

České vysoké učení technické v Praze
Fakulta elektrotechnická
Katedra teorie obvodů

Analýza a hodnocení chraptivosti v řečovém signálu

Dizertační práce

Adam Stráník

Praha, listopad 2013

Doktorský studijní program: Elektrotechnika a informatika
Studijní obor: Teoretická elektrotechnika

Školitel: doc. Ing. Roman Čmejla, CSc.

Abstrakt

Tato práce se zabývá objektivním automatickým hodnocením chraptivosti v hlasu. Chraptivost je jedním z prvních příznaků patologických změn v hlasovém traktu, proto je nutná včasná diagnostika příčin chraptivosti a kvalitní terapie, která zamezí jeho dalšímu poškození.

Pro účely této práce byly vytvořeny dvě databáze nahrávek s chraptivými, hlasy pořízené na Foniatrické klinice 1. LF UK a VFN v Praze. První databáze obsahuje 469 nahrávek prodloužené fonace hlásky /a/, druhá databáze obsahuje 593 nahrávek standardizovaného čteného textu *Podzim na Starém bělidle*. Obě databáze byly nezávisle na sobě ohodnoceny pěti odborníky z Foniatrické kliniky. Každý odborník hodnotil každou nahrávku z obou databází dvakrát s alespoň dvoutýdenním odstupem, aby se zamezilo paměťovému efektu hodnotitelů při opakovaném hodnocení. K subjektivnímu hodnocení byla použita modifikovaná stupnice subjektivního hodnocení GRBAS, a sice stupnice GRBT. Tyto stupnice hodnotí jednotlivé aspekty chraptivosti: G (celkový dojem z hlasu), R (chraplavost), B (dyšnost) a T (hlasová slabost nebo přemáhání hlasu).

Subjektivní hodnocení bylo podrobena statistickému vyhodnocení míry shody hodnotitele při opakovaném hodnocení a shody skupiny hodnotitelů. Vyhodnocení potvrdilo, že získané subjektivní hodnocení je dostatečně spolehlivé a může být použito k trénování modelů automatických klasifikátorů.

Byly zavedeny dvě klasifikační úlohy: hodnocení podle diskrétní a podle spojitě subjektivní stupnice GRBT. Hodnota výsledného diskrétního hodnocení je určena jako modus hodnocení daného parametru G, R, B nebo T pro každou nahrávku. Spojité hodnocení jako dáno průměrným hodnocením nahrávky pro daný parametr G, R, B nebo T.

Na jednotlivé nahrávky bylo aplikováno množství algoritmů pro analýzu zvukového signálu. Každá parametrizace poskytla časovou řadu hodnot pro danou nahrávku. Tyto časové průběhy byly dále parametrizovány jak popisem distribuce hodnot v rámci nahrávky, tak popisem časového průběhu. Pro nahrávky prodloužené fonace tak bylo vytvořeno 480 deskriptorů, a pro nahrávky čteného textu 201 deskriptorů.

Po redukci příznakového prostoru vhodnými metodami dolování dat byly vybrány maximálně 4 deskriptory, na základě kterých byla provedena automatická klasifikace míry chraptivosti podle stupnice GRBT. Diskrétní hodnocení prodloužené fonace hlásky /a/ dosáhlo přesnosti klasifikace od 52 % u parametru T do 75 % pro parametr B. Při diskrétním hodnocení nahrávek se čteným textem *Podzim na Starém bělidle* byly dosaženy přesnosti klasifikace od 65 % pro parametr T po 77 % pro parametr B. V úloze automatického hodnocení nahrávek podle spojitě stupnice jsou výsledky udávány pomocí Pearsonova lineárního korelačního koeficientu r . Pro prodlouženou fonaci dosahují hodnot od $r=0,42$ pro parametr T do $r=0,90$ pro parametr B. Čtený text dosahuje hodnot od $r=0,43$ pro parametr T do $r=0,92$ pro parametr B.

Abstract

This work is focused on objective and automatic evaluation of hoarseness in voice. Hoarseness is one of the first signs of pathological changes in the vocal tract. It is therefore necessary to have tools for early diagnosis of causes of hoarseness and good therapy tools.

The basis of the work are two speech databases recorded by the Department of Phoniatics, 1st Faculty of Medicine, Charles University and General Faculty Hospital in Prague. The first database contains 469 records of sustained phonation of vowel /a/, the second database contains 593 records of the standard read text *Podzim na Starém bělidle*. Both databases were independently evaluated by five experts from the Department of Phoniatics. Evaluation was carried out twice with at least two weeks time delay, to prevent the memory effect. A modified GRBAS (G – overall voice quality, R – roughness, B – breathiness, A – asthenicity and S – strain) scale was used, where the modification involves merging parameters A and S into parameter T (tension).

Subjective evaluations of records were subjected to statistical analysis by means of intra- and inter-rater agreement. Results of these analyses confirmed that subjective evaluation can be used as training data for automatic classifiers.

Two classification tasks were introduced: evaluation by discrete and continuous scale of GRBT. The discrete grade is a modus rating for each recording by the parameter whereas the continuous grade is an average rating by the parameter.

The above-mentioned recordings were subjected to a large amount of speech audio signal analysis algorithms, resulting in a time course of the parameters for a particular recording. These waveforms were then parameterized by means of distribution of values within the recording and by description of the time behavior. For each recording of sustained phonation, 480 descriptors were created, and for each recording of read text total of 201 descriptors were carried out.

After a reduction of the feature space by appropriate methods, a set of four descriptors was selected for each parameter from GRBT scale and used for the classification. The accuracy of classification according to the discrete GRBT scale for sustained vowel /a/ phonation ranged between 52 % for parameter T to 75 % for parameter B. For the read text the accuracy ranged from 65 % for parameter T to 77 % for parameter B. The classification according to the continuous scale, expressed by the Pearson's linear correlation coefficient r , ranged from $r=0.42$ for parameter T to $r=0.90$ for parameter B for sustained vowel /a/ phonation and from $r=0.43$ for parameter T to $r=0.92$ for parameter B for the read text.

Poděkování

Rád bych na tomto místě poděkoval doc. Ing. Romanu Čmejlovi, CSc. za vedení při práci, za cenné podněty a rady. Dále bych rád poděkoval Dr. Ing. Janu Vokřálovi z Foniatické kliniky 1. LF UK a VFN za připomínky k práci a pomoc při organizování subjektivního hodnocení odborníky z Foniatické kliniky. V neposlední řadě bych rád poděkoval těmto odborníkům za čas, který strávili při pečlivém hodnocení poměrně rozsáhlých databází.

Tento výzkum byl podporován z grantů GD102/08/H008; GACR 102/12/2230 a SGS12/185/OHK4/3T/13.

Obsah

Seznam obrázků	xii
Seznam tabulek	xv
Seznam zkratk a symbolů	xvii
1 Úvod	1
1.1 Hlas	1
1.1.1 Znělé hlásky	2
1.1.2 Neznělé hlásky	3
1.2 Chraptivost	3
1.3 Metody hodnocení chraptivosti	5
1.3.1 Subjektivní metody	5
1.3.2 Objektivní metody	7
2 Současný stav řešené problematiky	8
2.1 Používané parametrizace	8
2.1.1 Analýza hlasivkového tónu	8
2.1.2 Analýzy v časové oblasti	9
2.1.3 Analýzy ve spektrální oblasti	9
2.1.4 Analýzy v keprální oblasti	10
2.2 Používané databáze a metody subjektivního hodnocení	12
2.2.1 Syntetické signály	12
2.2.2 Komerční databáze	12
2.2.3 Lokální databáze zahraničních studií	13
2.3 Automatické hodnocení chraptivosti	14
3 Cíle práce	16
4 Akustické parametry pro analýzu chraptivosti	18
4.1 Analýza autokorelační funkce	18
4.2 Index dyšnosti BRI	21
4.3 Výraznost keprální špičky CPP	21
4.4 Poměr hlasivkového a šumového signálu GNE	23
4.5 Analýza generalizovaného variogramu GVA	27
4.6 Poměr amplitud harmonických složek H1A	29
4.7 Poměr energií ve frekvenčních pásmech HLR	30
4.8 Poměr energie harmonické a šumové složky HNR	31
4.8.1 Porovnání hodnot s Praatem	35
4.8.2 Vliv počtu iterací na výslednou hodnotu HNR	36
4.9 Spektrální páry LSP	38

4.9.1	Diference LSF – $dLSF_{x,y}$	40
4.9.2	Průměr LSF – $mLSF_{x,y}$	40
4.10	Spektrální deformace SD	41
4.11	Spektrální rolloff SR	41
4.12	Spektrální šířka SW	42
4.13	Odhady hlasivkové frekvence	42
4.13.1	ACRF0	43
4.13.2	CEPSF0	43
4.13.3	SPECF0	43
4.13.4	Ukázka průběhů hlasivkové frekvence	44
5	Databáze promluv	45
5.1	Metodika výběru nahrávek	45
5.2	Subjektivní hodnocení nahrávek	46
5.2.1	Metodika hodnocení nahrávek	46
5.2.2	Shoda hodnotitelů při opakovaném testu	46
5.2.3	Shoda skupiny hodnotitelů	47
5.3	Analýza subjektivního hodnocení	49
5.3.1	Přísnost opakovaného hodnocení	50
5.3.2	Vzájemná korelace parametrů GRBT	51
5.3.3	Shoda hodnotitelů při opakovaném hodnocení	52
5.3.4	Shoda skupiny hodnotitelů	53
6	Experimentální hodnocení chraptivosti	55
6.1	Parametrizace použité pro automatické hodnocení chraptivosti	55
6.1.1	Popis distribuce hodnot	55
6.1.2	Popis časového průběhu	55
6.1.3	Parametrizace pouze ve znělých částech	57
6.1.4	Metodika popisu parametrizací	58
6.2	Definice klasifikační úlohy	60
6.2.1	Diskrétní hodnocení	60
6.2.2	Spojité hodnocení	61
6.2.3	Použité metody dolování dat	62
6.2.4	Redukce příznakového prostoru pro klasifikaci	62
6.2.5	Použité klasifikátory	62
6.3	Výsledky automatického hodnocení prodloužené fonace	63
6.3.1	Diskrétní hodnocení	63
6.3.2	Spojité hodnocení	66
6.4	Výsledky automatického hodnocení čteného textu	70
6.4.1	Diskrétní hodnocení	70
6.4.2	Spojité hodnocení	74

7	Diskuze	79
7.1	Databáze promluv	79
7.2	Automatické hodnocení chraptivosti	79
7.2.1	Porovnání dosažených výsledků s publikovanými studii	80
7.2.2	Diskuze použitých akustických parametrů	81
8	Shrnutí výsledků a závěr	85
8.1	Shrnutí výsledků	87
8.2	Přehled výsledků dizertace	91
8.3	Závěry pro další rozvoj hodnocení chraptivosti	92
	Literatura	93
	Seznam vlastních publikací	107
9	Vzor formuláře pro hodnocení metodou CAPE-V	110
10	Obsah databáze	111
10.1	Fonace	111
10.2	Krátké věty	111
10.3	Různé	111
10.4	Vzor formuláře subjektivního hodnocení	113
11	Subjektivní hodnocení databáze	115
12	Porovnání subjektivního a objektivního hodnocení	122

Seznam obrázků

1.1	Popis mluvidel	2
3.1	Blokové schéma zamýšleného systému pro automatické hodnocení chraptivosti.	17
4.1	Průběh parametrů RPK a SPS pro zdravý a chraptivý hlas	19
4.2	Rozdíl mezi parametry RPK a SPS.	19
4.3	Ukázky autokorelačních funkcí pro nahrávku hodnocenou ($G_1R_2B_1T_{-1}$).	20
4.4	Porovnání průběhů BRI pro zdravý a chraptivý hlas	21
4.5	Porovnání průběhů CPP pro zdravý a chraptivý hlas.	22
4.6	Ukázka CPP v jednom mikrosegmentu.	23
4.7	Ukázka výpočtu GNE pro zdravý hlas.	25
4.8	Ukázka výpočtu GNE pro chraptivý hlas.	26
4.9	Porovnání průběhů GNE pro zdravý a chraptivý hlas.	27
4.10	Porovnání průběhů GVA a GVA_Topt pro zdravý a chraptivý hlas.	28
4.11	Histogramy GVA_Topt pro zdravý a chraptivý hlas.	28
4.12	Porovnání průběhů H1A pro zdravý a chraptivý hlas.	29
4.13	Ukázka detekce harmonických ve spektru.	30
4.14	Porovnání průběhů HLR pro zdravý a chraptivý hlas.	31
4.15	Základní princip algoritmu pro výpočet HNR.	32
4.16	Regiony zájmu ve spektru pro výpočet HNR.	33
4.17	Blokové schéma iterativního algoritmu pro výpočet HNR.	34
4.18	Kritické hodnoty hlasivkové frekvence v závislosti na délce mikrosegmentu.	35
4.19	Minimální délka DFT pro výpočet HNR v závislosti na odhadnuté hlasivkové frekvenci.	35
4.20	Porovnání průběhů HNR pro zdravý a chraptivý hlas.	36
4.21	Ukázka HNR v jednom mikrosegmentu.	36
4.22	Porovnání hodnot HNR mezi implementovaným algoritmem a Praatem.	37
4.23	Ukázka vlivu počtu iterací na výslednou hodnotu HNR.	37
4.24	Porovnání průběhů LSF pro zdravý a chraptivý hlas.	39
4.25	Ukázka LSF v jednom mikrosegmentu.	40
4.26	Porovnání průběhů SD pro zdravý a chraptivý hlas.	41
4.27	Porovnání průběhů SR pro zdravý a chraptivý hlas.	42
4.28	Porovnání průběhů SW pro zdravý a chraptivý hlas.	43
4.29	Porovnání průběhů f_0 pro zdravý a chraptivý hlas.	44
5.1	Korelační matice pro vzájemnou shodu hodnotitelů pro prodlouženou fonaci.	50
5.2	Korelační matice pro vzájemnou shodu hodnotitelů pro čtený text.	50

6.1	Ukázka střední hodnoty a směrodatné odchylky parametru GNE pro zdravý a chraptivý hlas	56
6.2	Ukázka trendu časového průběhu parametru GNE pro zdravý a chraptivý hlas	56
6.3	Ukázka časového průběhu parametru GNE zbaveného trendu pro zdravý a chraptivý hlas	57
6.4	Ukázka průběhu parametrizace δ	58
6.5	Porovnání detekce znělosti na krátkém úseku čteného textu.	59
6.6	Histogramy diskrétního hodnocení nahrávek prodloužené fonace metodou GRBT.	60
6.7	Histogramy diskrétního hodnocení nahrávek čteného textu metodou GRBT.	61
6.8	Histogramy spojitého hodnocení nahrávek prodloužené fonace metodou GRBT.	61
6.9	Histogramy spojitého hodnocení nahrávek čteného textu metodou GRBT.	61
6.10	Boxploty pro hodnoty parametrů použitých pro hodnocení diskrétního G u prodloužené fonace.	63
6.11	Boxploty pro hodnoty parametrů použitých pro hodnocení diskrétního R u prodloužené fonace.	64
6.12	Boxploty pro hodnoty parametrů použitých pro hodnocení diskrétního B u prodloužené fonace.	65
6.13	Boxploty pro hodnoty parametrů použitých pro hodnocení diskrétního T u prodloužené fonace.	65
6.14	Schéma rozhodovacího stromu pro klasifikaci diskrétního parametru T pro prodlouženou fonaci.	66
6.15	Ukázka klasifikace prodloužené fonace, parametr G.	67
6.16	Ukázka klasifikace prodloužené fonace, parametr R.	68
6.17	Ukázka klasifikace prodloužené fonace, parametr B.	69
6.18	Ukázka klasifikace prodloužené fonace, parametr T.	69
6.19	Boxploty pro hodnoty parametrů použitých pro hodnocení diskrétního G u čteného textu.	70
6.20	Schéma rozhodovacího stromu pro klasifikaci diskrétního parametru G pro čtený text.	70
6.21	Boxploty pro hodnoty parametrů použitých pro hodnocení diskrétního R u čteného textu.	71
6.22	Boxploty pro hodnoty parametrů použitých pro hodnocení diskrétního B u čteného textu.	72
6.23	Boxploty pro hodnoty parametrů použitých pro hodnocení diskrétního T u čteného textu.	73
6.24	Schéma rozhodovacího stromu pro klasifikaci diskrétního parametru T pro čtený text.	73
6.25	Ukázka klasifikace čteného textu, parametr G.	75
6.26	Ukázka klasifikace čteného textu, parametr R.	76
6.27	Ukázka klasifikace čteného textu, parametr B.	77

6.28	Ukázka klasifikace čteného textu, parametr T.	78
8.1	Schémata experimentálních systému pro automatické hodnocení chrap- tivosti podle diskrétní stupnice u (a) prodloužené fonace a (b) čteného textu.	87
8.2	Schémata experimentálních systému pro automatické hodnocení chrap- tivosti podle spojité stupnice u (a) prodloužené fonace a (b) čteného textu.	88
9.1	Vzor formuláře pro subjektivní hodnocení nahrávek metodou CAPE-V.	110
10.1	Průvodní formulář subjektivního hodnocení.	113
10.2	Vzor formuláře pro subjektivní hodnocení nahrávek.	114

Seznam tabulek

1.1	Lékařské diagnózy spojené s chraptivostí	4
2.1	Přehled studií využívajících databázi Kay Elemetrics.	13
4.1	Čas potřebný k výpočtu HNR pro různé počty iterací algoritmu.	37
5.1	Kvantifikace Cronbachova alfa slovním popisem.	47
5.2	Cohenovo kappa – příklad, kontingenční tabulky.	47
5.3	Fleissovo kappa – příklad frekvenční tabulky.	49
5.4	Náhled na hodnocení – korelační koeficienty.	50
5.5	Porovnání přísnosti subjektivního hodnocení při opakovaném testu.	51
5.6	Vzájemná korelace parametrů G, R, B, T.	51
5.7	Normovaná vzdálenost hodnocení parametrů G, R, B, T.	52
5.8	Výsledky Cronbachova alfa.	53
5.9	Výsledky Cohenova kappa.	53
5.10	Výsledky Fleissova kappa.	54
6.1	Seznam symbolů použitý k popisu parametrizace.	59
6.2	Matice záměn pro automatické objektivní hodnocení diskrétního G u prodloužené fonace.	64
6.3	Matice záměn pro automatické objektivní hodnocení diskrétního R u prodloužené fonace.	64
6.4	Matice záměn pro automatické objektivní hodnocení diskrétního B u prodloužené fonace.	65
6.5	Matice záměn pro automatické objektivní hodnocení diskrétního T u prodloužené fonace.	66
6.6	Matice záměn pro automatické objektivní hodnocení diskrétního G u čteného textu.	71
6.7	Matice záměn pro automatické objektivní hodnocení diskrétního R u čteného textu.	72
6.8	Matice záměn pro automatické objektivní hodnocení diskrétního B u čteného textu.	72
6.9	Matice záměn pro automatické objektivní hodnocení diskrétního T u čteného textu.	74
7.1	Srovnání výsledků automatické klasifikace s publikovanými výsledky.	80
7.2	Parametrizace použité k diskrétnímu hodnocení chraptivosti.	83
7.3	Parametrizace použité ke spojitému hodnocení chraptivosti.	84
8.1	Porovnání výsledků subjektivního a objektivního hodnocení pomocí Cohenova κ	90

10.1	Počty nahrávek pro jednotlivé úkoly.	112
11.1	Kontingenční tabulky subjektivního hodnocení pro sloučenou skupinu hodnotitelů pro prodlouženou fonaci.	116
11.2	Kontingenční tabulky subjektivního hodnocení pro hodnotitele A pro prodlouženou fonaci.	116
11.3	Kontingenční tabulky subjektivního hodnocení pro hodnotitele B pro prodlouženou fonaci.	117
11.4	Kontingenční tabulky subjektivního hodnocení pro hodnotitele C pro prodlouženou fonaci.	117
11.5	Kontingenční tabulky subjektivního hodnocení pro hodnotitele D pro prodlouženou fonaci.	118
11.6	Kontingenční tabulky subjektivního hodnocení pro hodnotitele E pro prodlouženou fonaci.	118
11.7	Kontingenční tabulky subjektivního hodnocení pro sloučenou skupinu hodnotitelů pro čtený text.	119
11.8	Kontingenční tabulky subjektivního hodnocení pro hodnotitele A pro čtený text.	119
11.9	Kontingenční tabulky subjektivního hodnocení pro hodnotitele B pro čtený text.	120
11.10	Kontingenční tabulky subjektivního hodnocení pro hodnotitele C pro čtený text.	120
11.11	Kontingenční tabulky subjektivního hodnocení pro hodnotitele D pro čtený text.	121
11.12	Kontingenční tabulky subjektivního hodnocení pro hodnotitele E pro čtený text.	121
12.1	Vážený průměr hodnot popisujících diskrétní klasifikaci.	122
12.2	Detailní přehled hodnot popisujících diskrétní klasifikaci.	123

Seznam zkratek a symbolů

$1 - \bar{P}_e$	míra shody mezi hodnotiteli dosažená náhodou
α	směrnice regresní přímky
α_C	Cronbachovo alfa
a_m	m-tý koeficient LPC analýzy
A	Aestehnicity (hlasová slabost)
$A(z)$	model hlasového traktu pomocí LPC analýzy
ACR[k]	autokorelační funkce
ACRF0	odhad hlasivkové frekvence pomocí autokorelační funkce
AMDF	Average Magnitude Difference Function
APQ	Amplitude Perturbation Quotient (koeficient amplitudové odchylky)
AR	autoregresní
B	Breathiness (dyšnost)
BPL	Band Pass Liftering
BRI	Breathiness Index (parametrizace signálu)
CAPE-V	Consensus Auditory Perceptual Evaluation of Voice (metoda subjektivního hodnocení hlasu)
CEP	Cepstrum
CEPSF0	odhad hlasivkové frekvence v keprální oblasti
CHNR	Cepstrum-based HNR (HNR odhadnuté v keprální oblasti)
CL	Curve Length (délka křivky)
corelF0cepsAcr	korelace mezi průběhem CEPSF0 a ACRF0
corelF0cepsSpec	korelace mezi průběhem CEPSF0 a SPECF0
corelF0specAcr	korelace mezi průběhem ACRF0 a SPECF0
CPP	Cepstral Peak Prominence
CWT	spojitá vlnková transformace
d	délka mezery mezi harmonickými částmi spektra při výpočtu HNR
δ	parametrizace časového průběhu dat
D_i	šumová část spektra
DFT	diskrétní Fourierova transformace
DSP	Digital Signal Processing (číslicové zpracování signálu)
EGG	elektroglottografie
$E_n(x[n])$	energie signálu $x[n]$
EPF	Energy Perturbation Factor
EPQ	Energy Perturbation Quotient
f_0	fundamentální frekvence, hlasivkový tón, někdy také <i>pitch</i>
f_s	vzorkovací frekvence signálu
FATR	Frequency Amplitude Tremor (frekvenční amplitudový tremor)
FDA	Frequency domain analysis (metoda odhadu f_0)
FN	False Negative (falešně negativní)

FP	False Positive (falsešně pozitivní)
G	Grade (celkový dojem z hlasu)
GMM	Gaussian Mixture Model (směs gausovských modelů)
GNE	Glottal-to-Noise Excitation ratio (míra vybuzení signálu hlasivkami)
GRBAS	metoda subjektivního hodnocení chraptivosti (G – Grade, R – Roughness, B – Breathiness, A – Aesthenicity, S – Strain)
GRBT	metoda subjektivního hodnocení chraptivosti (dtto, T – Tension)
GVA	Generalized Variogram Analysis (parametrizace signálu)
GVA_Topt	optimální posun okna při parametrizaci pomocí GVA
H	počet hodnotitelů subjektivních testů
H1A	poměr 1. a 2. harmonické signálu (parametrizace signálu)
HLR	Ratio of high- to mid/low-frequency energy (parametrizace signálu)
HOS	Higher Order Statistics (statistika vyšších řádů)
HNR	Harmonics-to-Noise Ratio
IDFT	inverzní (zpětná) diskrétní Fourierova transformace
\Im	imaginární část komplexního čísla
JIT	jitter (frekvenční nestabilita hlasivkového tónu)
K	počet opakovaného hodnocení subjektivního testu
κ_C	Cohenovo kappa
κ_F	Fleissovo kappa
L	řád LPC modelu
LPC	Linear Predictive Coding (lineární prediktivní kódování)
LPCC	kepstum získané pomocí LPC
LRE	Linear Regression Error (chyba lineární regrese)
LSF	Line Spectral Frequencies (parametrizace signálu)
LSP	Line Spectral Pairs (parametrizace signálu)
LVQ	Learning Vector Quantization (klasifikátor)
μ	střední hodnota souboru dat
M	délka okna (mikrosegmentu) signálu nebo medián souboru hodnot, viz kontext
M_m	m-tý spektrální moment
max	maximální hodnota souboru dat
maxabs	maximální hodnota absolutních hodnot souboru dat
MDVP	Multi-Dimensional Voice Program
MFCC	Mel-Frequency Cepstral Coefficients
min	minimální hodnota souboru dat
MPT	Maximum Phonation Time (délka fonace)
MWC	Mean Waveform Matching coefficient
N	délka DFT
NDE	Normalized Distance of Evaluation (normalizovaná vzdálenost hodnocení)
NNE	Normalized Noise Energy
ω_k	frekvence LSF při parametrizaci pomocí LSP
P	nahrávky chraptivých (Patologických) pacientů

p	hladina významnosti
$\bar{P} - \bar{P}_e$	míra skutečně dosažené shody mezi hodnotiteli
$Pr(a)$	relativní shoda mezi hodnotiteli
$Pr(e)$	hypotetická pravděpodobnost náhodně shody
$P(z)$	polynom použitý při výpočtu LSP
P_q	spektrální výkon q -té spektrální čáry
P_i	harmonická část spektra
PIP	Peak Intraoral Pressure (špičkový vnitroústní tlak)
PLI	Pathological Likelihood Index
PPQ	Pitch Perturbation Quotient (koeficient frekvenční odchylky)
PPF	Pitch Perturbation Factor
$Q(z)$	polynom použitý při výpočtu LSP
\bar{r}	průměr horního trojúhelníku korelační matice
r	Pearsonův lineární korelační koeficient
ρ	Spearmanův korelační koeficient
R	Roughness (chraplavost)
range	rozsah hodnot souboru dat
RAPT	Robust Algorithm for Pitch Tracking
RBF	Radial Basis Function (metoda transformace dat)
\Re	reálná část komplexního čísla
RPK	Pearson r at autocorrelation peak (parametrizace signálu)
σ	směrodatná odchylka souboru hodnot
σ_R	relativní směrodatná odchylka souboru hodnot
$s[n]$	periodická část hlasového signálu
S	Strain (hlasové napětí)
SD	Spectral Deformation (parametrizace signálu)
SEN	Senzitivita
SHI	shimmer (amplitudová nestabilita hlasivkového tónu)
SIFT	Simple Inverse Filter Tracking
SNR	Signal-to-Noise Ratio (odstup signálu od šumu)
SPE	Specificita
SPECF0	odhad hlasivkové frekvence ve frekvenční (spektrální) oblasti
SPS	Signal Periodicity Strength (parametrizace signálu)
SR	Spectral Rolloff (parametrizace signálu)
SVM	Support Vector Machine (klasifikační algoritmus)
SW	Spectral Width (parametrizace signálu)
Θ_k	kořen polynomu $P(z)$ nebo $Q(z)$ při parametrizaci pomocí LSP
T	vzorkovací perioda ($=1/f_s$) nebo Tension (hlasové napětí), viz kon- text
T_0	základní perioda signálu ($=1/f_0$)
TN	True Negative (správně negativní)
TP	True Positive (správně pozitivní)
UEP	Union of European Phoniaticians – Evropské sdružení foniatrů
V-RQOL	Voice Related Quality of Life (metoda subjektivního hodnocení hlasu pacientem)

VHI	Voice Handicap Index (metoda subjektivního hodnocení hlasu pacientem)
VRPA	Voice Range Profile Area (plocha hlasového pole)
$w[n]$	šumová část hlasového signálu
WT	diskrétní vlnková transformace
$x[n]$	diskrétně vzorkovaný signál
$x(t)$	signál v časové oblasti
$X[n]$	spektrum signálu $x[n]$
$\tilde{x}[n]$	kepstrum signálu $x[n]$
$x_{\Delta\Delta}[n]$	druhá derivace diskrétně vzorkovaného signálu
YIN	metoda odhadu hlasivkové frekvence (z „yin“ a „yang“ z orientální filozofie)
Z	nahrávky Zdravých pacientů

1. Úvod

Lidský hlas je základní komunikační nástroj, který využívá většina lidí v každodenním životě. Na základě hlasové komunikace si jedinci utváří základní náhled na protějšek, se kterým komunikují. Jakékoliv patologie, které se objeví na orgánech používaných při tvorbě hlasu, tzv. mluvidlech, se zpětně projeví ve větší či menší míře na výsledném akustickém signálu jevem zvaným chraptivost. Nejenže je vhodné tyto patologie co nejdříve odstranit, aby daný jedinec předešel vážnějším zdravotním komplikacím, ale podle výzkumu uveřejněného v Amir and Levine-Yundof [8] má subjektivní vjem posluchače zásadní vliv na posouzení mluvčího. Mluvčí s chraptivostí jsou častěji posuzováni jako nespolehliví, hloupí, sexuálně nepřítažliví a zlí. Nicméně v běžném životě je možné setkat se s opakem: mnoho herců či zpěváků si chraptivý hlas „pěstuje“, je jejich poznávacím znamením a jistě přitahuje mnoho diváků či posluchačů.

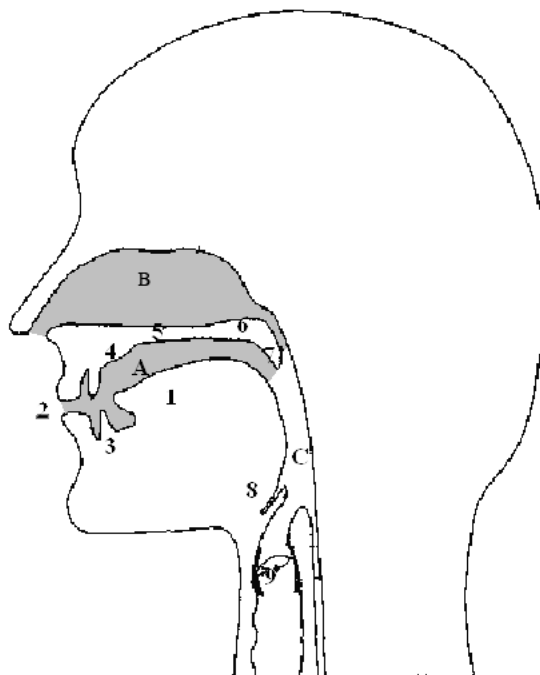
Péče o hlas je tedy velmi důležitá nejen pro hlasové profesionály, jako například televizní a rozhlasové moderátory, zpěváky a herce, ale i pro běžné lidi, kteří jsou „nuceni“ v každodenním životě hlas používat. Z toho důvodu je vhodné mít prostředky, které hlasové patologie co nejdříve odhalí, s jejichž pomocí lze takové patologie odstranit a v neposlední řadě je důležité mít prostředky pro kvalitní terapii, s jejíž pomocí se hlasový projev vrátí co nejdříve do normálu.

Následující odstavce v základních bodech popisují, jak dochází ke vzniku řečového signálu, je definován pojem chraptivosti a dále je nastíněn souhrn patologií hlasového ústrojí, které mohou vést ke vzniku chraptivosti. Zájemci o odbornější úvod do problematiky naleznou mnohem více informací v lékařské literatuře, anatomický popis především v otorinolaryngologických knihách, např. Hahn [54], Hybášek [67] a funkční vznik hlasu a produkci řeči ve fonetických knihách, např. Palková [113], Vokřál [143]. Při vzniku následujících odstavců byla využita výše zmíněná literatura.

1.1 Hlas

Hlasový signál je tvořen orgány, které jsou souhrnně nazývány hlasové ústrojí nebo mluvidla, viz obr. 1.1. Na tomto obrázku nejsou zobrazeny plíce, které se sice do mluvidel nezahrnují, nicméně jsou pro tvorbu hlasu velmi důležité: vytváří vzduchový proud, který prochází přes mluvidla a tím vytváří charakteristický zvuk – hlas.

Jednotlivé hlásky můžeme dělit například podle toho, jaké části mluvidel se při jejich tvorbě zapojují a tím pádem i jak dané hlásky znějí. Jedno z nejhrubších dělení nastane, rozdělíme-li hlásky na ty, které při tvorbě zapojují hlasivky (znělé hlásky) a na ty, které hlasivky nezapojují (neznělé hlásky).



Obrázek 1.1: Popis mluvidel; *A* – ústní dutina (*cavum oris*), *B* – nosní dutina (*cavum nasi*), *C* – hltan (*pharynx*), 1 – jazyk (*lingua*), 2 – rty (*labia*), 3 – zuby (*dentes*), 4 – dásňový oblouk se sklípky (*alveoli*), v nichž jsou usazeny zuby, 5 – tvrdé patro (*palatum durum*), 6 – měkké patro (*palatum velum*) 7 – čípek (*uvula*), 8 – hrtanová záklopka (*epiglottis*), 9 – hlasivky (*glottis*); Převzato z [2].

1.1.1 Znělé hlásky

Jako znělé označujeme ty hlásky, při jejichž vzniku kmitají hlasivky. Hlasivky jsou párové, jsou umístěny v nejužším místě hrtanu a jsou tvořeny hlasivkovým vazem a hlasivkovým svaelem, který při semknutí, tedy napnutí, tvoří překážku vzduchu proudícímu z plic a tím dochází ke zvýšení subglottického tlaku. Pokud vznikne takový tlak, že ho hlasivky neudrží, oddálí se od sebe a nahromaděný vzduch může pronikat do dalších částí hlasového traktu. Po otevření hlasivek dojde k poklesu subglottického tlaku a hlasivky se opět uzavřou. Tímto způsobem vznikají kvazi-periodické rázy, které vnímáme jako hlasivkový tón (někdy označovaný jako pitch perioda, fundamentální frekvence, f_0). Tato frekvence závisí především na délce a napětí hlasivek.

Zvukové rázy jsou dále modifikovány v dalších dutinách (hltan, nosní, ústní), které fungují jako rezonátory a dále zvuk frekvenčně „obarvují“ – vznikají tzv. formantové frekvence.

Z výše napsaného tedy vyplývá, že znělé hlásky jsou tvořeny směsí kvaziperiodických rázů způsobených kmitáním hlasivek a dále přidanou šumovou částí, která je způsobena prouděním vzduchu z plic. U zdravých hlasů je toto šumové proudění minimální.

1.1.2 Neznělé hlásky

Neznělé hlásky vznikají zužováním, respektive uzavřením hlasového traktu, při kterém vzniká charakteristický šum. Často se jedná o zúžení v oblasti dutiny ústní, kdy se jazyk zvedá proti měkkému patru, tvrdému patru nebo alveolům¹, případně je překážka tvořena rty. Hlasivky jsou u neznělých hlásek povolené a nepředstavují pro vzduch proudící z plic žádnou překážku.

I u neznělých hlásek však mohou být v signálu přítomné formantové frekvence.

1.2 Chraptivost

Jak je zmíněno výše v části 1.1.1, znělé hlásky jsou tvořeny směsicí kvaziperiodických rázů vzniklých kmitáním hlasivek a šumovou příměsí. U zdravých hlasů, kdy všechny části hlasového ústrojí pracují normálně a v přiměřeném rozsahu, je šumová příměs minimální. U patologických hlasů však vlivem nesprávné funkce alespoň jedné části hlasového ústrojí dochází k různým nepravidelnostem, které se odráží ve výsledném hlasovém projevu. Jeden z prvních příznaků nesprávné funkce hlasového ústrojí je chraptivost.

Příčiny vzniku chraptivosti můžeme rozdělit do následujících skupin (Vokřál [143]):

- **změna hmoty** nebo **fyzikálních vlastností** hlasivek – vede ke změnám pravidelnosti kmitů hlasivek, takový hlas je pak označován jako **hrubý**, **drsný** nebo **chraptivý**;
- **nedomykavost hlasivek** – hlasivky nejsou schopny úplného uzavření a tím pádem dochází ke stálému proudění vzduchu z plic; v hlasu je zvýšená šumová příměs, takový hlas je pak označován jako **dyšný**.

Je ovšem nutné zmínit, že jistá nepravidelnost kmitání hlasivek je zcela fyziologická, i proto je výše v textu uváděno „kvaziperiodické“ kmitání hlasivek a nikoliv periodické. V některých oborech je nepravidelnost kmitání hlasivek přímo vyžadována, viz např. vibrato při zpěvu.

Dále je nutné zmínit zvláštní případ nepravidelnosti kmitání hlasivek, tzv. diplofonii. Tímto termínem je označován stav, kdy každá hlasivka kmitá s jinou frekvencí (např. různá tuhost hlasivek, různé napětí hlasivek způsobené parézou zvrtného nervu apod.), takže výsledný zvukový vjem je stejný, jako kdyby zněly dva různé tóny najednou (Vokřál [143]).

Diagnózy při chraptivosti

V tab. 1.1 je uveden výčet lékařských diagnóz, které mají za příčinu chraptivý a / nebo dyšný hlas. Tyto diagnózy jsou zde uvedeny s ohledem na dostupnou databázi promluv, která byla pořízena autorem dizertační práce Vokřál [143], na kterou tato práce navazuje.

¹Dásňový výstupek u horních zubů, na němž se tvoří hlásky, na obr. 1.1 je označen číslem 4.

Tabulka 1.1: Lékařské diagnózy spojené s chraptivostí a jejich popis ([143]).

diagnóza	popis	vjem
polyp	slizniční výchlípka vyplněná vazivem nebo organizovaný hematom	diplofonie
uzlík	ohraničené ztlustění sliznice bránicí úplnému závěru hlasivkové štěrbiny	dyšnost
cysta	zapouzdřený útvar uvnitř hlasivky, může způsobit nerovný okraj	dyšnost
afonie	hlasivky se k sobě nepřibližují, označováno také jako bezhlasí	dyšnost
parézy zvratného nervu	vede k nedomykavosti hlasivek	dyšnost
stav po chordektomii	odstranění jedné z hlasivek, např. kvůli nádoru	dyšnost
papilomatóza	mnohačetné papilómy na sliznici hrtanu a na hlasivkách	dyšnost + chraplavost
neurofibromatosis	podobné jako papilomatóza, výrůstky jsou na nervových zakončeních	dyšnost + chraplavost
edém, Reinkeho edém	prosáknutí podslizničního vaziva; k úplnému uzávěru dochází často jen v místě edému; hlasivka s edémem může kmitat na jiné frekvenci	dyšnost + chraplavost
diafragma hrtanu	vrozená anomálie hrtanu, mezi hlasivkami je blána, která je omezuje v pohybu	dyšnost + chraplavost
hematom	krevní výlev do hlasivky (změna hmoty)	chraplavost
sklerodermatosis	změna podslizničního vaziva, mění se tuhost hlasivek	chraplavost
M. Urbach-Weithe	dochází ke střežení lipidů v hlasivkách, mění se jejich tuhost	chraplavost
hyperkinetická dysfonie	zvýšené napětí hlasivek - následek nadměrné hlasové námahy	chraplavost
hypokinetická dysfonie	snížené napětí hlasivek, buď jako následek hyperkinetické dysfonie, nebo samostatná příčina	chraplavost
insuficience hlasivkového svalu	má za následek snížené napětí hlasivek	chraplavost
spastická dysfonie, hlasová neuróza, psychogenní dysfonie	zvýšené napětí hlasivek, které vede až ke křečovitému (spastickému) uzavření hlasivek	chraplavost
stav po strumektomii	může vést ke sníženému napětí hlasivek	chraplavost
mutační hlasy	hlas přeskakuje z vysoké polohy do nízké	chraplavost

1.3 Metody hodnocení chraptivosti

Hodnocení kvality hlasu provádí podvědomě každý posluchač. Takové hodnocení je označováno jako subjektivní a znamená, že zatímco jeden posluchač považuje daný hlasový projev za normální, příjemný, jiný posluchač může daný hlas považovat za nepříjemný, s jistou dávkou patologie. V technickém světě je oproti tomu snaha o objektivní, strojové hodnocení, které je nezávislé na posouzení člověka.

1.3.1 Subjektivní metody

Subjektivní hodnocení kvality hlasu je založené na kvantifikaci percepčního vjemu posluchače. V průběhu let bylo publikováno několik měř, které slovně definovaly pojmy hodnocené v hlasovém projevu a snažily se o přibližný slovní popis jednotlivých „kvantizačních“ stupňů.

Nevýhodou subjektivních měř je nutnost značné zkušenosti testujícího, která beztak nezajistí stejné hodnocení totožné nahrávky s delším časovým odstupem, natož stejné hodnocení totožné nahrávky dvěma testujícími. Nicméně je nutné dodat, že zkušeni testující jsou ve svém hodnocení značně konzistentní, viz např. De Bodt et al. [29] nebo Karnell et al. [72]. V uvedených publikacích dosahují hodnotící velmi podobných výsledků. Ze závěrů v De Bodt et al. [29] a Karnell et al. [72] lze usoudit, že při opakovaném poslechu byl hodnotící ve většině případů přísnější – nahrávku hodnotil vyšším stupněm patologie. Podrobný rozbor hodnocení s časovým odstupem uvádí také Vokřál [143].

Při subjektivním posouzení chraptivého hlasu jsou nejčastěji posuzovány následující parametry (Vokřál [143]):

chraptivost	celkový dojem z hlasu jako celku
chraplavost	vjem spojený s nepravidelným kmitáním hlasivek
dyšnost	vjem spojený s mírou slyšitelného šumu v hlase, který vzniká v důsledku neúplného uzavření hlasivek
napětí v hlase	spojené se subjektivním vjemem slabosti nebo naopak přemáhání hlasu

Doporučení UEP

Evropské sdružení foniatrů (UEP – Union of European Phoniaticians) zavedlo šestibodovou stupnici pro hodnocení chraptivosti [1], která je definována následovně:

- 0 – normální hlas
- 1 – zastřený hlas
- 2 – mírná dysfonie
- 3 – středně těžká dysfonie
- 4 – těžká dysfonie
- 5 – afonie

Tato stupnice nerozlišuje mezi chraplavým a dyšným hlasem, celý dojem z hlasu popisuje pouze jednou hodnotou.

GRBAS

Metoda GRBAS (Hirano [63]) je doporučena Japonskou společností pro logopedii a obsahuje ve svém názvu první písmena hodnocených aspektů:

- G** rade – celkový dojem z hlasu jako celku
- R** oughness – chraplavost², míra nepravidelnosti kmitání hlasivek
- B** reathiness – dyšnost, míra šumové příměsi v hlase
- A** esthenicity – slabost
- S** train – napětí

Každý aspekt je hodnocen čtyřbodovou stupnicí v rozmezí 0-3 body, přičemž 0 bodů znamená, že daný aspekt není v hlasovém projevu přítomný, zatímco 3 body značí, že daný faktor je zastoupen v extrémní míře. Výsledné hodnocení je zapsáno například následovně: $G_2R_1B_3A_1S_1$.

Na stupnici GRBAS je vypracováno množství studií, ve kterých se autoři zabývají spolehlivostí metody GRBAS jako celku (Alpan et al. [6], Alpan et al. [7], Bhuta et al. [18], De Bodt et al. [29], Fredouille et al. [42], Godino-Llorente et al. [48], Godino-Llorente et al. [49], Lee and Hahn [77], Ma and Yiu [80], Maryn et al. [90], Saenz-Lechon et al. [126], Yu et al. [151]) a dále analyzují spolehlivost jednotlivých parametrů. Ve studiích De Bodt et al. [29] a Karnell et al. [72] shledali parametr G nejvíce spolehlivým a parametry A a S nejméně spolehlivé. Oproti tomu závěry uveřejněné v Wuyts et al. [148] ukazují parametr A jako druhý nejspolehlivější po parametru G. Dále De Bodt et al. [29] odkazuje na studie, ve kterých autoři na základě provedeného testu dospěli k závěru, že testující se ve svém hodnocení více liší u patologických než u zdravých hlasů. Nicméně jejich závěr je, že metoda GRBAS je dostatečně spolehlivým subjektivním hodnocením.

Godino-Llorente et al. [48] používá k ověřování navrženého parametru PLI (Pathological Likelihood Index) celkové známky promluvy, tzn. jsou sečtena jednotlivá dílčí hodnocení GRBAS. Např. promluva hodnocená $G_2R_0B_2A_1S_2$ získá celkovou známku 7. Takové hodnocení ovšem nezohledňuje jednotlivá dílčí hodnocení promluvy (dyšnost, chraplavost).

Modifikace GRBT byla zavedena v práci Dejonckere et al. [32], kde T označuje tonus nebo-li napětí v hlase. Hodnota parametru T může nabývat hodnot -2 až +2 body, kde záporná hodnota popisuje hlasovou slabost a kladná hodnota napětí v hlase. Stupnice GRBT je použita v práci Vokřál [143] a je dále používána v této práci.

CAPE-V

Metoda CAPE-V (Consensus Auditory Perceptual Evaluation of Voice, [11]) byla vyvinuta v roce 2002. Tato stupnice hodnocení obsahuje některé parametry shodné

²V češtině není překlad slova *roughness* z pohledu kvality hlasu jednoznačný – lze jej přeložit i termínem *drsnost*, případně *hrubost*. V rámci této práce však bude užíváno termínu *chraplavost*, který dle názoru autora této práce není citově zabarvený jako dva zmíněné alternativní překlady.

s metodou GRBAS, konkrétně G, R, B a S, a dále přidává „pitch“, (výšku hlasu) a „loudness“ (hlasitost). Jako novinku zavádí grafické znázornění výsledků, kdy jednotlivé parametry testující zakresluje jako vodorovné úsečky určité délky do formuláře, viz příloha 9.

Hodnocení metodou CAPE-V ve svých pracích použili např. Karnell et al. [72] nebo Lowell et al. [79].

V Karnell et al. [72] je také uveřejněna srovnávací studie, která porovnává metody GRBAS a CAPE-V. Z výsledků vyplývá, že obě metody jsou zhruba stejně přesné, nicméně díky grafickému znázornění výsledků v metodě CAPE-V je možné lépe posuzovat vývoj poruchy hlasu pacienta v čase během následujících sezení. Karnell et al. [72] také uvádí srovnání obou metod se subjektivním hodnocením pacienta (V-RQOL – Voice related quality of life a IPVI – Iowa Patient’s Voice Index).

Hodnocení prováděné pacientem

Metody V-RQOL (Voice Related Quality of Life, Hogikyan and Sethuraman [65]), VHI (Voice Handicap Index, Jacobson et al. [68], Švec et al. [141]) nebo IPVI jsou metody subjektivního hodnocení hlasu, které provádí sám pacient. Spočívají ve vyplnění formuláře s otázkami typu „Do jaké míry pociťujete potíže při mluvení“ apod.

1.3.2 Objektivní metody

Kontaktní metody

Existuje několik přístupů, kterými lze objektivně vyšetřit funkčnost hlasového ústrojí. Některé z těchto metod se zaměřují na vyšetření aerodynamických (např. spirometrie, pneumografie), jiné na měření elektrických veličin (elektroglottografie – EGG, elektromyografie) a jiné přímo vizuálně zobrazují různé části hlasového ústrojí při tvorbě hlasu (laryngoskopie, laryngostroboskopie, ...). Tyto metody poskytují přesný obraz stavu hlasového ústrojí a mají nezastupitelný význam při určení příčiny chraptivosti. Nicméně nejsou součástí analýz uvedených v této práci.

Bezkontaktní metody

Další skupina metod objektivního vyšetření hlasu je založená na analýze akustického signálu tvořeného hlasovým ústrojím. Z technického pohledu je jediná metoda, která zpětně neovlivňuje vznik hlasu, metoda akustická. Při akustické analýze je hlas pacienta snímán mikrofonom, digitalizován a následně číslicově zpracováván za účelem nalezení vhodných parametrů, které oddělují zdravé a patologické hlasy.

Jednou z největších výhod akustické analýzy hlasu je fakt, že jde o neinvazivní vyšetření, při kterém pacient nepociťuje žádné nepohodlí. Také finanční a hygienické nároky jsou neporovnatelně nižší v porovnání s ostatními zmíněnými metodami objektivního vyšetření hlasu.

Podrobný popis akustických parametrů použitých v této práci je uveden kapitole 4.

2. Současný stav řešené problematiky

2.1 Používané parametrizace

Objektivní akustická analýza chraptivého hlasu je typicky prováděna buď na prodloužené fonaci hlásky /a/, nebo na čteném textu. Zatímco analýza prodloužené fonace je technicky snazší, důležitější pro hodnocení chraptivosti v hlase má analýza spojitě řeči, tedy např. čteného textu.

Důvod, proč je analýza prodloužené fonace snazší, je poměrně zřejmý: v části 1.2 je popsána příčina chraptivosti jako změna fyzikálních vlastností hlasivek nebo jejich nedomykavost. Při prodloužené fonaci se tyto dvě příčiny naplno projeví a výsledný akustický signál jimi bude velmi ovlivněn. Bohužel, prodloužená fonace má blíže ke zpěvu než k běžné každodenní komunikaci. Při běžné komunikaci, kterou lze čteným textem poměrně dobře nasimulovat, dochází k většímu zatížení hlasového traktu jako celku. Hlasivky musí začínat kmitat v nepravidelných intervalech podle toho, zda je aktuální čtený text znělý nebo neznělý a také ostatní části mluvidel se musí pravidelně přestavovat v závislosti na právě artikulované hlásce. A právě při zatížení hlasového traktu mluvenou řečí se dysfunkce projeví více. Na druhou stranu je nutné se při automatické analýze s takovýmto neustálým přeuspořádáváním mluvidel a objeováním a ztracením fonace vypořádat.

2.1.1 Analýza hlasivkového tónu

Mezi nejstarší metody analýzy chraptivosti patří analýza parametrů spojených s hlasivkovým tónem f_0 . Tento postup zahrnuje detekci hlasivkového tónu a jeho následnou analýzu, jako určení jeho frekvenční (jitter – JIT) a amplitudové (shimmer – SHI) nestability a dále parametry jako APQ (Amplitude Perturbation Quotient – koeficient amplitudové odchylky), PPQ (Pitch Perturbation Quotient – koeficient frekvenční odchylky), FATR (Frequency Amplitude Tremor – frekvenční amplitudový tremor),... Obecně lze říci, že v posledních letech se od výzkumu v této oblasti ustupuje. Pokud některý z těchto parametrů autor používá, použil k jeho výpočtu některý z komerčních programů jako např. MDVP (Multi-Dimensional Voice Program), Dr. Speech, nebo některý z volně dostupných, ale odbornou komunitou uznávaných programů jako např. Praat (Boersma and Weenink [21]). Většina autorů pouze zmiňuje, že tyto parametrizace existují, ale dále se jejich zkoumáním a porovnáváním nezabývá. Tento trend potvrzuje Werth et al. [146], který tyto parametrizace aplikuje (za použití programu *Doctor Speech analyzing software* od firmy *Tiger Electronics Inc., USA*) na své databázi (cca 1700 hlasů) a dosahuje úspěšnosti klasifikace 75-90% (v závislosti na použitých klasifikátorech). Uvádí také normativní hodnoty a konstatuje, že SHI a HNR (Harmonics-to-Noise Ratio) se liší podle

pohlaví, zatímco JIT a NNE (Normalized Noise Energy) se neliší.

Jeden z důvodů, proč se postupně upouští od použití těchto parametrů je především fakt, že odhad f_0 v patologickém hlase je značně komplikovaná úloha a všechny výše zmíněné parametry na detekci f_0 přímo závisí. V literatuře je dostupných několik porovnání metod extrakce f_0 . Např. Manfredi et al. [82] porovnává několik metod (AR (Autoregressive), AMDF (Average Magnitude Difference Function), SIFT (Simple Inverse Filter Tracking), CEP (kepstrální), WT (vlnková) a CWT (spojitá vlnková)) jak na syntetických, tak na reálných signálech (nahrávky pacientů po laserové operaci hlasivek). Výsledky jsou prezentovány jako směrodatná odchylka chyby dané metody, přičemž jako správné f_0 použili hodnoty naměřené DSP systémem firmy *Kay Elemetrics*. Z prezentovaných výsledků nelze jednoznačně určit nejlepší metodu – pro různé pacienty poskytují nejlepší výsledky různé metody. Lze ovšem usoudit, že v průměru mají nejmenší chybu odhadu, okolo 6 %, metody *AMDF* a *CWT*. Další porovnání provedl Tavares et al. [138] (metody AR, YIN (de Cheveigné and Kawahara [30]), FDA), který se zaměřil na závislost přesnosti detekce f_0 na SNR, ale opět nepředkládá jednoznačné výsledky a ponechává závěr na čtenářích. Z prezentovaných výsledků se nejrobustněji jeví metoda FDA (Frequency Domain Analysis).

2.1.2 Analýzy v časové oblasti

Analýzy založené na lineárním prediktivním kódování (LPC) vyvolávají mezi autory rozporuplné názory. Zatímco Neto et al. [110] tvrdí, že LPC koeficienty samy o sobě dostatečně přesně oddělují patologické a zdravé hlasy (dokazuje experimentem) a Costa et al. [26] ukazuje, že i malé změny ve vokálním traktu se projeví na LPC koeficientech, Godino-Llorente et al. [47] tvrdí, že LPC parametrizace ze své podstaty předpokládá lineární model hlasového traktu, což právě u patologických hlasů splněno není. To je také jeden z důvodů, proč se rozvinula parametrizace chybového signálu lineární predikce (de Oliveira Rosa et al. [31], Jo et al. [69], Moore and Clements [100], Naylor et al. [109], Veeneman and BeMent [142]).

Zajímavý pokus realizoval Lee et al. [78], který z nahrávky v časové oblasti spočítal 3. a 4. normalizovaný centrální moment (šikmost a špičatost) a zjistil, že se pro zdravé a patologické hlasy liší.

2.1.3 Analýzy ve spektrální oblasti

Hledání subharmonických složek ve spektru signálu je další z metod analýzy chraptivosti. Vznik subharmonických složek není přesně objasněn, ale podle Omori et al. [112] je způsoben nestejnou hmotou hlasivek. V databázi 389 pacientů trpících různým stupněm chraptivosti objevil Omori et al. [112] subharmonické složky pouze u 20 nahrávek. Upozorňuje ale na případy, kdy byl jitter i shimmer v mezích normy, ale v signálu byly přítomné subharmonické složky a výsledný subjektivní dojem z hlasu byl patologický (chraplavý).

V poslední době je poměrně často používána parametrizace založená na parametru GNE – Glottal-to-Noise Excitation Ratio. Výhoda této parametrizace spočívá

v její nezávislosti na detekci f_0 . Parametr GNE představil v roce 1997 Michaelis et al. [97] a jeho výpočet je založen na vzájemné korelaci Hilbertových obálek signálu filtrovaného různými pásmovými propustmi. Jeho fyzikální význam spočívá ve schopnosti postihnout, zda daný úsek řečového signálu vznikl po vybuzení hlasivkami nebo frikativním šumem – čím více frikativního šumu v signálu je, tím menší bude vzájemná korelace mezi obáčkami. Fröhlich et al. [43] použil GNE v diagramu chraptivosti k měření míry šumové příměsi v hlase. Tento parametr zvolil z důvodu jeho větší vypovídací schopnosti při použití s hodnotami jitteru a shimmeru, než má například NNE nebo CHNR (cepstrum-based harmonic-to-noise ratio). Godino-Llorente et al. [49] se snažil určit normativní hodnoty pro tento parametr na databázi firmy *Kay Elemetrics*.

Diagram chraptivosti (hoarseness diagram), představený v práci Fröhlich et al. [43] je výsledkem úsilí o přehledné kvantitativní zobrazení kvality řeči, konkrétně periodičnosti a míry šumové příměsi. Diagram je dvourozměrný obrázek, kde jsou na jedné ose vyneseny parametry spojené s chraplavostí a na druhé ose parametry spojené s dyšností. Pozice na ose chraplavosti je dána kombinací parametrů jitteru, shimmeru a vzájemné korelace mezi následujícími glottálními cykly odhadnutými inverzní filtrací a pozice na ose dyšnosti je určena za pomoci parametru GNE. Fröhlich et al. [43] provádí experiment na čtyřech typech hlasů a ukazuje, jak jsou tyto hlasy rozmístěny po výsledném diagramu chraptivosti.

Markaki and Stylianou [87] se ve své práci snaží klasifikovat jednotlivé patologie (otok hlasivek, paralýza, polyp, ...). Jako parametrizaci používají tzv. modulační spektrum (modulation spectra), což je de facto spektrum spočítané ze spektra, resp. z časového vývoje určité spektrální čáry, nebo také signálu filtrovaného pásmovou propustí, jejíž centrální frekvence a šířka pásma odpovídá dané spektrální čáře. Po aplikaci algoritmů na redukci příznakového prostoru (původně získal matici 257x257 vzorků, po redukci 34x34 vzorků) klasifikovali za pomoci SVM s RBF s průměrnou přesností přes 90%. Analýze modulačního spektra se věnovali např. Markaki et al. [86], Vích and Vondra [140], Xiao et al. [149].

Zajímavou metodu detekce patologických hlasů představil Costa et al. [26], který používá měření entropie výkonového spektra signálu. Ve svém pokusu použil jak Shanonovu, tak relativní entropii. Závěr tohoto pokusu je, že relativní entropie poskytuje lepší klasifikační výsledky – 77% oproti 73%.

2.1.4 Analýzy v keprstrální oblasti

Kepstrální analýza akustického signálu je často používaná metoda (např. Arians-Londoño Ando et al. [10], Costa et al. [25, 27], Fraile et al. [40, 41], Godino-Llorente et al. [47], Lee et al. [78], Neto et al. [110], Tavares et al. [137]), kterou autoři používají ve snaze oddělit vliv zdroje buzení (hlasivky) a přenosového systému (hlasové ústrojí). Výhoda této parametrizace spočívá v možnosti výpočtu bez nutnosti odhadu f_0 . Autoři jsou ovšem nejednotní v otázce vhodné metody výpočtu keprstra ve smyslu rekurzivního výpočtu z LPC koeficientů, nebo výpočtu ze spektra. Tato nejednotnost vyplývá především z ambivalentního názoru na vhodnost LPC analýzy na patologických hlasech, viz výše v textu, proto jsou běžně k výpočtu používány

obě metody. Výhoda výpočtu ze spektra spočívá v přesnějším popisu buzení pomocí f_0 za cenu nepřesnějšího popisu přenosového systému (kvůli zohlednění šumu), zatímco při použití výpočtu z LPC koeficientů je lépe modelován vliv přenosového systému. Většina autorů se kloní k použití MFCC (Mel-Frequency Cepstral Coefficients), které lépe odráží vnímání lidského ucha, ale Fraile et al. [41] podotýká, že zatím není fyzikálně objasněno, jaký mají MFCC koeficienty vztah k patologii hlasu (nejasnosti panují především ohledně nelineární Mel-frekvenční transformace). Fraile et al. [41] používá MFCC koeficienty nejen k separaci zdravých a nemocných hlasů, ale také ukazuje, že je možné MFCC koeficienty použít i k odlišení pohlaví mluvčího a konstatuje větší rozptyl parametrů u žen než u mužů – tím pádem je u žen složitější dosáhnout správné klasifikace. Neto et al. [110], který se zabývá otoky hlasivek, používá LPCC (LPC keprstrální koeficienty) a MFCC koeficienty a konstatuje, že LPCC koeficienty nejlépe oddělí promluvy pacientů s otokem hlasivek od ostatních patologií, ale MFCC koeficienty nejlépe oddělí zdravé a patologické promluvy. V experimentu uvedeném v Godino-Llorente et al. [47] dosahují nejlepších výsledků pro 16 MFCC koeficientů. Tavares et al. [137] používá LPCC koeficienty a dále vážené keprstrální koeficienty (tzv. liftering, používá období pásmové propusti, BPL – band pass liftering). Podle prezentovaných výsledků dosahuje úspěšnosti klasifikace až 98% pro otoky a až 96% pro paralýzy hlasivek. Fraile et al. [41] představuje nový přístup výpočtu keprstrálních koeficientů. Tento nový výpočet zachovává robustnost MFCC parametrizace, ale při výpočtu není použita Mel-frekvenční transformace – výpočet MFCC je prováděn z vyhlazeného spektra signálu, proto je zachována fyzikální podstata výpočtu keprstra.

Parametrů MFCC použili pro analýzu a klasifikaci chraptivosti s velmi dobrými výsledky např. Fredouille et al. [42] a Saenz-Lechon et al. [126].

Je obecně uznáváno, že svou roli mají i první a druhá derivace keprstrálních koeficientů (označované jako Δ a $\Delta\Delta$), které de facto popisují časový vývoj. Autoři vychází z předpokladu, že patologický hlasový trakt bude více časově nestabilní a proto derivace keprstrálních koeficientů spočítaných ze signálu vzniklého v patologicky postiženém hlasovém traktu, budou vykazovat větší směrodatnou odchylku. Godino-Llorente et al. [47] uvádí výsledky studie, ze které vyplývá nedostatečná diferenciací $\Delta\Delta$ koeficientů v porovnání s Δ keprstrálními koeficienty.

Cepstral Peak Prominence (CPP, Hillenbrand and Houde [61]) je velmi rozšířený parametr pro analýzu jak prodloužené fonace, tak spojitě řeči (Alpan et al. [7], Awan et al. [12], Balasubramanium et al. [13], Herman-Ackah [59], Radish Kumar et al. [123], Lowell et al. [79]). Jeho výpočet spočívá v odečtení rozdílu amplitudy špičky v keprstru v rozmezí 3,3-16,7 ms (60-300 Hz) a regresní přímky, která aproximuje průběh keprstra. Parametr využívá faktu, že hlasy s dobře definovanou harmonickou strukturou obsahují jasněji definovanou keprstrální špičku, zatímco hlasy s větším množstvím šumové příměsi ne. Výsledky autorů ukazují jeho vhodnost při analýze chraptivosti.

Herman-Ackah [59] ukazuje nutnost výpočtu regresní přímky při určování CPP. Awan et al. [12] ve svém experimentu ukazuje, že intenzita při fonaci do jisté míry ovlivňuje výslednou hodnotu CPP.

Trend v analýze patologických hlasů směřuje spíše do roviny klasifikační než

parametrizační, to znamená, že nejsou vyvíjeny nové parametrizace, ale spíše jsou rozvíjeny metody dolování dat, které pomáhají při výběru vhodných parametrů (redukci příznakového prostoru) a jejich vzájemné kombinaci.

2.2 Používané databáze a metody subjektivního hodnocení

Používané databáze se jednak liší původem (syntetické signály, komerční databáze, vlastní databáze), počtem nahrávek a jejich původem (zdravé vzorky (Z) a patologické vzorky (P)) a způsobem hodnocení (subjektivní vs. objektivní hodnocení).

2.2.1 Syntetické signály

Řada autorů ve svých experimentech používá syntetické signály, tedy signály, které byly vygenerovány algoritmicky. Takový postup má výhodu v přesně definovaných parametrech, které mají být v rámci experimentu měřeny.

Hillenbrand [60] ve svém experimentu měřil HNR, JIT a SHI na přesně definovaných signálech. Z výsledků vyplývá, že se jednotlivé sledované parametry navzájem ovlivňují: přidáváním šumu do perfektně periodického signálu nejen že klesá HNR, ale roste JIT a SHI.

Michaelis et al. [97] při uvedení parametru GNE pracoval pouze se syntetickými signály. Až jeho kolegové v následujících pracích použili reálné signály od pacientů.

Murphy, který se intenzivně zabývá měřením šumové příměsi v hlase, často používá vygenerované signály: Murphy [103] pro odhad HNR ve frekvenční oblasti a Murphy and Akande [105], Murphy and Olatunji [106] pro odhad HNR v kepst-rální oblasti. Podobně jako Manfredi [83], který rovněž měří míru šumové příměsi v hlase pomocí parametru NNE.

2.2.2 Komerční databáze

Jako komerční databáze je označována taková databáze, kterou je možné volně zakoupit a použít k výzkumným nebo výukovým účelům. Vznikají v přísně kontrolovaných podmínkách a mají zaručenou výbornou kvalitu zvukových záznamů.

Databáze Kay Elemetrics

Databáze Kay Elemetrics je asi nejznámějším představitelem komerční databáze nahrávek. Vznikla ve spolupráci firmy *Kay Elemetrics* a *Massachusetts Eye and Ear Infirmary* [117] a obsahuje více než 1400 nahrávek od více než 700 lidí. Proces jejího nahrávání je přísně kontrolován (odhlučněná místnost, přesná pozice mluvčího a mikrofonu apod.). Každý nahrávaný fonoval hlásku /a/ alespoň tři vteřiny a dále četl foneticky vybalancovaný text označovaný jako „The rainbow passage.“ Z prodloužené fonace byla vybrána pouze ustálená část, čtený text je uveřejněn celý. K databázi je k dispozici seznam nahrávek a seznam akustických parametrů spočtených

pomocí software firmy *Kay Elemetrics*. Nahrávky nejsou subjektivně hodnoceny odborníky.

Autoři si pro své experimenty vybírají různé podmnožiny patologických hlasů. Databáze je originálně pořízena s $f_s=25$ nebo 50 kHz, ale autoři si data podvzorkovávají. Tab. 2.1 shrnuje studie, ve kterých byla databáze použita. Z tabulky je patrné, že autoři si z patologických nahrávek vybírají pouze ty, které se vztahují k jejich výzkumu.

Tabulka 2.1: Přehled studií využívajících databázi Kay Elemetrics.

studie	Z	P	poznámka
Costa et al. [25, 26, 27]	50	43	pouze otoky hlasivek, /a/ 1 až 3 sekundy
Fraile et al. [41, 40], Godino-Llorente et al. [47, 48, 49], Sáenz-Lechón et al. [130]	53	173	prodloužené /a/
Lee et al. [78]	50	600	prodloužené /a/, ale pořídili nahrávky 547 zdravých korejců, které měli ohodnocené odborníky, zde $f_s=16$ kHz
Markaki and Stylianou [87]	0	88	různé anamnézy
Neto et al. [110]	53	77	44 otoků na hlasivkách, 23 jiných patologií
Peng et al. [116]	39	177	prodloužené /a/
Shama et al. [131]	53	163	podvzorkovali na $f_s=16$ kHz, různé patologie, v článku seznam; prodloužené /a/
Tavares et al. [137]	53	99	otoky hlasivek a parézy zvratného nervu
Lowell et al. [79]	14	28	spojitá řeč
Hillenbrand and Houde [61]	5	50	/a/ a spojitá řeč
Qi et al. [120]	0	87	spojitá řeč

Databáze Japan Society of Logopedics and Phoniatics

V roce 1981 vydala Japan Society of Logopedics and Phoniatics, původce subjektivní stupnice GRBAS, databázi nahrávek 65 patologických hlasů. Společně s touto databází je k dispozici subjektivní hodnocení nahrávek výše zmíněnou metodu GRBAS.

Ve své práci ji použil Lee and Hahn [77], který k ní přidal 20 nahrávek zdravých korejských hlasů.

2.2.3 Lokální databáze zahraničních studií

Jako „lokální databáze zahraničních studií“ jsou označeny databáze promluv pacientů, které nejsou veřejně k dispozici a jsou pořízeny buď přímo autory příslušného

článku a jejich výzkumnými týmy, nebo se jedná o databázi nahrávek místní nemocnice.

Práce de Oliveira Rosa et al. [31] používá vlastní databázi obsahující 48 patologických a 25 zdravých hlasů. Nahrávali na počítač, použité parametry: $f_s=22050$ Hz, 16 bitů na vzorek. Nahrávali prodloužené /a/, /e/, /i/.

Martin et al. [88] používá vlastní databázi 70 patologických (9 mužů, 61 žen) a 20 zdravých hlasů (10 mužů, 10 žen). Analyzovali prodlouženou fonaci hlásky /a/, ve které analyzovali f_0 a s ním spojené parametry (jitter, shimmer). Autoři kladli zvláštní důraz na to, aby žádný z nahrávaných nebyl nikdy v minulosti hlasově trénovaný (zpěvák, herec, ...).

Omori et al. [112] používá vlastní databázi 389 patologických hlasů pro analýzu subharmonických složek v chraptivé řeči u prodloužené hlásky /a/. Po analýze nahrávek ovšem zjistil, že se subharmonické složky objevily pouze u 20 nahrávek pacientů.

Werth et al. [146] použil vlastní databázi promluv se 1700 patologickými a 580 zdravými hlasy. Databáze byla pořizována 9 let na místní klinice jako součást léčby pacientů. Databáze je hodnocena stupnicí GRBAS od tamějších odborníků. V článku je uveden seznam diagnóz a počet pacientů s danou diagnózou. Vzhledem k délce sběru dat byla databáze pořizována různými přístroji – to také nakonec vedlo k použití „jen“ 1673 hlasů (zdravých a patologických celkem). Hlavní důvod k redukci počtu nahrávek byl fakt, že část byla nahrávána se vzorkovací frekvencí $f_s=22050$ Hz a autoři se rozhodli, že nahrávky pořizené s vyšší vzorkovací frekvencí nebudou převzorkovávat na nižší frekvenci. Použitá vzorkovací frekvence je tedy $f_s=44100$ Hz.

2.3 Automatické hodnocení chraptivosti

Existuje několik publikací, ve kterých byla provedena klasifikace celkového dojmu z hlasu G podle stupnice GRBAS (Alpan et al. [6], Fredouille et al. [42], Lee and Hahn [77], Ma and Yiu [80]), avšak pouze jediná studie zahrnuje všechny sledované aspekty chraptivosti, tzn. i R, B, A a S (Saenz-Lechon et al. [126]).

Alpan et al. [6] testoval systém pro automatické hodnocení celkového dojmu z hlasu G. Systém založený na GVA (Generalized Variogram Analysis) a CPP (Cepstral Peak Prominence) testoval na databázi 251 promluv obsahujících prodlouženou fonaci hlásky /a/ a dvě krátké věty. Nahrávky byly subjektivně ohodnoceny pěti experty s alespoň pětiletou zkušeností se subjektivním hodnocením. Výsledek klasifikace je uveden pouze pomocí Pearsonova lineárního korelačního koeficientu a dosahuje hodnoty 0,79.

Fredouille et al. [42] navrhl systém pro automatické hodnocení chraptivosti založený na MFCC a směsi gausovských modelů (GMM – Gaussian Mixture Model). Pro experiment byla využita databáze 80 ženských hlasů, která byla ohodnocena stupnicí GRBAS třemi experty. Pro výslednou klasifikaci byl použit pouze parametr G. V práci jsou zveřejněny dva experimenty: v prvním byly rozlišovány pouze zdravé hlasy ($G=0$) od patologických hlasů ($G>0$) a ve druhém experimentu byly klasifikovány jednotlivé stupně G. První experiment dosáhl úspěšnosti klasifikace

92,5 %, zatímco druhý experiment „pouze“ 78,8%.

Lee and Hahn [77] představil automatický klasifikátor celkového dojmu z řeči G podle stupnice GRBAS, který používal statistiku vyšších řádů (HOS) aplikovanou na chybový signál lineární predikce. V experimentu byla použita databáze nahrávek japonského sdružení logopedů (Japan Society of Logopedics and Phonetics) obsahující 65 nahrávek patologických hlasů, ke které autoři přidali 20 nahrávek zdravých korejských hlasů. Z nahrávek byla použita pouze prodloužená fonace hlásky /a/. Klasifikátor založený na regresním stromu dosahoval přesnosti 93 %.

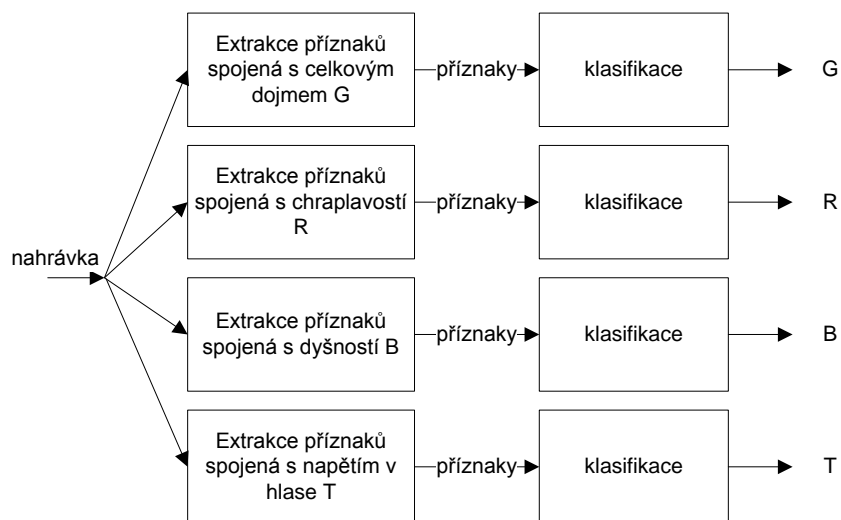
Ma and Yiu [80] použili *Multi-dimensional Voice Program* (MDVP), spirometický program *Aeroplane II* od firmy Kay Elemetrics a dále hlasové pole analyzované programem *Phong 1.0* od firmy AB Ntvalla DSP, za pomoci kterých se pokoušeli určit celkový dojem z hlasu G na nahrávkách krátkých vět. Databáze nahrávek obsahovala 112 osob, nahrávky byly subjektivně ohodnoceny čtyřmi studenty posledního ročníku medicíny (hlasové patologie). Ze širokého spektra použitých parametrů byly vybrány následující: délka fonace prodloužené hlásky /a/ (MPT – Maximum Phonation Time), špičkový vnitroústní tlak přechodu souhláska-samohláska /pi/ (PIP – Peak Intraoral Pressure), plocha hlasového pole (VRPA – Voice range Profile Area) a jitter (JIT). Výsledný klasifikátor ohodnotil správně 67,5 % nahrávek.

Saenz-Lechon et al. [126] navrhl automatický systém pro objektivní hodnocení všech parametrů ze stupnice GRBAS. Systém používal MFCC a algoritmus Learning Vector Quantization (LVQ). V experimentu bylo použito 648 nahrávek obsahujících prodlouženou fonaci hlásky /a/ a krátkou větu. Databáze nahrávek byla ohodnocena třemi experty pomocí metody GRBAS. Navržený systém dosahoval přesnosti okolo 65 % (od 55 % pro parametr A po 68 % pro parametr G).

3. Cíle práce

Na základě zjištění uvedených v předchozích dvou kapitolách lze shrnout cíle této práce do následujících bodů:

- **Implementace algoritmů pro popis akustického signálu pro objektivní hodnocení chraptivého hlasu.** Při znalosti používaných parametrizací je vhodné použít buď již implementované algoritmy, popřípadě provést vlastní implementace. Takto implementované algoritmy mohou vytvořit základ toolboxu pro automatické hodnocení chraptivosti. Implementované algoritmy jsou uvedeny v kapitole 4, ve které je rovněž provedeno porovnání jejich výstupů pro zdravý a chraptivý hlas.
- **Nalézt rozdíly v časových průbězích akustických parametrů použitých pro hodnocení chraptivosti.** Lze předpokládat, že zdravý hlas produkovaný zdravým hlasovým traktem bude vykazovat větší časovou stabilitu parametrů než hlas produkovaný poškozeným hlasovým traktem. Cílem je zavést parametry, které časový průběh, jeho trend a stabilitu popisují a experimentálně zjistit, zda se tyto průběhy liší u zdravých a chraptivých hlasů.
- **Analyzovat vhodnost implementovaných algoritmů v závislosti na subjektivním hodnocení odborníků.** Před použitím algoritmů je vhodné seznámit se s jejich schopnostmi diferenciací chraptivých hlasů. Algoritmy, které vykazují špatné výsledky je vhodné z dalších analýz vyřadit. Analýza použitých algoritmů je provedena v kapitole 6.
- **Sestavit experimentální systém pro objektivní hodnocení chraptivosti.** Zamýšlený experimentální systém pro objektivní hodnocení chraptivosti je zobrazen na obr. 3.1, ze kterého je patrné, že pro každý sledovaný aspekt chraptivosti (chraptivost, chraplavost, dyšnost, napětí v hlase) je nutné vybrat unikátní sadu parametrů a sestavit vhodný klasifikátor, jehož výstup bude co možná nejlépe korelovat se subjektivním hodnocením odborníků. Experimenty s výběrem příznaků a sestavováním vhodných klasifikátorů jsou uvedeny v kapitole 6.



Obrázek 3.1: Blokové schéma zamýšleného systému pro automatické hodnocení chraptivosti.

4. Akustické parametry pro analýzu chraptivosti

Na následujících stranách budou popsány parametry používané pro analýzu akustického signálu, které lze využít při popisu chraptivosti. Tyto parametry byly autorem práce implementovány v prostředí MATLAB ([92]) a následně využity v experimentální části práce.

4.1 Analýza autokorelační funkce

V práci Hillenbrand and Houde [61] byl zaveden parametr RPK (Pearson r at autocorrelation peak). Hodnota parametru RPK odpovídá amplitudě maxima autokorelační $ACR[k]$ v pásmu 3,3 až 16,7 ms (tzn. cca 60 až 300 Hz). Autokorelační funkce je dána vztahem

$$ACR[k] = \frac{1}{N - |k|} \sum_{n=0}^{N-1} x[n]x[n+k] \quad [-], \quad (4.1)$$

kde $ACR[k]$ je výsledná autokorelační funkce, N je délka vstupního signálu $x[n]$.

Parametr RPK je tedy určen vždy – vždy lze nalézt lokální maximum v daném rozsahu. Proto byl autorem této práce zaveden parametr SPS (Signal Periodicity Strength), který je určen podle algoritmu 1.

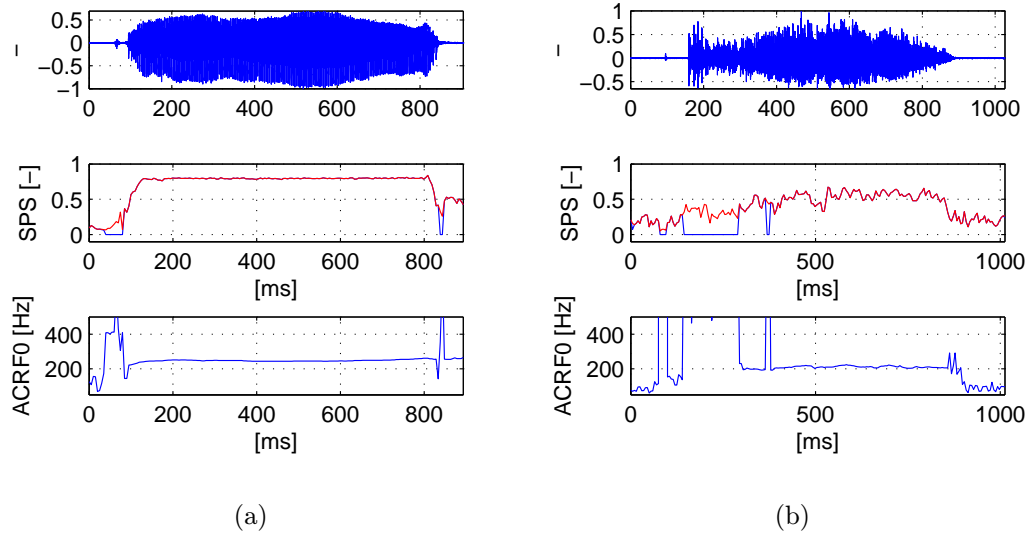
Algoritmus 1 Výpočet parametru SPS.

1. vypočítat autokorelační funkci;
 2. určit globální minimum autokorelační funkce;
 3. určit SPS jako hodnotu lokálního maxima, které je v čase dál (tzn. na nižší frekvenci) než lokální minimum;
-

Časový posun signálu, pro který autokorelační funkce nabývá globálního minima, je stav, kdy je signál v protifázi. Takový stav musí nastat, aby byl signál opět ve fázi, tzn. aby byla nalezena jeho perioda. Parametr SPS může nabývat nulových hodnot v případě, kdy nalezené maximum leží mimo hranice definované pro parametr RPK.

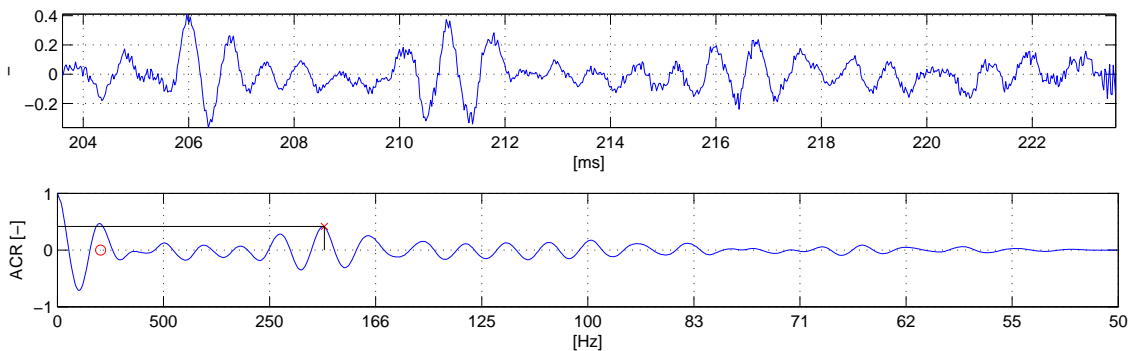
Na obr. 4.1 je zobrazen průběh parametrů RPK a SPS pro zdravý hlas (obr. 4.1(a)) a pro chraptivý hlas (obr. 4.1(b)). Z průběhů je patrné, že zdravý hlas dosahuje vyšších hodnot parametru RPK a SPS a je tedy periodičtější.

Z průběhů na obr. 4.1(b) je patrné, že dojde k situaci, kdy hodnota parametru RPK není stejná jako hodnota parametru SPS. Tento stav nastane v případě, kdy lokální maximum signálu, které určuje hodnotu SPS, leží mimo zvolené hranice 3,3 až 16,7 ms (60-300 Hz). Takový případ je zobrazen na obr. 4.2. Zde je patrné lokální maximum, které leží na nižší frekvenci než globální maximum a toto globální maximum je mimo oblast zájmu, tedy 60-300 Hz. Vrchol, který by odpovídal parametru



Obrázek 4.1: Ukázka průběhu parametrů RPK (červeně) a SPS (modře) pro (a) zdravý a pro (b) chraptivý hlas ($G_2R_2B_2T_1$). Dále je vyznačen odhad hlasivkové frekvence na základě detekce autokorelační špičky (ACRF0).

SPS leží na souřadnicích $[1250;0,47]^1$, zatímco vrcholek odpovídající parametru RPK je na souřadnicích $[200;0,42]$.

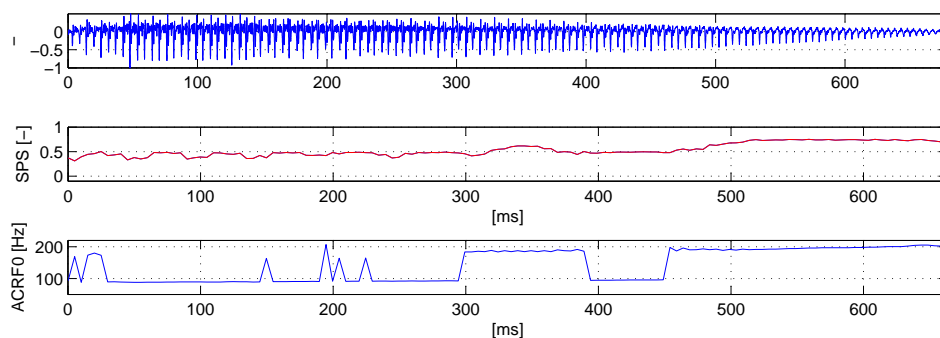


Obrázek 4.2: Ilustrace mikrosegmentu, pro který je dosažena rozdílná hodnota parametrů RPK (odpovídající maximum je vyznačeno červeným křížkem) a SPS (pozice je vyznačena červeným kolečkem).

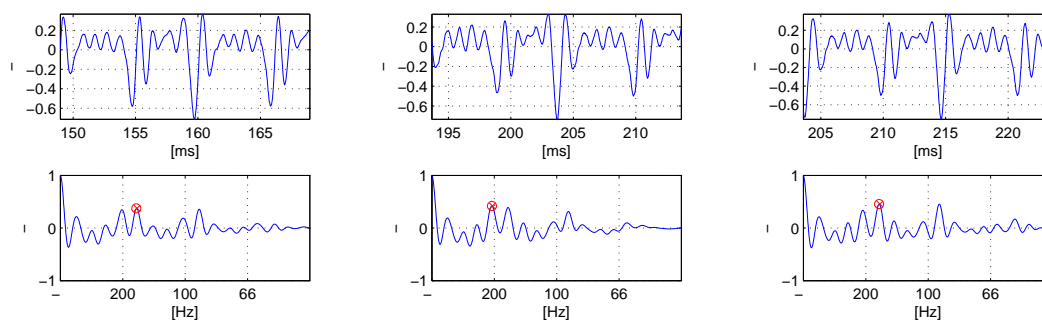
Obr. 4.3 zobrazuje jak celý signál, tak tři mikrosegmenty řečového signálu s chraptivým hlasem. Na časových průbězích zobrazených mikrosegmentů jsou dobře patrné jak amplitudové, tak frekvenční nestability hlasivkového tónu (tzn. Shimmer a Jitter). Tyto nestability mají za následek nestandardní průběh autokorelační funkce, ve které je patrné větší množství výrazných vrcholů. To se projeví nepřesnou detekcí

¹Souřadný prostor je definován v rozměrech $[\text{Hz};\text{ACR}]$, kde ACR je bezrozměrná hodnota autokorelační funkce v daném bodě.

hlasivkové frekvence, viz obr. 4.3(a) (průběh ACRF0 [Hz]) kde jsou patrné rychlé přeskoky mezi 100 a 200 Hz např. okolo 150 ms (detail viz obr. 4.3(b)) a okolo 200 ms (viz obr. 4.3(c) a 4.3(d)). Další důsledek takovýchto amplitudových a frekvenčních nestabilit je, že hodnoty parametrů SPS a RPK neodpovídají periodicitě hlasu².



(a)



(b)

(c)

(d)

Obrázek 4.3: Ukázky autokorelačních funkcí pro nahrávku hodnocenou ($G_1R_2B_1T_{-1}$): (a) původní akustický signál, průběhy parametrů SPS a RPK (jejich hodnoty jsou stejné, proto je viditelný pouze průběh parametru SPS zobrazený červenou barvou) a odhadnutá základní frekvence signálu ACRF0; (b), (c) a (d) jsou ukázky mikrosegmentů, ve kterých nestandardní časový průběh akustického signálu ovlivňuje správnou detekci základní frekvence a tím pádem i hodnoty parametrů RPK a SPS.

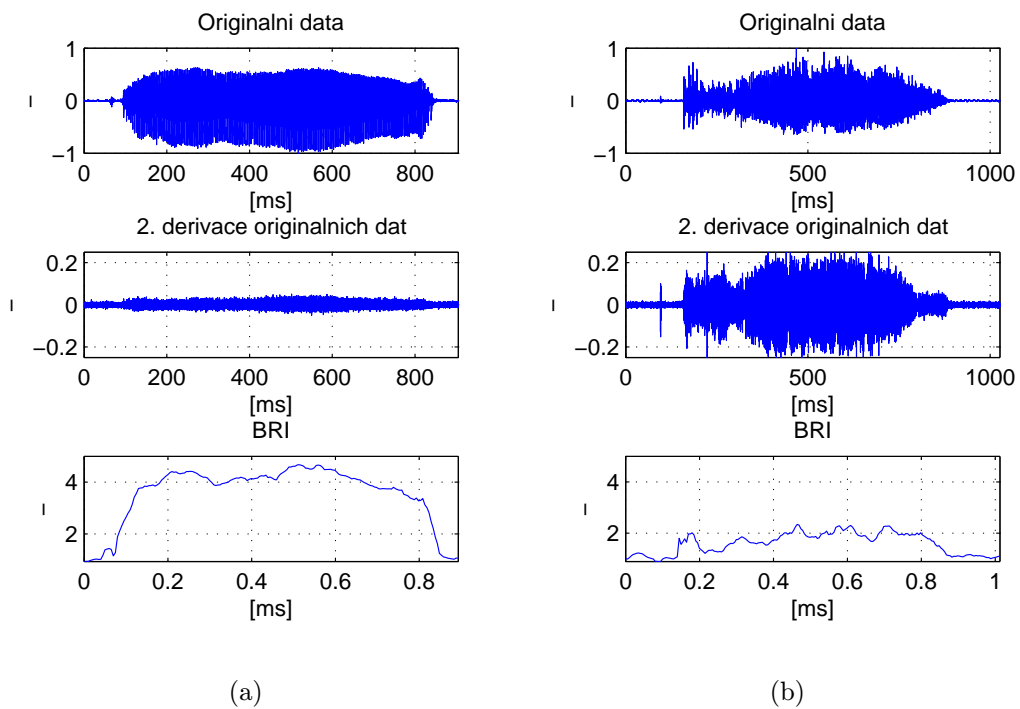
²Obecně lze říci, že čím více je hlas patologický, tím více frekvenčních a amplitudových nestabilit obsahuje. Z toho důvodu je odhad hlasivkové frekvence u hodně patologických hlasů za pomoci autokorelační funkce velmi nepřesný.

4.2 Index dyšnosti BRI

Parametr BRI (Breathiness Index) byl zaveden v práci Hillenbrand and Houde [61], kde je definován vztahem

$$\text{BRI} = \frac{\text{En}(x_{\Delta\Delta}[n])}{\text{En}(x[n])} \quad [-], \quad (4.2)$$

kde $\text{En}()$ je funkce která určí energii signálu a $x_{\Delta\Delta}[n]$ je druhá derivace časového průběhu signálu $x[n]$.



Obrázek 4.4: Ukázka průběhu parametru BRI pro 4.4(a) zdravý a pro 4.4(b) chraptivý hlas ($G_2R_2B_2T_1$).

Parametr postihuje fakt, že chraptivé hlasy mají více energie na vyšších frekvencích. První i druhou derivaci signálu lze chápat jako filtr typu horní propust a jak je patrné z obr. 4.4 (průběhy druhých derivací, druhý řádek), chraptivý hlas obsahuje více rychlých změn a proto jeho druhá derivace dosahuje vyšších hodnot a tím pádem i větší energie.

V literatuře nebyly nalezeny jiné odkazy na tento parametr. Jeho výhoda spočívá ve značné rychlosti výpočtu a dále ve faktu, že k jeho výpočtu není nutné znát f_0 .

4.3 Výraznost keprální špičky CPP

Parametr CPP (Cepstral Peak Prominence) byl zaveden v práci Hillenbrand and Houde [61] a je často používán jak pro analýzu prodloužené fonace, tak pro analýzu

spojité řeči (Alpan et al. [7], Awan et al. [12], Heman-Ackah et al. [58], Moers et al. [99], Murphy [104], Wolfe and Martin [147]).

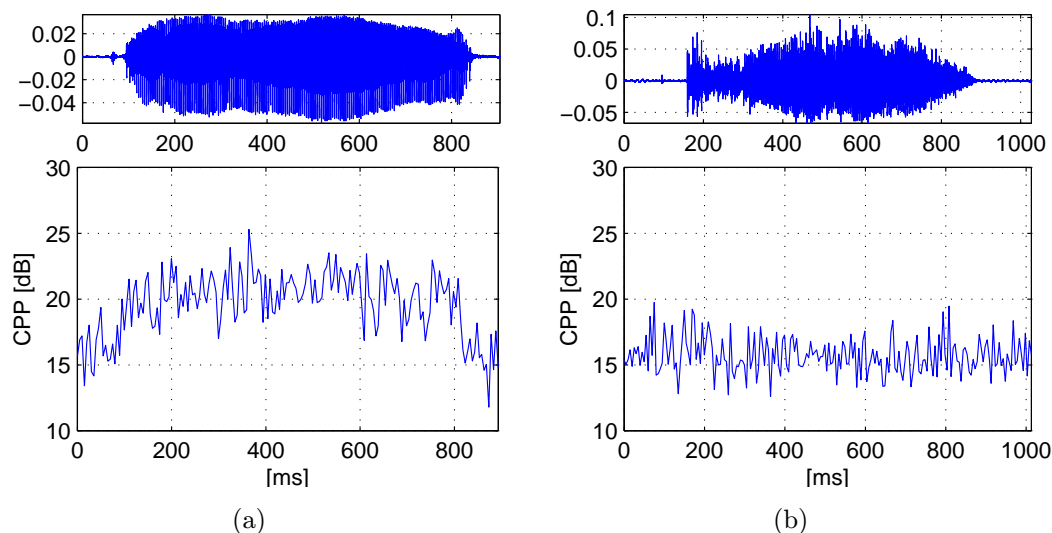
Kepstrum signálu je definováno jako inverzní Fourierova transformace logaritmu absolutní hodnoty spektra signálu

$$\tilde{x}[n] = \text{IDFT}(\log |\text{DFT}(x[n])|) \quad [-], \quad (4.3)$$

kde $\tilde{x}[n]$ je kepstrum signálu $x[n]$ a DFT a IDFT je dopředná a zpětná diskretní Fourierova transformace.

Prvních několik kepstrálních koeficientů popisuje přenosovou funkci hlasového traktu, zbylé kepstrální koeficienty jsou obrazem pravidelného buzení hlasového traktu periodickými pulzy hlasivek (Psutka et al. [119]).

Parametr CPP je definován jako vzdálenost kepstrální špičky (maxima v pásmu 3,3-16,7 ms, tzn. 60-300 Hz) od regresní přímky, která aproximuje kepstrum³. CPP využívá skutečnosti, že zdravý hlas s dobře definovanými hlasivkovými pulzy má v kepstru dobře definované špičky, chraptivý hlas má tyto špičky menší a tudíž méně vzdálené od regresní přímky prokládající kepstrum. To je ilustrováno na obr. 4.5, ze kterého je patrné, že zdravý hlas (obr. 4.5(a)) dosahuje vyšších hodnot CPP než chraptivý hlas (obr. 4.5(b)). Na obr. 4.6 je pak zobrazen typický průběh kepstra s vyznačenou kepstrální špičkou (červené kolečko) pro jeden mikrosegment zdravého hlasu (obr. 4.6(a)) a chraptivého hlasu (4.6(b)).

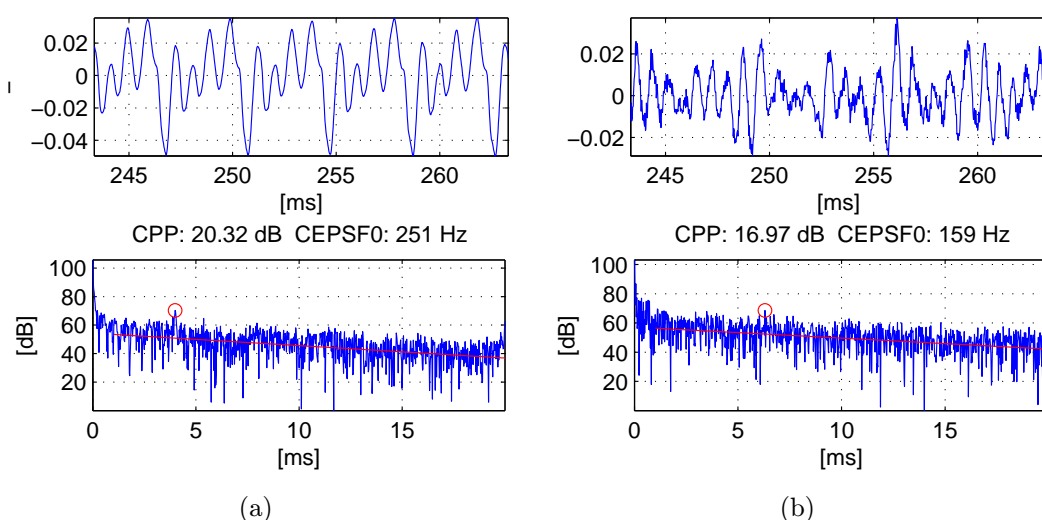


Obrázek 4.5: Ukázka průběhů CPP pro (a) zdravý hlas a (b) chraptivý hlas ($G_2R_2B_2T_1$) pro prodlouženou fonaci hlásky /a/.

Výhoda výpočtu parametru CPP spočívá ve faktu, že do jeho výpočtu přímo nevstupuje odhad hlasivkové frekvence signálu. Naopak, při jeho výpočtu je odhad hlasivkové frekvence proveden, viz dále v části 4.13.2. Chybný odhad není na

³Nutnost měření vzdálenosti mezi špičkou a regresní přímkou je rozebrána v práci Herman-Ackah [59]

škodu – předpokládá se, že špatný odhad hlasivkové frekvence u chraptivého hlasu je způsoben nedostatečně harmonickým buzením (Hillenbrand and Houde [61]). Navíc tento špatný odhad typicky nedosáhne tak vysoké hodnoty CPP, jako by dosáhl u zdravého hlasu.



Obrázek 4.6: Ukázka kepru s regresní přímkou a vyznačené maximum v kepru, ze kterého je odečteno CPP pro jeden mikrosegment signálu pro (a) zdravý hlas a (b) chraptivý hlas ($G_2R_2B_2T_1$) pro prodlouženou fonaci hlásky /a/.

4.4 Poměr hlasivkového a šumového signálu GNE

Michaelis et al. [97] představil parametr Glottal-to-Noise Excitation Ratio (GNE), jehož pomocí lze určit, do jaké míry je analyzovaný hlasový signál vybuzený hlasivkami. Testuje ho pouze na syntetických datech.

Fröhlich et al. [43] používá GNE v tzv. Hoarseness diagramu, což je zakreslování výsledků do prostoru [GNE; podobnost period⁴]. Používá databázi 1547 nahrávek, z nichž vybral 4 typické případy (dyšný nebo chraplavý hlas), na kterých ukazuje, kam do diagramu patří.

Fröhlich et al. [44] porovnává GNE s parametry popisujícími dyšnost: H1H2 (rozdíl energie harmonických), F1F3 (rozdíl energie formantů), NNE a podíly energií v různých pásmech, viz [44]. Používá databázi 454 nahrávek, která obsahuje nahrávky pacientů s nádory. Hodnocení (mechanismus poruchy) provedl expert na základě invazivního vyšetření pomocí stroboskopie.

Michaelis et al. [98] navazuje na Michaelis et al. [97] a zkouší GNE na reálných datech. Opět porovnává s NNE, CHNR, MWC (mean waveform matching coefficient), PPF (nebo EPF – period (energy) perturbation factor), PPQ (nebo EQP – period

⁴Podobnost period je určována na základě hodnot jitteru, shimmeru a korelace dvou následujících hlasivkových cyklů.

(energy) perturbation quotient). Používá databázi 535 nahrávek (447 patologických, 88 zdravých).

Fröhlich et al. [45] navazuje na Fröhlich et al. [43]. Používá databázi 425 nahrávek pořízenou na Göttingenské univerzitě. Databáze obsahuje 93 zdravých a 60 nahrávek nahraných šeptem (simulace afonických hlasů). Nahrávky obsahují samohlásky a standardizovaný text (ten ale nebyl použit). Hodnocení provedl expert na základě stroboskopického vyšetření.

Sáenz-Lechón et al. [130] zkoumá vhodnost a robustnost parametru při klasifikaci zdravý vs. patologický hlas na databázi Kay Pentax.

Godino-Llorente et al. [49] ukazuje použití parametru, testuje nastavení na databázi Kay Pentax a pokouší se zavést referenční hodnoty parametru. Ve své práci používá jen rozdělení zdravý vs. patologický hlas na části databáze Kay Pentax.

Výpočet GNE lze popsat algoritmem 2 (Michaelis et al. [97] a Godino-Llorente et al. [49]). Pro názornost odkazují příslušné kroky algoritmu 2 obr. 4.7 a obr. 4.8 – jednotlivé řádky v těchto obrázcích jsou kroky algoritmu (řádek a) až c) v daných obrázcích).

Algoritmus 2 Výpočet parametru GNE.

1. podvzorkovat vstupní signál na 10 kHz – při takto nízké vzorkovací frekvenci lépe vyniknou „Diracovy pulzy“ na místě glotálních pulzů v inverzně filtrovaném signálu; řečový signál v pásmu nad 5 kHz neobsahuje informace důležité pro výpočet GNE, Michaelis et al. [97]; viz řádek a)
 2. inverzní filtrace signálu pomocí LPC 13. řádu v okně délky 30 ms s posunem 10 ms; váhování Hannovým oknem; viz řádek b)
 3. pásmová filtrace inverzního signálu se zvolenou šířkou pásma a posunem centrální frekvence ve spektru; viz řádek c)
 4. výpočet Hilbertovy obálky pásmově filtrovaného signálu; viz řádek c), sloupec Hilbertova obálka
 5. z Hilbertových obálek pásmově filtrovaných signálů, jejichž centrální frekvence jsou od sebe dál než je polovina použité šířky pásma, jsou spočteny kroskorelační funkce
 6. výběr maximální korelace z vypočtených kroskorelací
-

Inverzní filtrace provedená v bodě 2. způsobí zploštění spektra signálu (viz obr. 4.7 a obr. 4.8, řádek b), sloupec frekvenční oblast). Tím pádem jsou si jednotlivé Hilbertovy obálky po pásmové filtraci podobnější.

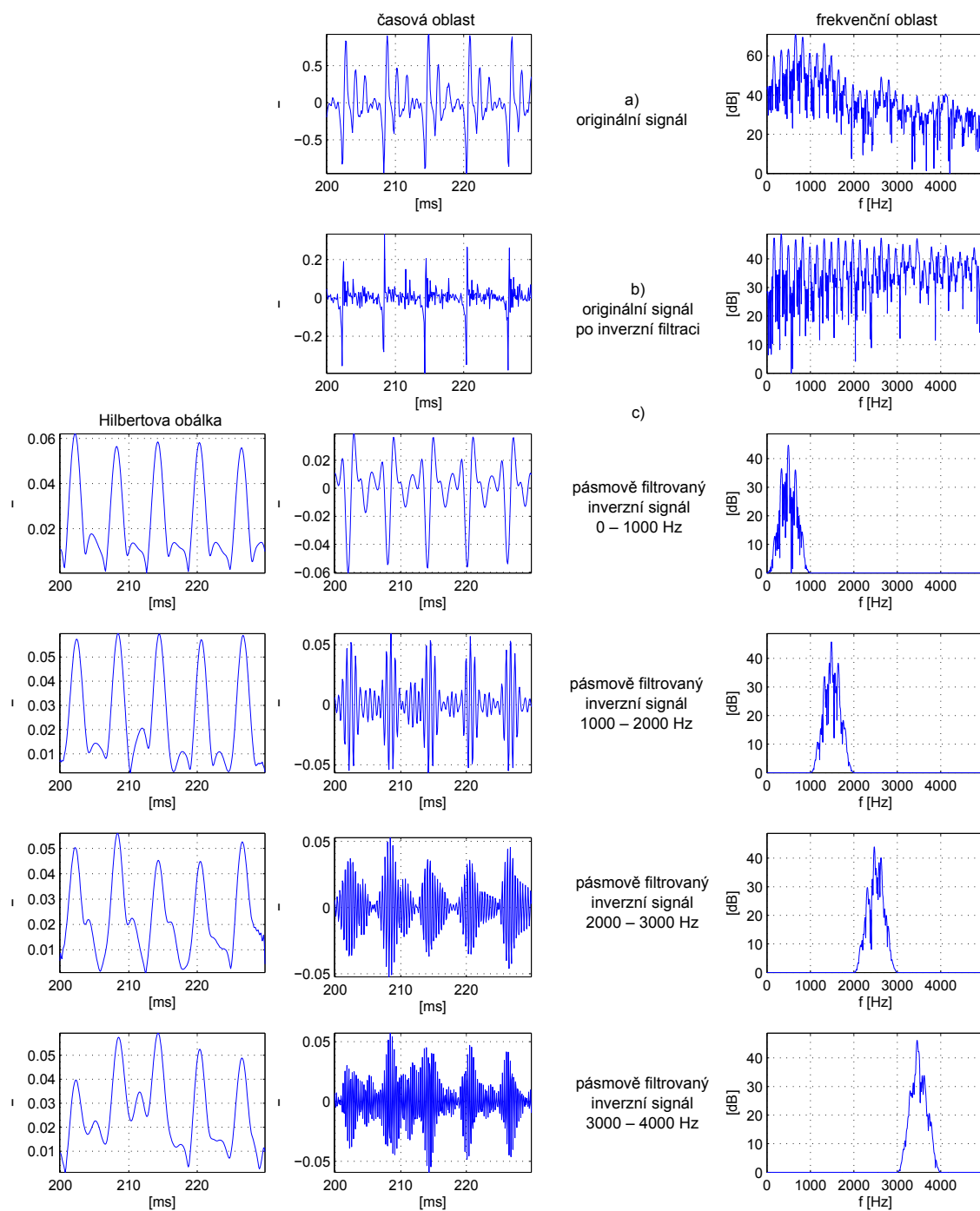
V bodě 5. výše popsaného postupu je kroskorelační funkce spočtena pro signály posunuté maximálně o $\pm 0,3$ ms (při použité f_s 10 kHz odpovídá ± 3 vzorky). Důvodem je fakt, že Hilbertovy obálky v jednotlivých frekvenčních pásmech mohou být fázově posunuté. Tento posun může být způsoben tím, že k vybuzení všech frekvenčních pásem nedojde v přesně stejném okamžiku a tím pádem inverzně filtrovaný signál není dokonalá delta funkce.

Godino-Llorente et al. [49] nepoužívá přímo bezrozměrné hodnoty GNE, které vyjadřují maximální hodnoty kroskorelací (tzn. rozmezí 0 až 1), ale logaritmickou

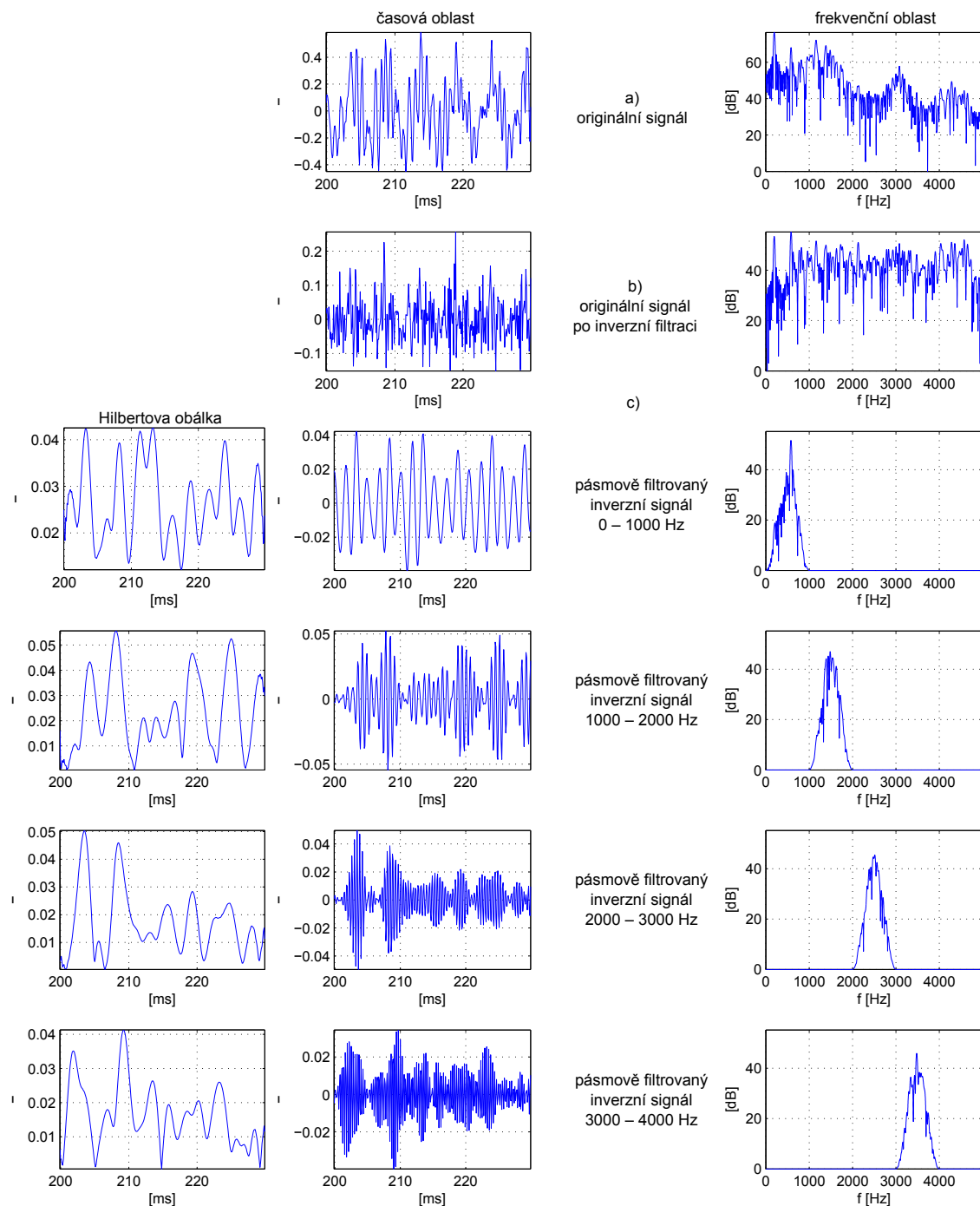
4.4. Poměr hlasivkového a šumového signálu GNE

stupnici podle vztahu $GNE_1 = 10 \log_{10}(1 - GNE)$. Takto upravené hodnoty GNE mají na databázi Kay Pentax normální rozdělení vhodnější pro testování hypotéz.

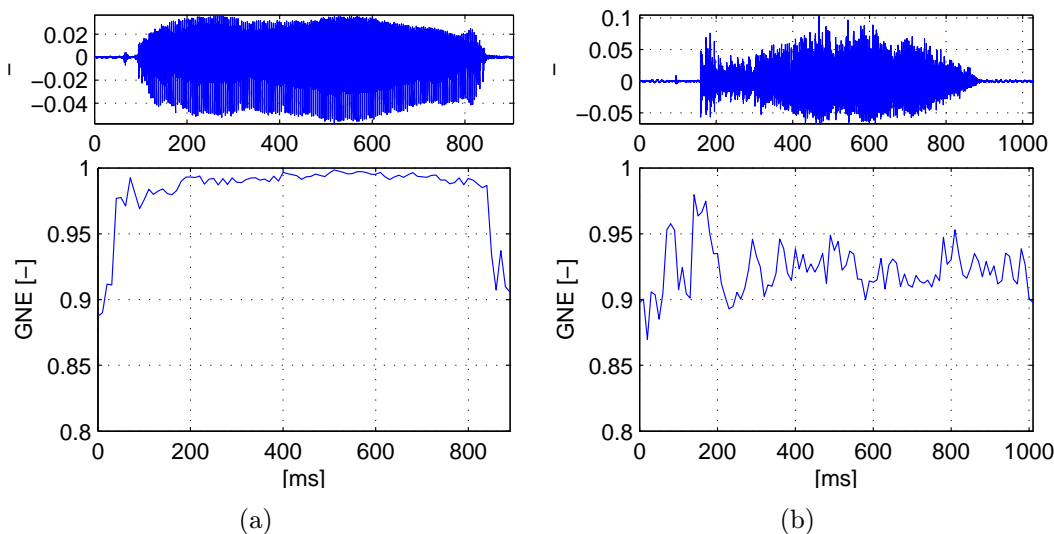
Na obr. 4.9 jsou zobrazeny průběhy parametru GNE pro zdravý (obr. 4.9(a)) a pro patologický (obr. 4.9(b)) hlas. Z obrázků je dobře patrné, že zdravý hlas dosahuje vyšších hodnot.



Obrázek 4.7: Ukázka výpočtu GNE pro zdravý hlas pro prodlouženou fonaci hlásky /a/.



Obrázek 4.8: Ukázka výpočtu GNE pro chraptivý hlas pro prodlouženou fonaci hlásky /a/.



Obrázek 4.9: Ukázka průběhů GNE pro (a) zdravý hlas a (b) chraptivý hlas ($G_2R_2B_2T_1$) pro prodlouženou fonaci hlásky /a/.

Výhoda parametru GNE spočívá v jeho absolutní nezávislosti na odhadu hlasivkové frekvence. V případě analýzy neznělého úseku je pouze dosahováno výrazně nižších hodnot parametru.

4.5 Analýza generalizovaného variogramu GVA

Analýzu hlasu pomocí generalizovaného variogramu (Generalized variogram analysis) popsal Kacha et al. [71] a dále ho použil Alpan et al. [6]. Obecně je variogram využíván spíše při analýze geologických dat (otřesy země apod., viz Kacha et al. [71]).

Použití tohoto parametru vychází z faktu, že periodický signál s periodou T_0 lze popsat vztahem $x(t) = x(t + kT_0)$, $k \in \mathbb{Z}$. Krátký úsek znělé části řečového signálu lze považovat za periodický signál a odchylku od periodicity takového signálu lze popsat vztahem

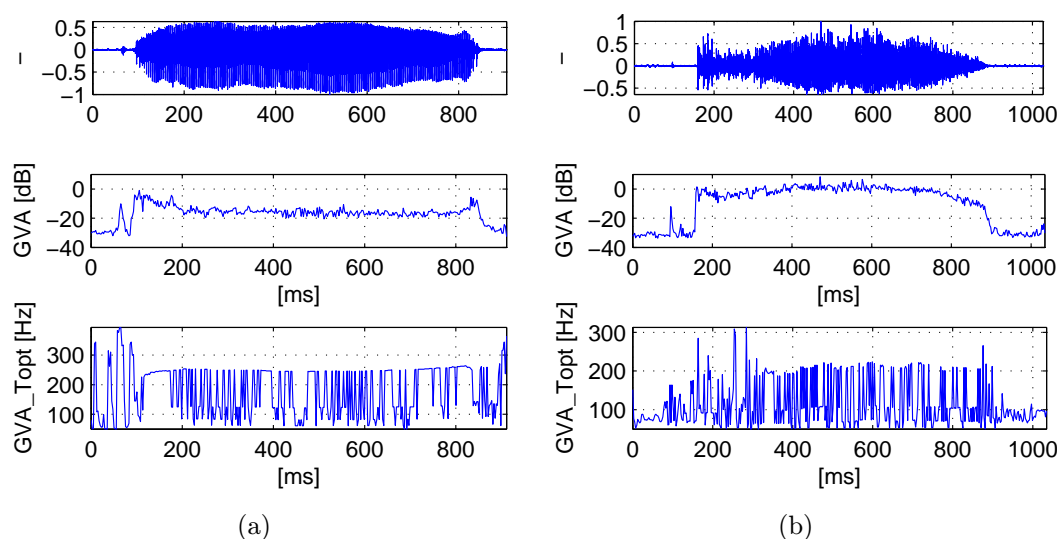
$$\text{GVA} = \min_T \left\{ \sum_{t=0}^M (x(t) - x(t+T))^2 \right\} \quad [-], \quad (4.4)$$

kde $-T_{\max} < T < -T_{\min}$, resp. $T_{\min} > T > T_{\max}$, t je čas uvnitř analyzovaného mikrosegmentu dlouhého M ms. Jedná se o sumu čtverců rozdílů mezi analyzovaným úsekem signálu a okolním signálem. Podle Alpan et al. [6] je T v rozmezí $\pm 2,5$ až ± 20 ms (tzn. mezi 50 až 400 Hz).

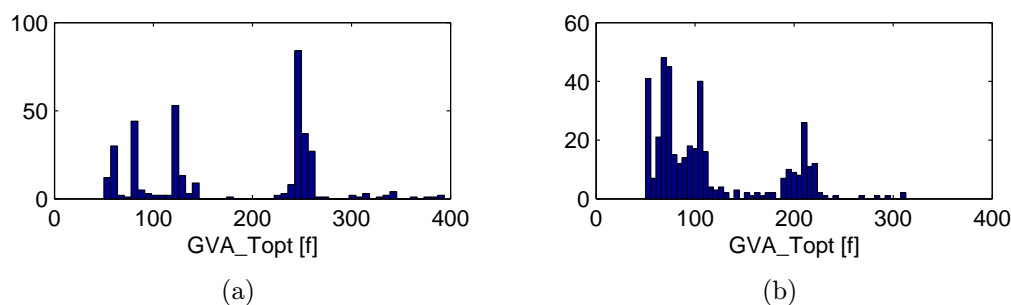
Při výše zmíněném rozmezí hodnot parametru T je použito okénko délky $M=5$ ms ($-2,5$ ms až $+2,5$ ms) a toto okénko je porovnáváno s jiným okénkem délky 5 ms v rozsahu 2,5 ms až 20 ms na obě strany (tzn. -20 až $-2,5$ ms a $2,5$ až 20 ms) a je hledán takový posun, pro který je suma čtverců rozdílů podle vztahu (4.4) mezi okénky nejmenší. Míra podobnosti je vyjádřena právě parametrem GVA. Optimální

posun, který určuje délku periody signálu je pak dán nalezenou hodnotou T , která je pro pozdější analýzu označována jako GVA_Topt .

Na obr. 4.10 jsou zobrazeny průběhy signálů GVA a GVA_Topt pro zdravý (obr. 4.10(a)) a pro chraptivý (obr. 4.10(b)) hlas. Dále jsou na obr. 4.11 zobrazeny histogramy parametru GVA_Topt , ze kterých lze odečíst, že pro zdravý hlas se hodnoty GVA_Topt více blíží hlasivkové frekvenci a jejím násobkům (obr. 4.11(a)), zatímco u chraptivého hlasu je distribuce GVA_Topt více rozptýlená (obr. 4.11(b)).



Obrázek 4.10: Ukázka průběhů GVA a GVA_Topt pro (a) zdravý hlas a (b) chraptivý hlas ($G_2R_2B_2T_1$) pro prodlouženou fonaci hlásky /a/. Hodnoty GVA byly pro vykreslení kvůli zvýšení přehlednosti zlogaritmovány, proto jsou uváděny s jednotkou [dB].



Obrázek 4.11: Histogramy GVA_Topt pro (a) zdravý hlas a (b) chraptivý hlas ($G_2R_2B_2T_1$) pro prodlouženou fonaci hlásky /a/.

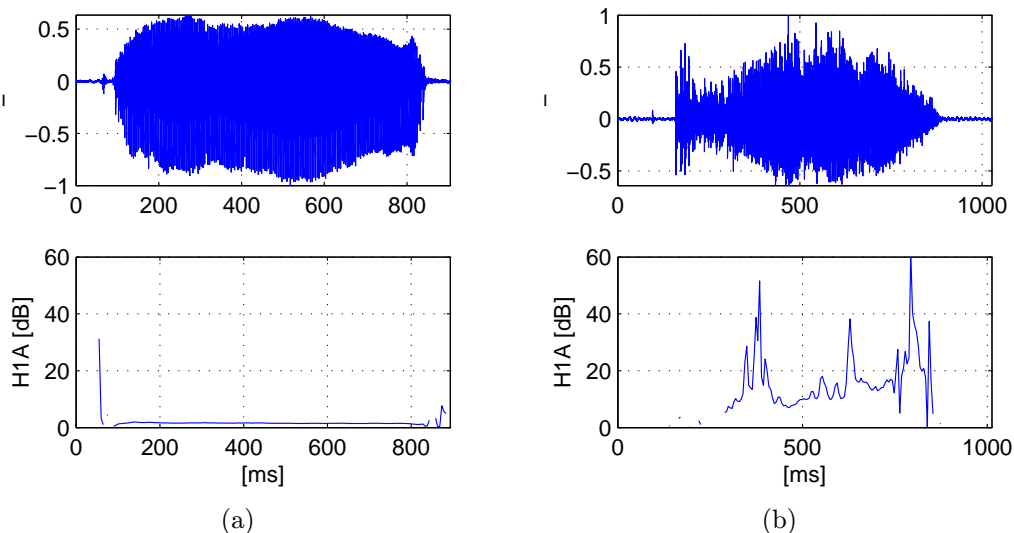
Pro výpočet parametru GVA není nutný odhad hlasivkové frekvence. Naopak na základě parametru GVA_Topt lze hlasivkovou frekvenci, byť velmi nepřesně, určit.

4.6 Poměr amplitud harmonických složek H1A

Parametr byl popsán v práci Hillenbrand and Houde [61], kde je jako důvod jeho zavedení popsán fakt, že časový průběh znělých částí dyšného hlasu je více podobný průběhu funkce sinus. Tento fakt se ve frekvenční oblasti projeví silnější 1. harmonickou signálu oproti zdravému hlasu.

Pro tento algoritmus je stěžejní automatická detekce harmonických částí signálu. Používaná metoda bude popsána dále v popisu algoritmu pro výpočet HNR, část 4.8.

Na obr. 4.12 je ukázka průběhu parametru H1A pro zdravý hlas (obr. 4.12(a)) a pro chraptivý hlas (obr. 4.12(b)). V místech, kde nebylo možné harmonické určit, není průběh parametru vykreslen.

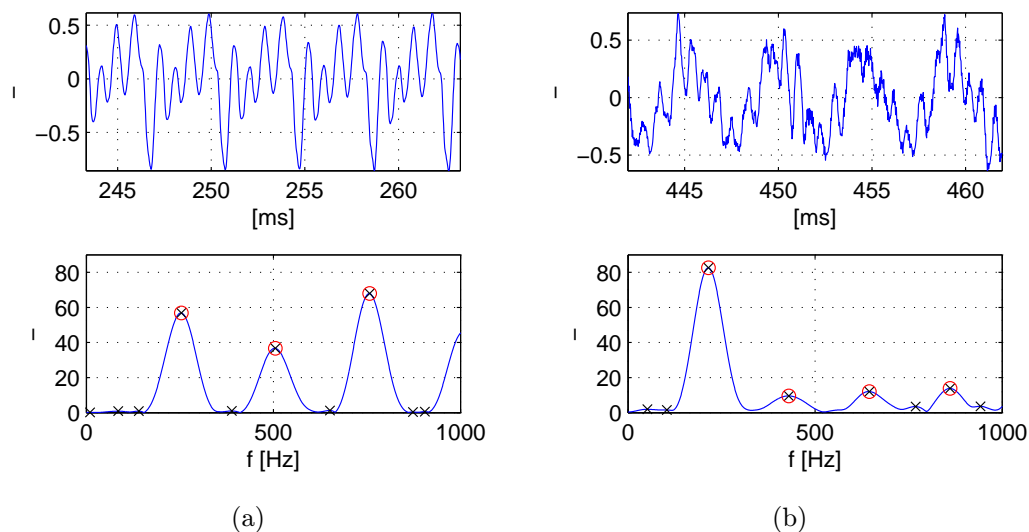


Obrázek 4.12: Ukázka průběhu parametru H1A pro (a) zdravý hlas a (b) chraptivý hlas ($G_2R_2B_2T_1$) pro prodlouženou fonaci hlásky /a/.

Z uvedených obrázků je na první pohled patrné, že zdravý hlas má poměr amplitud 1. a 2. harmonické vyrovnaný a tím pádem parametr H1A dosahuje značně nižších hodnot. Za povšimnutí stojí také mnohem vyrovnanější časový průběh parametru pro zdravý hlas.

Na obr. 4.13 jsou pak ukázky spekter s vyznačenými harmonickými. Zde je na první pohled patrný rozdíl amplitud 1. a 2. harmonické pro zdravý hlas (obr. 4.13(a)) a pro chraptivý hlas (obr. 4.13(b)).

Značnou nevýhodou parametru H1A je jeho závislost na odhadu harmonických ve spektru. Pokud tento krok selže, parametr H1A není určen správně.



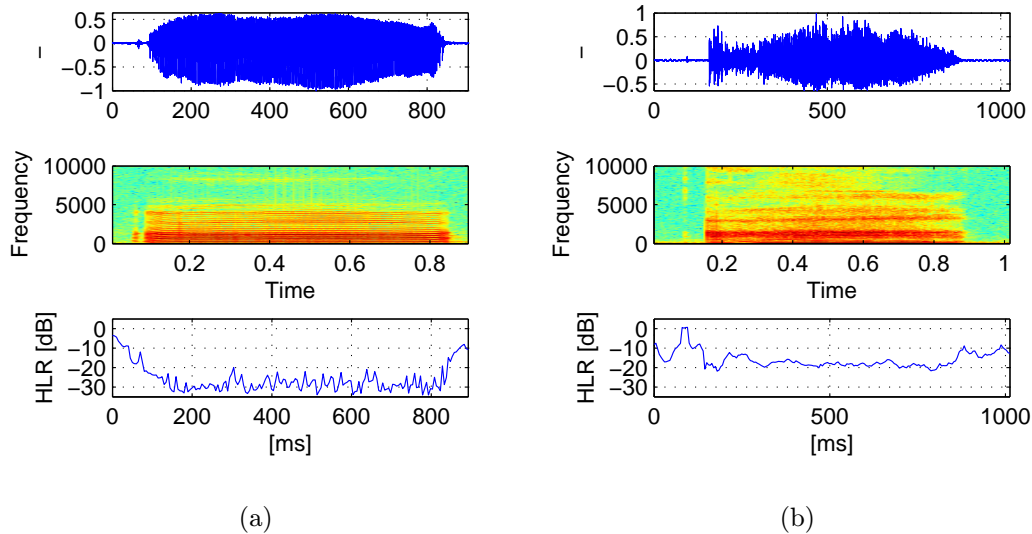
Obrázek 4.13: Ukázka detekce harmonických v jednom mikrosegmentu signálu pro (a) zdravý hlas a (b) chraptivý hlas ($G_2R_2B_2T_1$) pro prodlouženou fonaci hlásky /a/. Uvažované harmonické jsou červeně zakroužkované vrcholy.

4.7 Poměr energií ve frekvenčních pásmech HLR

Parametr HLR je popsán v práci Hillenbrand and Houde [61]. Jedná se o poměr energie ve spektru ve frekvenčním pásmu 0 až 4 kHz a 4 až $f_s/2$ kHz. Myšlenka u tohoto parametru je stejná jako u parametru BRI popsaného v části 4.2: chraptivé hlasy mají vyšší zastoupení energie na vyšších frekvencích. Na rozdíl od parametru BRI, který je počítán v časové oblasti, je parametr HLR počítán ve spektrální oblasti.

Na obr. 4.14 je ukázka průběhu parametru HLR pro zdravý hlas (obr. 4.14(a)) a pro chraptivý hlas (obr. 4.14(b)). Ze spektrogramů je patrné, že zdravý hlas obsahuje více energie na nižších frekvencích a tato energie je soustředěná hlavně do jednotlivých harmonických. Oproti tomu chraptivý hlas má ve spektru více šumový charakter – energie je více rozprostřena.

Podobně jako u parametru BRI není pro výpočet nutná znalost hlasivkové frekvence.



Obrázek 4.14: Ukázka průběhu parametru HLR pro (a) zdravý hlas a (b) chraptivý hlas ($G_2R_2B_2T_1$) pro prodlouženou fonaci hlásky /a/.

4.8 Poměr energie harmonické a šumové složky HNR

Míra energie v harmonických oblastech HNR (Harmonics-to-Noise Ratio) je jedním z velmi starých popisů chraptivosti. Vychází z popisu hlasu vztahem

$$x[n] = s[n] + w[n], \quad (4.5)$$

kde $x[n]$ je výsledný signál, $s[n]$ je periodická část signálu tvořená pravidelným kmitáním hlasivek a $w[n]$ je aperiodická (šumová) část signálu, která vzniká turbulentním prouděním vzduchu z plic. Zatímco ve zdravém hlase je dominantní periodická část $s[n]$, v patologickém hlase je kvůli špatnému uzavěru hlasivkové štěrbině periodická část $s[n]$ zastoupena méně na úkor šumové části $w[n]$.

Pokud dáme do poměru energie těchto částí, získáme parametr HNR podle vztahu

$$\text{HNR} = 10 \log \frac{\text{En}(s[n])}{\text{En}(w[n])} \quad [\text{dB}], \quad (4.6)$$

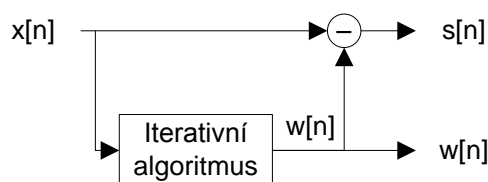
kde $\text{En}(s[n])$ je energie periodické části signálu a $\text{En}(w[n])$ je energie šumové části signálu.

Obecně existuje několik postupů, jak odhadnout energii periodické a šumové části signálu. Odhad lze provést v časové oblasti, jak navrhuje např. Yumoto et al. [152], který identifikuje jednotlivé hlasivkové cykly v časové oblasti, poté určí průměrný hlasivkový cyklus a na jeho základě určuje odchylky jednotlivých cyklů, z nichž určí míru šumové příměsi. Tento postup vyžaduje přesnou detekci hlasivkových pulzů a proto je nevhodný pro automatické použití, hlavně u těžce patologických hlasů.

Další možností je odhad HNR v kepstrální oblasti (tzv. CHNR, Fröhlich et al. [43], Michaelis et al. [98], Murphy [103], Murphy [104], Murphy and Akande [105], Murphy and Olatunji [106]). Tento způsob využívá nulování oblasti kepstra, která obsahuje informace o buzení a poté převedení kepstra zpět do frekvenční oblasti. Po tomto převodu je ve spektru obsažena pouze tzv. *baseline*, která určuje šumové pozadí signálu. Na základě znalosti šumového pozadí signálu je pak možné určit výsledné HNR. Neexistuje ovšem jednoznačná shoda, jakou část kepstra nulovat – není totiž zcela jasné, kolik koeficientů kepstra zodpovídá za buzení, více viz Murphy and Akande [105].

Manfredi [83] představuje možnost odhadu energie šumové části za pomoci vlnkové transformace.

V této práci je použit algoritmus navržený v práci Yegnanarayana et al. [150], který pracuje ve frekvenční oblasti a výstupem je odhad šumové složky signálu $w[n]$ v časové oblasti, viz obr. 4.15. Detailní postup algoritmu je shrnut v následujících odstavcích.



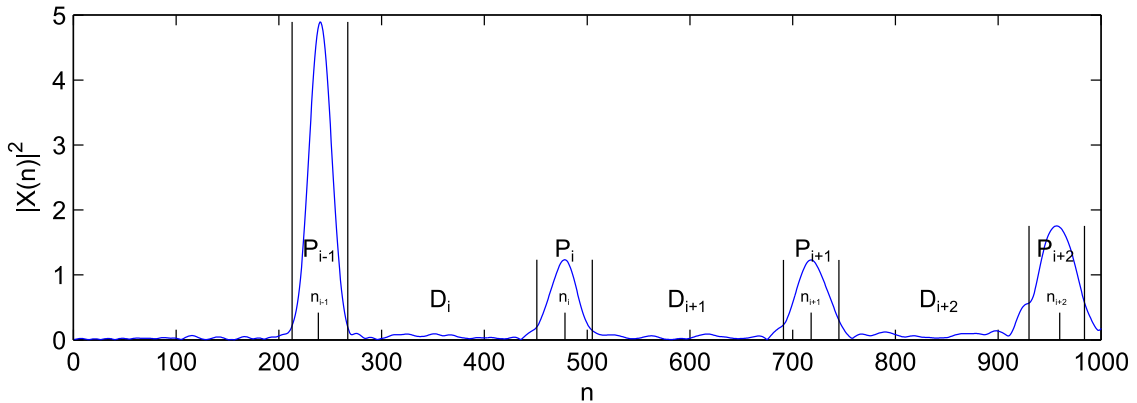
Obrázek 4.15: Základní princip algoritmu pro výpočet HNR. Blokové schéma iterativního algoritmu je zobrazeno na obr. 4.17.

Vstupní signál je segmentován na úseky délky M vzorků (mikrosegmenty jsou váhovány Hammingovým oknem), poté je na každý mikrosegment aplikována N bodová DFT ($N > M$). V amplitudovém spektru jsou následně nalezeny dva typy regionů, viz obr. 4.16:

- P_i – harmonická část spektra, která obsahuje jak energii periodického buzení (hlasivkové pulzy), tak přídavný šum; šířka regionů odpovídá délce DFT (N) a délce analyzovaného okna (M): $2N/M$
- D_i – mezera mezi harmonickými částmi spektra, která obsahuje pouze energii šumu; pokud mají být úseky D_i dlouhé alespoň d vzorků, musí být splněna podmínka

$$M \geq \frac{4N}{f_0NT - (d + 1)}, \quad (4.7)$$

kde M je délka analyzovaného Hammingova okna, N je délka použité DFT, f_0 je odhadnutá hlasivková frekvence v analyzovaném mikrosegmentu, T je vzorkovací perioda ($T = 1/f_s$) a d je požadovaná délka regionů D_i .



Obrázek 4.16: Regiony zájmu ve spektru pro výpočet HNR: oblasti obsahující harmonické buzení a šumovou oblast P_i a oblasti obsahující pouze šum D_i .

Oblasti P_i a D_i jsou určeny následovně:

$$P_i = \left\{ n \mid n_i - \frac{2N}{M} \leq n \leq n_i + \frac{2N}{M} \right\}, \quad (4.8)$$

$$D_i = \left\{ n \mid n_{i-1} + \frac{2N}{M} \leq n \leq n_i - \frac{2N}{M} \right\}, \quad (4.9)$$

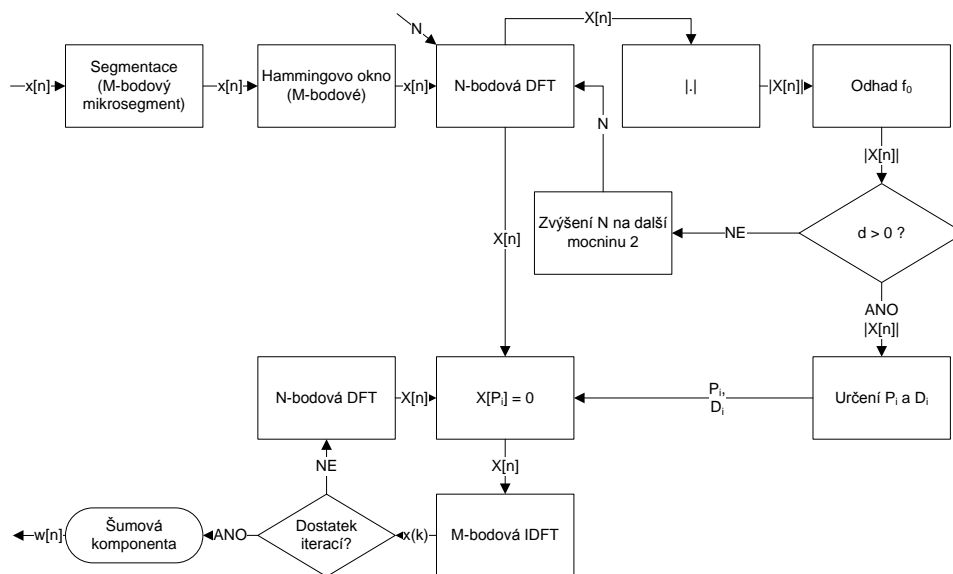
kde n je pořadí spektrální čáry a n_i je pozice i -té harmonické oblasti. Lokalizace harmonických regionů probíhá podle algoritmu 3.

Algoritmus 3 Určení harmonických oblastí spektra.

1. spočítat N bodové DFT
 2. určit lokální maxima spektra
 3. najít globální maximum spektra
 4. nastavit hranici 10 % amplitudy globálního maxima
 5. LM je lokální maximum, které je větší než zvolená mez a má nejmenší frekvenci
 6. **while** leží na poloviční frekvenci LM maximum, které má alespoň 40% amplitudy tohoto maxima **do**
 LM je toto menší maximum
 - end while**
 7. LM je pozice 1. harmonické signálu
-

Po identifikaci regionů P_i a D_i iterativní algoritmus opakovaně nuluje harmonické regiony P_i , spočítá M bodovou IDFT⁵ a opět N bodovou DFT, vynuluje oblasti P_i atd. dokud není splněn počet iterací, viz obr. 4.17. Yegnanarayana et al. [150] uvádí 10 jako dostatečný počet iterací algoritmu. Výstupem je odhad šumové komponenty řečového signálu v časové oblasti, viz obr. 4.15.

⁵Tedy kratší, dopředně je počítána N bodová DFT, důvod je uveden v Yegnanarayana et al. [150].



Obrázek 4.17: Blokové schéma iterativního algoritmu pro výpočet HNR.

Z výše uvedeného postupu vyplývá, že pro jisté kombinace M , N a f_0 může nastat situace, kdy se harmonické oblasti P_i překrývají, tzn. $d \leq 0$. Potom je nutné změnit délku N použité DFT. Po úpravě vztahu (4.7) lze psát

$$N \geq \frac{M(d+1)}{Mf_0T-4}, \quad (4.10)$$

která není definována pro

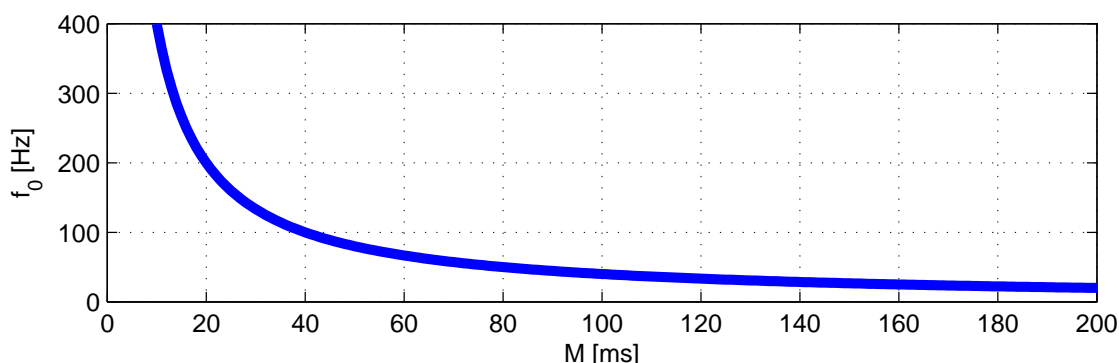
$$f_0 = \frac{4}{M_{\text{vzorky}}T} = \frac{4000}{M_{\text{ms}}}. \quad (4.11)$$

Obr. 4.18 zobrazuje závislost „kritické“ hlasivkové frekvence f_0 na délce analyzovaného mikrosegmentu v ms. Z grafu lze odečíst, že např. pro analýzu mikrosegmentu, ve kterém byla hlasivková frekvence odhadnuta na 100 Hz, je nutná délka mikrosegmentu alespoň 40 ms (lépe více).

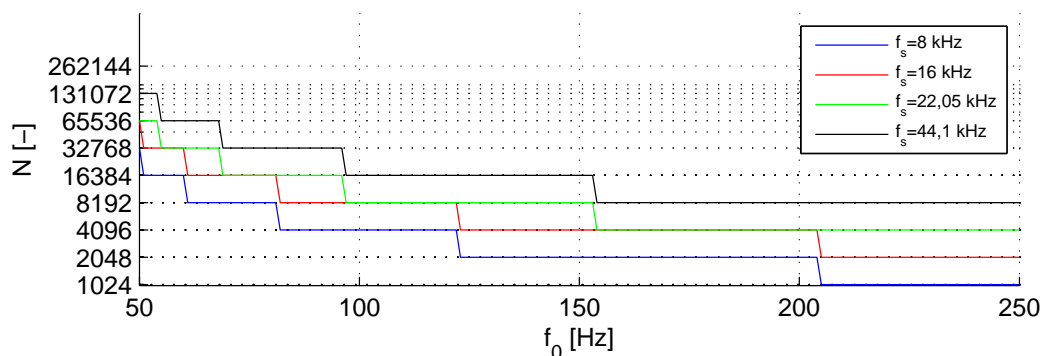
Nutná minimální délka DFT v závislosti na odhadnuté hlasivkové frekvenci je zobrazena na obr. 4.19. Na uvedeném obrázku jsou zobrazeny nejbližší vyšší mocniny dvou minimální délky N a to především z důvodu použití algoritmu FFT, který má nejmenší výpočetní náročnost právě pro signály, jejichž délka je mocnina dvou.

Na obr. 4.20 jsou zobrazeny ukázky průběhů HNR pro zdravý hlas (obr. 4.20(a)) a pro chraptivý hlas (obr. 4.20(b)). Z průběhů je patrné, že zdravý hlas dosahuje vyšších hodnot HNR než hlas chraptivý. Na obrázku je také zobrazen odhad hlasivkové frekvence SPECF0, který odpovídá 1. harmonické signálu, více v části 4.13.3.

Na obr. 4.21 jsou pak zobrazeny rozklady vstupního signálu na harmonickou část $s[n]$ a šumovou část $w[n]$. Z obrázku je patrné, že zdravý hlas (obr. 4.21(a)) má zanedbatelnou šumovou složku, zatímco chraptivý hlas (obr. 4.21(b)) ne.



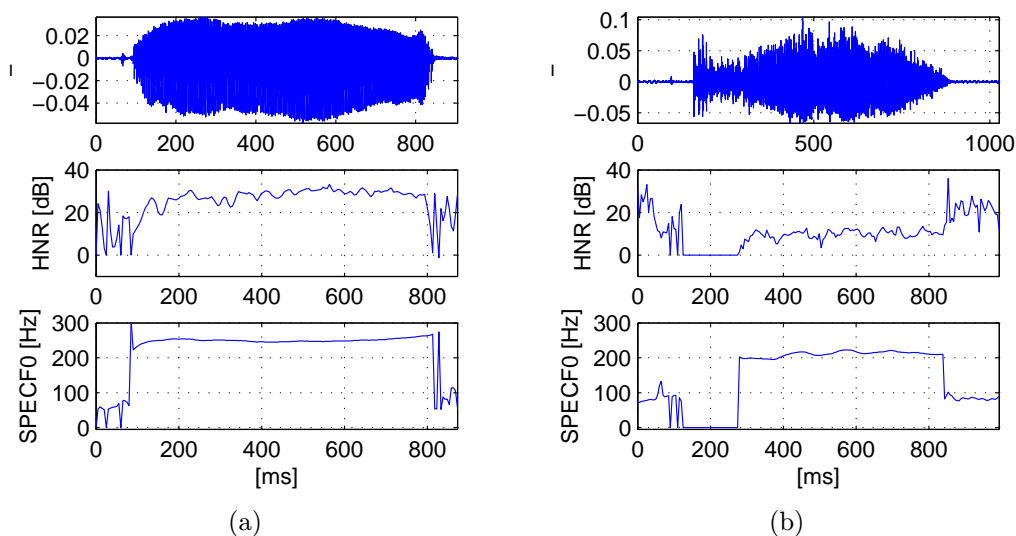
Obrázek 4.18: Kritické hodnoty hlasivkové frekvence v závislosti na délce mikrosegmentu.



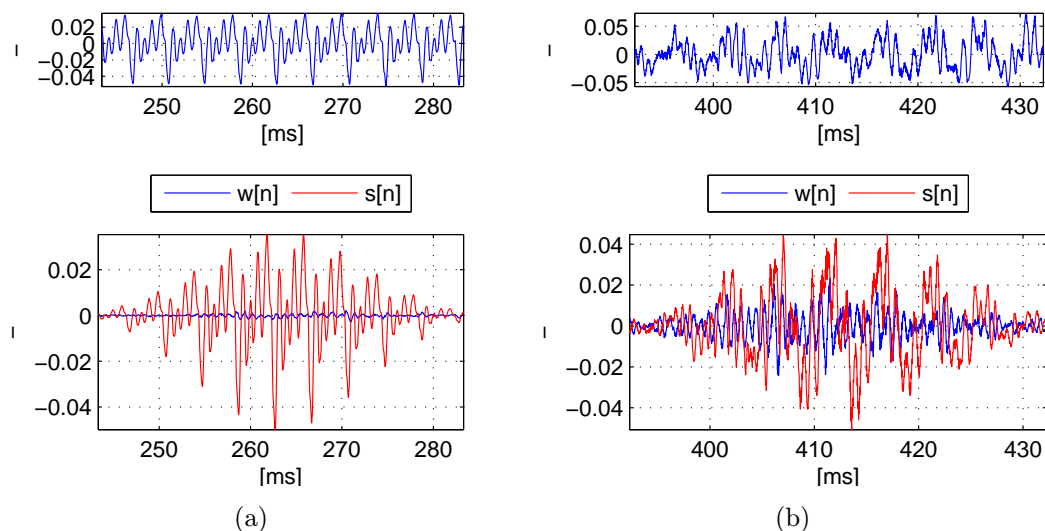
Obrázek 4.19: Minimální délka DFT pro výpočet HNR v závislosti na odhadnuté hlasivkové frekvenci.

4.8.1 Porovnání hodnot s Praatem

Použitý implementovaný algoritmus byl porovnán s výstupem z programu Praat (Boersma and Weenink [21]), který je široce používaný a odbornou komunitou uznávaný. K porovnání byla použita databáze Kay Pentax, která je popsána v části 2.2.2. Program používá algoritmus založený na analýze normalizované kroskorelační funkce, Boersma [19]. Porovnání je zobrazeno na obr. 4.22. Z obrázku je patrné, že je dosahováno dobré shody, Pearsonův lineární korelační koeficient dosahuje hodnoty 0,86. Dále je možné z obrázku usoudit, že pro méně patologické hlasy je dosahováno vyšší shody, se snižující se HNR jsou body rozptýlenější. Tyto rozdíly jsou způsobeny přítomností subharmonických složek ve spektru, která výrazně ovlivňují hodnoty HNR počítané použitým algoritmem. Subharmonické totiž leží v regionech, které mají obsahovat pouze šumovou energii, nejsou při výpočtu nulovány a zvyšují tím energii šumové složky.



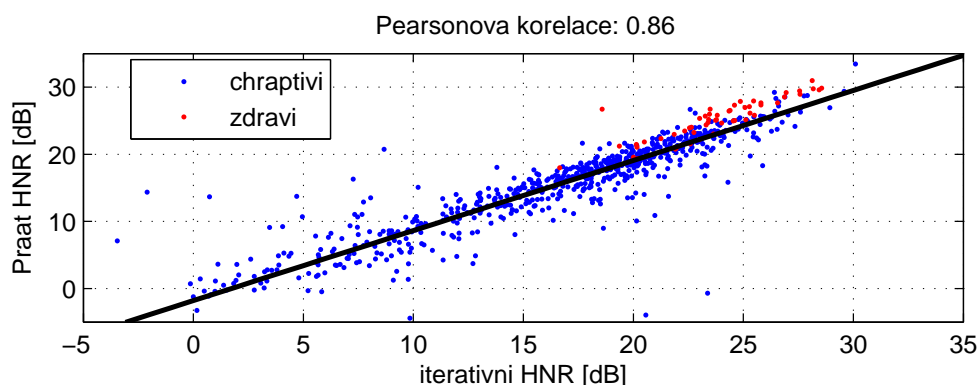
Obrázek 4.20: Ukázka průběhu parametru HNR pro (a) zdravý hlas a (b) chraptivý hlas ($G_2R_2B_2T_1$) pro prodlouženou fonaci hlásky /a/.



Obrázek 4.21: Ukázka odhadu šumové a periodické části signálu pro jeden mikrosegment signálu pro (a) zdravý hlas a (b) chraptivý hlas ($G_2R_2B_2T_1$) pro prodlouženou fonaci hlásky /a/.

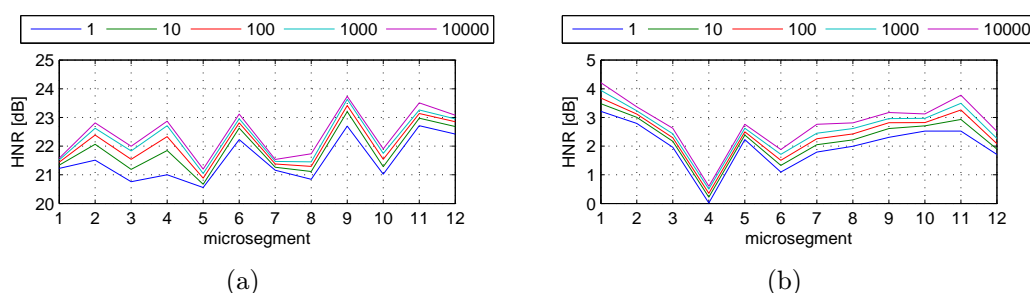
4.8.2 Vliv počtu iterací na výslednou hodnotu HNR

Na obr. 4.23 je ukázán vliv počtu iterací na výslednou odhadnutou hodnotu HNR. Z obrázku je patrné, že jak pro zdravý hlas (obr. 4.23(a)), tak pro chraptivý hlas (obr. 4.23(b)) je deset iterací dostatečný počet. Podle předpokladu platí, že čím více iterací je provedeno, tím je dosaženo vyšší hodnoty HNR. To je způsobeno častějším nulováním harmonických regionů a tím pádem poklesem energie v od-



Obrázek 4.22: Porovnání hodnot HNR mezi implementovaným algoritmem a Praatem.

hadnuté šumové části signálu. S rostoucím počtem iterací ovšem roste výpočetní náročnost algoritmu (viz tab. 4.1) – je nutné provádět víc převodů mezi časovou a spektrální oblastí. Z tab. 4.1 je také patrné, že pro deset iterací je algoritmus schopný pracovat v reálném čase.⁶



Obrázek 4.23: Ukázka vlivu počtu iterací na výslednou hodnotu HNR pro (a) zdravý hlas a (b) patologický hlas z databáze Kay Pentax.

Tabulka 4.1: Čas potřebný k výpočtu HNR pro různé počty iterací algoritmu. Délka testovacího signálu je 1 s.

počet iterací	1	10	100	1 000	10 000
doba odhadu [s]	0,36	0,65	5,32	51,86	518,73

⁶Doba výpočtu je kratší než délka signálu.

4.9 Spektrální páry LSP

Line Spectral Pairs (LSP) nebo Line Spectral Frequencies (LSF) je metoda parametrizace signálu používaná k jednoduchému popisu spektrálních vlastností signálu na základě lineární prediktivní analýzy. Hlavní výhodou použití tohoto způsobu parametrizace spočívá ve vyšší odolnosti vůči kvantizačnímu šumu – koeficienty LSP je možné při ukládání nebo posílání do přenosového systému kvantovat menším počtem bitů, aniž by byl výrazně zhoršen zpětně zrekonstruovaný signál. Koeficienty LSP a lineární prediktivní analýzy je možné mezi sebou poměrně jednoduše přepočítávat.

Popis hlasového signálu pomocí LSP je založen na válcovém modelu hlasového traktu – trakt je možné si představit jako sadu válců stejné délky, ale různého průměru, které jsou do sebe zasunuté. Pokud bude takovými válci proudit vzduch, bude docházet k různým rezonancím v závislosti na tom, zda bude tento model na konci otevřený nebo uzavřený (otevření a uzavření hlasivek, úst atp.), přičemž počet rezonančních frekvencí závisí na počtu válců, kterými je hlasový trakt modelován – na řádu modelu. Tyto rezonanční frekvence je možné popsat právě pomocí Line Spectral Frequencies (LSF), které přímo souvisí s LSP.

LSP je polynom, jehož kořeny jsou LSF. Pokud tyto kořeny seřadíme, dostaneme páry čísel⁷ (v tomto případě frekvencí), které popisují rezonanční frekvence hlasového traktu.

LSP je odvozeno z koeficientů LPC označovaných a_m . Pro výpočet koeficientů a vlastností LPC viz např. Psutka et al. [119].

Pokud známe LPC koeficienty a_m , můžeme zavést polynom

$$A(z) = 1 + a_1z + a_2z^2 + \dots + a_Lz^L, \quad (4.12)$$

což je LPC model L-tého řádu hlasového traktu pro daný signál.

Pokud zavedeme dva polynomy L+1 řádu $P(z)$ a $Q(z)$, které jsou antisymetrické⁸ a mají k polynomu $A(z)$ následující vztah

$$A(z) = \frac{P(z) + Q(z)}{2}, \quad (4.13)$$

získáme LSP polynomy. Pokud nalezneme kořeny těchto polynomů, získáme LSF.

Polynomy $P(z)$ a $Q(z)$ lze získat z následujících vztahů:

$$P(z) = A(z) - z^{-(L+1)}A(z^{-1}), \quad (4.14)$$

$$Q(z) = A(z) + z^{-(L+1)}A(z^{-1}). \quad (4.15)$$

Lze dokázat (McLoughlin [94] nebo Bäckström and Magi [14]), že kořeny LSP polynomů $P(z)$ a $Q(z)$ leží v z rovině na jednotkové kružnici a jsou navzájem proložené.

⁷Sudý a lichý kořen.

⁸Polynom $P(z)$ je antisymetrický k polynomu $Q(z)$, jestliže $Q(z) = -P(z)$.

Samotné frekvence ω_k lze z komplexních kořenů Θ_k polynomů $P(z)$ a $Q(z)$ vypočítat následovně

$$\omega_k = \tan^{-1} \left(\frac{\Re \Theta_k}{\Im \Theta_k} \right), \quad (4.16)$$

kde \Re je reálná a \Im je imaginární část komplexních kořenů Θ_k .

Dále je dokázáno, že vzhledem k tomu, že polynomy jsou navzájem antisymetrické, odpovídá jeden polynom modelu zavřeného hlasového traktu a druhý otevřenému.

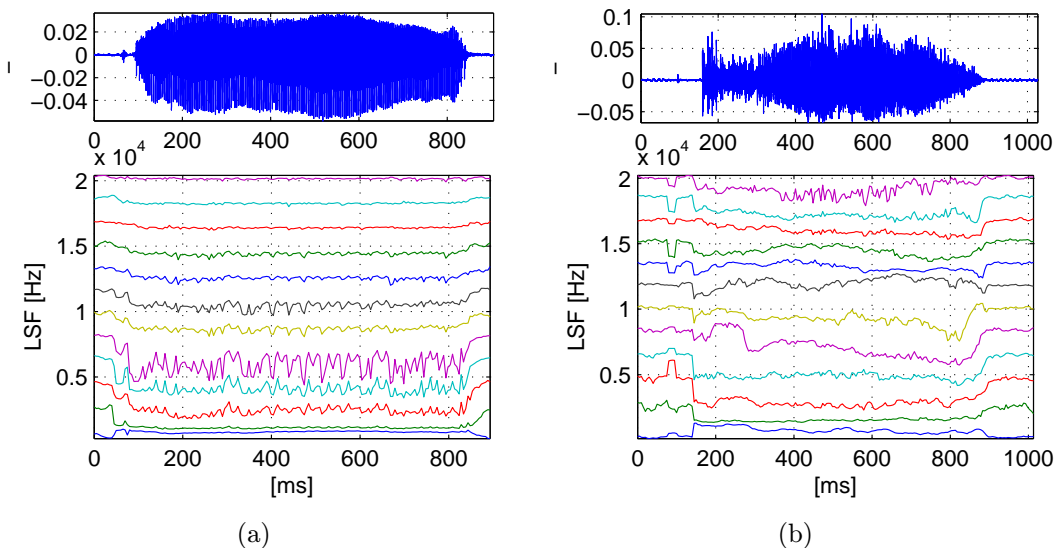
Pro matematická odvození a zpětné přepočty mezi LPC a LSP viz McLoughlin [94] nebo Bäckström and Magi [14].

Pozice LSF ukazují, kde jsou v signálu rezonanční frekvence – pokud jsou si odpovídající páry frekvenčně blízké, je mezi nimi významnější nárůst energie a naopak, čím jsou od sebe vzdálenější, tím větší lokální minimum energie mezi nimi je. Pokud v signálu nejsou obsaženy rezonanční frekvence, jsou LSF rovnoměrně rozmístěné.

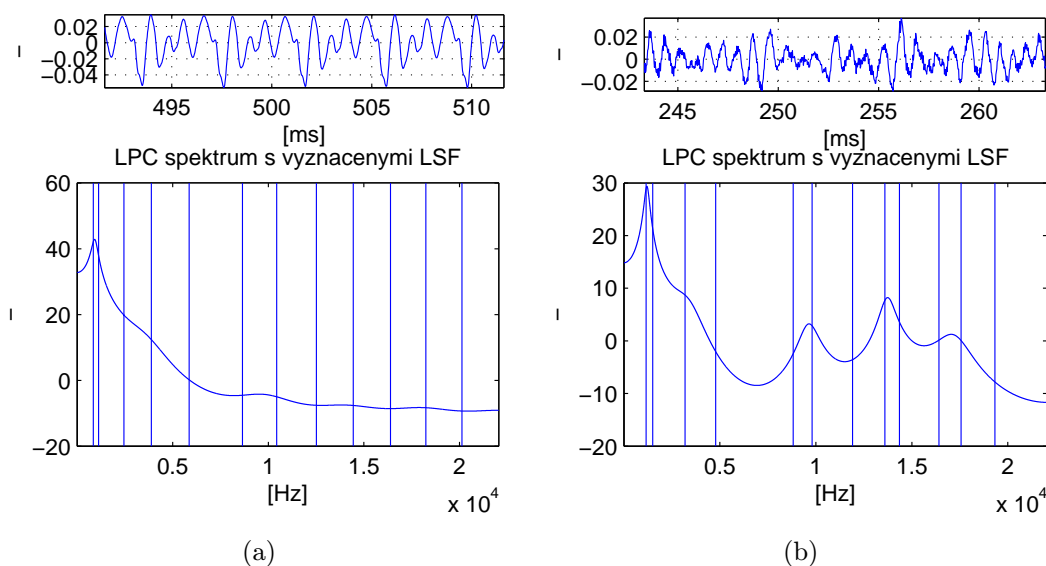
Ukázky průběhů LSF jsou zobrazeny na obr. 4.24. Z uvedených obrázků je patrné, že zatímco zdravý hlas na vyšších frekvencích neobsahuje významné množství energie (viz obr. 4.25(a)) a tím pádem jsou vyšší LSF ve spektru rovnoměrně rozmístěny, u patologického hlasu jsou i vyšší LSF rozmístěny okolo lokálních maxim energie (viz obr. 4.25(b)).

Pro výpočet LSF není nutný odhad hlasivkové frekvence signálu.

V Stráník and Čmejla [134] jsou definovány dva parametry založené na LSF, které popisují vzájemnou vzdálenost dvou korespondujících LSF a dále přibližnou frekvenci, okolo které jsou LSF rozmístěny.



Obrázek 4.24: Ukázka průběhů LSF pro (a) zdravý hlas a (b) chraptivý hlas ($G_2R_2B_2T_1$) pro prodlouženou fonaci hlásky /a/.



Obrázek 4.25: Ukázka odhadu LPC spektra s vyznačenými LSF pro jeden mikrosegment signálu pro (a) zdravý hlas a (b) chraptivý hlas ($G_2R_2B_2T_1$) pro prodlouženou fonaci hlásky /a/.

4.9.1 Diference LSF – dLSF_{x,y}

Parametr dLSF_{x,y} je dán vztahem

$$\text{dLSF}_{x,y} = \omega_k - \omega_{k-1}, \quad (4.17)$$

kde ω_k je frekvence vypočtená ze vztahu (4.16) a k je přirozené kladné sudé číslo nejvýše rovné řádu LSP polynomů a $x = k$, $y = k - 1$.

Tento parametr jednoduchým způsobem popisuje, jak významný je nárůst energie ve frekvenční oblasti mezi LSF – čím je dLSF_{x,y} menší, tím významnější je nárůst mezi x -tou a y -tou LSF a naopak, pokud je diference příliš vysoká, může se jednat o lokální minimum energie.

4.9.2 Průměr LSF – mLSF_{x,y}

Parametr mLSF_{a,b} je dán vztahem

$$\text{mLSF}_{x,y} = \frac{1}{2} (\omega_k + \omega_{k-1}), \quad (4.18)$$

kde ω je frekvence vypočtená ze vztahu (4.16) a k je přirozené kladné sudé číslo nejvýše rovné řádu LSP polynomů a kde $x = k$, $y = k - 1$.

Tento parametr zhruba odpovídá pozici lokálního maxima případně minima energie mezi LSP – o minimum se jedná, pokud je diference mezi LSP příliš velká.

4.10 Spektrální deformace SD

Parametr SD (Spectral Deformation) představil ve své práci zabývající se elektro-myografií Sinderby et al. [132]. Parametr byl dále využit v práci Frid and Lavner [46] k popisu frikativních hlásek. Parametr SD je definován podle vztahu

$$SD = \frac{\sqrt{M_2/M_0}}{M_1/M_0}, \quad (4.19)$$

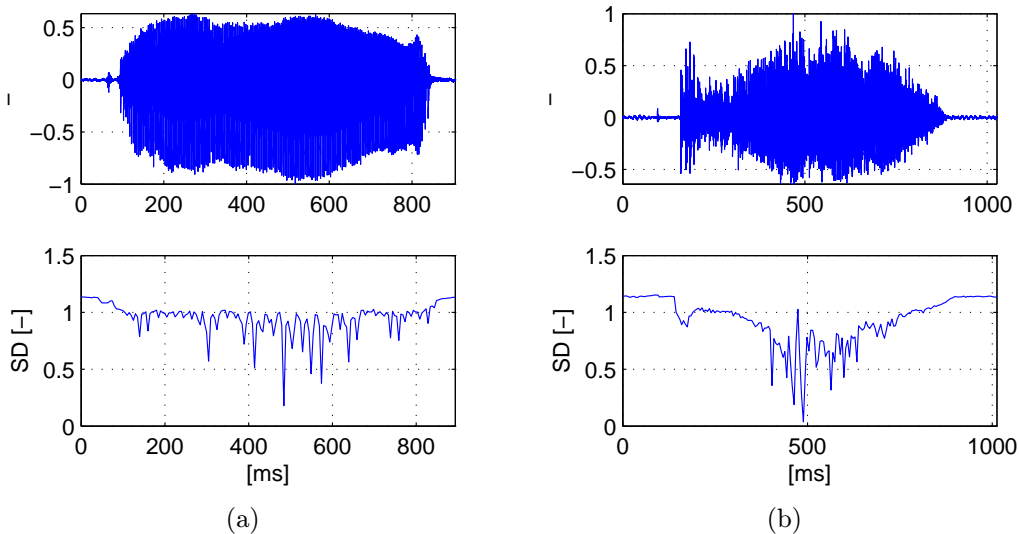
kde M_m je m -tý spektrální moment definovaný vztahem

$$M_m = \sum_n (P_n f_n^m), \quad (4.20)$$

kde P_n je spektrální výkon n -té spektrální čáry, která odpovídá frekvenci f_n

$$P_n = 20 \log(|X[n]|). \quad (4.21)$$

Ukázky průběhů SD jsou zobrazeny na obr. 4.26. Z obrázku je patrné, že pro zdravý hlas (obr. 4.26(a)) má SD vyrovnanější průběh než pro chraptivý hlas (obr. 4.26(b)).



Obrázek 4.26: Ukázka průběhů SD pro (a) zdravý hlas a (b) chraptivý hlas ($G_2R_2B_2T_1$) pro prodlouženou fonaci hlásky /a/.

4.11 Spektrální rolloff SR

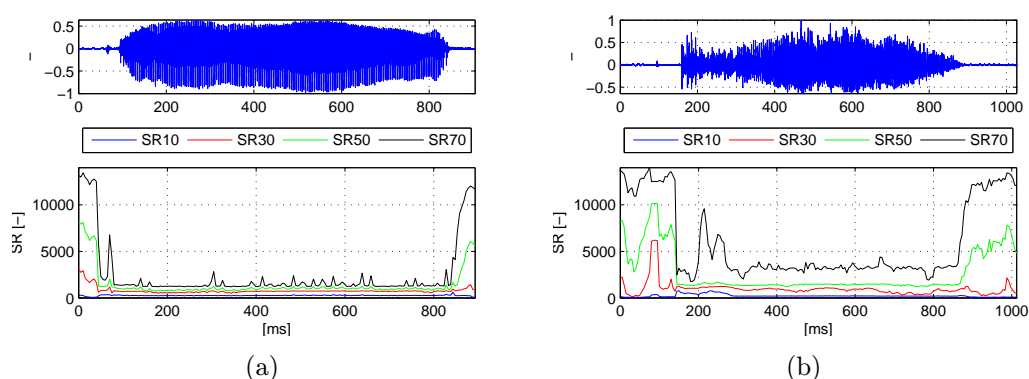
Spektrální rolloff (SR) je frekvence ve spektru signálu f_r , pod kterou je p procent energie signálu (Frid and Lavner [46]), viz následující vztah

$$\sum_{n=0}^{f_r} |X[n]| = p \sum_{n=0}^{N-1} |X[n]|. \quad (4.22)$$

V další práci je použito následující značení: SR10 znamená frekvenci, pod kterou je 10% celkové energie signálu, SR30 fekvenci, pod kterou je 30% celkové energie signálu atd.

Podle předpokladu mají chraptivé hlasy více energie na vyšších frekvencích a tím pádem budou hodnoty SR obecně vyšší v závislosti na tom, v jaké frekvenční oblasti je přídavný šum přítomný.

Ukázka průběhů parametru SR pro různá nastavení je zobrazena na obr. 4.27. Zde je patrné, že se průběhy liší především pro vyšší hodnoty p , zatímco pro zdravý hlas (obr. 4.27(a)) jsou průběhy SR10 a SR70 blízko u sebe a v průběhu fonace poměrně ustálené, u chraptivého hlasu (obr. 4.27(b)) tomu tak není.



Obrázek 4.27: Ukázka průběhů SR pro (a) zdravý hlas a (b) chraptivý hlas ($G_2R_2B_2T_1$) pro prodlouženou fonaci hlásky /a/.

4.12 Spektrální šířka SW

Parametr SW (Spectral Width) byl, podobně jako parametr SD popsany výše, představen v práci zabývající se elektromyografií Sinderby et al. [132] a dále využit v práci Frid and Lavner [46] k popisu frikativních hlásek. Parametr SW je definován podle vztahu

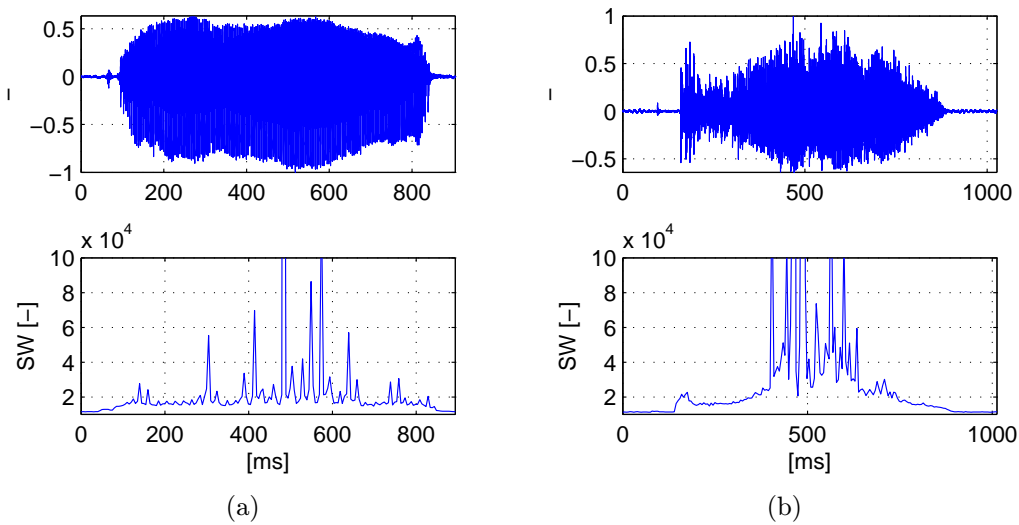
$$SW = \sqrt{\frac{M_2}{M_1} - \left(\frac{M_1}{M_0}\right)^2}, \quad (4.23)$$

kde M_m je definován ve vztahu (4.20).

Ukázky průběhů SW jsou zobrazeny na obr. 4.28. Z obrázku je patrné, že pro zdravý hlas (obr. 4.28(a)) má SW vyrovnanější průběh než pro chraptivý hlas (obr. 4.28(b)).

4.13 Odhady hlasivkové frekvence

V rámci práce nebyla snaha vyvíjet algoritmus, který by přesně určoval hlasivkovou frekvenci v řečovém signálu ať už při prodloužené fonaci nebo ve spojitě řeči.



Obrázek 4.28: Ukázka průběhů SW pro (a) zdravý hlas a (b) chraptivý hlas ($G_2R_2B_2T_1$) pro prodlouženou fonaci hlásky /a/.

V části 2.1.1 je uveden důvod: přesný automatický odhad f_0 je v patologickém hlasu velmi komplikovaný a v některých případech téměř nemožný.

Následující odhady f_0 jsou vedlejší produkty výpočtů parametrů popsaných výše.

4.13.1 ACRF0

Jedná se o odhad hlasivkového tónu v časové oblasti. Při výpočtu parametrů RPK a SPS (viz část 4.1) je nutné odečíst maximum v autokorelační funkci, které odpovídá hlasivkové frekvenci. Jako ACRF0 je označeno maximum vybrané funkcí SPS, tzn. pouze pokud lokální maximum, které následuje po globálním minimu leží v pásmu 3,3-16,7 ms (60-300 Hz). Tyto špičky jsou v obr. 4.1 a obr. 4.3 označeny červeným kolečkem.

4.13.2 CEPSF0

Jedná se o odhad hlasivkového tónu v keprální oblasti. Při výpočtu CPP (viz část 4.3) je nutné určit nejvyšší špičku v kepru. Tato špička by měla odpovídat hlasivkové frekvenci. V obr. 4.6 jsou tyto špičky vyznačeny červeným kolečkem.

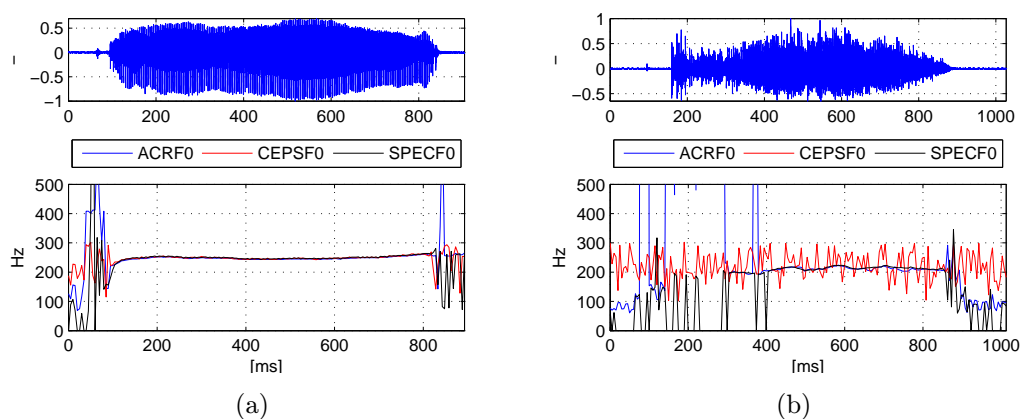
4.13.3 SPECFO

Jedná se o odhad hlasivkového tónu ve spektrální oblasti. Při výpočtu parametrů H1A a HNR je nutné určit harmonické oblasti ve spektru. Tyto oblasti by měly odpovídat hlasivkovému tónu. V obr. 4.13 jsou tyto harmonické vyznačeny červeným kolečkem⁹.

⁹Jako SPECFO je označena 1. harmonická, tzn. kolečko na nejnižší frekvenci.

4.13.4 Ukázka průběhů hlasivkové frekvence

Obr. 4.29 zobrazuje průběhy hlasivkové frekvence f_0 třemi výše zmíněnými metodami. Z uvedených průběhů je patrné, že u zdravého hlasu (obr. 4.29(a)) jsou při fonaci průběhy totožné, zatímco pro chraptivý hlas (4.29(b)) se průběhy liší. CEPSF0 se zásadně liší od zbylých dvou SPECF0 a ACRF0. V ojedinělých případech jsou rozdíly i mezi SPECF0 a ACRF0. Rozdíly jsou způsobeny především tím, že chraptivý hlas nemá tak dobře definovanou hlasivkovou frekvenci.



Obrázek 4.29: Ukázka průběhů f_0 odhadnutých metodami ACRF0, CEPSF0 a SPECF0 pro (a) zdravý hlas a (b) chraptivý hlas ($G_2R_2B_2T_1$) pro prodlouženou fonaci hlásky /a/.

Tyto diference vedou k zavedení korelací mezi jednotlivými odhady f_0 . Lze předpokládat, že zdravé hlasy budou dosahovat vyšších korelací než hlasy patologické. Korelace jsou označeny jako $corelF0cepsAcr$ pro korelaci mezi průběhem CEPSF0 a ACRF0, $corelF0cepsSpec$ pro korelaci mezi CEPSF0 a SPECF0 a $corelF0specAcr$ pro korelaci mezi ACRF0 a SPECF0.

5. Databáze promluv

V rámci práce byla použita část databáze nahrávek pacientů pořizovaná na Foniatické klinice 1. LF UK a VFN v Praze, která vznikala od roku 1977 do roku 1997. Zdraví i nemocní pacienti byli nahráváni v rámci standardního vyšetřovacího postupu. Nahrávky vznikaly v odhlučněné místnosti s hladinou šumu v pozadí menší než 18 dB SPL. Pacienti při nahrávkách zpravidla seděli, mikrofon byl umístěn na stojanu stojícím na podlaze, vzdálenost mikrofonu od úst nahrávaného byla cca 20-30 cm. Záznamový řetězec obsahoval kondenzátorový mikrofon, předzesilovač, mixážní pult a magnetofon, Vokřál [143].

Nahrávky obsahovaly dlouhé fonace vokálů /a/, /e/, /i/, /o/ a /u/, krátké věty, standardní čtený text *Podzim na Starém bělidle* a volné vyprávění. Kompletní popis typické nahrávky a počty nahrávek pro jednotlivé úkoly jsou uvedeny v příloze 10.

Nahrávky byly původně pořizovány na magnetofonové pásky a následně v roce 2010 digitalizovány profesionální firmou. Každá páska byla uložena jako samostatný WAV soubor se vzorkovací frekvencí 44100 Hz, 16 bitovým rozlišením a dvěma stopami (stereo). Každý takto vzniklý WAV soubor byl autorem ručně rozdělen na jednotlivé nahrávky pomocí programu CoolEdit [124]. Pro další analýzy byl stereo signál převeden na mono.

5.1 Metodika výběru nahrávek

Metodika výběru nahrávek je popsána v práci Vokřál [143].

1. Věk 15 až 60 let – vynechány případy, kde se může projevat vliv stárnutí hlasu a dětské hlasy. Avšak u mužů ve věku 15 let ještě nemusí ve všech případech dojít ke změně hlasu z dětské polohy do dospělé.
2. Zvukový záznam obsahuje prodlouženou fonaci hlásky /a/ a čtení standardního textu *Podzim na Starém bělidle*.
3. Přednost byla dána hlasům s čtenějším počtem zastoupení hlasů v téže skupině nebo diagnóze před ojedinělými.
4. Vzhledem k tomu, že počet možných analyzovaných případů by byl příliš rozsáhlý, byly skupiny s příliš velkým rozsahem zredukovány (zdravé hlasy, parézy zvrtného nervu, ...). Pacienti s opakovaným záznamem a podle subjektivního názoru stejným obrazem chraptivosti byli omezeni na jeden výskyt. Do analýz byly vzaty opakované záznamy pacientů se subjektivně se měnícím obrazem chraptivosti. Tyto záznamy byly chápány jako různé hlasové vzorky.

Na základě záznamů v práci Vokřál [143] bylo zpracováno celkem 55 pásek, ze kterých bylo rozděleno 1681 nahrávek. Z tohoto počtu bylo použito celkem 523 nahrávek, které byly dále rozděleny na jednotlivé úkoly popsané v příloze 10.

Pro experimentální část této práce byly použity nahrávky se dvěma úkoly:

- prodloužená fonace hlásky /a/ (celkem 469 nahrávek);
- čtení standardního textu *Podzim na Starém bělidle* (celkem 593 nahrávek).

5.2 Subjektivní hodnocení nahrávek

K vyhodnocení subjektivních testů hodnocených diskrétními hodnotami, které opakovaně hodnotila skupina hodnotitelů, jsou používány míry obecně známé jako *intra- a inter-rater agreement*.

V experimentu jsou k dispozici subjektivní hodnocení 5 odborníků z Foniatrické kliniky, kteří budou dále v textu označováni zkratkami *A, B, C, D, E*.

5.2.1 Metodika hodnocení nahrávek

Odborníci hodnotili jednak prodlouženou fonaci hlásky /a/ a jednak čtený text *Podzim na Starém bělidle*. Hodnocení nahrávek bylo prováděno modifikovanou metodou GRBAS. Cílem experimentu bylo získat podklady pro natrénování klasifikátorů a zároveň zjistit spolehlivost subjektivního hodnocení.

Modifikace metody GRBAS spočívala ve sloučení parametrů *A* a *S* do parametru *T* (Tension), který mohl nabývat hodnot od -2 do +2 bodů. Záporné hodnoty označovaly hlas subjektivně vnímaný jako slabý a kladné hodnoty označovaly hlas subjektivně vnímaný jako přemáhaný. Důvod k této úpravě je uveden ve Vokřál [143]: hodnotitelé mají v původní stupnici problém během jednoho poslechu určit všechny parametry a parametry *A* a *S* jsou označovány jako ty nejvíce nejednoznačné. Tato modifikace tedy měla usnadnit subjektivní hodnocení.

Obě databáze byly hodnoceny odděleně a každá z databází byla hodnocena dvakrát s alespoň dvoutýdenním odstupem. Pro subjektivní hodnocení byly nahrávky očíslovány pořadovými čísly a tato pořadová čísla byla pro opakované hodnocení změněna.¹

Metody hodnocení *intra- a inter-rater agreement*, čili míra shody jednoho hodnotitele v rámci opakovaného hodnocení a míra shody hodnocení mezi skupinou hodnotitelů budou rozebrány v následujících odstavcích.

5.2.2 Shoda hodnotitelů při opakovaném testu

K hodnocení shody hodnotitelů v rámci opakovaných testů se používají tzv. alfa statistiky. Ty slouží například k odhalení náhodného vyplnění dotazníku. Mezi nejznámější patří Cronbachovo alfa. Cronbachovo alfa (Cronbach [28]) je definováno vztahem

$$\alpha_C = \frac{K\bar{r}}{1 + (K-1)\bar{r}}, \quad (5.1)$$

kde K je počet hodnocení (pro jednou opakované hodnocení $K=2$, pro dvakrát opakované hodnocení $K=3$ atd.) a \bar{r} je průměr horního trojúhelníku korelační matice.

¹Pro vygenerování pořadových čísel nahrávek byla použita funkce *randperm* v MATLABu.

Koeficient α_C může nabývat hodnot $< 0; 1 >$, $\alpha_C \in \mathbb{R}$. Pro určení síly shody lze použít hodnoty uvedené v tab. 5.1. Obecně platí, že shoda vyšší než nula není náhodná, a dále platí, že čím vyšší shody je dosaženo, tím lépe.

Tabulka 5.1: Kvantifikace Cronbachova alfa slovním popisem.

α_C	slovní popis
$\alpha \geq 0,9$	vynikající
$0,8 \leq \alpha < 0,9$	dobrá
$0,7 \leq \alpha < 0,8$	přijatelná
$0,6 \leq \alpha < 0,7$	diskutabilní
$0,5 \leq \alpha < 0,6$	slabá
$\alpha < 0,5$	nepřijatelná

5.2.3 Shoda skupiny hodnotitelů

Pro určení míry shody ve skupině hodnotitelů jsou používány tzv. kappa statistiky. Mezi nejznámější patří Cohenovo kappa, které určuje míru shody dvou hodnotitelů a Fleissovo kappa, které popisuje míru shody skupiny H hodnotitelů.

Cohenovo kappa

K vyhodnocení shody mezi dvěma hodnotiteli se používá Cohenovo kappa definované jako (Cohen [24])

$$\kappa_C = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}, \quad (5.2)$$

kde $Pr(a)$ je relativní shoda mezi hodnotiteli a $Pr(e)$ je hypotetická pravděpodobnost náhodné shody.

Příklad výpočtu κ_C pro dvojici hodnotitelů A a B prvního testu prodloužené fonace hlásky /a/. Výpočet vychází z kontingenční tabulky, která je pro danou dvojici uvedena v tab. 5.2.

Tabulka 5.2: Cohenovo kappa - příklad, kontingenční tabulky pro první hodnocení dvojice odborníků A a B při hodnocení prodloužené fonace hlásky /a/, parametr G.

		A			
		0	1	2	3
B	0	35	41	4	0
	1	39	130	54	3
	2	0	23	71	8
	3	0	1	24	36

Relativní shoda mezi hodnotiteli (tzn. kdy se oba shodnou v hodnocení – hodnoty na diagonále kontingenční tabulky) je dána $Pr(a) = (35 + 130 + 71 + 36)/469 = 0,58$. Hodnota 469 je počet hodnocených nahrávek (suma kontingenční tabulky).

Pro hypotetickou pravděpodobnost náhodné shody je nutné určit podíl jednotlivých hodnocení pro každého z hodnotitelů. Např. B hodnotil známkou 0 $(35 + 41 + 4 + 0)/469 = 0,17$ nahrávek, zatímco A $(35 + 39 + 0 + 0)/469 = 0,16$ nahrávek. Náhodně se tedy mohli shodnout u $0,17 * 0,16 = 0,03$ nahrávek. Pokud sečtete dílčí shody pro jednotlivé stupně hodnocení, dostaneme $Pr(e) = 0,03 + 0,20 + 0,07 + 0,01 = 0,31$.

Hodnoty $Pr(a)$ a $Pr(e)$ dosadíme do vztahu (5.2) a získáme vzájemnou shodu určenou pomocí Cohenova kappa $\kappa_C = 0,39$.

Fleissovo kappa

Zatímco Cohenovo κ_C umí porovnat jen vzájemnou shodu dvou hodnotitelů, Fleissovo kappa (Fleiss [39]) umí porovnat shodu mezi více hodnotiteli. Je definováno vztahem

$$\kappa_F = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e}, \quad (5.3)$$

kde $1 - \bar{P}_e$ je míra shody dosažitelná náhodou, $\bar{P} - \bar{P}_e$ je míra skutečně dosažené shody. Pro další vysvětlení, nechť N je počet subjektů (vyšetřovaných nahrávek), n je počet hodnocení nahrávky a k je počet kategorií, kterými může být nahrávka hodnocena. Nahrávky jsou indexovány $i = 1, \dots, N$ a kategorie jsou indexovány $j = 1, \dots, k$. n_{ij} je počet hodnotitelů, kteří přiřadili i -tou nahrávku do j -té kategorie.

Postup výpočtu κ_F je následující:

1. zjistit podíl všech přiřazení do j -té kategorie – parametr p_j

$$p_j = \frac{1}{Nn} \sum_{i=1}^N n_{ij} \quad (5.4)$$

2. určit míru shody u i -té nahrávky (tzn. určit, kolik párů hodnotitelů hodnotilo do stejné kategorie ku všem možným párům hodnotitelů) – parametr P_i

$$P_i = \frac{1}{n(n-1)} \sum_{j=1}^k n_{ij}(n_{ij} - 1) \quad (5.5)$$

$$= \frac{1}{n(n-1)} \sum_{j=1}^k (n_{ij}^2 - n_{ij}) \quad (5.6)$$

$$= \frac{1}{n(n-1)} \left[\left(\sum_{j=1}^k n_{ij}^2 \right) - (n) \right] \quad (5.7)$$

3. vypočítat \bar{P} jako průměr P_i

$$\bar{P} = \frac{1}{N} \sum_{i=1}^N P_i \quad (5.8)$$

$$= \frac{1}{Nn(n-1)} \left(\sum_{n=1}^N \sum_{j=1}^k n_{ij}^2 - Nn \right) \quad (5.9)$$

4. nakonec vyjádřit \bar{P}_e

$$\bar{P}_e = \sum_{j=1}^k p_j^2 \quad (5.10)$$

Příklad výpočtu κ_F pro shodu mezi hodnotiteli v prvním kole hodnocení. Nejprve je nutné sestavit tabulku s počtem hodnocení v jednotlivých kategoriích pro každou nahrávku, viz tab. 5.3.

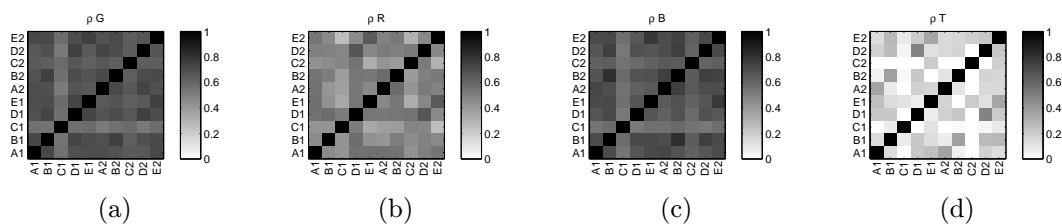
Tabulka 5.3: Fleissovo kappa – příklad frekvenční tabulky hodnocení pro jednotlivé nahrávky s vyznačenými shodami P_i a podíly přiřazení nahrávek do dané kategorie p_j v prvním kole hodnocení, parametr G.

i	kategorie				P_i
	0	1	2	3	
1	0	0	1	4	0,6
2	0	0	5	0	1
3	0	1	3	1	0,3
4	1	4	0	0	0,6
5	0	0	4	1	0,6
...		
469	5	0	0	0	1
total	541	924	623	257	
p_j	0,23	0,39	0,27	0,11	

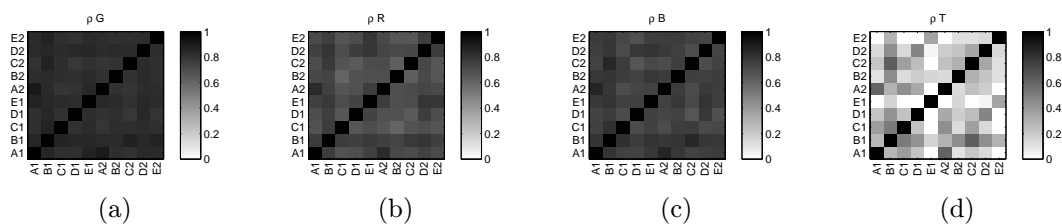
Hodnota $p_0 = 541/2345 = 0,23$, kde 2345 je počet všech hodnocení (počet nahrávek * počet hodnotitelů). P_1 je určeno podle vztahu (5.7) jako $P_1 = \frac{1}{5(5-1)}(0^2 + 0^2 + 1^2 + 4^2 - 5) = 0,6$. Po vyčíslení zbylých p_j a P_i lze snadno podle vztahu (5.8) a (5.10) dopočítat $\bar{P} = 0,50$ a $\bar{P}_e = 0,29$. Pak dosazením do vztahu (5.3) určíme, že $\kappa_F = 0,29$.

5.3 Analýza subjektivního hodnocení

Pro rychlý náhled nad hodnocením lze použít korelační matici Spearmanových korelačních koeficientů ρ , zobrazených v obr. 5.1 pro prodlouženou fonaci hlásky /a/ a v obr. 5.2 pro čtený text. Jejich sumarizace je uvedena v tab. 5.4.



Obrázek 5.1: Korelační matice Spearmanových korelačních koeficientů pro jednotlivé hodnotitele pro parametry G, R, B, T pro prodlouženou fonaci hlásky /a/.


 Obrázek 5.2: Korelační matice Spearmanových korelačních koeficientů pro jednotlivé hodnotitele pro parametry G, R, B, T pro čtený text *Podzim na Starém bělidle*.

Tabulka 5.4: Náhled na hodnocení – korelační koeficienty jednotlivých parametrů.

parametr	/a/				čtený text			
	G	R	B	T	G	R	B	T
$\bar{\rho}$	0,65	0,48	0,66	0,12	0,82	0,71	0,75	0,17
$\max \rho$	0,75	0,66	0,82	0,49	0,90	0,83	0,85	0,61
$\min \rho$	0,52	0,25	0,53	-0,18	0,77	0,61	0,65	-0,25

Pro prodlouženou fonaci hlásky /a/ je z tab. 5.4 a obr. 5.1 patrné, že průměrně je vyšších korelací dosahováno pro parametr G a B (vyšší hodnoty korelačního koeficientu jsou reprezentovány tmavší barvou). Poměrně dobrých korelací je dosahováno i pro parametr R, ale parametr T je téměř nekorelovaný. Podobných výsledků je dosahováno i pro čtený text.

Obr. 5.1 také ukazuje menší korelaci hodnotitele C na ostatní hodnotící (světlejší pruhy v korelačních maticích). Oproti tomu je vidět dobrá korelace prvního a opakovaného hodnocení hodnotitele E (tmavší čtverce na souřadnicích [E1;E2]). Na obr. 5.2 nejsou výraznější výkyvy jednotlivců tak dobře patrné.

Pro porovnání subjektivního hodnocení prodloužené fonace hlásky /a/ a čteného textu je z tab. 5.4 dobře patrné, že při hodnocení čteného textu dosahují hodnotitelé lepších výsledků.

5.3.1 Přísnost opakovaného hodnocení

Podle závěrů De Bodt et al. [29] a Karnell et al. [72] lze usoudit, že při opakovaném poslechu byl hodnotící ve většině případů přísnější – nahrávku hodnotil vyšším

stupněm patologie.

V rámci tohoto experimentu byl proveden neparametrický Mann–Whitneyho U test, který porovnává mediány dvou populací, tzn. hodnocení v prvním a ve druhém kole hodnocení. Výsledky testu jsou shrnuty v tab. 5.5, ze které lze odečíst, že jednotlivci nejsou přísnější ani méně přísní (kromě výjimky u hodnotitele A u prodloužené fonace a C u čteného textu), ovšem jako skupina jsou hodnotitelé pro prodlouženou fonaci hlásky /a/ méně přísní a pro čtený text naopak více přísní.

Ojedinelý případ nastal u hodnotitele B, který při hodnocení prodloužené fonace hlásky /a/ hodnotil výrazně jinak parametr T: v prvním hodnocení hodnotil hlasy více jako slabé, ve druhém naopak jako přemáhané.

Tabulka 5.5: Porovnání přísnosti subjektivního hodnocení při opakovaném testu pomocí Mann–Whitneyho U testu: \uparrow značí přísnější hodnocení, \downarrow značí méně přísné hodnocení a \cdot značí hodnocení beze změny. Záznamy označené * označují stav, kdy došlo ke změně znaménka u parametru T, tzn. změně hodnocení hlasové slabosti a napětí v hlase. Hladina významnosti testu $p=0,01$.

hodnotitel	/a/				čtený text			
	G	R	B	T	G	R	B	T
A	\downarrow	\downarrow	\downarrow	\cdot	\cdot	\cdot	\cdot	\cdot
B	\cdot	\cdot	\cdot	\uparrow^*	\cdot	\cdot	\cdot	\cdot
C	\cdot	\cdot	\cdot	\uparrow	\uparrow	\uparrow	\uparrow	\cdot
D	\cdot	\cdot	\cdot	\cdot	\uparrow	\cdot	\uparrow	\cdot
E	\cdot	\cdot	\downarrow	\cdot	\cdot	\cdot	\cdot	\cdot
všichni	\downarrow	\cdot	\downarrow	\downarrow	\uparrow	\uparrow	\uparrow	\cdot

5.3.2 Vzájemná korelace parametrů GRBT

V rámci experimentů byla prověřena i vzájemná korelace parametrů, tzn. do jaké míry na sebe korelují např. parametry G a R atd. Výsledky tohoto experimentu jsou shrnuty v tab. 5.6.

Tabulka 5.6: Vzájemná korelace parametrů G, R, B, T pro prodlouženou fonaci hlásky /a/ (horní trojúhelník) a pro čtený text (dolní trojúhelník).

	G	R	B	T
G	—	0,72	0,71	-0,05
R	0,86	—	0,44	0,06
B	0,81	0,68	—	-0,24
T	0,09	0,14	-0,04	—

Z tab. 5.6 je dobře patrné, že celkový dojem z hlasu G vždy nejvíce koreluje s hodnocením chraplavosti R, nicméně v porovnání s hodnocením dyšnosti B se

jedná o malé rozdíly. Oproti tomu korelace mezi hodnocením R a B je více rozdílná jak v případě hodnocení prodloužené fonace hlásky /a/, tak v případě hodnocení čteného textu. Hodnocení hlasového napětí nebo naopak hlasové slabosti T téměř nekoreluje s žádným dalším parametrem.

Další možnost pohledu na shodu hodnocení jednotlivých parametrů spočívá ve vyčíslení normované vzdálenosti hodnocení NDE (*Normalized Distance of Evaluation*) podle vztahu

$$\text{NDE}_{AB} = \frac{1}{N} \sum_{i=0}^{N-1} |A_i - B_i|, \quad (5.11)$$

kde N je počet hodnocených nahrávek a A , resp. B jsou jednotlivé hodnocené parametry G, R, B, T. Normování je prováděno především z důvodu nestejného počtu nahrávek prodloužené fonace hlásky /a/ a čteného textu. Čím nižší hodnota NDE, tím je menší rozdíl v hodnocení obou parametrů. Hodnoty NDE jsou vyčísleny v tab. 5.7.

Tabulka 5.7: Normovaná vzdálenost hodnocení parametrů G, R, B, T pro prodlouženou fonaci hlásky /a/ (horní trojúhelník) a pro čtený text (dolní trojúhelník).

	G	R	B	T
G	—	0,38	0,54	1,26
R	0,22	—	0,70	1,09
B	0,34	0,45	—	1,18
T	0,90	0,82	0,83	—

Z tab. 5.7 lze vyčíst, že celkový dojem z hlasu G je v obou případech nejbližší hodnocení chraplavosti R.

Z uvedených výsledků lze usuzovat, že hodnotitelé se nejvíce při hodnocení celkového dojmu z hlasu zaměřují na chraplavost, tedy na míru nepravidelnosti kmitání hlasivek. Z toho lze předpokládat, že parametry, které budou dobře popisovat chraplavost R budou dobře popisovat i celkový dojem z hlasu G.

5.3.3 Shoda hodnotitelů při opakovaném hodnocení

Tab. 5.8 obsahuje souhrnné výsledky shody hodnotitelů (konzistence jejich hodnocení). Test je proveden pomocí Cronbachova alfa. Z tab. 5.1 je patrné, že čím je vyšší hodnota Cronbachova alfa, tím je lepší vzájemná shoda hodnotitele v opakovaném testu.

Tab. 5.8 ukazuje, že hodnotitelé dosahují lepší konzistence při hodnocení čteného textu než prodloužené fonace. Nejlepší shody při opakovaném hodnocení je dosaženo pro hodnocení celkového dojmu z hlasu (G), následuje dyšnost (B), chraplavost (R) a napětí v hlase (T). Z výsledků je také možné odečíst výrazně lepší konzistenci v hodnocení parametru T pro čtený text než pro prodlouženou fonaci.

Tabulka 5.8: Výsledky Cronbachova alfa vyjadřující sílu shody mezi prvním a opakovaným subjektivním testem pro jednotlivé hodnotitele a pro skupinu hodnotitelů jako celek.

hodnotitel	Cronbachovo alfa							
	/a/				čtený text			
	G	R	B	T	G	R	B	T
<i>A</i>	0,80	0,72	0,81	0,50	0,95	0,92	0,93	0,73
<i>B</i>	0,84	0,68	0,85	0,52	0,91	0,85	0,87	0,58
<i>C</i>	0,72	0,64	0,67	0,10	0,90	0,80	0,85	0,56
<i>D</i>	0,85	0,79	0,80	0,65	0,90	0,86	0,83	0,66
<i>E</i>	0,86	0,78	0,87	0,48	0,91	0,89	0,85	0,51
všichni	0,84	0,73	0,82	0,50	0,86	0,81	0,82	0,58

5.3.4 Shoda skupiny hodnotitelů

Výpočet shody hodnocení mezi každými dvěma hodnotiteli na základě Cohenova kappa je uveden v tab. 5.9. Z uvedených hodnot je patrné, že dvojice hodnotitelů se lépe shodnou při hodnocení čteného textu. Výsledky shody u jednotlivých parametrů jsou srovnatelné. Zajímavostí je vysoká shoda u parametru T při hodnocení čteného textu.

Dosažené hodnoty Cohenova kappa jsou poměrně nízké, ovšem i takto nízké hodnoty ukazují, že se nejedná o náhodnou shodu.

Tabulka 5.9: Výsledky Cohenova kappa – vyhodnocení shody dvojic hodnotitelů pomocí Cohenova kappa v 1. a 2. kole hodnocení.

	/a/								čtený text							
	1. kolo				2. kolo				1. kolo				2. kolo			
	G	R	B	T	G	R	B	T	G	R	B	T	G	R	B	T
<i>A – B</i>	0,29	0,25	0,40	0,21	0,20	0,18	0,41	0,12	0,55	0,47	0,43	0,46	0,60	0,48	0,44	0,46
<i>A – C</i>	0,25	0,25	0,21	0,08	0,20	0,11	0,32	0,16	0,47	0,39	0,54	0,43	0,16	0,23	0,43	0,33
<i>A – D</i>	0,18	0,22	0,38	0,23	0,19	0,21	0,38	0,18	0,20	0,16	0,27	0,16	0,09	0,19	0,29	0,17
<i>A – E</i>	0,28	0,32	0,41	0,09	0,17	0,18	0,40	0,15	0,57	0,53	0,55	0,22	0,57	0,45	0,46	0,17
<i>B – C</i>	0,25	0,24	0,28	0,16	0,32	0,14	0,27	0,12	0,62	0,57	0,52	0,44	0,22	0,32	0,28	0,26
<i>B – D</i>	0,38	0,29	0,34	0,23	0,41	0,20	0,43	0,21	0,34	0,28	0,26	0,18	0,14	0,21	0,15	0,13
<i>B – E</i>	0,39	0,24	0,30	0,05	0,38	0,24	0,40	0,03	0,64	0,54	0,52	0,22	0,66	0,56	0,47	0,22
<i>C – D</i>	0,27	0,27	0,17	0,08	0,31	0,17	0,28	0,05	0,41	0,34	0,23	0,09	0,49	0,29	0,38	0,25
<i>C – E</i>	0,29	0,24	0,25	0,09	0,34	0,26	0,32	0,10	0,65	0,58	0,57	0,14	0,26	0,27	0,35	0,05
<i>D – E</i>	0,39	0,26	0,35	0,03	0,33	0,20	0,33	0,04	0,33	0,22	0,32	0,03	0,16	0,16	0,16	-0,01
max	0,39	0,32	0,41	0,23	0,41	0,26	0,43	0,21	0,65	0,58	0,57	0,46	0,66	0,56	0,47	0,46
min	0,18	0,22	0,17	0,03	0,17	0,11	0,27	0,03	0,20	0,16	0,23	0,03	0,09	0,16	0,15	-0,01
průměr	0,30	0,26	0,31	0,13	0,29	0,19	0,35	0,12	0,48	0,41	0,42	0,24	0,34	0,32	0,34	0,20
skupina	G		R		B		T		G		R		B		T	
	0,41		0,33		0,43		0,37		0,42		0,40		0,43		0,45	

Tab. 5.10 ukazuje vyhodnocení shody skupiny všech hodnotitelů pomocí Fleissova kappa κ_F .

I z těchto výsledků je patrné, že skupina je ve svém hodnocení více konzistentní u čteného textu.

Tabulka 5.10: Výsledky Fleissova kappa – vyhodnocení shody skupiny hodnotitelů pomocí Fleissova kappa.

	/a/				čtený text			
	G	R	B	T	G	R	B	T
1. kolo	0,29	0,26	0,30	0,09	0,48	0,38	0,36	0,18
2. kolo	0,27	0,18	0,35	0,08	0,41	0,36	0,37	0,19
1. a 2. kolo	0,29	0,23	0,33	0,12	0,45	0,39	0,37	0,22

6. Experimentální hodnocení chraptivosti

6.1 Parametrizace použité pro automatické hodnocení chraptivosti

K automatickému hodnocení chraptivosti byly použity akustické parametry popsané v kapitole 4. Nahrávky obsahující prodlouženou fonaci byly parametrizovány v okně délky 80 ms se 75 ms překryvem. Nahrávky obsahující čtený text byly analyzovány v okně délky 20 ms s 15 ms překryvem; při výpočtu spektra bylo použito Hammingovo okno příslušné délky. Parametr GNE byl v obou případech počítán podle doporučení uvedených Godino-Llorente et al. [49]: vstupní signál byl podvzorkován na 10 kHz, délka okna 60 ms, šířka pásmové propusti 1000 Hz a posun centrální frekvence pásmové propusti 300 Hz. Při výpočtu parametru GVA bylo použito okno délky 5 ms.

6.1.1 Popis distribuce hodnot

Pro lepší popis nebyla použita pouze střední hodnota (μ) a směrodatná odchylka (σ) použitých parametrů, ale i další parametrizace: medián hodnot (M) a relativní směrodatná odchylka (σ_R).

Medián byl zaveden především z toho důvodu, že není apriori zřejmé, zda rozdělení hodnot jednotlivých parametrizací v rámci nahrávky má normální rozdělení.

Relativní směrodatná odchylka σ_R je definována vztahem

$$\sigma_R = \frac{\sigma}{\mu}, \quad (6.1)$$

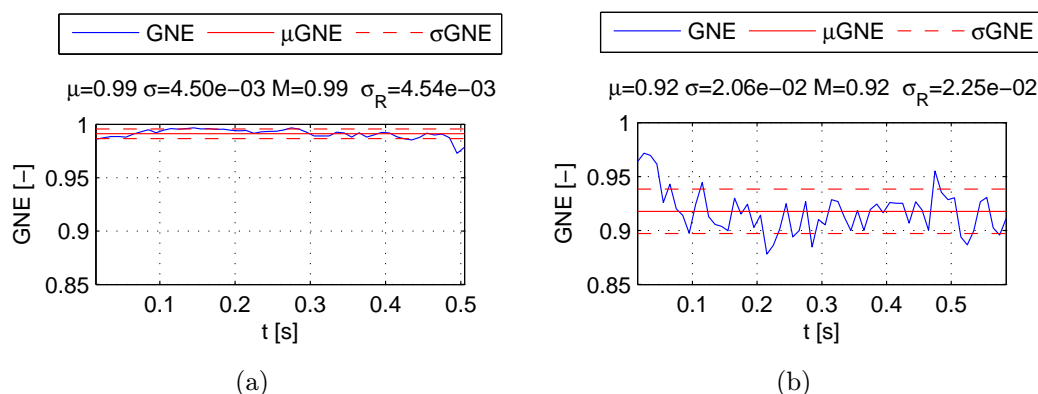
kde σ a μ je směrodatná odchylka a střední hodnota pro daný soubor hodnot.

Na obr. 6.1 jsou zobrazeny rozdíly střední hodnoty a směrodatné odchylky pro průběh parametru GNE pro zdravý hlas (obr. 6.1(a)) a pro chraptivý hlas (obr. 6.1(b)). Rozdíly jsou jasně patrné.

6.1.2 Popis časového průběhu

Pro získání dalších parametrizací byly použity popisy časového vývoje daného parametru. Důvody k zavedení takových popisů motivují následující předpoklady:

- lze předpokládat, že časové průběhy parametrů pro zdravý hlas nebudou obsahovat nenadálé rychlé změny;
- lze předpokládat, že časové průběhy parametrů budou pro zdravý hlas v rámci nahrávky vyrovnanější.



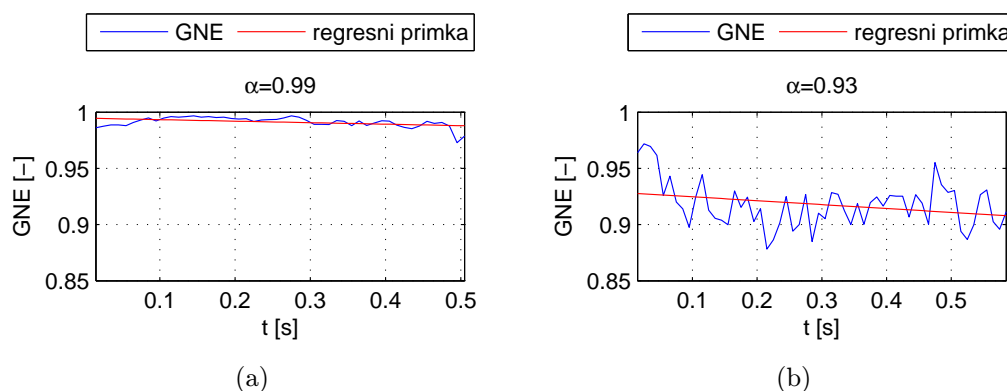
Obrázek 6.1: Ukázka střední hodnoty a směrodatné odchylky parametru GNE pro (a) zdravý a (b) chraptivý hlas.

Parametr délka křivky (CL - Curve Length) je definován podle vztahu (Pan et al. [114])

$$CL = \frac{1}{N} \sum_{n=1}^{N-1} |x[n-1] - x[n]|, [-] \quad (6.2)$$

kde $x[n]$ je vstupní signál délky N . Podle předpokladu by časově stabilnější průběhy měly být kratší, tzn. hodnoty parametru CL by mělo dosahovat nižších hodnot.

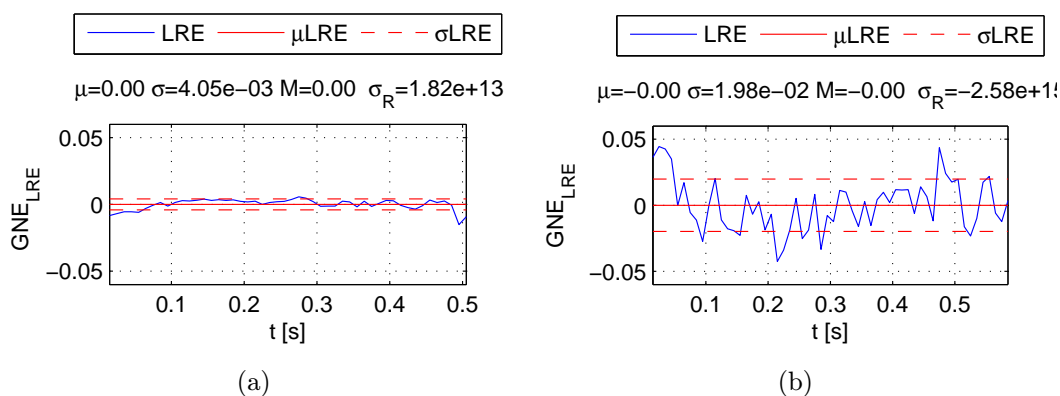
Popis časového průběhu za pomoci lineární regrese je další metodou použitou k popisu vlastností časového průběhu. Lineární koeficient regrese přímky α byl použit jako jeden z parametrů. Obr. 6.2 ukazuje rozdíly ve směrnici regrese přímky pro zdravý hlas (obr. 6.2(a)) a pro chraptivý hlas (obr. 6.2(b)). Z obrázku je patrné, že zdravý hlas má vyrovnanější průběh, zatímco chraptivý hlas dosahuje v závěru nahrávky nižších hodnot a tím pádem má regrese přímka větší sklon. Tento pokles v závěru může být způsoben větší únavou hlasového traktu dlouhou fonací.



Obrázek 6.2: Ukázka trendu časového průběhu parametru GNE pro (a) zdravý a (b) chraptivý hlas.

Popis pomocí LRE

Další parametrizace odpovídá popisu signálu zbaveného lineárního trendu. Tento parametr je nazván LRE (Linear regression error – chyba lineární regrese). Zavedení tohoto parametru je motivováno výše zmíněným faktem: u zdravého hlasu je předpokládán vyrovnanější, více lineární průběh parametrů v průběhu nahrávky. Na obr. 6.3 je zobrazen průběh parametru GNE popsáno pomocí LRE (značené GNE_{LRE}). Z popisu pomocí LRE jsou spočteny následující charakteristiky: směrodatná odchylka (LRE_{σ}), maximální a minimální hodnota (LRE_{max} a LRE_{min}), rozsah hodnot ($\text{LRE}_{\text{range}}$) a maximální vzdálenost od nulové hodnoty ($\text{LRE}_{\text{maxabs}}$). Střední hodnotu a medián LRE není nutné určovat; vzhledem ke způsobu výpočtu parametru LRE budou tyto hodnoty nulové.



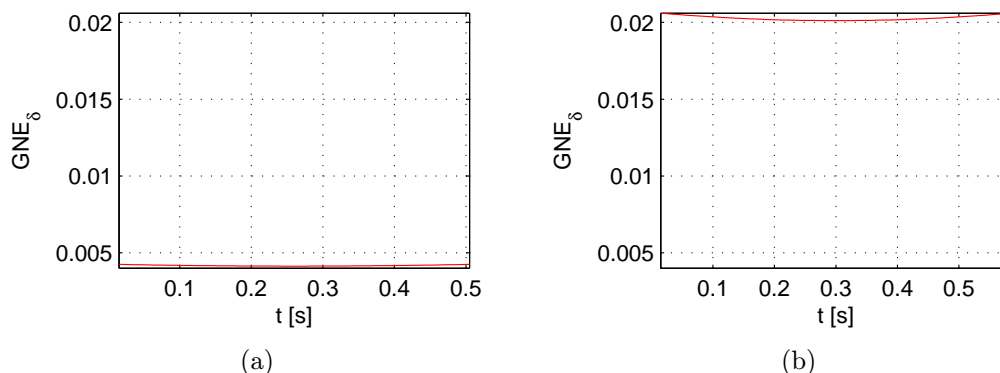
Obrázek 6.3: Ukázka časového průběhu parametru GNE zbaveného trendu pro (a) zdravý a (b) chraptivý hlas.

Popis pomocí δ

Jiná možnost popisu chyby predikce je implementována přímo v prostředí MATLAB. Jedná se o vektor $\delta[n]$, jednu z výstupních hodnot funkce *polyval*. V dokumentaci funkce je uvedeno: „ $\delta[n]$ je odhad směrodatné odchylky chyb při určení následujícího vzorku $x[n]$ ze vstupních dat“. Na obr. 6.4 je zobrazen vektor hodnot $\delta[n]$ pro průběh parametru GNE. Z obrázku je patrné, že pro zdravý hlas (obr. 6.4(a)) je dosaženo výrazně nižších hodnot než pro chraptivý hlas (obr. 6.4(b)). To je způsobeno větší časovou nestabilitou průběhu parametru pro chraptivý hlas. Z parametrizace δ jsou určeny následující hodnoty: střední hodnota (δ_{μ}), směrodatná odchylka (δ_{σ}), relativní směrodatná odchylka (δ_{σ_R}), minimální a maximální hodnota (δ_{min} a δ_{max}) a rozsah hodnot (δ_{range}).

6.1.3 Parametrizace pouze ve znělých částech

V rámci experimentálního objektivního hodnocení chraptivosti byly analyzovány i nahrávky obsahující čtený text. Takové nahrávky obsahují jak znělé, tak neznělé

Obrázek 6.4: Ukázka průběhu parametrizace δ pro (a) zdravý a (b) chraptivý hlas.

úseky. Některé z použitých parametrizací je nemožné určit na úseku neznělého signálu. Pro databázi nahrávek se čteným textem byly určeny znělé úseky signálu a vybrané parametrizace byly spočteny jak pro celou nahrávku, tak pouze pro znělé úseky signálu. Tyto parametrizace jsou v dalších experimentech označeny příponou *voiced* (tzn. znělý).

K detekci znělosti je použit algoritmus RAPT (Robust Algorithm for Pitch Tracking) představený v práci Talkin [136], který pracuje s normalizovanou kroskorelační funkcí (NCF – Normalized Crosscorrelation Function) a dále je využito množství dodatečných výpočtů, které odstraňují špatné detekce znělý/neznělý úsek. V práci Drugman and Alwan [35] je na základě experimentů RAPT označen jako nejlepší pro detekci znělých úseků jak spojité řeči, tak prodloužené fonace. Algoritmus je implementován např. v programu Wavesurfer (Sjölander and Beskow [133]) nebo v balíku skriptů pro MATLAB Voicebox (Brookes [22]).

Na obr. 6.5 je ukázka porovnání detekce znělosti algoritmem RAPT (implementace v programu Wavesurfer), autokorelační funkcí (implementace v programu Praat, Boersma and Weenink [21]) a ručním labelováním. Ruční labelování bylo prováděno na základě zkušeností autora této práce a to v místech, kde měl signál vizuálně alespoň náznak periodicity a v poslechu byl patrný náznak tónu. Z uvedeného obrázku je patrná nejlepší shoda mezi ručním labelováním a výstupem algoritmu RAPT.

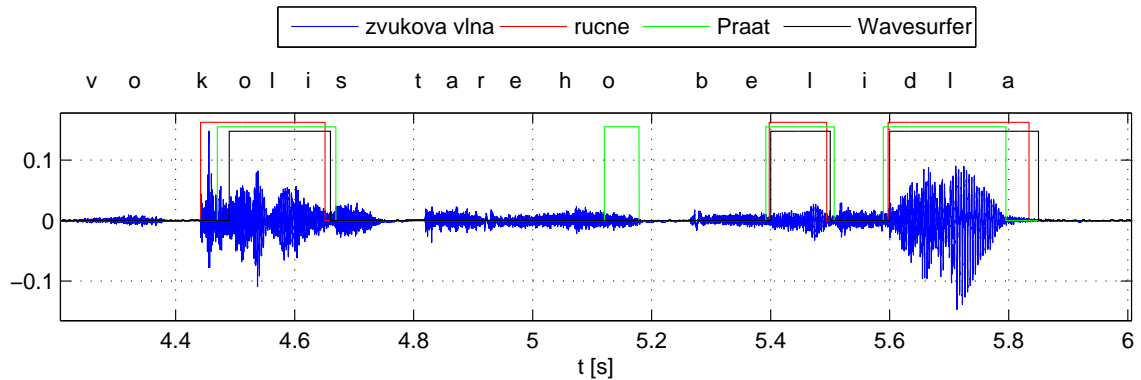
6.1.4 Metodika popisu parametrizací

Akustické parametrizace uvedené v kapitole 4 byly analyzovány výše popsanými metodami. Výsledný název parametru je složen ze jména parametru a použité dodatečné parametrizace. Seznam symbolů dodatečných parametrizací je uveden v tab. 6.1. Metodika tvoření výsledného popisu je prováděna podle následujícího klíče:

$$\text{PARAMETR}_{\text{voiced}_{\text{parametrizace}}}$$

kde „PARAMETR“ je jeden z parametrů uvedených v kapitole 4, „voiced“ určuje, zda byl parametr počítán pouze ve znělých úsecích (pokud není uveden, parametr byl počítán pro celý signál) a „parametrizace“ je jeden z parametrů popsaných v tab. 6.1.

6.1. Parametrizace použité pro automatické hodnocení chraptivosti



Obrázek 6.5: Porovnání detekce znělosti na krátkém úseku čteného textu.

Tabulka 6.1: Seznam symbolů použitý k popisu parametrizace. Výsledný popis je složen ze jména parametru (např. GNE) a použité dodatečné parametrizace. Příklad: střední hodnota parametru GNE spočtená pouze ve znělých úsecích vstupního signálu: GNE_{voiced}_{μ} .

zkratka	popis
μ	střední hodnota
σ	směrodatná odchylka
M	medián
σ_R	relativní směrodatná odchylka
CL	délka křivky
α	směrnice lineární regresní přímky
LRE	časový průběh parametru zbavený lineárního trendu
LRE_{σ}	směrodatná odchylka hodnot LRE
LRE_{\max}	maximální hodnota LRE
LRE_{\min}	minimální hodnota LRE
LRE_{range}	rozsah hodnot LRE
$LRE_{\max\text{abs}}$	maximální vzdálenost hodnoty LRE od nuly (LRE má nulovou střední hodnotu)
δ	odhad směrodatné odchylky chyb při určení následujícího vzorku $x[n]$ ze vstupních dat
δ_{μ}	střední hodnota δ
δ_{σ}	směrodatná odchylka δ
δ_{σ_R}	relativní směrodatná odchylka δ
δ_{\min}	minimální hodnota δ
δ_{\max}	maximální hodnota δ
δ_{range}	rozsah hodnot δ
voiced	parametrizace byla spočtena pouze ve znělých úsecích signálu

6.2 Definice klasifikační úlohy

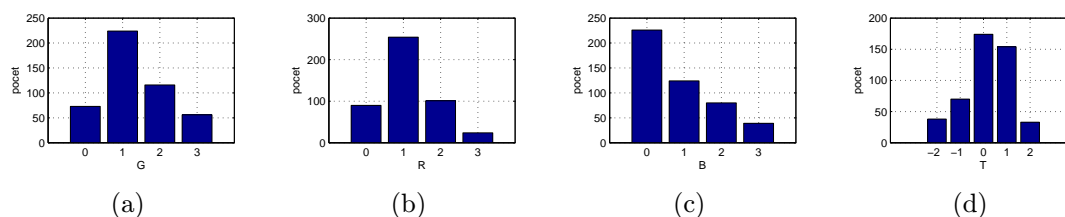
Cílem práce je nalézt klasifikátor, který s dostatečnou přesností určí míru chraptivosti podle stupnice GRBT. Jak je uvedeno v kapitole 3 na obr. 3.1, navrhovaný automatický systém pro hodnocení míry chraptivosti klasifikuje jednotlivé příznaky chraptivosti odděleně. Výstupem klasifikátoru je sada čtyř hodnot, které definují jednotlivé příznaky chraptivosti specifikované subjektivní metodou GBRT: celkový dojem z hlasu (G), míru nepravidelnosti kmitání hlasivek (R), míru šumové příměsi v hlasu (B) a vjem spojený s hlasovou slabostí, nebo naopak s přemáháním hlasu (T).

Byly zavedeny dvě klasifikační úlohy: hodnocení podle diskretní škály GRBT a hodnocení podle spojité škály GRBT.

6.2.1 Diskrétní hodnocení

Při diskretním hodnocení jsou pro učení klasifikátoru použity diskretní hodnoty jednotlivých ukazatelů stupnice GRBT. Tyto hodnoty jsou získány jako modus množiny deseti hodnocení každé nahrávky.¹

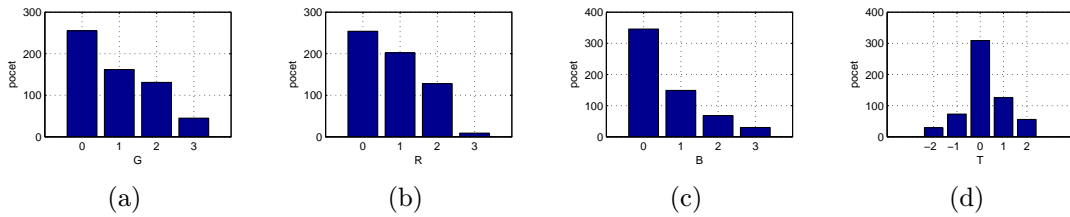
Na obr. 6.6 a obr. 6.7 jsou zobrazeny počty nahrávek pro jednotlivé stupně hodnocení pro prodlouženou fonaci (obr. 6.6) a pro čtený text (obr. 6.7). Při porovnání odpovídajících histogramů (tzn. G pro prodlouženou fonaci a pro čtený text) je možné si všimnout rozdílné strategie hodnocení prodloužené fonace a čteného textu. Zatímco při hodnocení prodloužené fonace je pro parametry G a R nejčastěji použito hodnocení jedním bodem (tzn. lehký příznak daného parametru), při hodnocení čteného textu je častěji použito hodnocení nulou bodů, tzn. bez příznaku. U parametru B však toto neplatí. Tyto rozdíly však mohou být způsobeny následujícím faktem: pro čtený text je k dispozici mnohem více nahrávek od pacientů, kteří byli původně označeni jako zdraví, viz tab. 10.1 v příloze. Z obrázků je ovšem také dobře patrné, že nejtěžší stupně $G=3$, $R=3$ a $B=3$ jsou v databázi zastoupeny nejméně. To je sice přirozený stav v populaci, nicméně pro trénování klasifikátoru to může představovat problém (může vzniknout největší chyba klasifikace).



Obrázek 6.6: Histogramy diskretního hodnocení nahrávek prodloužené fonace metodou GRBT, (a) G, (b) R, (c) B, (d) T.

Při použití diskretního hodnocení bude výsledné hodnocení ve stejném stylu, jaké použili odborníci, tzn. výsledná nahrávka bude hodnocena např. $G_1R_0B_2T_{-1}$.

¹Každá nahrávka byla hodnocena dvakrát pěti odborníky, viz popis v části 5.2.1.

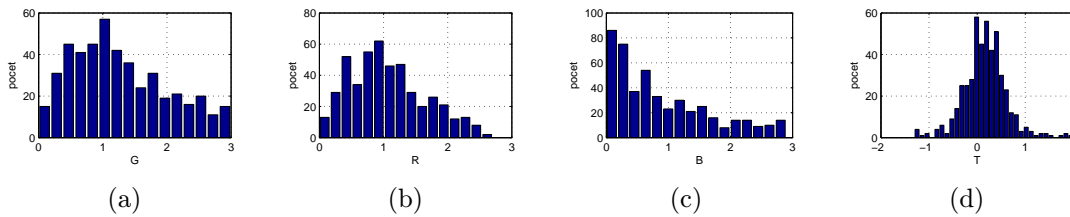


Obrázek 6.7: Histogramy diskrétního hodnocení nahrávek čteného textu metodou GRBT, (a) G, (b) R, (c) B, (d) T.

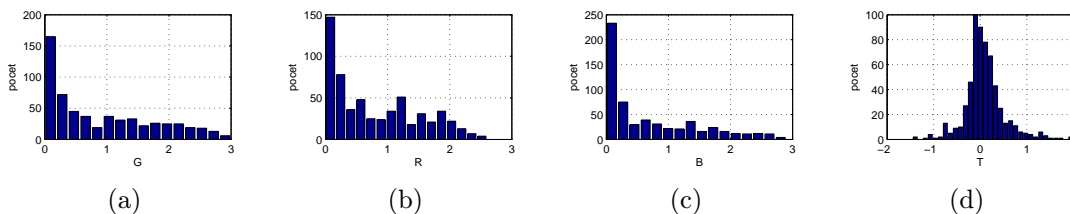
6.2.2 Spojité hodnocení

U spojitého hodnocení jsou jako vstup do klasifikátoru použity spojitě hodnoty GRBT, které vznikly jako průměrné hodnocení dané nahrávky. Průměr je určen z deseti hodnot subjektivního hodnocení.

Na obr. 6.8 a obr. 6.9 jsou zobrazeny distribuce hodnocení v databázi prodloužené fonace (obr. 6.8) a čteného textu (obr. 6.9).



Obrázek 6.8: Histogramy spojitého hodnocení nahrávek prodloužené fonace metodou GRBT, (a) G, (b) R, (c) B, (d) T.



Obrázek 6.9: Histogramy spojitého hodnocení nahrávek čteného textu metodou GRBT, (a) G, (b) R, (c) B, (d) T.

Při použití spojitého hodnocení bude výsledné skóre pro promluvu vypadat například následovně: $G_{1,23}R_{0,54}B_{1,97}T_{-0,23}$. Takové hodnocení vyjadřuje, že analyzovaná nahrávka má celkový dojem z hlasu horší než $G=1$, ale mnohem lepší než $G=2$, atd.

6.2.3 Použité metody dolování dat

Veškeré klasifikační úlohy a redukce příznakového prostoru byly provedeny za pomoci programu pro dolování dat WEKA ([56]). V rámci těchto úloh byla použita tzv. metoda křížové validace (z angličtiny *cross-validation*). Tato metoda je používána ke zjištění, jak bude výsledný model (v tomto případě klasifikátor) ovlivňován nezávislými vzorky. Během křížové validace je vstupní soubor dat rozdělen na N podmnožin. Poté je jedna z podmnožin použita pro natrénování modelu a zbylých $N-1$ podmnožin je použito pro testování modelu. V případě, že trénovaný model je klasifikátor, dojde k určení přesnosti klasifikátoru na neznámých datech. Tento postup je opakován a každá z podmnožin je postupně použita pro trénování modelu.

K testování v programu WEKA bylo při křížové validaci použito 10 podmnožin. Program WEKA při tvorbě jednotlivých podmnožin zachová poměry jednotlivých tříd (v tomto případě subjektivních hodnocení) v každé podmnožině. Výsledné skóre klasifikátoru určené programem WEKA je průměrné dosažené skóre z jednotlivých modelů.

6.2.4 Redukce příznakového prostoru pro klasifikaci

Pro každou nahrávku obsahující prodlouženou fonaci bylo spočteno celkem 480 parametrizací a pro nahrávky obsahující čtený text celkem 201 parametrizací. Kompletní seznam parametrů včetně jejich vlastností je na přiloženém CD.

Takto rozsáhlé příznakové prostory je nutné vhodným způsobem zmenšit – vyřadit parametry, které nesou redundantní informaci nebo jejich distribuce neodpovídá výsledné třídě.

V rámci této práce byla použita dvoufázová redukce příznakového prostoru. V prvním kroku byla jak pro spojitě, tak pro diskrétní hodnocení použita korelační analýza (funkce *CfsSubsetEval* v programu WEKA, Hall [55]). Tento algoritmus hledá takové parametry, které mají nejvyšší korelaci s danou třídou (tzn. subjektivním hodnocením G, R, B nebo T) a malou korelaci s ostatními parametry. Do druhého kroku redukce příznakového prostoru byly vybrány ty parametry, které se umístily ve všech analyzovaných podmnožinách při křížové validaci.

Algoritmus použitý v druhém kroku se lišil podle typu hodnocení (diskrétní vs. spojitě). Pro diskrétní hodnocení byl použit algoritmus založený na informačním zisku (funkce *InfoGainAttributeEval* v programu WEKA). Pro spojitě hodnocení byl použit algoritmus RELIEF (funkce *ReliefFAttributeEval* v programu WEKA, Kira and Rendell [74], Kononenko [75], Robnik-Sikonja and Kononenko [125]).

Jako konečný soubor příznaků byly vybrány maximálně čtyři příznaky, které se umístily nejlépe ve druhém kroku redukce příznakového prostoru.

6.2.5 Použité klasifikátory

V programu WEKA je implementované velké množství klasifikátorů. Některé jsou schopné pracovat pouze s diskrétními třídami a jiné pouze se spojitými třídami.

Pro diskrétní hodnocení byly použity tři klasifikátory: rozhodovací strom J48, který implementuje algoritmus C4.5 (Quinlan [121]); algoritmus k-NN (k nejbližších

sousedů)² a bayesovské rozhodování (John and Langley [70]). Výstupem z těchto klasifikátorů je matice záměn, ze které lze odečíst, jakým způsobem algoritmus klasifikoval známá vstupní data. Z matice záměn lze odečíst nejen přesnost klasifikátoru, ale i třídy, které způsobují největší chybu klasifikace.

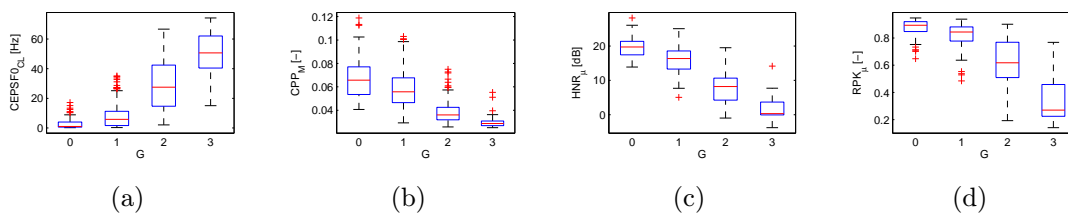
Pro spojitá hodnocení byly použity také tři klasifikátory, které hledají vhodnou kombinaci parametrů, která nejlépe koreluje se vstupními daty. Algoritmus M5Rules (Holmes et al. [66], Quinlan [122], Wang and Witten [144]) a rozhodovací strom M5P, který je založen na výše zmíněném algoritmu M5Rules (Quinlan [122], Wang and Witten [144]). Třetí použitý algoritmus je prostá lineární regrese. Výstupem z těchto algoritmů je Pearsonův lineární korelační koeficient r .

6.3 Výsledky automatického hodnocení prodloužené fonace

6.3.1 Diskrétní hodnocení

Parametr G

Po redukcí příznakového prostoru byly použity parametry: $\text{CEPSF0}_{\text{CL}}$, CPP_{M} , HNR_{μ} a RPK_{μ} . Rozdělení hodnot parametrů pro jednotlivé stupně hodnocení je na obr. 6.10. Výsledky klasifikace jsou uvedeny v tab. 6.2. Použitý klasifikátor: 10-NN. Tento algoritmus pracuje se vstupní množinou dat – schéma algoritmu není možné vytvořit.



Obrázek 6.10: Boxploty pro hodnoty parametrů použitých pro hodnocení diskrétního G u prodloužené fonace; (a) $\text{CEPSF0}_{\text{CL}}$, (b) CPP_{M} , (c) HNR_{μ} a (d) RPK_{μ} .

Parametr R

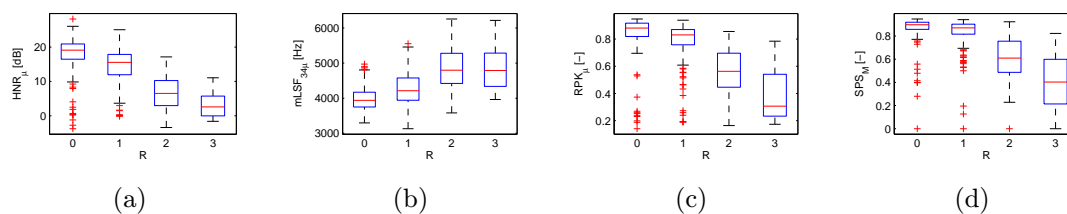
Po redukcí příznakového prostoru byly použity parametry: HNR_{μ} , $\text{mLSF}_{34\mu}$, RPK_{μ} a SPS_{M} . Rozdělení hodnot parametrů pro jednotlivé stupně hodnocení je na obr. 6.11. Výsledky klasifikace jsou uvedeny v tab. 6.3. Použitý klasifikátor: bayesovské rozhodování, hodnoty použité pro klasifikaci viz tab. 7.2, parametry R, prodloužená fonace, $\mu \pm \sigma$ pro jednotlivé stupně hodnocení.

²Pro klasifikační úlohu bylo zvoleno $k=10$.

Tabulka 6.2: Matice záměn pro automatické objektivní hodnocení diskretního G u prodloužené fonace.

		G			
		objektivní			
		0	1	2	3
subjektivní	0	43	49	0	0
	1	38	180	11	0
	2	0	35	61	9
	3	0	1	17	25

přesnost: 66 %
Cohenovo kappa [-]: 0,47


 Obrázek 6.11: Boxploty pro hodnoty parametrů použitých pro hodnocení diskretního R u prodloužené fonace; (a) HNR_μ, (b) mLSF_{34μ}, (c) RPK_μ a (d) SPS_M.

Tabulka 6.3: Matice záměn pro automatické objektivní hodnocení diskretního R u prodloužené fonace.

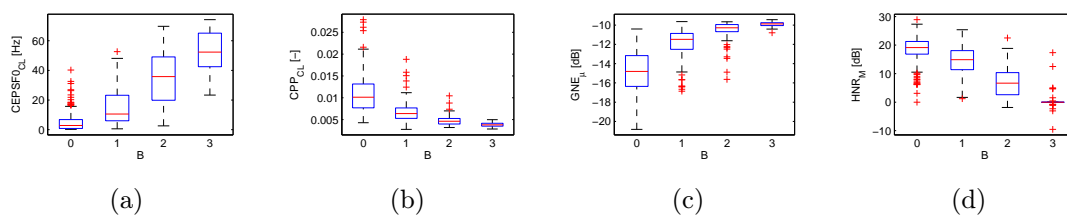
		R			
		objektivní			
		0	1	2	3
subjektivní	0	56	64	2	9
	1	21	193	11	8
	2	0	28	45	16
	3	0	1	8	7

přesnost: 64 %
Cohenovo kappa [-]: 0,42

Parametr B

Po redukci příznakového prostoru byly použity parametry: CEPSF0_{CL}, CPP_{CL}, GNE_μ a HNR_M. Rozdělení hodnot parametrů pro jednotlivé stupně hodnocení je na obr. 6.12. Výsledky klasifikace jsou uvedeny v tab. 6.4. Použitý klasifikátor: 10-NN.

6.3. Výsledky automatického hodnocení prodloužené fonace



Obrázek 6.12: Boxploty pro hodnoty parametrů použitých pro hodnocení diskretního B u prodloužené fonace; (a) $CEPSF0_{CL}$, (b) CPP_{CL} , (c) GNE_{μ} a (d) HNR_M .

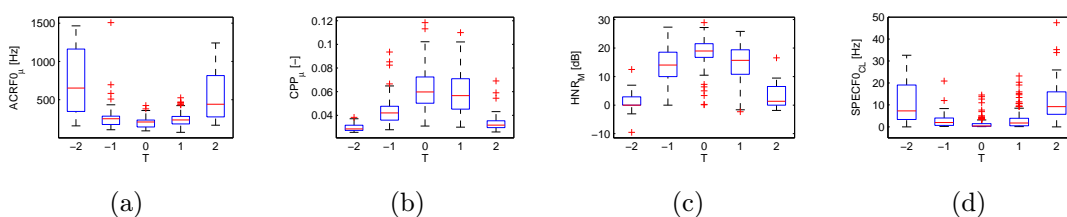
Tabulka 6.4: Matice záměn pro automatické objektivní hodnocení diskretního B u prodloužené fonace.

		B			
		objektivní			
		0	1	2	3
subjektivní	0	222	19	0	0
	1	43	72	9	0
	2	6	19	31	12
	3	0	2	9	25

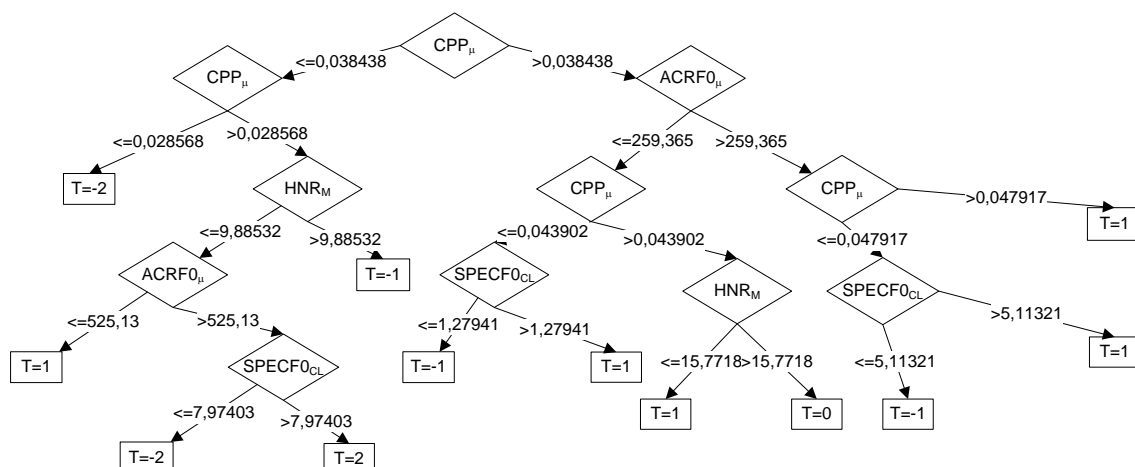
přesnost: 75 %
Cohnenovo kappa [-]: 0,59

Parametr T

Po redukcí příznakového prostoru byly použity parametry: $ACRF0_{\mu}$, CPP_{μ} , HNR_M a $SPECF0_{CL}$. Rozdělení hodnot parametrů pro jednotlivé stupně hodnocení je na obr. 6.13. Výsledky klasifikace jsou uvedeny v tab. 6.5. Použitý klasifikátor: J48 (viz obr. 6.14).



Obrázek 6.13: Boxploty pro hodnoty parametrů použitých pro hodnocení diskretního T u prodloužené fonace; (a) $ACRF0_{\mu}$, (b) CPP_{μ} , (c) HNR_M a (d) $SPECF0_{CL}$.



Obrázek 6.14: Schéma rozhodovacího stromu pro klasifikaci diskretního parametru T pro prodlouženou fonaci.

Tabulka 6.5: Matice záměn pro automatické objektivní hodnocení diskretního T u prodloužené fonace.

		T objektivní				
		-2	-1	0	1	2
subjektivní	-2	20	1	0	6	8
	-1	3	22	17	25	3
	0	1	16	136	27	1
	1	1	18	75	55	6
	2	5	2	0	11	10

přesnost: 52 %
Cohenovo kappa: 0,31

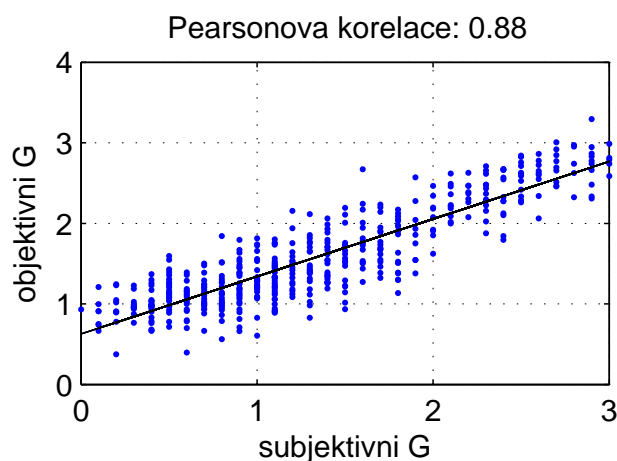
6.3.2 Spojité hodnocení

Parametr G

Po redukci příznakového prostoru byly použity parametry: BRI_{α} , $corelF0cepsAcr$, CPP_{μ} a HNR_{μ} . Výsledky klasifikace jsou uvedeny na obr. 6.15. Použitý klasifikátor: M5P (výsledný algoritmus viz Alg. 4).

Algoritmus 4 Objektivní klasifikace G u prodloužené fonace hlásky /a/.

$$G = 0.2334 * BRI_{\alpha} - 0.3225 * corelF0cepsAcr - 4.4607 * CPP_{\mu} - 0.0707 * HNR_{\mu} + 2.7834$$



Obrázek 6.15: Ukázka klasifikace prodloužené fonace, parametr G.

Parametr R

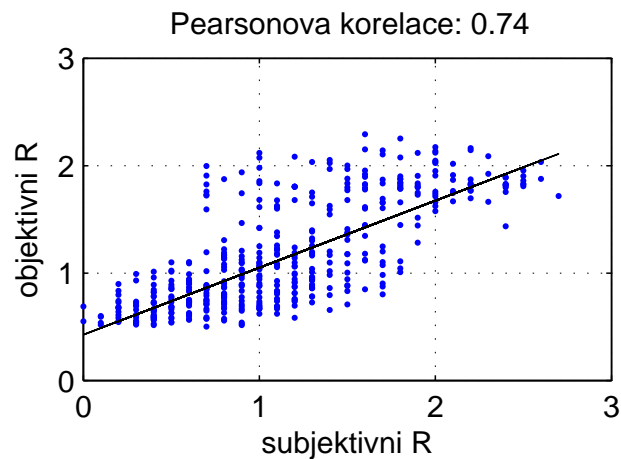
Po redukci příznakového prostoru byly použity parametry: $ACRF0_{\sigma_R}$, $CEPSF0_{\sigma_R}$, RPK_{μ} a $SPS_{LRE_{max}}$. Výsledky klasifikace jsou uvedeny na obr. 6.16. Použitý klasifikátor: M5P (výsledný algoritmus viz Alg. 5).

Algoritmus 5 Objektivní klasifikace R u prodloužené fonace hlásky /a/.

```

if  $RPK_{\mu} \leq 0,696$  then
  if  $RPK_{\mu} \leq 0,608$  then
     $R = 0.0359 * ACRF0_{\sigma_R} + 0.2439 * CEPSF0_{\sigma_R} - 0.0761 * RPK_{\mu} + 0.9599$ 
  *  $SPS_{LRE_{max}} + 1.5064$ 
  else
     $R = 0.2839 * ACRF0_{\sigma_R} + 0.3838 * CEPSF0_{\sigma_R} - 0.0761 * RPK_{\mu} + 0.3549$ 
  *  $SPS_{LRE_{max}} + 1.2012$ 
  end if
else
   $R = 0.0072 * ACRF0_{\sigma_R} + 1.8081 * CEPSF0_{\sigma_R} - 2.5726 * RPK_{\mu} - 0.6357 *$ 
  *  $SPS_{LRE_{max}} + 2.9139$ 
end if

```



Obrázek 6.16: Ukázka klasifikace prodloužené fonace, parametr R.

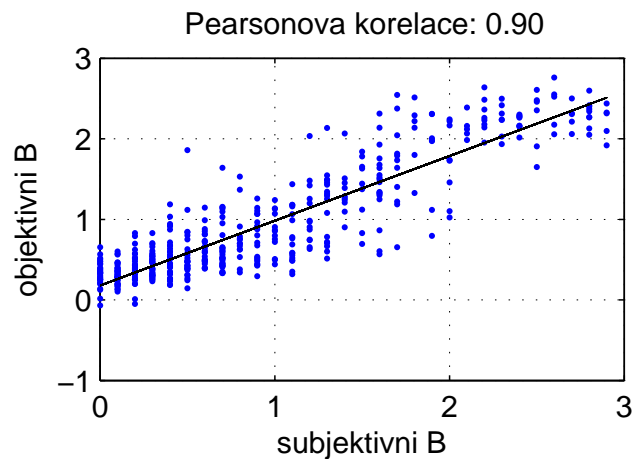
Parametr B

Po redukcí příznakového prostoru byly použity parametry: BRI_{α} , $CEPSF0_{CL}$, HLR_{μ} a $SR50_{\sigma_R}$. Výsledky klasifikace jsou uvedeny na obr. 6.17. Použitý klasifikátor: M5P (výsledný algoritmus viz Alg. 6).

Algoritmus 6 Objektivní klasifikace B u prodloužené fonace hlásky /a/.

```

if  $CEPSF0_{CL} \leq 19,654$  then
     $B = 0.1747 * BRI_{\alpha} + 0.0354 * CEPSF0_{CL} + 0.025 * HLR_{\mu} + 0.8804 * SR50_{\sigma_R}$ 
     $+ 1.0084$ 
else
    if  $CEPSF0_{CL} \leq 32,258$  then
         $B = 0.2744 * BRI_{\alpha} + 0.0063 * CEPSF0_{CL} + 0.0312 * HLR_{\mu} + 0.0666 * SR50_{\sigma_R}$ 
         $+ 2.074$ 
    else
         $B = 0.058 * BRI_{\alpha} + 0.0122 * CEPSF0_{CL} + 0.021 * HLR_{\mu} + 0.0666 * SR50_{\sigma_R}$ 
         $+ 1.9923$ 
    end if
end if
    
```



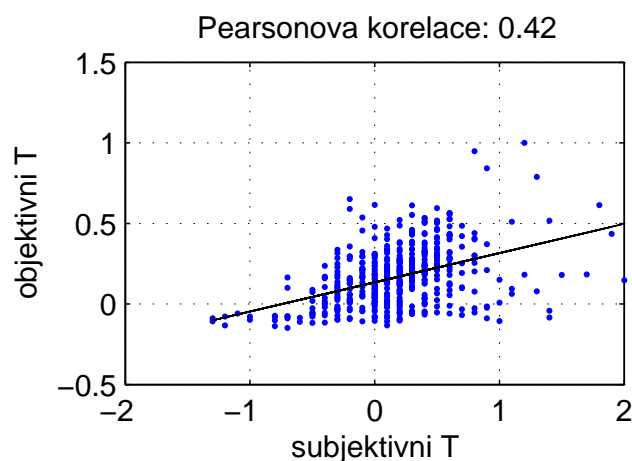
Obrázek 6.17: Ukázka klasifikace prodloužené fonace, parametr B.

Parametr T

Po redukci příznakového prostoru byly použity parametry³: $CPP_{\delta_{\max}}$, $RPK_{\delta_{\text{range}}}$. Výsledky klasifikace jsou uvedeny na obr. 6.18. Použitý klasifikátor: lineární regrese (výsledný algoritmus viz Alg. 7).

Algoritmus 7 Objektivní klasifikace T u prodloužené fonace hlásky /a/.

$$T = 30.0398 * CPP_{\delta_{\max}} + 64.3327 * RPK_{\delta_{\text{range}}} - 0.267$$



Obrázek 6.18: Ukázka klasifikace prodloužené fonace, parametr T.

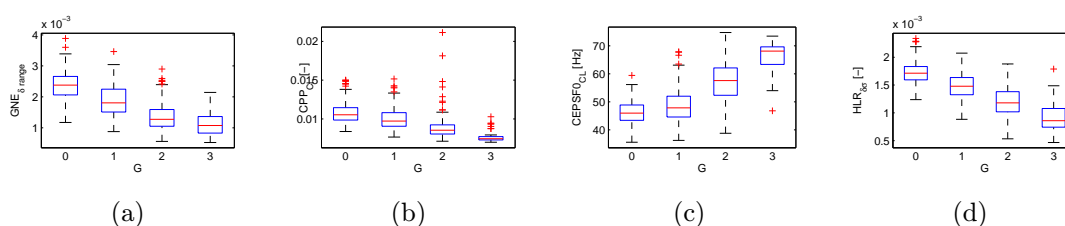
³Prvním krokem redukce příznakového prostoru víceparametrů neprošlo.

6.4 Výsledky automatického hodnocení čteného textu

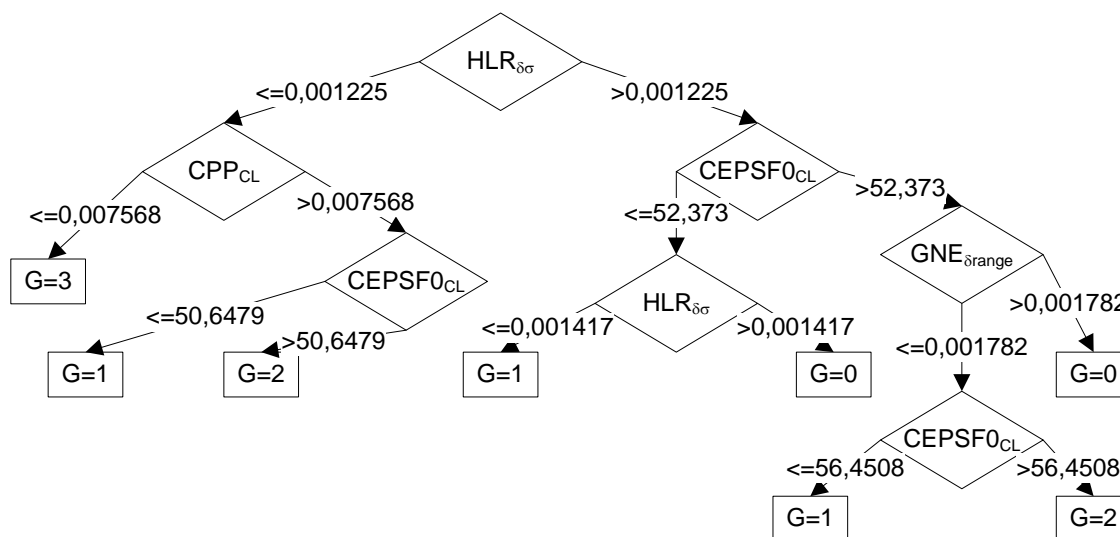
6.4.1 Diskrétní hodnocení

Parametr G

Po redukcí příznakového prostoru byly použity parametry: $GNE_{\delta_{range}}$, CPP_{CL} , $CEPSF0_{CL}$ a $HLR_{\delta\sigma}$. Rozdělení hodnot parametrů pro jednotlivé stupně hodnocení je na obr. 6.19. Výsledky klasifikace jsou uvedeny v tab. 6.6. Použitý klasifikátor: J48 (schéma viz obr. 6.20).



Obrázek 6.19: Boxploty pro hodnoty parametrů použitých pro hodnocení diskretního G u čteného textu; (a) $GNE_{\delta_{range}}$, (b) CPP_{CL} , (c) $CEPSF0_{CL}$ a (d) $HLR_{\delta\sigma}$.



Obrázek 6.20: Schéma rozhodovacího stromu pro klasifikaci diskretního parametru G pro čtený text.

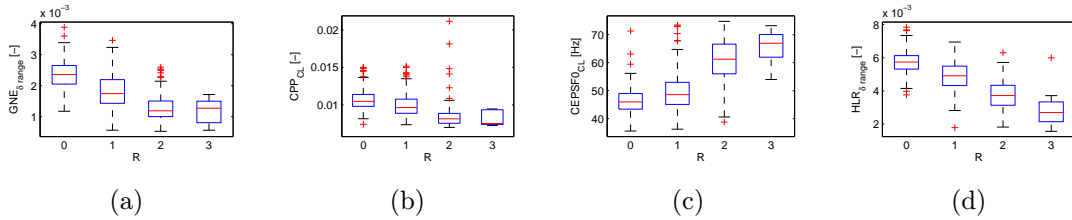
Tabulka 6.6: Matice záměn pro automatické objektivní hodnocení diskrétního G u čteného textu.

		G			
		objektivní			
		0	1	2	3
subjektivní	0	232	23	0	0
	1	81	63	17	1
	2	12	35	74	10
	3	0	2	16	27

přesnost: 67 %
Cohnenovo kappa [-]: 0,50

Parametr R

Po redukcí příznakového prostoru byly použity parametry: $GNE_{\delta_{range}}$, CPP_{CL} , $CEPSF0_{CL}$ a $HLR_{\delta_{range}}$. Rozdělení hodnot parametrů pro jednotlivé stupně hodnocení je na obr. 6.21. Výsledky klasifikace jsou uvedeny v tab. 6.7. Použitý klasifikátor: bayesovské rozhodování, hodnoty použité pro klasifikaci viz tab. 7.2, parametry R, čtený text, $\mu \pm \sigma$ pro jednotlivé stupně hodnocení.

Obrázek 6.21: Boxploty pro hodnoty parametrů použitých pro hodnocení diskrétního R u čteného textu; (a) $GNE_{\delta_{range}}$, (b) CPP_{CL} , (c) $CEPSF0_{CL}$ a (d) $HLR_{\delta_{range}}$.

Parametr B

Po redukcí příznakového prostoru byly použity parametry: $GNE_{voiced_{\mu}}$, $CEPSF0_{CL}$, $HLR_{\delta_{\sigma}}$ a VUR . Rozdělení hodnot parametrů pro jednotlivé stupně hodnocení je na obr. 6.22. Výsledky klasifikace jsou uvedeny v tab. 6.8. Použitý klasifikátor: bayesovské rozhodování, hodnoty použité pro klasifikaci viz tab. 7.2, parametry B, čtený text, $\mu \pm \sigma$ pro jednotlivé stupně hodnocení.

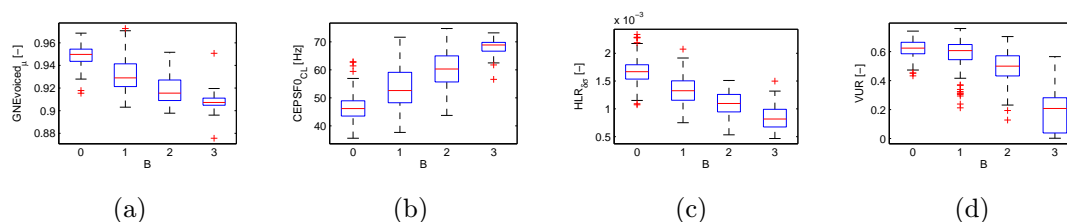
Parametr T

Po redukcí příznakového prostoru byly použity parametry: $GNE_{\delta_{range}}$, $GNE_{voiced_{\sigma}}$, $CEPSF0_{CL}$ a $HLR_{\delta_{\sigma}}$. Rozdělení hodnot parametrů pro jednotlivé stupně hodnocení

Tabulka 6.7: Matice záměn pro automatické objektivní hodnocení diskretního R u čteného textu.

		R			
		objektivní			
		0	1	2	3
subjektivní	0	213	39	2	0
	1	80	84	38	0
	2	5	20	101	2
	3	0	1	7	1

přesnost: 67 %
Cohnenovo kappa [-]: 0,50


 Obrázek 6.22: Boxploty pro hodnoty parametrů použitých pro hodnocení diskretního B u čteného textu; (a) GNEvoiced_μ, (b) CEPSPF0_{CL}, (c) HLR_{δσ} a (d) VUR.

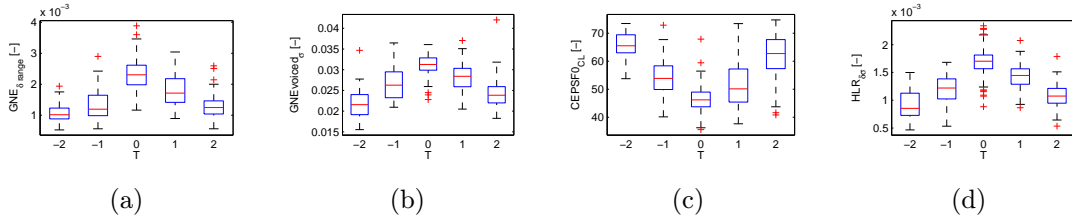
Tabulka 6.8: Matice záměn pro automatické objektivní hodnocení diskretního B u čteného textu.

		B			
		objektivní			
		0	1	2	3
subjektivní	0	322	24	0	0
	1	47	73	25	4
	2	3	20	35	10
	3	0	0	6	24

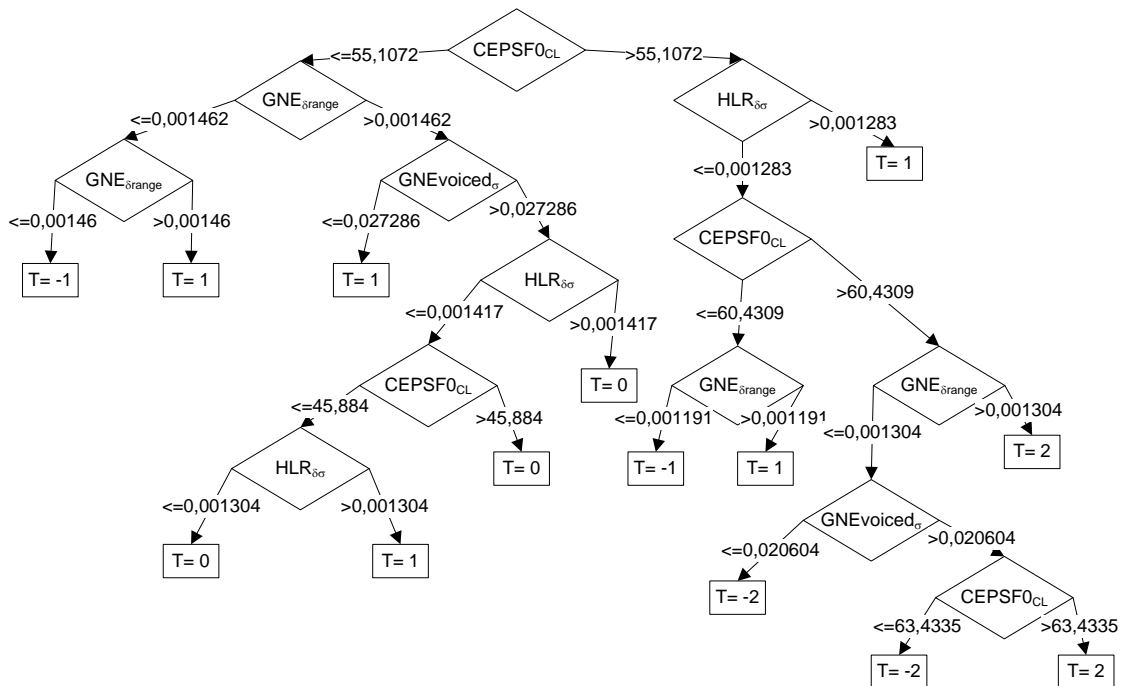
přesnost: 77 %
Cohnenovo kappa [-]: 0,59

je na obr. 6.23. Výsledky klasifikace jsou uvedeny v tab. 6.9. Použitý klasifikátor: J48 (viz obr. 6.24).

6.4. Výsledky automatického hodnocení čteného textu



Obrázek 6.23: Boxploty pro hodnoty parametrů použitých pro hodnocení diskretního T u čteného textu; (a) $GNE_{\delta_{range}}$, (b) $GNEvoiced_{\sigma}$, (c) $CEPSF0_{CL}$ a (d) $HLR_{\delta_{\sigma}}$.



Obrázek 6.24: Schéma rozhodovacího stromu pro klasifikaci diskretního parametru T pro čtený text.

Tabulka 6.9: Matice záměn pro automatické objektivní hodnocení diskrétního T u čteného textu.

		T				
		objektivní				
		-2	-1	0	1	2
subjektivní	-2	12	2	1	2	12
	-1	3	29	17	16	8
	0	0	6	281	21	1
	1	1	18	57	43	7
	2	13	4	3	16	20

přesnost: 65 %
 Cohnenovo kapa: 0,44

6.4.2 Spojité hodnocení

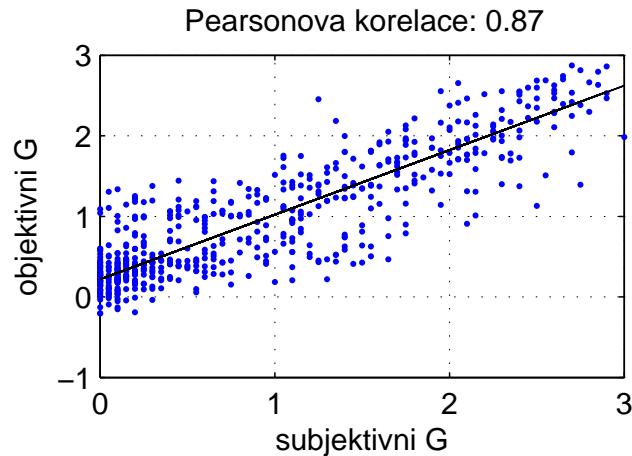
Parametr G

Po redukci příznakového prostoru byly použity parametry: $GNEvoiced_{\sigma}$, $CEPSF0_{CL}$, $HLR_{\delta_{range}}$ a $BRI_{\delta_{min}}$. Výsledky klasifikace jsou uvedeny na obr. 6.25. Použitý klasifikátor: M5Rules (výsledný algoritmus viz Alg. 8).

Algoritmus 8 Objektivní klasifikace G u čteného textu.

```

if  $CEPSF0_{CL} \leq 55,201$  AND  $HLR_{\delta_{range}} > 0,005$  then
     $G = -61.7256 * GNEvoiced_{\sigma} + 0.0126 * CEPSF0_{CL} - 234.7345 * HLR_{\delta_{range}} -$ 
     $0.0128 * BRI_{\delta_{min}} + 3.0427$ 
else
     $G = 0.0382 * CEPSF0_{CL} - 250.5227 * HLR_{\delta_{range}} - 1.1708 * BRI_{\delta_{min}} + 0.8266$ 
end if
    
```



Obrázek 6.25: Ukázka klasifikace čteného textu, parametr G.

Parametr R

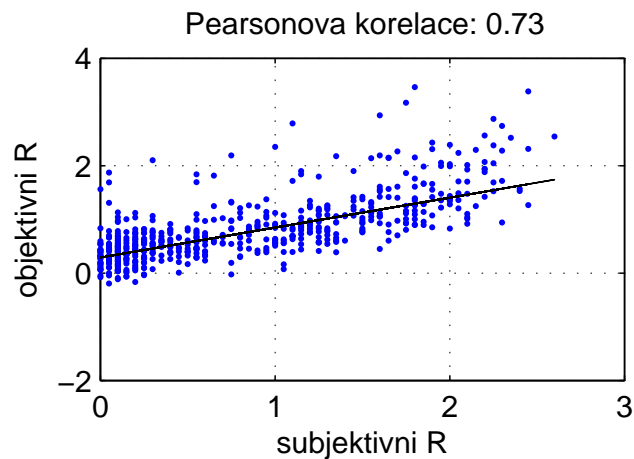
Po redukci příznakového prostoru byly použity parametry: $BRI_{\delta_{range}}$, $BRI_{\delta_{\sigma}}$, $RPK_{\delta_{range}}$ a HNR_M . Výsledky klasifikace jsou uvedeny na obr. 6.26. Použitý klasifikátor: M5P (výsledný algoritmus viz Alg. 9).

Algoritmus 9 Objektivní klasifikace R u čteného textu.

```

if  $HNR_M \leq 34.536$  then
  if  $BRI_{\delta_{\sigma}} \leq 0$  then  $R = -5566.1243 * BRI_{\delta_{range}} - 5661.6031 * BRI_{\delta_{\sigma}} - 2516.2436$ 
   $* RPK_{\delta_{range}} - 0.0023 * HNR_M + 2.4906$ 
  else
    if  $RPK_{\delta_{range}} \leq 0$  then  $R = 1195.6585 * BRI_{\delta_{range}} - 5661.6031 * BRI_{\delta_{\sigma}} -$ 
     $3577.4745 * RPK_{\delta_{range}} - 0.0023 * HNR_M + 1.7507$ 
    else  $R = 4275.5048 * BRI_{\delta_{range}} - 5661.6031 * BRI_{\delta_{\sigma}} - 13003.735 * RPK_{\delta_{range}}$ 
     $- 0.0537 * HNR_M + 3.6119$ 
    end if
  end if
else  $R = 726.7023 * BRI_{\delta_{range}} + 1313.6252 * BRI_{\delta_{\sigma}} - 18137.34 * RPK_{\delta_{range}} - 0.0331$ 
 $* HNR_M + 3.5877$ 
end if

```



Obrázek 6.26: Ukázka klasifikace čteného textu, parametr R.

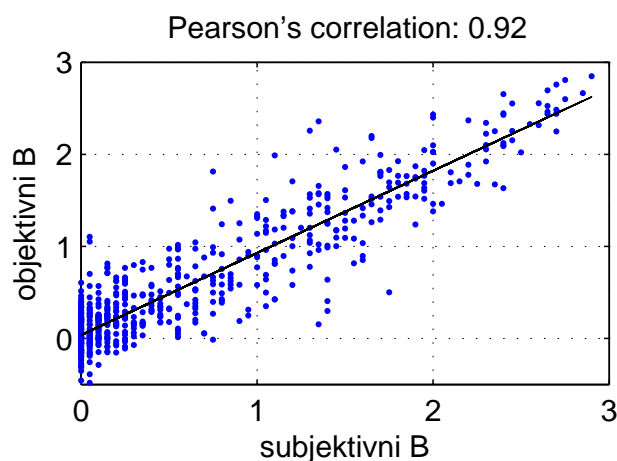
Parametr B

Po redukcí příznakového prostoru byly použity parametry: $GNEvoiced_{\mu}$, $CEPSF0_{CL}$, $HLR_{\delta_{\sigma}}$ a VUR . Výsledky klasifikace jsou uvedeny na obr. 6.27. Použitý klasifikátor: M5P (výsledný algoritmus viz Alg. 10).

Algoritmus 10 Objektivní klasifikace B u čteného textu.

```

if  $GNEvoiced_{\mu} \leq 0,933$  then
     $B = -0.8816 * GNEvoiced_{\mu} + 0.0349 * CEPSF0_{CL} - 705.9942 * HLR_{\delta_{\sigma}} - 0.9973$ 
     $* VUR + 1.5818$ 
else
    if  $HLR_{\delta_{\sigma}} \leq 0,002$  then  $B = -5.1104 * GNEvoiced_{\mu} + 0.02 * CEPSF0_{CL} -$ 
     $1138.2618 * HLR_{\delta_{\sigma}} - 0.0239 * VUR + 5.9901$ 
    else
        if  $GNEvoiced_{\mu} \leq 0,949$  then  $B = -0.6772 * GNEvoiced_{\mu} + 0.0147 *$ 
         $CEPSF0_{CL} - 102.2193 * HLR_{\delta_{\sigma}} + 1.2642 * VUR - 0.5171$ 
        else  $B = -0.6772 * GNEvoiced_{\mu} + 0.008 * CEPSF0_{CL} - 228.7807 * HLR_{\delta_{\sigma}}$ 
         $+ 0.0632 * VUR + 0.7364$ 
        end if
    end if
end if
end if
    
```



Obrázek 6.27: Ukázka klasifikace čteného textu, parametr B.

Parametr T

Po redukcí příznakového prostoru byly použity parametry⁴: $RPKvoiced_{\sigma}$, $SWvoiced_{\sigma_R}$ a $SDvoiced_{\sigma_R}$. Výsledky klasifikace jsou uvedeny na obr. 6.28. Použitý klasifikátor: M5P (výsledný algoritmus viz Alg. 11).

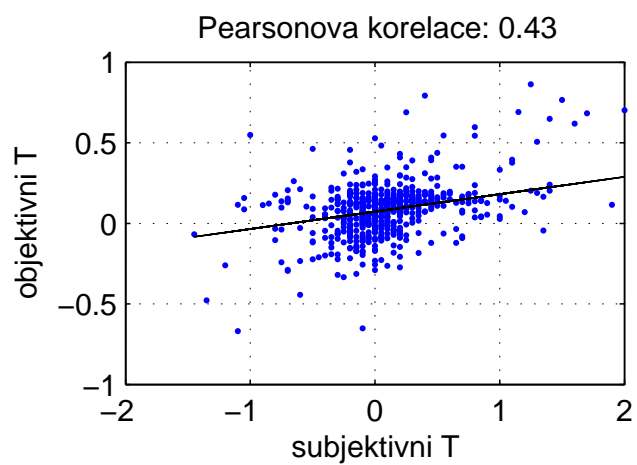
Algoritmus 11 Objektivní klasifikace T u čteného textu.

```

if  $SWvoiced_{\sigma_R} \leq 0,032$  then
     $T = 4.2953 * RPKvoiced_{\sigma} + 1.7248 * SWvoiced_{\sigma_R} + 66.7772 * SDvoiced_{\sigma_R} - 1.37$ 
else
    if  $SDvoiced_{\sigma_R} \leq 0,011$  then
         $T = 13.1632 * RPKvoiced_{\sigma} + 44.8967 * SWvoiced_{\sigma_R} - 284.4961 * SDvoiced_{\sigma_R} - 0.1327$ 
    else
         $T = 3.4491 * RPKvoiced_{\sigma} + 2.1942 * SWvoiced_{\sigma_R} - 5.6363 * SDvoiced_{\sigma_R} - 0.3744$ 
    end if
end if

```

⁴Prvním krokem redukce příznakového prostoru víceparametrů neprošlo.



Obrázek 6.28: Ukázka klasifikace čteného textu, parametr T.

7. Diskuze

7.1 Databáze promluv

Použitá databáze chraptivých promluv se svým rozsahem (469 nahrávek prodloužené fonace /a/ a 593 nahrávek čteného textu) řadí k větším databázím, viz část 2.2. Žádná z prací uvedená v části 2.2 také nepopisuje opakované hodnocení hodnotitelů.

Podle výsledků hodnocení se nepotvrdily závěry uvedené v De Bodt et al. [29] a Karnell et al. [72], kde popisují, že při opakovaném hodnocení jsou nahrávky hodnoceny vyšším stupněm patologie. K testování byl použit neparametrický Mann-Whitneyho U test, protože hodnocení nemají normální rozdělení (Lillieforsův test na 5% hladině významnosti). Výsledky testu jsou následující: jednotlivci jsou většinou ve svém hodnocení konzistentní, nejsou významně přísnější ani méně přísní. Hodnocení skupiny je v případě prodloužené fonace hlásky /a/ méně přísné u parametrů G, B a T, pro parametr R nedošlo k významné změně hodnocení. Naopak v případě hodnocení čteného textu je hodnocení skupiny přísnější u parametrů G, R a B, u parametru T nedošlo k významné změně hodnocení. Výsledky jsou shrnuty v tab. 5.5.

Míru shody skupiny hodnotitelů v opakovaném testu a shodu hodnocení skupiny je možné porovnat s výsledky uvedenými v De Bodt et al. [29]. V uvedené práci testovali vliv předchozích zkušeností se subjektivním hodnocením. K dispozici měli celkem 23 hodnotitelů s různou mírou zkušeností a 12 nahrávek, které obsahovaly jak prodlouženou fonaci hlásky /a/, tak čtený text. K hodnocení míry shody skupiny hodnotitelů při opakovaném testu použili Cohenovo kappa. Jejich skupina hodnotitelů dosáhla při opakovaném hodnocení lepšího skóre u parametrů G ($\kappa_C = 0,60$ vs. $\kappa_C = 0,41$ pro /a/, resp. $\kappa_C = 0,42$ pro čtený text). Pro parametry R je výsledek nejednoznačný ($\kappa_C = 0,35$ vs. $\kappa_C = 0,33$ pro /a/, resp. $\kappa_C = 0,40$ pro čtený text) a pro parametr B dosahují naši experti lepších výsledků ($\kappa_C = 0,38$ vs. $\kappa_C = 0,43$ pro /a/ i čtený text).

Kontingenční tabulky shrnující hodnocení našich expertů v prvním a ve druhém kole hodnocení pro obě databáze jsou uvedeny v příloze 11.

7.2 Automatické hodnocení chraptivosti

Byly provedeny experimenty s automatickým hodnocením chraptivosti na dvou databázích: na databázi obsahující pouze prodlouženou fonaci hlásky /a/ a na databázi obsahující standardizovaný čtený text *Podzim na Starém bělidle* z povídky Babička autorky Boženy Němcové. Obě databáze byly nezávisle na sobě dvakrát ohodnoceny skupinou pěti odborníků z Foniatrické kliniky 1. LF UK a VFN v Praze. Opakované hodnocení proběhlo s alespoň dvoutýdenním odstupem, aby bylo zamezeno tzv. paměťovému efektu. K subjektivnímu hodnocení byla použita metoda GRBT, která sloužila jako vstup při učení automatických klasifikátorů.

7.2.1 Porovnání dosažených výsledků s publikovanými studii

V tab. 7.1, jsou uvedeny studie zabývající se automatickým hodnocením chraptivosti a výsledky této práce.

Tabulka 7.1: Srovnání výsledků automatické klasifikace s publikovanými výsledky. /a/ – prodloužená fonace; CS – spojitá řeč; MPT – délka fonace prodloužené hlásky /a/; PIP – špičkový vnitroústní tlak přechodu souhláska-samohláska /pi/; JIT – jitter, VRPA – plocha hlasového pole; MFCC – Mel-frekvenční keprální koeficienty; LPC chyba – chybový signál lineární predikce; ostatní zkratky parametrů viz definice v této práci (kapitola 4 a tab. 6.1). Značka s označuje spojitě hodnocení, značka d diskrétní. Hodnoty přesnosti označené * jsou udávány pouze jako Pearsonův lineární korelační koeficient.

zdroj	databáze promluv nahrávek obsah	použité parametry	GRBAS	přesnost
Alpan et al. [6]	251 /a/, CS	GVA, CPP	G_d	0,79*
Fredouille et al. [42]	80 CS	MFCC	G_d	79 %
Lee and Hahn [77]	85 /a/	LPC chyba	G_d	93 %
Ma and Yiu [80]	112 CS	MPT, PIP, JIT, VRPA	G_d	67 %
			G_d	68 %
			R_d	63 %
Saenz-Lechon et al. [126]	648 /a/, CS	MFCC	B_d	57 %
			A_d	55 %
			S_d	63 %
		CEPSF0 _{CL} , CPP _M , HNR _{μ} , RPK _{μ}	G_d	66 %
		HNR _{μ} , mL ₃₄ _{μ} , RPK _{μ} , SPS _M	R_d	64 %
		CEPSF0 _{CL} , CPP _{CL} , GNE _{μ} , HNR _M	B_d	75 %
	469 /a/	ACRF0 _{μ} , CPP _{μ} , HNR _M , SPEC _{F0} _{CL}	T_d	52 %
		BRI _{α} , corel _{F0} cepsAcr, CPP _{μ} , HNR _{μ}	G_s	0,88*
		ACRF0 _{σ_R} , CEPSF0 _{σ_R} , RPK _{μ} , SPS _{LREmax}	R_s	0,74*
		BRI _{α} , CEPSF0 _{CL} , HLR _{μ} , SR50 _{σ_R}	B_s	0,90*
tato práce		CPP _{δ_{max}} , RPK _{δ_{range}}	T_s	0,42*
		GNE _{δ_{range}} , CPP _{CL} , CEPSF0 _{CL} , HLR _{$\delta\sigma$}	G_d	67 %
		GNE _{δ_{range}} , CPP _{CL} , CEPSF0 _{CL} , HLR _{δ_{range}}	R_d	67 %
		GNEvoiced _{μ} , CEPSF0 _{CL} , HLR _{$\delta\sigma$} , VUR	B_d	77 %
	593 CS	GNE _{δ_{range}} , GNEvoiced _{σ} , CEPSF0 _{CL} , HLR _{$\delta\sigma$}	T_d	65 %
		GNEvoiced _{σ} , CEPSF0 _{CL} , HLR _{δ_{range}} , BRI _{δ_{min}}	G_s	0,87*
		BRI _{δ_{range}} , BRI _{$\delta\sigma$} , RPK _{δ_{range}} , HNR _M	R_s	0,73*
		GNEvoiced _{μ} , CEPSF0 _{CL} , HLR _{$\delta\sigma$} , VUR	B_s	0,92*
		RPKvoiced _{σ} , SWvoiced _{σ_R} , SDvoiced _{σ_R}	T_s	0,43*

Přímé porovnání výsledků je poměrně komplikovaná záležitost. Všechny práce, které se zabývají automatickým hodnocením chraptivosti (viz tab. 7.1) hodnotí pouze celkový dojem z hlasu G. Pouze práce Saenz-Lechon et al. [126] hodnotí i ostatní parametry stupnice GRBAS. Zároveň každá z těchto studií používá jinou databázi nahrávek s jinou distribucí jednotlivých stupňů chraptivosti. Při přímém

porovnání dvou studií, které používají MFCC koeficienty, si lze všimnout naprosto odlišných hodnot přesnosti klasifikace: Fredouille et al. [42] dosahuje přesnosti 79 %, zatímco Saenz-Lechon et al. [126] „pouze“ 67 %, přičemž obě práce používají nahrávky čteného textu. To může být jednak způsobeno rozdílnou velikostí databází nahrávek (80 nahrávek vs 648 nahrávek) a jednak rozdílnou distribucí jednotlivých stupňů chraptivosti. V práci Fredouille et al. [42] je v databázi přesně 20 nahrávek na každý stupeň (tzn. 20 nahrávek pro $G=0$, 20 nahrávek pro $G=1, \dots$), databáze použitá v práci Saenz-Lechon et al. [126] obsahuje 433 nahrávek od zdravých lidí a 215 nahrávek od chraptivých. Výsledná distribuce hodnocení tak není uniformní jako v práci Fredouille et al. [42], ale blíží se distribuci hodnot v běžné populaci, tzn. podobně jako v této práci, viz histogramy hodnocení na obr. 6.6 a obr. 6.7.

Práce Ma and Yiu [80], která dosahuje srovnatelných výsledků klasifikace celkového dojmu z hlasu G jako tato práce, také používá databázi, ve které jsou jednotlivé stupně celkového dojmu zastoupeny rovnoměrně (cca 20 hlasů na každý stupeň). V uvedené práci je použita originální metodika hodnocení. Hodnotitelé byli nuceni hodnotit trénovací nahrávky tak dlouho, dokud nedosáhli vzájemné shody hodnocení alespoň 80 % a až poté hodnotili testovací nahrávky. Nicméně z prezentovaných výsledků je patrné, že tato metodika nemá na výsledné automatické hodnocení vliv.

Při porovnání ostatních parametrů, tzn. chraplavost R a dyšnost B , se studií Saenz-Lechon et al. [126] dosahuje tato práce srovnatelných klasifikačních výsledků (63 % vs 64 % pro prodlouženou fonaci resp. 67 % pro čtený text) pro parametr R a výrazně lepších výsledků pro parametr B (57 % vs 75 % pro prodlouženou fonaci resp. 77 % pro čtený text).

7.2.2 Diskuze použitých akustických parametrů

Z tab. 7.1 je možné odečíst parametry použité v této práci při automatickém hodnocení chraptivosti. Detailnější přehled vlastností těchto parametrů je uveden v tab. 7.2 pro diskrétní hodnocení a v tab. 7.3 pro spojitě hodnocení.

Tab. 7.2 obsahuje výsledky neparametrické verze testu ANOVA – Kruskal-Wallisova testu, schopnost parametru rozlišovat jednotlivé stupně daného parametru (*post-hoc* Bonferroniho korekce) a střední hodnoty a směrodatné odchylky pro jednotlivé stupně hodnocení. Pro spojitě hodnocení jsou v tab. 7.3 uvedeny míry korelace daného parametru se subjektivním hodnocením.

Za povšimnutí stojí časté využití parametru délky křivky odhadu hlasivkové frekvence, ať už v časové (ACRF0), spektrální (SPECF0) nebo kepsrální (CEPSF0) oblasti, což odpovídá předpokladu, že odhad hlasivkové frekvence je snažší u zdravého hlasu. Čím více je hlas vnímán jako patologický, tím větší je délka křivky: viz tab. 7.2, kde se střední hodnota těchto parametrů zvyšuje společně s rostoucím stupněm patologie a v tab. 7.3 je možné odečíst pozitivní hodnoty korelačního koeficientu.

Dalším často zastoupeným parametrem je Cepstral Peak Prominence (CPP). To je v souladu s předpokladem, protože tento parametr je hojně používán a dosahuje dobrých výsledků ([7], [12], [58], [61], [99], [104], [147]). Za povšimnutí stojí uplatnění jak popisu distribuce (μ , M), tak popisu časového průběhu (CL, δ_{\max}).

Parametr GNE je dalším často používaným parametrem ([44], [49], [95], [97], [130]). Výše zmíněné studie ovšem analyzovaly pouze prodlouženou fonaci nebo zpěv (Mecke and Sundberg [95]). V této práci byl parametr GNE použit jak na prodlouženou fonaci, tak na čtený text, který obsahuje znělé i neznělé úseky řeči. Navíc kromě řeči obsahuje také pauzy. Z toho důvodu byl parametr určován jak v celé nahrávce, tak pouze ve znělých úsecích řeči. Oba přístupy byly ve finálních klasifikátorech využity.

Zajímavostí je poměrně časté zastoupení parametrů BRI a HLR, které nejsou v literatuře příliš často zmiňovány (Hillenbrand and Houde [61]). Výhodou těchto parametrů je nenáročnost jejich výpočtu a jejich absolutní nezávislost na odhadu hlasivkové frekvence.

Při diskrétním hodnocení, kdy jsou k dispozici výsledky Kruskal-Wallisova testu a *post-hoc* Bonferroniho korekce pro určení schopnosti parametrů rozlišit jednotlivé stupně ukazatelů GRBT, je vhodné si všimnout jevu, kdy schopnost parametrů odlišit dva nejtěžší stupně patologie parametrů G, R, B, tzn. sloupec „2vs3“ v tab. 7.2, je značně menší než u ostatních stupňů. To může být způsobeno buď faktem, že nejtěžším stupněm patologie bylo ohodnoceno málo hlasů, nebo tím, že výsledný hlasový signál je narušený do takové míry, že z pohledu hodnot parametrů je velmi podobný. Oproti tomu rozlišení hodnot „1vs2“ je možné i na nejvyšší hladině významnosti ve všech případech. Rozlišení hlasů bez patologie a s lehkou patologií (sloupec „0vs1“) je ve většině případů možné také na nejvyšší hladině významnosti. Nikoliv však ve všech, což může být způsobeno tím, že stupněm 1 jsou hodnoceny i velmi málo patologické hlasy.

Samostatnou kapitolu tvoří hodnocení parametru T. Na obr. 6.13 a obr. 6.23 je možné si všimnout, že mediány a horní a dolní kvartily hodnot jsou pro hodnoty +1 a -1, resp. +2 a -2 velmi podobné. To může být způsobeno následujícím: zavedení parametru T mělo usnadnit hodnotitelům subjektivní hodnocení sloučením dvou parametrů (A – hlasová slabost a S – hlasové napětí) do jednoho parametru hlasové napětí (Vokřál [143]). V tomto parametru hodnotitelé hodnotí hlasovou slabost zápornými známkami a přemáhání hlasu známkami kladnými. Nemohou tudíž popsat hlas, který je zároveň slabý a přemáhaný, což je v praxi možné. To je patrné i z kontingenčních tabulek uvedených v tab. 11.1 a tab. 11.7 v příloze 11. Z nich je možné odečíst, že zatímco v prvním kole hodnocení byla nahrávka často hodnocena známkou +2, ve druhém kole byla často hodnocena známkou -2 a naopak. Tato nejednoznačnost se poté promítne do automatického hodnocení. Nicméně již v průběhu subjektivního hodnocení prodloužené fonace si skupina odborníků stěžovala, že pouze z prodloužené fonace je složité, až téměř nemožné, určit, zda je hlas přemáhaný nebo slabý. Při subjektivním hodnocení čteného textu jsou výsledky „lepší“, to je ovšem způsobeno tím, že valná většina hlasů byla hodnocena jako normální (tzn. ani jako přemáhaná, ani jako slabá).

Tabulka 7.2: Parametrizace použité k diskretnímu hodnocení chraptivosti. Výsledky neparametrické verze testu ANOVA – Kruskal-Wallisova testu. χ^2 je síla testu (pro prodlouženou fonaci, parametry G, R, B je $\chi^2(3, 468)$, pro parametr T je $\chi^2(4, 468)$); pro čtený text pro parametry G, R, B je $\chi^2(3, 592)$, pro parametr T je $\chi^2(4, 592)$; pro všechny testy $p < 0, 001$. -2vs-1, -1vs0, 0vs1, 1vs2 a 2vs3 je schopnost parametru rozlišit jednotlivá hodnocení (*post-hoc* Bonferroniho korekce), uvedené hodnoty jsou hladiny významnosti (NS - nesignifikantní, $p > 0, 05$).

parametr	χ^2	Kruskal-Wallis						typické hodnoty pro hodnocení															
		-2vs-1	-1vs0	0vs1	1vs2	2vs3	p<	-2	-1	0	1	2	3	μ	σ	μ	σ	μ	σ				
G; přesnost 66%, $\kappa_C = 0, 47$																							
CEPSF0 _{CL}	253,50	-	0,001	0,001	0,05	-	±	±	2,96E+00	±	3,69E+00	±	8,09E+00	±	8,08E+00	±	2,93E+01	±	1,81E+01	±	4,97E+01	±	1,55E+01
CPP _M	231,03	-	0,01	0,001	0,01	-	±	±	6,76E-02	±	1,70E-02	±	5,84E-02	±	1,80E-02	±	3,91E-02	±	1,05E-02	±	3,03E-02	±	5,98E-03
HNR _{μ}	276,95	-	0,001	0,001	0,01	-	±	±	1,97E+01	±	2,95E+00	±	1,60E+01	±	3,72E+00	±	8,11E+00	±	5,00E+00	±	1,65E+00	±	3,36E+00
RPK _{μ}	244,39	-	0,001	0,001	0,01	-	±	±	8,72E-01	±	6,20E-02	±	8,24E-01	±	8,14E-01	±	6,14E-01	±	1,76E-01	±	3,37E-01	±	1,52E-01
R; přesnost 64%, $\kappa_C = 0, 42$																							
HNR _{μ}	190,48	-	0,001	0,001	NS	-	±	±	1,78E+01	±	5,97E+00	±	1,47E+01	±	4,84E+00	±	6,70E+00	±	5,01E+00	±	2,85E+00	±	3,53E+00
mLSF _{34μ}	121,35	-	0,001	0,001	NS	-	±	±	4,00E+03	±	3,46E+02	±	4,26E+03	±	4,60E+02	±	4,85E+03	±	6,26E+02	±	4,82E+03	±	5,90E+02
RPK _{μ}	177,16	-	0,001	0,001	NS	-	±	±	8,23E-01	±	1,75E-01	±	7,91E-01	±	1,34E-01	±	5,57E-01	±	1,76E-01	±	3,77E-01	±	1,81E-01
SPSM _{μ}	154,24	-	0,01	0,001	NS	-	±	±	8,38E-01	±	1,87E-01	±	8,19-e1	±	1,70E-01	±	5,80E-01	±	2,36E-01	±	4,04E-01	±	2,56E-01
B; přesnost 75%, $\kappa_C = 0, 59$																							
CEPSF0 _{CL}	257,79	-	0,001	0,001	0,001	-	±	±	5,02E+00	±	6,23E+00	±	1,47E+01	±	1,18E+01	±	3,51E+01	±	1,78E+01	±	5,33E+01	±	1,32E+01
CPP _{CL}	269,08	-	0,001	0,001	0,001	-	±	±	1,03E-02	±	4,22E-03	±	6,85E-03	±	2,41E-02	±	4,85E-03	±	1,28E-03	±	3,85E-03	±	5,22E-04
GNE _{μ}	284,67	-	0,001	0,001	0,001	-	±	±	-1,19E+00	±	2,20E+00	±	-1,19E+00	±	1,60E+00	±	-1,06E+00	±	1,16E+00	±	-9,91E+00	±	2,74E-01
HNR _M	225,96	-	0,001	0,001	0,001	-	±	±	1,86E+01	±	4,53E+00	±	1,43E+01	±	4,78E+00	±	7,20E+00	±	5,84E+00	±	7,74E-01	±	4,20E+00
T; přesnost 52%, $\kappa_C = 0, 31$																							
ACRF0 _{μ}	134,08	0,001	0,001	0,001	0,001	-	±	±	7,36E+02	±	1,83E+02	±	2,66E+02	±	5,66E+01	±	2,40E+02	±	8,29E+01	±	5,92E+02	±	3,52E+02
CPP _{μ}	184,47	0,001	0,001	NS	0,001	-	±	±	2,97E-02	±	3,47E-03	±	4,41E-02	±	1,64E-02	±	5,87E-02	±	1,82E-02	±	3,51E-02	±	9,94E-03
HNR _M	186,37	0,001	0,001	0,001	0,001	-	±	±	1,22E+00	±	3,47E+00	±	1,40E+01	±	5,65E+00	±	1,47E+01	±	5,90E+00	±	3,27E+00	±	4,39E+01
SPECF0 _{CL}	120,58	0,05	0,001	0,001	0,001	-	±	±	2,43E+01	±	5,43E+01	±	2,80E+00	±	3,20E+00	±	3,05E+00	±	4,07E+00	±	1,30E+01	±	1,16E+01
G; přesnost 67%, $\kappa_C = 0, 50$																							
GNE _{δ,range}	302,66	-	0,001	0,001	NS	-	±	±	2,39E-03	±	4,56E-04	±	1,88E-03	±	4,77E-04	±	1,39E-03	±	4,52E-04	±	1,12E-03	±	3,48E-04
CPP _{CL}	241,92	-	0,001	0,001	0,001	-	±	±	1,08E-02	±	1,36E-03	±	1,01E-02	±	1,51E-03	±	8,97E-03	±	1,83E-03	±	7,71E-03	±	7,42E-04
CEPSF0 _{CL}	259,38	-	0,01	0,001	0,01	-	±	±	4,61E+01	±	3,92E+00	±	4,86E+01	±	6,13E+00	±	5,73E+01	±	7,58E+00	±	6,65E+01	±	5,25E+00
HLR _{δ,σ}	321,92	-	0,001	0,001	0,05	-	±	±	1,72E-03	±	1,91E-04	±	1,47E-03	±	2,22E-04	±	1,21E-03	±	2,64E-04	±	9,17E-04	±	2,60E-04
R; přesnost 67%, $\kappa_C = 0, 47$																							
GNE _{δ,range}	267,64	-	0,001	0,001	NS	-	±	±	2,37E-03	±	4,63E-04	±	1,82E-03	±	5,32E-04	±	1,30E-03	±	4,30E-04	±	1,16E-03	±	4,20E-04
CPP _{CL}	208,75	-	0,001	0,001	NS	-	±	±	1,07E-02	±	1,34E-03	±	1,00E-02	±	1,59E-03	±	8,56E-03	±	1,85E-03	±	8,10E-03	±	9,70E-04
CEPSF0 _{CL}	226,72	-	0,001	0,001	NS	-	±	±	4,63E+01	±	4,33E+00	±	4,97E+01	±	6,96E+00	±	6,04E+01	±	7,69E+00	±	6,55E+01	±	6,55E+00
HLR _{δ,range}	288,78	-	0,001	0,001	NS	-	±	±	5,74E-03	±	6,70E-04	±	4,84E-03	±	8,44E-04	±	3,74E-03	±	9,21E-04	±	2,97E-03	±	1,32E-03
B; přesnost 77%, $\kappa_C = 0, 59$																							
GNEvoiced _{μ}	307,43	-	0,001	0,001	NS	-	±	±	7,49E-01	±	8,53E-03	±	9,31E-01	±	1,45E-02	±	9,18E-01	±	1,21E-02	±	9,08E-01	±	1,10E-02
CEPSF0 _{CL}	262,21	-	0,001	0,001	NS	-	±	±	4,63E+01	±	4,34E+00	±	5,35E+01	±	7,56E+00	±	6,01E+01	±	7,20E+00	±	8,80E+01	±	3,72E+00
HLR _{δ,σ}	307,24	-	0,001	0,001	NS	-	±	±	1,67E-03	±	2,09E-04	±	1,33E-03	±	2,56E-04	±	1,09E-03	±	2,36E-04	±	8,61E-04	±	2,31E-04
VUR	167,17	-	0,01	0,001	0,05	-	±	±	6,22E-01	±	5,45E-02	±	5,83E-01	±	9,97E-02	±	4,80E-01	±	1,20E-01	±	2,01E-01	±	1,68E-01
T; přesnost 65%, $\kappa_C = 0, 44$																							
GNE _{δ,range}	289,19	NS	0,001	0,001	0,001	-	±	±	1,09E-03	±	3,30E-04	±	1,36E-03	±	4,70E-04	±	1,78E-03	±	4,79E-04	±	1,29E-03	±	4,00E-04
GNEvoiced _{σ}	239,86	0,05	0,001	0,001	0,001	-	±	±	9,09E-01	±	1,19E-02	±	9,26E-01	±	1,47E-02	±	9,37E-01	±	1,50E-02	±	9,21E-01	±	1,65E-02
CEPSF0 _{CL}	231,96	0,001	0,001	0,001	0,001	-	±	±	5,41E+01	±	6,82E+00	±	4,63E+01	±	4,20E+00	±	5,13E+01	±	8,08E+00	±	6,19E+01	±	7,80E+00
HLR _{δ,σ}	315,29	NS	0,001	0,001	0,001	-	±	±	9,13E-04	±	2,74E-04	±	1,21E-03	±	2,12E-04	±	1,69E-03	±	2,30E-04	±	1,08E-03	±	2,46E-04

Tabulka 7.3: Parametrizace použité ke spojitému hodnocení chraptivosti. Pearsonova lineární korelace r automatického hodnocení se subjektivním hodnocením je uvedena u příslušného popisku stupnice GRBT. Spearmanova korelace ρ jednotlivých parametrů se subjektivním hodnocením a příslušná hladina významnosti dosažené korelace je uvedena u každého parametru.

	parametr	ρ	$p <$		parametr	ρ	$p <$
	G; $r = 0,88$				G; $r = 0,87$		
	BRI $_{\alpha}$	0,29	0,001		GNEvoiced $_{\sigma}$	-0,68	0,001
	corelF0cepsAcr	-0,73	0,001		CEPSF0 $_{CL}$	0,63	0,001
	CPP $_{\mu}$	-0,71	0,001		HLR $_{\delta_{range}}$	-0,77	0,001
	HNR $_{\mu}$	-0,81	0,001		BRI $_{\delta_{min}}$	-0,25	0,001
prodloužená fonace	R; $r = 0,74$			čtený text	R; $r = 0,73$		
	ACRF0 $_{\sigma_R}$	0,51	0,001		BRI $_{\delta_{range}}$	-0,49	0,001
	CEPSF0 $_{\sigma_R}$	0,67	0,001		BRI $_{\delta_{\sigma}}$	-0,49	0,001
	RPK $_{\mu}$	-0,70	0,001		RPK $_{\delta_{range}}$	-0,59	0,001
	SPS $_{LRE_{max}}$	0,57	0,001		HNR $_M$	-0,44	0,001
	B; $r = 0,90$				B; $r = 0,92$		
	BRI $_{\alpha}$	0,27	0,001		GNEvoiced $_{\mu}$	-0,73	0,001
	CEPSF0 $_{CL}$	0,80	0,001		CEPSF0 $_{CL}$	0,68	0,001
	HLR $_{\mu}$	0,64	0,001		HLR $_{\delta_{\sigma}}$	-0,76	0,001
	SR50 $_{\sigma_R}$	0,31	0,001		VUR	-0,44	0,001
T; $r = 0,42$			T; $r = 0,43$				
CPP $_{\delta_{max}}$	0,42	0,001	RPKvoiced $_{\sigma}$	-0,17	0,001		
RPK $_{\delta_{range}}$	0,09	0,05	SWvoiced $_{\sigma_R}$	0,31	0,001		
			SDvoiced $_{\sigma_R}$	0,24	0,001		

8. Shrnutí výsledků a závěr

Základním cílem této práce bylo navrhnout automatický systém pro objektivní hodnocení míry chraptivosti v lidském hlase na základě analýzy akustického signálu. Chraptivost je jedním z prvních příznaků patologických změn v hlasovém traktu. Je proto vhodné mít prostředky ke včasné analýze a především k monitorování vývoje chraptivosti. K určení příčiny vzniku chraptivosti jsou k dispozici kontaktní metody, které mají při diagnostice chraptivosti nezastupitelnou roli, nicméně nejsou v této práci použity. Pro monitorování vývoje chraptivosti existuje celá řada bezkontaktních postupů a vědci z celého světa se snaží o jejich vylepšení. Tyto postupy jsou jak subjektivní, založené na zkušenostech odborníka, tak i objektivní, založené na analýze akustického signálu.

Na téma akustické analýzy řečového signálu existuje mnoho prací, které vyvíjí buď nové algoritmy, nebo vylepšují algoritmy již vyvinuté. Pro analýzu chraptivosti jsou důležité jak algoritmy detekce hlasivkové frekvence, tak algoritmy, které popisují řečový signál komplexněji, vyjadřují, do jaké míry je signál periodický, zda neobsahuje šum na frekvencích, na kterých by z fyziologického pohledu už být zastoupen neměl apod.

Některé metody analýzy akustického signálu musí být manuálně korigovány, jedná se např. o algoritmy, pro jejichž běh je nutná přesná znalost hlasivkové frekvence, ať už se jedná o odhad v Hz, nebo o přesnou identifikaci jednotlivých hlasivkových pulzů pro výpočet parametrů jakými jsou jitter a shimmer.

Cílem této práce bylo používat takové algoritmy, které nevyžadují žádné manuální zásahy během výpočtu či případné dodatečné korekce.

V práci byla využita originální databáze promluv pořizovaná na Foniatrické klinice 1. LF UK a VFN v Praze, která vznikala od 70. let 20. století jako součást standardního vyšetření jak zdravých, tak nemocných pacientů. Tato databáze byla roztržena a pro experimenty byly využity dvě části: soubor 469 nahrávek prodloužené fonace hlásky /a/ a dále 593 nahrávek čteného textu *Podzim na Starém bělidle*. Takto velké databáze se svým rozsahem řadí k rozsáhlejším databázím s dysfonickými promluvami, viz část 2.2, ve které jsou shrnuty databáze používané pro hodnocení chraptivosti.

Obě databáze byly nezávisle na sobě ohodnoceny skupinou pěti odborníků z Foniatrické kliniky. Každý odborník subjektivně ohodnotil obě databáze dvakrát. Opakované hodnocení není ve studiích zabývajících se automatickým hodnocením chraptivosti zmiňováno. Výsledky subjektivního hodnocení (shoda hodnotitele při opakovaném hodnocení, shoda skupiny hodnotitelů a vzájemná korelace parametrů) byly detailně rozebrány v kapitole 5. Z výsledků vyplývá, že hodnotitelé jsou při hodnocení konzistentnější u nahrávek čteného textu. To je v souladu se zpětnou vazbou od našich odborníků, kteří konstatovali, že z prodloužené fonace není možné vytvořit si kompletnější obrázek o celkovém dojmu z hlasu pacienta. V části 7.1 je provedeno detailní vyhodnocení a srovnání obsahu našich databází a jejich subjektivního hodnocení s ostatními pracemi, které se zabývají hodnocením chraptivosti. Z tohoto

srovnání vyplývá, že naši hodnotitelé dosahují při statistickém vyhodnocení shody skupiny a vzájemné shody při opakovaném hodnocení podobných výsledků jako odborníci ve srovnávaných pracích. Na základě dosažených výsledků lze konstatovat, že ohodnocené databáze lze použít pro trénování modelů automatického hodnocení chraptivosti.

V kapitole 4 jsou popsány akustické parametry použité v této práci, je vysvětlen princip jejich činnosti a je naznačeno, jak se liší průběh daného parametru u zdravého a u chraptivého hlasu.

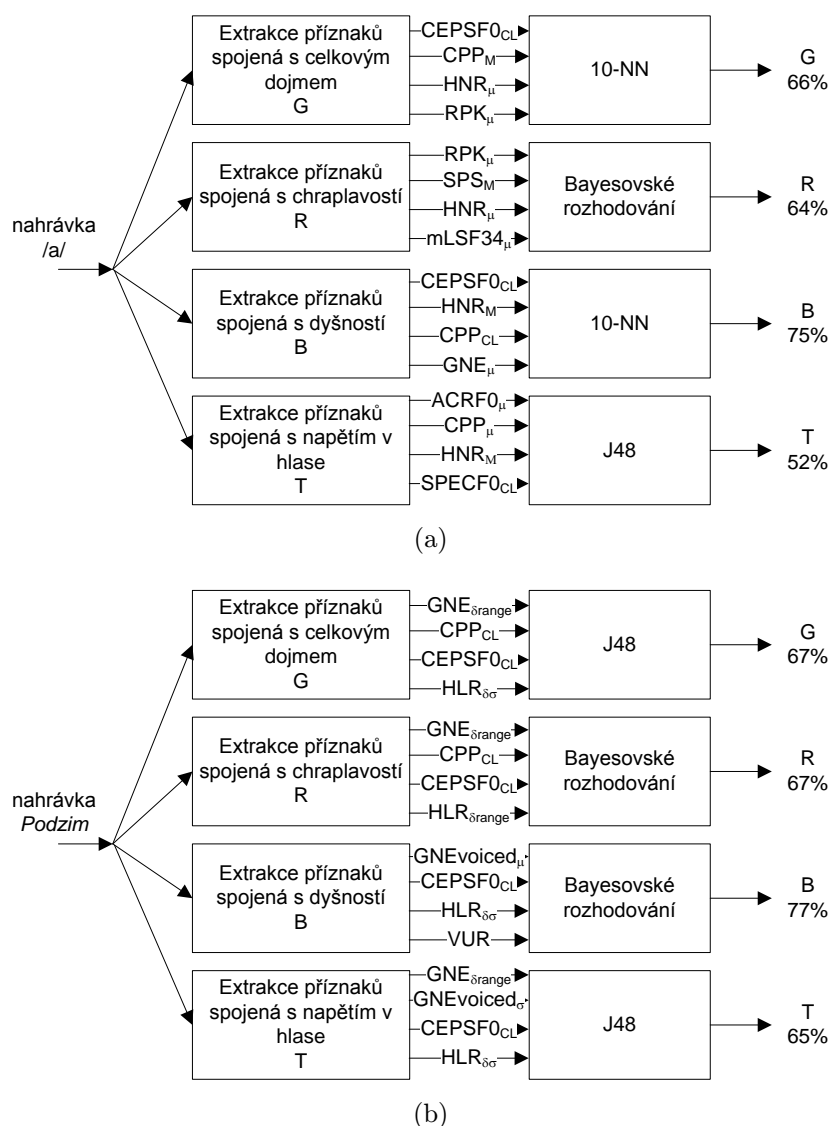
V kapitole 6 je zaveden popis pomocí akustických parametrů uvedených v kapitole 4. Je popsána nejen jejich distribuce střední hodnotou a směrodatnou odchylkou, jak je to v literatuře běžné, ale jsou zavedeny a definovány i popisy časových průběhů těchto parametrů. Tyto nestandardní popisy hrají významnou roli v později navržených automatických systémech pro hodnocení chraptivosti. Dále jsou v kapitole 6 popsány klasifikační úlohy, které jsou hlavní částí této práce: jednak objektivní hodnocení diskretními známkami tak, jak nahrávky hodnotí odborníci a jednak spojitou škálou vzniklou zprůměrováním subjektivního hodnocení.

Výsledky automatického hodnocení podle diskretní stupnice a schémata experimentálního systému pro hodnocení chraptivosti jsou zobrazeny na obr. 8.1. Z obrázku je patrné, že nejlepších výsledků je dosahováno při automatickém hodnocení dyšnosti B (75 % u prodloužené fonace a 77 % u čteného textu). Výsledky hodnocení celkového dojmu G a chraptivosti R jsou v obou případech srovnatelné a pohybují se okolo 66 %. Při hodnocení napětí v hlase T je dosahováno výrazně lepších výsledků u nahrávek čteného textu (52 % u prodloužené fonace vs 65 % u čteného textu). U nahrávek čteného textu je hodnocení parametru T srovnatelné s parametry G a R.

Schémat experimentálního systému pro automatické hodnocení chraptivosti podle spojitě stupnice jsou zobrazena na obr. 8.2. Za povšimnutí stojí, že dosažené hodnoty přesnosti klasifikace vyjádřené Pearsonovým lineárním korelačním koeficientem r jsou téměř totožné jak pro automatické hodnocení prodloužené fonace, tak pro automatické hodnocení čteného textu. Nejlepších výsledků je v obou případech opět dosaženo při hodnocení dyšnosti B ($r=0,90$ pro prodlouženou fonaci a $r=0,92$ pro čtený text). Následuje hodnocení celkového dojmu G ($r=0,88$ pro prodlouženou fonaci vs $r=0,87$ pro čtený text). Horších výsledků je dosaženo pro hodnocení chraptivosti R ($r=0,74$ pro prodlouženou fonaci vs $r=0,73$ pro čtený text), nicméně i tyto hodnoty jsou poměrně vysoké. Hodnocení napětí v hlase dosahuje nejhorších výsledků ($r=0,42$ pro prodlouženou fonaci vs $r=0,43$ pro čtený text). To je způsobeno převážně nejednoznačným zadáním při subjektivním hodnocení. Problém je detailněji popsán v části 7.2.

Použití popisu časových průběhů parametrů se ukázalo jako klíčové, jak je vidět z obr. 8.1 a obr. 8.2. Tyto parametrizace jsou ve výsledných klasifikátorech často používány a i samostatně dosahují dobrých výsledků z pohledu schopnosti odlišit jednotlivé stupně chraptivosti.

V části 7.2 lze nalézt srovnání se studii zabývajícími se automatickým hodnocením chraptivosti podle stupnice GRBAS. Vzhledem k tomu, že převážná většina studií hodnotí pouze celkový dojem z hlasu G, není při porovnání sloučení parame-

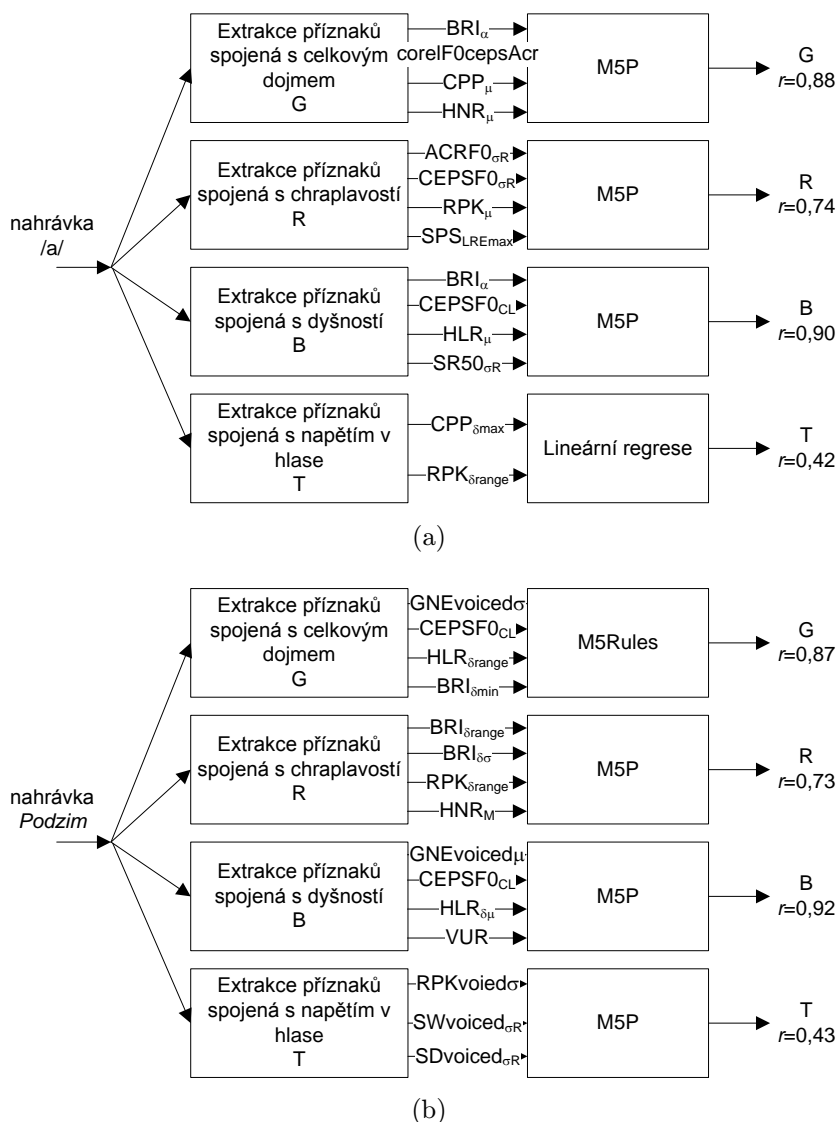


Obrázek 8.1: Schémata experimentálních systému pro automatické hodnocení chraplivosti podle diskrétní stupnice u (a) prodloužené fonace a (b) čteného textu.

tru A a S do parametru T ve stupnici GRBT překážkou. Tato práce navíc obsahuje automatické hodnocení dalších parametrů R, B a T, které jsou zahrnuty pouze do jediné studie. U těchto parametrů je v této práci dosaženo lepších výsledků. Hodnocení podle spojité stupnice není v jiných pracích zmiňováno.

8.1 Shrnutí výsledků

Na základě analýzy subjektivního hodnocení lze konstatovat, že parametr hlasové napětí T nemá smysl zahrnovat do automatické analýzy. To je způsobeno hlavně faktem, že parametrem T nelze popsat zároveň jak hlas slabý, tak hlas přemáhaný (slabé hlasy jsou hodnoceny zápornou známkou, přemáhané kladnou). Někteří hodnotitelé



Obrázek 8.2: Schémata experimentálních systému pro automatické hodnocení chraplavosti podle spojitě stupnice u (a) prodloužené fonace a (b) čteného textu.

tak upřednostňují při hodnocení hlasovou slabost, jiní přemáhání hlasu. Proto jsou výsledky subjektivního hodnocení tohoto parametru nejednoznačné. Poměrně dobrých výsledků vzájemné shody hodnotitelů a shody hodnotitele při opakovaném hodnocení u čteného textu je dosaženo především proto, že většina hlasů byla hodnocena známkou T=0 (tzn. hlas není vnímán ani jako slabý, ani jako přemáhaný). Tato skutečnost statistické hodnocení vychýlila.

Z tab. 5.6 lze odečíst, že při subjektivním hodnocení prodloužené fonace hlásky /a/ je poměrně velká korelace mezi celkovým dojmem z hlasu G a chraplavostí R ($\rho_{GR}=0,72$) a zároveň mezi celkovým dojmem z hlasu a dyšností B ($\rho_{GB}=0,71$). Na druhou stranu mezi chraplavostí a dyšností je korelace poměrně malá ($\rho_{RB}=0,44$). Obdobná situace platí i při hodnocení čteného textu: $\rho_{GR}=0,86$; $\rho_{GB}=0,81$ a $\rho_{RB}=0,68$. **Tyto výsledky potvrzují, že hodnocení celkového dojmu z hlasu G jsou**

směsí dojmů z chraplavosti R a dyšnosti B a zároveň že chraplavost R a dyšnost B jsou posuzovány více nezávisle.

Pro automatické **diskrétní hodnocení prodloužené fonace** se tato fakta odráží na použitých parametrech: při hodnocení celkového dojmu G a chraplavosti R jsou použity společné parametry HNR a RPK, a při hodnocení celkového dojmu G a dyšnosti B jsou použity parametry CEPSF0 a CPP. Zastoupení parametru HNR při hodnocení chraplavosti může být na první pohled zarážející, ovšem na základě provedené rešerše literatury lze konstatovat, že s přibývajícím nepravidelností kmitání hlasivek klesá hodnota HNR. To je způsobeno především větším "rozmažáním" harmonických oblastí spektra. Parametr RPK, který pracuje s autokorelační funkcí, také zohledňuje míru nepravidelnosti kmitání hlasivek, jak je ukázáno v kapitole 3: nepravidelné kmitání hlasivek způsobí, že autokorelační funkce obsahuje více výraznějších lokálních maxim. Tato výraznější lokální maxima způsobí, že hlavní lokální maximum dosahuje nižších hodnot. Při hodnocení dyšnosti se uplatní parametry CPP a CEPSF0 (odhad f_0 v keprální oblasti), který je vedlejším produktem výpočtu CPP. **Z výše uvedeného lze vyvodit závěr, že při subjektivním hodnocení prodloužené fonace diskrétní stupnicí se dyšnost více projeví v keprální oblasti.**

U automatického **hodnocení prodloužené fonace podle spojitě stupnice** nelze takovéto závěry jednoznačně vyvodit. Jediná shoda je u parametru BRI, který je použit jak pro hodnocení celkového dojmu G, tak pro dyšnost B, nicméně tento parametr dosahuje nízkých korelací se subjektivním hodnocením. Nejsilnější korelace se subjektivním hodnocením celkového dojmu je dosahováno parametrem HNR ($\rho=-0,81$), který ovšem není dále použit ani pro hodnocení chraplavosti R, ani dyšnosti B. Pro spojitě hodnocení chraplavosti R dosahuje nejvyšší korelace parametr RPK ($\rho=-0,70$), pro hodnocení dyšnosti B pak parametr CEPSF0 ($\rho=0,80$).

Automatické **hodnocení čteného textu podle diskrétní stupnice** je pro parametry G, R a B společný popis pomocí CEPSF0. Dále pro parametry G a R popis pomocí GNE a CPP. Zde je tedy situace opačná než při hodnocení prodloužené fonace, kdy popis v keprální oblasti byl lepší pro hodnocení dyšnosti. V případě parametrů G a B je společný popis pomocí HLR, který ukazuje, že zastoupení energie na vyšších frekvencích je důležitým faktorem pro klasifikaci.

Pro automatické **hodnocení čteného textu podle spojitě stupnice** jsou **parametry použité pro G a B téměř totožné**: GNE, CEPSF0 a HLR. V obou případech tyto parametry dosahují poměrně vysokých korelací se subjektivním hodnocením (od $\rho=0,63$ pro parametr CEPSF0 pro hodnocení G po $\rho=-0,77$ pro parametr HLR taktéž u hodnocení G). Pro automatické hodnocení chraplavosti R je použita zcela jiná sada parametrů, což je poměrně překvapivé, protože subjektivní hodnocení celkového dojmu G a chraplavosti R dosahuje vyšší korelace než subjektivní hodnocení celkového dojmu G a dyšnosti B ($\rho_{GR}=0,86$ vs $\rho_{GB}=0,81$). Tento jev může být způsoben použitím spojitě stupnice. Zajímavostí jsou poměrně nízké korelace u parametrů použitých pro hodnocení R (od $\rho=-0,44$ u parametru HNR po $\rho=-0,59$ u parametru RPK), nicméně výsledný klasifikátor dosahuje dobrých výsledků ($r=0,73$).

Tabulka 8.1: Porovnání výsledků subjektivního a objektivního hodnocení pomocí Cohnenova κ . Jako *subjektivní* je označeno porovnání hodnocení všech hodnotitelů v prvním kole ku hodnocení všech hodnotitelů ve druhém kole, jako *objektivní* je označeno porovnání subjektivního hodnocení pomocí diskrétní stupnice ku objektivnímu hodnocení navrženého experimentálního klasifikátoru.

		G	R	B	T
/a/	subjektivní	0,41	0,33	0,43	0,37
	objektivní	0,47	0,42	0,59	0,31
čtený text	subjektivní	0,42	0,40	0,43	0,45
	objektivní	0,59	0,50	0,59	0,44

K hodnocení výsledků klasifikace v porovnání se vstupními může být provedeno pomocí Cohnenova κ , které popisuje vzájemnou shodu dvou hodnotitelů (pro více detailů viz část 5.2.3), viz tab.8.1. Na základě dosažených hodnot lze konstatovat, že navržené klasifikátory jsou při hodnocení více podobné vložené předloze (tzn. subjektivnímu hodnocení), než hodnotitelé při opakovaném hodnocení téže nahrávky. Podobné výsledky vykazuje i popis pomocí senzitivity a specificity testu uvedený v příloze 12.

Práce může být shrnuta v následujících bodech:

- byla provedena rozsáhlá rešerše problematiky v několika oblastech:
 - příčiny vzniku chraptivosti;
 - akustické parametry používané při automatickém hodnocení chraptivosti;
 - přehled databází, jejich vlastností, rozsahu a způsobu subjektivního hodnocení chraptivosti;
- byly vytvořeny dvě rozsáhlé databáze nahrávek vhodných k analýze chraptivosti:
 - 469 nahrávek prodloužené fonace hlásky /a/;
 - 593 nahrávek standardizovaného čteného textu *Podzim na Starém; bělidle*
- vytvořené databáze byly opakovaně subjektivně ohodnoceny skupinou pěti odborníků;
- subjektivní hodnocení odborníků bylo detailně analyzováno vhodnými statistickými metodami, které hodnotí míru shody hodnotitele při opakovaném hodnocení a shodu skupiny hodnotitelů;
- bylo implementováno velké množství algoritmů pro analýzu akustického signálu, které lze použít při hodnocení chraptivosti (SPS, RPK, BRI, CPP, GNE, GVA, H1A, HLR, HNR, LSP, SD, SW, SR, CEPSF0, SPECF0, ACRF0);
- byly použity originální metody popisu akustického signálu na základě výše zmíněných parametrizací ;
- bylo provedeno statistické vyhodnocení vhodnosti implementovaných parametrizací k automatickému hodnocení chraptivosti;
- byl realizován experimentální systém pro automatické hodnocení chraptivosti

jak podle spojitě, tak podle diskrétní subjektivní stupnice, který dosahuje dobrých výsledků.

8.2 Přehled výsledků dizertace

Za konkrétní přínosy této práce považuji následující:

- Byly vytvořeny dvě databáze nahrávek s chraptivými hlasy
- Bylo provedeno subjektivní hodnocení nahrávek pro určení míry chraptivosti
 - obě výše zmíněné databáze byly nezávisle na sobě ohodnoceny skupinou pěti odborníků z Foniatrické kliniky 1. LF UK a VFN v Praze
 - každá nahrávka byla skupinou ohodnocena dvakrát s alespoň dvoutýdenním odstupem
 - statistickými metodami bylo ověřeno, že hodnotitelé jsou při opakovaném hodnocení konzistentní, nehodnotí tedy nahrávky nahodile; dále bylo ověřeno, že subjektivní hodnocení databáze je v rámci skupiny hodnotitelů dostatečně konzistentní a tím pádem je možné ho použít k nalezení automatických klasifikátorů; výsledky těchto testů jsou uvedeny v kapitole 5
- V prostředí MATLAB bylo implementováno velké množství algoritmů pro akustickou analýzu zvukového signálu; tyto algoritmy jsou popsány v kapitole 4
- Byly zavedeny originální metody popisu akustického signálu za pomoci výše zmíněných parametrů; tyto metody popisují nejen distribuci hodnot v rámci nahrávky, ale i časový průběh těchto hodnot
- Bylo provedeno statistické vyhodnocení vhodnosti parametrů ke klasifikaci
- Byly navrženy klasifikátory míry chraptivosti v hlase, které dosahují dobrých výsledků:
 - prodloužená fonace hlásky /a/
 - * G: diskretní: 66 %, spojitě 0,88
 - * R: diskretní: 64 %, spojitě 0,74
 - * B: diskretní: 75 %, spojitě 0,90
 - * T: diskretní: 52 %, spojitě 0,42
 - čtený text
 - * G: diskretní: 67 %, spojitě 0,87
 - * R: diskretní: 67 %, spojitě 0,73
 - * B: diskretní: 77 %, spojitě 0,92
 - * T: diskretní: 65 %, spojitě 0,43
- automatické hodnocení prodloužené fonace podle diskrétní stupnice podporuje hypotézu, že celkový dojem je dán kombinací chraplavosti a dyšnosti: dva parametry použité při hodnocení chraplavosti (HNR a RPK) a dva parametry použité při hodnocení dyšnosti (CEPSF0 a CPP) tvoří čtveřici parametrů pro hodnocení celkového dojmu z hlasu; pro hodnocení čteného textu toto neplatí

8.3 Závěry pro další rozvoj hodnocení chraptivosti

Budoucí práce v této oblasti budou spočívat v implementaci dalších algoritmů. Především by bylo vhodné ověřit vlastnosti Mel-frekvenčních keprálních koeficientů pro analýzu chraptivosti. Další zajímavou parametrizací je bezesporu tzv. modulační spektrum (Markaki and Stylianou [87]), které by mohlo přinést další možný pohled na patologickou řeč. Dále by bylo vhodné otestovat vhodnost detektoru náhlých změn ve spektru (Čmejla et al. [96]) k hodnocení chraptivosti. Lze totiž předpokládat, že patologický hlasový trakt bude více časově nestabilní a bude tím pádem obsahovat více změn ve spektru. Při dalších rešerších budou jistě objeveny nové parametrizace, které nemusí být apriori směřovány na analýzu akustických signálů, viz generalizovaný variogram v práci Kacha et al. [71].

Pro další vývoj by byl vhodný velmi robustní algoritmus detekce hlasivkových pulzů. To je velmi dlouho zkoumaný problém, jehož řešení je, jak se zdá, zatím v nedohlednu. Velmi patologické hlasy totiž z pohledu algoritmu nemají hlasivkovou frekvenci, nicméně při poslechu je možné jistý tón, který zřejmě odpovídá kmitání hlasivek nebo jiných struktur hlasového traktu, slyšet. Pokud by byl takový algoritmus k dispozici, bylo by možné se spolehnout na míry jako jsou jitter nebo shimmer i v automatickém systému. Jak bylo ukázáno v předešlých studiích, tak jitter i shimmer jsou velmi dobrými ukazateli míry chraptivosti v řeči.

Pro analýzy čteného textu by bylo vhodné zařadit rozpoznávač řeči, který by prováděl jemnější rozdělení vstupního signálu, než které bylo použito v této práci, tzn. znělé vs neznělé úseky.

Pro využití automatického systému pro objektivní hodnocení chraptivosti v klinické praxi by bylo vhodné implementovat klasifikátory navržené v této práci do samospustitelné aplikace, kterou by odborníci mohli využívat ve svých ordinacích.

Literatura

- [1] *Recommendations of the Union of European Phoniaticians for the assessment of the voice disorders*, 1979. 25 p.
- [2] http://commons.wikimedia.org/wiki/file:m_luvidla.png. online, 2009. URL <http://commons.wikimedia.org/wiki/File:Mluvidla.png>.
- [3] Matti Airas. Tkk aparat: An environment for voce inverse filtering and parametrization. *Logopedics Phoniatics Vocology*, 33:49–64, 2008. doi: 10.1080/14015430701855333.
- [4] Matti Airas and Paavo Alku. Emotions in short vowel segments: Effects of the glottal flow as reflected by the normalized amplitude quotient. In Elisabeth ANDRÉ, Laila DYBKJÆR, Wolfgang MINKER, and Paul HEISTERKAMP, editors, *Affective Dialogue Systems*, volume 3068 of *Lecture Notes in Computer Science*, pages 13–24. Springer Berlin / Heidelberg, 2004. URL http://dx.doi.org/10.1007/978-3-540-24842-2_2.
- [5] J.B. Alonso, J. De Leon, I. Alonso, and M.A. Ferrer. Automatic detection of pathologies in the voice by hos based parameters. *Eurasip Journal on Applied Signal Processing*, 2001(4):275–284, 2001.
- [6] A. Alpan, Y. Maryn, Kacha A., F. Grenez, and J. Schoentgen. Multi-band dysperiodicity analyses of disordered connected speech. *Speech Communication*, 53(1):131 – 141, 2011. doi: 10.1016/j.specom.2010.06.010.
- [7] A. Alpan, J. Schoentgen, Y. Maryn, F. Grenez, and P. Murphy. Assessment of disordered voice via the first rahmonic. *Speech Communication*, 54:655 – 663, 2012. doi: 10.1016/j.specom.2011.04.001. URL <http://www.sciencedirect.com/science/article/pii/S0167639311000628>.
- [8] O. Amir and R. Levine-Yundof. Listeners’ attitude toward people with dysphonia. *Journal of Voice*, 2013. In press.
- [9] Ofer Amir, Michael Wolf, and Noam Amir. A clinical comparison between two acoustic analysis softwares: Mdrv and praat. *Biomedical Signal Processing and Control*, 4(3):202–205, 2009. doi: 10.1016/j.bspc.2008.11.002. URL <http://www.sciencedirect.com/science/article/pii/S1746809408000773>.
- [10] J.D. Arians-Londoño Ando, J.I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruiz, and G. Castellanos-Domínguez. Automatic detection of pathological voices using complexity measures, noise parameters, and mel-cepstral coefficients. *Biomedical Engineering, IEEE Transactions on*, 58(2):370–379, 2011. doi: 10.1109/TBME.2010.2089052.

- [11] *Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V)*. ASHA Special Interest Division 3, Voice and Voice Disorders. URL <http://www.asha.org/uploadedFiles/ASHA/SIG/03/affiliate/CAPE-V-Purpose-Applications.pdf>.
- [12] Shaheen N. Awan, Ashley Giovinco, and Jennifer Owens. Effects of vocal intensity and vowel type on cepstral analysis of voice. *Journal of Voice*, 26(5): 670.e15 – 670.e20, 2012. doi: 10.1016/j.jvoice.2011.12.001.
- [13] R.K. Balasubramaniam, J.S. Bhat, S. Fahim III, and R. Raju III. Cepstral analysis of voice in unilateral adductor vocal fold palsy. *Journal of Voice*, 25: 326–329, 2011.
- [14] Tom Bäckström and Carlo Magi. Properties of line spectrum pair polynomials—a review. *Signal Processing*, 86(11):3286–3298, 2006. doi: 10.1016/j.sigpro.2006.01.010.
- [15] M. Benasid, J. Schoentgen, F. Bucella, and S. Ciocea. Estimation by means of wavelet analysis of the signal-to-noise ratio of dysphonic voices. *Proceedings of the ProRISC Workshop on Circuits, Systems and Signal Processing*, pages 47–51, 1997.
- [16] Ingrid Bendl. Computational-tools for the analysis of acoustical and physiological criteria of the singing voice. *Proceedings of the Second Vienna Talk*, pages 21–23, 2010. URL http://www.viennatalk.mdw.ac.at/papers/Pap_01_94_Bendl.pdf.
- [17] Frederic Bettens, Francis Grenez, and Jean Schoentgen. Estimation of vocal dysperiodicities in disordered connected speech by means of distant-sample bidirectional linear predictive analysis. *J. Soc. Am.*, 117(1):328–337, 2005. doi: 10.1121/1.1835511. URL <http://link.aip.org/link/?JAS/117/328/1>.
- [18] Tarika Bhuta, Linda Patrick, and James D. Garnett. Perceptual evaluation of voice quality and its correlation with acoustic measurements. *Journal of Voice*, 18(3):299–304, 2004. doi: 10.1016/j.jvoice.2003.12.004.
- [19] Paul Boersma. 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, 1993.
- [20] Paul Boersma. Should jitter be measured by peak picking or by waveform matching? *Folia Phoniatr Logop*, 61:305–308, 2009. doi: 10.1159/000245159.
- [21] Paul Boersma and David Weenink. Praat: doing phonetics by computer [computer program], 2011. URL <http://www.praat.org/>. Version 5.2.35.
- [22] Mike Brookes. Voicebox: Speech processing toolbox for matlab. online, 2011. URL <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html>.

- [23] Arturo Camacho and John G. Harris. A sawtooth waveform inspired pitch estimator for speech and music. *J. Soc. Am.*, 124(3):1638–1652, 2008. doi: 10.1121/1.2951592. URL <http://link.aip.org/link/?JAS/124/1638/1>.
- [24] Jacob Cohnen. A coefficient of agreement for nominal scales. *Educational and Physiological Measurement*, 20:37–46, 1960.
- [25] S.C. Costa, B.G. Aguiar Neto, and J.M. Fechine. Pathological voice discrimination using cepstral analysis, vector quantization and hidden markov models. In *BioInformatics and BioEngineering, 2008. BIBE 2008. 8th IEEE International Conference on*, pages 1–5, 2008. doi: 10.1109/BIBE.2008.4696783.
- [26] Silvana C. Costa, Suzete Correia, Hanniere Falcão, Náthalee Almeida, Benedito G. A. Neto, and Joseana M. Fechine. Pathological voice discrimination based on entropy measurements. *Proceedings of the 2008 ACM symposium on Applied computing*, pages 1424–1425, 2008.
- [27] Silvana C. Costa, Benedito G. A. Neto, Joseana M. Fechine, and Suzete. CORREIA. Parametric cepstral analysis for pathological voice assessment. *Proceedings of the 2008 ACM symposium on Applied computing*, pages 1410–1414, 2008.
- [28] Lee J. Cronbach. Coeficient alpha and the internal structure of tests. *Psychometrika*, 16:297–334, 1951.
- [29] Marc S. De Bodt, Floris. L. Wuyts, Paul H. Van de Heyning, and Christophe Croux. Test-retest study of the grbas scale: Influence of experience and professional background on perceptual rating of voice quaity. *Journal of Voice*, 11(1):74–80, 1997.
- [30] Alain de Cheveigné and Hideki Kawahara. Yin, a fundamental frequency estimator for speech and music. *J. Acoust. Soc. Am.*, 111(4):1917–1930, 2002. doi: DOI:10.1121/1.1458024. URL <http://dx.doi.org/doi/10.1121/1.1458024>.
- [31] Marcelo. de Oliveira Rosa, José Carlos Pereira, and Marcos Grellet. Adaptive estimation of residue signal for voice pathology diagnosis. *Biomedical Engineering, IEEE Transactions on*, 47(1):96–104, 2000. doi: 10.1109/10.817624.
- [32] P. H. Dejonckere, C. Obbens, G. M. de Moor, and G. H. Wieneke. Perceptual evaluation of dysphonia: reliability and relevance. *Folia Phoniatica*, 45:76–83, 1993.
- [33] P. H. Dejonckere, Patrick Bradley, Pais Clemente, Guy Cornut, Lise Crevier-Buchman, Gerhard Friedrich, Paul Van De Heyning, Marc Remacle, and Virginie Woisard. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments

- and evaluating new assessment techniques. *European Archives of Oto-Rhino-Laryngology*, 258:77–82, 2001. ISSN 0937-4477. URL <http://dx.doi.org/10.1007/s004050000299>. 10.1007/s004050000299.
- [34] M. Dogan, I. Midi, M.A. Yazici, I. Kocak, D. Günal, and M.A. Sehitoglu. Objective and subjective evaluation of voice quality in multiple sclerosis. *Journal of Voice*, 21(6):735–740, 2007.
- [35] Thomas Drugman and Abeer Alwan. Joint robust voicing detection and pitch estimation based on residual harmonics. In *INTERSPEECH*, pages 1973–1976. ISCA, 2011.
- [36] Jan Černocký. Zpracování řečových signálů – studijní opora. online, 2006. URL http://www.fit.vutbr.cz/study/courses/ZRE/public/opora/zre_opora.pdf.
- [37] Carlos Ferrer, Eduardo González, and María Hernández-Díaz. Evaluation of time and frequency domain-based methods for the estimation of harmonics-to-noise-ratios in voice signals. In José Martínez-Trinidad, Jesús Carrasco Ochoa, and Josef Kittler, editors, *Progress in Pattern Recognition, Image Analysis and Applications*, volume 4225 of *Lecture Notes in Computer Science*, pages 406–415. Springer Berlin / Heidelberg, 2006. doi: 10.1007/11892755“42. URL http://dx.doi.org/10.1007/11892755_42. ISBN: 978-3-540-46556-0.
- [38] Carlos A. Ferrer, Eduardo González, and Maria E. Hernández-Díaz. Correcting the use of ensemble averages in the calculation of harmonics to noise ratios in voice signals. *J. Soc. Am.*, 118(2):605–607, 2005. doi: 10.1121/1.1940450. URL <http://link.aip.org/link/?JAS/118/605/1>.
- [39] Joseph L. Fleiss. Measuring nominal scale agreement among many raters. *Physiological Bulletin*, 76(5):378–382, 1971.
- [40] R. Fraile, N. Sáenz-Lechón, J. I. Godino-Llorente, V. Oasma-Ruiz, and C Frenouille. Automatic detection of laryngeal pathologies in records of sustained vowels by means of mel-frequency cepstral coefficient parameters and differentiation of patients by sex. *Folia Phoniatr Logop*, 61:146–152, 2009. doi: 10.1159/000219950.
- [41] Rubén Fraile, Juan Ignacio Godino-Llorente, Nicolás Sáenz-Lechón, Víctor Oasma-Ruiz, and Pedro Gómez-Vilda. Automatic detection of laryngeal pathology on sustained vowels using short-term cepstral parameters: Analysis of performance and theoretical justification. In Ana Fred, Joaquim Filipe, and Hugo Gamboa, editors, *Biomedical Engineering Systems and Technologies*, volume 25 of *Communications in Computer and Information Science*, pages 228–241. Springer Berlin Heidelberg, 2009. doi: 10.1007/978-3-540-92219-3“17. URL http://dx.doi.org/10.1007/978-3-540-92219-3_17.

- [42] Corinne Fredouille, Gilles Pouchoulin, Alain Ghio, Joana Revis, Jean-Francois Bonastre, and Antoine Giovanni. Back-and-forth methodology for objective voice quality assessment: From/to expert knowledge to/from automatic classification of dysphonia. *EURASIP JOURNAL ON ADVANCES IN SIGNAL PROCESSING*, 2009.
- [43] Matthias Fröhlich, Dirk Michaelis, and Strube Hans Werner. Acoustic voice quality description: Case studies for different regions of the hoarseness diagram. *Advances in Quantitative Laryngoscopy*, 2:143–150, 1997.
- [44] Matthias Fröhlich, Dirk Michaelis, and Hans Werner Strube. Acoustic breathiness measures in the description of pathologic voices. In *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, volume 2, pages 937–940, may 1998. doi: 10.1109/ICASSP.1998.675420.
- [45] Matthias Fröhlich, Dirk Michaelis, Hans Werner Strube, and Eberhard Kruse. Acoustic voice analysis by means of the hoarseness diagram. *J Speech Lang Hear Res*, 43(3):706–720, 2000. URL <http://jslhr.asha.org/cgi/content/abstract/43/3/706>.
- [46] A. Frid and Y. Lavner. Acoustic-phonetic analysis of fricatives for classification using svm based algorithm. In *Electrical and Electronics Engineers in Israel (IEEEI), 2010 IEEE 26th Convention of*, pages 751 –755, 2010. doi: 10.1109/IEEEI.2010.5662110.
- [47] Juan I. Godino-Llorente, P. Gomez-Vilda, and M. Blanco-Velasco. Dimensionality reduction of a pathological voice quality assessment system based on gaussian mixture models and short-term cepstral parameters. *IEEE Transactions on Biomedical Engineering*, 53(10):1943–1953, 2006.
- [48] Juan I. Godino-Llorente, Pedro Gomez-Vilda, Fernando Cruz-Rolán, and Manuel Blanco-Velasco. Pathological likelihood index as a measurement of the degree of voice normality and perceived hoarseness. *Journal of Voice*, 24(6):667–677, 2010.
- [49] Juan I. Godino-Llorente, Víctor Osma-Ruiz, Sáenz-Lechón Nicolás, Pedro Gomez-Vilda, Manuel Blanco-Velasco, and Fernando. Cruz-Roldán. The effectiveness of the glottal to noise excitation ratio for the screening of voice disorders. *Journal of Voice*, 24(1):47–56, 2010.
- [50] Matthew Gordon and Ladefoged Peter. Phonation types: a cross-linguistic overview. *Journal of Phonetics*, 29(4):383–406, 2001. doi: 10.1006/jpho.2001.0147.
- [51] Svante Granqvist and Hammarberg Britta. The correlogram: A visual display of periodicity. *The Journal of the Acoustical Society of America*, 114(5):2934–2945, 2003. doi: 10.1121/1.1590972. URL <http://link.aip.org/link/?JAS/114/2934/1>.

- [52] Gary Groth-Marant. *Handbook of Physiological Assessment*. John Wiley & Sons, Inc., 2003. ISBN: 0-471-41979-6.
- [53] Stefan Hadjitodorov and Petar Mitev. A computer system for acoustic analysis of pathological voices and laryngeal diseases screening. *Medical Engineering & Physics*, 24(6):419–429, 2002. doi: 10.1016/S1350-4533(02)00031-0. URL <http://www.sciencedirect.com/science/article/pii/S1350453302000310>.
- [54] Aleš Hahn. *Otorinolaryngologie a foniatrie v současné praxi*. Grada, 2007.
- [55] M. A. Hall. *Correlation-based Feature Subset Selection for Machine Learning*. PhD thesis, University of Waikato, Hamilton, New Zealand, 1988.
- [56] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The weka data mining software: an update. *SIG-KDD Explor. Newsl.*, 11(1):10–18, Nov. 2009. doi: 10.1145/1656274.1656278.
- [57] Hai Hang and Jiaqiang Pan. Speech pitch determination based on hilbert-huang transform. *Signal Processing*, 86:792–803, 2006.
- [58] Yolanda D. Heman-Ackah, Deirdre D. Michael, and George S Goding Jr. The relationship between cepstral peak prominence and selected parameters of dysphonia. *Journal of Voice*, 16(1):20–27, 2002. doi: 10.1016/S0892-1997(02)00067-X. <http://www.sciencedirect.com/science/article/pii/S089219970200067X>.
- [59] Yolanda D. Herman-Ackah. Reliability of calculating the cepstral peak without linear regression analysis. *Journal of Voice*, 18(2):203–208, 2004. doi: 10.1016/j.jvoice.2004.01.005.
- [60] James Hillenbrand. A methodological study of perturbation and additive noise in synthetically generated voice signals. *J Speech Hear Res*, 30(4):448–461, 1987. URL <http://jslhr.asha.org/cgi/content/abstract/30/4/448>.
- [61] James Hillenbrand and Robert A. Houde. Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *J Speech Hear Res*, 39(2):311–321, 1996.
- [62] James Hillenbrand, Ronald A. Cleveland, and Robert L. Erickson. Acoustic correlates of breathy vocal quality. *J Speech Hear Res*, 37(4):769–778, 1994.
- [63] Minoru Hirano. *Clinical Examination of Voice*. Springer London, 1981.
- [64] Viveka L. Åhlander, Lars V. Malm, and Lucyna Schalén. Hoarseness as a sign of possible nonspecific mucosal hyperreactivity in vocal tract. *Journal of Voice*, 23(6):707–715, 2009. URL <http://www.scopus.com/inward/record.url?eid=2-s2.0-70350565509&partnerID=40&md5=2dd8249f121506941ac44c1c8e03c553>.

- [65] Norman D. Hogikyan and Girish Sethuraman. Validation of an instrument to measure voice-related quality of life (v-rqol). *Journal of Voice*, 13(4):559–569, 1999. doi: 10.1016/S0892-1997(99)80010-1.
- [66] Geoffrey Holmes, Mark Hall, and Eibe Frank. Generating rule sets from model trees. In *Twelfth Australian Joint Conference on Artificial Intelligence*. Springer, 1999.
- [67] Ivan Hybášek. *Ušní, nosní a krční lékařství*. Galén, 1999.
- [68] Barbara H. Jacobson, Alex Johnson, Cynthia Grywalski, Alice Silbergleit, Gary Jacobson, and Michael S. Benninger. The voice handicap index (vhi): Development and validation. *Am J Speech Lang Pathol*, 6:66–70, 1977.
- [69] Cheol-Woo Jo, Ho-Gyun Bang, and W.A. Ainsworth. Improved glottal closure instant detector based on linear prediction and standard pitch concept. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, volume 2, pages 1217–1220, 1996. doi: 10.1109/ICSLP.1996.607827.
- [70] George H. John and Pat Langley. Estimating continuous distributions in bayesian classifiers. In *In Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 338–345. Morgan Kaufmann, 1995.
- [71] A. Kacha, F. Grenez, J. Schoentgen, and K. Benmahammed. Dysphonic speech analysis using generalized variogram. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, volume 1, pages 917–920, 2005. doi: 10.1109/ICASSP.2005.1415264.
- [72] Michael P. Karnell, Sarah D. Melton, Jana M. Childes, Todd C. Coleman, Scott A. Dailey, and Henry T. Hoffman. Reliability of Clinician-Based (GRBAS and CAPE-V) and Patient-Based (V-RQOL and IPVI) Documentation of Voice Disorders. *Journal of Voice*, 21(5):576–590, 2007. doi: DOI:10.1016/j.jvoice.2006.05.001.
- [73] H. Kasuya, S. Ogawa, and Y. Kikuchi. An adaptive comb filtering method as applied to acoustic analyses of pathological voice. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '86.*, volume 11, pages 669–672, apr 1986. doi: 10.1109/ICASSP.1986.1168996.
- [74] Kenji Kira and Larry A. Rendell. A Practical Approach to Feature Selection. In Derek H. Sleeman and Peter Edwards, editors, *Ninth International Workshop on Machine Learning*, pages 249–256. Morgan Kaufmann, 1992.
- [75] Igor Kononenko. Estimating Attributes: Analysis and Extensions of RELIEF. In Francesco Bergadano and Luc De Raedt, editors, *European Conference on Machine Learning*, pages 171–182. Springer, 1994.

- [76] Jody Kreiman, Bruce R. Gerratt, Kristin Precoda, and Gerald S. Berke. Individual differences in voice quality perception. *JSHLR*, 35:512–520, 1992. <http://www.surgery.medsch.ucla.edu/glottalaffairs/papers/kreiman>
- [77] J. Lee and M. Hahn. Automatic assessment of pathological voice quality using higher-order statistics in the lpc residual domain. *Eurasip Journal on Advances in Signal Processing*, 2009.
- [78] Ji-Yeoun Lee, Sangbae Jeong, and Minsoo Hahn. Pathological voice detection using efficient combination of heterogenous features. *IEICE Transactions on Information and Systems*, E91-D(2):367–370, 2008.
- [79] S.Y. Lowell, R.H. Colton, R.T. Kelley, and S.A. Mizia. Predictive value and discriminant capacity of cepstral- and spectral-based measures during continuous speech. *Journal of Voice*, 2013. In Press.
- [80] E.P.-M. Ma and E.M.-L. Yiu. Multiparametric evaluation of dysphonic severity. *Journal of Voice*, 20(3):380–390, 2006.
- [81] Julia K. MacCallum, Aleksandra E. Olszewski, Yu Zhang, and Jack J. Jiang. Effects of low-pass filtering on acoustic analysis of voice. *Journal of Voice*, 25:15–20, 2011. doi: 10.1016/j.jvoice.2009.08.004.
- [82] C. Manfredi, M. D’Aniello, P. Brusciaglioni, and A. Ismaelli. A comparative analysis of fundamental frequency estimation methods with application to pathological voices. *Medical Engineering & Physics*, 22(2):135 – 147, 2000. doi: DOI:10.1016/S1350-4533(00)00018-7.
- [83] Claudia Manfredi. Adaptive noise energy estimation in pathological speech signals. *Biomedical Engineering, IEEE Transactions on*, 47(11):1538 –1543, 2000. doi: 10.1109/10.880107.
- [84] Claudia Manfredi, Leonardo Bocchi, and Giovanna Cantarella. A multipurpose user-friendly tool for voice analysis: Application to pathological adult voices. *Biomedical Signal Processing and Control*, 4(3):212–220, 2009. doi: 10.1016/j.bspc.2008.11.006. URL <http://www.sciencedirect.com/science/article/pii/S1746809408000815>.
- [85] P. Maragos. Fractal aspects of speech signals: dimension and interpolation. In *Acoustics, Speech, and Signal Processing, 1991. ICASSP-91., 1991 International Conference on*, volume 1, pages 417 –420, 1991. doi: 10.1109/ICASSP.1991.150365.
- [86] M. Markaki, Y. Stylianou, J.D. Arias-Londoño, and J.I. Godino-Llorente. Dysphonia detection based on modulation spectral features and cepstral coefficients. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 5162–5165, 2010. doi: 10.1109/ICASSP.2010.5495020.

- [87] Maria Markaki and Yannis Stylianou. Using modulation spectra for voice pathology detection and classification. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pages 2514–2517, sept. 2009. doi: 10.1109/IEMBS.2009.5334850.
- [88] David Martin, James Fitch, and Wolfe Virginia. Pathological voice type and the acoustic prediction of severity. *JSHLR*, 38:765–771, 1995.
- [89] Youri Maryn, Nelson Roy, Marc De Bodt, Paul Van Cauwenberge, and Paul Corthals. Acoustic measurement of overall voice quality: A meta-analysis. *JASA*, 126(5):2619–2634, 2009. doi: 10.1121/1.3224706.
- [90] Youri Maryn, Paul Corthals, Paul Van Cauwenberge, Nelson Roy, and Marc De Bodt. Toward improved ecological validity in the acoustic measurement of overall voice quality: Combining continuous speech and sustained vowels. *Journal of Voice*, 24(5):540–555, 2010. doi: 10.1016/j.jvoice.2008.12.014.
- [91] Mili M. Mathew and Jayashree S. Bhat. Soft phonation index – a sensitive parameter? *Indian Journal of Otolaryngology and Head & Neck Surgery*, 61(2): 127–130, 2009.
- [92] MATLAB. *version 8.0.0.783 (R2012b)*. The MathWorks Inc., 2012.
- [93] Philip McLeod and geoff Wyvill. A smarter way to find pitch. *Proc. International Computer Music Conference*, pages 138–141, 2005.
- [94] Ian Vince McLoughlin. Line spectral pairs. *Signal Processing*, 88(3):448–467, 2008. doi: 10.1016/j.sigpro.2007.09.003.
- [95] Ann-Christine Mecke and Johan Sundberg. Gender differences in childrens’s singing voices: Acoustic analyses and results of a listening test. *J. Soc. Am.*, 127(5):3223–3231, May 2010.
- [96] Roman Čmejla, Jan Ruzs, Petr Bergl, and Jan Vokřál. Bayesian changepoint detection for the automatic assessment of fluency and articulatory disorders. *Speech Communication*, 55(1):178–189, 2013. doi: 10.1016/j.specom.2012.08.003.
- [97] D. Michaelis, T. Gramss, and H. W. Strube. Glottal-to-noise excitation ratio – a new measure for describing pathological voices. *ACUSTICA*, 83:700–706, 1997.
- [98] Dirk Michaelis, Matthias Fröhlich, and Strube Hans Werner. Selection and combination of acoustic features for the description of pathologic voices. *J. Soc. Am.*, 103(3):1628–1639, 1998.
- [99] Cornelia Moers, Bernd Möbius, Frank Rosanowski, Elmar Nöth, Ulrich Eysholdt, and Tino Haderlein. Vowel- and text-based cepstral analysis of chronic hoarseness. *Journal of Voice*, 26(4):416–424, 2012. doi: 10.1016/j.jvoice.2011.05.001.

- [100] E. Moore and M. Clements. Algorithm for automatic glottal waveform estimation without the reliance on precise glottal closure information. In *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, volume 1, pages 101–104, 2004. doi: 10.1109/ICASSP.2004.1325932.
- [101] Peter Murphy and Olatunji Akande. Cepstrum-based estimation of the harmonics-to-noise ratio for synthesized and human voice signals. In Marcos Faundez-Zanuy, Léonard Janer, Anna Esposito, Antonio Satue-Villar, Joseph Roure, and Virginia Espinosa-Duro, editors, *Nonlinear Analyses and Algorithms for Speech Processing*, volume 3817 of *Lecture Notes in Computer Science*, pages 150–160. Springer Berlin / Heidelberg, 2005. doi: 10.1007/11613107_13. URL http://dx.doi.org/10.1007/11613107_13. ISBN: 978-3-540-31257-4.
- [102] Peter Murphy and Olatunji Akande. Cepstrum-based harmonics-to-noise ratio measurement in voiced speech. In Gérard CHOLLET, Anna ESPOSITO, Marcos FAUNDEZ-ZANUY, and Maria MARINARO, editors, *Nonlinear Speech Modeling and Applications*, volume 3445, pages 87–93. Springer Berlin / Heidelberg, 2005. URL http://dx.doi.org/10.1007/11520153_9.
- [103] Peter J. Murphy. Perturbation-free measurement of the harmonics-to-noise ratio in voice signals using pitch synchronous harmonic analysis. *The Journal of the Acoustical Society of America*, 105(5):2866–2811, 1999. doi: 10.1121/1.426901.
- [104] Peter J. Murphy. Periodicity estimation in synthesized phonation signals using cepstral harmonic peaks. *Speech Communication*, 48(12):1704 – 1713, 2006. doi: 10.1016/j.specom.2006.09.001.
- [105] Peter J. Murphy and Olatunji O. Akande. Quantification of glottal and voiced speech harmonics-to-noise ratios using cepstral-based estimation. In *NOLISP*, volume 3817, pages 150–160, 2005. doi: http://dx.doi.org/10.1007/11613107_13.
- [106] Peter J. Murphy and Akande O. Olatunji. Noise estimation in voice signals using short-term cepstral analysis. *J. Soc. Am.*, 121(3):1679–1690, 2007. doi: 10.1121/1.2427123. URL <http://link.aip.org/link/?JAS/121/1679/1>.
- [107] K. Mustafa and Ian C. Bruce. Robust formant tracking for continuous speech with speaker variability. *Audio, Speech, and Language Processing, IEEE Transactions on*, 14(2):435–444, 2006. doi: 10.1109/TSA.2005.855840.
- [108] Hiroshi Muta, Thomas Baer, Kikuju Wagatsuma, Teruo Muraoka, and Fukuda Hiroyuki. A pitch-synchronous analysis of hoarseness in running speech. *J. Acoust. Soc. Am.*, 84(4):1292–1301, 1988. doi: DOI:10.1121/1.396628.

- [109] Patrick A. Naylor, Anastasis Kounoudes, Jon Gudnason, and Mike Brookes. Estimation of glottal closure instants in voiced speech using the dypsa algorithm. *Audio, Speech, and Language Processing, IEEE Transactions on*, 15(1): 34–43, 2007. doi: 10.1109/TASL.2006.876878.
- [110] Benedito G. A. Neto, Silvana C. Costa, Joseana M. Fechine, and Menaka Muppa. Feature estimation for vocal fold edema detection using short-term cepstral analysis. *Proceedings of the 7th IEEE International Conference on Bioinformatics and Bioengineering*, pages 1158–1162, 2007.
- [111] Lucas Léon Oller. *Analysis of Voice Signals for the Harmonics-to-Noise Crossover Frequency*. PhD thesis, KTH - School of Computer Science and Communication (CSC), Department of Speech, Music and Hearing, 2008. URL <http://www.speech.kth.se/prod/publications/files/3270.pdf>.
- [112] Koichi Omori, Hisayoshi Kojima, Rajesh Kakani, David H. Slavit, and Stanley M. Blaugrund. Acoustic characteristics of rough voice: Subharmonics. *Journal of Voice*, 11(1):40–47, 1997.
- [113] Zdena Palková. *Fonetika a fonologie češtiny*. Univerzita Karlova, 1994.
- [114] Yaozhang Pan, Shuzhi Sam Ge, Feng Ru Tang, and A. Al Mamun. Detection of epileptic spike-wave discharges using svm. In *Control Applications, 2007. CCA 2007. IEEE International Conference on*, pages 467–472, oct. 2007. doi: 10.1109/CCA.2007.4389275.
- [115] M. P. Paulraj, Sazali Bin Yaacob, Ahamad Nazri Abdullah, and Natraj Sathees Kumar. Segmentation of voiced portion for voice pathology classification using fuzzy logic. *Challenges and Innovations in Information Technology-2010*, pages 561–566, 2010. URL http://publicweb.unimap.edu.my/~paul/pdfs/papers/2010_CIIT_sathees.pdf.
- [116] Ce Peng, Wenxi Chen, Xin Zhu, Baiukun Wan, and Daming Wei. Pathological voice classification based on a single vowel’s acoustic features. *Seventh International Conference on Computer and Information Technology*, 2007.
- [117] Kay PENTAX, Massachusetts Eye, and Ear Infirmary. Disordered voice database and program, model 4337. online.
- [118] Verma Piyush, Pal Manisha, and Raj Anoop. Objective acoustic analysis of voice improvement after phonosurgery. *Indian Journal of Otolaryngology and Head & Neck Surgery*, 62:131–137, 2010. doi: 10.1007/s12070-010-0024-6. URL <http://dx.doi.org/10.1007/s12070-010-0024-6>.
- [119] Josef Psutka, Luděk Müller, Jindřich Matoušek, and Vlasta Radová. *Mluvíme s počítačem česky*. Academia, Praha, 2006. ISBN 80-200-1309-1.

- [120] Yingyong Qi, Robert E. Hillman, and Claudio Milstein. The estimation of signal-to-noise ratio in continuous speech for disordered voices. *J. Soc. Am.*, 105(4):2532–2535, 1999. doi: 10.1121/1.426860. URL <http://link.aip.org/link/?JAS/105/2532/1>.
- [121] Ross Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.
- [122] Ross J. Quinlan. Learning with continuous classes. In *5th Australian Joint Conference on Artificial Intelligence*, pages 343–348, Singapore, 1992. World Scientific.
- [123] B. Radish Kumar, J.S. Bhat, and N. Prasad. Cepstral analysis of voice in persons with vocal nodules. *Journal of Voice*, 24(6):651–653, 2010.
- [124] R. Riley. *Cool Edit 2000 a Cool Edit Pro 2: Strih, efekty, procesory*. Computer Press, Brno., 2004.
- [125] Marko Robnik-Sikonja and Igor Kononenko. An adaptation of Relief for attribute estimation in regression. In Douglas H. Fisher, editor, *Fourteenth International Conference on Machine Learning*, pages 296–304. Morgan Kaufmann, 1997.
- [126] N. Saenz-Lechon, J.I. Godino-Llorente, V. Osma-Ruiz, M. Blanco-Velasco, and F. Cruz-Roldan. Automatic assessment of voice quality according to the grbas scale. In *Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE*, pages 2478–2481, 2006. doi: 10.1109/IEMBS.2006.260603.
- [127] Jungo Sawa and Toshihiko Morikawa. Interrater reliability for multiple raters in clinical trial of ordinal scale. *Drug Information Journal*, 41:595–605, 2007.
- [128] Jean Schoentgen and Fabrizio Bucella. Acoustic analysis of dysphonic voices: Descriptors and methods. *LARYNX*, 37-46:37–46, 1997.
- [129] Nicolás Sáenz-Lechón, V. Osma-Ruiz, J. I. Godino-Llorente, M. Blanco-Velasco, F. Cruz-Rolán, and J. D. ARIALS-LONDOÑO. Effects of audio compression in automatic detection of voice pathologies. *Biomedical Engineering, IEEE Transactions on*, 55(12):2831–2835, dec. 2008. doi: 10.1109/TBME.2008.923769.
- [130] Nicolás Sáenz-Lechón, V. Osma-Ruiz, R. Fraile, J. Godino-Llorente, and P. Gómez-Vilda. Screening voice disorders with the glottal-to-noise excitation ratio. 2009. URL http://oa.upm.es/3398/1/INVE_MEM_2008_54786.pdf.
- [131] Kumara Shama, Anantha Krishna, and Miranjan U. Cholayya. Study of harmonics-to-noise ratio and critical-band energy spectrum of speech as acoustic indicators of laryngeal and voice pathology. *EURASIP J. Appl. Signal Process.*, pages 50–50, 2007.

- [132] C. Sinderby, L. Lindström, and A. E. Grassino. Automatic assessment of electromyogram quality. *J Appl Physiol*, 75(5):1803–1815, 1995.
- [133] K. Sjölander and J. Beskow. *Wavesurfer [Computer program] (Version 1.8.5)*. URL <http://www.speech.kth.se/wavesurfer>.
- [134] Adam Stráník and Roman Čmejla. Popis spektrálních vlastností sibilantů /s/, /š/, /z/ a /ž/ na základě lsp parametrizace. *Akustické listy*, 16(4):9–14, 2010.
- [135] Adam Stráník and Roman Čmejla. An analysis of iterative algorithm for estimation of harmonics-to-noise ratio in speech. In *Technical Computing Prague 2011*, pages 1–7, 2011.
- [136] D. Talkin. A robust algorithm for pitch tracking (rapt). In W. B. KLEIN and K. K. PALIVAL, editors, *Speech Coding and Synthesis*, pages 459–518. Elsevier, 1995.
- [137] Raissa Tavares, Monteiro Nathália, Suzete Correia, Silviana C. Costa, Benedito G. A. Neto, and Joseana M. Fachine. Optimizing laryngeal pathology detection by using combined cepstral features. *Proceedings of 20th International Conference on Acoustics, ICA 2010*, pages 23–27, 2010.
- [138] Tiago F. Tavares, Jayme G. A Barbedo, and Amauri Lopes. Performance evaluation of fundamental frequency estimation algorithms. *Proceedings of the International Workshop on Telecommunications - IWT*, pages 94–97, 2009. URL <http://www.dca.fee.unicamp.br/~tavares/iwt2009.pdf>.
- [139] S.A. Thati, B. Bollepalli, P. Bhaskararao, and B. Yegnanarayana. Analysis of breathy voice based on excitation characteristics of speech production. In *Signal Processing and Communications (SPCOM), 2012 International Conference on*, pages 1–5, 2012. doi: 10.1109/SPCOM.2012.6290015.
- [140] Robert Vích and Martin Vondra. Speech spectrum envelope modeling. In *Proceedings of the 2007 COST action 2102 international conference on Verbal and nonverbal communication behaviours*, COST 2102'07, pages 129–137. Springer-Verlag, 2007. URL <http://portal.acm.org/citation.cfm?id=1783474.1783491>.
- [141] J. G. Švec, M. Lejska, J. Frostová, M. Zábrodský, J. Dršata, and P. Král. Česká verze dotazníku voice handicap index pro kvantitativní hodnocení hlasových potíží vnímaných pacientem. *Otorinolaryng. a Foniat.*, 58(3):132–139, 2009.
- [142] D. Veeneman and S. BeMent. Automatic glottal inverse filtering from speech and electroglottographic signals. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 33(2):369 – 377, apr 1985. ISSN 0096-3518. doi: 10.1109/TASSP.1985.1164544.
- [143] Jan Vokřál. *Akustické parametry chraptivosti*. PhD thesis, České vysoké učení technické v Praze, Fakulta elektrotechnická, 1998.

- [144] Y. Wang and I. H. Witten. Induction of model trees for predicting continuous classes. In *Poster papers of the 9th European Conference on Machine Learning*. Springer, 1997.
- [145] C. Wendt and A. P. Petropulu. Pitch determination and speech segmentation using the discrete wavelet transform. *IEEE International Symposium on Circuits and Systems*, 2:45–48, 1996.
- [146] Katrin Werth, Daniel Voigt, Michael Döllinger, Ulrich Eysholdt, and Jörg LOHSCHELLER. Clinical value of acoustic voice measures: a retrospective study. *European Archives of Oto-Rhino-Laryngology*, 267(8):1261–1271, 2009.
- [147] Virginia Wolfe and David Martin. Acoustic correlates of dysphonia: type and severity. *Journal of Communication Disorders*, 30(5):403–416, 1997. doi: 10.1016/S0021-9924(96)00112-8.
- [148] Floris L. Wuyts, Marc S. De Bodt, and Van de Heyning Paul H. Is the reliability of a visual analog scale higher than an ordinal scale? an experiment with the grbas scale for the perceptual evaluation of dysphonia. *Journal of Voice*, 13(4):508–517, 1999. doi: DOI:10.1016/S0892-1997(99)80006-X.
- [149] Xiong Xiao, Chng Eng Siong, and Li Haizhou. Normalization of the speech modulation spectra for robust speech recognition. *Audio, Speech, and Language Processing, IEEE Transactions on*, 16(8):1662–1674, nov. 2008. ISSN 1558-7916. doi: 10.1109/TASL.2008.2002082.
- [150] B. Yegnanarayana, Christophe d’Alessandro, and Vassilis Darsinos. An iterative algorithm for decomposition of speech signals into periodic and aperiodic components. *IEEE Transactions on Speech and Audio Processing*, 6(1):1–11, 1998.
- [151] Ping Yu, Ouaknine Maurice, Revis Joana, and Giovanni Antoine. Objective voice analysis for dysphonic patients: A multiparametric protocol including acoustic and aerodynamic measurements. *Journal of Voice*, 15(4):529–542, 2001. doi: 10.1016/S0892-1997(01)00053-4. URL <http://www.sciencedirect.com/science/article/pii/S0892199701000534>.
- [152] Eiji Yumoto, Yumi Sasaki, and Hiroshi Okamura. Harmoics-to-noise ratio and physiological measurement of the degree of hoarseness. *JSHLR*, 27:2–6, 1984.

Seznam vlastních publikací

Práce vztahující se k tématu dizertace

Impaktované časopisy

- Stráník, A. - Čmejla, R. - Vokřál, J.: Acoustic parameters for classification of breathiness in continuous speech according to the GRBAS scale. In *Journal of Voice* (in press). [34%-33%-33%]

Recenzované časopisy

- Stráník, A. - Čmejla, R.: Popis spektrálních vlastností sibilantů /s/, /š/, /z/ a /ž/ na základě LSP parametrizace. *Akustické listy*. 2010, roč. 16, č. 4, s. 9-14. ISSN 1212-4702. [50%-50%]

WOS publikace

- Stráník, A. - Čmejla, R.: Spectral Approximations for Sibilant Classification. In *2011 International Conference on Applied Electronics*. Plzeň: Západočeská univerzita v Plzni, 2011, p. 385-388. ISBN 978-80-7043-987-6. [50%-50%]

Monografie

- Roubíčková, J. - Hedánek, J. - Stráník, A.: *Test 3F. Dysartrický profil*. 3. vyd. Praha: Galén, 2011. 86 s. ISBN 978-80-7262-714-1. [34%-33%-33%]

Vyzvané přednášky

- Stráník, A.: Vyštření dysartrie: Dysartrický profil - Test 3F. [Nepublikovaná přednáška]. Institut postgraduálního vzdělávání ve zdravotnictví. 2009-11-28. [100%]

Patenty

Ostatní publikace

- Stráník, A. - Čmejla, R. - Vokřál, J.: Hoarseness Assessment: Selection of Acoustical Parameters for Analysis of Sustained Vowel /a/ Phonation. In *Pan-European Voice Conference Prague 2013 (PEVOC)*. [34%-33%-33%]
- Stráník, A. - Čmejla, R.: Objektivizace chraptivosti - porovnání výpočtu HNR ve frekvenční oblasti s Praatem. In *LETNÍ DOKTORANDSKÉ DNY 2012*. Praha: ČVUT, 2012, s. 113-119. ISBN 978-80-01-05050-7. [50%-50%]

- Stráník, A. - Čmejla, R.: Parametrizace prodloužené fonace pro hodnocení chraptivosti. In *Novinky ve foniatrii*. Praha 5, Na bělidle 34, 150 00: Nakladatelství Galén, 2012, s. 126-128. ISBN 978-80-7262-940-4. [50%-50%]
- Stráník, A. - Čmejla, R.: Parametrizace průběhu křivky pro hodnocení chraptivosti. In *20th Annual Conference Proceeding's Technical Computing Bratislava 2012*. Praha: Humusoft, 2012, . ISBN 978-80-970519-4-5. [50%-50%]
- Stráník, A.: Comparison of Harmonics-to-noise Ratio Estimated in Frequency Domain with Praat. In *POSTER 2012 - 16th International Student Conference on Electrical Engineering*. Praha: Czech Technical University in Prague, 2012, p. 1-5. ISBN 978-80-01-05043-9. [100%]
- Stráník, A. - Čmejla, R.: Assessment of Hoarseness by Means of Analysis of Acoustic Signals. In *Czech-German Workshop on Speech Pathology and Biological Signals - Proceedings*. Prague: CTU, Faculty of Electrical Engineering, Department of Circuit Theory, 2012, p. 62-63. ISBN 978-80-01-05164-1. [50%-50%]
- Stráník, A. - Čmejla, R.: Hodnocení chraptivosti na základě analýzy akustického signálu řeči. In *Sborník 85. akustického semináře*. Praha: Nakladatelství ČVUT, 2012, s. 51-58. ISBN 978-80-01-05133-7. [50%-50%]
- Čmejla, R. - Rusz, J. - Bauer, L. - Lustyk, T. - Nejepsová, M. - et al.: Analýza patologického hlasu a řeči v laboratoři SAMI ČVUT. In *Novinky ve foniatrii*. Praha 5, Na bělidle 34, 150 00: Nakladatelství Galén, 2012, s. 28-30. ISBN 978-80-7262-940-4.
- Stráník, A. - Čmejla, R.: An Analysis of Iterative Algorithm for Estimation of Harmonics-To-Noise Ratio in Speech. In *19th Annual Conference Proceedings Technical Computing Prague 2011*. Technická 5, 16628 Praha: Vydavatelství VŠCHT Praha, 2011, p. 1-7. ISBN 978-80-7080-794-1. [50%-50%]
- Rusz, J. - Čmejla, R. - Stráník, A. - Janča, R.: Komplexní měření plicních funkcí s využitím spirometrie. In *19th Annual Conference Proceedings Technical Computing Prague 2011*. Technická 5, 16628 Praha: Vydavatelství VŠCHT Praha, 2011, díl 102, s. 1-5. ISBN 978-80-7080-794-1. [25%-25%-25%-25%]
- Rusz, J. - Čmejla, R. - Bartošek, J. - Janda, J. - Lustyk, T. - et al.: Assessment of voice and speech impairment. In *Workshop 2011, CTU Student Grant Competition in 2010 (SGS 2010)*. Praha: ČVTVS, 2011, p. 1-6.
- Stráník, A. - Čmejla, R.: Possibilities of Automated Assessment of /s/. In *Technical Computing Bratislava 2010*. Bratislava: RT systems, s.r.o, 2010, p. 1-5. ISBN 978-80-970519-0-7. [50%-50%]
- Stráník, A. - Čmejla, R.: Analysis of Fricative Consonant /s/ in Dysarthria Test. In *20th Czech - German Workshop on Speech Processing*. Prague: Institute of Photonics and Electronics AS CR, 2010, p. 58-69. ISBN 978-80-86269-21-4. [50%-50%]
- Stráník, A.: Assessment of Fricative Consonants on PC: Analysis of Recording Scheme. In *Krátky 2010*. Brno: Brno University of Technology, 2010, p. 146-149. ISBN 978-80-214-4139-2. [100%]
- Stráník, A.: Design of Real-time Signal Processing Framework in C#.NET. In *Digital Technologies 2010*. Žilina: TU v Žilině, 2010, p. 1-4. ISBN 978-80-554-

- 0304-5. [100%]
- Stráník, A.: Klasifikace mezi /s/ a /š/ na základě parametrizace vstupního signálu pomocí LSF. In *Analýza a zpracování řečových a biologických signálů - sborník prací 2010*. Praha: České vysoké učení technické v Praze, 2010, s. 92-98. ISBN 978-80-01-04680-7. [100%]
 - Stráník, A. - Čmejla, R.: Možnosti automatického hodnocení při syčení. In *8. ČESKO-SLOVENSKÝ FONIATRICKÝ KONGRES*. Bratislava: Samedi s.r.o., 2010, s. 25-26. ISSN 1337-2181. [50%-50%]
 - Stráník, A.: Návrh frameworku pro zpracování signálů v reálném čase v prostředí .NET. In *Analýza a zpracování řečových a biologických signálů - sborník prací 2009*. Praha: České vysoké učení technické v Praze, 2009, s. 82-89. ISBN 978-80-01-04474-2. [100%]

Citace

- Roubíčková, J. - Hedánek, J. - Stráník, A.: *Test 3F. Dysartrický profil*. 3. vyd. Praha: Galén, 2011. 86 s. ISBN 978-80-7262-714-1.
 - Košťálová, M. et al.: Test 3F Dysartrický profil – normativní hodnoty řeči v češtině. In *Česká a slovenská neurologie a neurochirurgie*. 2013, roč. 76, č. 5, s. 614-618. ISSN: 1210-7859; 1802-4041 (elektronická verze)
 - Eliasova, I. - Mekyska, M. - Kostalova, R. - Marecek, R. - Smekal, Z. - Rektorova, I.: Acoustic evaluation of short-term effects of repetitive transcranial magnetic stimulation on motor aspects of speech in Parkinson's disease. In *Journal of Neural Transmission*. 2013, 120(4), s. 597-605

Práce nevztahující se k tématu dizertace

Impaktované časopisy

Recenzované časopisy

WOS publikace

Ostatní publikace

9. Vzor formuláře pro hodnocení metodou CAPE-V

Hospital # _____ Date: _____

The following parameters of voice quality will be rated upon completion of the following tasks:

1. Sustained vowels /a/ and /i/ for 3-5 seconds duration each.
2. Sentence production:

The blue spot is on the key again.	We eat eggs every Easter.
How hard did he hit him	My mama makes lemon muffins.
We were away a year ago	Peter will keep at the peak.

3. Spontaneous speech. **Prompt as follows:**
 - a. **New Patient:** Tell me about when your voice problem began, what you were noticing, and what you have done about it.
 - b. **Return Patients:** Tell me what's happened with your voice since last time you were here. What treatment have you had? Did it help?

Legend: C =Consistent I = Intermittent MI = Mildly Deviant MO = Moderate Deviant SE = Severely Deviant

Overall Severity	_____	_____	_____	C I	/100
	MI	MO	SE		
Roughness	_____	_____	_____	C I	/100
	MI	MO	SE		
Breathiness	_____	_____	_____	C I	/100
	MI	MO	SE		
Strain	_____	_____	_____	C I	/100
	MI	MO	SE		
Pitch	(Indicate the nature of the abnormality): _____			C I	/100
	MI	MO	SE		
Loudness	(Indicate the nature of the abnormality): _____			C I	/100
	MI	MO	SE		
_____	_____	_____	_____	C I	/100
_____	_____	_____	_____	C I	/100

Comments about resonance: Normal Other (Provide description): _____

Additional Features (For example, diplophonia, fry, falsetto, asthenia, aphonia, pitch instability, tremor, wet/gurgly, or other relevant terms)

G ___ R ___ B ___ A ___ S ___

Clinician _____

Obrázek 9.1: Vzor formuláře pro subjektivní hodnocení nahrávek metodou CAPE-V.

10. Obsah databáze

Tabulka 10.1 shrnuje počty nahrávek pro jednotlivé úkoly.

10.1 Fonace

Pacient byl instruován, aby přečetl české samohlásky přirozeným hlasem – běžná hlasitost a výška hlasu. Poté byl požádán, aby fonoval hlásku /a/ jak nejdéle dokáže, opět s přirozenou intenzitou a výškou hlasu. Nakonec měl pacient vyzpívat co možná největší rozsah na hlásku /a/, případně /na/ a to od nejhlubšího tónu po nejvyšší tón a zase zpět k nejhlubšímu tónu.

10.2 Krátké věty

Pacient přečetl šest krátkých vět. Tyto věty jsou uvedeny v Tabulce 10.1.

10.3 Různé

Mezi různé byly zařazeny následující úkoly:

- čtení standardního textu *Podzim na starém bělidle* (text je uveden níže);
- zdůrazněný text – tato pasáž obsahuje dvakrát nahrávku textu *Podějte mi pásku*, nejprve normálním hlasem s přirozenou intenzitou a podruhé zakřičené s největší možnou intenzitou;
- spontánní promluva (nebyla vyjmuta);
- zpěv (nebyl vyjmut).

Podzim na Starém bělidle. V okolí Starého bělidla začínalo být smutno a ticho, les byl světlejší, stráž žloutla, vítr a vlny odnášely chomáče suchého listí bůh ví kam. Ozdoba sadu uschována byla v komoře. (34 slov)

Tabulka 10.1: Počty nahrávek pro jednotlivé úkoly. Označení *zdravý* a *patologický* je převzato z původní práce (Vokřál [143]), pro pozdější subjektivní hodnocení nebyl na toto označení brán zřetel.

úkol	zdravý	patologický	celkem
fonace			
/a/	72	397	469
/e/	71	397	468
/i/	71	400	471
/o/	71	398	469
/u/	71	396	467
výdrž /a/	174	188	362
rozsah /a/	167	91	258
krátké věty			
Vašek pil vodu	70	407	477
Fotografujeme růže	67	382	449
Pojedeme na výlet	2	61	63
Maminku bolí zub	67	383	450
Kočka nese čtyři myši	67	384	451
U lavice dítě stálo	67	379	446
různé			
Podzim	237	356	593
zdůrazněný text	9	100	109
celkem			6002

10.4 Vzor formuláře subjektivního hodnocení

Na 10.1 je zobrazen průvodní formulář k subjektivnímu hodnocení s definicí hodnocených pojmů. Obr. 10.2 obsahuje vzor formuláře pro subjektivní hodnocení nahrávek.

Subjektivní hodnocení hlásky /a/

Jedná se o vyjmutou hlásku /a/ za čtení všech samohlásek.

Na přiloženém CD disku jsou zvukové nahrávky. Každá nahrávka obsahuje jednu fonaci hlásky /a/ a poté dvě vteřiny ticha. V této době ticha se zapisují výsledky subjektivního hodnocení.

Hodnocení na škále GRBT

obecně:

0 – normální

1 – mírná

2 – značná

3 – extrémní, těžká

G – celkový dojem z hlasu

R – míra nepravidelnosti kmitání hlasivek či náhodných fluktuací

B – dyšnost, míra šumové příměsi v důsledku neúplného uzávěru glottis

T – napětí v hlase

-2 – značná hlasová slabost (insuficience, hypokinéza)

-1 – mírná hlasová slabost

0 – normální hlas

1 – mírné napětí v hlase

2 – největší přemáhání hlasu (hyperkinéza, spasticita)

Děkujeme za hodnocení

Obrázek 10.1: Průvodní formulář subjektivního hodnocení s definicí pojmů a vymezením hodnocení.

pořadí	G	R	B	T	pořadí	G	R	B	T	pořadí	G	R	B	T
1					43					85				
2					44					86				
3					45					87				
4					46					88				
5					47					89				
6					48					90				
7					49					91				
8					50					92				
9					51					93				
10					52					94				

Obrázek 10.2: Vzor formuláře pro subjektivní hodnocení nahrávek.

11. Subjektivní hodnocení databáze

Kontingenční tabulky shrnující subjektivní hodnocení prodloužené fonace hlásky /a/ a čtení standardního textu *Podzim na Starém bělidle* pro jednotlivé hodnotitele a pro skupinu hodnotitelů. U každé tabulky je uvedeno Cronbachovo alfa Cohenovo kappa.

Prodloužená fonace

Tabulka 11.1: Kontingenční tabulky hodnocení parametrů G, R, B, T pro sloučenou skupinu hodnotitelů pro prodlouženou fonaci hlásky /a/.

		G						R						B			
		2. kolo						2. kolo						2. kolo			
		0	1	2	3			0	1	2	3			0	1	2	3
1. kolo	0	269	185	7	1	1. kolo	0	290	252	18	2	1. kolo	0	736	188	24	2
	1	179	558	162	6		1	181	681	176	13		1	256	329	89	4
	2	16	228	342	77		2	22	241	277	45		2	52	177	224	42
	3	1	12	107	195		3	5	20	63	59		3	3	10	74	135
		Cronbachovo alfa: 0,84 Cohnenovo kappa: 0,41						Cronbachovo alfa: 0,73 Cohnenovo kappa: 0,33						Cronbachovo alfa: 0,82 Cohnenovo kappa: 0,43			

		T				
		2. kolo				
		-2	-1	0	1	2
1. kolo	-2	81	27	4	21	28
	-1	27	175	87	178	18
	0	0	51	563	188	8
	1	6	96	183	386	47
	2	13	15	9	60	74
		Cronbachovo alfa: 0,50 Cohnenovo kappa: 0,37				

Tabulka 11.2: Kontingenční tabulky hodnocení parametrů G, R, B, T pro hodnotitele A pro prodlouženou fonaci hlásky /a/.

		G						R						B			
		2. kolo						2. kolo						2. kolo			
		0	1	2	3			0	1	2	3			0	1	2	3
1. kolo	0	147	38	2	0	1. kolo	0	103	46	6	0	1. kolo	0	133	24	3	0
	1	70	49	23	1		1	55	116	19	0		1	68	64	13	0
	2	10	40	57	19		2	6	59	41	6		2	18	50	40	2
	3	0	0	3	10		3	0	6	4	2		3	2	5	12	35
		Cronbachovo alfa: 0,80 Cohnenovo kappa: 0,35						Cronbachovo alfa: 0,72 Cohnenovo kappa: 0,32						Cronbachovo alfa: 0,81 Cohnenovo kappa: 0,40			

		T				
		2. kolo				
		-2	-1	0	1	2
1. kolo	-2	21	8	4	3	5
	-1	1	14	37	23	2
	0	0	18	77	34	0
	1	3	19	60	82	8
	2	7	4	3	15	21
		Cronbachovo alfa: 0,50 Cohnenovo kappa: 0,26				

Tabulka 11.3: Kontingenční tabulky hodnocení parametrů G, R, B, T pro hodnotitele B pro prodlouženou fonaci hlásky /a/.

		G						R						B			
		2. kolo						2. kolo						2. kolo			
		0	1	2	3			0	1	2	3			0	1	2	3
1. kolo	0	36	42	0	0	1. kolo	0	76	53	4	2	1. kolo	0	211	42	3	1
	1	38	134	42	2		1	53	99	45	9		1	43	43	20	1
	2	0	33	59	16		2	11	27	41	8		2	4	18	24	12
	3	0	2	14	51		3	5	7	9	20		3	0	0	7	40
		Cronbachovo alfa: 0,84 Cohenovo kappa: 0,42						Cronbachovo alfa: 0,66 Cohenovo kappa: 0,28						Cronbachovo alfa: 0,85 Cohenovo kappa: 0,48			

		T				
		2. kolo				
		-2	-1	0	1	2
1. kolo	-2	33	5	0	5	5
	-1	12	54	21	85	9
	0	0	11	35	33	0
	1	1	26	17	81	9
	2	5	1	0	2	19
		Cronbachovo alfa: 0,52 Cohenovo kappa: 0,30				

Tabulka 11.4: Kontingenční tabulky hodnocení parametrů G, R, B, T pro hodnotitele C pro prodlouženou fonaci hlásky /a/.

		G						R						B			
		2. kolo						2. kolo						2. kolo			
		0	1	2	3			0	1	2	3			0	1	2	3
1. kolo	0	44	55	4	1	1. kolo	0	39	85	5	0	1. kolo	0	220	65	10	0
	1	26	124	35	2		1	16	184	27	1		1	38	45	15	0
	2	6	59	54	19		2	4	50	49	1		2	15	20	21	2
	3	1	6	23	19		3	0	2	4	2		3	0	3	15	0
		Cronbachovo alfa: 0,72 Cohenovo kappa: 0,29						Cronbachovo alfa: 0,64 Cohenovo kappa: 0,30						Cronbachovo alfa: 0,67 Cohenovo kappa: 0,30			

		T				
		2. kolo				
		-2	-1	0	1	2
1. kolo	-2	0	0	0	6	10
	-1	0	3	6	32	5
	0	0	3	93	65	5
	1	0	6	68	103	13
	2	0	4	5	29	13
		Cronbachovo alfa: 0,10 Cohenovo kappa: 0,16				

Kapitola 11. Subjektivní hodnocení databáze

Tabulka 11.5: Kontingenční tabulky hodnocení parametrů G, R, B, T pro hodnotitele D pro prodlouženou fonaci hlásky /a/.

G					R					B							
		2. kolo						2. kolo						2. kolo			
		0	1	2	3			0	1	2	3			0	1	2	3
1. kolo	0	26	28	1	0	1. kolo	0	34	38	2	0	1. kolo	0	80	29	7	1
	1	23	156	24	0		1	29	143	40	3		1	54	84	21	3
	2	0	56	49	17		2	1	47	60	21		2	12	46	49	22
	3	0	3	34	52		3	0	3	22	26		3	1	2	21	37
Cronbachovo alfa: 0,85 Cohnenovo kappa: 0,41					Cronbachovo alfa: 0,79 Cohnenovo kappa: 0,34					Cronbachovo alfa: 0,80 Cohnenovo kappa: 0,36							

T						
		2. kolo				
		-2	-1	0	1	2
1. kolo	-2	27	14	0	7	8
	-1	14	103	13	35	2
	0	0	12	23	9	0
	1	2	43	15	97	14
	2	1	5	0	7	18
Cronbachovo alfa: 0,65 Cohnenovo kappa: 0,40						

Tabulka 11.6: Kontingenční tabulky hodnocení parametrů G, R, B, T pro hodnotitele E pro prodlouženou fonaci hlásky /a/.

G					R					B							
		2. kolo						2. kolo						2. kolo			
		0	1	2	3			0	1	2	3			0	1	2	3
1. kolo	0	16	22	0	0	1. kolo	0	38	30	1	0	1. kolo	0	92	28	1	0
	1	22	95	38	1		1	28	139	45	0		1	53	93	20	0
	2	0	40	123	15		2	0	58	86	9		2	3	43	90	4
	3	0	1	33	63		3	0	2	24	9		3	0	0	19	23
Cronbachovo alfa: 0,86 Cohnenovo kappa: 0,47					Cronbachovo alfa: 0,78 Cohnenovo kappa: 0,35					Cronbachovo alfa: 0,87 Cohnenovo kappa: 0,48							

T						
		2. kolo				
		-2	-1	0	1	2
1. kolo	-2	0	0	0	7	8
	-1	0	1	10	3	0
	0	0	7	335	47	3
	1	0	2	23	23	3
	2	0	1	1	7	3
Cronbachovo alfa: 0,48 Cohnenovo kappa: 0,29						

Čtený text

Tabulka 11.7: Kontingenční tabulky hodnocení parametrů G, R, B, T pro sloučenou skupinu hodnotitelů pro čtený text *Podzim na Starém bělidle*.

		G				R				B							
		2. kolo				2. kolo				2. kolo							
		0	1	2	3			0	1	2	3			0	1	2	3
1. kolo	0	761	233	9	1	1. kolo	0	703	223	6	3	1. kolo	0	1125	248	26	4
	1	363	446	188	9		1	389	636	196	6		1	372	406	141	13
	2	19	178	369	108		2	17	227	362	75		2	6	168	212	62
	3	0	4	112	165		3	2	10	49	61		3	0	7	54	121
		Cronbachovo alfa: 0,86 Cohnenovo kappa: 0,42				Cronbachovo alfa: 0,81 Cohnenovo kappa: 0,40				Cronbachovo alfa: 0,82 Cohnenovo kappa: 0,43							

		T				
		2. kolo				
		-2	-1	0	1	2
1. kolo	-2	50	15	1	13	34
	-1	21	199	59	75	24
	0	10	164	1128	110	4
	1	9	131	296	367	97
	2	7	4	2	27	118
		Cronbachovo alfa: 0,58 Cohnenovo kappa: 0,45				

Tabulka 11.8: Kontingenční tabulky hodnocení parametrů G, R, B, T pro hodnotitele A pro čtený text *Podzim na Starém bělidle*.

		G				R				B							
		2. kolo				2. kolo				2. kolo							
		0	1	2	3			0	1	2	3			0	1	2	3
1. kolo	0	238	22	1	0	1. kolo	0	247	32	0	0	1. kolo	0	324	17	0	0
	1	11	104	19	0		1	16	141	23	0		1	26	82	9	0
	2	0	34	126	5		2	0	49	79	1		2	1	66	53	3
	3	0	0	13	20		3	0	0	4	1		3	0	0	5	7
		Cronbachovo alfa: 0,95 Cohnenovo kappa: 0,74				Cronbachovo alfa: 0,92 Cohnenovo kappa: 0,67				Cronbachovo alfa: 0,93 Cohnenovo kappa: 0,63							

		T				
		2. kolo				
		-2	-1	0	1	2
1. kolo	-2	0	1	0	0	0
	-1	0	36	32	1	0
	0	0	20	479	2	0
	1	0	0	9	8	0
	2	0	0	0	4	1
		Cronbachovo alfa: 0,73 Cohnenovo kappa: 0,53				

Tabulka 11.9: Kontingenční tabulky hodnocení parametrů G, R, B, T pro hodnotitele B pro čtený text *Podzim na Starém bělidle*.

		G				R					B						
		2. kolo				2. kolo					2. kolo						
		0	1	2	3						0	1	2	3			
1. kolo	0	65	0	0	0	1. kolo	0	31	0	0	0	1. kolo	0	105	0	0	0
	1	251	32	1	0		1	250	88	2	0		1	230	79	6	0
	2	16	92	47	0		2	13	108	61	3		2	3	56	64	8
	3	0	4	76	9		3	1	5	22	9		3	0	1	17	24
		Cronbachovo alfa: 0,91 Cohenovo kappa: 0,05				Cronbachovo alfa: 0,85 Cohenovo kappa: 0,07					Cronbachovo alfa: 0,87 Cohenovo kappa: 0,27						

		T				
		2. kolo				
		-2	-1	0	1	2
1. kolo	-2	0	0	0	0	0
	-1	5	2	0	0	1
	0	10	63	186	16	2
	1	5	27	169	50	37
	2	1	0	0	2	17
		Cronbachovo alfa: 0,58 Cohenovo kappa: 0,14				

Tabulka 11.10: Kontingenční tabulky hodnocení parametrů G, R, B, T pro hodnotitele C pro čtený text *Podzim na Starém bělidle*.

		G				R					B						
		2. kolo				2. kolo					2. kolo						
		0	1	2	3						0	1	2	3			
1. kolo	0	100	84	2	0	1. kolo	0	47	67	1	0	1. kolo	0	119	92	14	1
	1	26	98	64	3		1	50	163	61	4		1	39	86	42	8
	2	3	13	76	38		2	3	31	78	48		2	1	13	58	25
	3	0	0	10	76		3	0	3	9	28		3	0	3	23	69
		Cronbachovo alfa: 0,90 Cohenovo kappa: 0,45				Cronbachovo alfa: 0,80 Cohenovo kappa: 0,31					Cronbachovo alfa: 0,85 Cohenovo kappa: 0,40						

		T				
		2. kolo				
		-2	-1	0	1	2
1. kolo	-2	34	9	0	9	22
	-1	12	126	15	28	6
	0	0	40	35	16	0
	1	3	62	32	104	9
	2	1	1	1	2	26
			Cronbachovo alfa: 0,56 Cohenovo kappa: 0,39			

Tabulka 11.11: Kontingenční tabulky hodnocení parametrů G, R, B, T pro hodnotitele D pro čtený text *Podzim na Starém bělidle*.

		G						R						B			
		2. kolo						2. kolo						2. kolo			
		0	1	2	3			0	1	2	3			0	1	2	3
1. kolo	0	181	23	0	0	1. kolo	0	197	25	2	3	1. kolo	0	297	35	0	0
	1	70	144	20	0		1	65	136	27	1		1	70	99	11	1
	2	0	35	58	13		2	1	34	49	15		2	1	31	11	6
	3	0	0	11	38		3	1	2	13	22		3	0	3	8	20
		Cronbachovo alfa: 0,90 Cohnenovo kappa: 0,58						Cronbachovo alfa: 0,86 Cohnenovo kappa: 0,53						Cronbachovo alfa: 0,83 Cohnenovo kappa: 0,50			

		T				
		2. kolo				
		-2	-1	0	1	2
1. kolo	-2	13	4	1	3	6
	-1	1	5	4	14	5
	0	0	2	222	23	0
	1	0	4	81	146	8
	2	2	3	0	16	30
		Cronbachovo alfa: 0,66 Cohnenovo kappa: 0,53				

Tabulka 11.12: Kontingenční tabulky hodnocení parametrů G, R, B, T pro hodnotitele E pro čtený text *Podzim na Starém bělidle*.

		G						R						B			
		2. kolo						2. kolo						2. kolo			
		0	1	2	3			0	1	2	3			0	1	2	3
1. kolo	0	177	104	6	1	1. kolo	0	181	99	3	0	1. kolo	0	280	104	12	3
	1	5	68	84	6		1	8	108	83	1		1	7	60	73	4
	2	0	4	62	52		2	0	5	95	8		2	0	2	26	20
	3	0	0	2	22		3	0	0	1	1		3	0	0	1	1
		Cronbachovo alfa: 0,91 Cohnenovo kappa: 0,37						Cronbachovo alfa: 0,89 Cohnenovo kappa: 0,48						Cronbachovo alfa: 0,85 Cohnenovo kappa: 0,36			

		T				
		2. kolo				
		-2	-1	0	1	2
1. kolo	-2	3	1	0	1	6
	-1	3	30	8	32	12
	0	0	39	206	53	2
	1	1	38	5	59	43
	2	3	0	1	3	44
		Cronbachovo alfa: 0,51 Cohnenovo kappa: 0,50				

12. Porovnání subjektivního a objektivního hodnocení

Tabulka 12.1: Vážený průměr hodnot popisujících diskrétní klasifikaci. Porovnání výsledků subjektivního a objektivního hodnocení pomocí senzitivity (SEN) a specificity (SPE) a metrik TP, FP, TN, FN.

		objektivní						subjektivní					
		TP	FN	FP	TN	SEN	SPE	TP	FN	FP	TN	SEN	SPE
<i>/a/</i>	G	0,66	0,34	0,21	0,79	0,66	0,79	0,58	0,42	0,19	0,81	0,58	0,81
	R	0,64	0,36	0,23	0,77	0,64	0,77	0,56	0,44	0,24	0,76	0,56	0,76
	B	0,75	0,25	0,15	0,85	0,75	0,85	0,61	0,39	0,18	0,82	0,61	0,82
	T	0,52	0,48	0,21	0,79	0,52	0,79	0,55	0,45	0,17	0,83	0,55	0,83
<i>text</i>	G	0,67	0,33	0,17	0,83	0,69	0,83	0,59	0,41	0,17	0,83	0,59	0,83
	R	0,67	0,33	0,18	0,82	0,67	0,82	0,59	0,41	0,20	0,80	0,59	0,80
	B	0,77	0,23	0,15	0,85	0,77	0,85	0,63	0,37	0,20	0,80	0,63	0,81
	T	0,65	0,35	0,18	0,82	0,65	0,82	0,63	0,37	0,16	0,84	0,63	0,84

Tabulka 12.2: Detailní přehled hodnot popisujících diskrétní klasifikaci. Porovnání výsledků subjektivního a objektivního hodnocení pomocí senzitivity (SEN) a specificity (SPE) a metrik TP, FP, TN, FN.

		objektivní						subjektivní						
		TP	FN	FP	TN	SEN	SPE	TP	FN	FP	TN	SEN	SPE	
/a/	G	0	0,47	0,53	0,10	0,90	0,47	0,90	0,58	0,42	0,10	0,90	0,58	0,90
		1	0,79	0,21	0,35	0,65	0,79	0,65	0,62	0,38	0,30	0,70	0,62	0,71
		2	0,58	0,42	0,08	0,92	0,58	0,92	0,52	0,48	0,16	0,84	0,52	0,84
		3	0,58	0,42	0,07	0,92	0,58	0,98	0,62	0,38	0,04	0,96	0,62	0,96
	R	0	0,43	0,57	0,06	0,94	0,43	0,94	0,52	0,48	0,12	0,88	0,52	0,88
		1	0,83	0,17	0,36	0,61	0,83	0,61	0,65	0,35	0,40	0,60	0,65	0,60
		2	0,51	0,49	0,06	0,95	0,51	0,95	0,47	0,53	0,15	0,85	0,47	0,85
		3	0,44	0,56	0,07	0,93	0,44	0,93	0,40	0,60	0,03	0,97	0,40	0,97
	B	0	0,92	0,08	0,22	0,79	0,92	0,79	0,78	0,22	0,22	0,78	0,78	0,78
		1	0,58	0,42	0,17	0,88	0,58	0,88	0,49	0,51	0,23	0,77	0,49	0,78
		2	0,46	0,54	0,05	0,96	0,46	0,96	0,45	0,55	0,10	0,90	0,45	0,90
		3	0,69	0,31	0,03	0,97	0,69	0,97	0,61	0,39	0,02	0,98	0,61	0,98
T	-2	0,57	0,43	0,02	0,98	0,57	0,98	0,50	0,50	0,02	0,98	0,50	0,98	
	-1	0,31	0,69	0,09	0,91	0,31	0,91	0,36	0,64	0,10	0,90	0,36	0,90	
	0	0,75	0,25	0,32	0,68	0,75	0,68	0,70	0,30	0,18	0,82	0,70	0,82	
	1	0,36	0,65	0,22	0,78	0,36	0,78	0,54	0,46	0,28	0,72	0,54	0,73	
	2	0,36	0,64	0,04	0,96	0,36	0,96	0,43	0,58	0,05	0,95	0,43	0,95	
text	G	0	0,91	0,09	0,28	0,73	0,91	0,73	0,76	0,24	0,20	0,82	0,76	0,81
		1	0,39	0,61	0,14	0,86	0,39	0,86	0,44	0,56	0,21	0,79	0,44	0,79
		2	0,57	0,44	0,07	0,93	0,57	0,93	0,55	0,45	0,14	0,86	0,55	0,87
		3	0,60	0,40	0,20	0,98	0,60	0,98	0,59	0,41	0,04	0,96	0,59	0,96
	R	0	0,94	0,16	0,25	0,75	0,84	0,75	0,75	0,25	0,20	0,80	0,75	0,80
		1	0,42	0,58	0,15	0,85	0,42	0,85	0,52	0,48	0,27	0,74	0,52	0,74
		2	0,79	0,21	0,10	0,90	0,79	0,90	0,53	0,47	0,11	0,89	0,53	0,89
		3	0,11	0,89	0,00	1,00	0,11	1,00	0,50	0,50	0,03	0,97	0,50	0,97
	B	0	0,93	0,07	0,20	0,80	0,93	0,80	0,80	0,20	0,24	0,76	0,80	0,76
		1	0,49	0,51	0,10	0,90	0,49	0,90	0,44	0,56	0,21	0,79	0,44	0,79
		2	0,52	0,89	0,06	0,94	0,52	0,94	0,47	0,53	0,09	0,91	0,47	0,91
		3	0,80	0,20	0,02	0,98	0,80	0,98	0,67	0,34	0,03	0,97	0,67	0,97
T	-2	0,41	0,59	0,03	0,97	0,41	0,97	0,44	0,56	0,02	0,98	0,44	0,98	
	-1	0,39	0,60	0,06	0,94	0,38	0,94	0,51	0,49	0,12	0,88	0,51	0,88	
	0	0,91	0,09	0,28	0,73	0,91	0,73	0,80	0,20	0,23	0,77	0,80	0,77	
	1	0,34	0,66	0,12	0,88	0,34	0,88	0,41	0,59	0,11	0,89	0,41	0,89	
	2	0,36	0,64	0,05	0,95	0,36	0,95	0,75	0,25	0,06	0,94	0,75	0,94	