter.

The model of modal  $F_0$  needed to be initiated for prior measurement and the proper setting of parameters that define the behaviour of the Kalman filter. The initial model was estimated by the following process. The signal was resampled to 3 kHz and analysed in a sliding window 75 milliseconds in length and a 7.5 milliseconds step with Gaussian weighting. Real cepstrum was performed using a fast Fourier transformation (FFT) with 4096 samples. Only cepstrum corresponding to the frequency range from 50 to 500 Hz was analyzed. Each position of the analysing window was described via the value and location of the maximal peak in cepstrum. Only peaks with a value higher than the 90th percentile of all values were accepted, as they represent phonation with high quality and thus can be associated with modal voice. The location of the selected peaks was recalculated to frequency. The mean and standard deviation of the initial  $F_0$  model were determined by the median and median absolute deviation, respectively, of the frequencies of selected peaks. The median absolute deviation was rescaled to be a quantile of the standard deviation. The standard deviation of the  $F_0$  model was limited to be always higher than 10 Hz due to numerical problems with the representation of small probabilities of outlying values.

### *Detection of modal $F_0$*

The signal was resampled to 3 kHz and processed in windows corresponding to 10 periods of the initial model of modal  $F_0$ . The step of the sliding window was set to be 10% of window length. Gaussian weighting with a length of six standard deviations was applied to increase the resolution of the Gaussian interpolation in further harmonic analysis (Gasior and Gonzales 2004). Longer analysing windows do not deteriorate the temporal resolution and should be preferred due to the rapid tapering of the Gaussian window function. The windowed signal was zero-padded to the length of 4096 samples, and the FFT was calculated for each position of the analysing window. The single-sided amplitude spectrum was normalized to a unity sum. Local extremes were localised within the spectrum. The minimal distance between local minimis or local maximis was conditioned to 25 Hz. Less extreme values violating the condition were discarded. The location and amplitudes of local extremes were refined via Gaussian interpolation (Gasior and Gonzales 2004). Local maximis with prominence higher than 5 dB were selected as candidates for modal  $F_0$ .

Occasionally, a candidate for modal  $F_0$  can be too noisy to be accepted or may be completely missing. Therefore, additional candidates for modal  $F_0$  were reconstructed from a weighted frequency histogram of seven harmonics (Schroeder 1968). Weights of the harmonics were set accordingly to the inverse of the harmonic number, e.g., the weight of the fifth harmonic was 1/5. The following conditions were established to prevent inflation of the fractions of even-order harmonics. All entries with these conditions were rejected. The set of candidates for modal  $F_0$  was extended by a new candidate only when the total weight of a candidate was higher than 0.75, when at least 30% of the harmonics constituting an additional candidate for modal  $F_0$  were

odd, and when the candidate was located more than half an octave from other, already accepted, candidates for modal  $F_0$ .

Only candidates ranging from 50 to 500 Hz were accepted for further analysis. The harmonic series up to the 7<sup>th</sup> harmonic was calculated for each candidate. Local extremes were matched to harmonic series with a tolerance of less than one semitone. Extremes out of tolerance were not accepted. The amplitude of matched local minimis was negated, emphasizing the harmonic structure constituted by the maxims of spectral peaks and minimis in between. Each candidate was described by the probability calculated as the mean amplitude of the extremes matched to its harmonic series. The probabilities of candidates were then compared with the probability model for modal  $F_0$ . Modal  $F_0$  was selected from the candidates as the one with maximal likelihood.

### ***Analysis of subharmonics***

Subharmonics were analysed in each position of the analysing window used for the detection of modal  $F_0$ . The harmonic series  $F_0/2$  was computed from the modal  $F_0$  detected in the position of the analysing window. The harmonic series was then described using the subharmonic-to-harmonic ratio (SHR), calculated as a ratio of even multiples of  $F_0/2$  and odd multiples of  $F_0/2$  (Sun and Xu 2002) via the following equation:

$$\text{SHR} = \frac{\sum_{i=1}^N A(F_0 \cdot i - F_0/2)}{\sum_{i=1}^N A(F_0 \cdot i)}, \quad \text{Equation 3}$$

where  $A$  is the amplitude of a given frequency obtained by the Fourier transform,  $F_0$  is the detected modal fundamental frequency,  $i$  is index of harmonic, and  $N$  represents the maximal number of harmonics. The proposed method analysed series up to seven harmonics. Each amplitude of a given frequency was estimated as the amplitude of the nearest local extrema in the amplitude spectrum refined using Gaussian interpolation (Gasior and Gonzales 2004). Only matches in tolerance of less than one semitone were accepted. The amplitudes of the rejected matches were set to zero, and the sign of the local minim was set to negative. This heuristic compensates for the possible influence of perturbation. Subharmonics were identified when the SHR exceeded the critical value of 0.1, determined by the perceptual experiments of Bergan and Titze (2001).

### ***Adaptation of a statistical model for modal $F_0$***

The variation in modal  $F_0$  was modeled as a linear system of the first order corrupted by stationary Gaussian noise. We assume that  $x_t$ , representing modal  $F_0$ , relates to the previous state  $x_{t-1}$  according to the following equation:

$$x_t = F \cdot x_{t-1} + v_t, \quad \text{Equation 4}$$

where  $F$  represents the state transition model, and  $v_t$  is the normally distributed process noise with zero mean and covariance  $Q$ . The state  $x_t$  can be observed as  $z_t$  according to:

$$z_t = H \cdot x_t + w_t, \quad \text{Equation 5}$$

where  $H$  provides a mapping of the true state into the observed space, and  $w_t$  is the normally distributed observation noise with zero mean and covariance  $R_t$ .

Modal  $F_0$  was predicted by a constant velocity model described by position (i.e.,  $F_0$ ) and velocity (i.e., melodic change):

$$x = \begin{bmatrix} f_0 \\ \dot{f}_0 \end{bmatrix}, \quad \text{Equation 6}$$

where  $f_0$  refers to modal  $F_0$  and  $\dot{f}_0$  its derivation. The state transition matrix represents the transition between consecutive positions of the analyzing window:

$$F = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}, \quad \text{Equation 7}$$

where  $T$  is the period between consecutive positions of the sliding window (i.e., step of the sliding window). The control matrix  $H$  passes only the modal  $F_0$ , which is the only value measured:

$$H = \begin{bmatrix} 1 & 0 \end{bmatrix}. \quad \text{Equation 8}$$

The state of  $f_0$  was treated as noise-free. The estimation of melody was to be imperfect with no relation to  $f_0$ . The process noise covariance  $Q$  can then be defined as:

$$Q = \begin{bmatrix} 0 & 0 \\ 0 & \sigma_{f_0}^2 \end{bmatrix}, \quad \text{Equation 9}$$

where  $\sigma_{f_0}^2$  is the variance of the initial modal  $F_0$  model. The scope of our adapted model was controlled by the covariance of measurement noise,  $R$ . The initial value of  $R_{t=0}$  was determined by the uncertainty  $\sigma_{f_0}^2$  of the initial modal  $F_0$  model:

$$R_{t=0} = \sigma_{f_0}^2. \quad \text{Equation 10}$$

The initial error covariance matrix  $P_{t=0}$  relies on the error of initial measurement determined by the variance of the initial modal  $F_0$  model:

$$P_{t=0} = \begin{bmatrix} \sigma_{f_0}^2 & 0 \\ 0 & 0 \end{bmatrix}. \quad \text{Equation 11}$$

The predicted state  $\hat{x}_{t|t-1}$  and predicted error covariance  $P_{t|t-1}$  were calculated via the following equations:

$$\hat{x}_{t|t-1} = F \cdot \hat{x}_{t-1|t-1}, \quad \text{Equation 12}$$

$$P_{t|t-1} = F \cdot P_{t-1|t-1} \cdot F^T + Q, \quad \text{Equation 13}$$

where the modal  $F_0$  model was predicted with state  $\hat{x}_{t|t-1}$  representing the mean and the first element of the error covariance  $P_{t|t-1}$  representing variance of the model. Then the value  $z_t$  was measured using the predicted modal  $F_0$  model. Next, state  $\hat{x}_{t|t}$  of modal  $F_0$  and the error covariance  $P_{t|t}$  were updated via the following equations:

$$K_t = P_{t|t-1} \cdot H^T \cdot (H_t \cdot P_{t|t-1} \cdot H^T + R_t)^{-1}, \quad \text{Equation 14}$$

$$P_{t|t} = P_{t|t-1} - K_t \cdot H \cdot P_{t|t-1}, \quad \text{Equation 15}$$

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t(z_t - H \cdot \hat{x}_{t|t-1}), \quad \text{Equation 16}$$

where  $K_t$  represent the Kalman gain. The Kalman filter was initialized for every new phonation and reset to initial settings every time vocalization was interrupted.

### Algorithm outcome

The algorithm provides a measurement of modal  $F_0$  and its distribution, as predicted by the Kalman filter. Extreme values of  $F_0$  can occur only when one glottal pulse is analyzed, typically on the border of the voiced interval. Therefore, all  $F_0$  values that were distributed eight standard deviations from the prediction of the Kalman filter were rejected, and the position was reclassified as unvoiced. The resulting  $F_0$  time course was smoothed by a median filter of the 3<sup>rd</sup>

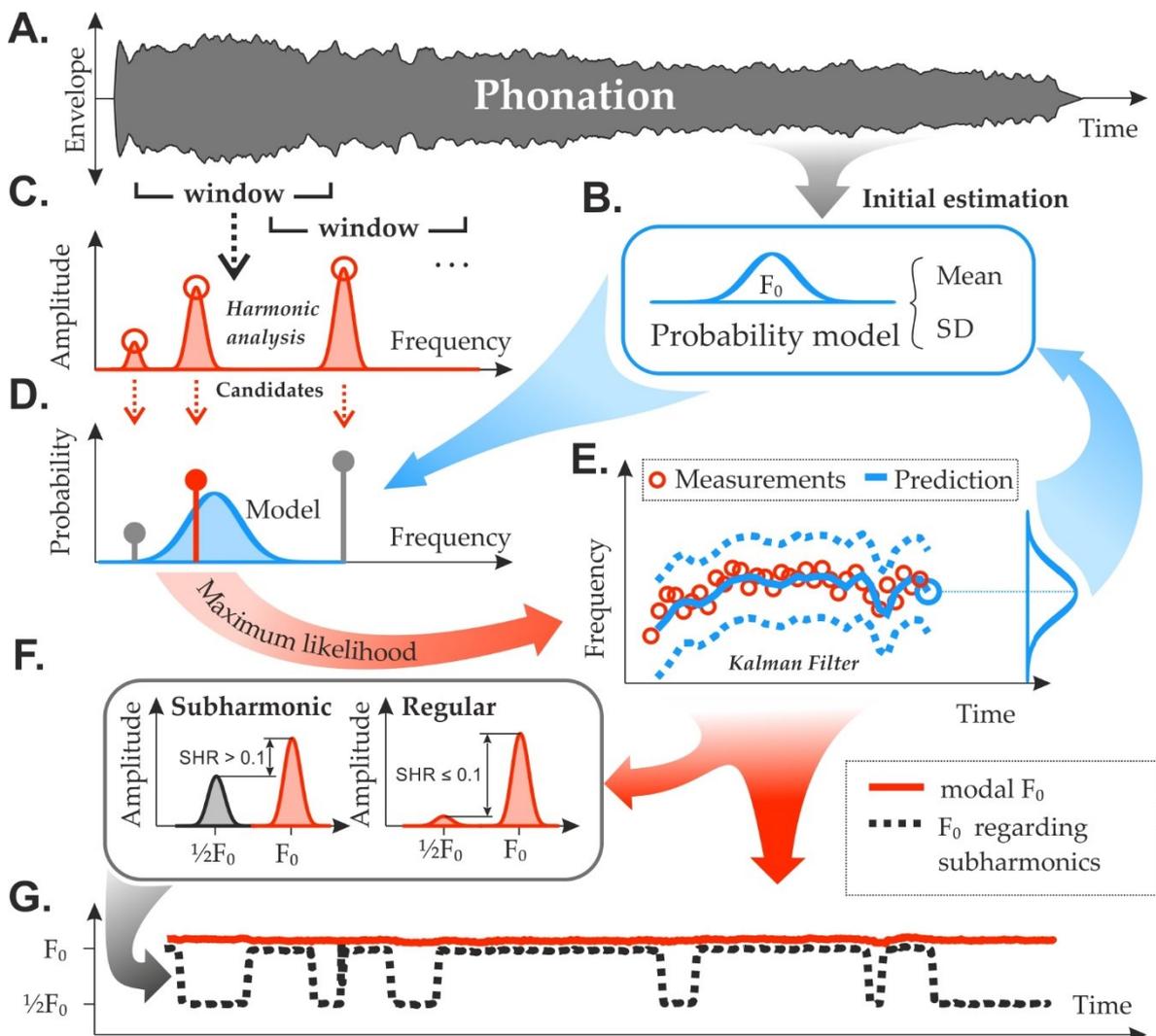


Figure 2: Process diagram illustrating the analysis of modal and subharmonics vibrations.

The loudness envelope represents a sample of the analysed signal (A). Parameters of the statistical model describing modal  $F_0$  were estimated via cepstral analysis (B). The signal was processed inside the sliding window via the harmonic analysis of the spectrum (C). A candidate for modal fundamental frequency was selected accordingly to the probability model of the modal voice (D). The Kalman filter updated the probability model based on a new measurement of modal  $F_0$  (E). Subharmonic intervals were recognised via analysis of the subharmonic-to-harmonic ratio (F). The resulting track of modal fundamental frequency (solid red line) and track recalculated from the presence or absence of subharmonics (dashed black line) are illustrated in the graph (G).

**Abbreviations:**  $F_0$  = fundamental frequency, SHR = subharmonic-to-harmonic ration, SD = standard deviation.

order, which decreased the short-time imprecision while maintaining an outstanding time resolution. The resulting  $F_0$  represents modulation by the laryngeal muscles without the influence of possible subharmonic vibrations, i.e., intended melody.

Subharmonic intervals were identified via the following decision process. Decisions concerning subharmonics were smoothed by a median filter of the 7<sup>th</sup> order. All voiced intervals between subharmonics that were shorter than 300 milliseconds were reclassified as subharmonics when phonation was not interrupted by a pause. All subharmonics shorter than 50 milliseconds were rejected. The positions of subharmonics were described by time labels.

The algorithm measures modal  $F_0$ , the positions of regular and subharmonic intervals, and the time course of the SHR.

## **SPEECH FEATURES**

### ***Degree of vocal arrests (DVA)***

An abnormal contraction of the laryngeal muscles causes intermittent voice stoppage. Increased DVA can be associated especially with spasmodic or choreatic movements of laryngeal muscles (Manfredi et al. 1996, García et al. 2011).

DVA was measured as the proportion of unvoiced intervals detected by segmentation to the total time of performance. Voiced intervals initiated after 90% of total phonation time were removed from the analysis in order to avoid the influence of fatigue or weak respiratory flow. Only unvoiced intervals situated between the onset of phonation and termination of the last voiced interval were included.

### ***Maximum phonation time (MPT)***

The economy of respiration as well as coordination between phonatory and respiratory control affect directly the longest time one can phonate.

MPT was calculated as the total duration of all voiced intervals detected by segmentation.

### ***Standard deviation of $F_0$ (stdF0)***

Involuntary movements of laryngeal muscles can affect the geometrical and mechanical properties of vocal folds directly and thus modulate into melody. Increased variation of modal  $F_0$  can imply involuntary movements of vocal folds since each speaker was instructed to produce a steady tone. Additionally, an unstable modal register that changes gradually into a pulse regime can also be measured as increased variation.

The variation in modulation by the laryngeal muscles was described as the standard deviation of modal  $F_0$  (Hlavnička et al. 2019). The standard deviation was estimated as the median absolute deviation rescaled to a quantile of the standard deviation.

### ***The proportion of subharmonic intervals (PSI)***

Subharmonic intervals can be more dominant in phonation for various reasons. Some healthy individuals may show subharmonics without any underlying pathology. When a person's voice start to become more prone to subharmonics, it may indicate changes in the mass or control of vocal folds. A neurogenic origin of subharmonics is well known. Nevertheless, current terminology regarding subharmonics is very ambiguous and describe subharmonics via three different auditory-perceptual characteristics, namely, diplophonia, harsh voice, or a sudden shift of pitch called pitch break, depending on the perceived depth of alternation and duration. The degree to which

perceived intervals of subharmonics dominate without respect to the depth of alternation was enumerated by PSI.

PSI was calculated as the ratio between the total duration of subharmonic intervals per total duration of voicing (Hlavnička et al. 2019).

### ***Location of subharmonic intervals (LSI)***

Muscle fatigue is one of the possible factors that may influence subharmonic intervals and should be considered, especially when subharmonic intervals occur at the end of phonation.

LSI was calculated as the initial time of the first detected subharmonic interval occurring in the course of phonation (Hlavnička et al. 2019).

### ***Standard deviation of the power spectral density (stdPSD)***

Changing the positioning of articulators causes changes in the resonant characteristics of the speech apparatus that can be captured in the spectrum. Although some speakers may move their articulators during sustained phonation into a more comfortable position or a position demanding less respiratory flow, excess movements of articulators are preeminently involuntary. The position of the articulators determines unique resonance characteristics of the speech apparatus that can be captured in the spectrum. Moving articulators makes the spectrum more variable. Increased values of the feature stdPSD indicate the increased variability of the spectrum and thus the severity of involuntary movements. The method was developed by the author of this thesis and inspired by collaborating with speech-language pathologist Hana Růžičková, who exploited sustained vowels successfully for the perceptual examination of tongue dystonia.

The signal was decimated to 8 kHz, as only frequencies up to 4 kHz were subject to analysis. The signal was analyzed only in voiced intervals using a sliding window 100 milliseconds in length, a 10 millisecond step, and hamming weighting. It was normalized to unity power in each position of the analyzing window. The power spectral density was estimated by the bank of 16 linearly spaced triangular filters in each position of the analyzing window, and the standard deviation of the power was calculated for each frequency band. The value of the stdPSD was calculated as the mean value of the standard deviations describing all 16 frequency bands.

### ***Degree of hypernasality (EFn\_M) and intermittent hypernasality (EFn\_SD)***

The impaired neuromuscular control of the elevator muscle of the soft palate increases the involvement of the nasal cavity in the process of speech production. Acoustically, velopharyngeal insufficiency can be perceived as hypernasality. When the elevator muscle of the soft palate moves involuntarily, the degree of hypernasality varies accordingly and hypernasality is then intermittent. The resonance of the nasal cavity can distort the formant structure significantly, which can be measured directly as the attenuation of a specific frequency band. Hypernasality can be measured on the sustained vowel /I/.

The method used for the assessment of hypernasality was developed originally by Novotný et al. (2016) and is provided here for consistency and to clarify implementation details. Frequencies below 65 Hz were filtered out using a high-pass Chebychev filter of the 4<sup>th</sup> order to prevent possible disruptions from popping or a main hum. The signal was decimated to 8 kHz. Only voiced intervals situated between 10% and 90% of the signal time course were analyzed. This trimming reduces the influence of unstable vocal activity at the beginning of phonation and weak expiratory flow at the end of phonation. The signal was filtered with a band-pass Butterworth filter of the 3<sup>rd</sup> order with a passband from 890.9 Hz to 1122.5 Hz. The energy of the filtered signal was

calculated inside the sliding window of 60 milliseconds in length and a 5 milliseconds step. All measurements of the energy were normalized by the total energy of the signal and expressed via a logarithmical scale. The overall degree of hypernasality was described by the mean energy of the filtered signal (EFn\_mean), calculated by using the median. The variability of hypernasality was estimated as the standard deviation of the energy of the filtered signal (EFn\_SD). Note that the significant effect of hypernasality on frequencies around 1 kHz was discovered by Novotný (2016) using a 1/3-octave analysis. The proposed implementation applies only to one filter of the desired frequency band in order to decrease the computational burden. Suggested refinements, including a different sampling rate, had no measurable influence on the resulting features. The proposed implementation and the original method were perfectly correlated for all available data.

### ***Jitter, shimmer, and Harmonics-to-noise ratio (HNR)***

Jitter, shimmer, and HNR are well-established metrics that measure the perturbation of the vocal fold vibration in terms temporal instability, amplitude instability, and additive noise, respectively. Increased values of perturbation are associated with dysphonia, namely, hoarseness.

Intervals of regular vibrations were preferred for perturbation analysis in order to avoid bias resulting from the alternation of subharmonic vibrations related to different speech conditions. Jitter, shimmer, and HNR were analyzed in a sliding window with a length calculated to be 10 periods of the estimated modal  $F_0$  and a step corresponding to one period of the modal  $F_0$  given by the initial model. A template waveform with length equal to the period of modal  $F_0$  was selected at the beginning of analyzing window (see Figure 3). The normalized cross-correlation between the template waveform and rest of the window was defined by the following equation:

$$y_{cc}[k] = \frac{\sum_{n=1}^M (g[n] - \bar{g}) \cdot (x[n - k + 1] - \bar{x}_k)}{\sqrt{\sum_{n=1}^M (g[n] - \bar{g})^2 \cdot \sum_{n=1}^M (x[n - k + 1] - \bar{x}_k)^2}}, \quad \text{Equation 17}$$

where  $y_{cc}[k]$  is  $k$ -th sample of the normalized cross-correlation between the template  $g$  and signal  $x$ . The bar above  $g$  and  $x$  indicates the corresponding average value. The equation was implemented using Lewis's approach (1995). The template was normalized to zero mean and unity variance ( $\dot{g}$ ):

$$\dot{g} = \frac{g - \bar{g}}{\sigma(g)}, \quad \text{Equation 18}$$

allows the definition of the cross-correlation function to be simplified to the following equation:

$$y_{cc}[k] = \frac{\sum_{n=1}^M \dot{g} \cdot (x[n - k + 1] - \bar{x}_k)}{\sqrt{\sum_{n=1}^M (x[n - k + 1] - \bar{x}_k)^2}}, \quad \text{Equation 19}$$

where the denominator is a normalization factor realized as a sliding variance. The convolution was realized via multiplication in the frequency domain. Local HNR was calculated from maximal peaks of the normalized cross-correlation detected at local maxims with a minimal distance of 80% of the detected modal  $F_0$ . HNR was expressed in logarithmical scale. Local jitter was computed from the distance between adjacent maximal peaks, and local shimmer was determined as the difference in measurements of the normalization factor at times of adjacent maximal peaks of normalized cross-correlations. Each position of the analyzing window was described by the

median values of the perturbation measurements. Overall values of jitter, shimmer, and HNR were estimated as the median of all values measured inside the sliding window. The median was preferred to reduce the influence of erroneous extreme values. The method was conceived by author of this thesis and has not been published yet.

## 2.3.2 Rhythm test

### SEGMENTATION

Segmentation of the rhythm task can be tricky in dysarthria even though only isolated syllables are subject to detection. Syllables can be separated imperfectly, and intervals between syllables are not always regular. Syllables can also be articulated imprecisely with varying loudness. Increased noise, such as respirations, incomplete occlusion, and tongue clicks, can be expected. The segmentation algorithm was designed to overcome these obstacles via sensitive syllable identification, followed by the self-correction of false positives utilized by outlier detection and verification. The method was developed by author of this thesis and published in paper by Rusz et al. (2015A). Figure 4 illustrates the segmentation process.

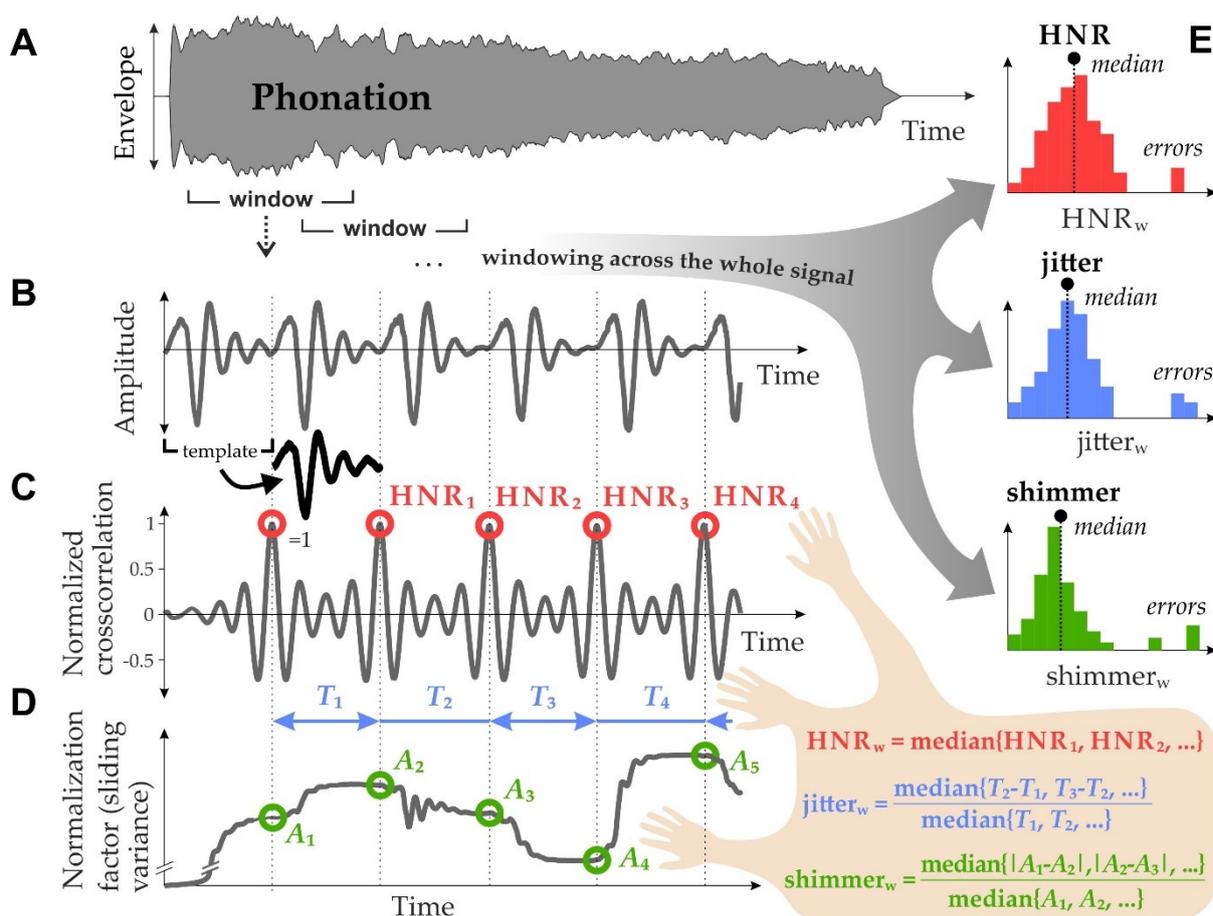


Figure 3: Illustration of perturbation analysis.

A sample of the speech signal is plotted as a loudness envelope (A). The template was selected from the beginning of the sliding window (B) and matched with the rest of the window in terms of the normalized cross-correlation (C). Maxims of cross-correlations served as a measurement of HNR, whereas periods between maxims allowed the jitter to be measured (C). Shimmer was determined from the normalization factor of the cross-correlation function (D). The resulting values of jitter, shimmer, and HNR were calculated as median measurements gathered from all positions of the sliding window (E).

**Abbreviations and symbols:** HNR = harmonics-to-noise ratio,  $T$  = period,  $A$  = amplitude.

### ***Syllable identification***

The sampling rate of the signal was reduced to 10 kHz in order to decrease the computational cost of high-frequency components, which are redundant for the detection of syllable nuclei. The signal was analyzed in a sliding window 10 milliseconds in length with a 3 milliseconds step and hamming weighting. Further, the signal was parameterized by 12 Mel-frequency cepstral coefficients. The spectrum of syllables is skewed more towards low frequencies, whereas the spectrum of pauses is flatter. Therefore, the first three MFCCs representing the low-frequency envelope of the spectrum were selected for recognition of syllable nuclei. The MFCC were analyzed in the recognition window, ensuring fast adaptation to the spectrum variability due to articulatory deficits. A recognition window 4 seconds in length and 800 millisecond steps ensure that every position of the window will contain at least one syllable. The K-means algorithm with two components was preferred for the cluster analysis because maximizing the distance between clusters can be more robust in cases when there are interfering clusters or additional clusters constituted by non-speech sounds. Modeling the data with a Gaussian mixture model may prove problematic because the expectation-maximization (EM) algorithm may converge into local optima in these situations. A cluster of syllables was identified as the component with the higher mean of the first MFCC associated with the power of the signal. The decision was smoothed by a median filter of the fifth order. Only syllables longer than 30 milliseconds and pauses longer than 80 milliseconds were accepted. Figure 4 illustrates the process of syllable identification.

### ***Outlier detection***

Audible inspirations and other non-speech sounds may be detected occasionally as syllables when their spectrum is more similar to the spectrum of syllables than to the spectrum of pauses. Therefore, the identified syllables should be compared to each other and described as one observation by computing the mean of each of the first three MFCCs. The following procedure is intended to remove these false positives by identification of inliers represented by syllables and outliers represented by false positives.

A more general description of the algorithm is provided here in order to illustrate the process of outlier identification comprehensively. Outliers have extreme values and thus can be identified as observations with low probability. The problem can arise when the total number of observations is low and/or the number of outliers is high. Outliers can bias the probability model estimated on all the data and make results unreliable. Therefore, it is preferable to estimate the model on all the data and identify inliers with increased certainty using the very tough criteria of the 30<sup>th</sup> percentile. Inliers can then define a new probability model more reliably. The set of inliers can updated using the new probability model and a less rigorous 50<sup>th</sup> percentile, and the process of model redefinition and the update of inliers can be repeated until no new outliers are found.

The original study (Rusz et al. 2015A) applied the Mahalanobis distance; however, any multivariate probability model or metric can serve just as well. The Mahalanobis distance is defined by the following equation:

$$D_M(x_s) = \sqrt{(x_s - \mu)^T \cdot S^{-1} \cdot (x_s - \mu)}, \quad \text{Equation 20}$$

where  $D_M(x_s)$  is the Mahalanobis distance between observation of syllable  $x_s$  and the distribution of syllables  $X_s$ ,  $S$  is covariance matrix of the distribution  $X_s$ , and  $\mu$  is mean of the distribution  $X_s$ . Percentiles were tested via conditioning  $D_M^2(x_s)$  using values determined from  $\chi^2$  distribution as

$\chi_L^2(q)$ , where the degrees of freedom  $L$  represent the number of dimensions analyzed and  $q$  is the percentile (30<sup>th</sup> or 50<sup>th</sup>) used for the initial estimation of inliers and the further update, respectively.

**Outlier verification**

The algorithm above will identify all extreme observations, which may include not only respirations, but also dissimilarly articulated syllables, typically too loud or silent syllables. Accordingly, the loudness of outliers was verified. Only the frequency band from 100-500 Hz, filtered using a Chebychev’s filter of the fifth order, was analyzed because respirations usually manifest very low energy in this band. The signal was squared and filtered via a moving average

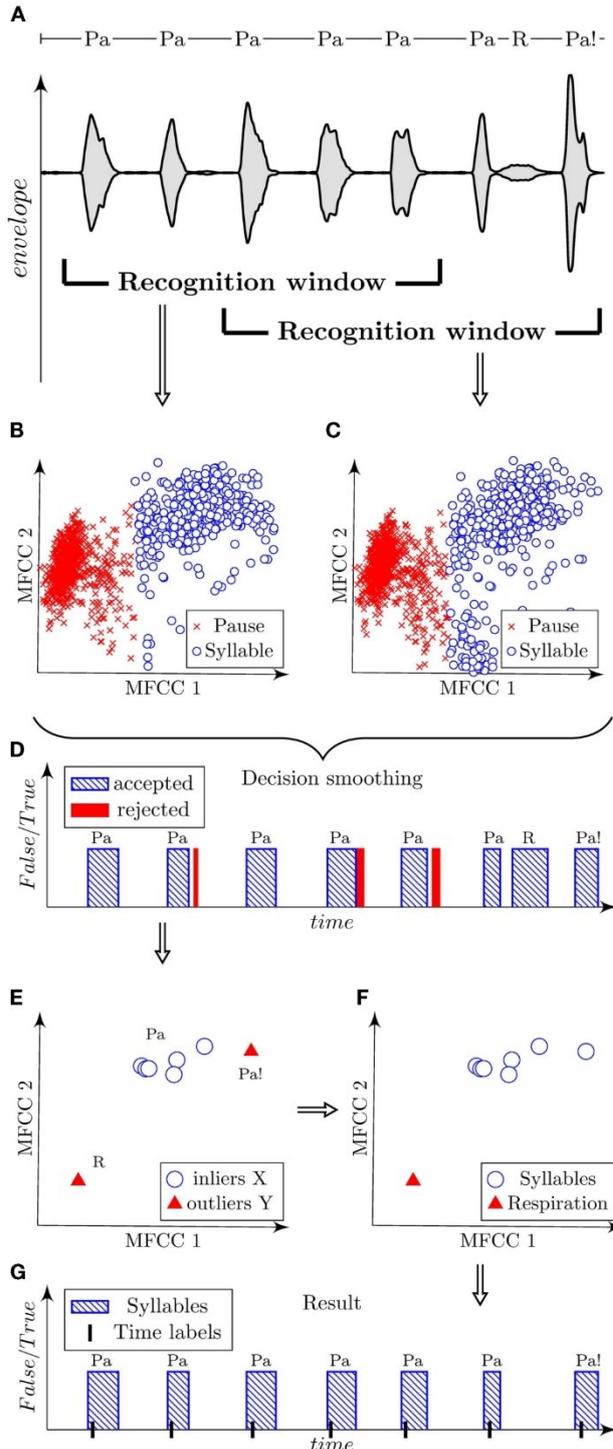


Figure 4: Process diagram of syllable identification. (A) A sample rhythm task containing the syllables ‘Pa’, dissimilar syllables ‘Pa!’, and respirations ‘R’ plotted as the envelope of sound pressure level with marked positions of sliding recognition windows. (B, C) Clusters of detected syllables marked with red ‘x’ marks and clusters of pauses marked with blue circles in parametric space. (D) Time course of smoothed decision with rejected intervals marked as red filled areas. (E) Syllables described as individual observations in parametric space with inliers highlighted as blue circles and outliers highlighted as red triangles. (F) Accepted syllables, including the dissimilar one, marked as blue circles, and respiration verified as the outlier plotted as a red triangle. (G) The resulting recognition of syllables plotted as blue hatched areas and corresponding time labels of energy peaks marked on the time axis.

**Abbreviations:** MFCC = Mel-frequency cepstral coefficients.  
**Copyright notice:** This figure was designed by the author of this thesis and published in the research by Ruzs et al. (2015A) under the terms of the Creative Commons Attribution License (CCBY). All authors of the original paper (Ruzs et al. 2015A) share co-authorship.

with a 10 millisecond length and 3 milliseconds overlap, plus hamming weighting. Each inlier  $P_x$  and outlier  $P_y$  was described by their mean power or maximum power, respectively. Outliers were rejected when their maximum power  $P_y$  was higher than 95% of the inliers' mean powers  $P_x$ . Testing of the 95% population level was performed using Chebychev's inequality, which is defined as follows:

$$P_{Y(i)} > \mu(P_X) - 4\sigma(P_X), \quad \text{Equation 21}$$

where  $P_{Y(i)}$  denotes individual observations of outliers,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

### Time labels

Each syllable was described with a label corresponding to the time of the highest filtered energy peak.

## SPEECH FEATURES

### Rhythm acceleration (RA) and rhythm instability (RI)

Oral festination associated with hypokinetic dysarthria is related to timing disturbances in the basal ganglia affected by parkinsonian neurodegeneration. Increased values of RA indicate pace acceleration. Additionally, the pace can be more irregular as a result of decreased control over the speech apparatus. However, hyperkinetic and ataxic movements are more significant causes of disturbed regularity of pace. An increased value of RI is associated with a less regular rhythm.

RA and RI were calculated using regression analysis. The duration of the intervals between consecutive syllables, hereby referred to as syllable gaps, and its time of occurrence was regressed by a polynomial of first order (see Figure 5). RA was determined as the negative slope of the regression line. Values of RA higher than zero indicate acceleration. The RI was computed as

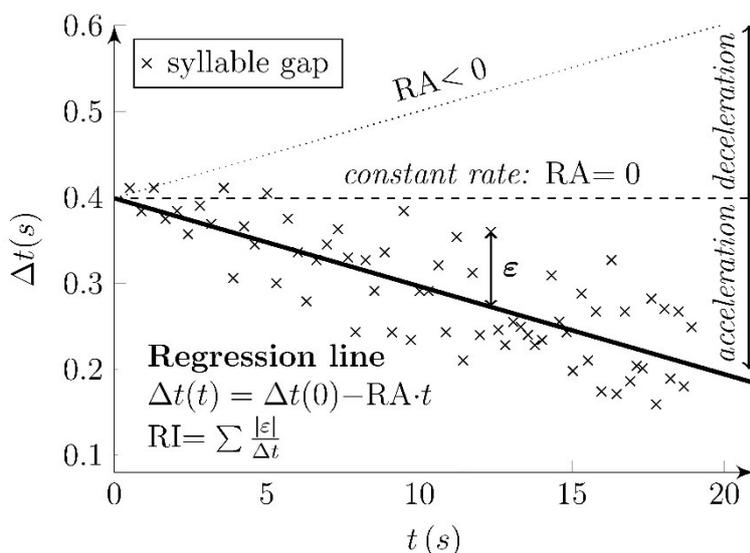


Figure 5: Illustration of designed rhythm features.

Gaps between detected syllables are plotted in the time course as 'x' marks. Note that the slope of the regression line has a negative sign in the defining equation in order to help an examiner with the interpretation of results by associating positive values of RA with acceleration.

**Symbols and abbreviations:**  $\Delta t$  = interval duration between consecutive syllables, RA = rhythm acceleration, RI = rhythm instability,  $\epsilon$  = residuals of the regression model.

**Copyright notice:** This figure was designed by the author of this thesis and published in the research by Rusz et al. (2015A) under the terms of the Creative Commons Attribution License (CCBY). All authors of the original paper (Rusz et al. 2015A) share co-authorship.

the sum of absolute residuals determined by the difference between the observed value of the syllable gap and value predicted by the regression line divided by the total duration of speech.

### 2.3.3 Connected speech

#### SEGMENTATION

The basic activities of the speech apparatus represented by voiced speech, unvoiced speech, respiration, and pause were detected by the following process (see Figure 6) published by Hlavnička et al. (2017A). The signal was preprocessed and parameterized. A cluster analysis was then applied in the sliding recognition window, ensuring adaptation to the changing quality of a speech performance. Voiced speech, unvoiced speech, respiration, and pauses were separated sequentially using various spaces of parameters. Such an approach makes it possible to use a simple Gaussian mixture model (GMM) for the description of otherwise complicated and imperfectly separable mixtures.

#### *Preprocessing*

The signal was decimated to 8 kHz sampling rate. The main hum, popping, and other possible disruptions were filtered out by a 4th order high-pass Chebychev filter with a cut-off frequency at 130 Hz. Additionally, high frequencies were emphasized by an infinite impulse response filter with coefficients [1 0.9] in order to improve recognition of unvoiced intervals.

#### *Parameterization*

Parameters PWR, ACR, and ZCR were computed inside a sliding window of 15 milliseconds in steps of 5 milliseconds using the following equations:

$$\text{PWR} = \frac{1}{N} \sum_{n=1}^N x^2[n] \cdot h[n], \quad \text{Equation 22}$$

$$R_x[k] = \frac{1}{N \cdot \sigma_x^2} \sum_{n=1}^N (x[n] - \mu_x) \cdot (x[n+k] - \mu_x), \quad \text{Equation 23}$$

$$\text{ACR} = \frac{1}{M-1} \sum_{k=1}^M (R_x[k] - \overline{R_x})^2, \quad \text{Equation 24}$$

$$\text{ZCR} = \frac{1}{N-1} \sum_{n=1}^{N-1} |\text{sign}(R_x[n+1]) - \text{sign}(R_x[n])|, \quad \text{Equation 25}$$

$$\text{sign}(R_x[n]) = \begin{cases} 1, & R_x[n] \geq 0 \\ -1, & R_x[n] < 0, \end{cases} \quad \text{Equation 26}$$

where  $x$  is a signal in a window of length  $N$ ,  $h$  represents the hamming window,  $R_x$  denotes the normalized autocorrelation function,  $M$  is length of one-sided autocorrelation function shortened to 75%,  $\sigma_x$  symbolizes the standard deviation of the signal, and  $\mu_x$  describes the mean of the signal. All parameters were described with a logarithmical scale to compensate for their log-normality. One-sided  $R_x$  was shortened to 75% to reduce estimation error.  $R_x$  was preferred for computation of ZCR because all voiced intervals, including vowels and consonants, can be then described by the unimodal normal distribution. ACR was determined as the variance of  $R_x$ . The first five of the 24 linear-frequency cepstral coefficients (LFCC) were used to describe the low frequency envelope of the power spectral density.

#### *Sequential separation*

Intervals of voiced speech, unvoiced speech, respiration, and pauses were recognized in a given order following the process of cluster analysis inside the recognition window. The cluster analysis

assumed the GMM of the parametric space. An optimal number of mixtures in GMM was determined from the corresponding highest value of Calinski-Harabasz index computed over the range  $\langle 2; 3 \rangle$ . The parameters of the GMM were estimated using the EM algorithm. Clustering was performed using a Bayesian discriminant. The resulting decision was smoothed with a set of rules derived from the natural timing of the speech apparatus and the assumption that unvoiced speech accompanies voiced speech in the Indo-European language family.

The voiced speech was determined in a recognition window 20 seconds in length with 6-second steps. A cluster of voiced speech was determined inside the parametric space of PWR, ACR, and ZCR as the one with the highest mean PWR. A median filter of the 5th order smoothed the detection. Voiced segments shorter than 30 milliseconds were reclassified as voiceless while regarding natural limitations to control abduction and adduction of the vocal folds within a shorter period.

Unvoiced speech comprises intervals of unvoiced consonants. The unvoiced speech was classified in voiceless intervals shorter than 300 milliseconds by using a recognition window of 60 seconds in length and 20-second steps. A long recognition window guarantees that the sample size of the less frequent unvoiced consonants will be sufficient. The first five LFCCs were the preferred parameters for cluster analysis because both unvoiced speech and environmental noise represent random signals and can be distinguished well in the spectrum. A cluster of unvoiced speech with the highest mean for the first LFCC was identified in relation to the power of the signal. Only unvoiced speech longer than 5 milliseconds in a distance shorter than 30 milliseconds was accepted, all other unvoiced speech was reclassified as pause intervals.

Respirations were analyzed in the remaining speechless intervals longer than 200 milliseconds. Respirations were determined in the space of the first five LFCCs. The component with the highest mean for the first LFCC was classified as respiration. Only respirations longer than 40 milliseconds were accepted. The distance of respiration to the nearest interval of voiced speech was conditioned with a threshold of 30 milliseconds, which stems from the fact that respirations are bounded with silence, as lungs stop during the reversion of airflow. Respirations bounded with no pause longer than 30 milliseconds were thus reclassified as unvoiced speech. Intervals between respirations shorter than 400 milliseconds were classified as respirations. When less than two respirations are detected, then candidates for respiration longer than 200 milliseconds were reclassified as respirations. This rule was added to correct for situations in which a speaker with a severe speech disorder breaks general assumptions.

Pauses comprise all intervals longer than 30 milliseconds which are not voiced speech or unvoiced speech. Note that respirations are a subset of pauses.

The outcome of the segmentation are labels describing the start time and end time of each detected interval of voiced speech, unvoiced speech, respiration, and pause.

## **SPEECH FEATURES**

### ***Resonant frequency attenuation (RFA)***

The main hypothesis behind RFA is that articulatory decay and mumbling will reduce the prominence of acoustic resonances in the speech signal (Rusz et al. **2015B**). Less prominent resonances are surrounded naturally by shallow valleys. The decreased depth of valleys between formants is related to overall sound propagation and radiation and thus can indicate subliminal articulatory imperfections that cannot be captured using standard measurements of formant frequencies, such as vowel space area.

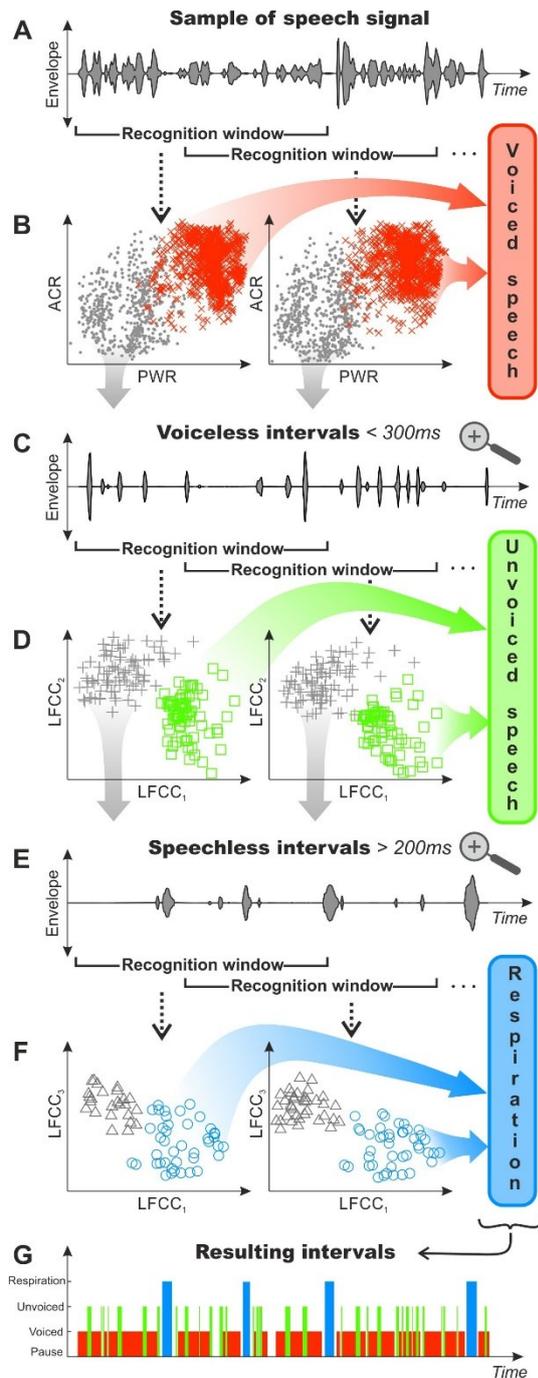


Figure 6: Automated segmentation of connected speech.

(A) Speech signal analyzed inside sliding recognition window. (B) Clusters of voiced speech marked by red 'x' marks were determined in the parametric space of ACR, PWR, and ZCR. (C) Remaining voiceless intervals shorter than 300 milliseconds were classified into unvoiced speech and pauses inside the recognition window. (D) Recognized clusters of unvoiced speech marked by green squares and pause intervals marked as grey '+' marks. (E) Speechless intervals longer than 200 milliseconds analyzed inside recognition window. (F) Clusters of detected intervals of respiration plotted as blue 'o' marks and pauses without respiration marked as grey triangles. (G) Time course of the resulting classification into voiced, unvoiced, respiratory, and pause intervals.

**Abbreviations:** ACR= variance of autocorrelation function, PWR = signal power, LFCC = linear-frequency cepstral coefficients, *ms* = milliseconds.

**Copyright notice:** This figure was designed by the author of this thesis and published in the research by Hlavnička et al. **2017A** under the terms of the Creative Commons Attribution 4.0 International License. All co-authors of the original paper (Hlavnička et al. **2017A**) share co-authorship.

The RFA was analysed in a sliding window of 50 milliseconds in length with 12.5-millisecond steps. Only the voiced speech was analysed. The power spectral density (PSD) was computed for every position of the sliding window using a bank of 24 linearly-spaced, overlapping filters in the band from 200 Hz to 4 kHz. An abnormal vocal source or nasal resonance may affect the prominence of resonances too. Nonetheless, their influence can be compensated by the cepstral liftering of PSD. The PSD was described using a logarithmic scale and transformed into cepstrum (CPSD) using a discrete cosine transformation. The low-frequency components of PSD, such as the power of the signal and gradual attenuation, are contained approximately within the first three coefficients of CPSD. High-frequency components representing sudden changes in the spectrum are contained approximately within coefficients higher than 10 in order. Therefore, only

values of coefficients ranging from 4 to 9 in order were preserved, and other values of CPSD were set to zero. Cepstrally lifted PSD (LPSD) was computed by the inverse discrete cosine transform of CPSD. The RFA was measured as the difference between the first local minima of LPSD and its consecutive maxima. With respect to the logarithmic scale, RFA represents the ratio of the second resonance and its preceding valley. The resulting value of RFA was calculated as the mean value of RFA calculated for all positions of the sliding window. The RFA was invented by the author of this thesis and described in the publication by Rusz et al. (2015B).

### ***The rate of speech timing (RST)***

Reduced range of movement in hypokinetic dysarthria can disturb the timing and coordination of speech subsystems considerably. As a result, the speech rate can be accelerated or slowed, phonemes are imprecisely articulated or omitted, and voicing can interfere in unvoiced speech or pauses. This complex deficit manifests as a reduced stream of voiced, unvoiced and pause intervals. RST was designed by Hlavnička et al. (2017A) to measure the rate of voiced, unvoiced and pause intervals and is influenced by the syllabic rate as well as the overall quality of a speech performance. Although RST can serve as a proxy measurement for syllabic rate, the results of RST should always be interpreted with respect to the above.

The total number of voiced, unvoiced, and pause intervals was accumulated during the time course. The time course was modeled by a regression line, which reduces bias made by extreme values. RST was determined as the gradient of the regression line (Hlavnička et al. 2017A).

### ***Net speech rate (NSR)***

Dysfunctional speech manifests frequently as slow speech, not only due to the slowness of individual movements, but also as compensatory mechanism for increasing the intelligibility of speech. The NSR is standard measurement which has been used for decades by speech pathologists. NSR can be measured only when the number of syllables is a priori known. Thus, NSR was analysed only for the text reading task.

The total number of syllables was divided by the total duration of speech, including only detected voiced and unvoiced intervals (Hlavnička et al. 2017A).

### ***Acceleration of speech timing (AST)***

Acceleration of speech production results from increasing the rate of speech movements or decreasing the range of speech movements. The stream of voiced, unvoiced, and pause intervals can be evaluated by RST; thus, measuring changes in RST could quantify acceleration. Hypothetically, both a reduction in RST and increase in RST over the course of time can indicate acceleration, which extends the assumption of reduced RST in acceleration presented by the original study of Hlavnička et al. (2017A). Additionally, other factors, such as fatigue, can also manifest in reducing the stream of voiced, unvoiced, and pause intervals. Therefore, the feature requires thorough interpretation in the broad context of other symptoms.

The speech run was spitted into two halftimes with a 25% overlap in order to smooth the transition and decrease the influence of the speech content. The AST was determined as the difference between the RST computed in each halftime divided by the total duration of a speech utterance.

### ***Duration of pause intervals (DPI)***

Difficulties in initiating speech have a considerable effect on pause duration. Alternatively, the statistical distribution of pause duration can be biased by omitted short pauses, thereby making

prolongation of pauses even more noticeable. The DPI reflects hypokinesia of the movements involved in initiating speech and pause production.

The DPI was calculated as the median length of pause intervals (Hlavnička et al. 2017A).

### ***Entropy of speech timing (EST)***

Any healthy speaker has the ability to produce a variety of sounds. When the control and coordination of speech become more limited, the speech becomes more ordered and predictable. The arsenal of speech movements can be categorized crudely as voiced speech, unvoiced speech, pause, and respiration. Accordingly, a decreased entropy of observed categories can indicate impaired coordination between subsystems or insufficient control over one or more subsystems of speech. Voiced speech, which represents a fundamental component of speech production, may tend to dominate the speech at the expense of other types of speech intervals.

The EST was computed as Shannon entropy applied on incidences of speech intervals according to the following equation (Hlavnička et al. 2017A):

$$\text{EST} = -\frac{n_v}{n_t} \cdot \log_2 \left( \frac{n_v}{n_t} \right) - \frac{n_u}{n_t} \cdot \log_2 \left( \frac{n_u}{n_t} \right) - \frac{n_p}{n_t} \cdot \log_2 \left( \frac{n_p}{n_t} \right) - \frac{n_r}{n_t} \cdot \log_2 \left( \frac{n_r}{n_t} \right), \quad \text{Equation 27}$$

where  $n_v$  is number of voiced intervals,  $n_u$  is number of unvoiced intervals,  $n_p$  is number of pause intervals,  $n_r$  is number of respiratory intervals, and  $n_t$  means the total number of speech intervals. Each interval was accounted for as one observation. Note that pauses bounding respiration were accounted only once, as respiration is a subset of pauses.

### ***Duration of unvoiced stops (DUS)***

The production of unvoiced stops represents the most rapid task for articulators. Articulation of unvoiced stops is thus a valuable marker of speech control and coordination. Performance of unvoiced stops is commonly measured with voice onset time, which unfortunately does not reflect pure articulatory precision but rather coordination of articulation and phonation. The DUS was designed by Hlavnička et al. (2017A) under the assumption that the explosion of poorly articulated stops is more likely to be accompanied by turbulent noise. Unlike VOT, which identifies the position of the burst, DUS identifies the stop consonant as the interval of the impulse and/or noise. Rapidness and precision of articulatory movements can be then quantified based simply on the duration of the stop consonant. An increased value of DUS reflects increased friction, and extreme values can indicate spirantization of unvoiced stops.

The DUS requires detection of unvoiced stops, which demonstrate a significantly shorter duration. The duration of unvoiced stops and fricatives have a bimodal normal distribution. The parameters of the distribution can be estimated using the EM algorithm and classified with a Bayesian discriminant. The DUS was computed as the median duration of unvoiced stops.

### ***Decay of unvoiced fricatives (DUF)***

Speech pathologists commonly observe the decay of articulatory precision during speech run in hypokinetic dysarthria. A wide variety of movements contributes to articulation, making the measurement of articulatory decay a very difficult task. Frictions represent a specific articulatory movement that can be identified easily and quantified in voiceless fricatives. The level of high-frequency components in unvoiced fricatives (>2.5 kHz) is in direct relation to the level of friction.

The DUF measures the gradual decay of high-frequency bulk and thus quantifies the possible decay of the performance.

The speech run was spitted into two halftimes with a 25% overlap, ensuring a smooth transition and decreasing the influence of the speech content. Fricatives were determined from the durational distribution of unvoiced consonants as described in section 2.3.3 CONNECTED SPEECH in chapter DURATION OF UNVOICED STOPS (DUS), page 36. Every interval of unvoiced fricative was parameterized using 24 MFCCs. The ratio between the low and high Mel-frequency bands was approximated by the second MFCC. The mean value of the second MFCC coefficient was computed from all unvoiced fricatives for both halftimes. The DUF was then computed as the difference between the mean second MFCCs in the two halftimes divided by the total duration of the speech. The DUF was scaled in parts per thousand to increase convenience in reading. The DUF was developed by author of this thesis and described in the publication by Hlavnička et al. (2017A).

### ***Duration of voiced intervals (DVI)***

Decreased control of the laryngeal muscles and coordination of the laryngeal and supra-laryngeal muscles may manifest via voicing that interferes or continues within voiceless intervals, including unvoiced speech or pauses. Voiced intervals are prolonged as a result.

The DVI was computed as the mean duration of voiced intervals detected by the segmentation (Hlavnička et al. 2017A).

### ***Gaping in between voiced intervals (GVI)***

Examination of the pause production in the phenomena described above (see DVI) can provide deep insight into the vocal folds' ability to abduct and adduct. The pauses bounded by voicing represent pure activity to adduct when vocal folds block the airflow to stop voicing or to abduct in the case where vocal folds stop voicing without blocking the airflow. The adduction is more dominant in short pauses, hereby referred to as gaps, whereas abduction can be performed naturally in long pauses between words or sentences, hereby referred to as formal pauses. The rate of the gaps in between voiced intervals reflects the ability of the vocal folds to stop voicing via adduction (Hlavnička et al. 2017A). Decreased gapping may indicate limited control over vocal fold adductors.

The distribution of pauses in between voiced intervals is a bimodal mixture of gaps and formal pauses. The parameters of these mixtures can be estimated using an EM algorithm. Gaps in between voiced intervals can then be identified via Bayesian discriminant analysis as the component with the shorter mean duration. The GVI was computed as the number of gaps per total speech time according to the original publication by Hlavnička et al. (2017A).

### ***Rate of speech respiration (RSR)***

Decreased range or control of respiratory movements, inefficient air-flow management during speech production, or an impaired ventilatory pattern can lead to an increased rate of speech respiration.

The following computational procedure for the RST aims to decrease the influence of falsely detected respiratory intervals. Each respiration event was described by the mean time between the start and end of the respiratory intervals. The time between consecutive respiratory events represents a period between respirations. The RSR was estimated as an inversion of the median respiratory period and expressed in respirations per minute (Hlavnička et al. 2017A).

***Pause intervals per respiration (PIR)***

Speech is structured uniquely by ventilation patterns. Respiratory needs, phonatory and articulatory control as well as grammar and cognition contribute collectively to the resulting breath groups. The collaboration between respiration and the other subsystems can be disturbed particularly by the ability to control the respiratory airflow, which can be captured by a decreased number of pauses per breath group.

The PIR was calculated as the median number of pauses between detected respiratory intervals (Hlavnička et al. **2017A**).

***Relative loudness of respiration (RLR)***

The respiratory airstream produces turbulent noise as it flows through respiratory airways and the oral cavity. The loudness of the inspiratory noise is related to the respiratory force and can be increased by an obstruction in the airways, such as a constricted laryngeal muscle in hyperkinetic dysarthria. Unfortunately, the measurement of loudness requires a reference signal or calibration of the recording system, which is not convenient in practical applications. Using speech loudness as a reference for measuring the loudness of respiration can compensate for the unknown gain of the recording system and be used to evaluate the differences between the expiratory and inspiratory effort represented by speech and detected respirations, respectively.

The signal was decimated to 8 kHz because a band of higher frequency does not contribute to the measured effect. It was squared and filtered with a moving average of 15 milliseconds in length. Loudness was computed by expressing the resulting power envelope via a logarithmic scale. The RLR was calculated as the difference between the median loudness of respiratory intervals and median loudness of voiced speech. Note that there are studies (Hlavnička et al. **2017A**, **2017B**) that have utilized all of the speech intervals, including voiced and unvoiced speech, which has been found to be less sensitive and biased considerably by the loudness of unvoiced consonants.

***Latency in respiratory exchange (LRE)***

Exhalation and inhalation involve groups of respiratory and accessory muscles which must coordinate perfectly during the conversion from exhalation to inhalation. Expiratory movements must stop and inspiratory movements must be initiated properly. The ability to initiate inspiration can be substantially deteriorated, particularly in the later stages of movement disorders. Problematic initiation of respiration manifests in an increased latency between exhalation and inhalation. In general, speech in the Indo-European language family is carried out by exhalation, with the small exception of rare ingressive speech sounds used for feedback words or expression of emotions, e.g., 'Huh!'. Detected respiratory intervals can be definitely assumed to be inspirations. Given the above, the measurement of latency between intervals of detected speech and respiration can indicate the problematic initiation of inspiration (Hlavnička et al. **2017A**).

All respiratory intervals were paired with preceding intervals of speech. Latency was then determined as the difference between the detected start of respiratory intervals and the detected end of the preceding speech interval (Hlavnička et al. **2017A**). The LRE was calculated as the mean of all latencies.

***Standard deviation of power (stdPWR)***

Abnormal variations in loudness can be observed in any dysarthria, with the exception of unilateral motor neuron dysarthria (Duffy **2013**). In addition to a neurological impairment, a psychological disorder or habitual impairment may be the cause of the variations observed. Abnormal variation

of loudness reflects typically poor respiratory-phonatory coordination and control. A variation in excessive loudness is prominent in ataxic and hyperkinetic dysarthria and is related mostly to the momentary hyperadduction of the vocal folds or the effects of dystonia on respiratory support (Freed **2011**, Duffy **2013**). Other dysarthrias may manifest substantial decreased loudness variation, which is also called monoloudness, as a result of weakened laryngeal or respiratory muscles (Duffy **2013**). Monoloudness may be co-currently present in hyperkinetic dysarthria when dystonic contractions wane (Freed **2011**). Variation of loudness is categorized commonly as a prosodic feature.

The signal was decimated to 8 kHz because higher frequencies are redundant for further analysis. The squared signal was filtered by a moving average of 20 millisecond in length and expressed using a logarithmic scale. The feature stdPWR was established as the standard deviation of the resulting loudness envelope computed on all voiced intervals.

### ***Standard deviation of fundamental frequency (stdF0)***

The melody of the voice is modulated by very fine movements of the vocal folds. Inspection of vocal melody can yield accurate insights into patients' abilities to control laryngeal muscles. Variation in melody reflects the ability to contract and/or relax muscles controlling the vocal folds. A limited range of motion in the laryngeal musculature due to weak laryngeal control or tenseness of the laryngeal muscles decreases variation in melody. A perceptual feature of decreased melody variation, referred to as monopitch, is associated with the hypoadduction of the vocal folds. Monopitch can occur in various dysarthrias and is one of the most prominent characteristics of hypokinetic dysarthria. Excess melody variations can be the result of the involuntary movements of the laryngeal muscles in dystonia, a medical condition describing sustained or repetitive muscle contractions. Excessive melody variation is a typical feature of hyperkinetic dysarthria.

Modal vibrations of vocal folds were estimated using an automated algorithm (see section ANALYSIS OF THE MODAL AND SUBHARMONIC VIBRATIONS OF VOCAL FOLDS, page 21). The detected time course of modal  $F_0$  was expressed in semitones in order to compensate for the differences in variability between lower- and higher-pitched voices. The feature stdF0 was implemented as the standard deviation of detected modal  $F_0$  in semitones estimated via the median absolute deviation.

## 2.3.4 Diadochokinetic test

### **SEGMENTATION**

Segmentation of the speech signal in a diadochokinetic test aims to describe the position of the unvoiced stop consonants via the time of burst and voiced intervals via the time of voice onset and time of occlusion. Identification of individual syllables can be a very difficult task in severe dysarthria when syllables are articulated in a diverse fashion and voicing continues between syllables. However, the precise detection of a burst in embarrassing conditions of increased noise between syllables represents the biggest challenge in segmentation. Even precisely detected bursts can be valueless when voice onset is detected inaccurately because bursts are examined only via the durations of the intervals between burst and voice onset while following speech features.

The method for the automated detection of syllables (Rusz et al. **2015A**) was adopted for segmentation of the diadochokinetic test due to the supreme accuracy of syllable detection. The only modification of the procedure undertaken was the use of a shorter recognition window of 0.3 s, which allowed faster adaptation to rapid articulatory movements. The method (Rusz et al.

**2015A)** detects approximate intervals of syllables with a precision limited by the step of the parameterization window, which is acceptable for the evaluation of rhythm, but not for the precise evaluation of rapid diadochokinetic movements. Thus, the refinement of the voice onsets of detected syllables is provided. Additionally, robust detection of the burst is also introduced in order to make the assessment reliable for practical application. These extensions of the original method (Rusz et al. **2015A**) were designed by the author of this thesis and have not been published yet.

### ***Refinement of voice onset detection***

The onset of each syllable detected by the method (Rusz et al. **2015A**) was refined using the following procedure. The signal sampled at 8 kHz was analyzed in the interval from 20 milliseconds prior to syllable onset to occlusion. It was then filtered by the integrator with an integration constant of 0.95. The integration aimed to highlight vocal pulses and suppress high frequencies of burst. Subsequently, the power envelope was computed from the squared signal using a moving average 5 milliseconds in length and Gaussian weighting. The power was scaled in logarithms to compensate for its log normality. The zero-crossings of the integrated signals were described by their respective powers and classified into three clusters using the k-means algorithm. Clusters corresponded to pause, instabilities preceding voice onset, and voiced interval. The cluster with the highest mean rank of power was labeled as a voiced cluster. The decision was smoothed by a median filter of the fifth order. Voice onset was detected as the first voiced zero-crossing over the course of time.

### ***Detection of burst***

Bursts were detected for each voice onset in the interval from 75 milliseconds preceding voice onset to voice onset. When occlusion of the previous syllable interfered into the interval, a shorter frame beginning at the time of occlusion was analyzed. A new, unpublished method for the detection of bursts was developed by author of this thesis in order to increase the reliability of the results for severe dysarthria, for which the established method, based on the analysis of a spectrogram, was not suitable (Novotný et al. **2014, 2015**).

The detection of impulses using a spectrogram is a popular and very intuitive method, but it ignores completely the importance of phase in the localization of an impulse. In the limited case of a signal only when a Dirac impulse is present, the magnitude of the spectrum is flat and the slope of the unwrapped phase spectrum determines the position of the impulse. Every burst has an impulse-like nature which must be emphasized for the proper localization of the burst. Given the above, a magnitude spectrum is obsolete for burst detection and can be set to a constant value, just as in the case of the magnitude spectrum for a perfect impulse. Finally, the position of the burst can be reconstructed purely from the phase via the following equations:

$$X[k] = \sum_{n=0}^{N-1} x_n \cdot e^{-\frac{2\pi jkn}{N}}, \quad \text{Equation 28}$$

$$X[k] = |X[k]| \cdot e^{j\theta}, \quad \text{Equation 29}$$

$$y_n = \frac{1}{N} \sum_{k=0}^{N-1} e^{\frac{2\pi jkn}{N} + \theta}, \quad \text{Equation 30}$$

where  $x_n$  denotes the  $n$ -th sample of signal  $x$ ,  $X[k]$  is the  $k$ -th sample of Fourier transformation of the signal,  $j$  means an imaginary unit,  $\theta$  is the phase, and  $y_n$  is the  $n$ -th sample of the signal reconstructed from the phase. The reconstructed signal  $y$  preserves only non-stationary elements of the signal, i.e., harmonic signals occurring from leakage or during the articulation of voiced stop

consonants suppressed by the reconstruction. It is convenient to discard samples around the beginning and end of the reconstructed signal, as they may reflect the edges of the boxcar function.

Only the absolute values of the reconstructed signal were analysed because the sign of the peak does not add any additional information to a decision. All absolute values were normalized to a unity sum, which allows the interpretation of the value as the probability of a burst position. Each probability value was compared with the expected distribution of voice-onset-times obtained from the manually labeled database. The distribution of the durations between bursts and voice onsets was modeled by a gamma distribution with a shape parameter of 3.4 and inverse scale parameter of 0.015. The burst was labeled at the position with maximal likelihood.

## **SPEECH FEATURES**

### ***Voice onset time (VOT)***

VOT is a well-established metric, which is defined as the duration of the interval between the release of a stop consonant and the initiation of vocal fold vibration. Supralaryngeal muscles releasing stop consonant and laryngeal muscles initiating vocalization must be synchronized within a few dozen milliseconds of VOT. The value of the VOT can be positive, zero, or negative, depending on the position of the consonant. A positive VOT is associated with unvoiced consonants, and a negative VOT is associated with a voiced consonant. The key factor determining the duration of the VOT in voiceless stops is the ability of the laryngeal muscles to initiate voicing. Disrupted control over articulators may contribute to a significant deviation in the VOT. Generally, an abnormal VOT can be associated with the disrupted coordination of the laryngeal and supralaryngeal muscles. In this thesis, the VOT has been assigned to the subsystem of phonation in accordance with the literature (Duffy **2013**), regarding the fact that only voiceless stops were the subject of analysis.

The VOT was measured as the median duration of the intervals between detected bursts and following voice onsets.

### ***Diadochokinetic rate (DDKR)***

The rate of alternating movements in the diadochokinetic test is traditionally used by speech pathologists to assess overall oral motor function. The DDKR is defined as the number of syllables spoken in a given time period. Decreased DDKR refers to deteriorated articulatory performance.

The DDKR was estimated as the inversion of the median duration between consecutive voice onsets. Median was preferred in order to increase robustness against misdetections.

### ***Vowel duration (VD)***

The slowness of repetitive movements with excessive vocal emphasis typical of ataxic dysarthria can propagate into the prolongation of vowels, as measured by VD.

The VD was estimated as the mean duration of detected voiced intervals.

### ***Diadochokinetic irregularity (DDKI)***

Involuntary movements can superimpose intended movements of the vocal tract, making the pace of the alternating motion more irregular. Increased values of DDKI can be accounted preeminently to involuntary movements of the speech apparatus, but the contribution of disturbed timing should also be assessed, as timing deficits, such as the acceleration of speech in PD, may project into the overall irregularity.

The DDKI was estimated as the standard deviation of the measured durations between consecutive voice onsets.

### ***The standard deviation of power (stdPWR)***

Diadochokinesis represents a task performed within one respiratory cycle as a series of isolated syllables. Airflow between syllables can be stopped rapidly only by a complete blockage of airflow using articulators or adduction of the vocal folds. Steady loudness then arises naturally from the fact that the speaker does not have to vary his respiratory effort to perform the DDK task. Thus, difficulties with respiratory and laryngeal/supralaryngeal control can be the hypothetical cause. The values of the feature can be increased substantially due to incoordination or involuntary movements. The literature associates increased values typically with ataxic and hyperkinetic dysarthria (Kent et al. 2000, Hartelious et al. 2003), but they can also be expected in other dysarthrias of various neurogenic origin, such as stroke (Kent et al. 1999).

The feature stdPWR in DDK was calculated using the same methodology as was used for stdPWR in connected speech, as described in section 2.3.3 CONNECTED SPEECH in chapter STANDARD DEVIATION OF POWER (STDPWR), page 38.

## 2.4 MODELING OF SPEECH PATTERNS

### 2.4.1 Normalization

A statistical model was established in order to increase the interpretability of individual features and speech patterns (see Figure 7). Normalization was realized by comparing a measured value of each speech feature with normative values measured on a group of HC matched to the sex and age characteristics of the examined speaker. Normative values were estimated via the following process.

A sample of at least 30 healthy individuals with ages and/or sex similar to each examined speaker was selected from the precomputed database of healthy controls using the following rules:

- *If the number of speakers with the same sex and maximal age difference less than 5 years in the database of healthy controls is greater than 30, consider this sample as a matched group.*
- *Else, apply the same rule with the condition of maximal age difference less than 10 years.*
- *Else, apply the same rule with the condition of maximal age difference less than 20 years.*

The threshold rules are more convenient than the simple selection of an exact number of similar speakers because it allows the selection of a larger group of speakers with a sufficient margin of error. Normative data were calculated for an age series from 30 to 75 years with a step of one year separately for each sex.

Each speech feature was modeled by a cumulative distribution function of values measured on matched healthy individuals. The distribution function was described as a normal distribution, log-normal distribution, or gamma distribution (see APPENDIX B: NORMALIZED VALUES OF SPEECH FEATURES, page 105). The mean and standard deviation of the normal distribution were estimated by the median and rescaled median absolute deviation, respectively. The parameters of the lognormal distribution were determined similarly by using values transformed via base 10 logarithms, and the parameters of the gamma distribution were estimated via maximum likelihood.

Further, the parameters of the distribution sorted by age were zero-phase filtered by a moving average with a duration of 5 years in order to reduce the random variation of the sample.

The probability of each speech feature was calculated by inspection of the corresponding lower tail, upper tail, or both tails of the cumulative distribution function, when speech abnormality was related to decreased value, increased value, or both directions, respectively. Normalized values were also expressed in terms of z-scores. Z-scores of features with non-normal distributions were calculated by transforming the estimated probability.

### 2.4.2 Combination of probabilities

The evaluation of the speech dimension or speech patterns requires that the evidence of individual speech features be combined. The normalized values of speech features expressed in a z-score can be combined according to Lipták (1958):

$$Z \sim \frac{\sum_{i=1}^G w_i \cdot Z_i}{\sqrt{\sum_{i=1}^G w_i^2}}, \tag{Equation 31}$$

where  $Z$  represents the resulting data fusion,  $Z_i$  is the z-score of hypothesis  $i$ ,  $w_i$  is the weight assigned to hypothesis  $i$ , and  $G$  is the total number of hypotheses. The presented equation is a weighted version of the z-transform test introduced by Stouffer et al. (1949), which is referenced by some authors as the weighted Stouffer’s method, weighted z-test, or weighted inverse normal method. Lipták’s method reduces to Stouffer’s method in the case when all of the weights of the hypotheses are equal. The original motivation of the weighting was to combine the results of independent studies with regard to their power. Division by the squared sum of squares ensures a

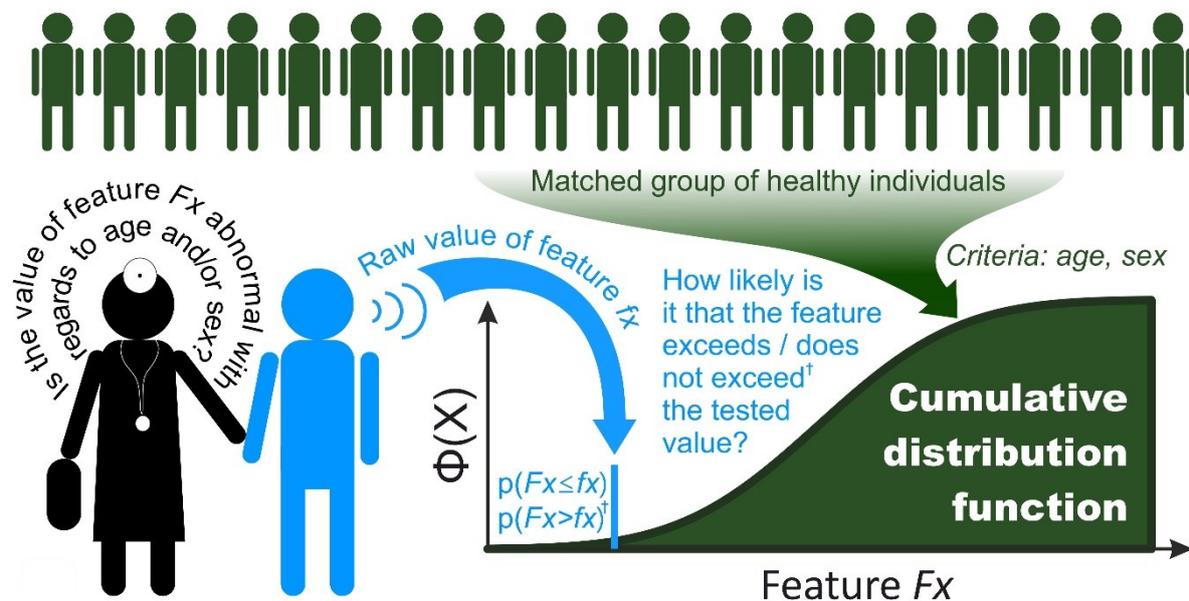


Figure 7: Illustration of the normalization process.  
 † The choice between the cumulative distribution function  $\Phi(F_x)$  describing  $p(F_x \leq f_x)$ , complementary cumulative distribution function  $1 - \Phi(F_x)$  describing  $p(F_x > f_x)$ , or analysis of both tails depended on the hypothesis and design of each feature specified in section 2.3 Acoustic analysis, page 20.  
**Symbols and abbreviations:**  $F_x$  = a generic speech feature,  $f_x$  = measured value of a generic speech feature,  $\Phi$  = cumulative distribution function,  $p$  = probability.

unity variance for the  $Z$  statistic. The combined probability of the null hypothesis corresponding to  $Z$  can then be calculated directly from the inverse cumulative standard normal distribution.

Unlike a simple weighted average of z-scores, Lipták's method allows one to constitute stronger support for a decision when the effect is evidenced simultaneously in multiple hypotheses. Even insignificant results (e.g.,  $p > 0.05$ ) can be combined into a significant one (e.g.,  $p < 0.05$ ) using Lipták's approach. Therefore, particular attention should be paid to assumptions of independence. Many speech abnormalities can propagate into multiple speech features, which is not typically the result of the multidimensionality of a disorder or the dependence of the hypotheses. For illustration, a speaker with hoarseness is more likely to have an increased harmonics-to-noise ratio as well as shimmer, even though no causality between the metrics has been established. Nevertheless, some speech features, namely EFn\_m and EFn\_SD, are dependent by definition, since an increased EFn\_SD implies an increased EFn\_M. Although, the assumption of independence is violated in these rare cases, the root of the sum of squares in the denominator of Equation 31 is still desired for more practical reasons. The root of the sum of squares of weights allows the introduction of negative weights without an additional correction of the sign. Lipták's approach is thus employed here as a convenient tool for information fusion that provides an interpretable insight into complex patterns of acoustic speech features.

### 2.4.3 Pattern analysis

Speech patterns were analyzed via the aggregation of probabilities according to Lipták (1958). Here, we depart from the original idea of the method (Lipták 1958) and combine probabilities based on the significance of the observed effect in the analyzed pattern. This new approach of pattern recognition hence referred as supervised weighted fusion of z-scores (SWFZ) has not been utilized previously. A pattern is represented as a linear combination of features defined by Equation 31, which makes results highly interpretable and allows the straightforward decomposition of a speech pattern. The weight of an individual speech feature can be seen as an enumerated importance of the feature for a description of a pattern, or, in other words, the degree to which we trust the feature in terms of the recognition of a pattern. The training of the pattern by this methodology requires only that the weights of individual features be estimated. The estimation of weights can follow the assumed importance of a feature for distinguishing dysarthria—a comprehensive summary of characteristics and their importance can be found in Duffy (2013)—or can be realized by minimizing a cost function on the training dataset. Note that both solutions can be similarly successful, but we will focus here on the estimation from the dataset because it requires no expert knowledge about relations between clinical characteristics and acoustic features.

Weights were estimated on the training dataset using a batch gradient descend and the following procedure. A binary hypothesis about the presence or absence of the pattern for one particular speaker was introduced, and the level of significance was set to  $p < 0.05$ . Then the combined z-score was mapped to the binary decision via a logistic function defined as:

$$f(Z) = \frac{1}{1 + e^{-s \cdot (Z - Z_0)}}, \quad \text{Equation 32}$$

where  $Z$  is the aggregated z-score,  $Z_0$  is the one-tailed z-score corresponding to the level of significance, and  $s$  is steepness of the curve. Steepness was set to the value of 1. Weights were

randomly initialized and optimized using a batch gradient descend with cost function  $J$ , which was defined as:

$$J(W) = \frac{1}{2} \sum_{k=1}^M (Y_k - \hat{Y}_k)^2, \quad \text{Equation 33}$$

where  $W$  represents the set of optimized weights,  $i$  denotes one particular speaker out of  $M$  speakers total,  $Y_k$  is the reference label of the speaker, and  $\hat{Y}_k$  is the predicted label determined by the logistic function of the aggregated z-score  $f(Z_k)$ . The gradient was estimated analytically from Equation 31, and the learning rate was set to  $10^{-4}$ . A vanishing gradient problem was prevented by saturation of the z-score to the maximal absolute value of 5 standard deviations. The saturation was employed also in testing and routine applications in order to diminish the effect of individual extremes on the combined z-score.

The aggregated z-score of the pattern can be calculated for each speaker from the known z-scores of individual speech features measured on the speaker and the weights of individual speech features from the analyzed pattern using Equation 31. The value of the aggregated z-score represents the salience of the speech pattern and can also be interpreted in terms of the probability of the speech pattern. A hypothesis concerning the presence of the speech pattern can be then tested by a defined level of significance.

#### 2.4.4 Pattern decomposition

When the weights of individual features of the pattern are known, any speech pattern of an individual speaker can be decomposed using the following procedure. Features with a positive product of  $w_i$  and  $Z_i$  make a speech pattern more salient. To the contrary, features with a negative product of  $w_i$  and  $Z_i$  balance a speech pattern to normality. The degree to which a speech feature with the positive product of  $w_i$  and  $Z_i$  influences the speech pattern of an individual speaker's contribution can be computed by the following equation:

$$C_i \sim \frac{w_i \cdot Z_i}{\sum_{i=1}^G w_i \cdot Z_i}, \quad \text{Equation 34}$$

where  $C_i$  is the relative contribution of the speech feature  $i$  with weight  $w_i$  and the individual speaker's performance  $Z_i$  to the speech pattern supported by  $G$  speech features with a positive product of  $w_i$  and  $Z_i$ . The scores of features can be grouped together, which allows the individual speech pattern to be assessed from a different perspective, e.g., speech dimension. Grouped contributions can be estimated by averaging when the number of descriptors varies across groups. The contribution of a feature or a dimension to the speech pattern illustrates the composition of the speech disorder in a clear and understandable way.

#### 2.4.5 Excitatory and inhibitory speech patterns

Here, excitatory and inhibitory speech patterns are introduced in order to simplify the categorization of acoustic speech abnormalities into a comprehensible form that will not interfere with established categories of dysarthria. The idea of inhibitory and excitatory speech patterns was coined by the author of this thesis after the recognition of a considerable overlap between

dysarthria characteristics and the pointlessness of categorization based purely on instrumentation data (see section 1.3 EXAMINATION OF DYSARTHRIA, page 4). Inhibitory and excitatory tendencies in speech patterns can be seen as two antagonistic forces competing persistently underneath any speech movement. When the balance between inhibition and excitation in motor control circuitry is disrupted, the motor activity tends to be reduced or exaggerated, respectively. An inhibitory speech pattern can be associated with hypokinesia of speech movements, whereas an excitatory speech pattern can be associated with hyperkinesia of speech movements. The typical characteristics of an inhibitory speech pattern can be demonstrated on hypokinetic movement disorders, such as PD, whereas an excitatory speech pattern can be demonstrated on hyperkinetic movement disorders, such as HD.

Note that the nomenclature for the dysarthria categories differentiates hyperkinetic and ataxic dysarthria, but movement disorders include ataxia as a special form of hyperkinetic movement disorder. The definition of excitatory and inhibitory patterns of speech features introduced here relates to movements; thus, ataxic dysarthria is assigned to the excitatory pattern. Indeed, discoordination associated with ataxic dysarthria manifests naturally with the exaggeration of movements.

Speech tendencies that are related rather to the severity of the speech disorder were pooled in an unspecific speech pattern introduced for completing the set of speech patterns. Unspecific speech patterns also include features that otherwise represent inhibition but are used frequently as a strategy for the compensation of a speech disability, such as slow NSR. Acoustic features employed in this thesis for a description of these patterns are listed in section 2.7 CLASSIFICATION EXPERIMENT, page 48.

## 2.5 INTERPRETATION AND VISUALIZATION OF RESULTS

Several factors should be taken into account by the examiner in order to avoid a flawed inference concerning a speech disorder. First, a normalized result reflects the cumulative probability of the raw value in the healthy population and should be interpreted as such. The value of a normalized feature refers to the rareness of the raw value in the healthy population and does not imply malfunction automatically. An abnormal value of a speech feature may be observed even in a healthy speaker demonstrating a speech idiosyncrasy. The examiner can deduce a speech deficiency from the low incidence of the feature in the healthy population and/or concurrence of multiple abnormal speech features. Second, the performance of the task can be influenced by incorrect instructions or a lack of understanding of the correct instructions. The diagnosis of a severe speech disorder based on the performance of a single speech task should be questioned because a severe speech disorder will more likely manifest in more than a single task. The reiteration of the speech task in question is highly recommended in order to confirm a suspicion about a speech disorder arising from a single task. A consistently abnormal result in a single task for different speakers may indicate an issue with the examiner. The resulting speech patterns should be interpreted in the context of the overall examination with regard to pattern composition. When the speech pattern is not well-spread, the reliability of the finding should be considered.

The complicated process of interpretation cannot be realized by the fully automated approach because the diagnosis of a speech disorder is inferred by more factors than just the

acoustic features. Therefore, the emphasis is on the lucid reporting of results. The results of a speech analysis were mediated in a comprehensive report with following qualities (see Figures in section 3.5 CASE STUDIES, page 61). First, normal values defined by  $p > 0.05$  were plotted using green rounded shapes, such as circles or leaf-like shapes. Abnormal values defined by  $p < 0.05$  were emphasized with red-cornered shapes, such as squares or triangles. Individual features were arranged in a radar chart according to their hypothesized dominant speech dimension. Speech features were labeled with an abbreviation and a numerical code corresponding to speech task. Each speech dimension was evaluated by the combined probability (see section 2.4.2 COMBINATION OF PROBABILITIES, page 43) of all of the corresponding speech features with equal weighting. The normality or abnormality of the speech dimension was indicated by the shape and color of the dimension labels on the radar chart. Features were also combined with equal weights to corresponding speech tasks. The performance of each task was plotted in a bar chart. Labels of speech tasks were denoted with a numerical code, which provided a reference for the numerical codes for each task for the individual speech features in the radar chart. The brief results of the speech pattern analysis were plotted in the bar graph. A table of results then provided a reference for all abbreviations, values, and descriptions. As a result, an examiner can find the significant features on the top of the table easily, as all items in the table were sorted by their significance. Z-scores were also reported in the table, supplementing the information about the resulting trend. Results which were higher than norm had a positive z-score; to the contrary, a negative z-score meant that the result was less than the norm. The table also suggests a defined interpretation of significant results with respect to the trend of an abnormality. When a subject is studied longitudinally, the results summarized in one file can be viewed individually by selecting a recording session. Clinicians may be interested more in the development of individual features over time. Therefore, longitudinal graphs of individual features were also provided by following the same visual philosophy described above. Finally, a quick overview of significant speech patterns was offered via a pie chart and table reporting on the contributions of the speech dimensions and features associated with a speech pattern.

The visual form of the report aimed to satisfy busy clinicians as well as patients uninitiated in speech analysis. A fully automated procedure for the generation of a described graphical report was implemented. The report document was coded in HyperText Markup Language (HTML), with figures using Scalable Vector Graphics (SVG).

## 2.6 STATISTICAL ANALYSIS

The normality of speech features was tested using the Kolmogorov-Smirnov test. Normalized z-scores of speech features were preferred for the analysis over raw values of speech features, as they allow groups to be compared regardless of age and sex differences. All disease groups were compared only to the HC in order to test for the presence or absence of a speech symptom within the group, which not only simplified the interpretation of the results but also reduced the number of comparisons, thus preventing unnecessarily wide confidence intervals. A clinician can consider the rate of progression, medication, or disease duration from an abnormality of an acoustic feature but could hardly infer a differential diagnosis based the severity of an individual acoustic feature. Thus, a comparison between diseases could provide redundant or even misleading information for practical application. A group comparison of normally or log-normally distributed features was conducted via a one-way ANOVA, followed by the Westfall-Young procedure for multiple comparisons. Gamma-distributed features were compared via the Kruskal-Wallis test, followed by a many-to-one comparison according to Gao et al. (2008) with a Hochberg's step-up procedure.

The correlation of normalized speech features with clinical scales was carried out via Pearson's correlation coefficient. The level of significance was set to 0.05. A statistical comparison intended to demonstrate the different trends in speech disorders for given groups was carried out in order to provide a summary for interpreting the results of individual speakers. In this scenario, the rejection of the hypotheses is more favorable for the determination of whether a symptom is group-specific or not. Therefore, the family-wise error rate of the feature set was not controlled in order to avoid inflation of the type II error.

## 2.7 CLASSIFICATION EXPERIMENT

A classification experiment was introduced in order to estimate the incidence of speech patterns and to compare the performance of the SWFZ for the modeling of speech patterns with state-of-the-art classification techniques. Only inhibited, excited, and unspecific speech patterns were subject to analysis. Hypothetically, all three patterns may be present in a single subject. Therefore, the classification categories were defined by following rules that assigned only one category to a subject.

- Subjects with no inhibitory, excitatory, or unspecific speech pattern were categorized as no pattern present.
- Subjects with only an inhibitory speech pattern were categorized as inhibitory.
- Subjects with only an excitatory speech pattern were categorized as excitatory.
- Subjects with both inhibitory and excitatory speech patterns were categorized as mixed.
- Subjects with no inhibitory and/or excitatory speech patterns showing an unspecific speech pattern were categorized as unspecific.

Inhibitory and excitatory speech patterns are more vital for the description of trends in speech movements than an unspecific speech pattern that is only complementary and could be expected in any disease. Therefore, inhibitory and excitatory speech patterns were assigned a higher priority than unspecific speech patterns. The incidence of each speech pattern, including none, inhibitory, excitatory, mixed, and unspecific, in a group was calculated as the proportion of subjects labeled with the respective category within the group.

A subsample of speakers that manifested a speech disorder was selected from the dataset for classification experiment via the following criteria. Only PDU and PDT subjects with speech item of UPDRS III equal to 1 or higher were included. Only MSA and PSP subjects with speech item of NNIPPS equal or higher than 1 were added to the subsample. Only HDU and HDT subjects with speech item of UHDRS equal of 1 or higher were accepted for the subsample. Only MS and CA subjects with mild dysarthria identified by a speech-language pathologist were added to the subsample. All HC and RBD subjects were included unconditionally.

An inhibitory speech pattern was described by the stdF0 measured on the reading passage and monologue and the stdPWR measured on the reading passage. A model of an inhibitory speech pattern was trained on subjects with PDU and PDT versus subjects with HC. An excitatory speech pattern consisted of stdPSD, stdF0, and DVA, all measured on sustained phonation of the vowel /A/. A model of an excitatory speech pattern was trained on subjects with HDU, HDT, MSA, and CA versus HC. An unspecific speech pattern was based on features that hypothetically represented an unspecific speech abnormality, were related to the severity of a speech disorder, or were found to be abnormal in various diseases: VOT, DDKR, and DDKI measured on the

diadochokinetic task; RI measured on the rhythm task; EFn\_M measured on the sustained phonation of vowel /I/; jitter and HNR measured on the sustained phonation of the vowel /A/; DUS, RFA, and DPI measured on both the reading passage and monologue; GVI measured on the monologue; and NSR measured on the reading passage. A model of an unspecific speech pattern was trained on subjects with PDU, PDT, HDU, HDT, CA, and MS versus HC. All models of speech patterns were always estimated on the training dataset. The number of subjects with a disease and HC in the training dataset was balanced by oversampling.

The overall incidence of speech patterns in the dataset was estimated via the leave-one-out cross-validation process. Each subject was excluded from the dataset and speech patterns were trained on the remaining subjects accordingly to PATTERN ANALYSIS, 44. Then trained speech patterns were applied to the excluded subject, and the p-value was calculated. The excluded subject was tested at the significance level 0.05 for each pattern analyzed and labeled with the decision. The process was repeated iteratively for all of the subjects in the database.

Leave-one-out cross-validation allows the unambiguous decision for each subject to be estimated, but models across all iterations are highly correlated, as only one subject is missing during each training iteration. Therefore, a randomized cross-validation realized by the process below was preferred for the estimation of the performance variability and a comparison with state-of-the-art classifiers. Iteratively, the dataset was stratified randomly in a training sample containing 75% of the subjects for each group and a testing sample containing 25% of the subjects for each group. In each iteration, the classifier was trained only on training data, and the trained model was applied to testing data. The incidence of speech patterns was calculated on the training sample. The process was repeated for 30 iterations. The SWFZ was compared with the following classifiers:

- **Neural network** with one hidden layer of five neurons, a positive linear transfer function in the hidden layer, and a sigmoid activation function at the output layer trained using the mean squared error loss function.
- **Naïve Bayes** classifier with probabilities calculated via kernel density estimation using a Gaussian window.
- **SVM** with radial basis function optimized via a grid search. The optimal combination was selected as the one with the maximal sum of incidences of associative diseases, e.g., incidences of inhibitory speech patterns for PD.

All classifiers were trained and tested within the same iteration, i.e., the random samples were identical for all classifiers. The incidences of speech patterns estimated by all classifiers were compared across all repetitions using Friedman's test. Finally, the overall performance was described as the mean and standard deviation of individual incidences across all iterations.

## 2.8 QUESTIONNAIRE FEEDBACK FROM CLINICIAN

The proposed methodology for speech assessment was implemented in © MATLAB (MathWorks, Natick, Massachusetts, USA), including a simple graphical user interface (see APPENDIX C: SOFTWARE APPLICATION, page 115). The application was used experimentally by experienced speech pathologist Hana Růžičková starting in July 2017 and ending in September 2018. The design of the application was developed systematically to meet the specific requirements of clinical

practice. The applicability of the system was investigated via a questionnaire (see Appendix D: Questionnaire feedback, page 121) in terms of customer satisfaction (questions 1-5), clinical applicability (questions 6-10), interpretability of provided results (questions 11-15), benefits (questions 16-20), and limitations (questions 21-25). Additional information was gathered through a series of conversations and observations of clinical examinations. The questionnaire was answered September 22, 2018, more than one month after the release of the final version. Speech pathologist Hana Růžičková scored every question on the scale from -5 to +5, where a more positive value represents a more positive answer, and, conversely, a more negative value means a more negative answer. The sign of the scores was corrected to reflect the overall positivity or negativity of the performance. Scores were expressed as a percentage, where -5 corresponded to a 0 % and +5 to a 100% performance. Specific comments were also recorded.

# 3

## RESULTS

*God is just giving me here my real perfect kind of a chance to just see and to just feel exactly how my own mother saw and felt.*

*–Woody Guthrie, Personal Correspondence, 1956*

A thorough evaluation of the methodology in terms of accuracy, a comparison of the groups in the database, a comparison of the SWFZ with conventional classifiers for pattern recognition, and a demonstration of clinical applicability is provided in this chapter. Generally, an accuracy analysis was performed across various diseases, including mild to severe stages of speech disorder, in order to demonstrate the validity of the methodology. The data used for evaluation were mostly a subset of the database presented in this thesis. Nevertheless, additional recordings that did not belong to the presented database were included to increase the sample size and make analysis more challenging. An analysis of the database is presented here in the context of previous findings or hypotheses in order to make the results more informative. The classification experiment provides information on how the proposed speech patterns are distributed across various diseases with regards to the comparability with current methods for pattern recognition. Clinical applicability is considered via the questionnaire survey of the clinician and two case studies, which are examined in light of the patients' histories and the author's suggested interpretations of two complex acoustic assessments.

## 3.1 TRACKING THE ACCURACY OF THE ANALYSIS

### 3.1.1 Connected speech

Accuracy was tested on intervals of respiration and speech/pause, where speech includes both voiced and unvoiced intervals of speech in order to simplify the presentation of the result. Segmentation was evaluated on 271 recordings of passages being read and monologues selected randomly from a dataset containing HC, RBD, PD, MSA, PSP, and HD patients. Intervals of speech, pause, and respiration obtained by manual segmentation were compared with the outcome of the automated segmentation in terms of the F1-score. Additionally, the detection accuracy was measured on the voice activity detector ITU-T G729B (International Telecommunication Union **1996**) and the pause detector for dysarthria by Rosen et al. (**2010**). The accuracy of detection was evaluated via F-score. Only detected labels paired with reference labels within tolerance were considered to be true positives. All detected labels outside tolerance were marked as false positives. All remaining reference labels were treated as false negatives. The tolerance for pauses was defined as a quarter of the duration of the corresponding pause. The tolerance for respiration was determined to be the duration of the corresponding respiration. Each recording was evaluated via F-score. More information about manual segmentation and evaluation can be found in the study by Hlavnička et al. (**2017**).

Figure 8 illustrates the accuracy of the automated segmentation. The proposed method, with an efficiency of pause detection of  $69.1 \pm 20.3\%$  outperformed the pause detector by Rosen et al. (**2010**), with an efficiency of  $32.8 \pm 12.3\%$ , and the ITU-T G.729B (International Telecommunication Union **1996**), with an efficiency of  $35.4 \pm 7.4\%$ , across all pause lengths. Respiration was detected efficiently with a score of  $73.8 \pm 20.3\%$ .

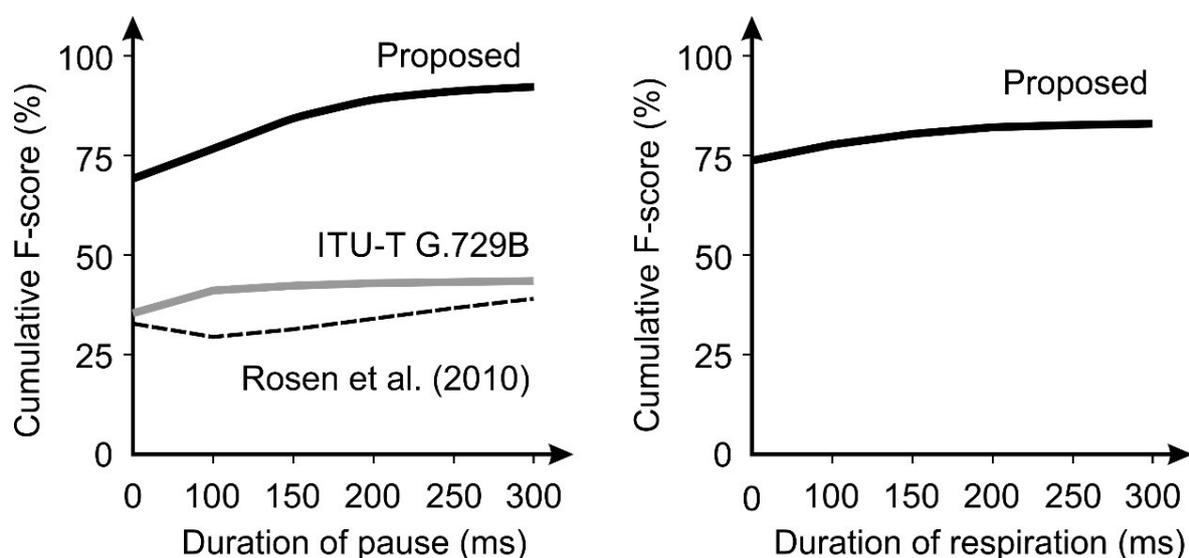


Figure 8: Detection efficiency of pause and respiratory intervals in connected speech. The score is plotted as a cumulative function of interval length, i.e., the accuracy for 100 milliseconds in length was calculated on all intervals longer than 100 milliseconds. (A) Comparison of the proposed method with a pause detector by Rosen et al. (**2010**) and the ITU-T G.729B (International Telecommunication Union **1996**). (B) The efficiency of the proposed method in detection of respiratory intervals.

### 3.1.2 Rhythm

The detection of syllable nuclei was evaluated by comparing the outcome of the automated algorithm with the manually identified position of syllables in 207 recordings obtained from 109 subjects. The database included HC, PD, MSA, PSP, and HD subjects and subjects with ephedrone Parkinsonism. The error rate was calculated as the number of error detections divided by the total number of syllables. The accuracy was determined as the complement of the error out of 100%. The overall accuracy was estimated as the mean accuracy across all recording in the database. More information about the database, manual segmentation, the evaluation process, and deep analysis can be found in the study by Rusz et al. (2015A).

The algorithm showed a very high overall accuracy of  $99.6 \pm 2.0\%$ . The majority of the errors consisted of misclassified respirations.

### 3.1.3 Diadochokinetic task

The accuracy of segmentation was evaluated on a very large dataset of 698 recordings. All possible data were included in order to demonstrate the reliability of the method. Therefore, more groups than the database used in the rest of the dissertation were covered. The enhanced dataset consisted of 317 recordings of patients with Parkinson's disease, of which 258 recordings were of patients treated by deep brain stimulation, 76 recordings were done of patients with dystonia in both the ON and OFF periods, 191 recordings were made of healthy speakers, 78 recordings were from speakers with Huntington's disease, and 36 recordings were made of patients with rapid eye movement sleep behavior disorder. The majority of the patients manifested severe dysarthria. Most subjects were represented by two recordings. The maximal number of recordings for one subject was limited to three.

All recordings were segmented manually by Michal Novotný. The rules for segmentation were described in detail by Novotný et al. (2014). The method proposed in this thesis was compared with the method by Novotný et al. (2015), which represents the advanced version of the original paper (Novotný et al. 2014). Additionally, the accuracy of the teager energy operator (TEO) by Hansen et al. (2010), Bayesian step change-point detector (BSCD) by Čmejla et al. (2001), and Bayesian autoregressive change-point detector by Čmejla et al. (2004), representing the state-of-the-art in burst detection, were investigated.

The accuracy of burst detection, voice onset detection, and accuracy of speech features were examined independently in order to demonstrate the propagation of errors into the resulting speech features. Bursts were detected in the interval preceding each voice onset determined by manual segmentation, which ensured that the reported results would not be biased by the accuracy of voice onset detection. Bursts and voice onsets were both evaluated using the empirical cumulative distribution of the absolute difference between the reference label and detection. The errors of each recording were averaged to prevent the unequal influence of recordings with a high number of syllables. Errors were then expressed via cumulative distribution to increase intelligibility of comparison. All parameters of models, such as the order of linear predictive coding or possible systematical errors, were compensated to obtain best results. Speech features were evaluated via Pearson's correlation coefficient between features computed from reference labels and features computed from detected labels.

The evaluation of detection accuracy is illustrated in Figure 9. The proposed method outperformed other methods in terms of the detection of bursts and voice onset. The increased precision can also be observed in a higher correlation between the computed and reference speech features (see Table 3).

### 3.1.4 Sustained vowels

The accuracy of segmentation was evaluated on a database containing 22 HC subjects (11 men, 11 women), 22 patients with RBD (11 men, 11 women), 22 patients with PD (10 men, 12 women), 21 patients with MSA (9 men, 12 women), 18 patients with PSP (12 men, 6 women), and 20 patients with HD (9 men, 11 women). Each recording was segmented manually into voice and silence categories based on the inspection of the oscillogram and spectrogram. Only a periodic signal with a fundamental frequency from 50 to 500 Hz was labeled as voiced. Manual labels were then compared with labels obtained by automated segmentation. Only intervals with error less than 100 milliseconds were accepted as true positives. Undetected intervals of voice were considered to be false negatives. This rigorous approach was preferred because the speech features of the task can be influenced significantly by any misdetection. Undetected intervals of silence were accounted for as false positives. Each recording was evaluated for precision and recall and with an F-score. Overall scores were averaged across all recordings in the dataset. The results of the proposed method were compared with PRAAT in standard settings (voicing threshold of 0.45) and PRAAT with settings adjusted to a threshold similar to the one used in the proposed segmentation (voicing threshold of 0.24) and a much lower threshold for demonstrating the trend (voicing threshold of 0.2).

The accuracy of speech features was evaluated on 505 synthetic replicas of the sustained vowels /A/ and /I/. Synthetic replicas were preferred, as their parameters are perfectly known and are not biased by an error in measurement. Replicas were synthesized from parameters measured semi-automatically on the database. Values of modal  $F_0$ , jitter, shimmer, HNR, the position and

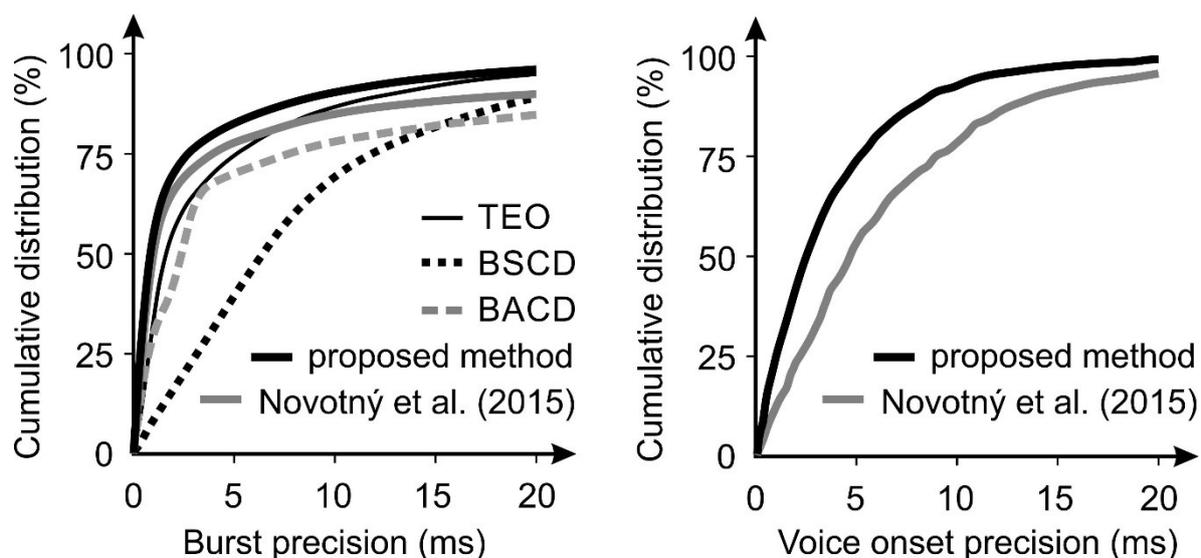


Figure 9: Cumulative distribution of segmentation errors in the diadochokinetic task.

The values of the cumulative distribution indicate how many recordings showed an error lower than the requested precision value, e.g., TEO showed an error of burst detection lower than 5 milliseconds for approximately 75% of recordings.

**Abbreviations:** BACD = Bayesian autoregressive change-point detector (Čmejla et al. 2004), BSCD = Bayesian step change-point detector (Čmejla et al. 2001), TEO = teager energy operator (Hansen et al. 2010), ms = milliseconds.

shape of each glottal pulse, the position of each subharmonic interval, as well as the depth of alternation of subharmonics were known for each synthetic signal. More information about synthesis as well as evaluation can be found in Hlavnička et al. (2019).

The accuracy of  $F_0$  detection was estimated by the mean semitone error (ME), standard deviation of error in semitones (SDE), root mean square error in semitones (RMSE), and median absolute semitone error (MAE), as defined below:

$$e_n = 12 \cdot \log_2 \hat{u}_n - 12 \cdot \log_2 u_n, \quad \text{Equation 35}$$

$$\text{ME} = \frac{1}{N} \sum_{n=1}^N e_n, \quad \text{Equation 36}$$

$$\text{SDE} = \sqrt{\frac{1}{N} \sum_{n=1}^N (e_n - \text{ME})^2}, \quad \text{Equation 37}$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{n=1}^N (e_n)^2}, \quad \text{Equation 38}$$

$$\text{MAE} = \text{median}|e_n|, \quad \text{Equation 39}$$

where  $e_n$  is the difference between the  $n$ -th estimation of the measured value  $\hat{u}_n$  and the reference value  $u_n$ , and  $N$  is total number of measurements. Only intervals that showed consent between the voiced/unvoiced decisions made by the algorithm and the reference were used in order to prevent bias caused by errors in voiced/unvoiced decisions. The results were compared with a large set of publicly available detectors. As some detectors may be susceptible to subharmonics and some not, the metrics was calculated on both modal  $F_0$  and  $F_0$  with regards to the subharmonics, i.e., modal  $F_0$  was corrected to  $F_0/2$  during subharmonic intervals. The results with the lowest overall RMSE were selected as representative of the performance of the detector.

The accuracy of subharmonic detection was evaluated via F-score, precision, and recall, where only the edges of intervals within 100 milliseconds tolerance around the reference value were accepted as true positives. All redundant detections or detections outside tolerance were accounted for as false positives. All undetected edges of intervals were considered to be false negatives. Additionally, the presence or absence of subharmonics detected by the algorithm was compared with the reference. All results were averaged across the database.

The proposed method for the estimation of jitter, shimmer, and HNR based on normalized cross-correlation was compared with PRAAT (Boersma and Weenink 2018). The error of prediction was estimated as the absolute difference between measurement and reference.

|               | Novotný et al. (2014) | Proposed method |
|---------------|-----------------------|-----------------|
| $\rho$ (VOT)  | 0.34                  | 0.73            |
| $\rho$ (DDKR) | 0.94                  | 0.98            |
| $\rho$ (DDKI) | 0.91                  | 0.88            |
| $\rho$ (VD)   | 0.75                  | 0.94            |

Table 3: Correlations between the reference and automated speech features.

**Abbreviations:**  $\rho$  = Pearson's correlation coefficient, VOT = voice onset time, DDKR = diadochokinetic rate, DDKI = diadochokinetic irregularity, VD = vowel duration. The computations of these features are described closely in section 2.3.4 DIADOCHOKINETIC TEST in chapter SPEECH FEATURES, page 41.

The evaluation of segmentation is summarized in Table 4. The proposed method achieved sufficient accuracy. The PRAAT showed good accuracy after lowering the decision threshold towards reduced periodicity. The default threshold performed with poor precision.

Figure 10 illustrates the accuracy of  $F_0$  detection compared with publicly available detectors. The majority of classifiers showed a lower RMSE for  $F_0$  regarding subharmonics, which limits their applicability in the assessment of phonatory dysfunction. The proposed method showed high accuracy for detection of modal  $F_0$  as well as for detection of subharmonic intervals (F-score= 91.59%  $\pm$  21.23 standard deviation (SD), mean precision= 92.43%  $\pm$  22.04 SD, and recall= 90.21%  $\pm$  23.19 SD). The detector decided reliably if subharmonics were present or absent in the recording (97.03% accuracy, 99.4% sensitivity, and 93.37% specificity).

Table 5 summarizes the median errors for each feature measured by PRAAT and the proposed method. The median was preferred in order to avoid bias caused by extreme values consequent to the erroneous tracking of the fundamental frequency.

Hypernasality measures were not validated within the scope of this thesis since the method was not developed by the author of the thesis, but the results of the original work will be cited here for completeness. Novotný et al. (2016) compared automated features with the perceptual ratings of 37 speakers with HC, 37 speakers with PD, and 37 speakers with HD and reported a strong correlation with perceptual ratings for EFn\_M ( $r=0.87$ ,  $p<0.001$ ) as well as for EFn\_SD ( $r=0.79$ ,  $p<0.001$ ).

## 3.2 STATISTICAL ANALYSIS

Normative data, information about units, and the distributions of individual speech features measured on HC can be found in APPENDIX A: NORMATIVE DATA FOR THE CZECH LANGUAGE, page 95. Characteristics of the normalized values of all speech features and the results of omnibus tests are provided in APPENDIX B: NORMALIZED VALUES OF SPEECH FEATURES, page 105. Table 6 and Table 7 summarize the comparison of all disease groups to HC. The hypotheses for all features are indicated for each tested tail in the interpretation column in Table 6 and Table 7. The vast majority of speech features showed a significant abnormality for at least one disease group. Only the features of RA in rhythm, EFn\_SD in the sustained vowel /I/, AST in the monologue, and DUF in the reading passage and monologue showed no significant effect in the omnibus test. Insignificant effects that relate to characteristics of a disease or were found significant in previous studies or may be hypothetically present are marked in Tables 1 and 2 in order to avoid misinterpretation of results due to the randomness of the sample. Table 6 and Table 7 can then be used as a guide through the complicated trends of acoustic speech features. No correlations between acoustic measures and clinical scales were found.

|                   | Precision     | Recall       | F1            |
|-------------------|---------------|--------------|---------------|
|                   | Mean / SD     | Mean / SD    | Mean / SD     |
| Proposed          | 99.41 / 4.45  | 99.72 / 2.94 | 99.49 / 3.49  |
| PRAAT (TH = 0.45) | 77.6 / 32.02  | 96.84 / 11.6 | 82.12 / 27.02 |
| PRAAT (TH = 0.24) | 88.61 / 23.43 | 98.2 / 8.94  | 91.23 / 18.81 |
| PRAAT (TH = 0.20) | 89.54 / 22.45 | 98.13 / 9.34 | 91.9 / 18.01  |

Table 4: Segmentation accuracy in sustained vowels expressed in percent.

**Abbreviations:** SD = standard deviation, TH = voicing threshold setting in PRAAT.

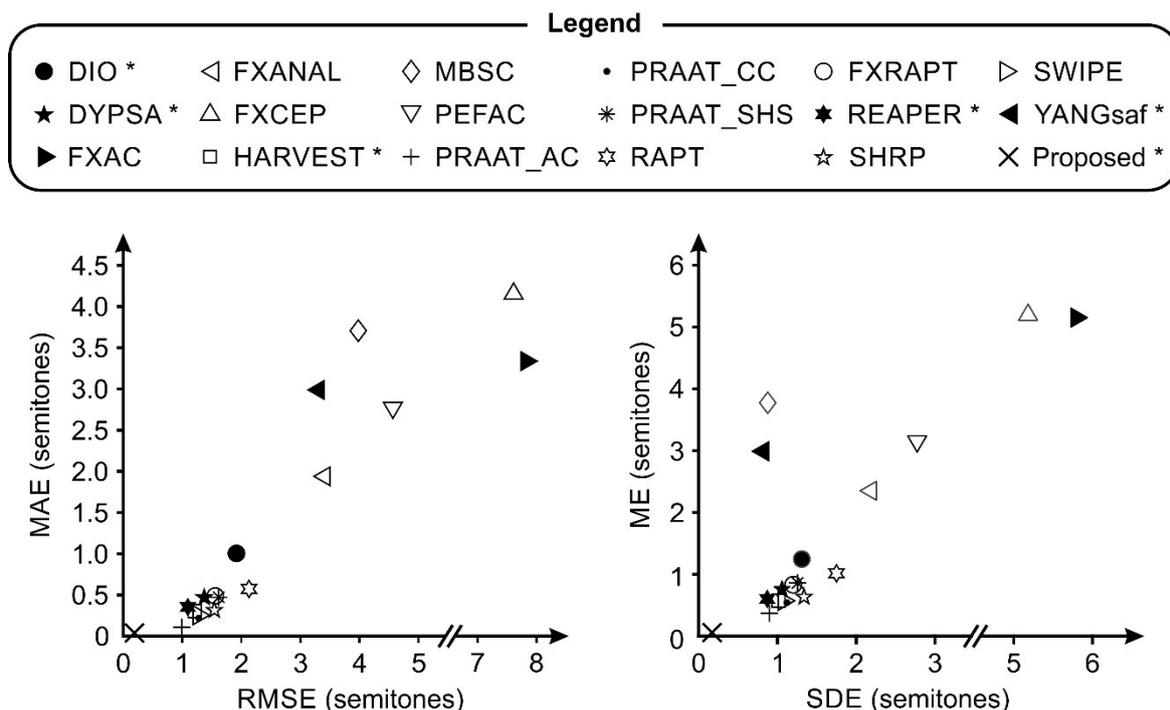


Figure 10: Accuracy of  $F_0$  detection by the proposed method and publicly available detectors.

The asterisk symbol (\*) denotes that modal  $F_0$  was preferred as the reference  $F_0$  for evaluation.

**Abbreviations:** DIO = detector by Morise et al. (2009) incorporated in the WORLD vocoder (2018), DYPSA = Dynamic Programming Projected Phase-Slope Algorithm by Naylor et al. (2007) released in VOICEBOX (2018), FXAC = autocorrelation of the cubed signal without tracking provided in the Speech Filing System (2018), FXANAL = autocorrelation of the cubed signal with tracking by Secrest and Doddington (1983) available in the Speech Filing System (2018), FXCEP = Noll’s (1967) cepstrum-based detector implemented in Speech Filing System (2018), Harvest = detector by Morise (2017) available in the WORLD vocoder (2018), MBSC = multiband summary correlogram by Tan and Alwan (2013) available as shareware code (MBSC, 2018), PEFAC = Pitch Estimation Filter with Amplitude Compression by Gonzales and Brookes (2014) available in VOICEBOX (2018), PRAAT\_AC = autocorrelation-based detector with comprehensive post-processing implemented in PRAAT (Boersma and Weenink 2018), PRAAT\_CC = cross-correlation version of the PRAAT\_AC algorithm (Boersma and Weenink 2018), PRAAT\_SHS = subharmonic summation on logarithmic frequency mantissa (Hermes et al. 1988) provided by PRAAT (Boersma and Weenink 2018), RAPT = autocorrelation-based detector with a robust algorithm for pitch tracking by Talkin (1995) available in VOICEBOX (2018), FXRAPT = robust algorithm for pitch tracking by Talkin (1995) applied on a normalized cross-correlation available in Speech Filing System (2018), REAPER = detector David Talkin developed at Google (2018), SHRP = Subharmonic-to-Harmonic Ratio Procedure by Sun (2002) and available online (2018), SWIPE = Sawtooth Waveform Inspired Pitch Estimator by Camacho et al. (2008) provided in the author’s dissertation (Camacho 2007), YANGsaf = detector from Yet Another Glottal source analysis framework (YANGsaf 2018) developed by Kawahara et al. (2016) at Google.

|          | Error of jitter | Error of shimmer | Error of HNR |
|----------|-----------------|------------------|--------------|
| PRAAT    | 0.37            | 4.72             | 2.36         |
| Proposed | 0.02            | 2.03             | 1.69         |

Table 5: Median prediction errors measured on the database of synthetic phonations.

**Abbreviations:** HNR = harmonics-to-noise ratio.

### 3.3 CLASSIFICATION EXPERIMENT

Comparison of the proposed method with state-of-the-art classifiers is summarized in Table 8. The repeated measures Friedman’s test failed to reject the hypothesis that type of classifier has no effect on estimated incidences [ $\chi^2(3)=1.33$ ,  $p=0.72$ ]. The incidence of speech patterns in the database recognized by the supervised weighted fusion of z-scores using leave-one-out cross-validation is illustrated in Figure 11.

| Feature              | Interpretation  | RBD | PDU | PDT | MSA | PSP | HDU | HDT | CA  | MS  |
|----------------------|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| <b>DDK</b>           |   |     |     |     |     |     |     |     |     |     |
| VOT                  | ↑ Disrupted coordination of laryngeal and supralaryngeal muscles. Decreased ability of laryngeal muscles to initiate voicing.   | -   | **  | -   | *** | *** | *** | *** | *** | *** |
| DDKR <sup>1</sup>    | ↓ Decreased rate of articulation.   | *   | +   | *   | *** | *** | *** | *** | *** | **  |
| VD                   | ↑ Slow movements and excessive vocal emphasis manifested by abnormally prolonged vowels.  | -   | -   | -   | *** | *** | *** | *** | *** | -   |
| DDKI                 | ↑ Pace of alternating motion is more irregular due to impaired timing, planning, or involuntary movements.  | *   | *   | *** | *** | *** | *** | *** | *** | *** |
| stdPWR               | ↑ Excess loudness variation due to involuntary movements of respiratory muscles or discoordinated phono-respiration.  | -   | -   | -   | *** | *** | *** | *** | *** | -   |
| <b>Rhythm</b>        |   |     |     |     |     |     |     |     |     |     |
| RA <sup>2</sup>      | ↑ Accelerating pace, also called oral festination.  | -   | ++  | ++  | ++  | ++  | -   | -   | -   | -   |
| RI                   | ↑ Irregular pace due to decreased speech motor control, discoordination, impaired timing, or presence of involuntary movements.   | -   | -   | *   | *** | *** | *** | *** | *** | -   |
| <b>Sustained //</b>  |   |     |     |     |     |     |     |     |     |     |
| EFn_M                | ↑ Increase hypernasality due typically to impaired control over elevator muscle of the soft palate.   | -   | -   | -   | **  | **  | *** | *** | -   | -   |
| EFn_SD <sup>3</sup>  | ↑ Intermittent hypernasality due to involuntary movements of elevator muscle of the soft palate.  | -   | -   | -   | -   | -   | +   | +   | -   | -   |
| <b>Sustained /A/</b> |   |     |     |     |     |     |     |     |     |     |
| DVA                  | ↑ Voicing stops suddenly due to abnormal laryngeal muscle contraction.  | -   | -   | -   | **  | -   | **  | *** | -   | -   |
| stdPSD               | ↑ Involuntary movements of articulators, preeminently the tongue.   | -   | -   | -   | *** | -   | *** | *** | *** | -   |
| MPT                  | ↓ Weak laryngeal and respiratory musculature.   | -   | -   | **  | **  | *   | *** | *** | -   | -   |
| stdFO                | ↑ Excess variation of fundamental frequency due to involuntary movements of laryngeal muscles or deteriorated motor control.  | -   | -   | -   | *** | *   | *** | *** | *** | *** |
| Jitter               | ↑ Unstable periods of glottal pulses. Associated with hoarseness.   | -   | -   | -   | -   | *   | *** | *   | -   | -   |
| Shimmer              | ↑ Unstable amplitudes of glottal pulses. Associated with hoarseness.  | -   | -   | -   | *** | *** | *** | *   | *** | -   |
| HNR                  | ↓ Increased noise due to turbulent airflow in vocal folds. Associated with hoarseness.  | -   | *   | -   | *** | *** | *** | *** | *** | *** |
| PSI <sup>4</sup>     | ↑ Vocal folds vibrate with alternating period, amplitude, or both. Geometrical or mechanical asymmetry of vocal folds. Associated with rough voice, pitch breaks, or special case of diplophonia. | -   | -   | -   | +   | +   | *   | -   | -   | -   |
| LSI                  | ↓ Vocal folds started subharmonic vibrations early in the course of the phonation. Vocal folds are either more prone to subharmonics or neuromuscular control of vocal folds is deteriorated.     | -   | -   | -   | *   | *** | **  | *** | -   | -   |

Table 6: Summary of acoustic features measured on diadochokinetic task, rhythm, and sustained vowels.

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ , -  $p > 0.05$ , ++ not significant in the sample but can be distinguishing when present, + not significant in the sample but can be expected, up arrows indicate an upper-tail hypothesis (increased when severe), down arrows indicate a lower-tail hypothesis (decrease when severe).

<sup>1</sup> findings of slow diadochokinetic rate are not consistent in the literature. Rusz et al. (2011) and Novotný et al. (2014) reported a significantly slower rate on cohorts of early, untreated PD patients. Harel et al. (2004) suggested that the slow rate may not be expected when compensatory strategies are implemented.

<sup>2</sup> acceleration of speech rate was observed in parkinsonian speech in studies by Skodda et al. (2010) and Rusz et al. (2015A). Acceleration of speech is a specific feature of hypokinetic dysarthria that is distinguishing but not invariably present (Duffy 2013).

<sup>3</sup> intermittend hypernasality is a symptom of hyperkinetic dysarthria (Duffy 2013). Increased EFn\_SD in HD was observed by Novotný et al. (2016).

<sup>4</sup> increased proportion of subharmonics can be expected in APS possibly due to spasticity (Hlavnička et al. 2019).

**Abbreviations:** RBD = rapid eye movement sleep behavior disorder, PDU = untreated Parkinson's disease, PDT = treated Parkinson's disease, MSA = multiple system atrophy, PSP = progressive supranuclear palsy, HDU = untreated Huntington's disease, HDT = treated Huntington's disease, CA = cerebellar ataxia, MS = multiple sclerosis, VOT = voice onset time, DDKR = diadochokinetic rate, VD = vowel duration, DDKI = diadochokinetic irregularity, stdPWR = standard deviation of power, RA = rhythm acceleration, RI = rhythm instability, EFn\_M = degree of hypernasality, EFn\_SD = intermittend hypernasality, DVA = degree of vocal arrests, stdPSD = standard deviation of power spectral density, MPT = maximum phonation time, stdFO = standard deviation of fundamental frequency, HNR = harmonics-to-noise ratio, PSI = proportion of subharmonic intervals, LSI = location of subharmonic intervals.

| Feature          | Interpretation   | Task      | RBD | PDU | PDT | MSA | PSP | HDU | HDT | CA  | MS  |
|------------------|--|-----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| EST              | ↓ Reduced stream of voiced, unvoiced, and pause intervals. Typically in consequence to reduced range of speech movements and/or decreased syllabic rate.                               | Reading   | -   | -   | -   | -   | **  | *** | *** | -   | -   |
|                  |  | Monologue | -   | -   | -   | *   | -   | *** | *** | -   | -   |
| RST              | ↓ Reduced stream of voiced, unvoiced, and pause intervals. Typically in consequence to reduced range of speech movements and/or decreased syllabic rate.                               | Reading   | -   | **  | -   | *** | *** | *** | *** | *** | -   |
|                  |  | Monologue | **  | **  | -   | *** | *** | **  | -   | *** | -   |
| AST <sup>1</sup> | ↑ Accelerating stream of voiced, unvoiced, and pause intervals resulting from increasing rate of speech movements.   | Reading   | -   | +   | +   | +   | +   | -   | -   | -   | -   |
|                  |  | Monologue | -   | -   | -   | -   | -   | -   | -   | -   | -   |
|                  | ↓ Decelerating stream of voiced, unvoiced, and pause intervals due to fatigue or decreased range of speech movements.  | Reading   | -   | +   | +   | **  | +   | *** | +   | +   | +   |
|                  |  | Monologue | -   | -   | -   | -   | -   | -   | -   | -   | -   |
| DPI <sup>2</sup> | ↑ Difficulties in initiating speech and/or omission of short pauses.   | Reading   | +   | *   | +   | *** | *** | *** | *** | **  | *   |
|                  |  | Monologue | +   | *   | +   | *** | *** | **  | *   | -   | -   |
| DVI              | ↑ Voicing interferes or continues within voiceless intervals. Decreased control of laryngeal muscles and coordination of laryngeal and supra-laryngeal muscles.                        | Reading   | -   | -   | -   | *** | *** | *** | *** | *** | -   |
|                  |  | Monologue | -   | -   | -   | *** | *** | -   | -   | **  | -   |
| GVI              | ↓ Decreased ability of vocal folds to stop voicing by adduction.   | Reading   | -   | -   | -   | *** | *** | *** | *** | -   | -   |
|                  |  | Monologue | *   | **  | -   | *** | *** | **  | **  | -   | -   |
| DUS              | ↑ Imperfect articulation of unvoiced stops. Unvoiced stops are prolonged or, for more extreme values, spirantized.   | Reading   | -   | -   | -   | *** | *** | *** | *** | -   | -   |
|                  |  | Monologue | -   | -   | -   | *** | *** | *** | *** | -   | *** |
| DUF              | ↑ Gradual weakening of friction in unvoiced fricatives.  | Reading   | -   | -   | -   | -   | -   | -   | -   | -   | -   |
|                  |  | Monologue | -   | -   | -   | -   | -   | -   | -   | -   | -   |
| RFA              | ↓ Acoustic resonances are less prominent due to articulatory imperfections such as mumbling.   | Reading   | -   | *** | -   | -   | -   | *   | -   | -   | *** |
|                  |  | Monologue | -   | **  | -   | -   | -   | **  | -   | -   | **  |
| RLR              | ↑ Excess inspiratory effort and/or obstruction in upper airways during inspiration.  | Reading   | -   | -   | -   | -   | -   | **  | -   | -   | -   |
|                  |  | Monologue | -   | -   | -   | -   | -   | *   | -   | -   | -   |
|                  |  | Monologue | -   | -   | -   | -   | -   | -   | -   | -   | -   |
| PIR              | ↓ Decreased pausing within breath groups. Decreased ability to control respiratory airflow.  | Reading   | -   | -   | -   | **  | *** | *** | *** | -   | -   |
|                  |  | Monologue | -   | -   | -   | *** | *** | *   | **  | -   | -   |
| RSR              | ↑ Increased rate of speech respiration. Inefficient respiration due to weakness of respiratory muscles or restricted range of movements. Imbalanced homeostasis may also be the cause. | Reading   | -   | -   | **  | -   | *** | -   | **  | -   | -   |
|                  |  | Monologue | -   | -   | -   | -   | **  | *** | *   | -   | -   |
| LRE              | ↑ Decreased ability to reverse from expiration to inspiration, especially difficulties in initiating inspiration.  | Reading   | -   | -   | -   | *** | *** | *   | *   | -   | -   |
|                  |  | Monologue | -   | -   | -   | *** | *** | *   | *** | **  | -   |
| stdPWR           | ↑ Excess loudness variation due to involuntary movements or deteriorated motor control.  | Reading   | -   | -   | -   | -   | -   | +   | +   | *   | -   |
|                  |  | Monologue | -   | -   | -   | -   | -   | *** | +   | +   | -   |
|                  | ↓ Abnormally low variation of loudness, also called monoloudness.  | Reading   | -   | *   | -   | -   | -   | -   | -   | -   | *** |
|                  |  | Monologue | -   | -   | -   | -   | -   | -   | -   | -   | -   |
| stdF0            | ↑ Excess variation of fundamental frequency due to involuntary movements of laryngeal muscles.   | Reading   | -   | -   | -   | -   | -   | -   | -   | -   | -   |
|                  |  | Monologue | -   | -   | -   | -   | -   | -   | -   | -   | -   |
|                  | ↓ Abnormally low variation of pitch, also called monopitch.  | Reading   | *** | *** | *** | *** | *** | *** | -   | -   | *** |
|                  |  | Monologue | -   | **  | **  | *   | -   | -   | -   | -   | -   |
| NSR              | ↓ Decreased syllable rate.   | Reading   | -   | -   | -   | *   | *** | *** | *** | *** | **  |

Table 7: Summary of acoustic features measured on connected speech.

\*\*\* p<0.001, \*\* p<0.01, \* p<0.05, - p>0.05, ++ not significant in the sample but is distinguishing when present, + not significant in the sample but may be present, up arrows indicate an upper-tail hypothesis (increased when severe), down arrows indicate a lower-tail hypothesis (decrease when severe)

<sup>1</sup> both tails can be expected in Parkinsonism since acceleration of speech can be achieved by both an increased speech rate and decreased range of movements and may not be always present (Duffy 2013). The feature was marked rather as possibly present with regard to possible influence of fatigue and lack of reference for the evaluation of reliability.

<sup>2</sup> inappropriate silences is an established feature of hypokinetic dysarthria, and prolongation of pauses was included in summary by Duffy (2013) and described in RBD patients by Hlavnička et al. (2017).

<sup>3</sup> excess loudness variations are associated with hyperkinetic dysarthria as well as ataxic dysarthria (Duffy 2013).

**Abbreviations:** RBD = rapid eye movement sleep behavior disorder, PDU = untreated Parkinson's disease, PDT = treated Parkinson's disease, MSA = multiple system atrophy, PSP = progressive supranuclear palsy, HDU = untreated Huntington's disease, HDT = treated Huntington's disease, CA = cerebellar ataxia, MS = multiple sclerosis, EST = entropy of speech timing, RST = acceleration of speech timing, AST = acceleration of speech timing, DPI = duration of pause intervals, DVI = duration of voiced intervals, GVI = gaping in between voiced intervals, DUS = duration of unvoiced stops, DUF = decay of unvoiced fricatives, RFA = resonant frequency attenuation, RLR = relative loudness of respiration, PIR = pause intervals per respiration, RSR = rate of speech respiration, LRE = latency in respiratory exchange, stdPWR = standard deviation of power, stdF0 = standard deviation of fundamental frequency, NSR = net speech rate. "Reading" references the reading passage task.

|                       | None          | Unspecific    | Inhibitory    | Excitatory    | Mixed         |
|-----------------------|---------------|---------------|---------------|---------------|---------------|
|                       | Mean / SD     |
| <b>SWFZ</b>           |               |               |               |               |               |
| HC                    | 91.97 / 3.05  | 2.96 / 1.86   | 3.00 / 1.51   | 2.07 / 1.64   | 0.00 / 0.00   |
| RBD                   | 72.08 / 15.29 | 9.17 / 9.81   | 18.75 / 11.72 | 0.00 / 0.00   | 0.00 / 0.00   |
| PDU                   | 32.50 / 22.88 | 13.33 / 17.04 | 54.17 / 22.82 | 0.00 / 0.00   | 0.00 / 0.00   |
| PDT                   | 25.56 / 22.63 | 10.00 / 15.54 | 64.44 / 23.05 | 0.00 / 0.00   | 0.00 / 0.00   |
| MSA                   | 19.33 / 14.37 | 14.67 / 14.79 | 41.33 / 14.79 | 11.33 / 12.52 | 13.33 / 14.22 |
| PSP                   | 21.11 / 28.34 | 41.11 / 20.87 | 32.22 / 25.50 | 0.00 / 0.00   | 5.56 / 12.63  |
| HDU                   | 3.33 / 12.69  | 18.33 / 27.80 | 23.33 / 25.37 | 53.33 / 31.98 | 1.67 / 9.13   |
| HDT                   | 0.00 / 0.00   | 8.33 / 18.95  | 0.00 / 0.00   | 83.33 / 27.33 | 8.33 / 18.95  |
| CA                    | 4.44 / 11.52  | 57.78 / 28.94 | 15.56 / 19.04 | 22.22 / 18.22 | 0.00 / 0.00   |
| MS                    | 28.33 / 19.65 | 28.89 / 18.54 | 21.67 / 16.46 | 21.11 / 18.54 | 0.00 / 0.00   |
| <b>Naïve Bayes</b>    |               |               |               |               |               |
| HC                    | 90.28 / 3.71  | 3.57 / 2.24   | 4.84 / 2.55   | 1.31 / 1.33   | 0.00 / 0.00   |
| RBD                   | 72.08 / 12.58 | 10.00 / 7.63  | 17.92 / 11.22 | 0.00 / 0.00   | 0.00 / 0.00   |
| PDU                   | 42.50 / 27.97 | 0.83 / 4.56   | 56.67 / 27.02 | 0.00 / 0.00   | 0.00 / 0.00   |
| PDT                   | 33.33 / 23.16 | 1.11 / 6.09   | 65.56 / 23.95 | 0.00 / 0.00   | 0.00 / 0.00   |
| MSA                   | 22.00 / 16.90 | 6.67 / 9.59   | 44.67 / 18.71 | 12.00 / 13.49 | 14.67 / 14.79 |
| PSP                   | 22.22 / 18.22 | 33.33 / 21.44 | 33.33 / 27.68 | 0.00 / 0.00   | 11.11 / 15.98 |
| HDU                   | 0.00 / 0.00   | 1.67 / 9.13   | 25.00 / 25.43 | 66.67 / 27.33 | 6.67 / 17.29  |
| HDT                   | 0.00 / 0.00   | 0.00 / 0.00   | 0.00 / 0.00   | 86.67 / 22.49 | 13.33 / 22.49 |
| CA                    | 5.56 / 12.63  | 48.89 / 25.87 | 15.56 / 22.71 | 30.00 / 25.30 | 0.00 / 0.00   |
| MS                    | 40.00 / 16.14 | 12.78 / 11.32 | 27.78 / 16.57 | 17.78 / 14.47 | 1.67 / 5.09   |
| <b>SVM</b>            |               |               |               |               |               |
| HC                    | 93.57 / 3.07  | 1.83 / 1.39   | 3.33 / 2.08   | 1.27 / 1.50   | 0.00 / 0.00   |
| RBD                   | 69.17 / 15.99 | 10.83 / 11.24 | 20.00 / 13.77 | 0.00 / 0.00   | 0.00 / 0.00   |
| PDU                   | 35.83 / 22.44 | 9.17 / 12.25  | 55.00 / 23.12 | 0.00 / 0.00   | 0.00 / 0.00   |
| PDT                   | 41.11 / 22.63 | 5.56 / 12.63  | 53.33 / 25.67 | 0.00 / 0.00   | 0.00 / 0.00   |
| MSA                   | 20.67 / 12.30 | 19.33 / 19.29 | 36.00 / 20.61 | 14.67 / 12.79 | 9.33 / 10.15  |
| PSP                   | 17.78 / 20.96 | 36.67 / 20.25 | 37.78 / 27.31 | 0.00 / 0.00   | 7.78 / 14.34  |
| HDU                   | 1.67 / 9.13   | 33.33 / 30.32 | 5.00 / 15.26  | 60.00 / 30.51 | 0.00 / 0.00   |
| HDT                   | 0.00 / 0.00   | 5.00 / 15.26  | 0.00 / 0.00   | 71.67 / 31.30 | 23.33 / 25.37 |
| CA                    | 14.44 / 16.80 | 50.00 / 20.99 | 14.44 / 20.87 | 18.89 / 18.94 | 2.22 / 8.46   |
| MS                    | 36.11 / 20.57 | 20.56 / 14.31 | 22.22 / 14.73 | 19.44 / 17.00 | 1.67 / 5.09   |
| <b>Neural network</b> |               |               |               |               |               |
| HC                    | 92.63 / 2.93  | 2.68 / 1.97   | 3.05 / 2.34   | 1.60 / 1.37   | 0.05 / 0.26   |
| RBD                   | 77.92 / 15.29 | 7.92 / 11.12  | 14.17 / 13.02 | 0.00 / 0.00   | 0.00 / 0.00   |
| PDU                   | 55.00 / 25.76 | 0.83 / 4.56   | 44.17 / 26.82 | 0.00 / 0.00   | 0.00 / 0.00   |
| PDT                   | 45.56 / 29.66 | 6.67 / 13.56  | 47.78 / 31.18 | 0.00 / 0.00   | 0.00 / 0.00   |
| MSA                   | 23.33 / 15.83 | 19.33 / 19.99 | 38.00 / 28.45 | 9.33 / 14.61  | 10.00 / 13.65 |
| PSP                   | 25.56 / 29.92 | 32.22 / 29.66 | 33.33 / 33.90 | 1.11 / 6.09   | 7.78 / 14.34  |
| HDU                   | 3.33 / 12.69  | 11.67 / 21.51 | 16.67 / 23.97 | 63.33 / 34.57 | 5.00 / 15.26  |
| HDT                   | 0.00 / 0.00   | 6.67 / 21.71  | 0.00 / 0.00   | 81.67 / 27.80 | 11.67 / 21.51 |
| CA                    | 22.22 / 23.71 | 47.78 / 22.63 | 8.89 / 17.36  | 20.00 / 22.49 | 1.11 / 6.09   |
| MS                    | 51.67 / 22.04 | 16.11 / 14.83 | 17.22 / 14.17 | 13.33 / 11.91 | 1.67 / 5.09   |

Table 8: Incidences of speech patterns by randomized stratified cross-validation.

All values are expressed in percent.

**Abbreviations:** SD = standard deviation, SVM = support vector machine, HC = healthy control, RBD = rapid eye movement sleep behavior disorder, PDU = untreated Parkinson's disease, PDT = treated Parkinson's disease, MSA = multiple system atrophy, PSP = progressive supranuclear palsy, HDU = untreated Huntington's disease, HDT = treated Huntington's disease, CA = cerebellar ataxia, MS = multiple sclerosis, SVM = support vector machine, SWFZ = supervised weighted fusion of z-scores.

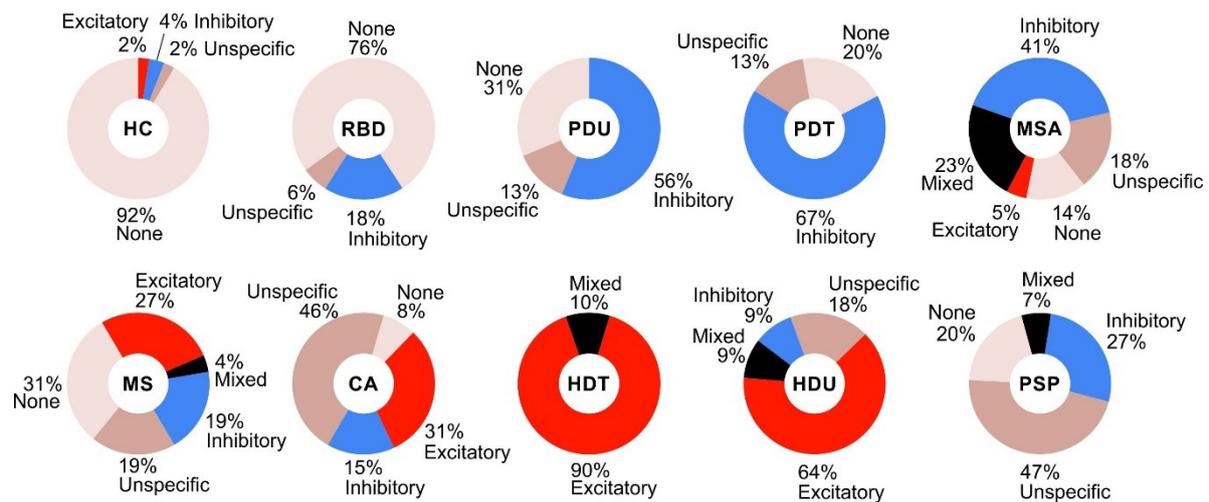


Figure 11: Incidences estimated by the leave-one-out cross-validation experiment.

**Abbreviations:** HC = healthy control, RBD = rapid eye movement sleep behavior disorder, PDU = untreated Parkinson's disease, PDT = treated Parkinson's disease, MSA = multiple system atrophy, PSP = progressive supranuclear palsy, HDU = untreated Huntington's disease, HDT = treated Huntington's disease, CA = cerebellar ataxia, MS = multiple sclerosis.

## 3.4 QUESTIONNAIRE FEEDBACK

The experienced clinical speech pathologist rated the software implementation of the proposed method (see APPENDIX D: QUESTIONNAIRE FEEDBACK, page 121) with a mean score of 92% regarding customer satisfaction (questions 1-5), 89% regarding clinical relevance (questions 6-10), 96% regarding interpretability of the provided results (questions 11-15), 90% regarding overall benefits (questions 16-20), and 88% regarding limitations of application (questions 21-25), where 100% corresponds to usefulness and 0% to uselessness. Detailed answers can be found in APPENDIX D: QUESTIONNAIRE FEEDBACK, page 121.

## 3.5 CASE STUDIES

The proposed methodology is demonstrated in this section on two case studies. Both subjects agreed with the recording and provided informed consent. The neurological diagnosis was conducted by a neurologist experienced in motor disorders. A speech-language-swallowing examination, including the recording of audio signals, was conducted by a speech pathologist. All signals were analyzed automatically using the proposed methodology and interpreted by the author of this thesis.

### 3.5.1 Case A

*Characterization:* Male born in 1932.

#### NEUROLOGICAL DIAGNOSIS

*Date:* November 30, 2017.

*Diagnosis:* Patient suffers from tremor dominant idiopathic Parkinson's disease. Hypokinetic dysarthria worsens over time. Dyskinesia in orofacial region probably induced by levodopa.

*Subjective:* Tremor at rest of right upper limb and less of right lower limb first manifested approximately in 2015. Patient is aware of reduced intelligibility. Medication lessens tremor, but its effect on speech is minimal.

*Pharmacological anamnesis:* Nakom mitte, PK-Merz.

*Follow-up diagnosis (July 28, 2018):* Parkinson's disease and possible vascular encephalopathy.

### **SPEECH-LANGUAGE-SWALLOWING PATHOLOGY DIAGNOSIS**

*Date:* January 23, 2018.

*Diagnosis:* Patient is lucid, cooperative; communicates verbally, fluently, coherently with sufficient information value. No indices of language disorder were found (not a target of examination). Neurologist recommended examination in order to inspect motor function of speech and swallowing movements.

*Subjective:* Worsening of speech was observed for last 2 months—mumbling, wheezing, unintelligible speech. Patient was cold 14 days ago; had inflammation of the tooth and mucosa in oral cavity—burning sensation, allegedly monitored by dentist. No difficulties in breathing and swallowing. Involuntary oral movements—patient bites himself sometimes.

*Facial movements:* Hypomimia, mild side facial-asymmetry, no paresis of *nervus facialis*, hypokinesia of mimic muscles. Patient shows no difficulties in keeping lips closed, protrusion and grinning are symmetrical, isolated lateral oral movements—left is limited. Diadochokinesis is hypokinetic, bradykinetic; deteriorated coordination—speech indicates oral apraxia. Lower jaw can move in elevation and depression. Protrusion is limited (laterally right is better but laterally left is limited). Coordination of complex rotation movements is worsened. Tonus of muscles of mastication is hard to evaluate. Resistance of jaw muscles against pressure of hand is sufficient. Scars from cheek biting are visible—scars are not atrophic. No pathological cover on tongue's surface was found. Protrusion of tongue was normal. Elevation of tongue is limited in all parts. Soft palate is symmetrical at rest and elevation. Pharyngeal reflex is present. Tactile sensitivity of oral cavity is sufficient. Involuntary movements of orofacial musculature were not present during examination—intermittent presence is suspected according to documented neurological examination.

*Phono-respiration:* Respiration at rest is regular—nose is involved. Maximal phonation time was 14 seconds. Hoarseness and tremolo were observed in prolonged phonation. Patient manifested hypophonia in connected speech. Durations of breath groups in connected speech are sufficient. Phono-respiration is intermittently discoordinated during speech; reserve volume is not fully expired, mild hypernasality, Peak cough flow was 344 l/m.

*Phonetics:* High-arched palate and cross-bite occlusion (influential to alveolar fricatives) were pre-morbid. Dysarthria manifests by weakened occlusive, decreased intelligibility—especially changing speech rate with tendencies to mumbling, monoloudness, and monopitch.

*Deglutition:* Head posture and body posture are voluntarily controlled. Self-reliance during eating is sufficient. Calorie intake is adequate. Lunch takes approximately 30 minutes. Patient lost 13 kg within last two years unintentionally (originally 93 kg weight, 180 cm height). Patient eats all kinds of food with consistency—no type is avoided, no thickening agents are used, no sipping. High temperature, infection of airways, gastroesophageal reflux, regurgitation, heartburn, and odynophagia were treated.

*Salivation:* Saliva gathers in right corner of lips causing occasional drooling. Speech is affected by retained saliva. Swallowing reflex initiates with latency. Elevation of larynx is sufficient. No gurgle or reflexive cough was present in phonation after deglutition.

*Liquids:* Patient is able to swallow liquids continually without thickening agent. Swallowing reflex is initiated with latency. Elevation of larynx is sufficient. Reflexive cough after deglutition was present and effective.

*Volume test:* 30ml per 2 deglutition (reflexive cough); 20ml per 1 deglutition (reflexive cough). Normative value for a man is 30ml per 2 deglutition.

*Speed test:* 100ml per 12 deglutition within 19 seconds (reflexive cough was not present). Normative value for a man is 100ml per 10 seconds. Considerably reduced speech of deglutition and size of bolus (circa 8.3ml) increased coordination of deglutition. Deglutition of solid foods was not examined. Patient mentioned that reflexive cough was present also after eating food with mixed consistency, such as soup, especially.

*Conclusion:* R47.1 hypokinetic dysarthria (suspected combination with hyperkinetic dysarthria); R13 dysphagia.

*Recommendations:* The patient was told to keep the regime and compensatory actions during eating and drinking—using thickening agents, especially. Weight must be monitored regularly. If unintended weight loss continues, the patient should be examined by a nutritionist and sipping should be evaluated. Motor therapy of speech and deglutition will be conducted by a speech pathologist.

*Videofluoroscopic examination of degustation on January 30, 2018:* Patient manifests silent aspiration during swallowing of fluids (Rosenbek 8)—residuals remain in airways. Contrast agent gets in contact with vocal folds during swallowing of thickened consistence (yogurt, Rosenbek 5)—contraindication with residuum. Contrast agent remain above vocal folds, and residuum is noticeable when swallowing solid food (sponge cake with barium contrast agent, Rosenbek 3). Moderate dysphagia, according to Daniels.

## **THERAPY OF SPEECH AND SWALLOWING**

Patient manifests no perceptual deficit of speech and written language. Main deficits are bradyphrenia, problems with memory, dysexecutive syndrome (limiting for therapy—little work can be done during one session). Patient trains at home regularly, but frequently erroneously or ineffectively. Therapy is focused primarily on deglutition and strengthening respiration. Some sessions were cancelled due to patient's injuries.

*Therapy session on February 2, 2018:* Speech motor training of articulators, strengthening laryngeal and pharyngeal muscles (to improve elevation of larynx during swallowing).

*Therapy session on March 3, 2018:* Speech motor training of articulators, strengthening laryngeal and pharyngeal muscles, training of supraglottic swallowing.

*Therapy session on April 17, 2018:* Speech motor training of articulators, strengthening laryngeal, pharyngeal muscles, and glottis, training respiration—air stacking for improvement of maximum insufflation capacity, phonation training focused on optimization of vocal register and strengthening of phonation, whistle register (elevation of larynx), fixation of supraglottic swallowing.

*Therapy session on May 15, 2018:* Speech motor training of articulators, training of phono-respiration– air stacking, phonation training focused on optimization of vocal register, threshold lung muscle trainer for both inspiration and expiration (delay caused by revision physician), inspiration threshold was recommended for beginning in order to become more familiar with the tools.

*Therapy session on June 26, 2018:* Patient forgot all tools at home. The training continued similarly to previous session. Only speech motor training was performed at home according to printed illustration of exercises. Patient was not able to train other exercises at home.

### **ACOUSTIC ANALYSIS**

Reading passage, rhythm task, diadochokinetic task, and sustained vowels were recorded during the initial speech-language-swallowing examination on January 23, 2018 and during the follow-up therapy sessions on March 6, May 15, and June 26, 2018. All signals were recorded and processed using the methodology and equipment described in section 2 Method, page 17. The automatically generated report is illustrated in Table 9, Figure 12, Figure 13, Figure 14, and Figure 15.

*Prosody:* Variation of melody was reduced significantly in the reading passage. Although variation of loudness was not abnormal, the patient manifested a significant inhibitory speech pattern. The flow of voiced, unvoiced, and pause intervals was also reduced, whereas the net speech rate remained normal, which indicates that the reduced range of speech movements was the relevant cause. In addition, the increased organization of speech regarding the occurrence of voiced, unvoiced, pause, and respiratory intervals suggests that the voiced speech had tendencies to dominate the speech flow, possibly due to impaired phonatory control, as detected by other speech features. Difficulties in initiating speech was another prosodic feature related to impaired phono-respiration. Increased duration of vowels in the diadochokinetic test point out that prosody was possibly influenced by the slowness of articulatory movements.

*Articulation:* Spirantization of unvoiced stops was observed in both tasks of connected speech. Additionally, the diadochokinetic task was performed at a normal rate but with irregular rhythm. However, no irregularity was found in the performance of the rhythm task, which suggests that the irregularity in diadochokinesis resulted from compensation of difficulties in articulation.

*Resonance:* A significantly increased degree of hypernasality was observed in the sustained vowel /I/. Hypernasality was steady with no indication of involuntary movements.

*Phonation:* Decreased gapping in between voiced intervals and prolongation of voiced intervals was found in both tasks dealing with connected speech, which, together with the significant prolongation of pauses in connected speech and the increased VOT in the diadochokinetic task, indicates deteriorated control over adduction and abduction of the vocal folds. A subharmonic vibrational regime manifested early in the course of phonation, lowering the likelihood that the phonatory deficit originates in abnormal respiration. Instabilities in the amplitude and waveform of the vibrations were also observed. The impaired phonatory control can be attributed to the hoarseness together with the possible influence of inflammation since the patient was cold two weeks before the examination.

*Respiration:* The increased rate of speech respiration during the monologue suggests a decreased effectivity of respiration. Breath groups were performed with fewer pauses, suggesting bad economy of respiration as well as decreased function of vocal folds as a valve for opening and closing the airway during exhalation, which is supported by findings of abnormal GVI and DVI.

| Symbol  | Task                 | Value | P-value | Z-score | Description                                    | Interpretation  |
|---------|----------------------|-------|---------|---------|--|---|
| DVI     | Monologue            | 419   | 0.0001  | 5.69    | Duration of voiced intervals (ms)              | Voicing interferes with or continues within voiceless intervals. Decreased control of laryngeal muscles and coordination of laryngeal and supra-laryngeal muscles.                          |
| DUS     | Monologue            | 58.4  | 0.0001  | 4.72    | Duration of unvoiced stops (ms)                | Imperfect articulation of unvoiced stops. Unvoiced stops are prolonged or, for more extreme values, spirantized.  |
| DVI     | Reading passage      | 352   | 0.0001  | 4.5     | Duration of voiced intervals (ms)              | Voicing interferes with or continues within voiceless intervals. Decreased control of laryngeal muscles and coordination of laryngeal and supra-laryngeal muscles.                          |
| RST     | Monologue            | 224   | 0.0001  | -4.16   | Rate of speech timing (intervals/s)            | Reduced stream of voiced, unvoiced, and pause intervals. Typically in consequence to reduced range of speech movements and/or decreased syllabic rate.                                      |
| EFn_M   | Sustained vowel /I/  | -31.6 | 0.0001  | 4.04    | Hypernasality mean (dB)                        | Increased hypernasality due typically to impaired control over elevator muscle of the soft palate.  |
| EST     | Reading passage      | 1.52  | 0.0002  | -3.85   | Entropy of speech timing (-)                   | Impaired coordination between subsystems, or insufficient control over one or more subsystems of speech, e.g., voiced speech tends to dominate speech typically in severe dysarthria.       |
| HNR     | Sustained vowel /A/  | 10.2  | 0.0002  | -3.61   | Harmonic to noise ratio (dB)                   | Increased noise due to turbulent airflow in vocal folds. Associated with hoarseness.  |
| LSI     | Sustained vowel /A/  | 3.87  | 0.0017  | -2.94   | Location of first subharmonic interval (s)     | Vocal folds started subharmonic vibrations early in the course of the phonation. Vocal folds are either more prone to subharmonics or neuromuscular control of vocal folds is deteriorated. |
| DUS     | Reading passage      | 35.9  | 0.0018  | 2.92    | Duration of unvoiced stops (ms)                | Imperfect articulation of unvoiced stops. Unvoiced stops are prolonged or, for more extreme values, spirantized.  |
| GVI     | Reading passage      | 24.4  | 0.0019  | -2.9    | Gaping in-between voiced intervals (pause/min) | Decreased ability of vocal folds to stop voicing by adduction.  |
| RST     | Reading passage      | 274   | 0.0026  | -2.8    | Rate of speech timing (intervals/s)            | Reduced stream of voiced, unvoiced, and pause intervals. Typically in consequence to reduced range of speech movements and/or decreased syllabic rate.                                      |
| GVI     | Monologue            | 18.1  | 0.0055  | -2.54   | Gaping in-between voiced intervals (pause/min) | Decreased ability of vocal folds to stop voicing by adduction.  |
| VOT     | Diadochokinetic task | 37.3  | 0.0072  | 2.45    | Voice Onset Time (ms)                          | Disrupted coordination of laryngeal and supralaryngeal muscles. Decreased ability of laryngeal muscles to initiate voicing.   |
| VD      | Diadochokinetic task | 78.3  | 0.0088  | 2.38    | Vowel duration (ms)                            | Slow movements and excessive vocal emphasis manifested by abnormally prolonged vowels.  |
| DPI     | Monologue            | 273   | 0.0113  | 2.28    | Duration of pause intervals (ms)               | Difficulties in initiating speech and/or omission of short pauses.  |
| RSR     | Monologue            | 25.8  | 0.0122  | 2.25    | Rate of speech respiration (respirations/min)  | Increased rate of speech respiration. Inefficient respiration or imbalanced homeostasis.  |
| DDKI    | Diadochokinetic task | 65.1  | 0.0123  | 2.25    | Diadochokinetic irregularity (ms)              | Pace of alternating motion is more irregular due to involuntary movements of speech apparatus or impaired timing.   |
| RLR     | Monologue            | -17.2 | 0.0144  | 2.45    | Relative loudness of respiration (dB)          | Excess inspiratory effort and/or obstruction in upper airways during inspiration.   |
| Shimmer | Sustained vowel /A/  | 5.54  | 0.0184  | 2.09    | Shimmer (%)                                    | Unstable amplitudes of glottal pulses. Associated with hoarseness.  |
| stdFO   | Reading passage      | 1.2   | 0.019   | -2.35   | Standard deviation of F0 (semitones)           | Abnormally low variation of pitch, also called monopitch  |
| PIR     | Monologue            | 2     | 0.0216  | -2.02   | Pause intervals per respiration (-)            | Decreased pausing within breath groups. Decreased ability to control respiratory airflow.   |
| DPI     | Reading passage      | 220   | 0.0238  | 1.98    | Duration of pause intervals (ms)               | Difficulties in initiating speech and/or omission of short pauses.  |
| RLR     | Reading passage      | -19.5 | 0.0705  | 1.81    | Relative loudness of respiration (dB)          | N/A   |
| RFA     | Monologue            | 7.73  | 0.0877  | -1.36   | Resonant frequency attenuation (dB)            | N/A   |

Table 9: Summary of most severe speech features of case A measured at the first recording session.

Findings were sorted by ascending p-values. The table represents an illustrative capture of the automated report. Only significant results and two insignificant features were included for illustration. Tasks were renamed according to the notation used in the thesis. Note that the reported interpretation was assigned automatically following definitions derived from Table 6 and Table 7.

**Abbreviations:** N/A = not available—marking insignificant results.

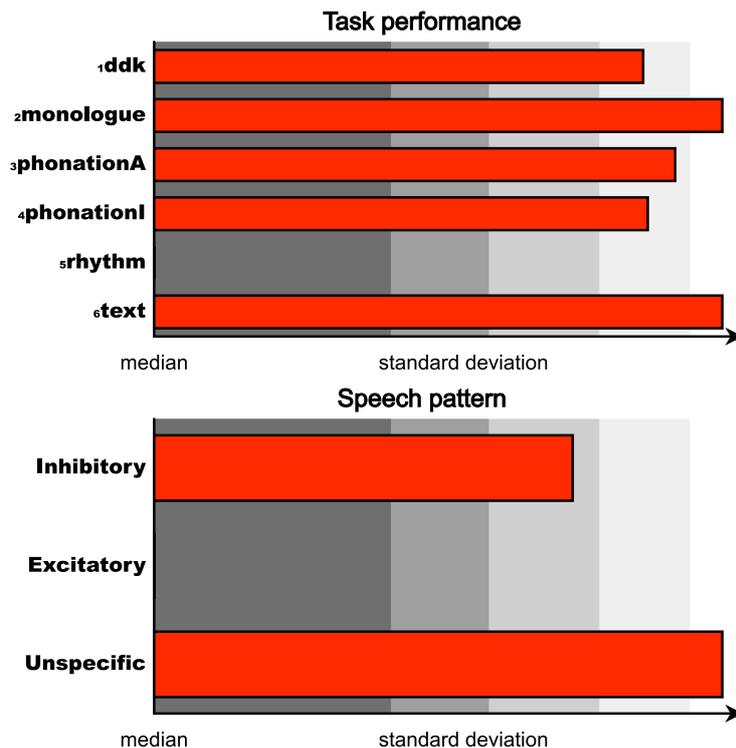
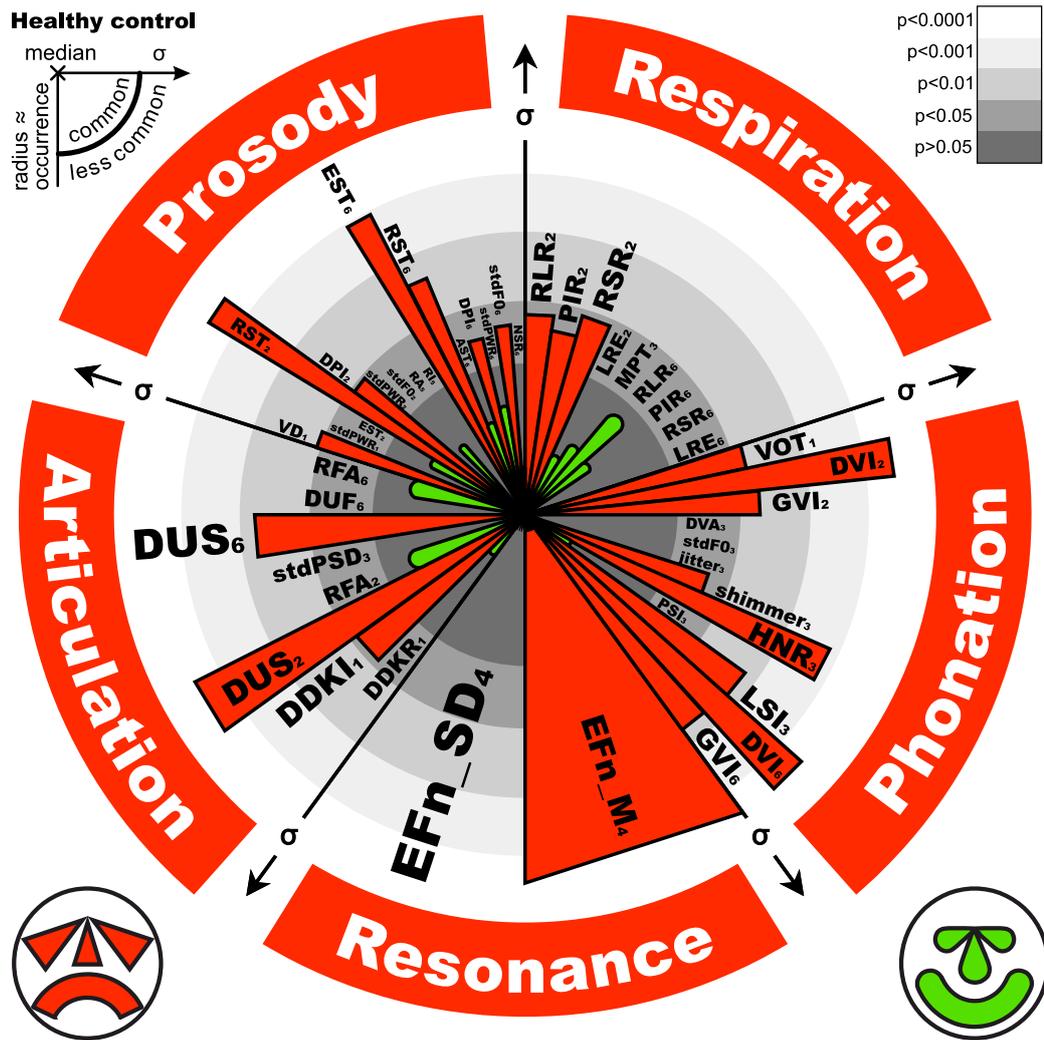


Figure 12: Illustrated results of case A measured at the first recording session.

**Abbreviations:** AST = acceleration of speech timing, DDKI = diadochokinetic irregularity, DDKR = diadochokinetic rate, DPI = duration of pause intervals, DUF = decay of unvoiced fricatives, DUS = duration of unvoiced stops, DVA = degree of vocal arrests, DVI = duration of voiced intervals, EFn\_M = degree of hypernasality, EFn\_SD = intermittent hypernasality, EST = entropy of speech timing, GVI = gapping in between voiced intervals, HNR = harmonics-to-noise ratio, LRE = latency in respiratory exchange, LSI = location of subharmonic intervals, MPT = maximum phonation time, NSR = net speech rate, PIR = pause intervals per respiration, PSI = proportion of subharmonic intervals, RA = rhythm acceleration, RI = rhythm instability, RLR = relative loudness of respiration, RFA = resonant frequency attenuation, RSR = rate of speech respiration, RST = acceleration of speech timing, stdFO = standard deviation of fundamental frequency, stdPSD = standard deviation of power spectral density, stdPWR = standard deviation of power, VD = vowel duration, VOT = voice onset time, DDK = diadochokinetic task, phonationA = sustained vowel /A/, phonationI = sustained vowel /I/, text = reading passage. Features were measured on tasks corresponding to subscripted numbers.

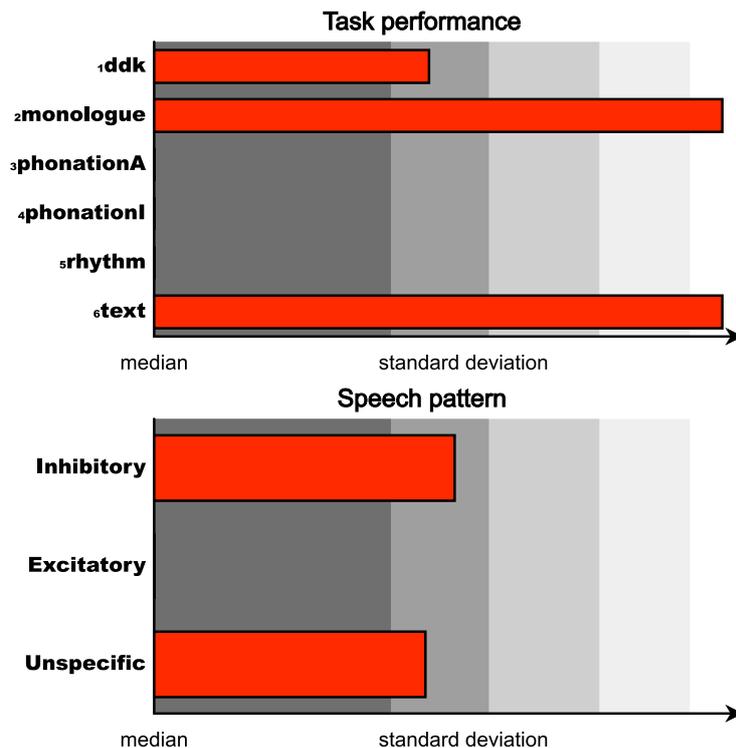
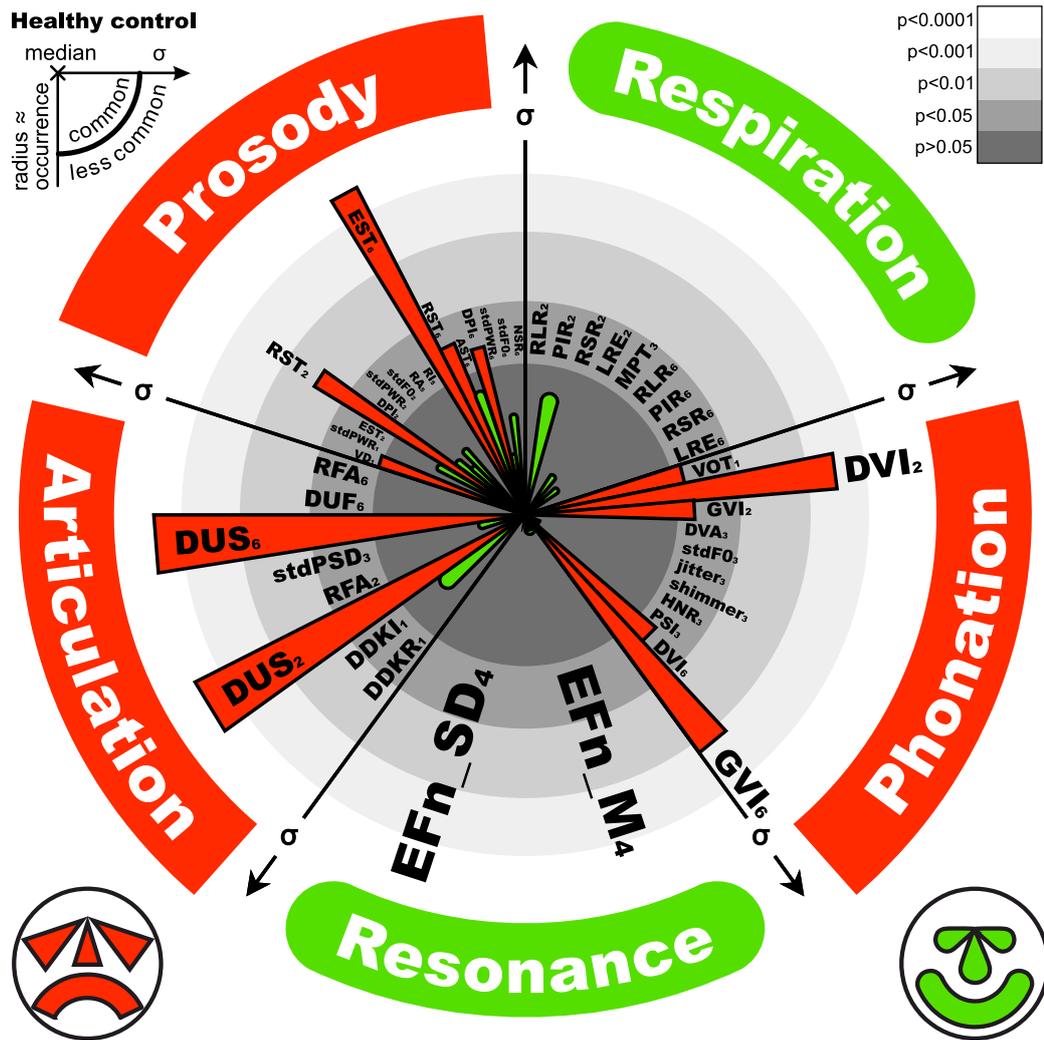


Figure 13: Illustrated results of the case A measured at the last recording session.

**Abbreviations:** AST = acceleration of speech timing, DDKI = diadochokinetic irregularity, DDKR = diadochokinetic rate, DPI = duration of pause intervals, DUF = decay of unvoiced fricatives, DUS = duration of unvoiced stops, DVA = degree of vocal arrests, DVI = duration of voiced intervals, EFn\_M = degree of hypernasality, EFn\_SD = intermittent hypernasality, EST = entropy of speech timing, GVI = gaping in between voiced intervals, HNR = harmonics-to-noise ratio, LRE = latency in respiratory exchange, LSI = location of subharmonic intervals, MPT = maximum phonation time, NSR = net speech rate, PIR = pause intervals per respiration, PSI = proportion of subharmonic intervals, RA = rhythm acceleration, RI = rhythm instability, RLR = relative loudness of respiration, RFA = resonant frequency attenuation, RSR = rate of speech respiration, RST = acceleration of speech timing, stdF0 = standard deviation of fundamental frequency, stdPSD = standard deviation of power spectral density, stdPWR = standard deviation of power, VD = vowel duration, VOT = voice onset time, DDK = diadochokinetic task, phonationA = sustained vowel /A/, phonationI = sustained vowel /I/, text = reading passage. Features were measured on tasks corresponding to subscripted numbers.

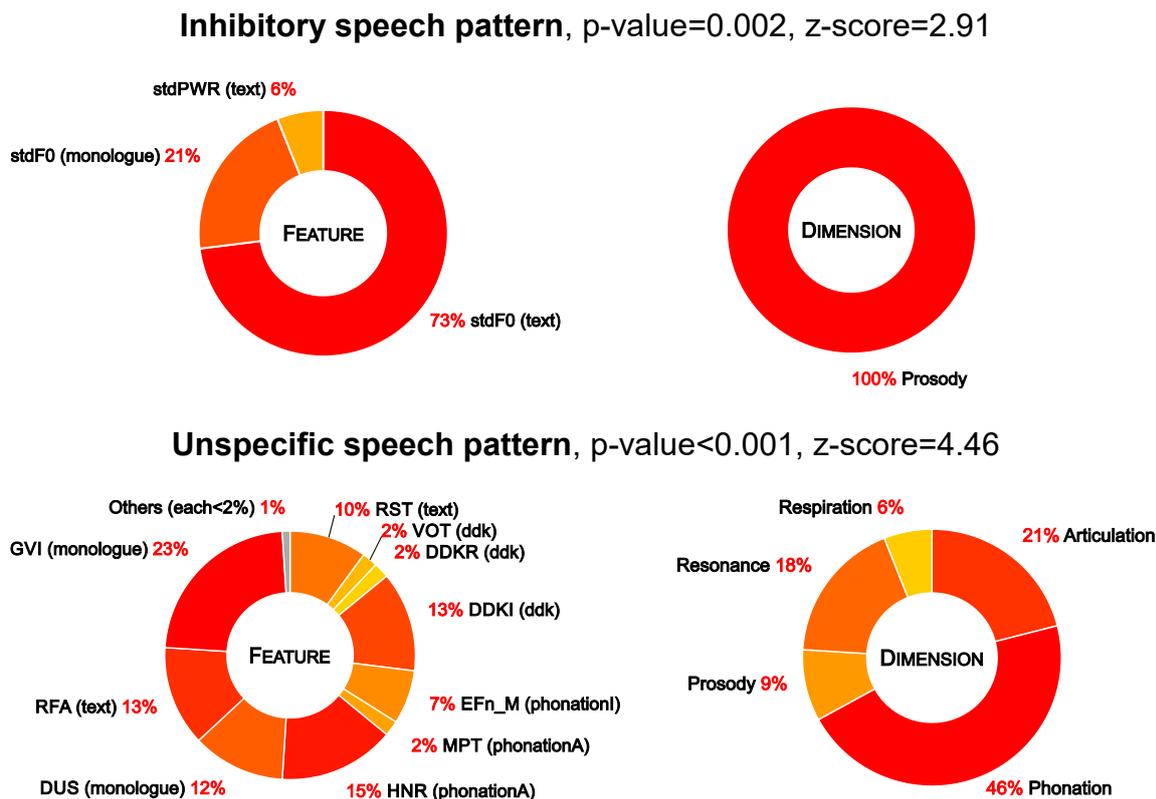


Figure 14: Speech patterns of the case A measured at the first recording session.

Percent corresponds to the contribution of the feature and dimension to the overall salience enumerated by the z-score. Contribution was estimated according to Equation 34.

**Abbreviations:** DDKI = diadochokinetic irregularity, DDKR = diadochokinetic rate, DUS = duration of unvoiced stops, EFn\_M = degree of hypernasality, GVI = gapping in between voiced intervals, HNR = harmonics-to-noise ratio, MPT = maximum phonation time, RFA = resonant frequency attenuation, RST = rate of speech timing, stdF0 = standard deviation of fundamental frequency, stdPWR = standard deviation of power, VOT = voice onset time, DDK = diadochokinetic task, phonationA = sustained vowel /A/, phonationI = sustained vowel /I/, text = reading passage.

Finally, inspirations were loud in the monologue task, probably as a result of the gathered saliva described in the speech-language-swallowing pathology diagnosis.

*Longitudinal follow-up:* A considerable improvement in speech performance was observed in individual speech features as well as overall dimensions and speech patterns. Prosody did improve in terms of loudness and melody variation. The rate of speech timing increased after the first session, but the values were still below the norm. Several prosodic aspects, such as the entropy of speech timing, duration of voiced intervals in diadochokinetic task, and duration of pauses varied over time, reflecting the erratic qualities of the speaker's performance that originate from interactions between the various dimensions and the speaker's compensation for the speech disability. Spirantization was present over all sessions, but the increase in the regularity of the performance of the diadochokinetic task indicated a subtle improvement of articulation, whereas VOT did improve only after the first session and then remained on the border of abnormality. Additionally, gradual improvement of formant resonances was observed despite the fact that the resonances were normally prominent at the time of initial examination. The velopharyngeal insufficiency disappeared completely after the first session, and the degree of hypernasality converged towards modal values. Hoarseness disappeared completely after the first recording session, suggesting the possible effect of cured coldness. Other measures related to phonation also dropped after the first session but remained abnormal with random variation. The subharmonics

arose later in consecutive sessions and were not even present in the last recording. Finally, an increased speech respiratory rate as well as abnormal loudness of respiration declined over time and reached modal values at the final recording session, whereas breath grouping during the monologue did improve after the first therapy session and remained insignificant for all following sessions.

*Conclusions:* The patient manifested significant inhibitory tendencies, weak respiration, impaired phonatory control, and an unspecific speech pattern with prominent phonatory and articulatory deficits. All dimensions of speech improved by some degree during therapy, and respiration and resonance showed significant improvements. Nevertheless, factors beyond the speech-swallowing

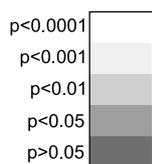
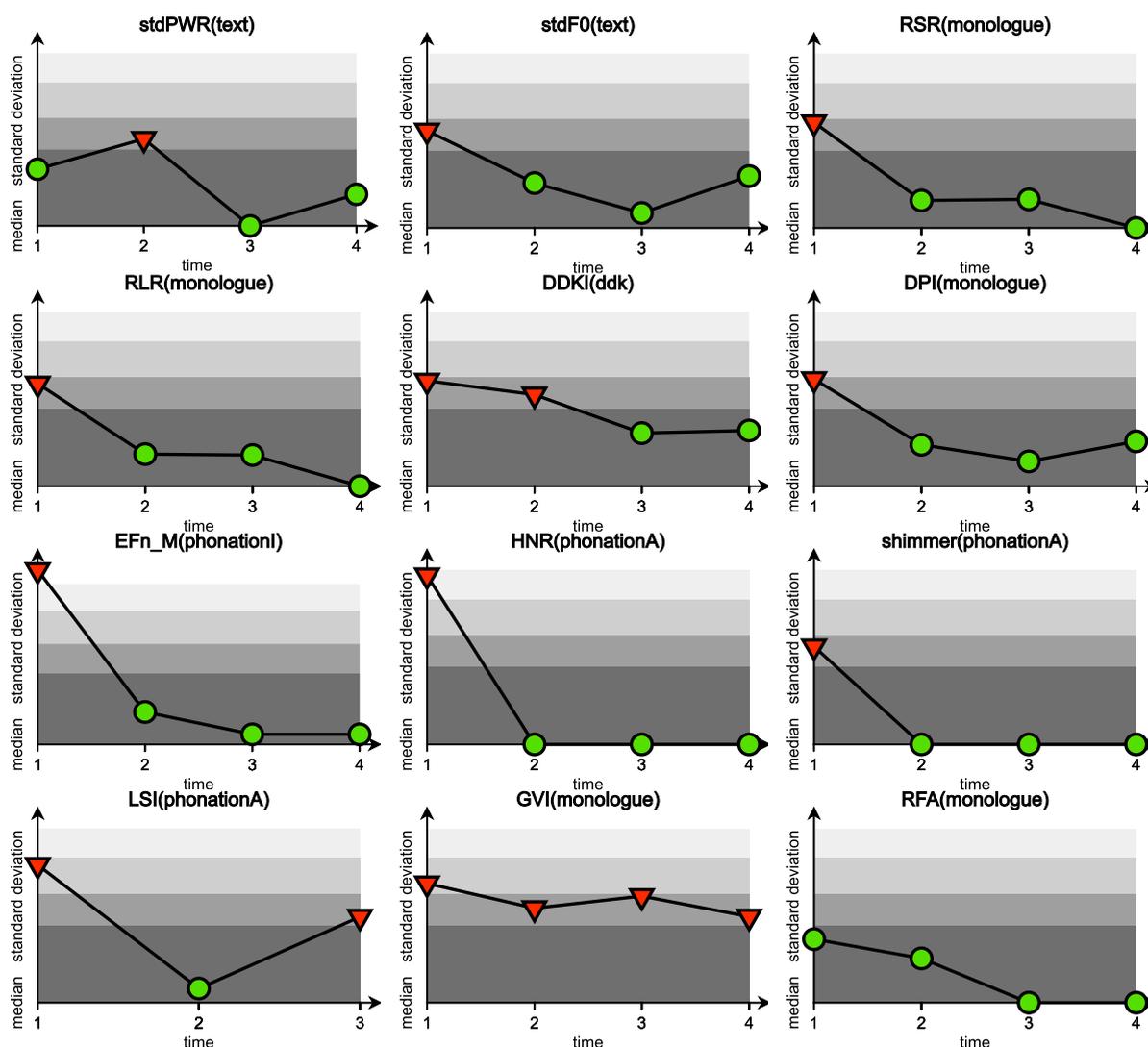


Figure 15: Longitudinal data of selected speech features measured on case A. Note that no subharmonics were present in the last recording session; thus, no value of LSI was measured.

Time stamps correspond to the order of the recording sessions.

**Abbreviations:** DDKI = diadochokinetic irregularity, DPI = duration of pause intervals, EFn\_M = degree of hypernasality, GVI = gapping in between voiced intervals, HNR = harmonics-to-noise ratio, RFA = resonant frequency attenuation, RLR = relative loudness of respiration, RSR = rate of speech respiration, stdF0 = standard deviation of fundamental frequency, stdPWR = standard deviation of power, DDK = diadochokinetic task, phonationA = sustained vowel /A/, phonationI = sustained vowel /I/, text = reading passage.

therapy, such as medication and becoming well again, could have contributed to the better performance, especially when considering the suddenness of the change between the first and second recording sessions.

### 3.5.2 Case B

*Characterization:* Male born in 1980.

#### **NEUROLOGICAL DIAGNOSIS**

*Date:* April 19, 2017.

*Diagnosis:* Huntington's disease, moderate stage, genetically verified in June 2017, generalized chorea, behavioral disorder, attention-deficit disorder, anxiety.

*Pharmacological anamnesis:* Argofax, Xanax, Risperidon, Valprocit.

#### **SPEECH-LANGUAGE-SWALLOWING PATHOLOGY DIAGNOSIS**

*Date:* July 19, 2017.

*Diagnosis:* Patient communicates verbally with borderline fluency. Information value is limited—patient is not capable of describing his anamnesis (help of his mother was required). Patient is able to react to verbal impetus appropriately; complies with roles in dialogue and holds weak eye contact. Spoken and written language is preserved. Level of speech disorder corresponds to the communication-cognitive deficits associated with the neurological diagnosis. Examination was focused on motor function of speech and swallowing. Progress of the therapy was slow due to the deteriorated cognition and memory problems—patient requires written instructions for training at home. Fast complex training was ineffective.

*Subjective:* The patient is aware of difficulties with intonation of melody and speech tempo—both vary involuntarily. Patient has problems characterizing and describing events. Patient chokes during drinking and eating—food falls out of patient's mouth frequently; eating takes longer than previously (circa 15-20 minutes). Patient is able to eat whole portion. Gathering of saliva with occasional drooling do not wake him up at night—patient negates saliva leaks. Occasional pyrosis. Coordination of movements is worsened. Patient feels stressed frequently, or on the contrary, lethargic or depressive. Weight loss is not apparent (height 182 cm, weight 75 kg). Patient negates nasal penetration, odynophagia, tightness in the throat during swallowing, and regurgitation.

*Facial movements:* Face is symmetrical. No grimaces were present at the time of examination. The patient is involuntarily chattering teeth. Keeping lips closed is not difficult for the patient. Protrusion and grinning is symmetrical. Patient suffers from dysdiadochokinesia, difficulties in lateral movements—lingual apraxia. Mandibular movements are preserved in elevation and depression, protrusion and lateral movements only partially. Rotation movements are strongly affected. Tonus of muscles of mastication is sufficient. The patient manifests buccal apraxia. Tongue is white and shows no atrophy, no scars from cheek biting, and normal protrusion. Patient cannot straighten the protruding tongue. The patient compensates the involuntary retraction of the tongue by rotating the tongue down to chin. Isolated elevation of the tip, upper side, and root of the tongue is unaffected. Lateral movements and coordination of repetitive movements are deteriorated. Tonus of the tongue is appropriate. No involuntary movements were observed in

relaxed tongue. Soft palate is symmetrical in rest and in elevation. Palatal and pharyngeal reflex were present.

*Phono-respiration:* Respiration at rest is regular. No dysrhythmia was observed. Maximal phonation time is shortened to 3-4 seconds. Excess pitch variation (not a tremolo) was present. Phonation is not affected by hoarseness. Breath groups were shortened in connected speech. The patient manifests no abnormal nasality.

*Phonetics:* Articulation is slurred—deteriorated coordination of articulation in connected speech; intact only isolated (except thrill consonants). Tempo is variable. No dysfluencies are present. The patient manifests abnormal pitch variation in connected speech.

*Deglutition:* Head posture and body posture are voluntarily controlled. Self-reliance during eating is lowered. Fine motor control is deteriorated—cutting food is problematic. Consequently, patient prefer big mouthful. Calorie intake is adequate. No thickening agents or sipping are required.

*Salivation:* No gathering of saliva was observed during examination. Mucosa is sufficiently wet without stickiness (subject mentioned occasional gathering of saliva). Voluntary deglutition of saliva is problematic—swallowing apraxia.

*Liquids:* Patient drinks in single shots. Head goes from hyperextension to extension after gathering bolus. Swallowing reflex is initiated with latency. Elevation of larynx is sufficient. No perceptible changes in phonation or reflexive cough were present after deglutition. Deteriorated coordination of swallowing reflex is assumed. Leaking is suspected. No tachyphagia or aerophagia were observed.

*Volume test:* 30 ml / 15 s—corresponds to normative data for males.

*Speed test:* 100 ml / 15 s—slower than normative values of 100 ml / 10 s in consequence to drinking continually, which was too risky for the patient (100 ml in three shots using 6 deglutition). Deglutition of solid foods was not examined.

*Conclusion:* R47.1 hyperkinetic dysarthria, R48.2 oral apraxia and swallowing apraxia, R13 mild dysphagia in all phases of deglutition.

*Recommendations:* The patient should undergo speech motor training of articulators and phono-respiration, strengthening laryngeal and pharyngeal muscles, guidance about optimal eating regime, and training cognitive functions.

*Videofluoroscopic examination of degustation (November 28, 2017):* Swallowing was intermittently discoordinated during drinking normal bolus of fluids (Rosenbek 7). Patient then aspirates below the level of glottal folds. Contrast agent remains in airways despite reflexive cough. Deglutition of thickened and solid foods (Rosenbek 10) was normal.

## **THERAPY OF SPEECH AND SWALLOWING**

The therapy will focus on the preservation of swallowing functions. It is important to identify the breakpoint when percutaneous endoscopic gastrostomy is required and to gently prepare and inform the patient about the situation. Caloric intake must be checked regularly to avoid malnutrition. Motor control and coordination of respiration during deglutition (and intensity of reflexive cough) is the primal goal. Articulation and prosody are unachievable targets with regard to limited time of sessions (approximately 30 minutes) and patient's cognitive deficits (aboulia).

*Therapy session on September 31, 2017:* The patient was hospitalized at psychiatry up to date of the session. The patient did learn all exercises according to the photographic guide focused on training lip closure that was provided on July 17, 2017. A new exercise for the mastication muscles was added to the training.

*Therapy session on November 28, 2017:* Speech motor training of lips, jaw, and tongue was conducted. Results of Videofluoroscopic examination were explained to the patient. The patient was informed about the regime and compensatory actions for safe swallowing. The nutrition specialist revised the dietary regime of the patient in order to prevent weight loss.

*Therapy session on January 9, 2018:* Speech motor training conducted in previous session was extended by training of soft palate. Phono-respiration was exercised. Air stacking was also performed, but it proved to be too demanding on coordination and the nasal emission of air. Patient started visiting Ergoactive–communication group and ergotherapy (fine motor control–handcraft). Visiting Ergoactive was recommended because the group meetings are more frequent than the therapy sessions covered by health insurance.

*Therapy session on February 13, 2018:* Speech motor training was conducted. Phono-respiration was exercised. Air stacking was performed–difficulties in coordination. Patient is capable of performing only simplified exercises without nasal emissions–improved resonance.

*Therapy session on March 6, 2018:* Speech motor training was conducted. Phono-respiration was exercised, particularly the coordination between phonation and respiration. Phonation was trained. The patient trained humming–optimization of vocal register and continuity of phonation.

*Therapy session on April 3, 2018:* Exercising was similar to previous session. Mendelsons maneuver was trained unsuccessfully (poor coordination and problems with cognition)–this exercise is definitely unsuitable. Voluntary coordination of respiration and deglutition was trained successfully.

*Therapy session on May 30, 2018:* Speech motor training was conducted. Phono-respiration was exercised, particularly the coordination between phonation and respiration. Vocal register and continuity of phonation was trained. Swallowing control was exercised–spirometer shows no abnormality. Disease progression is noticeable compared to conditions 2 months ago in discoordination of respiration and increased chorea (tongue, especially).

## ACOUSTIC ANALYSIS

Recordings were gathered on July 26, 2017 (for educational purposes) and during therapy sessions on March 6 and May 30, 2018. The acoustic signals from reading a passage, and the rhythm, diadochokinetic, and sustained vowels tasks were recorded and processed by the techniques described in 2 METHOD, page 17. The results and visualizations generated by the fully automated approach are summarized in Table 10, Figure 16, Figure 17, Figure 18, and Figure 19.

*Prosody:* The net speech rate was significantly reduced. The flow of connected speech was frequently interrupted by long pauses, which can be associated not only with difficulties in initiating speech and the omission of short pauses but also with the cognitive deficits mentioned in the speech-language-swallowing pathology diagnosis.<sup>2</sup> The flow of voiced, unvoiced, and pause

---

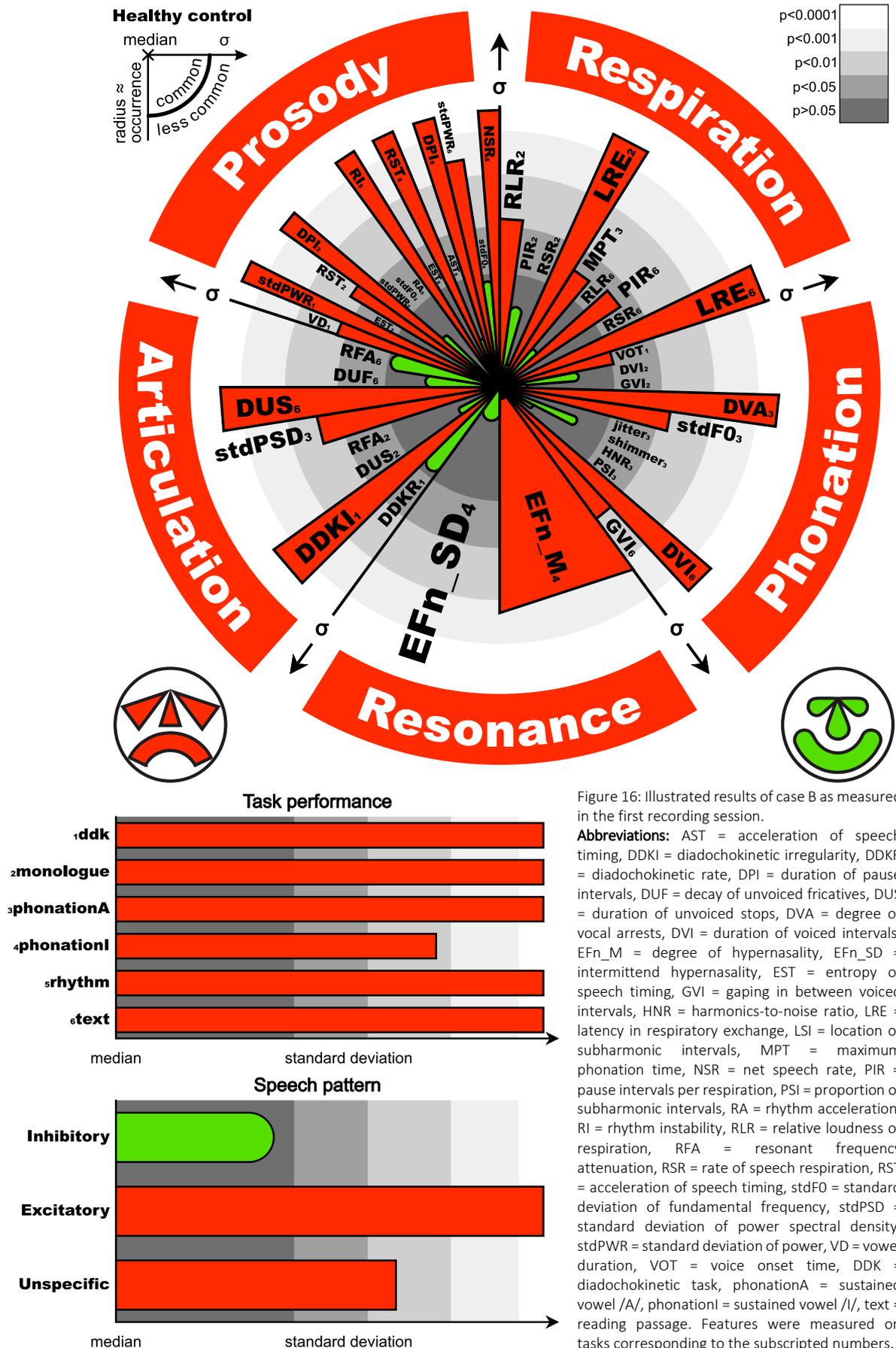
<sup>2</sup> The automated report does not include cognitive deficits on the list of possible causes by default because this touchy issue can be implied only with a caution regarding the true cognitive abilities observed by an examiner.

| Symbol | Task                 | Value | P-value | Z-score | Description                                       | Interpretation  |
|--------|----------------------|-------|---------|---------|---|---|
| stdPWR | Diadochokinetic task | 17.8  | 0       | 29.8    | Standard deviation of speech loudness (dB)        | Excess loudness variation due to involuntary movements of respiratory muscles or discoordinated phonorespiration.   |
| LRE    | Monologue            | 661   | 0       | 10.8    | Latency of respiratory exchange (ms)              | Prolonged pause between expiration and inspiration. Decreased ability to reverse from expiration to inspiration, especially difficulties in initiating inspiration. |
| DPI    | Reading passage      | 397   | 0       | 20.6    | Duration of pause intervals (ms)                  | Difficulties in initiating speech and/or omission of short pauses.  |
| NSR    | Reading passage      | 0.505 | 0.0001  | -8.2    | Net syllable rate (syllable/s)                    | Decreased syllable rate.  |
| DVI    | Reading passage      | 430   | 0.0001  | 7.95    | Duration of voiced intervals (ms)                 | Voicing interferes or continues within voiceless intervals. Decreased control of laryngeal muscles and coordination of laryngeal and supra-laryngeal muscles.       |
| LRE    | Reading passage      | 337   | 0.0001  | 7.59    | Latency of respiratory exchange (ms)              | Decreased ability to reverse from expiration to inspiration, especially difficulties in initiating inspiration.   |
| DDKI   | Diadochokinetic task | 258   | 0.0001  | 5.62    | Diadochokinetic irregularity (ms)                 | Pace of alternating motion is more irregular due to involuntary movements of speech apparatus or impaired timing.   |
| DUS    | Reading passage      | 49.4  | 0.0001  | 5.56    | Duration of unvoiced stops (ms)                   | Imperfect articulation of unvoiced stops. Unvoiced stops are prolonged or, for more extreme values, spirantized.  |
| RI     | Rhythm               | 28.1  | 0.0001  | 5.34    | Rhythm instability (%)                            | Irregular pace due to decreased control over speech apparatus or presence of involuntary movements.   |
| RST    | Reading passage      | 186   | 0.0001  | -4.95   | Rate of speech timing (intervals/s)               | Reduced stream of voiced, unvoiced, and pause intervals. Typically in consequence to reduced range of speech movements and/or decreased syllabic rate.              |
| DVA    | Sustained vowel /A/  | 4.31  | 0.0001  | 4.48    | Degree of vocal arrest (%)                        | Voicing stops suddenly due to impaired control over laryngeal muscles.  |
| DPI    | Monologue            | 329   | 0.0001  | 3.89    | Duration of pause intervals (ms)                  | Difficulties in initiating speech and/or omission of short pauses.  |
| stdPWR | Reading passage      | 6.69  | 0.0005  | 3.53    | Standard deviation of speech loudness (dB)        | Excess loudness variation.  |
| EFn_M  | Sustained vowel /I/  | -31.6 | 0.0006  | 3.27    | Hypernasality mean (dB)                           | Increased hypernasality due typically to impaired control over elevator muscle of the soft palate.  |
| stdPSD | Sustained vowel /A/  | 2.97  | 0.0039  | 2.66    | Standard deviation of power spectral density (dB) | Involuntary movements of articulators, preeminently the tongue.   |
| RST    | Monologue            | 254   | 0.0061  | -2.51   | Rate of speech timing (intervals/s)               | Reduced stream of voiced, unvoiced, and pause intervals. Typically in consequence to reduced range of speech movements and/or decreased syllabic rate.              |
| stdFO  | Sustained vowel /A/  | 0.529 | 0.0065  | 2.48    | Standard deviation of F0 (semitones)              | Excess variation of fundamental frequency due to involuntary movements of laryngeal muscles. Perceptual feature is called excess pitch variation.                   |
| VD     | Diadochokinetic task | 60.5  | 0.007   | 2.46    | Vowel duration (ms)                               | Slow movements and excessive vocal emphasis manifested by abnormally prolonged vowels.  |
| RLR    | Monologue            | -32.2 | 0.0073  | -2.68   | Relative loudness of respiration (dB)             | Decreased inspiratory effort.   |
| GVI    | Reading passage      | 22.3  | 0.0091  | -2.36   | Gaping in-between voiced intervals (pause/min)    | Decreased ability of vocal folds to stop voicing by adduction.  |
| PIR    | Reading passage      | 2.5   | 0.0185  | -2.09   | Pause intervals per respiration (-)               | Decreased pausing within breath groups. Decreased ability to control respiratory airflow.   |
| MPT    | Sustained vowel /A/  | 7.21  | 0.0227  | -2      | Maximum phonation time (s)                        | Weakened control over respiratory and/or laryngeal musculature.   |
| VOT    | Diadochokinetic task | 23.6  | 0.0472  | 1.67    | Voice Onset Time (ms)                             | Disrupted coordination of laryngeal and supralaryngeal muscles. Decreased ability of laryngeal muscles to initiate voicing.   |
| RFA    | Reading passage      | 8.69  | 0.0531  | -1.62   | Resonant frequency attenuation (dB)               | N/A   |
| DDKR   | Diadochokinetic task | 5.83  | 0.0585  | -1.57   | Diadochokinetic rate (syllables/s)                | N/A   |

Table 10: Summary of the most severe speech features of case B measured in the first recording session.

Findings were sorted by ascending p-value. The table represents the illustrative capture of the automated report. Only significant results and two insignificant features were included to illustrate the design of the automated summary. Tasks were renamed according to the notation used in the thesis. The reported interpretation was assigned automatically following the simplified definitions derived in Table 6 and Table 7.

**Abbreviations:** N/A = not available—marking insignificant results.



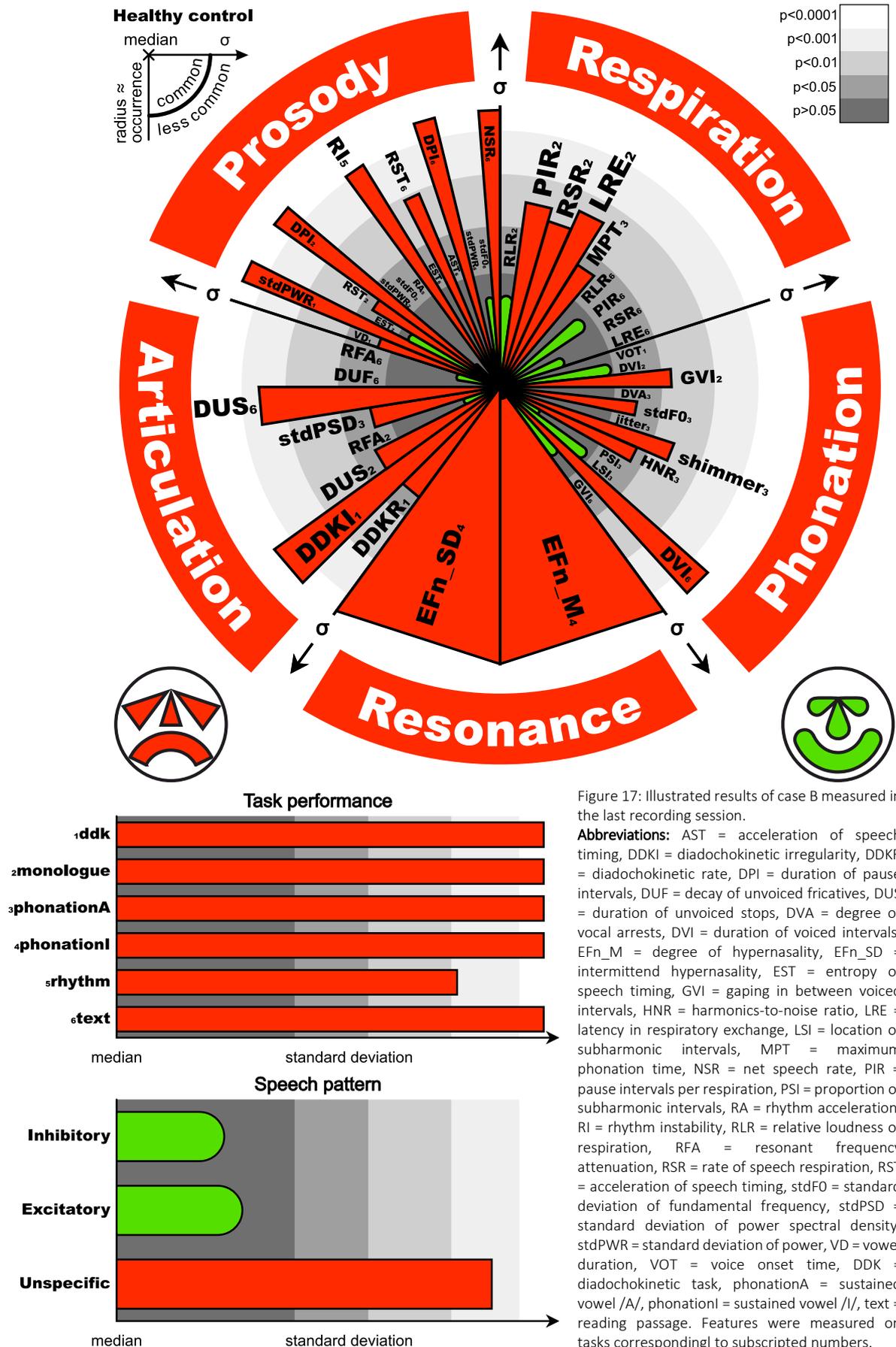


Figure 17: Illustrated results of case B measured in the last recording session.

**Abbreviations:** AST = acceleration of speech timing, DDKI = diadochokinetic irregularity, DDKR = diadochokinetic rate, DPI = duration of pause intervals, DUF = decay of unvoiced fricatives, DUS = duration of unvoiced stops, DVA = degree of vocal arrests, DVI = duration of voiced intervals, EFn\_M = degree of hypernasality, EFn\_SD = intermittent hypernasality, EST = entropy of speech timing, GVI = gapping in between voiced intervals, HNR = harmonics-to-noise ratio, LRE = latency in respiratory exchange, LSI = location of subharmonic intervals, MPT = maximum phonation time, NSR = net speech rate, PIR = pause intervals per respiration, PSI = proportion of subharmonic intervals, RA = rhythm acceleration, RI = rhythm instability, RLR = relative loudness of respiration, RFA = resonant frequency attenuation, RSR = rate of speech respiration, RST = acceleration of speech timing, stdFO = standard deviation of fundamental frequency, stdPSD = standard deviation of power spectral density, stdPWR = standard deviation of power, VD = vowel duration, VOT = voice onset time, DDK = diadochokinetic task, phonationA = sustained vowel /A/, phonationI = sustained vowel /I/, text = reading passage. Features were measured on tasks corresponding to subscripted numbers.

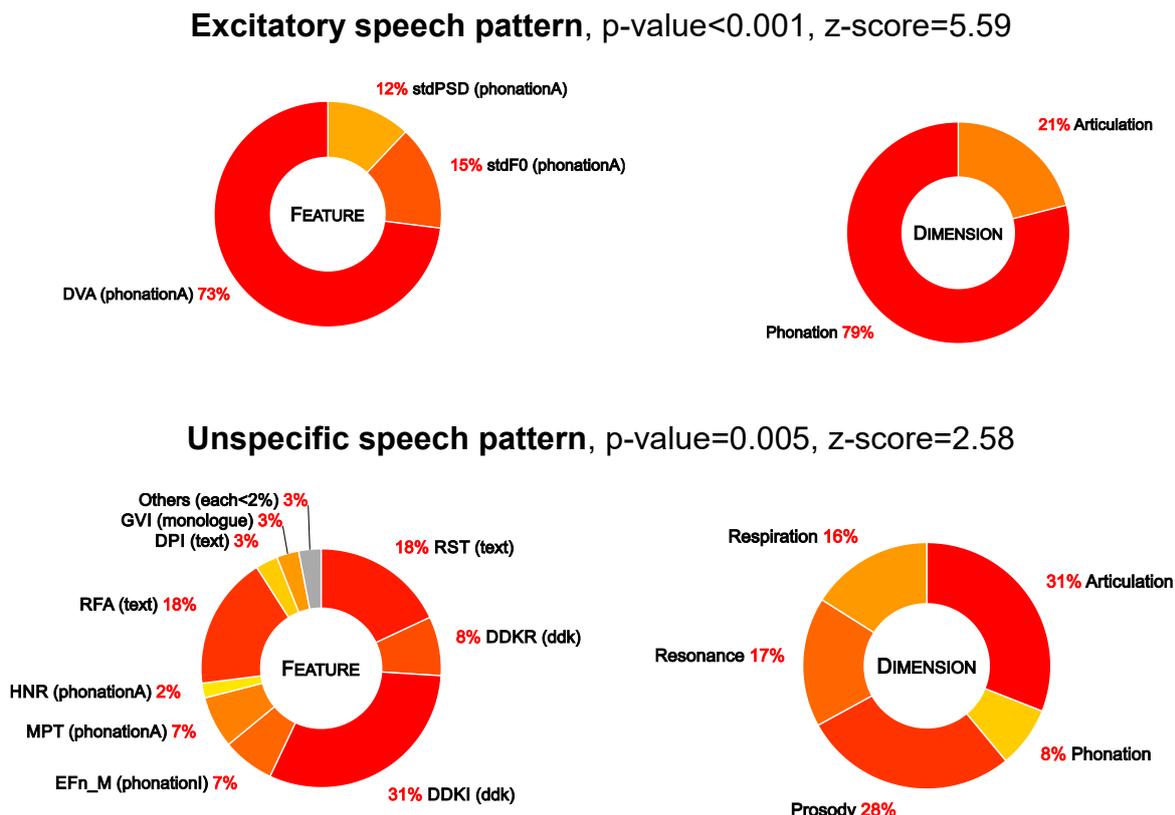


Figure 18: Speech patterns for case B measured in the first recording session.

Percent corresponds to the contribution of the feature or dimension to the overall salience enumerated by the z-score. Contribution was estimated according to Equation 34.

**Abbreviations:** DDKI = diadochokinetic irregularity, DDKR diadochokinetic rate, DPI = duration of pause intervals, DVA = degree of vocal arrests, EFn\_M = degree of hypernasality, GVI = gapping in between voiced intervals, HNR = harmonics-to-noise ratio, MPT = maximum phonation time, RFA = resonant frequency attenuation, RST = acceleration of speech timing, stdF0 = standard deviation of fundamental frequency, stdPSD = standard deviation of power spectral density, stdPWR = standard deviation of power, DDK = diadochokinetic task, phonationA = sustained vowel /A/, phonationI = sustained vowel /I/, text = reading passage. RST = acceleration of speech timing,

intervals was significantly reduced, which reflects the overall slowness of the speech rate rather than the limited range of movements since the z-score of the NSR was almost doubled compared to the RST. The loudness variation was significantly increased during the reading passage as well as in diadochokinetic task, indicating involuntary movements of the respiratory muscles. Involuntary movements that were also present in other speech dimensions are the most probable cause of the increased irregularity of the rhythm task and the prolongation of vowels in the diadochokinetic task.

*Articulation:* Spirantization of unvoiced stops was observed while reading the passage, which suggests deteriorated control over the fine movements of articulators. Resonances were less prominent, almost reaching the level of significance. Similarly, the diadochokinetic rate was rather slow but insignificantly so. The articulatory disability manifested strongly in the irregularity of the diadochokinetic task. The repetitive articulatory rate was most likely modulated by the involuntary movements of articulators that were clearly present in the steady articulation of the sustained vowel /A/.

*Resonance:* The degree of hypernasality was significantly increased in the sustained vowel /I/. Velopharyngeal insufficiency was not intermittent at the time of the initial recording session.

*Phonation:* In addition to deteriorated control over adduction of the vocal folds manifested in the prolongation of voiced intervals and reduced pausing of voiced intervals in the reading passage

and increased VOT in the diadochokinetic task, phonation was heavily impaired by the involuntary movements of the laryngeal muscles, causing vocal arrests and excess melody variation in the steady phonation of the vowel /A/.

*Respiration:* The maximum phonation time, together with decreased pausing in breath groups, indicates low inspiratory capacity and impaired control over phono-respiration. Significant prolongation of the interval between expiration and respiration in both tasks of connected speech ( $z$ -scores $>7$ ) suggests discoordination of the respiratory muscles as well as difficulties in initiating phonation in terms of the decreased inspiratory effort during the reading passage.

*Longitudinal follow-up:* Improvement of the net speech rate was observed in the course of therapy. Accordingly, flow of voiced, unvoiced, and pause intervals increased in reading the passage. A better performance in the last session compared to the first one was observed also in the

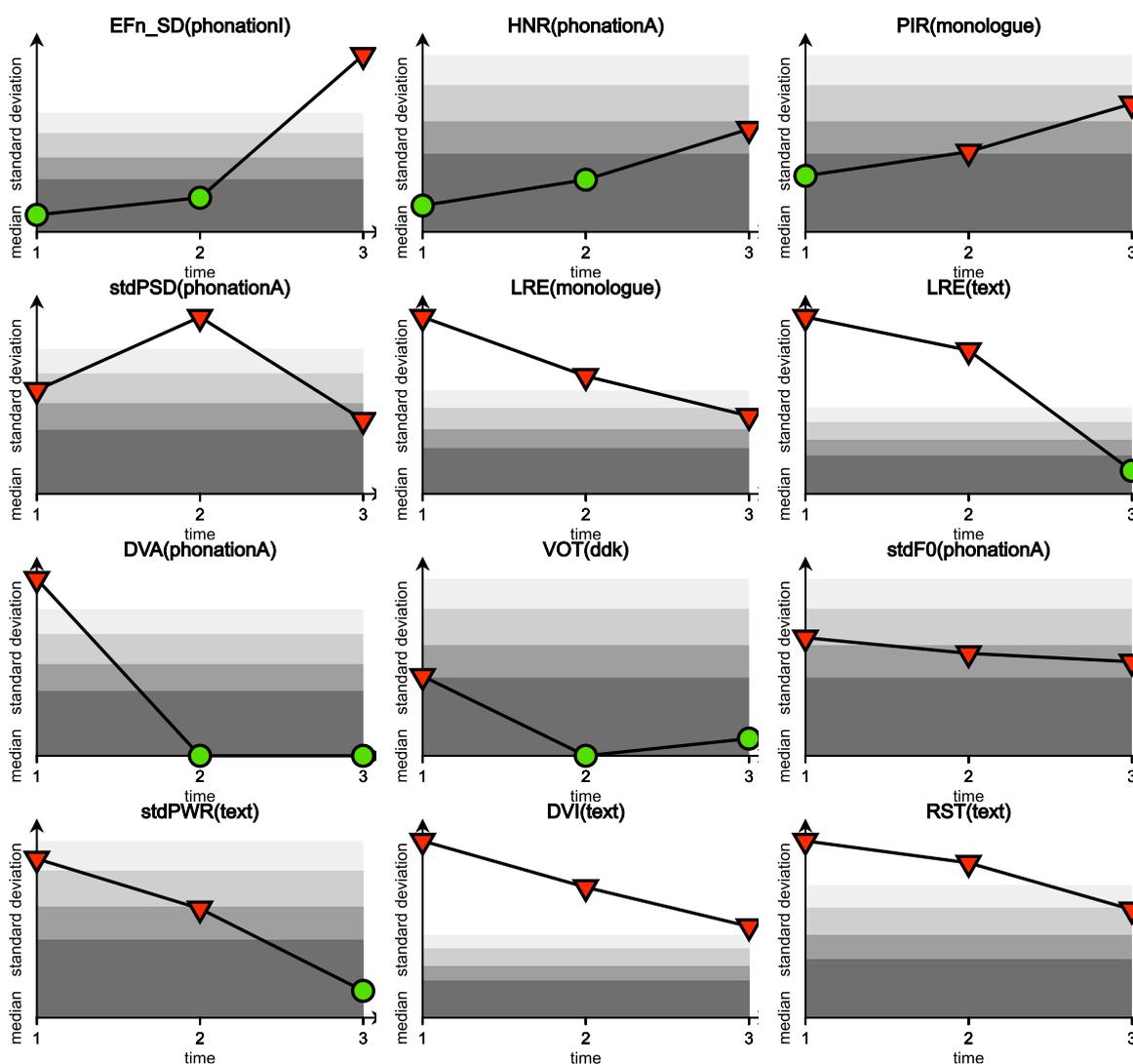


Figure 19: Longitudinal data of selected speech features measured on case B. Note that no subharmonics were present in the last recording session; thus, no value of LSI was measured. Time stamps correspond to the order of the recording sessions. **Abbreviations:** DVA = degree of vocal arrests, DVI = duration of voiced intervals, EFn\_SD = intermittent hypernasality, HNR = harmonics-to-noise ratio, LRE = latency in respiratory exchange, PIR = pause intervals per respiration, RST = acceleration of speech timing, stdPSD = standard deviation of power spectral density, stdPWR = standard deviation of power, DDK = diadochokinetic task, phonationA = sustained vowel /A/, phonationI = sustained vowel /I/, text = reading passage.

monologue, but the trend showed great deviation in the second session. Interestingly, the variation in loudness decreased over the course of therapy and dropped below the level of significance in the last session. The rhythm task also became more regular. Although the therapy was not focused on prosody itself, the motor speech exercises and training of phono-respiration affected the speech rhythm as well as the loudness variation positively. The only aspect of prosodic that worsened considerably over time was the duration of the pause, possibly reflecting the progression of the disease. Other prosodic features showed either no change or fluctuating quality. Involuntary movements and impaired control over articulators affected articulation over all sessions with varying degrees. Intermittent hypernasality was observed with increasing intensity over time, which raises the question of whether the increased degree of hypernasality observed in the first and last sessions—contradicting perceptual findings—was a consequence of the waxing and waning character of dystonia. The patient established better control over phonation after the first recording session, completely eliminating vocal arrests and stabilizing the pitch in the sustained vowel /A/. The VOT dropped suddenly to values in the normal range after the first recording session. Pausing between voiced intervals and shortening voiced intervals in the reading passage suggests the improvement of phonation as well, but the performance was not consistent with the findings from connected speech. Hoarseness did increase gradually in a clear trend and exceeded the level of significance in the last recording session. An advance in respiration control, possibly resulting from therapy, did manifest notably in the decreased latency between expiration and inspiration trending in both tasks of connected speech. A less unambiguous trend was observed in the increased inspiratory effort in the monologue. The rate of respiration and pausing in breath groups worsened only in the monologue, whereas breath groups in the reading passage improved mildly in the last session, which could be attributed to a random variation in performance. The maximum phonation time did fluctuate around the level of significance.

# 4

## DISCUSSION

*And while it would be absurd to suggest that Huntington's disease made Woody Guthrie a brilliant songwriter, Dr. Whittier (and, later, Marjorie Guthrie herself) would wonder aloud if the disease hadn't worked like a drug on Woody, as a creative spur (in much the same that some artists use alcohol and other drugs), enhancing his natural rhythmicity, forcing the brain to continually rewire itself as cells died, forcing new, wonderful, and unexpected synaptic pathways to open (which also led to some unexpected and not so wonderful behavior), forcing the brain to become—in effect—more creative to survive; and then, after a point, exhausted and starving for energy, the synapses and ganglia short-circuiting ... preventing him from concentrating on anything, making him fidgety, antsy, causing him to lose perspective and, eventually, his creative sense of himself.*

—Joe Klein, *Woody Guthrie: A Life*, 1980

Complex automated acoustic assessment of dysarthria comparable in comprehensiveness to this thesis has not been presented before. Currently, the impact of neurodegeneration on speech motor control is known only by perceptual measures, which are commonly considered to be subjective. The link between objectivity of speech assessment and acoustic measures—recognized despite the limited interchangeability of different recording systems and processing methods—is de facto a chimaera, since speech performance itself manifests a high degree of randomness. Of course, any acoustic feature can be measured objectively on one particular signal when recorded and analysed by a clearly defined process—and its objectivity can be indisputable—, but the performance of the speaker may vary by age, sexual dimorphism, social background, and other unpredictable factors, such as momentary emotions. In this situation, what matters the most for interpretation is not the exact value of the feature but the enumerated uncertainty of abnormality. This chapter discusses the methodology founded upon this idea from the perspective of an individual acoustic analysis, findings, and speech patterns; concretizes scenarios in which the methodology could be applicable, and suggests directions for further development.

## 4.1 ACOUSTIC ANALYSIS

### 4.1.1 Connected speech

#### SEGMENTATION

The method used for the segmentation of connected speech into the categories of voiced, unvoiced, pause, and respiratory intervals was based on unsupervised learning. The rationale behind the design was to adapt the decision to a variable level of non-speech noise and to make segmentation more versatile for the processing of recordings with environmental noise. Note that only clean recordings were selected for the database, but increased noise can be expected in any real recording situation. The proposed methodology showed superior performance in the detection of pause intervals in comparison with other available methods and sufficient performance in detection of respiratory intervals. Pauses shorter than 100 milliseconds were deemed erroneous due to the increased levels of turbulent noise that preserved the spectral characteristics of the previous phoneme. Classification accuracy was decreased considerably in the APS and HD groups (c.f. Hlavnička et al. 2017) due to an increased level of non-speech noise produced by the speakers. The most problematic source of errors in pauses longer than 100 milliseconds was loud respirations, which may have shown energy levels and a spectral envelope comparable to unvoiced fricatives. The classification of respirations as unvoiced fricatives and vice versa were the most prominent types of errors in the detection of respirations and pauses longer than 100 milliseconds. Unpublished experiments with a filtered signal showed that the methodology could adapt to various spectral characteristics to a certain degree. Recognition seems to be sensitive to high-frequency bands, which can be exploited for improving detection accuracy.

An analysis of errors suggests that unsupervised learning can be a very effective solution for segmentation of connected speech affected by dysarthria. Supervised learning would require a much larger database with speakers along the severity spectrum of a speech disorder because non-speech noises are more frequent in more severe speech disorder. In addition, the manual labeling of all types of intervals in a large database can be too time consuming for supervised training. Finally, the adaptive method presented in this thesis can be applied to data gathered by various recording systems without any additional adjustments, unlike in supervised learning, which would require device-independent descriptors or a database recorded across various recording systems. Both adaptation of unsupervised learning as well as a clear definition of categories by supervised learning is desirable; thus, a hypothetical combination could be the ultimate solution.

#### SPEECH FEATURES

The battery of speech features related to prosody, articulation, phonation, and respiration demonstrated the importance of the task for the examination of dysarthria. Monoloudness, monopitch, the decreased rate of gaping in voiced speech, articulatory imperfections, prolonged pauses, and a reduced stream of voiced, unvoiced, and pause intervals were identified as general acoustic manifestations of hypokinetic dysarthria in connected speech. The acceleration of speech was not significant in the PD sample, which, however, does not negate the possibility of individual incidences, since the symptom may not be present in all patients (Duffy 2013). Gradual weakening of friction in unvoiced fricatives is rather an individual characteristic that may not be expected for a significant portion of PD patients and, as in the case of the acoustic measure of speech acceleration, its relation to PD is only hypothetical.

Unsurprisingly, the majority of features showed general trends towards inhibition across the various groups. Connected speech is a challenging task that requires the extensive involvement of all subsystems of speech. Thus, inhibition as a common manifestation of dysarthria is not only a natural response but also a frequent compensatory mechanism in a speech disability. Based on the results of the cross-sectional design of the thesis, inhibitory speech manifestations may develop early in the course of PD, which makes them a valuable indicator of speech motor status. It seems that the early presence of inhibition unrelated to compensation for a speech disability makes these speech features uniquely parkinsonian, which, in the context of other symptoms, such as idiopathic REM sleep behaviour disorder, olfaction, and face akinesia, could support early recognition of high-risk individuals (Postuma et al. **2009**, Postuma et al. **2012**).

Excitatory tendencies, such as excess loudness variations in HD and SCA and increased loudness of respiration in HD, suggest that connected speech is probably the only task that is capable of exposing both inhibitory and excitatory trends within a single aspect of speech, e.g., significantly increased stdPWR in HD and SCA versus significantly decreased stdPWR in PD.

It is of significance to note that trends were not comparably distributed between tasks. For illustration, the majority of groups manifested monopitch in reading the passage, but only subjects with Parkinsonism demonstrated monopitch in the monologue. The reading passage is advantageous, as the standardized content reduces the variance of measured values and thus increases the sensitivity of the feature. Contrarily, the monologue provides more freedom in speech expression as well as breath economy. Finally, reading the passage and performing the monologue are noninterchangeable tasks with immeasurable importance for the analysis of all groups, with special consideration given to PD.

## 4.1.2 Rhythm

### SEGMENTATION

The proposed segmentation of the rhythm task represents another example of an algorithm that achieved very high accuracy via unsupervised learning. Unsupervised learning executed inside a sliding recognition window, together with an additional analysis of the extremities, seems to be advantageous for overcoming the unpredictable manifestations of dysarthric speech for more than just the rhythm task. An additional analysis of the relations between segmentation accuracy and values of speech features is outside the scope of this thesis, but, generally, one can imagine that the results for speech features may be substantially skewed when different parts of the signal are measured or, even more likely, if the key data are positions of segments, as in the rhythm task. Although the proposed segmentation proved to be reliable, we can never expect an error-free performance for all signals in any task. Therefore, all features of all tasks in the thesis were designed to prevent the possible influence of misdetections. In the rhythm task, a regression analysis increased the robustness against failures of segmentation; in other tasks, the preference for nonparametric descriptors, such as the median and median absolute deviation, increased the robustness. It is noteworthy that these refinements may also decrease the sensitivity to short-term extremities in a speaker's performance, thus, they were not applied invariably to all features. However, the current rhythm metrics from Skodda et al. (**2010**) were overly influenced by the quality of small sequences of syllables, although irregularities and accelerations developed through the course of whole phonations. In other words, the proposed approach prevents the occasional influence of other subsystems of speech, such as respiration or articulation, on the overall rhythm

metrics. Here, an insightful reader may notice that even these small refinements in the computation of speech features are oriented towards the decomposition of the speech processes, which was introduced as the leading principle of this thesis.

### **SPEECH FEATURES**

The acceleration of speech showed no significant trend in PD, which contradicts a previous study by Rusz et al. (2015A). The database in the thesis represents an extended sample of speakers analyzed by Rusz et al. (2015A), excluding the Manganese-induced Parkinsonism caused by abuse of methcathinone (Ephedrone). The discrepancy can be accounted for by a different sample rather than the normalization procedure, as the normalized values examined for the subsample corresponding to the original study (Rusz et al. 2015A) showed a significant effect for the PD group. Duffy (2013) highlighted the acceleration of speech as a distinguishing feature of hypokinetic dysarthria, albeit it may not be constantly present. Indeed, the analysis of the subsample indicated variability in the presence of acceleration, which, based on an established interpretation (Duffy 2013), does not prevent the consideration of the acceleration of speech measured on an acoustic signal as a feature of hypokinetic dysarthria.

The widespread incidence of irregularities in the performance of the rhythm task indicates the unspecific character of the feature. Interestingly, irregularity was salient in diseases associated with discoordination or involuntary movements. Additionally, its significant presence in the PDT group can be accounted for by decreased speech motors control. The automated method can track the irregularity with great accuracy, but the feature has limited potential in terms of explaining what exactly the cause is; thus, the underlying pathophysiology must be deduced based on other, more specific, findings or patterns.

### 4.1.3 Sustained vowels

#### **SEGMENTATION**

The segmentation of sustained vowels traditionally employs supervised learning due to the strong relation between signal quality and the likelihood of the voice being present. However, the main reason for the preference for the supervised approach is that the decision can be defined by the technical limitations of the quantitative analysis, typically the pitch detection. Indeed, the normalized coefficient of the autocorrelation function is tested by a threshold of 0.45 in PRAAT (Boersma 1993), which limits the minimal measurable harmonics-to-noise ratio to approximately -0.9 dB. The threshold of the normalized coefficient of the autocorrelation function in the MDVP is set to the clinically more relevant level of 0.29 (Deliyski 1993), i.e., approximately a -3.9 dB harmonics-to-noise ratio. Boersma and Weenink, in the online documentation of PRAAT (2018), advocate the comparability of these thresholds since MDVP does not perform an accurate interpolation and correction of the window, according to Boersma (1993). One may argue that these thresholds are far beyond the observable harmonics-to-noise ratio; thus, they may not influence the results negatively in any considerable way. However, short-term extremities are, in fact, quite common in HNR, despite high average values for a speaker, because any instability in the period, such as an extreme jitter or alternating periods, can cancel the periodic order, and a very low HNR can be then be measured locally, even in conditions of very low additive noise. The analysis of different thresholds in PRAAT demonstrated that decision levels are indeed too optimistic and adjustment could significantly improve the accuracy. Unfortunately, a diminished threshold can never solve the above limitation of the autocorrelation function, but additional

parameters could. Therefore, two more parameters were used in addition to the crudely estimated HNR. The proposed segmentation employed a simple decision logic based on physiological assumptions to achieve superior performance in the detection of voiced intervals. The segmentation was designed specifically for sustained vowels, but can be applicable to any other task with the assumption that the recording will contain at least one interval of vocalization longer than the duration of the parametric window (75 milliseconds). The segmentation of sustained vowels, as well as several other technologies presented in this thesis, assumes low environmental noise. A complex acoustic analysis requires a quiet room much as a chemical analysis requires a clean test tube; this analogy will be valid for as long as acoustic features are measured directly without additional correction logic including quality selection or complicated noise reduction. Although some analyses are principally robust to environmental noise, such as adaptive segmentation or pitch detection, a complex assessment including perturbation measurements should not be performed on noisy or de-noised signals. This general limitation was deduced not only from the nature of the processing methods but also from unpublished experiments with noise-added and de-noised signals.

### **SPEECH FEATURES**

The steady task of sustained phonation provided a unique opportunity to measure involuntary movements of articulators as well as laryngeal muscles sensitively. Vocal arrests, as the coarsest indicators of involuntary movements, were observed in the MSA and HD groups. Excitatory trends were also present in the articulation of the vowel /A/ for MSA, HD, and CA. Finally, the fine movements of the laryngeal muscles showed higher variances in APS, HD, CA, and MS. Note that this analysis was based on a newly introduced approach that distinguishes vibrations modulated by laryngeal muscles and the effect of subharmonic vibrations. Therefore, the variation of  $F_0$  does not cover pitch jumps caused by subharmonics.

Strong excitatory trends in sustained vowels for CA and MS can be explained by ataxia that could increase the variability of steady movements due to inaccurate targeting. An inability to target pitch to a visual contour was found to be related to ataxia in experiments by McClean et al. (1987). Nevertheless, a generalization of this hypothesis to the unsteadiness of phonation and articulation in CA and MS would require additional experiments.

Increased perturbation is a common denominator of phonatory dysfunction, which, according to our results, can be expected in all types of dysarthrias analyzed in the thesis. Still, only HNR was found to be significant in PD and can be related to the low efficiency of the glottal movements that convert the respiratory flow into vibrations rather than the simple cycle-to-cycle stability, as described by jitter and shimmer. The instability in the vibrational regime in APS and HD, as manifested by the early incidence of subharmonics, was more dominant in HD. An increased proportion of subharmonics can trick the mind into perceiving a pitch break, which is an established symptom of hyperkinetic dysarthria (Duffy 2013). The neurogenic origin of subharmonics in HD can be associated with excitatory movements. Other factors, such as the disintegration of speech motor control and respiratory insufficiency determined by shortened phonation time, can contribute to the overall instability of phonation in PD, APS, and HD.

Hypernasality was present in APS and HD, in accordance with a previous study (Novotný et al. 2016). Nevertheless, the normative values of the matched healthy controls were significantly higher (t-test,  $p < 0.001$ ) for both males and females compared to the data published in the original study (Novotný et al. 2016). Furthermore, intermittent hypernasality was not significantly present

in any group. This discrepancy is caused by a different segmentation methodology. The segmentation in the original study (Novotný et al. 2016) was based on PRAAT in default settings, whereas this thesis determines voiced speech by an algorithm with different criteria.

This case demonstrates why the standardization of acoustic analysis is necessary for the introduction of acoustic analysis into clinical practice. Any adjustment in the algorithm may bias the normative data considerably. For this reason, statistical modeling is beneficial for any speech feature, as it can compensate for a systematic shift in values. Anyway, all normative acoustic data should be compared only to values calculated with the corresponding code because even similarly defined acoustic features can be incomparable when implemented differently. This general statement is based on analyses of accuracy and experiments with adjustments of methods evaluated in this thesis as well as evidence in the literature. For illustration, an analysis of sustained vowels is comparable typically for  $F_0$  characteristics but incomparable for the majority of other measurements (Bielamowicz et al. 1996, Oğuz et al. 2011, Burris et al. 2017). The unpredictable comparability of results by various algorithms is the rationale for this general limitation of acoustic analysis.

#### 4.1.4 Diadochokinetic test

##### SEGMENTATION

The diadochokinetic test and rhythm task are both, in essence, syllable repetition tasks. From the segmentation point of view, the biggest difference is the rapid of pace in diadochokinesis versus the self-determined pace of the rhythm test. The obstacles in the segmentation of diadochokinesis are analogous to the rhythm task; therefore, the advantageousness of the unsupervised approach represented here by the algorithm for segmentation of rhythm task can also be scrutinized on recordings of diadochokinesis. As diadochokinesis requires precise identification of voice onset and the additional detection of bursts, the algorithm for the segmentation of rhythm was adjusted to the rapid tempo of the diadochokinetic task, detected boundaries of voice onset were also refined by the unsupervised approach, and a new robust method for the detection of a burst was introduced—no other changes to the original algorithm were implemented.

The comparison of detectors demonstrated that the unsupervised approach improves the accuracy of segmentation substantially. A detailed analysis of errors revealed that the adaptive approach is more robust to the situation in which voicing continues between syllables, some syllables are quiet, and the rhythm is more irregular. The precise identification of voiced onset proved to be a vital factor that lead to the improvement of the VOT estimation. As the period between glottal pulses in VOT can take up to approximately 20 milliseconds (i.e., a pitch of 50 Hz), even an error of one pulse can lead to disastrous results. The selection of the initial pulse via unsupervised clustering introduced a desirable adaptability into the detection of voice onset and hence boosted the accuracy of the decision in the deteriorated signal.

In addition to voice onset, the precision of VOT relies on the precision of burst detection. A comparison of the available technologies for burst detection highlighted the convenience of phase-only reconstruction for the emphasis of the impulsive nature of a burst. Although interesting properties of phase-only reconstruction have been described by Oppenheim and Lim (1981), its implication for the localisation of impulses has been seriously overlooked. The precision was also improved by the statistical modeling of burst position, which diminished the importance of impulsive artefacts that may occur at the initiation of voicing or around the occlusion of the

previous vowel. Finally, a comparison of the errors in voice onset detection and burst detection stressed the importance of the precise detection of voice onset, which is commonly marginalized but has a considerable impact on the measured VOT and higher levels of errors than found in burst detection.

Studies by Novotný (2014, 2015) did evaluate accuracy on a small portion of the database, as the larger dataset was previously not available. The database used for evaluation included very challenging recordings; thus, these results do not downgrade the algorithm by Novotný (2015) but rather highlight that the algorithm should be used only on less severe speech disorders. The detection of voice onset time is very sensitive to the precision of both burst and voice onset, as an error of estimated voice onset time grows when the detections of burst and voice onset are both simultaneously imprecise. The results of the correlation analysis suggest that there is a room for improvement, but, as the proposed method outperformed the others, and the majority of features were strongly correlated with the reference for a variety of diseases, the methodology was considered to be preferable for experimental use in clinical applications.

### **SPEECH FEATURES**

The overall performance of diadochokinesis deteriorated in all types of dysarthria. The diadochokinesis of the RBD group, as well as that for other diseases, was slow and irregular. Based on the correlation analysis, some of the RBD speakers possibly preferred adjusting their motion rate to momentary articulation capacity, and other speakers may have simply slowed the overall rate. In addition to compensation strategies, prolongation of vowels due to slow movements or vocal emphasis also contributed to the slow alternating motion rate in APS, HD, and CA. Phonation in APS, HD, and CA was not only poorly controlled by the laryngeal muscles but was also driven by abnormal respiratory movements, as the variance in loudness increased abnormally during diadochokinesis. The increased VOT suggests that difficulties in initiating vocalization develop early in the course of PD and possibly have a good response to medication (c.f. prolongation of pauses when reading the passage and performing the monologue). Note that the interpretation of VOT is not consistent amongst authors, but findings across other tasks suggest that phonation is more explanatory, which is in accordance with Duffy (2013). This paragraph demonstrated that interpretation of diadochokinetic task in the context of other diseases relies strongly on overall speech tendencies measured on tasks other than diadochokinesis. Since alternating motions cover a wide variety of motions, the decomposition of speech processes is rather difficult in diadochokinesis. Nevertheless, speech pathologists value this task, in particular, because it can describe the overall articulatory performance in simple terms.

## **4.2 SPEECH PATTERNS**

The sensitivity of speech to neurological lesions and possible medications manifested in various trends in acoustic speech features. Generally, untreated groups showed abnormalities in more acoustic speech features, especially in connected speech, than the group treated for PD as well as the HD group. The effect of the increased duration of PDT compared to PDU (*t*-test,  $p < 0.001$ ) can be rejected, as to accept it would imply the opposite—a more frequent speech impairment for the PDT group. Overall, speech abnormalities were more salient in APS than PDT, suggesting a faster progression in APS.

Despite the considerable overlap of symptoms across various diseases and treatment groups, the central tendencies associated with the inhibition and excitation of speech movements

distinguished several groups of diseases, notably PD and HD. Inhibition was identified as a dominant speech pattern in diseases that manifest hypokinetic dysarthria or mixed dysarthria with hypokinetic components, namely, PDU, PDT, and APS. Inhibition was observed more frequently in RBD than in HC (z-test,  $p < 0.001$ ), suggesting that an inhibitory speech pattern may indicate prodromal Parkinsonian neurodegeneration. Based on an analysis of weights by the newly introduced methodology, monopitch was found to be more important for the recognition of the pattern followed by monoloudness. Interestingly, in agreement with our findings, monopitch and monoloudness were ranked by DAB as the first and third most important speech dimensions constituting hypokinetic dysarthria, respectively. Note that the inhibitory pattern was not meant to substitute for the pattern of hypokinetic dysarthria, but it can be generally associated with it.

Diseases associated with hyperkinetic and/or ataxic dysarthria or mixed dysarthria with hyperkinetic and/or ataxic component frequently showed an excitatory pattern. All HDT patients and the majority of HDU patients showed an excitatory pattern, which can be linked to the involuntary movements in HD. The presence of an excitatory pattern in MS and CA results presumably from the excessive range and force of speech movements associated with ataxia. It is noteworthy that excitation was observed more frequently in patients with MSA compared to patients with PSP (z-test,  $p < 0.01$ ), which emphasizes the contribution of this approach for the improved recognition of APS.

The hallmarks of ataxic and hyperkinetic dysarthria fall into the same category of excitatory speech pattern, which diverges from the categorization introduced by the DAB to categorize movement disorders. The rationale was given by the definition introduced in the thesis (see section 2.4.5 EXCITATORY AND INHIBITORY SPEECH PATTERNS, page 45). In addition, simple acoustic measures cannot distinguish exaggerated speech movements due to involuntary movements and discoordination. Acoustic features that are commonly associated with ataxic dysarthria, such as an irregular diadochokinetic rate (Kent et al. **2000**), measure ataxia indirectly and require thorough interpretation with regards to other findings. Although clinicians can recognize ataxic dysarthria reliably via perceptual judgements, even without proper knowledge concerning a patient's language abilities (Kent et al. **1998**), a simple acoustic measure can hardly substitute for the exquisite processing of the human auditory cortex in this task. Even the perceptual speech characteristics of ataxic dysarthria overlap frequently with hyperkinetic dysarthria (see Table 15-4 by Duffy **2013**). Ataxic movements can be deduced from the co-occurrence of acoustic features, but individual acoustic features for ataxia would require knowing the target. The intentions of the speaker, such as chosen loudness, the position of articulators, or pitch, must be compared with the speaker's performance in order to quantify ataxic movement. Hypothetically, the target can be predefined as in the experiments by McClean et al. (**1987**) or modeled based on recognized or predicted contents of speech, which would lead to a very sophisticated algorithm. This hypothesis crystalized from a deep analysis of processing methods and is mentioned here to justify the proposed definition of inhibitory and excitatory speech patterns. The development of a new method for the analysis of ataxia is far beyond the scope of this thesis, and the database is unsuitable for validation; thus, this paragraph is intended only to clarify the situation and to inspire future development of acoustic speech features for ataxia.

Generally, the overall deterioration of speech motor control can lead directly or via a compensatory mechanism to inhibitory manifestations of acoustic features. In addition, excitatory patterns of acoustic speech features may arise when deteriorated motor control induces instability into speech production. Therefore, speech features must be selected thoroughly in order to

increase the informative value of the pattern for diagnostic purposes, especially due to the increased overlap in the advanced stages of speech disorders.

Inhibition can be measured principally on tasks with an increased variance in speech movements, such as connected speech that is rich in melody, loudness levels, and stress patterns. In other words, high activity of speech subsystems is required from the speaker to measure inhibition. Conversely, excitation can be measured for a task with a reduced variance in speech movements, such as sustained vowels that are performed steadily with the low overall activity of speech subsystems. In sustained vowels, the target is to perform with the lowest variability possible. Thus, any discoordination or involuntary movement can be measured directly as increased variance.

A simple combination of sustained vowels for the measurement of excitatory speech movements and connected speech for the measurement of inhibitory speech movements proved to be successful strategy with which to estimate incidences. Although the diadochokinetic test and rhythm task may increase the incidence of patterns within disease groups, the possible overlap between patterns would degrade the diagnostic value for more severe speech disorders.

Various classifiers, including naïve Bayes based on a kernel density estimation, support a vector machine with radial basis function, or a shallow neural network can recognize the proposed speech patterns. Notably, a novel approach based only on the weighted fusion of z-scores showed a performance in recognising speech patterns that was comparable to the other classifiers despite its simplicity. The biggest advantage of the proposed pattern recognition is that it can decompose a speech pattern into the contributions of individual speech features, making its results highly interpretable. The majority of classifiers serve as so-called “black boxes” that provide only decisions without any explanation. The contribution of features calculated by the proposed methodology explains decision fully and allows the user to judge the reliability of the prediction in terms of other factors not in the computational model, such as speech idiosyncrasies. Consequently, the results of the classification obtained by the method described in this thesis are transparent and make it possible to interpret the results in the context of other findings, patients’ histories, or socioeconomic backgrounds. Additionally, the accordance between the trained pattern and hypotheses can be tested easily by checking the weights of individual speech features, which is priceless for preventing overtraining. Another advantage of the method is that it allows the clinician to consider the severity of the speech pattern anomalies proportionally to the enumerated value of the z-score or p-value. The classification experiment used a significance level of 0.05, but any other level can be tested as well to control the sensitivity of the decision easily in various situations, such as a high-throughput screening of the population vs. screening within the high-risk group. The presented methodology for classification complies fully with the demands for a medical-grade classifier for the recognition of speech patterns.

### 4.3 CLINICAL APPLICABILITY

Compliance with the requirements of a clinical examination, as summarized in the introduction to this thesis, and the experimental use of the software evaluated by the survey were the key sources for the discussion of the proposed methodology in terms of clinical applicability. Answers to questions presented in the guidelines “What To Ask When Evaluating Any Procedure, Product, or Program” by The American Speech-Language-Hearing Association (2018) were embedded into the frame of the discussion in order to facilitate ranking of the presented method.

The methodology was designed as an acoustic analyzer of dysarthria for speech pathologists and neurologists. Based on an analysis of different recording devices (Rusz et al. 2018), only a recording system (type and manufacturer) similar to the one that was used for recording the control database should be used for a complex assessment of dysarthria. If a different recording system is desired, then a new control database must be recorded on the device or a similar brand since recording devices have different spectral characteristic which can influence the acoustic analysis, with a small exception being intonation- and segmentation-based descriptors (Rusz et al. 2018). The hardware and system requirements of the software are determined by © MATLAB (MathWorks, Natick, Massachusetts, USA) and may vary depending on the version. Only basic computer literacy is necessary for the execution of the analysis, i.e., the management of files and folders, interaction with the simple graphical user interface, and opening html files in a web browser. Depending on the recording procedure, trimming audio files in the audio editor may be required, especially when the examination was recorded into one file. All requisites can be trained within a brief session. No training material has been developed currently for the developmental stage of this project. Thus, only those individuals trained personally can be considered to be qualified.

The database analyzed in the thesis comprised various stages of hypokinetic, hyperkinetic, and mixed dysarthria with hypokinetic, spastic, ataxic, or hyperkinetic components. The methodology is applicable for neurodegenerative diseases associated with a dysfunctional basal ganglia and cerebellum. Although MS and CA can also affect the brainstem and spinal cord, the analysis would require a larger dataset and specific evaluation to support the extended application of the methodology. Recommended diseases can be analyzed and interpreted reliably, but, for others, interpretation is not possible due to the lack of knowledge about the incidence of acoustic speech manifestations. Indeed, few anomalies were identified in patients with another diagnoses during the experimental use by a speech pathologist. Unfortunately, an extension of the database would be required for clarification of these unexpected findings. Since many speech parameters can vary based on language, especially features measured on connected speech, the normative data provided by the thesis are applicable only to Czech native speakers. When a new language is desired, a special control group in the new language must be recorded to obtain valid normative data.

Based on the survey, the proposed methodology's software was appreciated as user-friendly, offering a high degree of interpretability and clinical relevance. The most positively rated aspect was the increased intelligibility of results induced by the statistical modeling of measured values and patterns. Furthermore, all models were linked to hypotheses concerning the underlying speech pathologies; thus, no deep understanding of digital signal processing and machine learning is required for the interpretation of the results. The statistical modeling proposed in the thesis eliminates the effects of sexual dimorphism and age and allows the clinician to consider the degree of abnormality without prior knowledge of the normative data. Therefore, the methodology can be employed with ease by speech pathologists, neurologists, and other clinicians.

The methodology does not suggest any diagnostic option. Only basic speech tendencies are suggested to the clinician, since more data than an acoustic signal is required for the diagnosis of a motor speech disorder. Based on the presented findings, speech patterns evolve with disease progression, individual response to therapy, specific compensation of speech disability, and the patient's emotional state, all of which can hardly be modeled based on the current database and knowledge about the mechanisms behind speech patterns. Generally, all of the variables that a

clinician considers in a diagnosis must be incorporated into a computational model to automate the diagnosis of dysarthria. Building such a universal model would require tremendous longitudinal data for various diagnoses, medications, and languages. Finally, acoustic methods can be introduced into ambulances as something that analyzes not diagnoses, simply because clinicians are intuitively aware of these limitations.

The thesis introduced a method that extends and objectifies acoustic speech symptoms of neurodegeneration with a high degree of interpretability. Several novel features, such as descriptors of subharmonics, are capable of deciphering phenomena that can be hardly or vaguely described by perceptual features. The form of the proposed report was highly appreciated for its lucidity. The dizzying array of various characteristics can be reviewed quickly, simplifying the examination of speech for an examiner and allowing him or her to fully concentrate on the patient. The longitudinal charts, in particular, pleased the clinician as well as the patients since a simple image is capable of capturing the patient's performance with more lucidity than mere words. The encouraging effect of a plotted performance is well known, and manual charting is recommended by prominent speech pathologists, despite the fact that manual charting can be bothersome (Dworkin 1991). The proposed methodology eliminates completely the manual recording of scores. The clinician can then fully interact with the patient, and the patient is no longer preoccupied with the resulting score (Dworkin 1991). In summary, the methodology can help clinicians to identify the critical breakdown in the hierarchy of speech production and track speech changes over time with ease.

The accuracy analysis conducted on a large database of various disorders demonstrated that the methodology for the processing of acoustic signals used for the calculation of speech features represent the most precise method available up to the date of this thesis. Currently, no comparable methodology for a complex acoustic analysis that employs statistical modeling of speech features and has been validated on a large dataset of various diseases is available. The application is currently not publicly available, but a demo version will be released in the future.

## 4.4 LIMITATIONS AND FUTURE STEPS

Only recordings of sustained vowels, the rhythm task, the diadochokinetic test, and connected speech were subjected to analysis. These selected tasks can assess basic speech abnormalities associated with diagnoses in the database in a complex fashion (Duffy 2013). Nevertheless, they represent only a portion of the possible tasks that can contribute to the examination of speech. Highly specialized isolated tasks, such as the fast repetition of syllables or prolongation of unvoiced fricatives, were requested by speech pathologists during the development of the algorithms because they can be convenient for tracking the individual effects of speech therapy. Unfortunately, development of algorithms covering these tasks was not possible because the database was originally built for analysis of diagnostic features, not for speech therapy; thus, these specialized tasks were not recorded in the database.

The cross-sectional design of the database is another factor that limits the hypothesized benefits of the methodology for tracking speech therapy. However, the qualities of the methodology for tracking speech changes were verified during experimental use. Future evolution of the application will most likely be centered on longitudinal data and isolated tasks since these topics are desirable for clinical practice.

Currently, the normative data are restricted to the Czech language. Although the isolated task may be robust against the effect of the language, the use of the presented normative data on other languages will presumably lead to invalid results. Only a thorough analysis across multiple languages could warrant language independency. Generally, language influences acoustic features in connected speech strongly (Maidment **1976**, Ramus et al. **1999**), and an exception can be hardly envisaged for all proposed metrics. To conclude, the future database must cover multiple languages to verify extended applicability, since many of the speech manifestations that are salient in the Czech language may not be salient in other languages and vice versa.

Acoustic features based on simple statistics, such as mean values and the dispersion of parameters, were preferred in the thesis because more complex processing would require either laborious analysis, such as the evaluation of stress patterns, or reference measures, such as those obtained with electromyography. Despite the simplicity of the proposed metrics, clinicians may not understand all novel features correctly. Namely, the entropy of speech timing seems to be a very problematic term for anybody with a background only in speech pathology. Addressing the relation between therapeutic effects and speech features will be necessary to induce intuition concerning these complicated terms.

The biggest limitation identified during experimental use was related to the fact that speech was recorded into the flash card, then manually trimmed and analyzed on different devices, which is laborious and does not allow for controlling the quality of the recording process. Since the recording system used for collecting the database and in the experimental application had to be similar, these technical obstacles could not be addressed during the experimental use. Nevertheless, this limitation can be removed in the future through the incorporation of the recording process into the analyzing application.

Currently, the normative data can be compared only to recordings captured by a recording system of the same brand as that used in the thesis. Spectral measures, such as RFA, can be influenced considerably by the spectral characteristic of the recording system. Unpublished experiments with Wiener filtering of the data captured by a smartphone demonstrated that the spectral characteristic can be compensated for sufficiently, but only if the microphone is placed steadily as described in section 2.2 RECORDING PROCESS, page 19. An analysis of the compensation filter showed that any adjustment of the microphone position can lead to an alteration in the measured spectrum, which can be problematic, especially when recording via smartphone since position is not as strictly defined as it is when a headset is used. Technically, the majority of obstacles that prevent the wider use of acoustic analysis were tackled by the thesis, except the effect of the microphone. The short-sighted solution can be based simply on the recording of a control database using the preferred device. The sophisticated solution could compensate for spectral characteristics via the reference signal, which would require a thorough analysis of signals recorded simultaneously using several devices.

Only the fundamental issues of the thesis were addressed here. The survey on the experimental use identified some minuscule problems that relate more to the implementation of the methodology and thus can be solved programmatically, e.g., the scaling of the longitudinal graphs. The thesis was not meant to be about implementation of software and does not provide the code with regard to the developmental stage of the project. Due to limited human resources, only one clinician was involved into the experimental use. Nevertheless, the clinician's exceptional commentaries definitely helped to identify the critical deficits and strengths of the proposed system. The questionnaire was included in the thesis rather as a basis for discussion of applicability

than as a tool for the evaluation of the possible clinical impact of the method. Only long-term experimental use and further development in cooperation with clinicians could clarify whether or not the proposed approach has the potential to revolutionize the art of dysarthria assessment.



# 5

## CONCLUDING REMARKS

*Any damn fool can get complicated, but it takes a genius to attain simplicity.*

*–Pete Seeger, Woody Guthrie Folk Songs, 1963*

The thesis summarizes the author’s experiences, ideas, and discoveries in the form of a comprehensive recipe for the acoustic analysis of dysarthria covering all levels of acoustic examination from capturing the data, through innovative methods for digital signal processing and the modeling of speech features to the design of a highly intelligible report. Principally, the thesis contributes to the knowledge-driven approach through hand-designed acoustic features and opposing the current trend in data-driven acoustic analysis. The results of the accuracy analysis suggest the crucial role of adaptability in the digital processing of disordered speech. The models of speech patterns presented in the thesis overcame the fundamental limitation of acoustic analysis by reducing the effects of age and sexual dimorphism. Additionally, a novel approach for pattern recognition and decomposition achieved a comparable performance to other established recognition methods and showed increased transparency and interpretability of results. The statistical analysis of acoustic features stressed the importance of speech patterns due to the overlap between individual characteristics of dysarthrias. The inhibitory and excitatory speech patterns introduced by the thesis qualify as an efficient way to describe elementary tendencies of speech movements detectable by simple acoustic measures. In summary, the thesis defined clinically applicable solutions for acoustic analysis that may extend the current battery of tests used for the examination of speech from the prodromal to developed stages of neurodegenerative diseases.



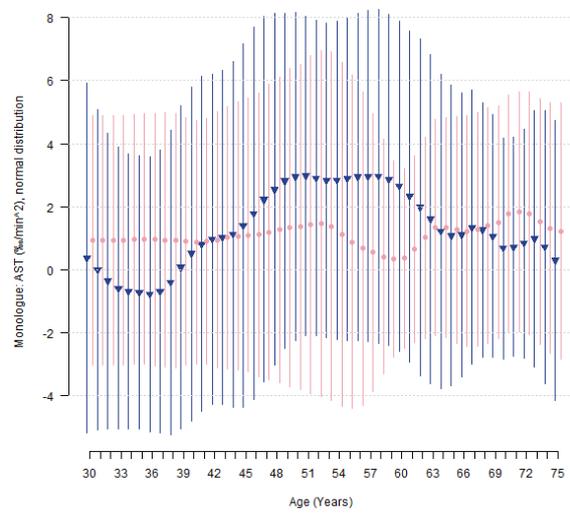
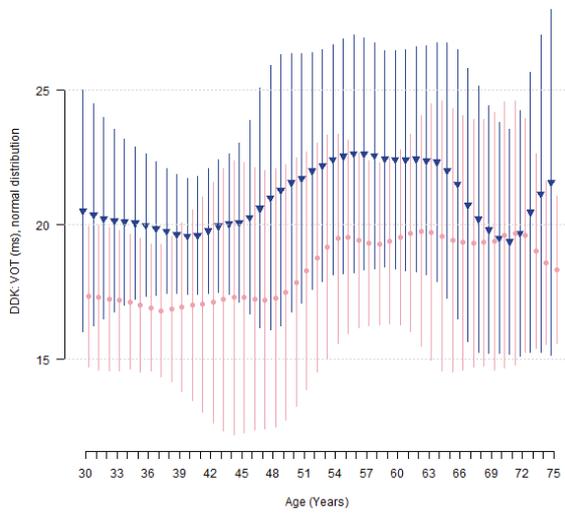
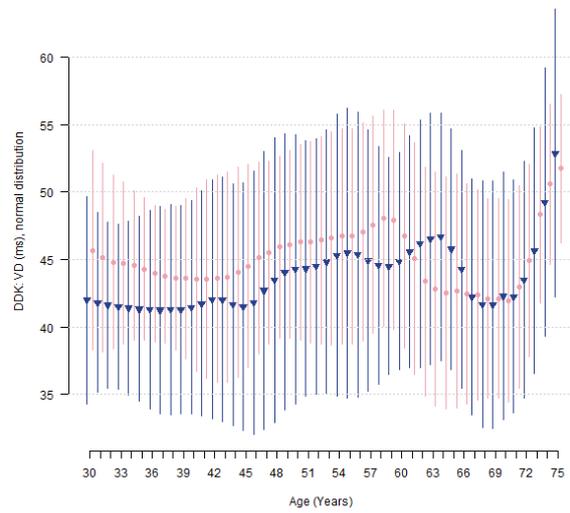
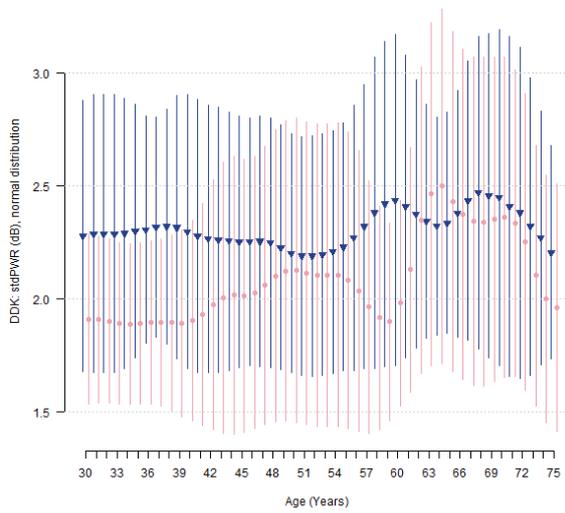
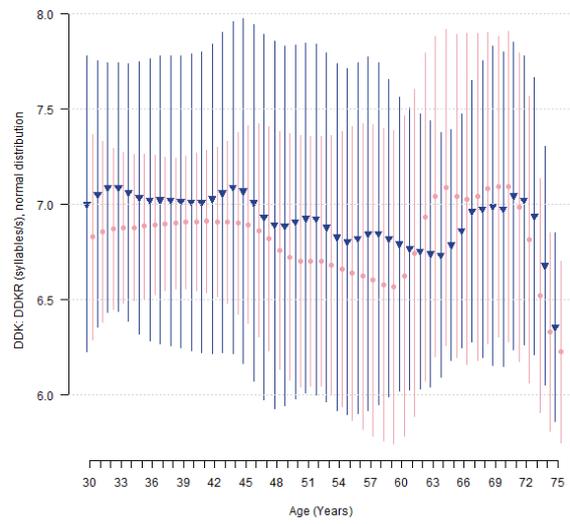
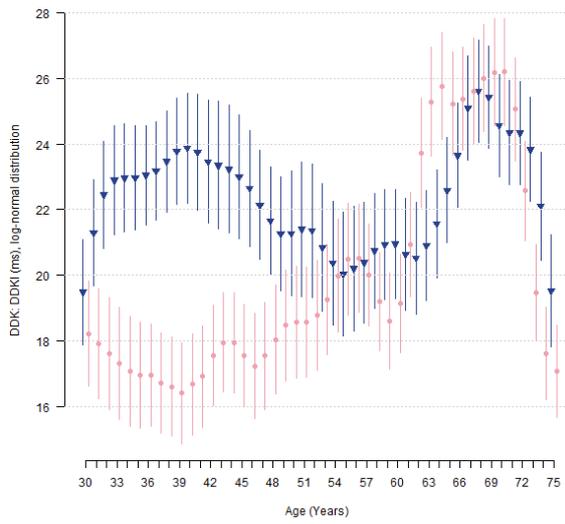
# APPENDIX A:

## NORMATIVE DATA FOR THE CZECH LANGUAGE

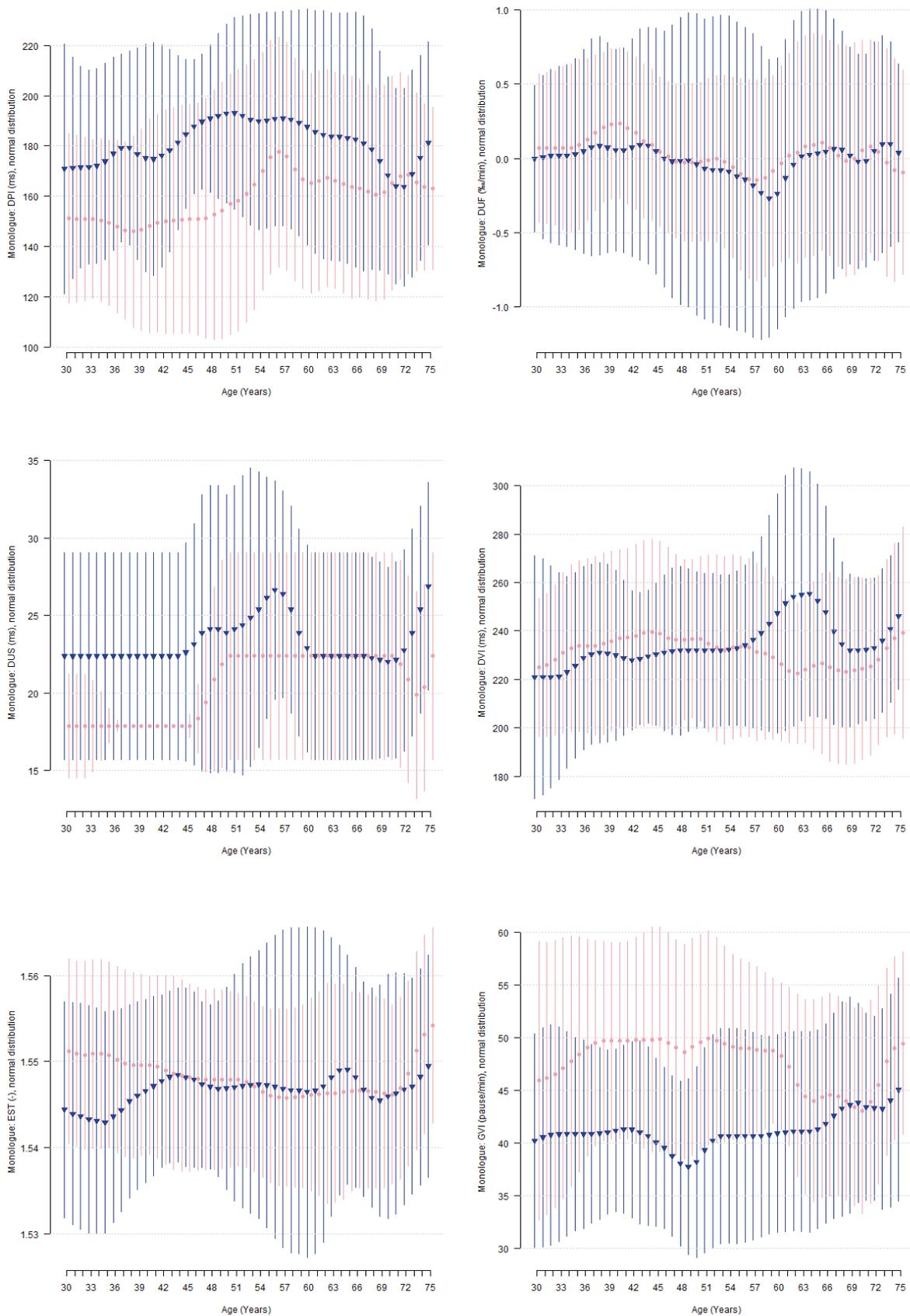
**N**ormative data are plotted on following pages separately for males (dark blue) and females (light pink). Values for the males are located on the left side of ticks and values for the females on the right side of ticks within the same figures. The parameters of the lognormal distribution were exponentiated in order to plot the results in original scale. Points denote the mean values of the normal and lognormal distributions or the shape parameter of the gamma distribution. Error bars illustrate the standard deviations of the normal and lognormal distributions or the scale parameter of the gamma distribution. Features, described by task and abbreviation, are listed in alphabetical order.

**Abbreviations:** DDKI = diadochokinetic irregularity, DDKR = diadochokinetic rate, stdPWR = standard deviation of power, VD = vowel duration, VOT = voice onset time, AST = acceleration of speech timing, DPI = duration of pause intervals, DUF = decay of unvoiced fricatives, DUS = duration of unvoiced stops, DVI = duration of voiced intervals, EST = entropy of speech timing, GVI = gaping in between voiced intervals, LRE = latency in respiratory exchange, PIR = pause intervals per respiration, RFA = resonant frequency attenuation, RLR = relative loudness of respiration, RSR = rate of speech respiration, RST = acceleration of speech timing, stdF0 = standard deviation of fundamental frequency, stdPWR = standard deviation of power, DVA = degree of vocal arrests, HNR = harmonics-to-noise ratio, LSI = location of subharmonic intervals, MPT = maximum phonation time, PSI = proportion of subharmonic intervals, stdPSD = standard deviation of power spectral density, EF<sub>n</sub>\_M = degree of hypernasality, EF<sub>n</sub>\_SD = intermittend hypernasality, NSR = net speech rate, RA = rhythm acceleration, RI = rhythm instability.

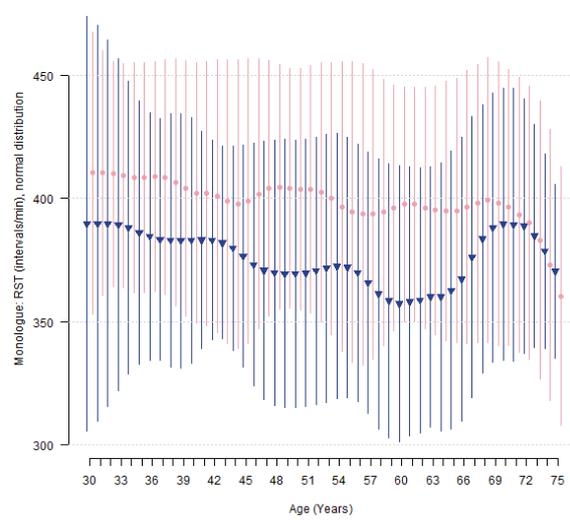
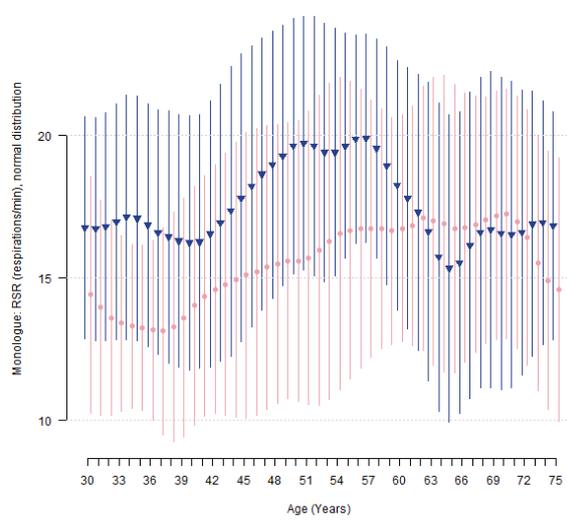
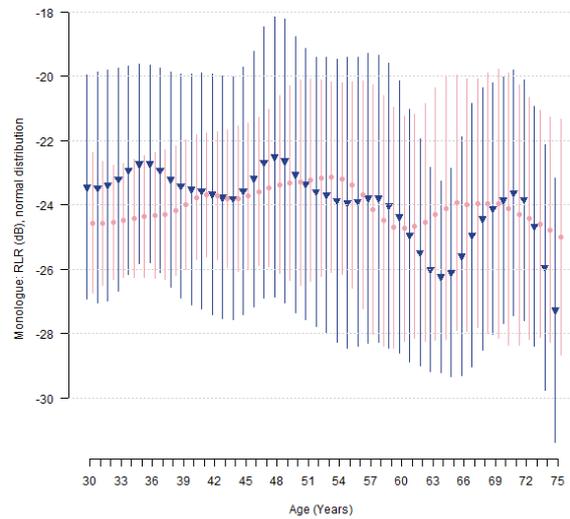
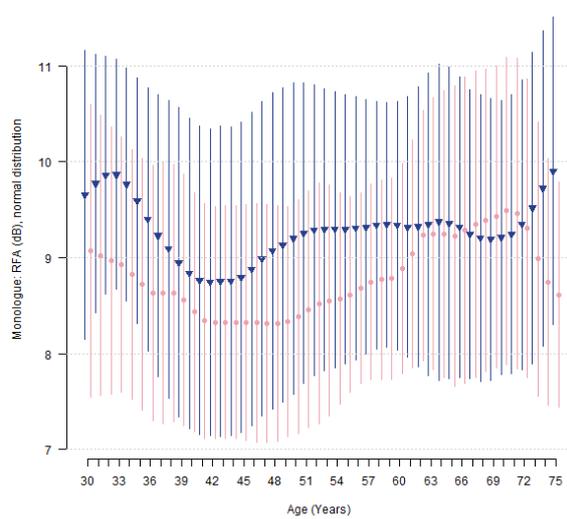
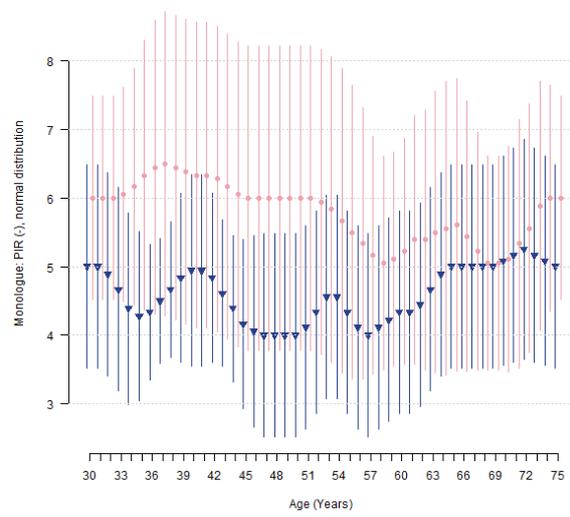
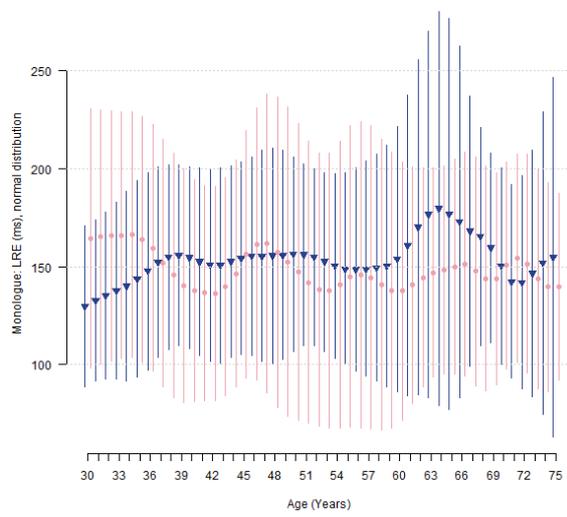
# Appendix A: Normative data for the Czech language



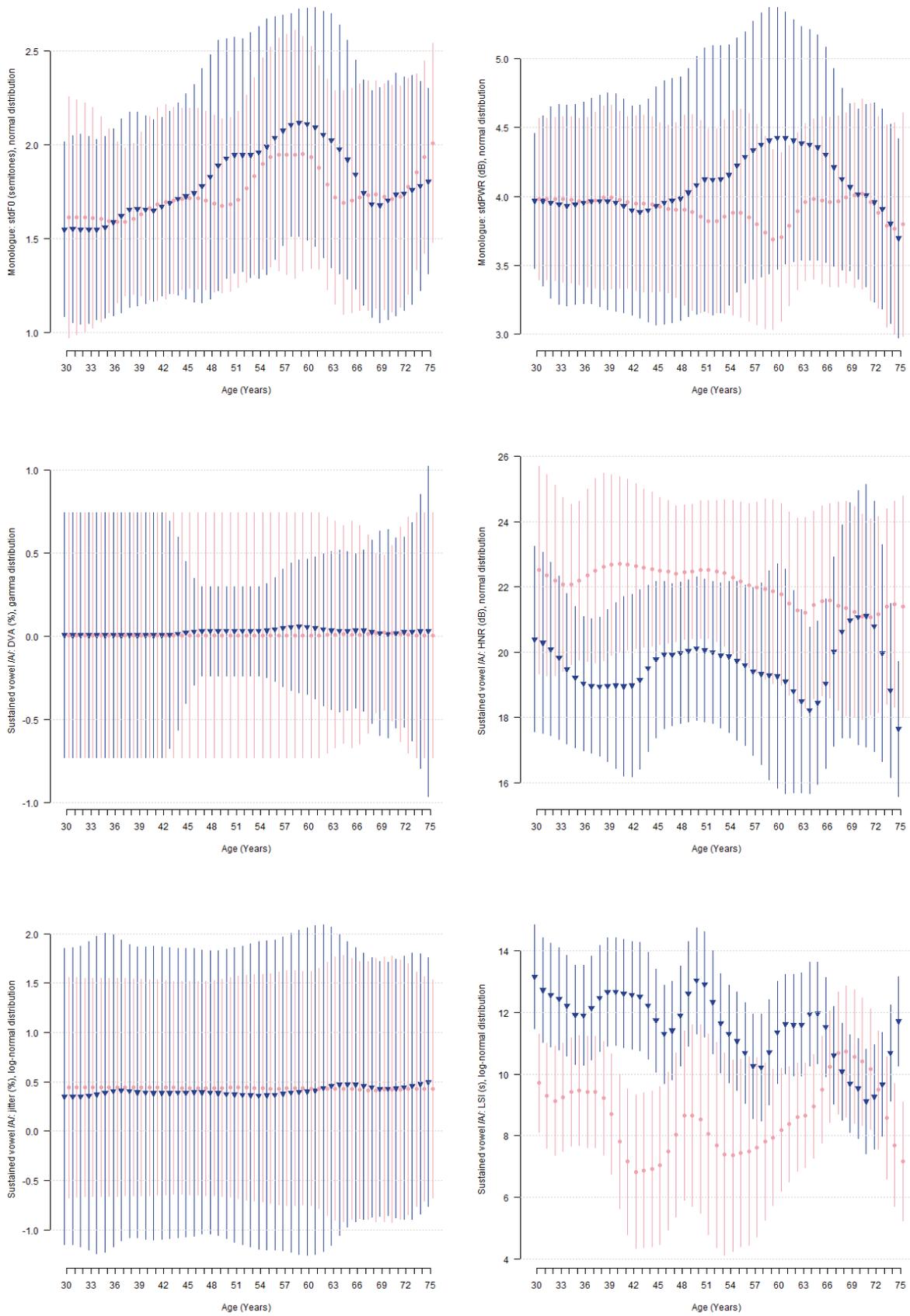
# Appendix A: Normative data for the Czech language



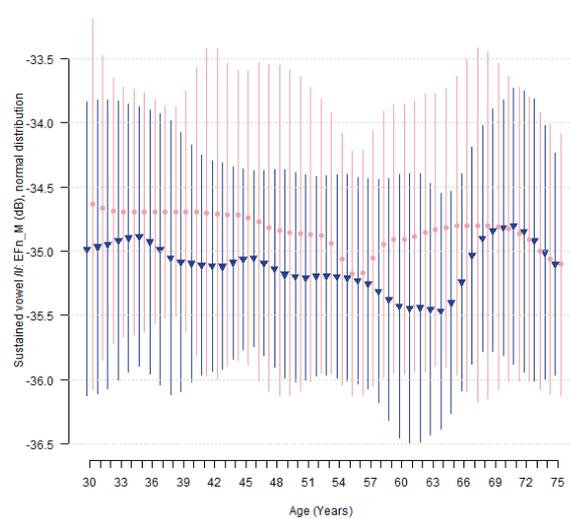
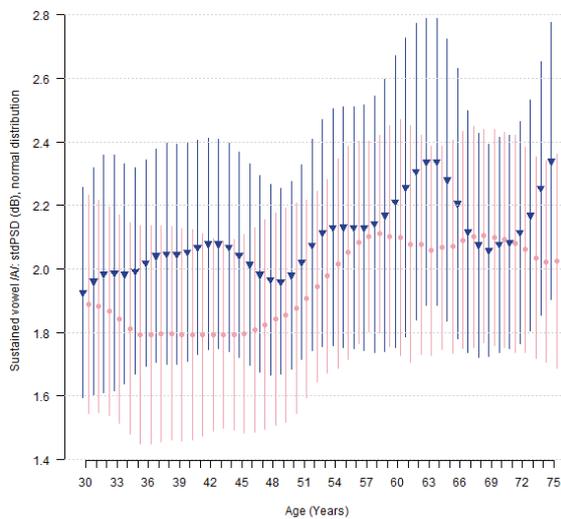
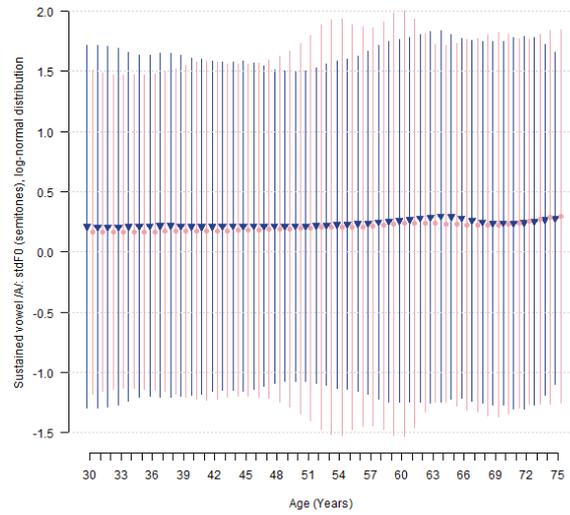
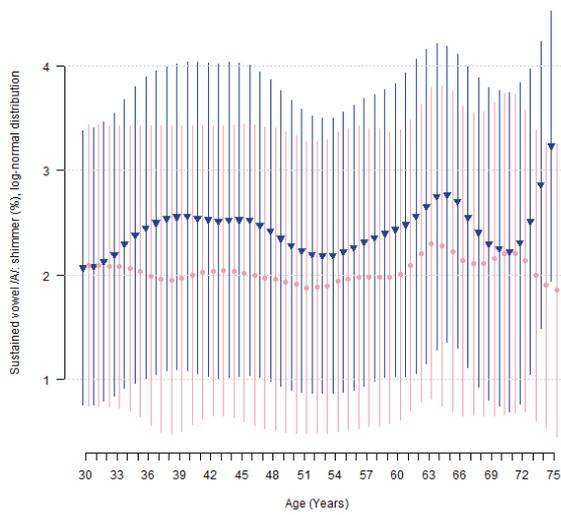
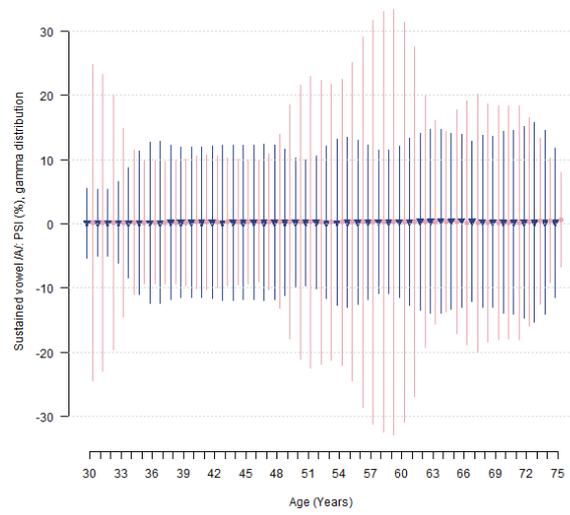
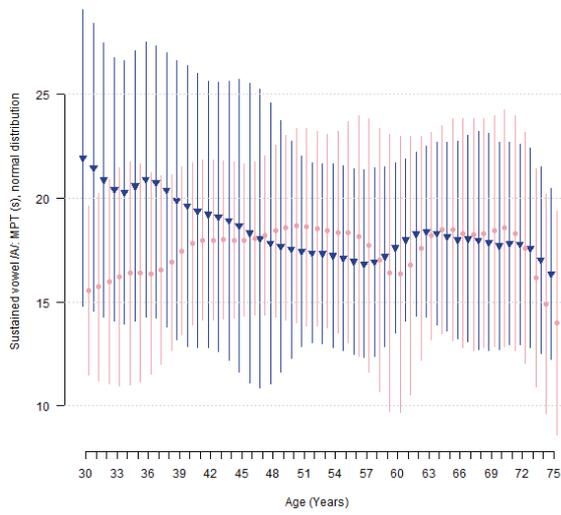
# Appendix A: Normative data for the Czech language



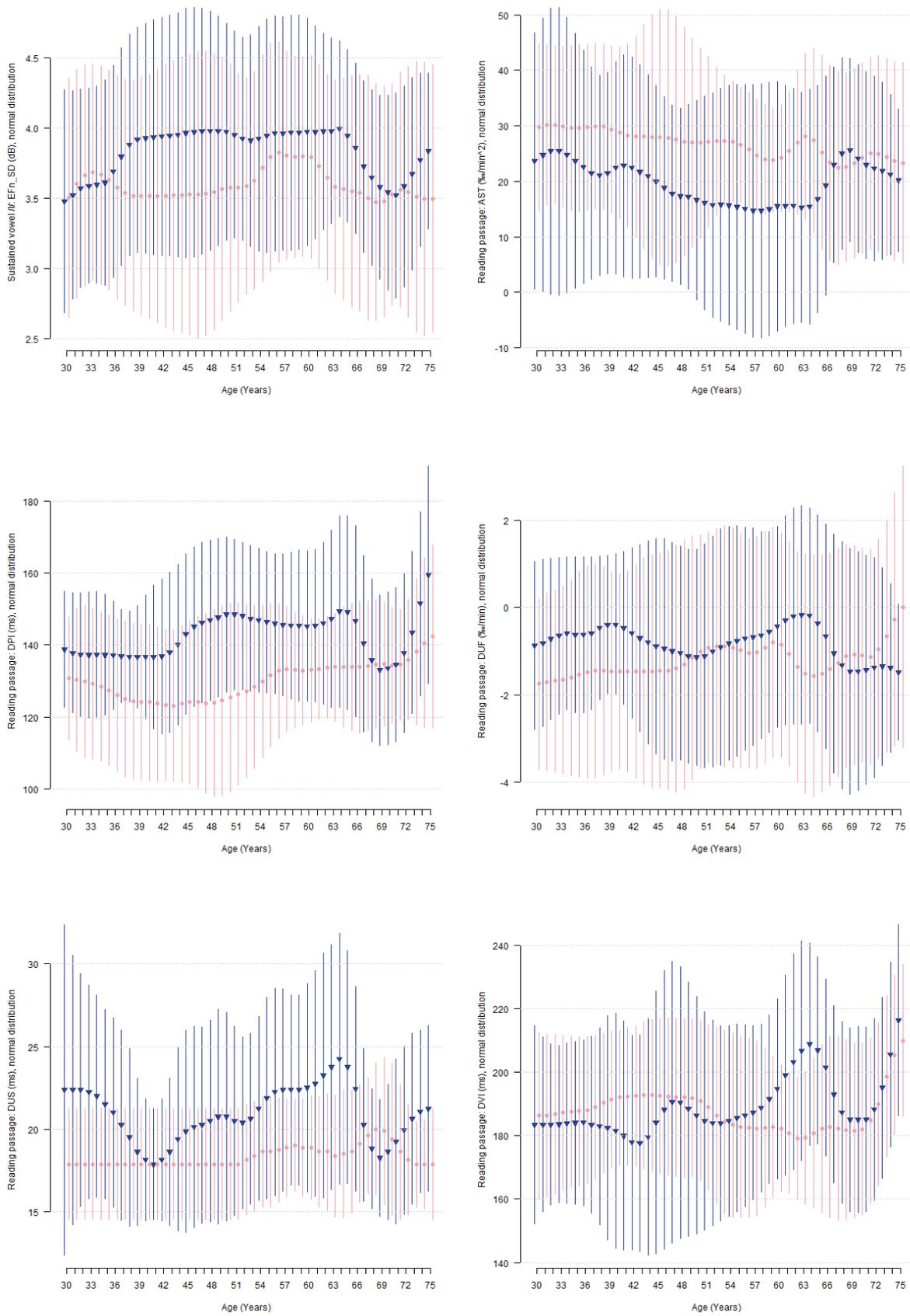
# Appendix A: Normative data for the Czech language



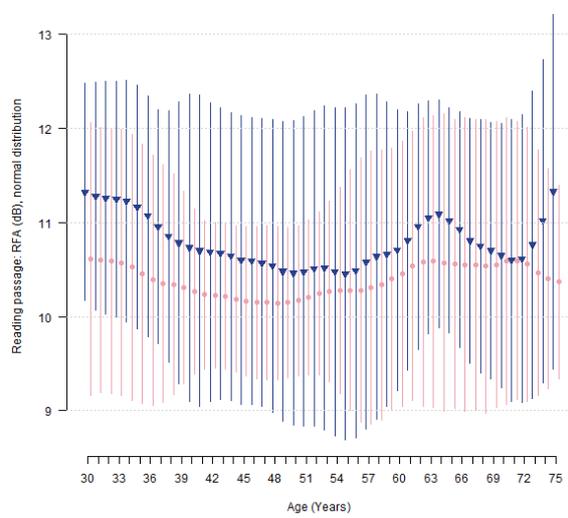
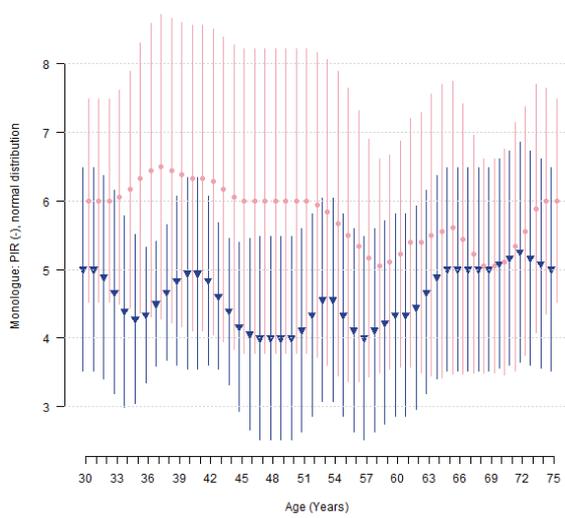
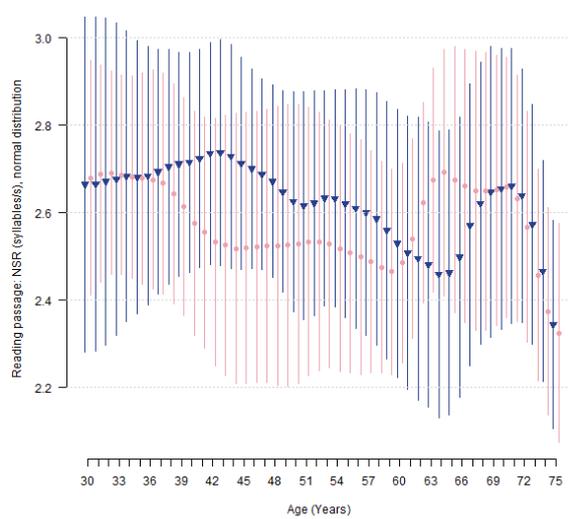
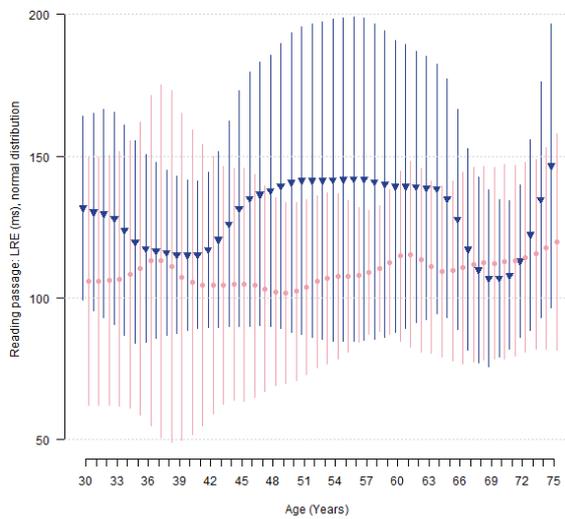
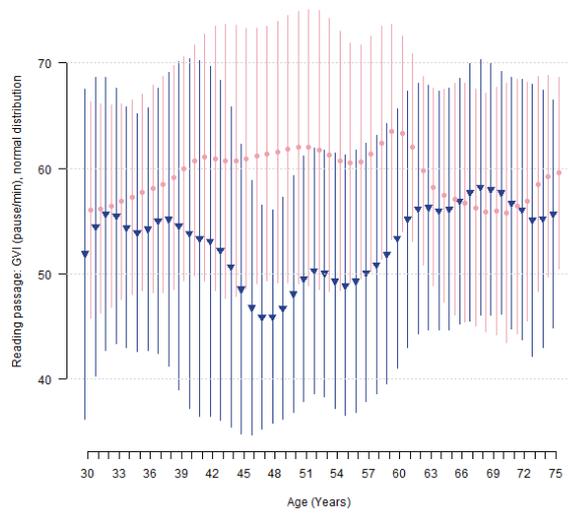
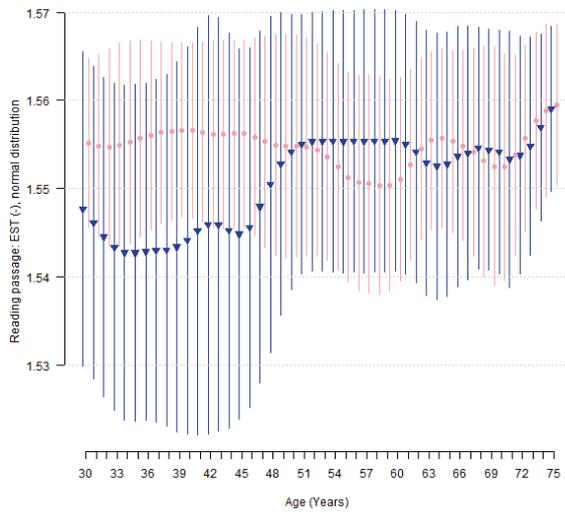
# Appendix A: Normative data for the Czech language



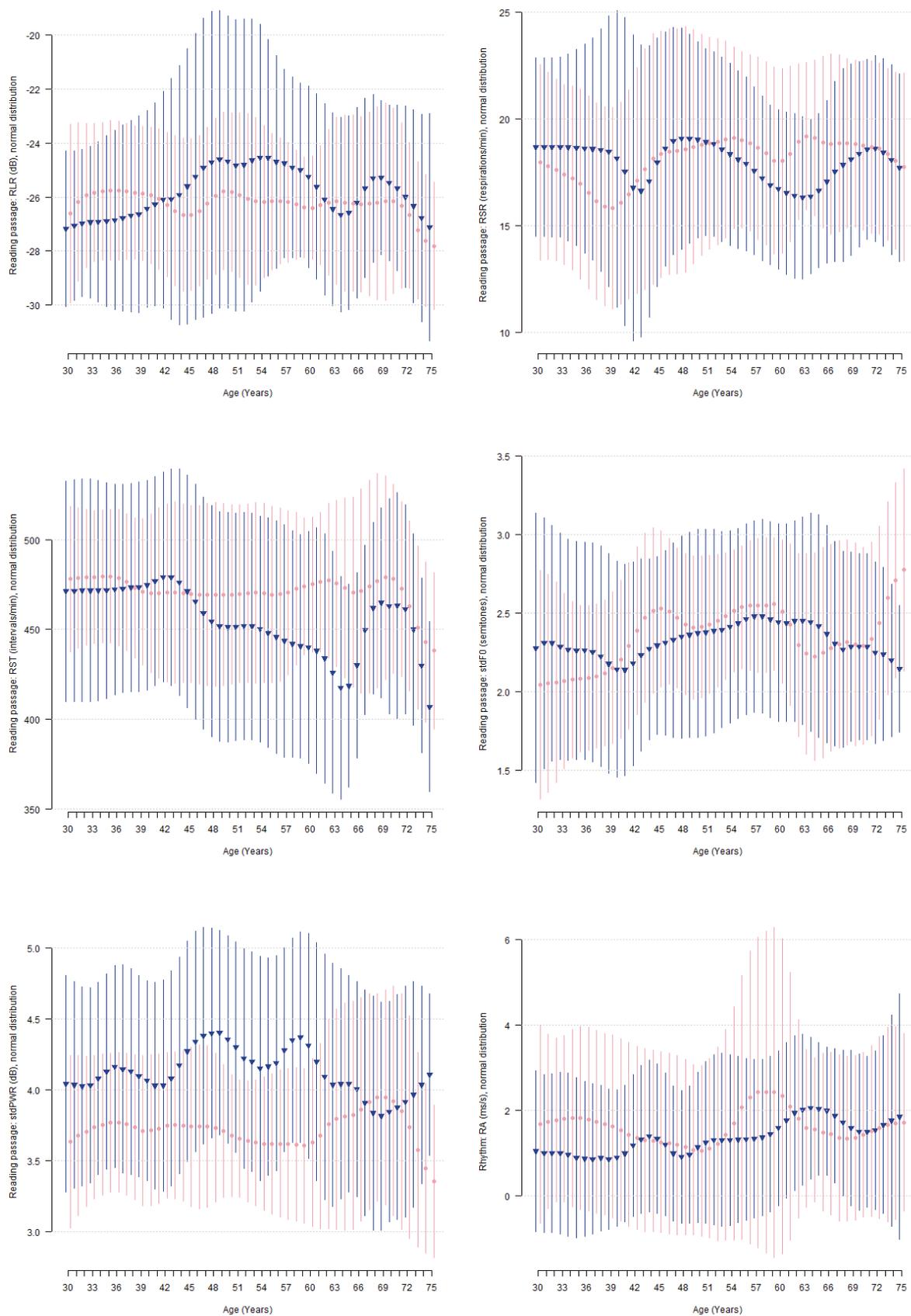
# Appendix A: Normative data for the Czech language

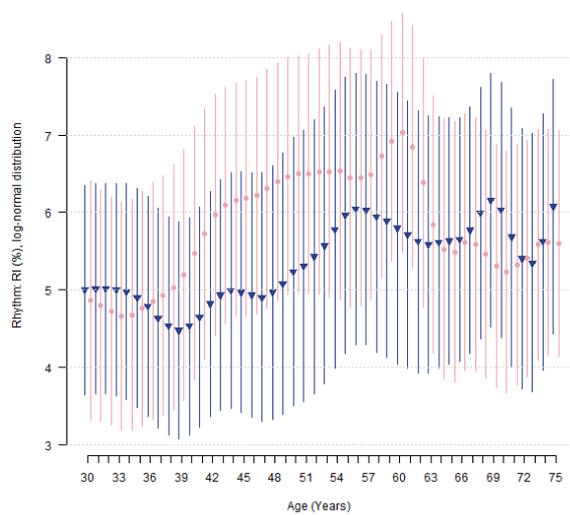


# Appendix A: Normative data for the Czech language



# Appendix A: Normative data for the Czech language





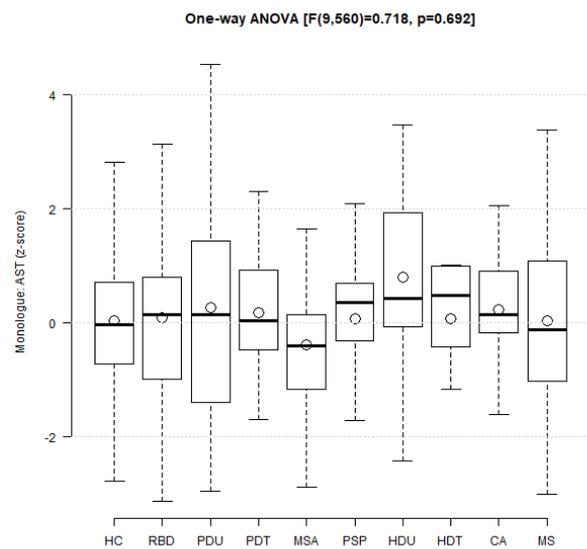
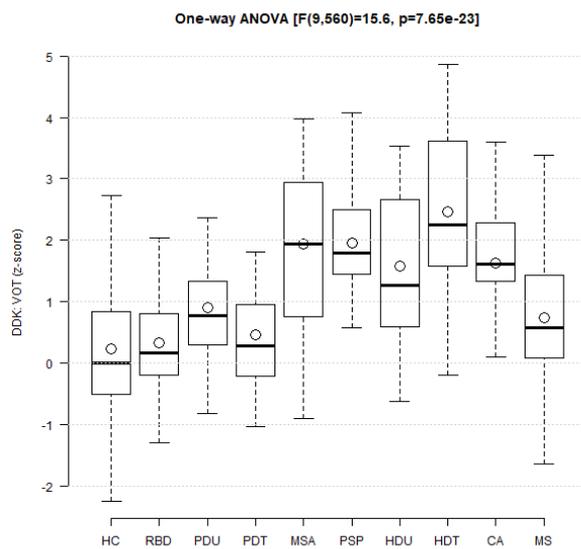
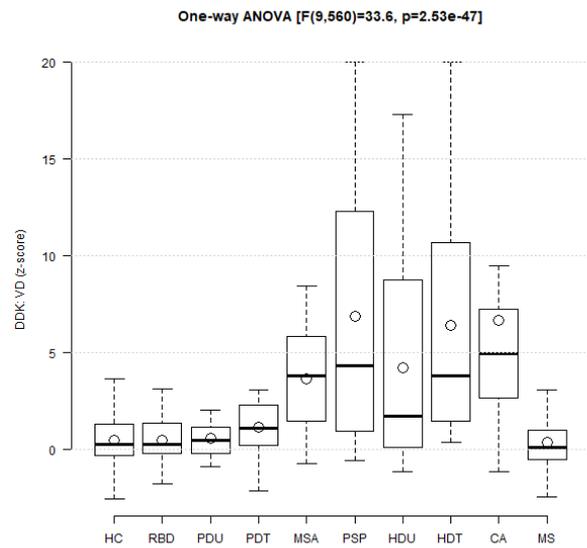
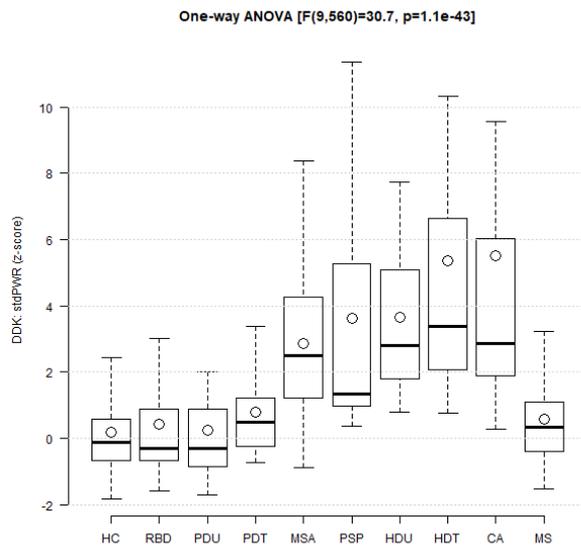
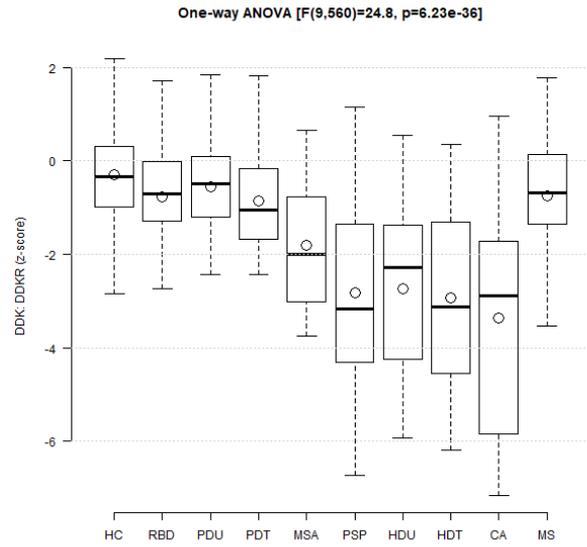
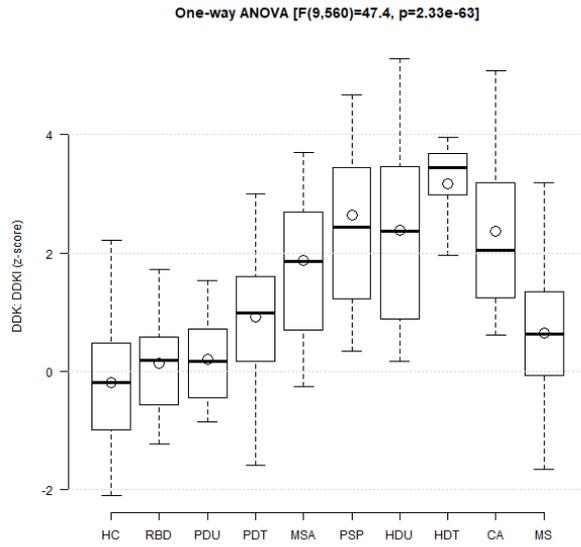
# APPENDIX B:

## NORMALIZED VALUES OF SPEECH FEATURES

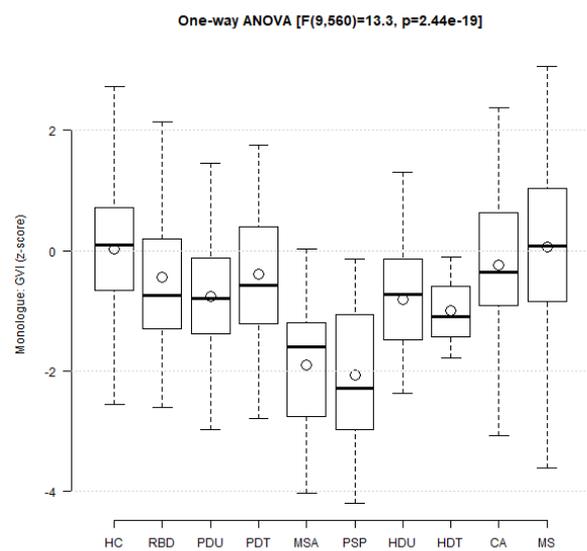
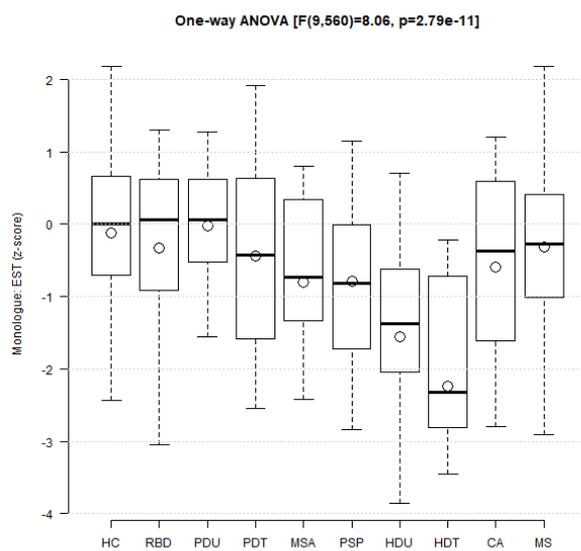
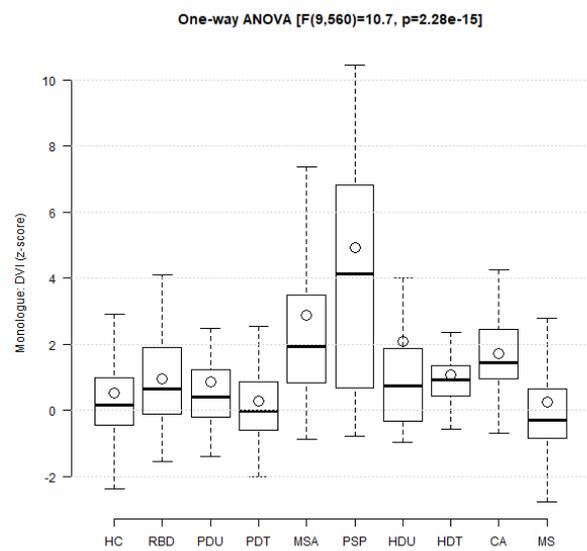
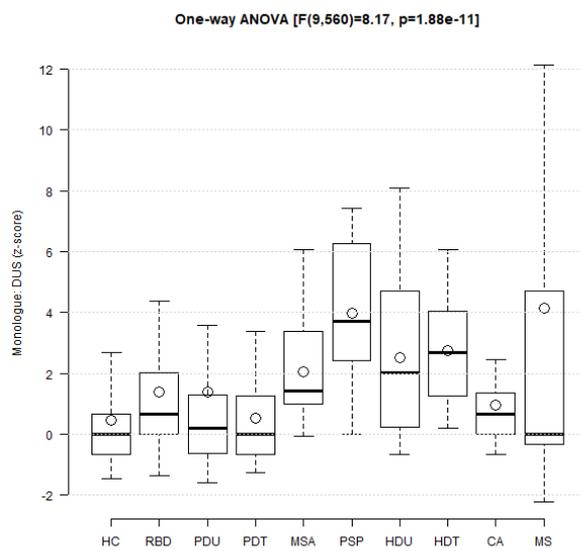
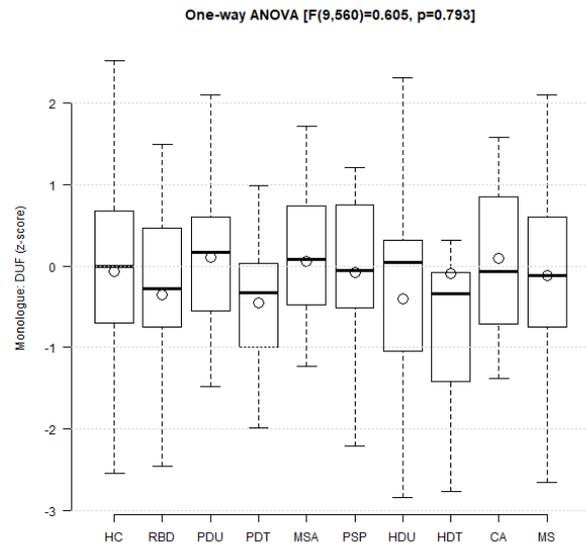
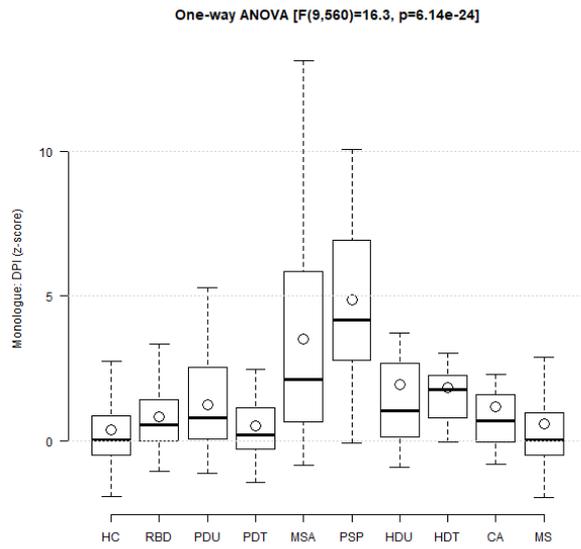
The normalized values of features are summarized in boxplots representing the characteristics of individual groups. Mean values are marked with circles. In order to increase the readability of the graphs, outliers were not depicted and all values were limited in range from -20 to 20. Features, described by task and abbreviation, are listed in alphabetical order.

**Abbreviations of groups:** HC = healthy control, RBD = rapid eye movement sleep behavior disorder, PDU = untreated Parkinson's disease, PDT = treated Parkinson's disease, MSA = multiple system atrophy, PSP = progressive supranuclear palsy, HDU = untreated Huntington's disease, HDT = treated Huntington's disease, CA = cerebellar ataxia, MS = multiple sclerosis.

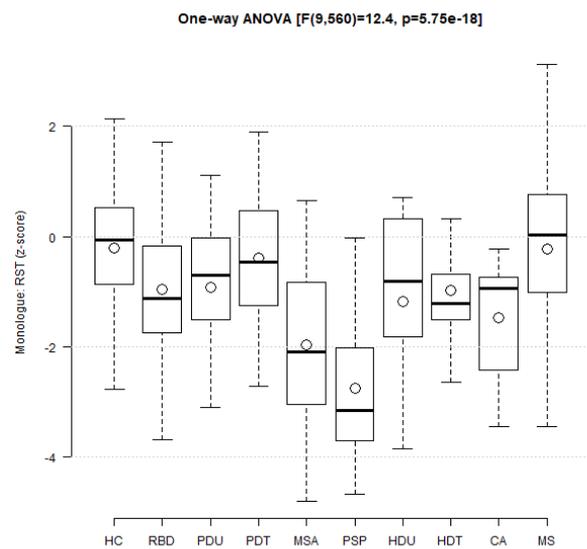
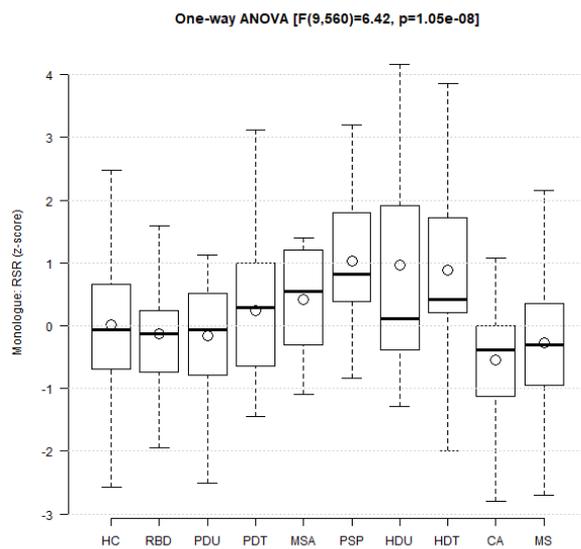
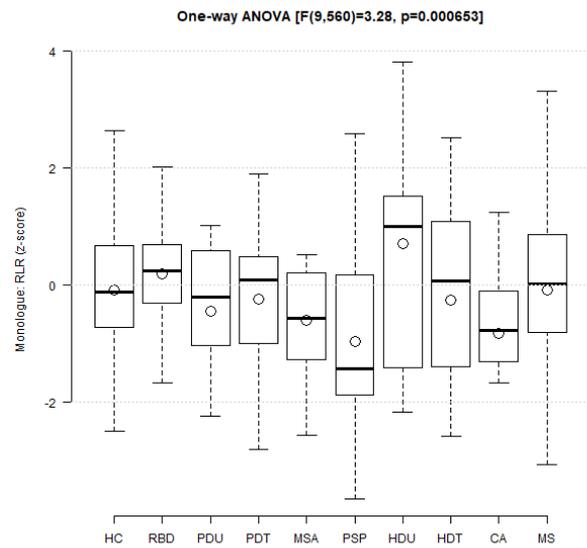
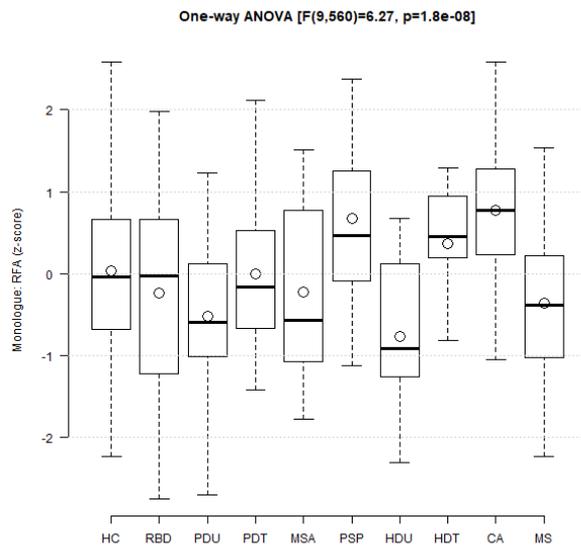
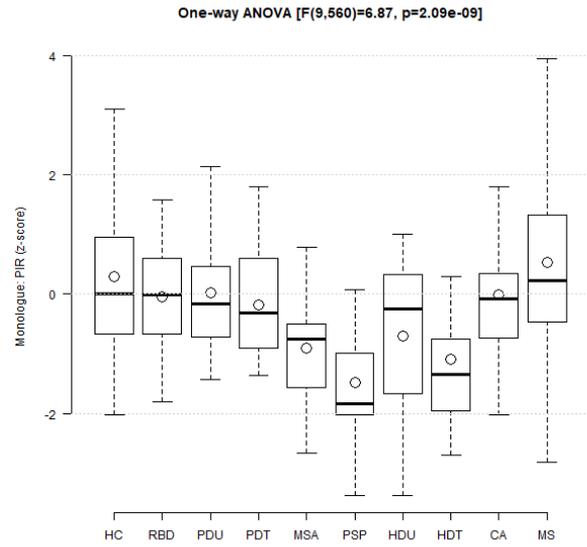
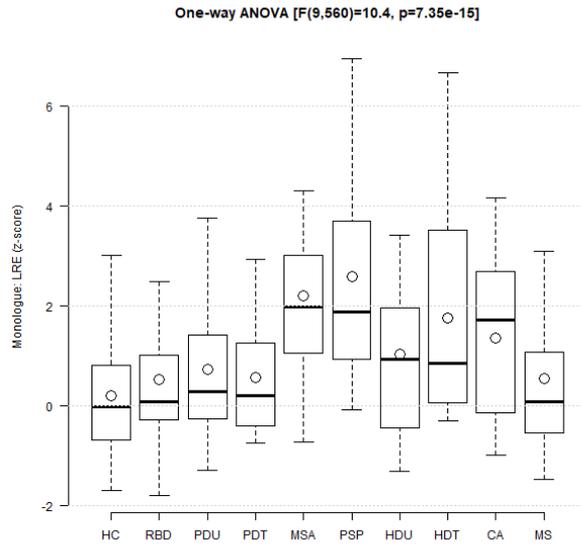
**Abbreviations of features:** DDKI = diadochokinetic irregularity, DDKR = diadochokinetic rate, stdPWR = standard deviation of power, VD = vowel duration, VOT = voice onset time, AST = acceleration of speech timing, DPI = duration of pause intervals, DUF = decay of unvoiced fricatives, DUS = duration of unvoiced stops, DVI = duration of voiced intervals, EST = entropy of speech timing, GVI = gaping in between voiced intervals, LRE = latency in respiratory exchange, PIR = pause intervals per respiration, RFA = resonant frequency attenuation, RLR = relative loudness of respiration, RSR = rate of speech respiration, RST = acceleration of speech timing, stdF0 = standard deviation of fundamental frequency, stdPWR = standard deviation of power, DVA = degree of vocal arrests, HNR = harmonics-to-noise ratio, LSI = location of subharmonic intervals, MPT = maximum phonation time, PSI = proportion of subharmonic intervals, stdPSD = standard deviation of power spectral density, EFn\_M = degree of hypernasality, EFn\_SD = intermittent hypernasality, NSR = net speech rate, RA = rhythm acceleration, RI = rhythm instability.



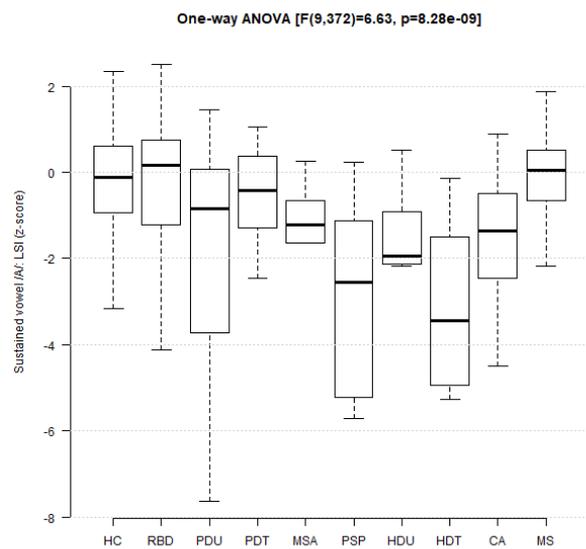
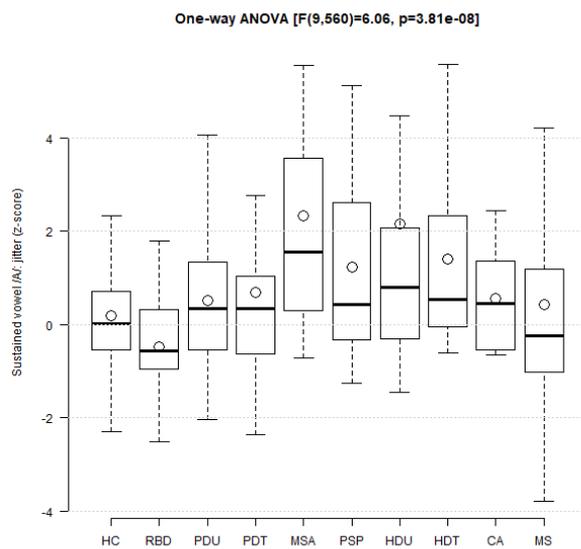
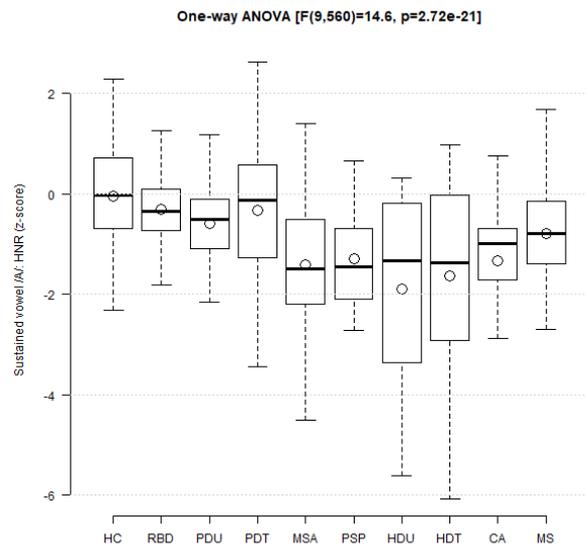
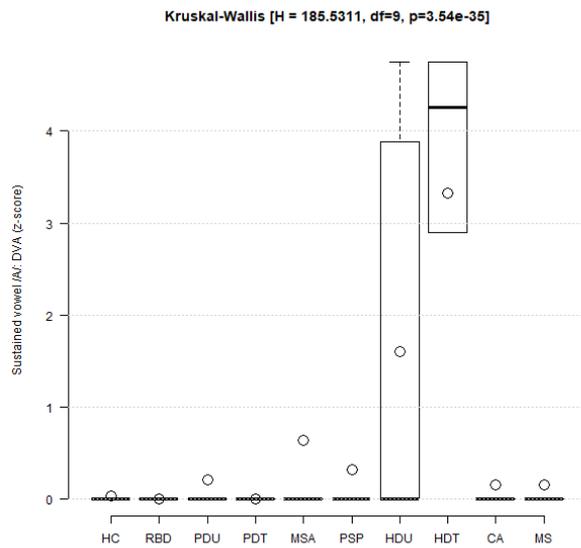
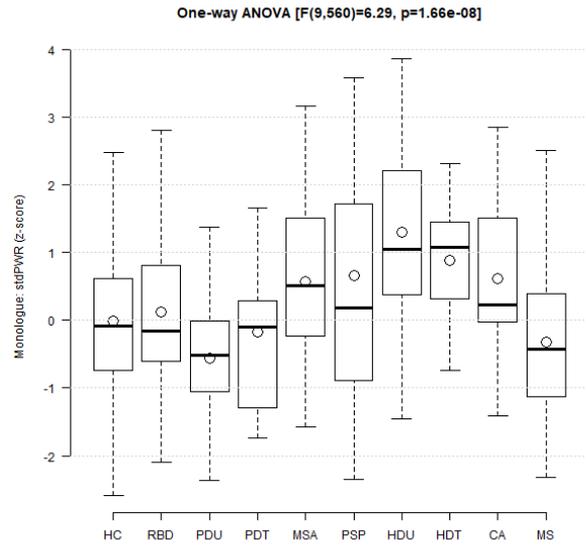
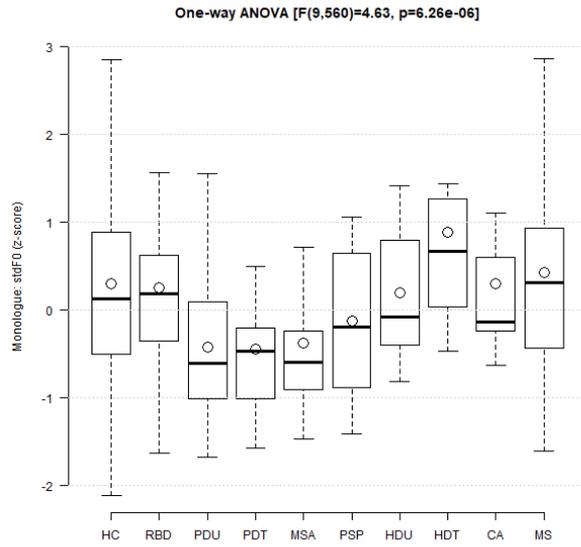
## Appendix B: Normalized values of speech features



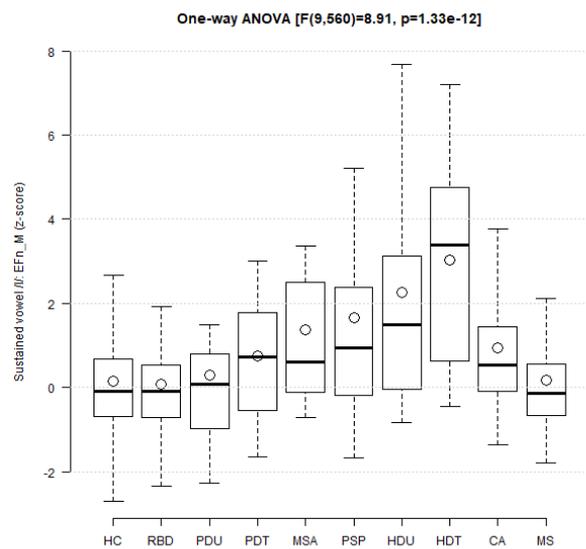
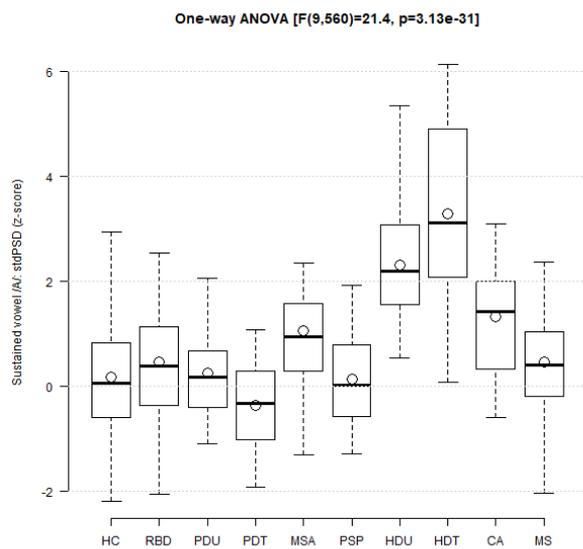
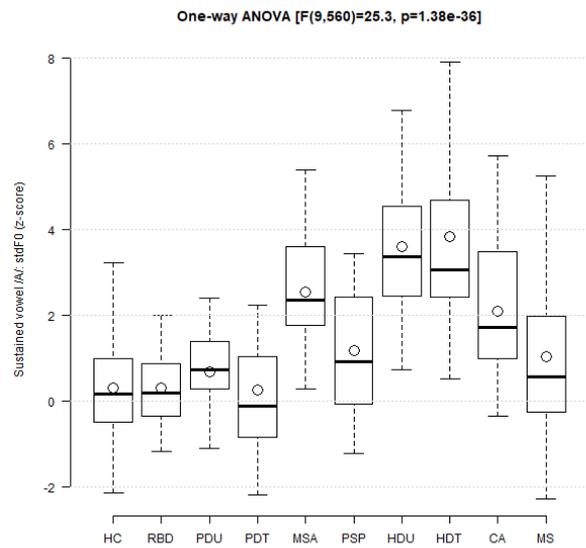
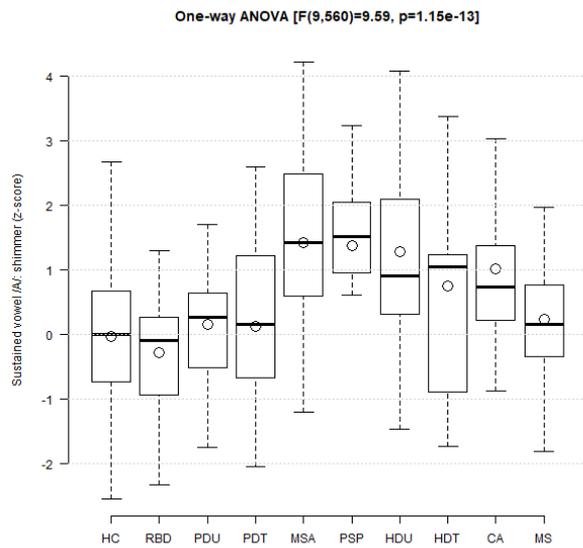
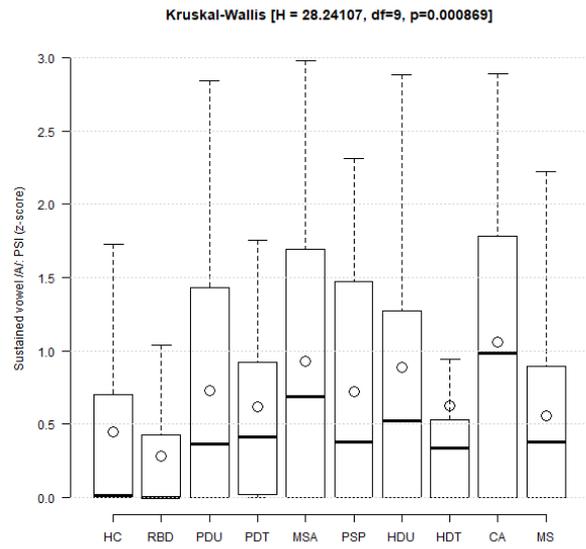
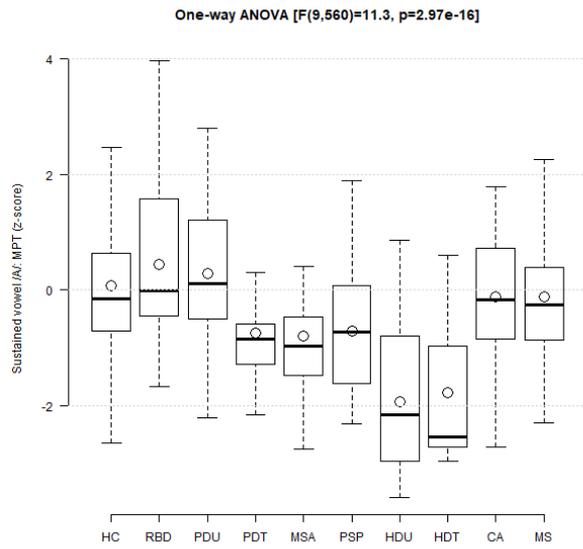
Appendix B: Normalized values of speech features



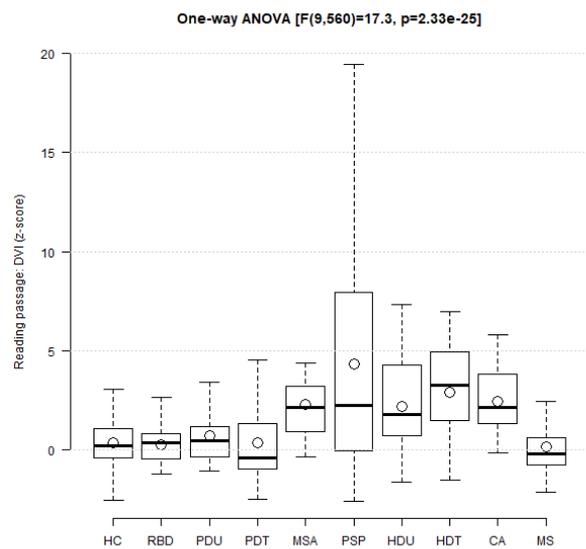
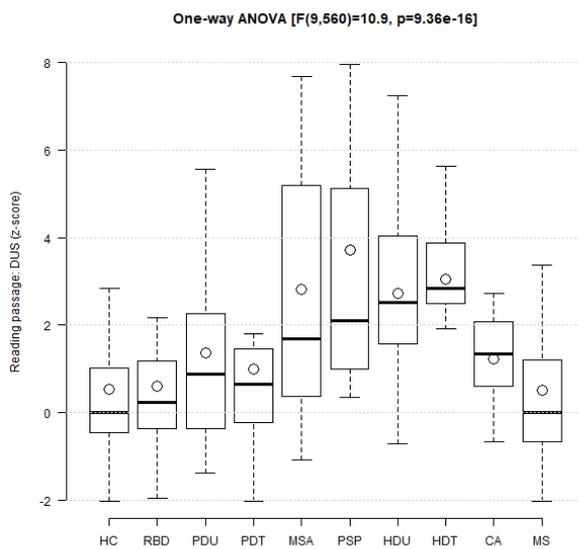
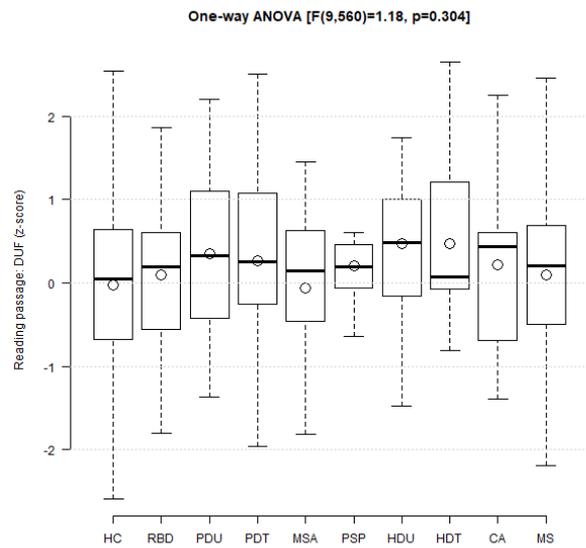
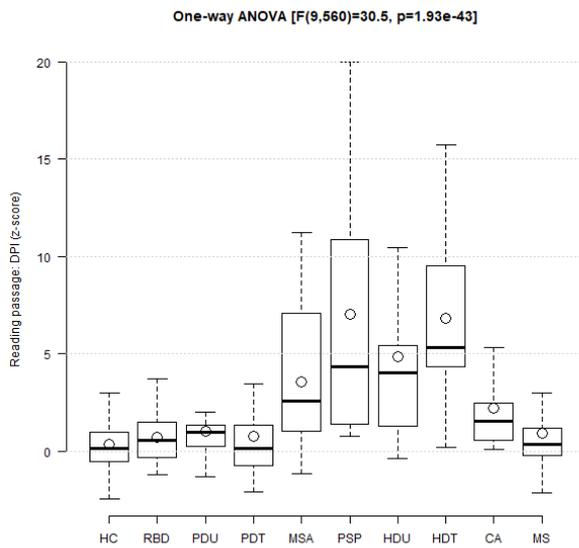
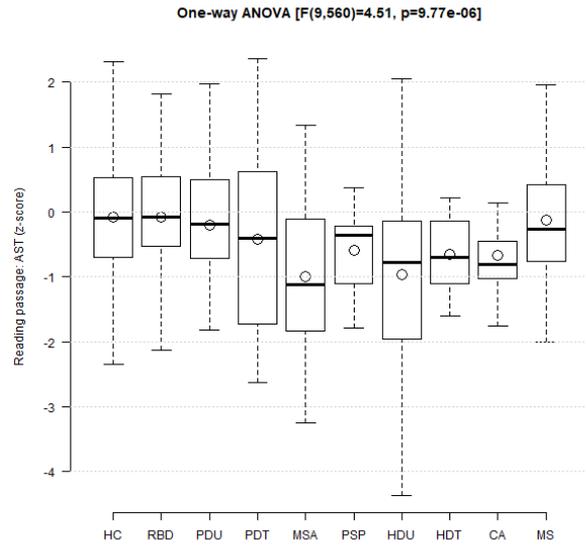
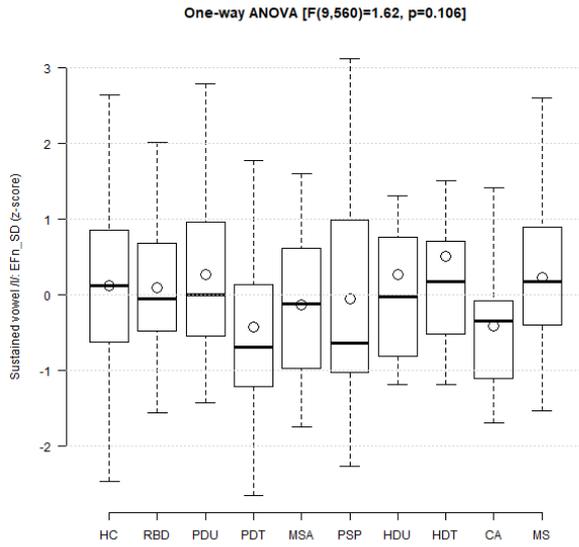
Appendix B: Normalized values of speech features



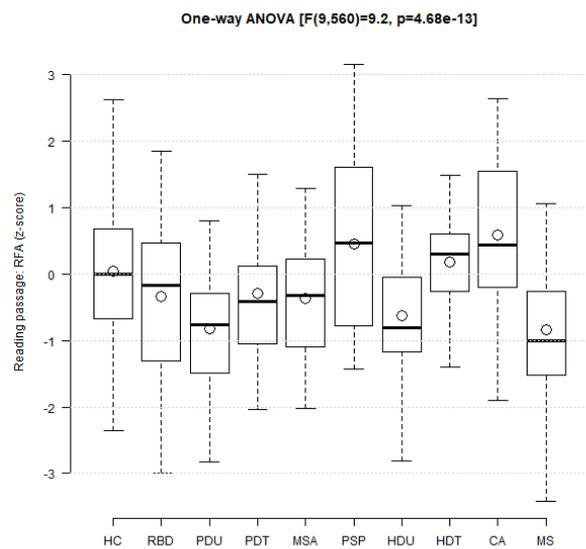
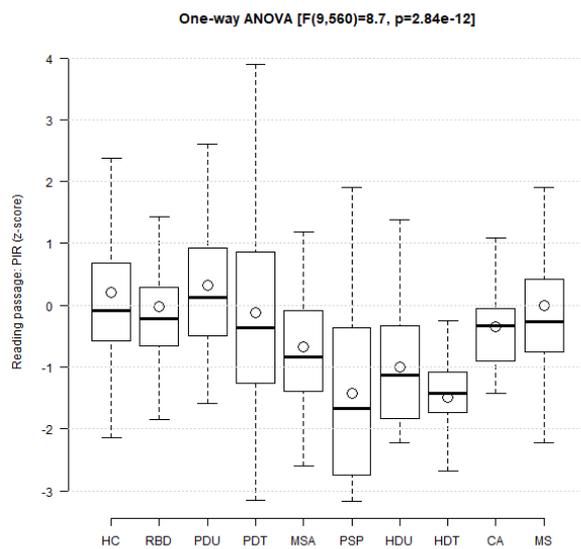
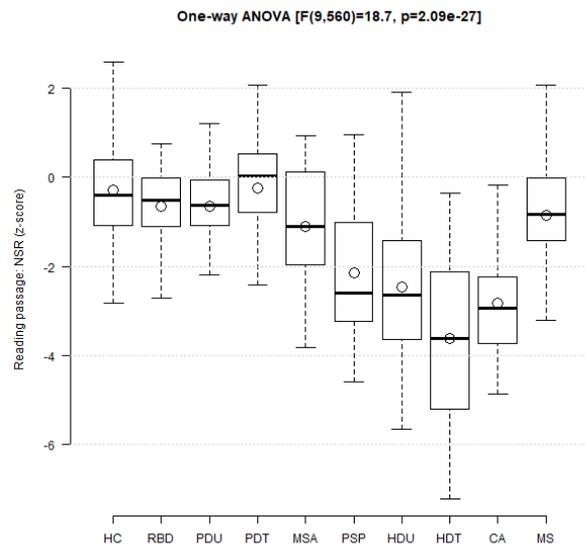
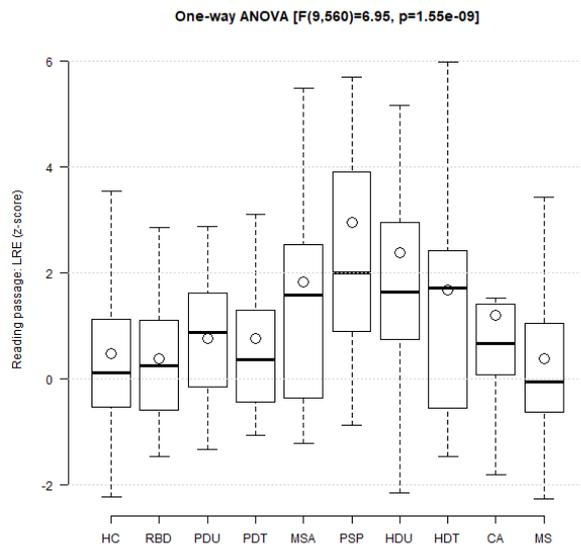
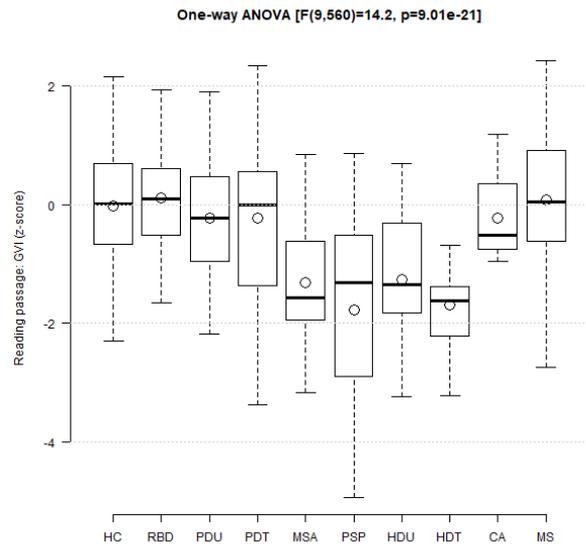
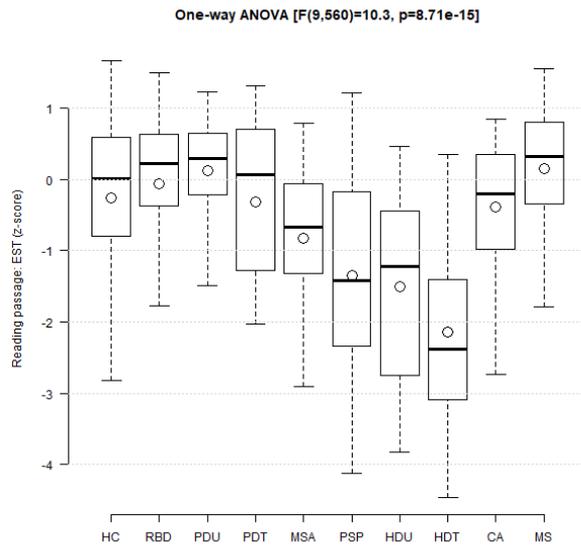
# Appendix B: Normalized values of speech features



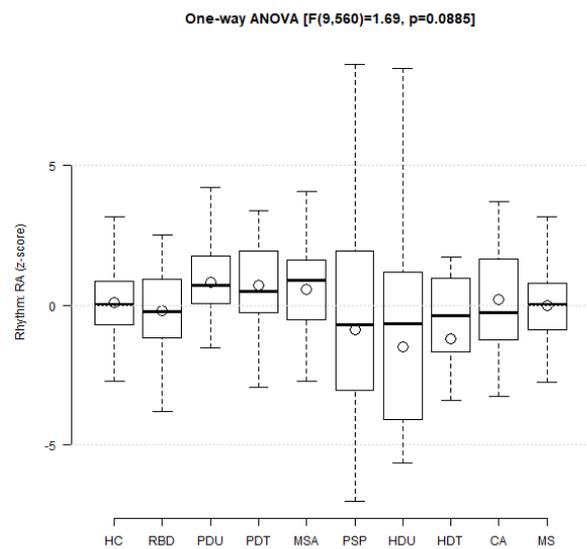
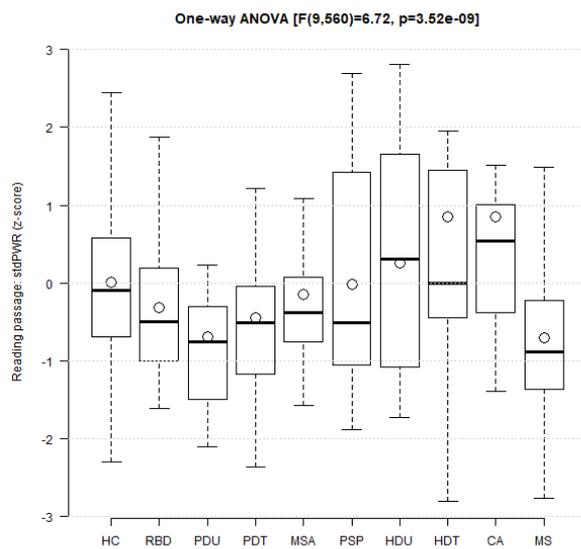
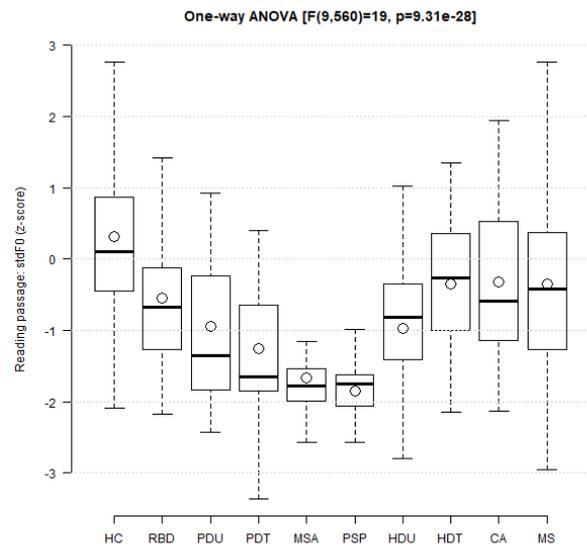
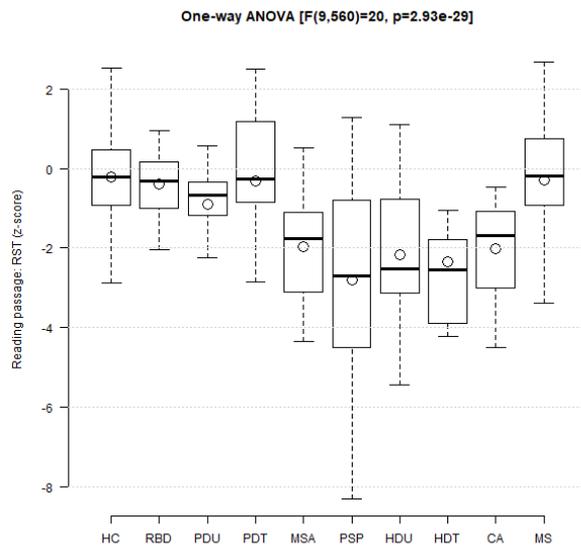
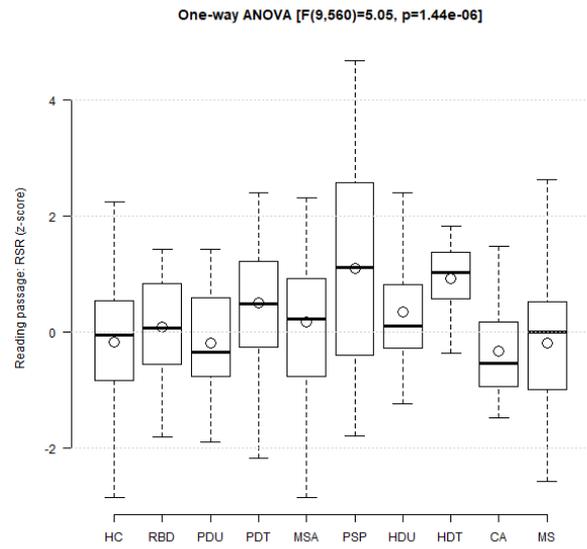
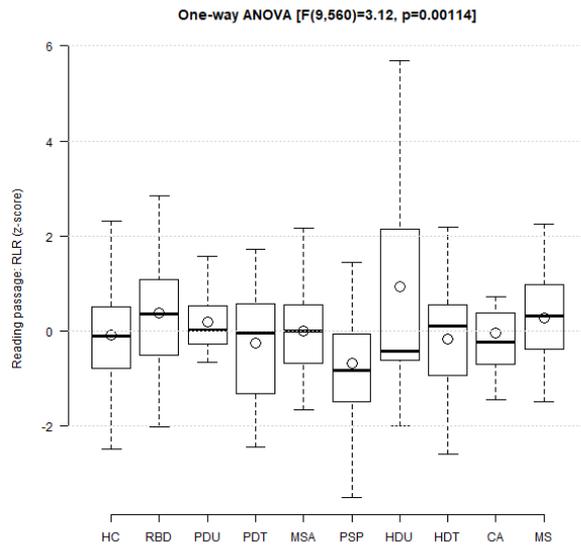
Appendix B: Normalized values of speech features

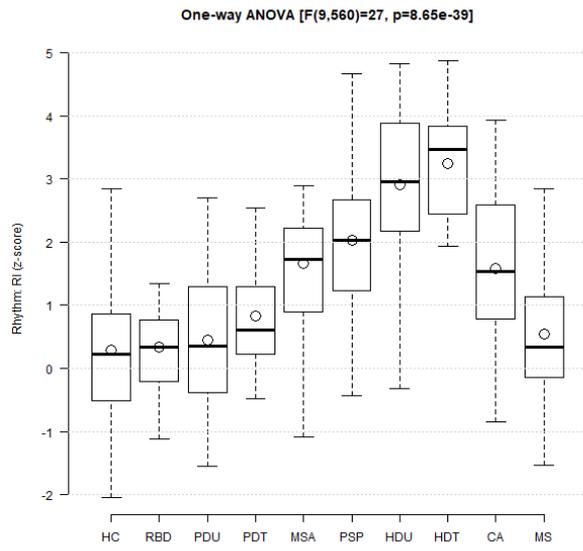


Appendix B: Normalized values of speech features



Appendix B: Normalized values of speech features





# APPENDIX C: SOFTWARE APPLICATION

The minimalistic graphical user interface (GUI) was designed to reduce the demands on the computer literacy of the user. The interactions between the application and user were reduced using the following approach. Since the application requires numerous MATLAB scripts, a batch file (dysan.bat) was coded for setting path to m-files, starting MATLAB and the GUI. All m-files can be saved in other than the working directory (specified by the path in dysan.bat). The working directory is defined by the location of dysan.bat, which can be changed simply by copying dysan.bat into a new location. The application does not require a user to link all relations between individual recordings and task and subjects manually, because the application determines all the relations from locations and names of the recordings automatically. All recordings must be located and named by following standard.

The working directory contains a directory called database. The database directory is divided into subfolders of tasks. The subfolder ddk corresponds to diadochokinetic task, rhythm to rhythm task, phonationA to sustained vowel /A/, phonationI to sustained vowel /I/, monologue to monologue task, and text to reading passage. Each recording can be located only within the subdirectory of the corresponding task. A task of the recording is thus determined by its location within the directory structure. Examples of locations can be found in Table C1.

Each speaker is identified by a unique code in the format “nameNUMBER,” where the name must consist of only alphabetic characters, and NUMBER must only consist of numeric characters<sup>3</sup>. Additionally, a session can be defined numerically in the format “nameNUMBER\_xR,” where x is a numeric character representing the order of the session, and the character ‘R’ is a suffix used for the increased readability of the code. Sessions can also be defined by date in the format “nameNUMBER\_dd.MM.yyyy” or “nameNUMBER\_yyyy-MM-dd,” where dd is the day, MM represents the month, and yyyy is the year. Examples of codes can be found in Table C1. All recordings of a speaker have to be named according to the speaker’s identification code. Multiple repetitions of a task can be distinguished by adding a single alphabetic character to the code as a suffix, e.g., recordings belonging to speaker HC101 could have HC101a, HC101b, and HC101c and so forth under their name, or recordings of speaker HC102\_1R could have names HC102\_1Ra, HC102\_1Rb, HC102\_1Rc and so forth. All recordings that are located in the same subfolder of the task and are named with similar speaker’s identification code but different suffix belong to the same speaker performing multiple repetition of the task. In summary, the code defines the speaker and his session. All recordings of one speaker within the session must have a similar code, whereas only the suffix can vary to identify repetitions. Of course, no suffix is required when the speaker performed the task only once. See table Table C1 for more examples.

When dysan.bat is executed in an otherwise empty working directory, a directory structure of the database, including subdirectories of tasks, is generated automatically. Of course, a user can create the directory structure on their own. The execution dysan.bat will start the GUI illustrated in Figure C1. Since age and gender are required for the exploitation of all features provided by the methodology, the user is allowed to define these characteristics in the table datalist.csv or manually via the GUI. The table datalist.csv can be generated via the button “Generate datalist.csv” in the help dialogue or manually using the following standard. The first column of the table represents the code name of the speaker, and the second column denotes the group— here, any combination of alphanumeric characters are appropriate for entry, e.g., HD or MJ12. Next, the third column specifies gender using the abbreviations M for male and F for female. Finally, the fourth column defines age in years. Currently, the application is limited to only the Macintosh format for csv-files.

The user is allowed to analyze all of the data or individual speakers by selecting an option in the pop-up menu (see “Select subject” in Figure C1). The default option “- all subjects –“ performs an analysis across all of the subjects in the database. The user is not allowed to define speakers’ characteristics via the GUI, and all characteristics are gathered from datalist.csv. All options will show up after clicking on the pop-up menu (see Figure C2), and the user can then select a group for analysis. Speakers with recordings spanning more than one session are marked as longitudinal. If, for example, the user wishes to select an individual speaker, then they would position the cursor over that speaker’s code name, highlighting their choice, and click. Their confirmed choice would then label the pop-up menu after the menu collapses (see Figure C3). The user could then define the age and gender of this subject.

---

<sup>3</sup> The length of a code is not limited by the application. However, the filename must be shorter than 255 characters.

Analysis will begin after user clicks on the “START” button. Then all controls will become unactive with the exception of the “QUIT” button, which interrupts all processes and closes the application. The controls are reactivated after all of the requested results have been generated. Raw values are provided in the table results.csv. When all of the characteristics of a speaker are available, a report based on normalized values is generated and placed into an html-file under the code of the analyzed subject. When a speaker was recorded in multiple sessions but no characteristics are known, the report is based on just the raw values of the speech features. No report is generated for subjects with unknown characteristics recorded within one session.

The application saves data into a temporary folder “./var” that is created after the start of an analysis and removed after quitting an application.

| File or folder  | Description   |
|---|---|
| <b>The application’s files and folders</b>  |   |
| ./dysan.bat   | Batch file starts MATLAB, set paths, and starts graphical interface   |
| ./datalist.csv  | Table of speakers’ characteristics (optional).  |
| ./var/  | Temporary directory used during processing and removed after application exit.  |
| <b>The database paths</b>   |   |
| ./database  | Directory of database   |
| ./database/text/  | Subdirectory for recordings of reading the passage  |
| ./database/monologue/   | Subdirectory for recordings of monologue  |
| ./database/rhythm/  | Subdirectory for recordings of rhythm task  |
| ./database/ddk/   | Subdirectory for recordings of diadochokinetic task   |
| ./database/phonationA/  | Subdirectory for recordings of sustained vowel /A/  |
| ./database/phonationI/  | Subdirectory for recordings of sustained vowel /I/  |
| <b>Examples of recordings</b>   |   |
| <i>Recordings of various tasks by HC101</i>   |   |
| ./database/rhythm/HC101.wav   | Example of the single performance of rhythm task by subject HC101. Note that suffix is optional in this case since single performance does not need to be distinguished.  |
| ./database/phonationI/HC101.wav   | Example of the single performance of sustained vowel /I/ by subject HC101.  |
| ./database/ddk/HC101a.wav   | Example of the first repetition of diadochokinetic task by subject HC101. Note that suffix was required to identify repetition (cf. previous example of single performance of rhythm task and reading passage). |
| ./database/ddk/HC101b.wav   | Example of the second repetition of diadochokinetic task by subject HC101.  |
| ./database/ddk/HC101c.wav   | Example of the third repetition of diadochokinetic task by subject HC101.   |
| <i>Various tasks by HC205 recorded within session 1R</i>  |   |
| ./database/rhythm/HC205_1Ra.wav   | Example of first repetition of rhythm by subject HC205 recorded within session 1R   |
| ./database/rhythm/HC205_1Rb.wav   | Example of first repetition of rhythm by subject HC205 recorded within session 1R   |
| ./database/rhythm/HC205_1Rc.wav   | Example of first repetition of rhythm by subject HC205 recorded within session 1R   |
| ./database/text/HC205_2R.wav  | Example of single performance of reading passage by subject HC205 recorded within session 1R. Note that suffix is not required to identify single performance of the task.                                      |
| <i>Various tasks by HC205 recorded within session 2R</i>  |   |
| ./database/rhythm/HC205_2Ra.wav   | Example of first repetition of rhythm by subject HC205 recorded within session 1R   |
| ./database/rhythm/HC205_2Rb.wav   | Example of second repetition of rhythm by subject HC205 recorded within session 1R  |
| ./database/monologue/HC205_2Ra.wav  | Example of first repetition of sustained vowel /I/ by subject HC205 recorded within session 1R  |
| ./database/monologue/HC205_2Rb.wav  | Example of second repetition of sustained vowel /I/ by subject HC205 recorded within session 1R   |
| <i>Various tasks by MSA423 recorded on Christmas Eve 2018 with date specified using yyyy-MM-dd format</i> |   |
| ./database/ddk/MSA423_2018-12-24.wav  | Example of single performance of diadochokinetic task by subject MSA102 recorded on Christmas Eve 2018. Note that suffix is not required in this case.  |
| ./database/phonationA/MSA423_2018-12-24a.wav  | Example of first repetition of reading passage by subject MSA102 recorded on Christmas Eve 2018   |
| ./database/phonationA/MSA423_2018-12-24b.wav  | Example of second repetition of reading passage by subject MSA102 recorded on Christmas Eve 2018  |
| <i>Various tasks by PD534 recorded on Christmas Eve 2018 with date specified using dd.MM.yyyy format</i>  |   |
| ./database/ddk/PD534_24.12.2018.wav   | Example of single performance of diadochokinetic task by subject PD534 recorded on Christmas Eve 2018. Note that suffix is not required in this case.   |
| ./database/phonationA/PD534_24.12.2018a.wav   | Example of first repetition of reading passage by subject PD534 recorded on Christmas Eve 2018  |
| ./database/phonationA/PD534_24.12.2018b.wav   | Example of second repetition of reading passage by subject PD534 recorded on Christmas Eve 2018   |

Table C1: Summary of files and folders within the working directory.  
Working directory is symbolized by dot/slash (./).

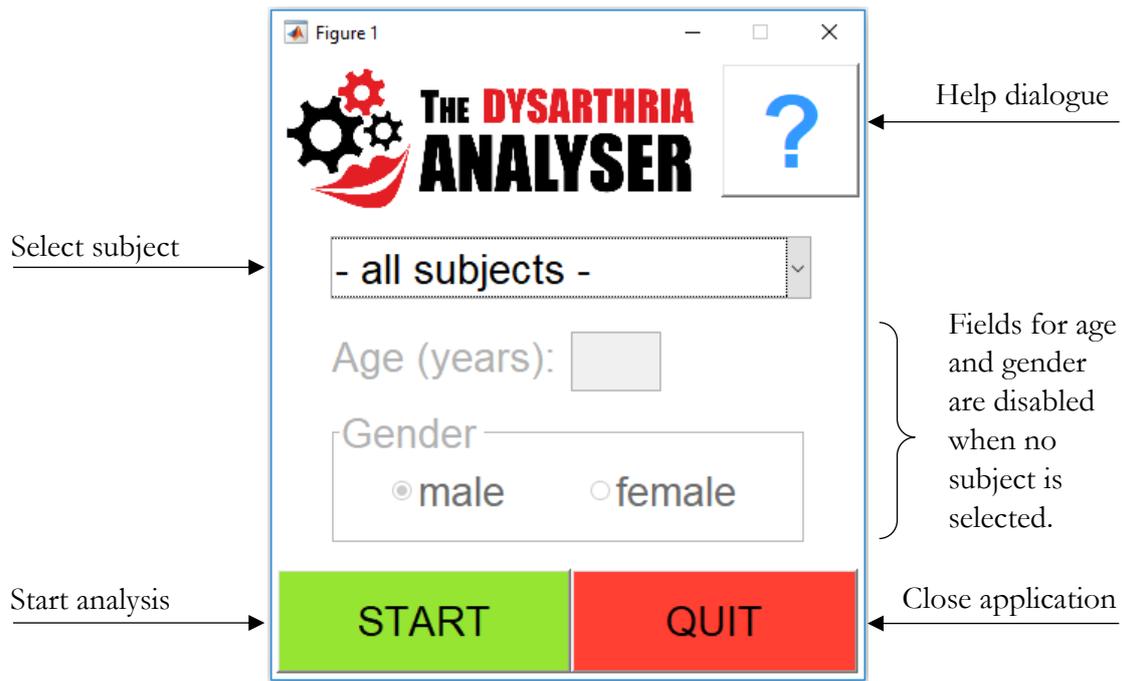


Figure C1: Screenshot of the application after start.

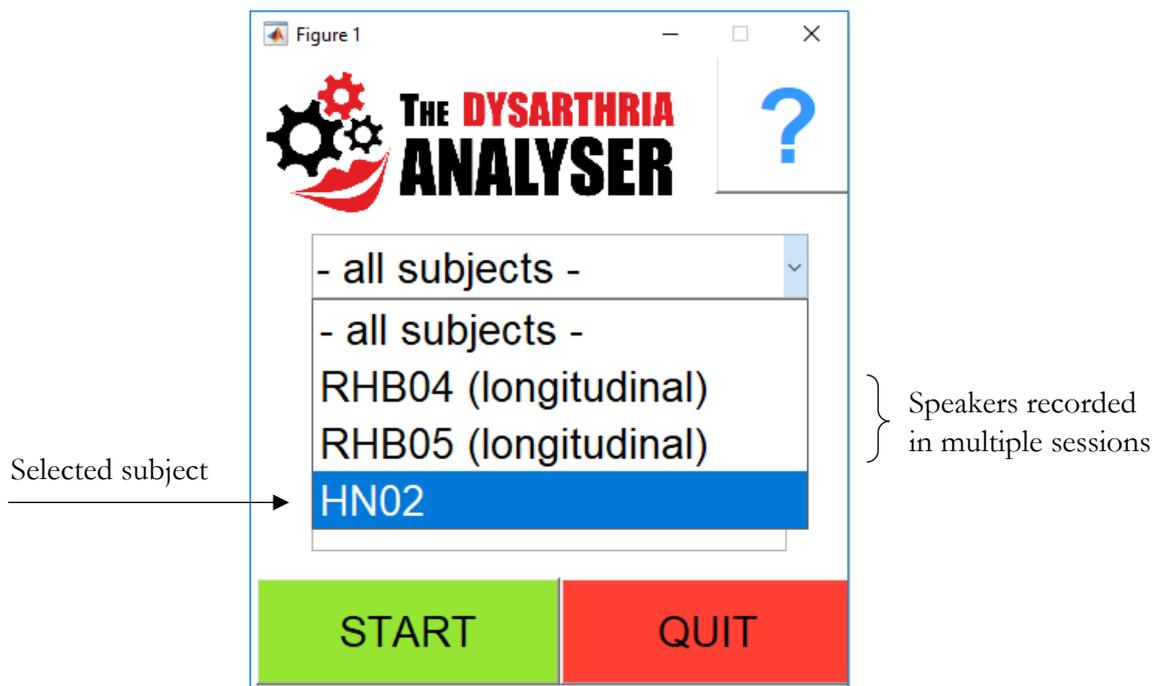


Figure C2: Screenshot of the application during the selection of a subject.

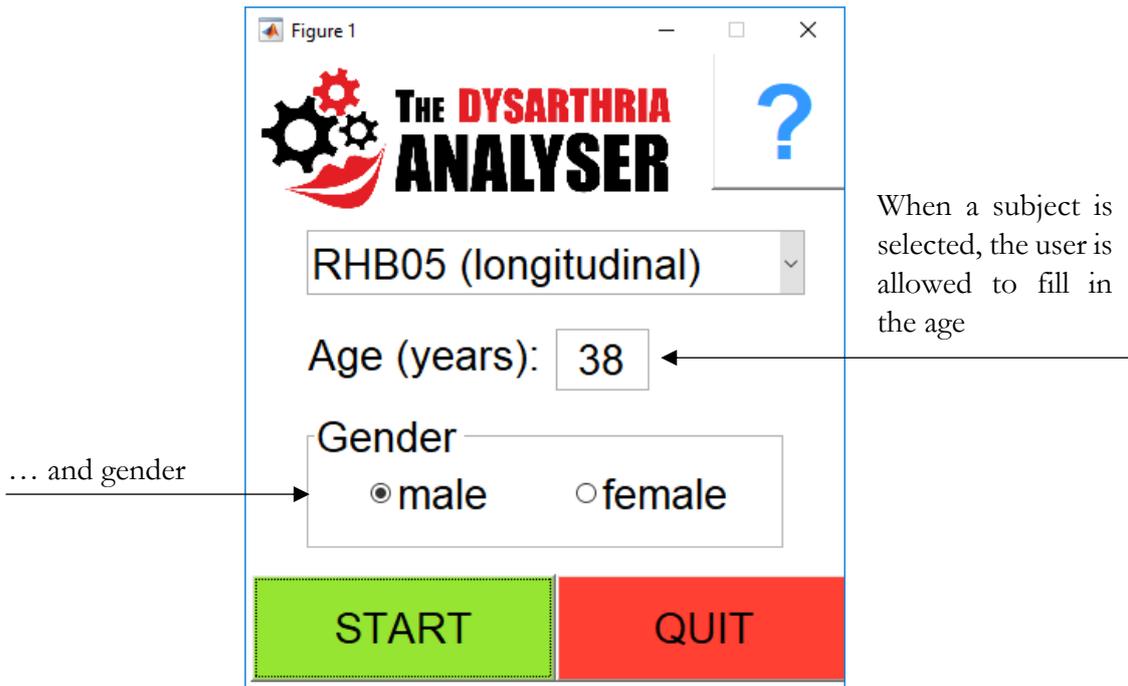


Figure C3: Screenshot of the application after the selection of the subject. Note that the fields for age and gender are activated.

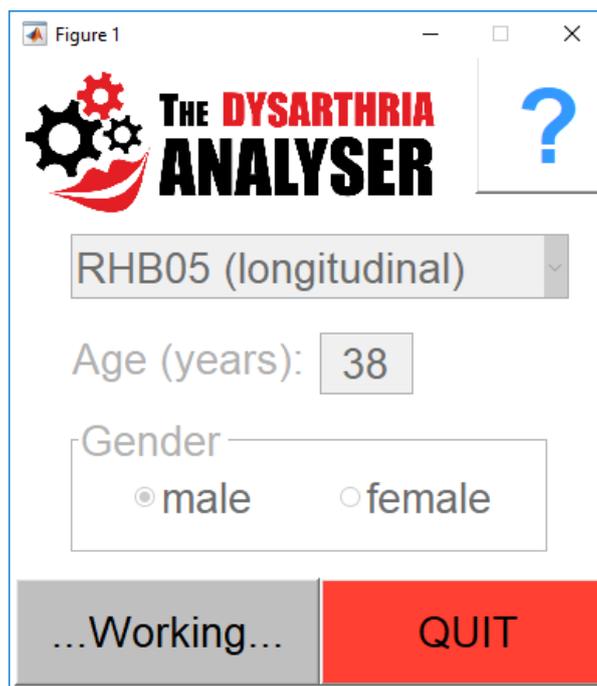


Figure C4: Screenshot of the application during the processing of the data. Notice that all control elements are inactive except the "Quit" button.



# APPENDIX D: QUESTIONNAIRE FEEDBACK

The filled in questionnaire is presented on the following pages.

Name: Hana Růžičková

Date: 22.9.2018

Occupation: Speech pathologist

## Mark your answer

| Question   | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 | ? | Detailed answer (optional)   |
|--|----|----|----|----|----|---|---|---|---|---|---|---|--|
| 1. How easy was it to perform acoustic analysis? Do not consider the manual trimming of recordings. (higher score=easier)              |    |    |    |    |    |   |   |   |   |   | 5 |   | The software was user friendly—only error I made was in renaming recordings, i.e., manual trimming, everything else was OK. I highly appreciate the suggested descriptions and interpretations of parameters.  |
| 2. Was the application a burden on your attention? Do not consider the manual trimming of recordings. (lower score =more burden)       |    |    |    |    |    |   |   |   |   | 4 |   |   | Results were not available immediately, because it was not possible to trim recordings during the session with the patient. Consequently, I had problems remembering results and explaining them to my patient during the next session. In other words, the problem was that the days for examination and acoustic analysis differed due to the manual trimming of recordings, which distracted my attention. I can write notes, of course, but that takes too much time and can hardly be performed regularly in clinical practice. To be honest, I cannot answer this question properly, but the software brings so many benefits that it is worthy despite this limitation. |
| 3. Did you find any features that made the analysis time consuming? If any, comment on them, please. (lower score=more time consuming) |    |    |    |    |    |   |   |   | 3 |   |   |   | The interpretation of complex pause characteristics was too difficult for me—phonation, respiration, and phrasing have considerable influence on the result—I have to inspect various pause characteristics in different locations within the diagrams when I want to consider pause characteristics as a whole. Despite my willingness, I still have not memorize all of the abbreviations of the acoustic features. Therefore, I have to consult the table, which slows me down.   |
| 4. Do you appreciate that the analysis for individual recordings was fully automated?  |    |    |    |    |    |   |   |   |   |   | 5 |   | Yes, considerably.   |
| 5. Do you think that the proposed analyses meet the requirements of clinical practice? Consider function, simplicity, etc.             |    |    |    |    |    |   |   |   |   | 4 |   |   | I find the software to be very beneficial, but disproportional—prosodic features (pause, melody) are overly empathized. There are not many features correlating with facial movements or dysfluency (palilalia, hesitation phenomena, saccades). Once, I had a problem with the microphone. A few parts of the signal were noisy, and I lost one recording completely (it was even more troublesome, as I did not make notes during the session). Some indication of a problem with the recording system would be very helpful to clinicians.  |

Mark your answer

| Question  | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 | ? | Detailed answer (optional)   |
|---|----|----|----|----|----|---|---|---|---|---|---|---|--|
| 6. Mark how much the analyses provide results in concert with your clinical judgement. (higher score -> more agreeable)                   |    |    |    |    |    |   |   |   |   | 4 |   |   | Yes, for majority of cases—especially, diseases upon which the application was developed. Diseases that were not validated for the application showed some anomalies.  |
| 7. Did the analyses highlight any overlooked aspect of speech disorders?  |    |    |    |    |    |   |   |   |   | 4 |   |   | Entropy of speech timing is quite difficult for me to understand and interpret—the parameter is very complex.  |
| 8. Do you think that the application can help to address critical issues in speech disorders and track the progression of speech therapy? |    |    |    |    |    |   |   |   |   |   | 5 |   | Yes, considerably.   |
| 9. Did you find that the analyses could provide a quick and reliable summary of trends in a speech disorder?                              |    |    |    |    |    |   |   |   | 3 |   |   |   | I do not find it quick enough—clinical practice require that recordings be analyzed on the same day, which is not easy due to the manual trimming of recordings. The feedback in speech therapy is limited then.   |
| 10. Do you think that the use of the application could support or improve diagnostic decisions?   |    |    |    |    |    |   |   |   | 3 | 4 |   |   | I propose a score between 3 and 4. The proposed software is definitely applicable. I regard it only as a supporting tool because diagnosis requires one to consider other factors, such as socioeconomic status, mental complications due to diagnosis, etc., that can influence many speech dimensions, especially prosody. |

Mark your answer

| Question   | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 | ? | Detailed answer (optional)  |
|--|----|----|----|----|----|---|---|---|---|---|---|---|---|
| 11. Did the visualization help you to gain a comprehensible insight into speech disabilities?  |    |    |    |    |    |   |   |   |   |   | 5 |   | Yes, considerably. I really appreciate it.  |
| 12. Did the visual symbolism, e.g., red / green color and rounded / cornered shape, make it easier to read results?  |    |    |    |    |    |   |   |   |   |   | 5 |   | The orientation was perfect on the monitor. Diagrams are less readable when printed in black and white (an option for black-and-white printing would be very helpful).  |
| 13. Do you find the normalized values, i.e., z-scores and probabilities, more interpretable than the raw values of speech features?                                |    |    |    |    |    |   |   |   |   |   | 5 |   | Yes, I find it more convenient.   |
| 14. Did the interpretation suggested by the application show a clinical validity?  |    |    |    |    |    |   |   |   |   | 4 |   |   | Mostly yes—when I did not agreed, I considered the anamnesis of the patient; thus, deviations from the assumed model.   |
| 15. Did the longitudinal graphs of the normalized results illustrate the course of progression or effect of therapy in a clearer way than a reading of raw values? |    |    |    |    |    |   |   |   |   | 4 |   |   | Yes—it is more illustrative and quick. Everybody can understand it very well. Problems arise occasionally when a patient shows an abnormality outside a given range. Then the trend is clear and the table must be read, which can be time consuming. |

Mark your answer

| Question  | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 | ? | Detailed answer (optional)  |
|---|----|----|----|----|----|---|---|---|---|---|---|---|---|
| 16. Was the use of application more beneficial than time consuming?   |    |    |    |    |    |   |   |   | 3 |   |   |   | I record the patient in one room, trim the recordings in the second room, and sometimes analyze in another room, which is time demanding. Such a complex situation happens in my clinical practice and can be generally expectable. |
| 17. Do you find the application valuable for tracking speech quality over time? Consider the encouraging effect on a patient.                                       |    |    |    |    |    |   |   |   |   |   | 5 |   | It is incredibly helpful and objective feedback for me as a clinician and for the patient (motivation).   |
| 18. Does the application include new speech features that can improve your diagnostic / treatment decision?   |    |    |    |    |    |   |   |   |   |   | 5 |   | Previously, I classified subharmonics as hoarseness. Thanks to the application, I can recognize them as a special symptom and make my diagnosis more detailed.  |
| 19. Does the application provide any speech feature that have no equivalent in terms of auditory-perception and that can enrich your insight into speech disorders? |    |    |    |    |    |   |   |   | 3 |   |   |   | I still have problems understanding entropy.  |
| 20. Did an overview provided by the application allow you to focus on individual aspects of speech in more details than you usually do?                             |    |    |    |    |    |   |   |   |   | 4 |   |   | Yes and no—I intend to reflect on a patient’s preferences and priorities in addition to the objective outcome of the acoustic analysis.   |

Mark your answer

| Question  | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 | ? | Detailed answer (optional)   |
|---|----|----|----|----|----|---|---|---|---|---|---|---|--|
| 21. Would you prefer representation of results other than through a probability / z-score? If so, please specify your choice.   |    |    |    |    |    |   |   |   |   |   | 5 |   | No, I find the current representation to be suitable.  |
| 22. Would you appreciate an indication of a mild / moderate / severed speech abnormality rather than the normal / rare indication used in the application?  |    |    |    |    |    | 0 |   |   |   |   |   |   | I do not mind.   |
| 23. Do you think that integrated recording or automated recognition of tasks would make the application more attractive to clinicians compared to the time-consuming manual trimming used in the experimental version?                      |    |    |    |    |    |   |   |   |   |   | 5 |   | Yes, absolutely.   |
| 24. Would you appreciate a more supervised approach, such as user-controlled and being able to adjust critical parts of the analytical process, i.e., manual selection of the analyzed signal?  |    |    |    |    |    |   |   |   |   | 4 |   |   | It seems attractive to me—only if I would be able to organize it. I would also some training to use it, of course. |
| 25. Do you find the classification excitatory vs. inhibitory too limiting for a clinical application or did you find some convenience and applicability in it? If you are interested in no or another classification, please comment on it. |    |    |    |    |    |   |   |   |   |   | 5 |   | Yes, I find it suitable.   |

# REFERENCES

- American Academy of Sleep Medicine. (2014). International classification of sleep disorders—third edition (ICSD-3). Darien, IL: American Academy of Sleep Medicine.
- American Speech-Language-Hearing Association (2018). What to ask when evaluating any procedure, product, or program. Retrieved online from <https://www.asha.org/slp/evaluate/>
- Auzou, P., Ozsancak, C., Jan, M., Menard, J.F., Eustache, F., & Hannequin, D. (2000). Evaluation of motor speech function in the diagnosis of various forms of dysarthria. *Revue Neurologique*, 156, 47-52.
- Baken, R.J. (1990). Irregularity of vocal period and amplitude: A first approach to the fractal analysis of voice. *Journal of Voice*, 4, 185-197.
- Bandini, A., Giovannelli, F., Orlandi, S., Barbagallo, S.D., Cincotta, M., Vanni, P., Chiaramonti, R., Borgheresi, A., Zaccara, G., & Manfredi, C. (2015). Automatic identification of dysprosody in idiopathic Parkinson's disease. *Biomedical Signal Processing and Control*, 17, 47-54.
- Benesty, J., Sondhi, M.M., & Huang, Y. (Eds.). (2007). *Springer handbook of speech processing*. Heidelberg, Germany: Springer-Verlag.
- Bergan, C.C., & Titze, I.R. (2001). Perception of pitch and roughness in vocal signals with subharmonics. *Journal of Voice*, 15, 165-175.
- Berisha, V., Liss, J., Huston, T., Wisler, A., Jiao, Y., & Eig, J. (2017). Float like a butterfly, sting like a bee: Changes in speech preceded Parkinsonism diagnosis for Muhammad Ali. *Proceedings of the Interspeech 2017: Situated interaction*, Stockholm, Sweden (pp. 1809-1813). Adelaide, Australia: Causal Productions.
- Bielamowicz, S., Kreiman, J., Gerratt, B.R., Dauer, M.S., & Berke, G.S. (1996). Comparison of voice analysis systems for perturbation measurement. *Journal of Speech, Language, and Hearing Research*, 39, 126-134.
- Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences* 17, 97-110.
- Boersma, P. & Weenink, D. (2018). Praat: Doing phonetics by computer [Computer program]. Version 5.3.51, retrieved 20 February 2018 from <http://www.praat.org/>
- Brendel, B., Synofzik, M., Ackermann, H., Lindig, T., Schölderle, T., Schöls, L., & Ziegler, W. (2015). Comparing speech characteristics in spinocerebellar ataxias type 3 and type 6 with Friedreich ataxia. *Journal of Neurology*, 262, 21-26.
- Burris, C., Vorperian, H.K., Fourakis, M., Kent, R.D., & Center, W. (2011). *Acoustic analysis software: A quantitative and qualitative comparison of four systems* (Doctoral dissertation). Madison, Wisconsin: University of Wisconsin.
- Camacho, A. (2007). *SWIPE: A sawtooth waveform inspired pitch estimator for speech and music*. Gainesville, Florida: University of Florida.

- Camacho, A., & Harris, J.G. (2008). A sawtooth waveform inspired pitch estimator for speech and music. *The Journal of the Acoustical Society of America*, 124, 1638-1652.
- Compston, A., & Coles, A. (2008). Multiple sclerosis. *Lancet*, 372, 1502–1517.
- Čmejla, R., & Sovka, P. (2001). Estimation of boundaries between speech units using Bayesian changepoint detectors. *Proceedings of the 4th International Conference on Text, Speech and Dialogue, Železná ruda, Czech Republic* (pp. 291-298). Heidelberg, Germany: Springer-Verlag.
- Čmejla, R., & Sovka, P. (2004). Recursive Bayesian autoregressive changepoint detector for sequential signal segmentation. *Proceedings of the 12th European Signal Processing Conference EUSIPCO-2004, Vienna, Austria* (pp. 245-248). Tampere, Finland: Suvisoft.
- Darley, F.L., Aronson, A.E., & Brown, J.R. (1969a). Differential diagnostic patterns of dysarthria. *Journal of Speech, Language, and Hearing Research*, 12, 246-269.
- Darley, F.L., Aronson, A.E., & Brown, J.R. (1969b). Clusters of deviant speech dimensions in the dysarthrias. *Journal of Speech, Language, and Hearing Research*, 12, 462-496.
- Deliyski, D.D. (1993). Acoustic model and evaluation of pathological voice production. *Proceedings of the 3rd European Conference on Speech Communication and Technology EUROSPEECH'93, Berlin* (pp. 1969-1972). Grenoble, France: International Speech and Communication Association.
- Duffy, J.R. (2013). *Motor speech disorders: Substrates, differential diagnosis, and management*. St. Luis, Missouri: Mosby.
- Dworkin, J.P. (1991). *Motor speech disorders: A treatment guide*. St. Louis, Missouri: Mosby-Year Book, Inc.
- Edwards, L., & Veale, M. (2018). Enslaving the algorithm: From a “Right to an explanation” to a “Right to better decisions”? *IEEE Security & Privacy*, 16, 46-54.
- Fearnley, J.M., & Lees, A. J. (1991). Ageing and Parkinson's disease: Substantia nigra regional selectivity. *Brain*, 114, 2283-2301.
- Fletcher, S.G. (1972). Time-by-count measurement of diadochokinetic syllable rate. *Journal of Speech, Language, and Hearing Research*, 15, 763-770.
- Freed, D. (2011). *Motor speech disorders: Diagnosis & treatment*. Boston, Massachusetts: Cengage.
- Fonville, S., Van Der Worp, H.B., Maat, P., Aldenhoven, M., Algra, A., & Van Gijn, J. (2008). Accuracy and inter-observer variation in the classification of dysarthria from speech recordings. *Journal of Neurology*, 255, 1545-1548.
- Gao, X., Alvo, M., Chen, J., & Li, G. (2008). Nonparametric multiple comparison procedures for unbalanced one-way factorial designs. *Journal of Statistical Planning and Inference*, 138, 2574-2591.
- García, M. J.V., Cobeta, I., Martín, G., Alonso-Navarro, H., & Jimenez-Jimenez, F.J. (2011). Acoustic analysis of voice in Huntington's disease patients. *Journal of Voice*, 25, 208-217.
- Gasior, M., and Gonzalez, J.L. (2004). Improving FFT frequency measurement resolution by parabolic and Gaussian interpolation (CERN-AB-Note-2004-021). Geneva, Switzerland: European Council for Nuclear Research.

- Gerratt, B.R., Till, J.A., Rosenbek, J.C., Wertz, R.T., Boysen, A.E., Moore, C.A., Yorkston, K.M. and Beukelman, D.R. (1991). Use and perceived value of perceptual and instrumental measures in dysarthria management. In Moore, C.A., Yorkston, K.M. & Beukelman, D.R. (Eds.), *Dysarthria and Apraxia of Speech: Perspectives on Management*. Baltimore: Paul H. Brookes.
- Gilman, S., Wenning, G.K., Low, P.A., Brooks, D.J., Mathias, C.J., Trojanowski, J.Q., Wood, N.W., Colosimo, C., Dürr, A., Fowler, C.J., & Kaufmann, H. (2008). Second consensus statement on the diagnosis of multiple system atrophy. *Neurology*, 71, 670-676.
- Gonzalez, S., & Brookes, M. (2014). PEFAC-a pitch estimation algorithm robust to high levels of noise. *IEEE/ACM Transactions on Audio, Speech and Language Processing TASLP*, 22, New Your, U.S.A. (pp. 518-530). Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.
- Goodman, B., & Flaxman, S. (2016). EU regulations on algorithmic decision-making and a “right to explanation”. *AI Magazine*, 38, arXiv:1606.08813.
- Hansen, J.H., Gray, S.S., & Kim, W. (2010). Automatic voice onset time detection for unvoiced stops (/p/,/t/,/k/) with application to accent classification. *Speech Communication*, 52, 777-789.
- Harel, B.T., Cannizzaro, M.S., Cohen, H., Reilly, N., & Snyder, P. J. (2004). Acoustic characteristics of Parkinsonian speech: A potential biomarker of early disease progression and treatment. *Journal of Neurolinguistics*, 17, 439-453.
- Hariharan, M., Polat, K., & Sindhu, R. (2014). A new hybrid intelligent system for accurate detection of Parkinson's disease. *Computer Methods and Programs in Biomedicine*, 113, 904-913.
- Hartelius, L., Carlstedt, A., Ytterberg, M., Lillvik, M., & Laakso, K. (2003). Speech disorders in mild and moderate Huntington disease: Results of dysarthria assessments of 19 individuals. *Journal of Medical Speech-Language Pathology*, 11, 1-15.
- Hermes, D. J. (1988). Measurement of pitch by subharmonic summation. *The Journal of the Acoustical Society of America*, 83, 257-264.
- Hess, W. (1983). *Pitch determination of speech signals: Algorithms and devices*, 3. Heidelberg, Germany: Springer-Verlag.
- Hillenbrand, J., & Houde, R. A. (1996). Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *Journal of Speech, Language, and Hearing Research*, 39, 311-3
- Hlavnička, J., Čmejla, R., Tykalová, T., Šonka, K., Růžička, E., & Ruzs, J. (2017a). Automated analysis of connected speech reveals early biomarkers of Parkinson's disease in patients with rapid eye movement sleep behaviour disorder. *Scientific Reports*, 7, 12.
- Hlavnička, J., Tykalová, T., Čmejla, R., Klempíř, J., Růžička, E., & Ruzs, J. (2017b). Dysprosody differentiate between Parkinson's disease, progressive supranuclear palsy, and multiple system atrophy. *Proceedings of the Interspeech 2017: Situated interaction*, Stockholm, Sweden (pp. 1844-1848). Adelaide, Australia: Causal Productions.
- Hlavnička, J., Čmejla, R., Klempíř, J., Růžička, E., & Ruzs, J. (2019). Acoustic tracking of pitch, modal, and subharmonic vibrations of vocal folds in Parkinsonism. Manuscript submitted for publication.

- Ho, A. K., Iannsek, R., Marigliani, C., Bradshaw, J. L., & Gates, S. (1998). Speech impairment in a large sample of patients with Parkinson's disease. *Behavioural neurology*, 11, 131-137.
- Hughes, A.J., Daniel, S.E., Kilford, L., & Lees, A. J. (1992). Accuracy of clinical diagnosis of idiopathic Parkinson's disease: A clinico-pathological study of 100 cases. *Journal of Neurology, Neurosurgery & Psychiatry*, 55, 181-184.
- Huntington Study Group. (1996). Unified Huntington's disease rating scale: Reliability and consistency. *Movement Disorders*, 11, 136-142.
- International Telecommunication Union. Standardization sector of ITU (1996). ITU-T G.729, WTSC-96, Geneva, Switzerland, 1-39.
- Interspeech 2012 (September 9-13, 2012). The 13th Annual Conference of the International Speech Communication Association. Portland, Oregon.
- Interspeech 2015 (September 6-10, 2015). The 16th Annual Conference of the International Speech Communication Association. Dresden, Germany.
- Iranzo, A., Molinuevo, J.L., Santamaría, J., Serradell, M., Martí, M.J., Valldeoriola, F., & Tolosa, E. (2006). Rapid-eye-movement sleep behaviour disorder as an early marker for a neurodegenerative disorder: A descriptive study. *The Lancet Neurology*, 5, 572-577.
- Iranzo, A., Fernández-Arcos, A., Tolosa, E., Serradell, M., Molinuevo, J.L., Valldeoriola, F., Gelpi, E., Vilaseca, I., Sánchez-Valle, R., Lladó, A., & Gaig, C. (2014). Neurodegenerative disorder risk in idiopathic REM sleep behavior disorder: Study in 174 patients. *PLoS One*, 9, e89741.
- Jiao, Y., Berisha, V., Tu, M., & Liss, J. (2015). Convex weighting criteria for speaking rate estimation. *IEEE/ACM transactions on Audio, Speech, and Language Processing*, 23, 1421-1430.
- Kent, R.D., Kent, J.F., Duffy, J., & Weismer, G. (1998). The dysarthrias: Speech-voice profiles, related dysfunctions, and neuropathology. *Journal of Medical Speech-Language Pathology*, 6, 165-211.
- Kent, R.D., Duffy, J., Kent, J.F., Vorperian, H.K., & Thomas, J.E. (1999). Quantification of motor speech abilities in stroke: Time-energy analyses of syllable and word repetition. *Journal of Medical Speech-Language Pathology*, 7, 83-90.
- Buder, E.H. (2000). Acoustic analysis of voice quality: a tabulation of algorithms 1902-1990. In R.D. Kent, & M.J. Ball (Eds.), *Voice quality measurement* (pp. 119-244). San Diego: Singular Publishing Group.
- Kent, R.D., Kent, J.F., Duffy, J.R., Thomas, J.E., Weismer, G., & Stuntebeck, S. (2000). Ataxic dysarthria. *Journal of Speech, Language, and Hearing Research*, 43, 1275-1289.
- Kim, Y., Kent, R.D., Kent, J.F., & Duffy, J.R. (2010). Perceptual and acoustic features of dysarthria in multiple system atrophy. *Journal of Medical Speech-Language Pathology*, 18, 66-71.
- Kluin, K.J., Foster, N.L., Berent, S., & Gilman, S. (1993). Perceptual analysis of speech disorders in progressive supranuclear palsy. *Neurology*, 43, 563-563.
- Kluin, K.J., Gilman, S., Lohman, M., & Junck, L. (1996). Characteristics of the dysarthria of multiple system atrophy. *Archives of Neurology*, 53, 545-548.

- Kurtzke, J.F. (1983). Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). *Neurology*, 33, 1444-1452.
- Lewis, J.P. (1995). Fast normalized cross-correlation. *Vision Interface*, 10, 120-123.
- Lipták, T. (1958). On the combination of independent tests. *Magyar Tudományos Akadémia, Alkalmazott Matematikai Intézetének Közleményei*, 3, 171-197.
- Lisker, L., & Abramson, A.S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20, 384-422.
- Liss, J.M., White, L., Mattys, S.L., Lansford, K., Lotto, A.J., Spitzer, S.M., & Caviness, J.N. (2009). Quantifying speech rhythm abnormalities in the dysarthrias. *Journal of Speech, Language, and Hearing Research*, 52, 1334-1352.
- Little, M.A., McSharry, P.E., Roberts, S.J., Costello, D.A., & Moroz, I.M. (2007). Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *BioMedical Engineering OnLine*, 6, 23.
- Little, M.A., McSharry, P.E., Hunter, E.J., Spielman, J., & Ramig, L.O. (2009). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Transactions on Biomedical Engineering*, 56, 1015-1022.
- Litvan, I., Agid, Y., Calne, D., Campbell, G., Dubois, B., Duvoisin, R.C., Goetz, C.G., Golbe, L.I., Grafman, J., Growdon, J.H., & Hallett, M. (1996). Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): Report of the NINDS-SPSP international workshop. *Neurology*, 47, 1-9.
- Logemann, J. A., Fisher, H. B., Boshes, B., & Blonsky, E. R. (1978). Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients. *Journal of Speech and Hearing Disorders*, 43, 47-57.
- Lowit, A. (2014). Quantification of rhythm problems in disordered speech: A re-evaluation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369, 20130404.
- Maidment, J.A. (1976). Voice fundamental frequency characteristics as language differentiators. *Speech and Hearing: Work in progress*, 2.
- Manfredi, C., Cantarella, G., Migali, N., Berlusconi, A., & Maraschi, B. (1996). Assessing the effectiveness of botulinus treatment in spasmodic dysphonia. *Measurements*, 117, 219-224.
- Martens, H., Dekens, T., Van Nuffelen, G., Latacz, L., Verhelst, W., & De Bodt, M. (2015). Automated speech rate measurement in dysarthria. *Journal of Speech, Language, and Hearing Research*, 58, 698-712.
- MBSC [Computer software]. (2018). Retrieved from <http://www.seas.ucla.edu/spapl/shareware.html>
- McClean, M.D., Beukelman, D.R., & Yorkston, K.M. (1987). Speech-muscle visuomotor tracking in dysarthric and nonimpaired speakers. *Journal of Speech, Language, and Hearing Research*, 30, 276-282.
- Morise, M., Kawahara, H., & Katayose, H. (2009). Fast and reliable F0 estimation method based on the period extraction of vocal fold vibration of singing voice and speech. *Proceedings of the*

Audio Engineering Society Conference: 35th International Conference: Audio for Games, London, England (pp. 77-81). New York, New York: Curran Associates.

Morise, M. (2017). Harvest: A high-performance fundamental frequency estimator from speech signals. Proceedings of the Interspeech 2017: Situated interaction, Stockholm, Sweden (pp. 2321-2325). Adelaide, Australia: Causal Productions.

Naylor, P.A., Kounoudes, A., Gudnason, J., & Brookes, M. (2007). Estimation of glottal closure instants in voiced speech using the DYPSA algorithm. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 15, 34-43.

Noll, A.M. (1967). Cepstrum pitch determination. *The Journal of the Acoustical Society of America*, 41, 293-309.

Novotný, M., Ruzs, J., Čmejla, R., & Růžička, E. (2014). Automatic evaluation of articulatory disorders in Parkinson's disease. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22, 1366-1378.

Novotný, M., Pospíšil, J., Čmejla, R., & Ruzs, J. (2015). Automatic detection of voice onset time in dysarthric speech. Proceedings of the 40th IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2015, Brisbane, Australia (pp. 4340-4344). Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Novotný, M., Ruzs, J., Čmejla, R., Růžicková, H., Klempíř, J., & Růžička, E. (2016). Hypernasality associated with basal ganglia dysfunction: Evidence from Parkinson's disease and Huntington's disease. *PeerJ*, 4, e2530.

Oğuz, H., Kiliç, M.A., & Şafak, M.A. (2011). Comparison of results in two acoustic analysis programs: Praat and MDVP. *Turkish Journal of Medical Sciences*, 41, 835-841.

Ontaneda, D., Thompson, A.J., Fox, R.J., & Cohen, J.A. (2017). Progressive multiple sclerosis: Prospects for disease therapy, repair, and restoration of function. *The Lancet*, 389, 1357-1366.

Oppenheim, A.V., & Lim, J.S. (1981). The importance of phase in signals. *Proceedings of the IEEE*, 69, 529-541.

Orozco-Arroyave, J.R., Hönig, F., Arias-Londoño, J.D., Vargas-Bonilla, J.F., Daqrouq, K., Skodda, S., Ruzs, J., & Nöth, E. (2016). Automatic detection of Parkinson's disease in running speech spoken in three different languages. *The Journal of the Acoustical Society of America*, 139, 481-500.

Orozco-Arroyave, J.R., Vásquez-Correa, J.C., Vargas-Bonilla, J.F., Arora, R., Dehak, N., Nidadavolu, P.S., Christensen, H., Rudzicz, F., Yancheva, M., Chinaei, H. and Vann, A. (2018). NeuroSpeech: An open-source software for Parkinson's speech analysis. *Digital Signal Processing*, 77, 207-221.

Parliament and Council of the European Union (2016). General Data Protection Regulation act of the European Union 2016/679.

Payan, C.A., Viallet, F., Landwehrmeyer, B.G., Bonnet, A.M., Borg, M., Durif, F., Lacomblez, L., Bloch, F., Verny, M., Fermanian, J., & Agid, Y. (2011). Disease severity and progression in progressive supranuclear palsy and multiple system atrophy: Validation of the NNIPPS–Parkinson Plus Scale. *PLoS One*, 6, e22293

- Polman, C.H., Reingold, S.C., Banwell, B., Clanet, M., Cohen, J.A., Filippi, M., Fujihara, K., Havrdova, E., Hutchinson, M., Kappos, L. and Lublin, F.D. (2011). Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Annals of Neurology*, 69, 292-302.
- Postuma, R.B., Gagnon, J.F., Vendette, M., & Montplaisir, J.Y. (2009). Markers of neurodegeneration in idiopathic rapid eye movement sleep behaviour disorder and Parkinson's disease. *Brain*, 132, 3298-3307.
- Postuma, R.B., Lang, A.E., Gagnon, J.F., Pelletier, A., & Montplaisir, J.Y. (2012). How does Parkinsonism start? Prodromal Parkinsonism motor changes in idiopathic REM sleep behaviour disorder. *Brain*, 135, 1860-1870.
- Ramus, F., Nespors, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265-292.
- Rosen, K., Murdoch, B., Folger, J., Vogel, A., Cahill, L., Delatycki, M., & Corben, L. (2010). Automatic method of pause measurement for normal and dysarthric speech. *Clinical Linguistics & Phonetics*, 24, 141-154.
- Rossi, M., Perez- Lloret, S., Doldan, L., Cerquetti, D., Balej, J., Millar Vernetti, P., Hawkes, M., Cammarota, A., & Merello, M. (2014). Autosomal dominant cerebellar ataxias: A systematic review of clinical features. *European Journal of Neurology*, 21, 607-615.
- Rusz, J., Čmejla, R., Růžicková, H., & Růžicka, E. (2011). Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. *The Journal of the Acoustical Society of America*, 129, 350-367.
- Rusz, J., Hlavnička, J., Čmejla, R., & Růžicka, E. (2015a). Automatic evaluation of speech rhythm instability and acceleration in dysarthrias associated with basal ganglia dysfunction. *Frontiers in Bioengineering and Biotechnology*, 3, 104.
- Rusz, J., Hlavnička, J., Tykalová, T., Busková, J., Ulmanová, O., Růžicka, E., & Šonka, K. (2015b). Quantitative assessment of motor speech abnormalities in idiopathic rapid eye movement sleep behavior disorder. *Sleep medicine* 19, 141-147.
- Rusz, J., Hlavnička, J., Tykalová, T., Novotný, M., Dušek, P., Šonka, K., & Růžicka, E. (2018). Smartphone allows capture of speech abnormalities associated with high risk of developing Parkinson's disease. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26, 1495-1507.
- Sandoval, S., Berisha, V., Utianski, R.L., Liss, J.M., & Spanias, A. (2013). Automatic assessment of vowel space area. *The Journal of the Acoustical Society of America*, 134, EL477-EL483.
- Schmitz-Hübsch, T., Du Montcel, S.T., Baliko, L., Berciano, J., Boesch, S., Depondt, C., ... & Kremer, B. (2006). Scale for the assessment and rating of ataxia: Development of a new clinical scale. *Neurology*, 66, 1717-1720.
- SHRP [Computer software]. (2018). Retrieved from: <https://www.mathworks.com/matlabcentral/fileexchange/1230-pitch-determination-algorithm>.
- Sondhi, M. (1968). New methods of pitch extraction. *IEEE Transactions on Audio and Electroacoustics*, 16, 262-266.

- Speech Filling System [Computer software]. (2018). Retrieved from <http://www.phon.ucl.ac.uk/resource/sfs/>
- Sun, X. (2002). Pitch determination and voice quality analysis using subharmonic-to-harmonic ratio. Proceedings of the 27th IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2002, Orlando, Florida (pp. 333-336). Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.
- Talkin, D. (1995). A robust algorithm for pitch tracking (RAPT). *Speech coding and synthesis*, 495, 518.
- Talkin, D. (2018). REAPER [Computer software]. Retrieved from <https://github.com/google/REAPER>
- Tan, L.N., & Alwan, A. (2013). Multi-band summary correlogram-based pitch detection for noisy speech. *Speech Communication*, 55, 841-856.
- Titze, I.R., & Alipour, F. (2006). *The myoelastic aerodynamic theory of phonation*. Iowa city, Iowa: National Center for Voice and Speech.
- Tsanas, A., Little, M.A., McSharry, P.E., Spielman, J., & Ramig, L.O. (2012). Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. *IEEE Transactions on Biomedical Engineering*, 59, 1264-1271.
- Tsanas, A., Zañartu, M., Little, M.A., Fox, C., Ramig, L.O., & Clifford, G.D. (2014). Robust fundamental frequency estimation in sustained vowels: Detailed algorithmic comparisons and information fusion with adaptive Kalman filtering. *The Journal of the Acoustical Society of America*, 135, 2885-2901.
- Tykalova, T., Ruz, J., Klempir, J., Cmejla, R., & Ruzicka, E. (2017). Distinct patterns of imprecise consonant articulation among Parkinson's disease, progressive supranuclear palsy and multiple system atrophy. *Brain and Language*, 165, 1-9.
- Sachin, S., Shukla, G., Goyal, V., Singh, S., Aggarwal, V., & Behari, M. (2008). Clinical speech impairment in Parkinson's disease, progressive supranuclear palsy, and multiple system atrophy. *Neurology India*, 56, 122.
- Saxena, M., Behari, M., Kumaran, S.S., Goyal, V., & Narang, V. (2014). Assessing speech dysfunction using BOLD and acoustic analysis in Parkinsonism. *Parkinsonism & Related Disorders*, 20, 855-861.
- Schalling, E., Hammarberg, B., & Hartelius, L. (2007). Perceptual and acoustic analysis of speech in individuals with spinocerebellar ataxia (SCA). *Logopedics Phoniatrics Vocology*, 32(1), 31-46.
- Schalling, E., & Hartelius, L. (2013). Speech in spinocerebellar ataxia. *Brain and Language*, 127, 317-322.
- Schenck, C.H., Bundlie, S.R., & Mahowald, M.W. (1996). Delayed emergence of a parkinsonian disorder in 38% of 29 older men initially diagnosed with idiopathic rapid eye movement sleep behavior disorder. *Neurology*, 46, 388-393.
- Schroeder, M.R. (1968). Period histogram and product spectrum: New methods for fundamental-frequency measurement. *The Journal of the Acoustical Society of America*, 43, 829-834.

- Secrest, B., & Doddington, G. (1983, April). An integrated pitch tracking algorithm for speech systems. *Proceedings of the 27th IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 1983*, Boston, Massachusetts (pp. 1352-1355). Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.
- Scripture, E.W. (1902). *The elements of experimental phonetics*. New York, New York: Charles Scribner's Sons, London: Edward Arnold.
- Skodda, S., Flasskamp, A., & Schlegel, U. (2010). Instability of syllable repetition as a model for impaired motor processing: Is Parkinson's disease a "rhythm disorder"? *Journal of Neural Transmission*, 117, 605-612.
- Skodda, S., Visser, W., & Schlegel, U. (2011). Acoustical analysis of speech in progressive supranuclear palsy. *Journal of Voice*, 25, 725-731.
- Skodda, S., Schlegel, U., Klockgether, T., & Schmitz-Hübsch, T. (2013). Vowel articulation in patients with spinocerebellar ataxia. *International Journal of Speech & Language Pathology and Audiology*, 1, 61-71.
- Stouffer, S.A., Suchman, E.A., DeVinney, L.C., Star, S.A., & Williams Jr, R.M. (1949). *The American soldier: Adjustment during army life*. (Studies in social psychology in World War II, Vol 1). Princeton, New Jersey: Princeton University Press.
- Till, J.A. (1995). Diagnostic goals and computer-assisted evaluation of speech and related physiology. *Special Interest Division 2, Neurophysiology and Neurogenic Speech and Language Disorders*, American Speech-Language-Hearing Association, 5.
- Trapp, B.D., & Nave, K.A. (2008). Multiple sclerosis: An immune or neurodegenerative disorder?, *Annual Review of Neuroscience*, 31, 247-269.
- Tsanas, A., Zañartu, M., Little, M.A., Fox, C., Ramig, L.O., & Clifford, G.D. (2014). Robust fundamental frequency estimation in sustained vowels: Detailed algorithmic comparisons and information fusion with adaptive Kalman filtering. *The Journal of the Acoustical Society of America*, 135, 2885-2901.
- Vaiciukynas, E., Verikas, A., Gelzinis, A., & Bacauskiene, M. (2017). Detecting Parkinson's disease from sustained phonation and speech signals. *PloS one*, 12, e0185613.
- Van der Graaff, M., Kuiper, T., Zwinderman, A., Van de Warrenburg, B., Poels, P., Offeringa, A., Van der Kooi, A., Speelman, H., & De Visser, M. (2009). Clinical identification of dysarthria types among neurologists, residents in neurology and speech therapists. *European Neurology*, 61, 295-300.
- VOICEBOX [Computer software]. (2018). Retrieved from <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html>
- Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law*, 7, 76-99.
- Weismer, G. (2006). *Motor speech disorders: Essays for Ray Kent*. San Diego, California: Plural Publishing.
- WORLD [Computer software]. (2018). Retrieved from: <https://github.com/mmorise/World/>.

YANGsaf (2018). Retrieved from: [https://github.com/google/yang\\_vocoder](https://github.com/google/yang_vocoder)

Yorkston, K.M., Miller, R.M., & Strand, E.A. (1995). Management of speech and swallowing in degenerative diseases. Tucson, Arizona: Communication Skill Builders.

Zyski, B.J., & Weisiger, B.E. (1987). Identification of dysarthria types based on perceptual analysis. *Journal of Communication Disorders*, 20, 367-378.

# LIST OF AUTHOR'S PUBLICATIONS AND RECOGNITION

All authors hold equal share in the joined publications.

## PUBLICATIONS RELATED TO THE DOCTORAL THESIS

### Articles in journals with impact factor

Rusz, J., Hlavnička, J., Tykalová, T., Busková, J., Ulmanová, O., Růžička, E., & Šonka, K. (2015). Quantitative assessment of motor speech abnormalities in idiopathic rapid eye movement sleep behavior disorder. *Sleep medicine* 19, 141-147.

*J.H. developed the new speech measurement, performed the analysis, and revised the final version of the manuscript.*

Hlavnička, J., Čmejla, R., Tykalová, T., Šonka, K., Růžička, E., & Rusz, J. (2017). Automated analysis of connected speech reveals early biomarkers of Parkinson's disease in patients with rapid eye movement sleep behaviour disorder. *Scientific Reports*, 7, 12.

*J.H. designed the experiment, developed all the technologies presented in the paper, evaluated the methods, analysed the data, and wrote the manuscript. J.R. supervised the work and contributed to the experimental design and analysis of the results.*

Rusz, J., Novotný, M., Hlavnička, J., Tykalová, T., & Růžička, E. (2017). High-accuracy voice-based classification between patients with Parkinson's disease and other neurological diseases may be an easy task with inappropriate experimental design. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 24, 1100-1108.

*J.H. Conducted the classification experiment, analyzed the results, and revised the final version of the manuscript.*

Rusz, J., Beňová, B., Růžičková, H., Novotný, M., Tykalová, T., Hlavnička, J., Uher, T., Vaněčková, M., Andělová, M., Novotná, K., Kadrnožková, L., & Horáková, D. (2018). Characteristics of motor speech phenotypes in multiple sclerosis. *Multiple sclerosis and related disorders*, 19, 62-69.

*J.H. evaluated quality of articulation measurements and revised the final version of the manuscript.*

Rusz, J., Hlavnička, J., Tykalová, T., Novotný, M., Dušek, P., Šonka, K., & Růžička, E. (2018). Smartphone allows capture of speech abnormalities associated with high risk of developing Parkinson's disease. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26, 1495-1507.

*J.H. measured the smartphone characteristics, evaluated the processing methods, analysed the acoustic signals, and participated on the writing of the manuscript.*

## Articles in peer-reviewed journals

Rusz, J., Hlavnička, J., Čmejla, R., & Růžička, E. (2015). Automatic evaluation of speech rhythm instability and acceleration in dysarthrias associated with basal ganglia dysfunction. *Frontiers in Bioengineering and Biotechnology*, 3, 104.

*J.H. developed the presented technology, performed analysis of the data, participated on the writing of the manuscript.*

## Other articles indexed by the SCOPUS

Hlavnička, J., Tykalová, T., Čmejla, R., Klempíř, J., Růžička, E., & Rusz, J. (2017). Dysprosody differentiate between Parkinson's disease, progressive supranuclear palsy, and multiple system atrophy. *Proceedings of the Interspeech 2017: Situated interaction, Stockholm, Sweden* (pp. 1844-1848). Adelaide, Australia: Causal Productions.

*J.H. conducted the experiment, evaluated quality of the applied processing, performed the analysis, and wrote the manuscript.*

Hlavnička, J. (2015). Dynamical characteristics of speech apparatus in Huntington's disease. *Lékař a technika* 45, 88-92.

*J.H. exploited the data recorded for previous studies and did it all alone.*

## Other articles and abstracts

Hlavnička, J., Rusz, J., & Čmejla, R. (2014). Automatické hodnocení pauz v řeči u Parkinsonovy nemoci. *Proceedings of the 22nd Annual Conference Technical Computing Bratislava 2014, Bratislava, Slovak Republic* (pp. 1-6). Prague, Czech Republic: Institute of Chemical Technology.

*J.H. developed and evaluated the technology and wrote the manuscript.*

Hlavnička, J., Čmejla, R., & Rusz, J. (2015). Rychlost a rytmus v řeči u Parkinsonovy nemoci. *Proceedings of the Letní doktorandské dny 2015, Prague, Czech Republic* (pp. 67-72). Prague, Czech technical university in Prague.

*J.H. developed and tested the methodology and wrote the manuscript.*

Hlavnička, J. (2015). Dynamical characteristics of speech apparatus in Huntington's disease. Proceedings of the 19th International Scientific Student Conference POSTER 2015, Prague, Czech Republic (pp. BI07). Prague, Czech Republic: Czech technical university in Prague.

*J.H. prepared this study solely by himself utilizing the data recorded for previous studies.*

Hlavnička, J., Čmejla, R., & Rusz, J. (2016). Robustní detektor základní frekvence hlasivek pro dysartrickou řeč. Letní doktorandské dny 2016, 47-52.

*J.H. designed the experiment based on the data from previous studies, invented the new technology, and wrote the extended abstract.*

Rusz, J., Růžička, E., Šonka, K., Hlavnička, J., Tykalová, T., Bušková, J., Ulmanová, O. (2015) Motor speech impairment indicates prodromal neurodegeneration in REM sleep behaviour disorder. Abstracts of the Nineteenth International Congress of Parkinson's Disease and Movement Disorders, San Diego, California (pp. LBA17). Hoboken, New Jersey: Wiley-Blackwell.

*J.H. analysed the acoustic signals and revised the final version of the abstract.*

Hlavnička, J., Čmejla, R., Klempíř, J., Růžička, E., & Rusz, J. (2019). Acoustic tracking of pitch, modal, and subharmonic vibrations of vocal folds in Parkinsonism. Manuscript submitted for publication.

*J.H. conceived the design of the experiment, developed all the methods, evaluated the accuracy, conducted the statistical analysis, and wrote the manuscript.*

## OTHER PUBLICATIONS

### Other articles and abstracts

Hlavnička, J., Tykalová, T., Bačáková, B., Baxa, M., Čmejla, R., Motlík, J., Klempíř, J., & Rusz, J. (2016). Hoarseness can be found in vocalisations of both human as well as genetically modified minipig model of Huntington's disease. Presented on the European Huntington's Disease Network (EHDN2017), Hague, Netherlands (Suppl. 1, pp. A32). London, Great Britain: Journal of Neurology, Neurosurgery & Psychiatry.

*J.H. devised the processing of the recordings, performed the analysis, and wrote the abstract.*

### Other articles indexed by the SCOPUS

Tykalová, T., Hlavnička, J., Macáková, M., Baxa, M., Čmejla, R., Motlík, J., Klempíř, J., & Rusz, J. (2015). Grunting in genetically modified minipig animal model for Huntington's disease. Česká a slovenská neurologie a neurochirurgie, 61-65.

*J.H. participated on one of the recording sessions, performed the qualitative analysis of the recordings, and revised the final version of the manuscript.*

## CITATIONS INDEXED IN THE WEB OF SCIENCE AND SCOPUS

Citing publications are listed under the titles of cited publications. The list of citations does not include publications that had at least one author in common with the cited publication.

### **CHARACTERISTICS OF MOTOR SPEECH PHENOTYPES IN MULTIPLE SCLEROSIS**

Noffs, G., Perera, T., Kolbe, S.C., Shanahan, C.J., Boonstra, F.M.C., Evans, A., Butzkueven, H., van der Walt, A. and Vogel, A.P. (2018). What speech can tell us: A systematic review of dysarthria characteristics in Multiple Sclerosis. *Autoimmunity reviews*.

Fazeli, M., Moradi, N., Soltani, M., Naderifar, E., Majdinasab, N., Latifi, S. M., & Dastoorpour, M. (2018). Dysphonia Characteristics and Vowel Impairment in Relation to Neurological Status in Patients with Multiple Sclerosis. *Journal of Voice*.

Mefferd, A. S., Lai, A., & Bagnato, F. (2019). A first investigation of tongue, lip, and jaw movements in persons with dysarthria due to multiple sclerosis. *Multiple sclerosis and related disorders*, 27, 188-194.

### **DYSPROSDY DIFFERENTIATE BETWEEN PARKINSON'S DISEASE, PROGRESSIVE SUPRANUCLEAR PALSY, AND MULTIPLE SYSTEM ATROPHY**

Pell, E. (2018). Progressive supranuclear palsy: longitudinal study by acoustical analysis of speech. *Revista de Investigacion en Logopedia*, 8, 115-128.

### **HIGH-ACCURACY VOICE-BASED CLASSIFICATION BETWEEN PATIENTS WITH PARKINSON'S DISEASE AND OTHER NEUROLOGICAL DISEASES MAY BE AN EASY TASK WITH INAPPROPRIATE EXPERIMENTAL DESIGN**

Brabenec, L., Mekyska, J., Galáž, Z., & Rektorová, I. (2017). Speech disorders in Parkinson's disease: early diagnostics and effects of medication and brain stimulation. *Journal of Neural Transmission*, 124, 303-334.

Mirarchi, D., Vizza, P., Tradigo, G., Lombardo, N., Arabia, G., & Veltri, P. (2017). Signal Analysis for Voice Evaluation in Parkinson's Disease. *Proceedings of Healthcare Informatics (ICHI)*, Park City, Utah (pp. 530-535). Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

### **AUTOMATED ANALYSIS OF CONNECTED SPEECH REVEALS EARLY BIOMARKERS OF PARKINSON'S DISEASE IN PATIENTS WITH RAPID EYE MOVEMENT SLEEP BEHAVIOUR DISORDER**

Bhat, S., Acharya, U. R., Hagiwara, Y., Dadmehr, N., & Adeli, H. (2018). Parkinson's disease: Cause factors, measurable indicators, and early diagnosis. *Computers in biology and medicine*, 102, 234-241.

Dashtipour, K., Tafreshi, A., Lee, J., & Crawley, B. (2018). Speech disorders in Parkinson's disease: pathophysiology, medical management and surgical approaches. *Neurodegenerative disease management*, 8, 337-348.

Ward, R.M., & Kelty-Stephen, D. G. (2018). Bringing the nonlinearity of the movement system to gestural theories of language use: Multifractal structure of spoken English supports the compensation for coarticulation in human speech perception. *Frontiers in physiology*, 9, 1152.

Lipsmeier, F., Taylor, K. I., Kilchenmann, T., Wolf, D., Scotland, A., Schjodt-Eriksen, J., Cheng, W. Y., Fernandez-Garcia, I., Siebourg-Polster, J., Jin, L., Soto, J., Verselis, L., Boess, F., Koller, M., Grundman, M., Monsch, A. U., Postuma, R. B., Ghosh, A., Kremer, T., Czech, C., Gossens, C., & Lindemann, M. (2018). Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 Parkinson's disease clinical trial. *Movement disorders*, 33, 1287-1297.

Lowit, A., Marchetti, A., Corson, S., & Kuschmann, A. (2018). Rhythmic performance in hypokinetic dysarthria: Relationship between reading, spontaneous speech and diadochokinetic tasks. *Journal of communication disorders*, 72, 26.

Siddharth. A., Visanji, N.P., Mestre, T.A., Tsanas, A., AlDakheel, A., Connolly, B.S., Gasca-Salas, C., Kern, D.S., Jain, J., Slow, E.J., Faust-Socher, A., Lang, A.E., Little, M.A., & Marras, C. (2018). Investigating voice as a biomarker for leucine-rich repeat kinase 2-associated Parkinson's disease. *Journal of Parkinson's disease*, 8, 503-510.

Mičianová, E., Straka, I., Kušnírová, A., Valkovič, P., & Cséfalvay, Z. (2018). Zrozumiteľnosť reči a klinické parametre u pacientov s Parkinsonovou chorobou. *Česká a slovenská neurologie a neurochirurgie*, 81, 586-592.

Alimuradov, A.K., Tychkov, A., Yu., Kuzmin, A.V., Churakov, P.P., Ageykin, A.V., Vishnevskaya, G.V. (2019). Signal analysis algorithm for mental disorders diagnostic system: Pitch frequency detection and measurement. *International Journal of Embedded and Real-Time Communication Systems*, 10, Pages 22-47.

Arias-Vergara, T., Vasquez-Correa, J. C., Orozco-Arroyave, J. R., Klumpp, P., & Nöth, E. (2018). Unobtrusive monitoring of speech impairments of Parkinson's disease patients through mobile devices. *Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, Alberta, Canada (pp. 6004-6008). Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Alimuradov, A., Tychkov, A., Kuzmin, A., Churakov, P., Ageykin, A., & Vishnevskaya, G. (2017). Measurement of speech signal patterns under borderline mental disorders. *Proceedings of 21st Conference of Open Innovations Association (FRUCT)*, Helsinki, Finland (pp. 26-33). Helsinki, Finland: Fruct.

#### **QUANTITATIVE ASSESSMENT OF MOTOR SPEECH ABNORMALITIES IN IDIOPATHIC RAPID EYE MOVEMENT SLEEP BEHAVIOUR DISORDER**

Siddharth. A., Visanji, N.P., Mestre, T.A., Tsanas, A., AlDakheel, A., Connolly, B.S., Gasca-Salas, C., Kern, D.S., Jain, J., Slow, E.J., Faust-Socher, A., Lang, A.E., Little, M.A., & Marras, C. (2018). Investigating voice as a biomarker for leucine-rich repeat kinase 2-associated Parkinson's disease. *Journal of Parkinson's disease*, 8, 503-510.

Högl, B., Stefani, A., & Videnovic, A. (2018). Idiopathic REM sleep behaviour disorder and neurodegeneration—an update. *Nature Reviews Neurology*, 14, 40.

Brabenec, L., Mekyska, J., Galáž, Z., & Rektorová, I. (2017). Speech disorders in Parkinson's disease: early diagnostics and effects of medication and brain stimulation. *Journal of Neural Transmission*, 124, 303-334.

Jeancolas, L., Benali, H., Benkelfat, B.E., Mangone, G., Corvol, J.C., Vidailhet, M., Lehericy, S. & Petrovska-Delacrétaz, D. (2017). Automatic detection of early stages of Parkinson's disease through acoustic voice analysis with mel-frequency cepstral coefficients. *Proceedings of the Advanced Technologies for Signal and Image Processing (ATSIP)*, Fez, Morocco (pp. 1-6). Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Postuma, R. B. (2016). Voice changes in prodromal Parkinson's disease: Is a new biomarker within earshot?. *Sleep medicine*, 19, 148-149.

Cushnie-Sparrow, D., Adams, S., Abeyesekera, A., Pieterman, M., Gilmore, G., & Jog, M. (2018). Voice quality severity and responsiveness to levodopa in Parkinson's disease. *Journal of communication disorders*, 76, 1-10.

Zhou, L., Zhu, L., & Liu, J. (2018). From rapid eye movement sleep behavior disorder to Parkinson's disease: Possible predictive markers of conversion. *ACS Chemical Neuroscience*. doi: 10.1021/acschemneuro.8b00388.

#### **AUTOMATIC EVALUATION OF SPEECH RHYTHM INSTABILITY AND ACCELERATION IN DYSPARTHRIAS ASSOCIATED WITH BASAL GANGLIA DYSFUNCTION**

Kashyap, B., Pathirana, P.N., Horne, M., Power, L., Szmulewicz, D. (2018). Quantitative Assessment of Syllabic Timing Deficits in Ataxic Dysarthria. *Proceedings of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Honolulu, Hawaii (pp. 425-428). Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Lowit, A., Marchetti, A., Corson, S., & Kuschmann, A. (2018). Rhythmic performance in hypokinetic dysarthria: Relationship between reading, spontaneous speech and diadochokinetic tasks. *Journal of communication disorders*, 72, 26.

Klempíř, O., & Krupička, R. (2018). Machine learning using speech utterances for Parkinson's disease detection. *Lékař a technika*, 48, 66-71.

Godino-Llorente, J. I., Shattuck-Hufnagel, S., Choi, J. Y., Moro-Velázquez, L., & Gómez-García, J. A. (2017). Towards the identification of Idiopathic Parkinson's disease from the speech. *New articulatory kinetic biomarkers*. *PloS one*, 12, e0189583.

Barkmeier-Kraemer, J. M., & Clark, H. M. (2017). Speech–Language Pathology Evaluation and Management of Hyperkinetic Disorders Affecting Speech and Swallowing Function. *Tremor and Other Hyperkinetic Movements*, 7, 1-19. doi: 10.7916/D8Z32B30.

## **AWARDS**

The publication by Hlavnička et al. (2017) was awarded first prize in the contest of original publications in the field of neurology and neuroscience by the *Herenerin nadační fond* in December 2017.