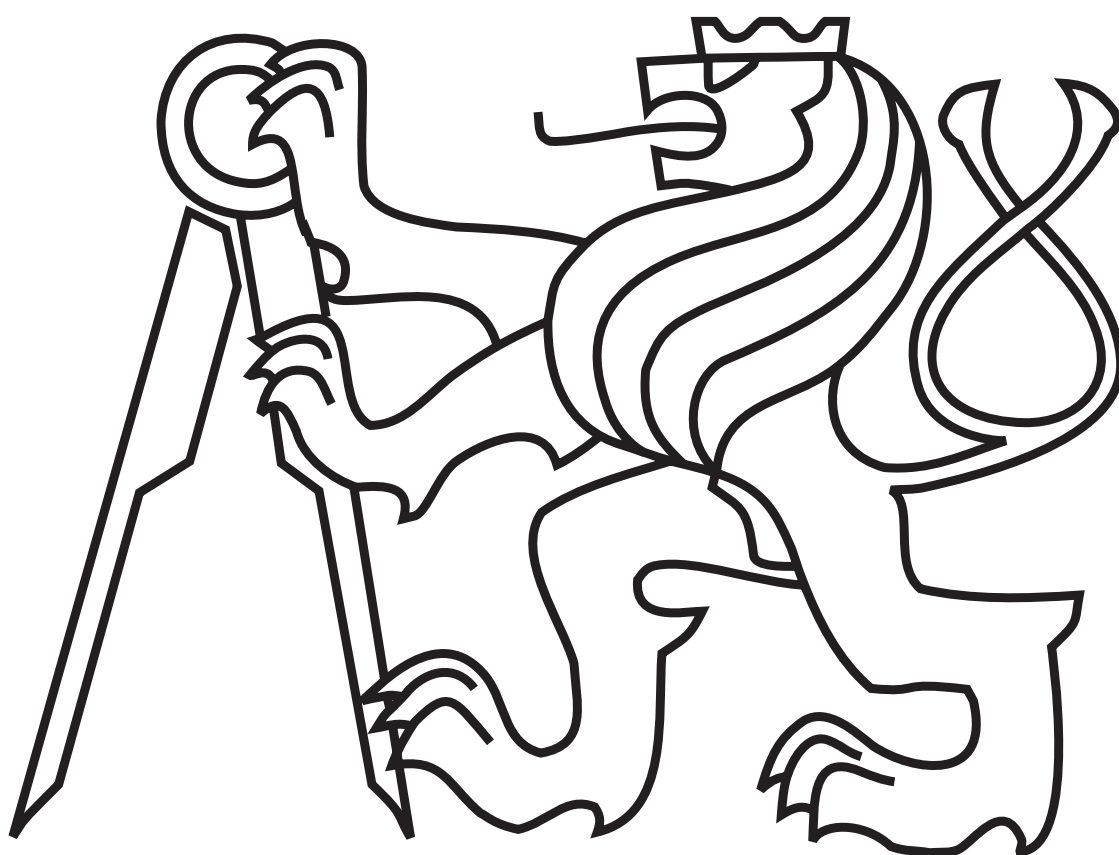


**ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE**



**TEZE K DIZERTAČNÍ PRÁCI**

Dizertační práce byla vypracována v prezenční formě doktorského studia na katedře Teorie obvodů Fakulty elektrotechnické ČVUT v Praze.

Uchazeč: Adam Stráník  
Katedra teorie obvodů  
České vysoké učení technické v Praze  
Technická 2, Praha 6, 166 27

Školitel: doc. Ing. Roman Čmejla, CSc.  
Katedra teorie obvodů  
České vysoké učení technické v Praze  
Technická 2, Praha 6, 166 27

Oponenti:

Teze byly rozeslány dne:

Obhajoba dizertace se koná dne ..... v ..... hod. před komisí pro obhajobu dizertační práce ve studijním oboru Teoretická elektrotechnika v zasedací místnosti č. .... Fakulty elektrotechnické ČVUT v Praze.

S disertací je možno se seznámit na děkanátu Fakulty elektrotechnické ČVUT v Praze, na oddělení pro vědu, výzkum a zahraniční styky, Technická 2, Praha 6.

předseda komise pro obhajobu dizertační práce  
ve studijním oboru  
Teoretická elektrotechnika  
Fakulta elektrotechnická ČVUT, Technická 2, Praha 6

České vysoké učení technické v Praze  
Fakulta elektrotechnická  
Katedra teorie obvodů

**Adam Stráník**

**ANALÝZA A HODNOCENÍ CHRAPTIVOSTI V ŘEČOVÉM  
SIGNÁLU**

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

Teze dizertace k získání titulu „doktor“, ve zkratce „Ph.D.“

Praha, listopad 2013

# Obsah

<b>1</b>	<b>Stav zkoumané problematiky</b>	<b>1</b>
1.1	Subjektivní metody hodnocení chraptivosti . . . . .	1
1.2	Objektivní metody hodnocení chraptivosti . . . . .	3
1.2.1	Kontaktní metody . . . . .	3
1.2.2	Bezkontaktní metody . . . . .	3
1.3	Automatické hodnocení chraptivosti . . . . .	3
<b>2</b>	<b>Cíle práce</b>	<b>5</b>
<b>3</b>	<b>Metodika</b>	<b>6</b>
3.1	Databáze promluv . . . . .	6
3.2	Subjektivní hodnocení . . . . .	6
3.3	Akustické parametry . . . . .	8
3.4	Popisy průběhů parametrů . . . . .	10
3.5	Dolování dat . . . . .	13
<b>4</b>	<b>Výsledky</b>	<b>14</b>
4.1	Subjektivní hodnocení . . . . .	14
4.2	Automatické hodnocení chraptivosti . . . . .	15
<b>5</b>	<b>Závěry a další cíle</b>	<b>20</b>
5.1	Závěry pro další rozvoj hodnocení chraptivosti . . . . .	21
	<b>Literatura</b>	<b>22</b>
	<b>Seznam vlastních publikací</b>	<b>27</b>
	<b>Summary</b>	<b>30</b>
	<b>Resumé</b>	<b>31</b>

# 1. Stav zkoumané problematiky

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 přešel vážnějším zdravotním komplikacím, ale podle výzkumu uveřejněného v Amir and Levine-Yundof [3] 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.

Příčiny vzniku chraptivosti můžeme rozdělit do následujících skupin (Vokřál [34]):

- **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**.

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.

## 1.1 Subjektivní metody hodnocení chraptivosti

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 zkušenosti testující jsou ve svém hodnocení značně konzistentní, viz např. De Bodt et al. [6] nebo Karnell et al. [17]. V uvedených publikacích dosahují hodnotící velmi podobných výsledků. Ze závěrů v De Bodt et al. [6] a Karnell et al. [17] 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 [34].

Při subjektivním posouzení chraptivého hlasu jsou nejčastěji posuzovány následující parametry (Vokřál [34]):

- 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

Metoda GRBAS (Hirano [15]) 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 – chraplavost<sup>1</sup>, míra nepravidelnosti kmitání hlasivek
- B** reathiness – dyšnost, míra šumové příměsi v hlase
- A** esthenticity – 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. [1], Alpan et al. [2], Bhuta et al. [4], De Bodt et al. [6], Fredouille et al. [9], Godino-Llorente et al. [11], Godino-Llorente et al. [12], Lee and Hahn [20], Ma and Yiu [21], Maryn et al. [23], Saenz-Lechon et al. [30], Yu et al. [36]) a dále analyzují spolehlivost jednotlivých parametrů. Ve studiích De Bodt et al. [6] a Karnell et al. [17] shledali parametr G nejvíce spolehlivým a parametry A a S nejméně spolehlivé. Oproti tomu závěry uveřejněné v Wuyts et al. [35] ukazují parametr A jako druhý nejspolehlivější po parametru G. Dále De Bodt et al. [6] 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. [11] 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).

---

<sup>1</sup>V 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.

**Modifikace GRBT** byla zavedena v práci Dejonckere et al. [7], 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 [34] a je dále používána v této práci

## 1.2 Objektivní metody hodnocení chraptivosti

### 1.2.1 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.

### 1.2.2 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.

## 1.3 Automatické hodnocení chraptivosti

Existuje několik publikací, ve kterých byla provedena klasifikace celkového dojmu z hlasu G podle stupnice GRBAS (Alpan et al. [1], Fredouille et al. [9], Lee and Hahn [20], Ma and Yiu [21]), avšak pouze jediná studie zahrnuje všechny sledované aspekty chraptivosti, tzn. i R, B, A a S (Saenz-Lechon et al. [30]).

Alpan et al. [1] 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. [9] 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 [20] 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 [21] 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.

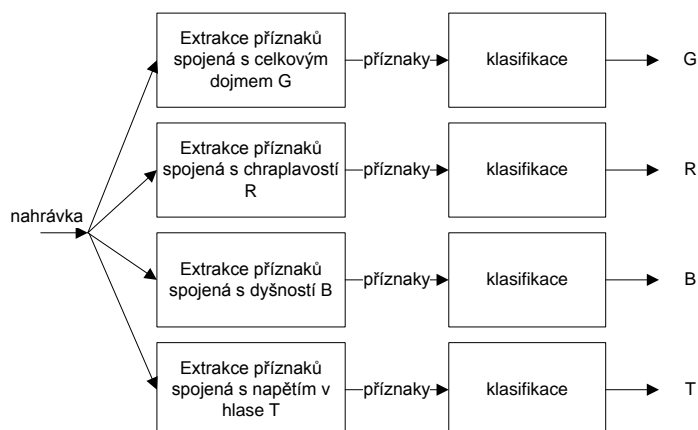
Saenz-Lechon et al. [30] navrhl automatický systém pro objektivní hodnocení všech parametrů ze stupnice GRBAS. Systém používal MFCC a algoritmus Learning Vector Quantization (LVQ). V experimentu bylo použito 648 nahrávek obsahujících prodlouženou fonaci hlásky /a/ a krátkou větu. Databáze nahrávek byla ohodnocena třemi experty pomocí metody GRBAS. Navržený systém dosahoval přesnosti okolo 65 % (od 55 % pro parametr A po 68 % pro parametr G).



## 2. Cíle práce

Na základě provedených rešerší 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.
- **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.
- **Sestavit experimentální systém pro objektivní hodnocení chraptivosti.** Zamýšlený experimentální systém pro objektivní hodnocení chraptivosti je zobrazen na obr. 2.1, ze kterého je patrné, že pro každý sledovaný aspekt chraptivosti (chraptivost, chraplavost, dyšnost, napětí v hlasu) je nutné vybrat unikátní sadu parametrů a sestavit vhodný klasifikátor, jehož výstup bude co možná nejlépe korelovat se subjektivním hodnocením odborníků.



Obrázek 2.1: Blokové schéma zamýšleného systému pro automatické hodnocení chraptivosti.

## 3. Metodika

### 3.1 Databáze promluv

V rámci práce byla použita část databáze nahrávek pacientů pořizovaná na Foniatrické 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 [34].

Nahrávky obsahovaly dlouhé fonace vokálů /a/, /e/, /i/, /o/ a /u/, krátké věty, standardní čtený text *Podzim na Starém bělidle* a volné vyprávění.

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 [28]. Pro další analýzy byl stereo signál převeden na mono.

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).

### 3.2 Subjektivní hodnocení

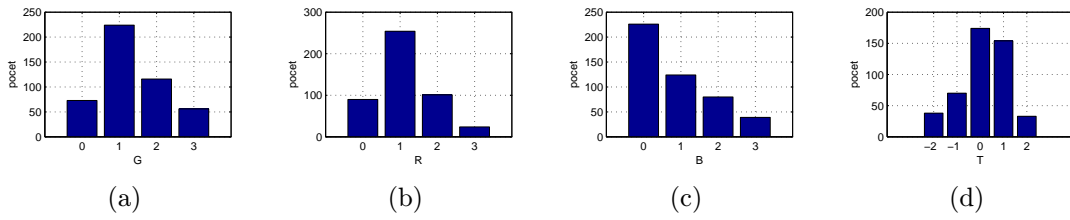
Subjektivní hodnocení bylo provedeno pěti odborníky z Foniatrické kliniky 1. LF UK a VFN v Praze. Obě databáze byly hodnoceny odděleně a každá z databází byla hodnocena dvakrát s alespoň dvoutýdenním odstupem. Pro subjektivní hodnocení byly nahrávky očíslovány pořadovými čísly a tato pořadová čísla byla pro opakované hodnocení změněna.<sup>1</sup>

Pro natrénování klasifikátorů je nutné mít každou nahrávku ohodnocenou jistým „zlatým standardem,“ tzn. tak, jak by danou nahrávku ohodnotil průměrný odborník.<sup>2</sup> Byly zavedeny dva přístupy pro získání tohoto zlatého standardu: diskrétní hodnocení nahrávky dané jako modus hodnocení pro danou nahrávku a spojitě hodnocení nahrávky, které je dané aritmetickým průměrem hodnocení.

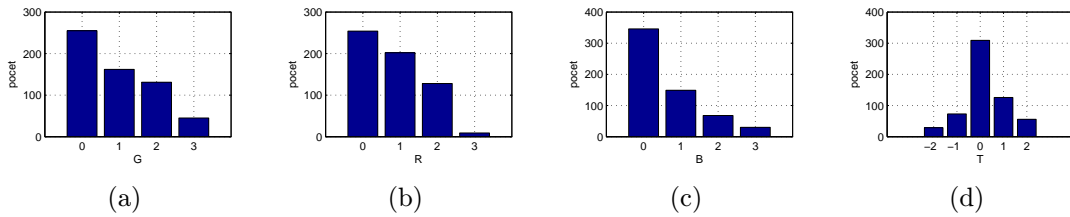
Při použití diskrétního hodnocení bude výsledné hodnocení ve stejném stylu, jaké použili odborníci, tzn. výsledná nahrávka bude hodnocena např.  $G_1R_0B_2T_{-1}$ .

<sup>1</sup>Pro vygenerování pořadových čísel nahrávek byla použita funkce *randperm* v MATLABu.

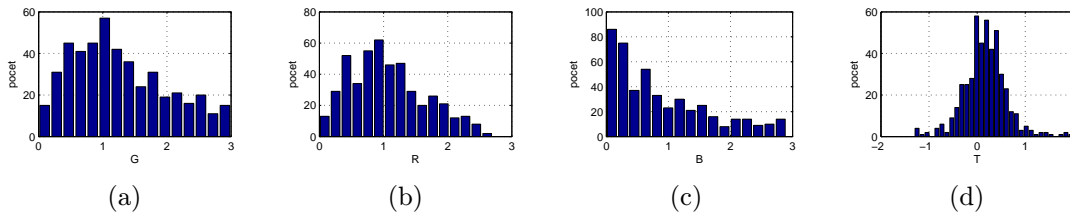
<sup>2</sup>Průměrný odborník není myšleno z pohledu odbornosti, ale z pohledu množství.



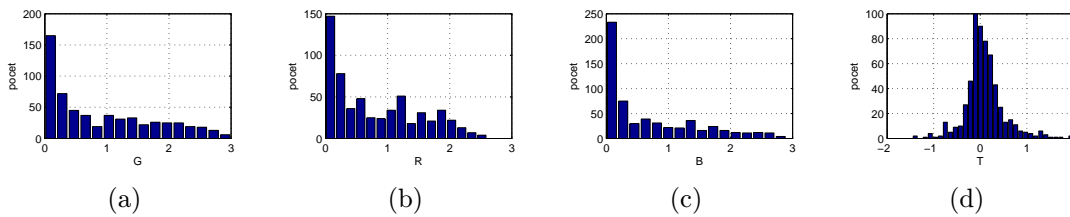
Obrázek 3.1: Histogramy diskrétního hodnocení nahrávek prodloužené fonace meto-  
dou GRBT, (a) G, (b) R, (c) B, (d) T.



Obrázek 3.2: Histogramy diskrétního hodnocení nahrávek čteného textu meto-  
dou GRBT, (a) G, (b) R, (c) B, (d) T.



Obrázek 3.3: Histogramy spojitého hodnocení nahrávek prodloužené fonace meto-  
dou GRBT, (a) G, (b) R, (c) B, (d) T.



Obrázek 3.4: Histogramy spojitého hodnocení nahrávek čteného textu meto-  
dou 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.

Na obr. 3.1 a obr. 3.2 je zobrazeno rozložení výsledného hodnocení nahrávek podle diskrétní stupnice a na obr. 3.3 a obr. 3.4 podle spojitě stupnice.

### 3.3 Akustické parametry

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 ([24]) a následně využity v experimentální části práce.

**Analýza autokorelační funkce** je zastoupena parametry RPK (Pearson  $r$  at autocorrelation peak, Hillenbrand and Houde [14]) a SPS (Signal Periodicity Strength). Oba parametry hledají maximum v autokorelační funkci, které popisuje míru periodicity signálu. Parametr RPK nabývá hodnoty vždy – jedná se o maximum autokorelační funkce v rozmezí 3,3 až 16,7 ms. Parametr SPS nabývá nenulové hodnoty, pokud leží maximum signálu určené parametrem RPK za globálním minimem autokorelační funkce. Podle předpokladu mají zdravé hlasy dobře definovanou strukturu a tím pádem budou parametry HLR a SPS nabývat vyšších hodnot.

**Index dyšnosti BRI** (Hillenbrand and Houde [14]) dává do poměru energii druhé derivace časového průběhu signálu a energii originálního signálu. Podobně pracuje i parametr HLR (High- to mid/low frequency energy ratio, Hillenbrand and Houde [14]), který dává do poměru energii v pásmu 0 až 4 kHz ku energii v pásmu 4 až  $f_s/2$  kHz. Podle předpokladu zdravé hlasy neobsahují energii na vyšších frekvencích a tím pádem budou dosahovat jiných poměrů energií.

**Výraznost keprstrální špičky CPP** (Cepstral Peak Prominence, Hillenbrand and Houde [14]) popisuje výraznost špičky v keprstru signálu. Keprstrum signálu vzniká nelineární transformací signálu – díky aplikaci logaritmu na spektrum signálu dojde k oddělení přenosu (obraz hlasového traktu) a buzení signálu (hlasivkové pulzy). Zdravý hlas s dobře definovanou hlasivkovou strukturou tak bude v keprstru obsahovat výraznější špičku, která odpovídá periodickému hlasivkovému buzení. Parametr CPP vyjadřuje vzdálenost této keprstrální špičky od regresní přímky, která keprstrum prokládá.

**Poměr hlasivkového a šumového signálu GNE** (Glottal-to-Noise Excitation Ratio, Michaelis et al. [26]) určuje, do jaké míry je analyzovaný signál vybuzený hlasivkami. Princip výpočtu je následující: vstupní řečový signál je inverzně filtrován (tím dojde ke zploštění spektra signálu), je spočteno spektrum signálu, které je následně pásmově filtrováno. Jednotlivé pásmově filtrované signály jsou převedeny zpět do časové oblasti a je spočtena jejich Hilbertova obálka. Parametr GNE pak určuje nejlepší korelaci mezi těmito Hilbertovými obálkami. Algoritmus využívá faktu, že v inverzně filtrovaném signálu zdravého hlasu s dobře definovanými hlasivkovými pulzy, vzniknou v místě těchto pulzů „Diracovy pulzy“ (nikoliv dokonalé, ale poměrně úzké strmé špičky). Tyto špičky jsou schopné vybudit celé spektrum signálu a tím pádem jsou si jednotlivé Hilbertovy obálky po pásmové filtraci velmi podobné. Patologický hlas, který není dobře vybuzen hlasivkami, tyto pulzy neobsahuje a tím pádem není spektrum signálu vybuzeno rovnoměrně. Výsledné pásmově filtrované Hilbertovy obálky signálu si tím pádem podobné nejsou a hlas dosahuje nižších hodnot parametru GNE.

**Analýza generalizovaného histogramu GVA** (Generalized Variogram analysis, Kacha et al. [16]) je parametrizace, která využívá faktu, že periodický signál

lze popsat vztahem  $x(t) = x(t + kT_0)$ ,  $k \in \mathbb{Z}$  a  $x(t)$  je signál s periodou  $T_0$ . Parametr GVA hledá minimální vzdálenost (tzn. maximální podobnost) mezi krátkým okénkem signálu (typicky 5 ms) a okolím tohoto okénka (typicky  $\pm 20$  ms). Zvolený rozsah je schopný nalézt signály s hlasivkovou frekvencí  $f_0$  v rozmezí 50 až 400 Hz.

**Poměr amplitud harmonických složek H1A** byl popsán v práci Hillenbrand and Houde [14]. Motivací k zavedení tohoto popisu je fakt, že hlasy vnímané jako dyšené mají časový průběh podobnější funkci sinus než hlasy zdravé. Díky této podobnosti s funkcí sinus mají výrazně silnější 1. harmonickou ve spektru než hlasy, které vnímané jako dyšené nejsou.

**Poměr energie harmonické a šumové složky HNR** (Harmonics-to-Noise Ratio) je velmi starý popis akustického signálu. Opět využívá faktu, že patologické hlasy mají větší zastoupení energie v šumové části signálu. V této práci je implementován algoritmus, který iterativně ve spektru signálu nuluje harmonické složky. Tím pádem dojde k odhadu šumové části signálu v časové oblasti. Tuto část signálu lze v časové oblasti odečíst od původního signálu a tím získat odhad periodické části signálu opět v časové oblasti. Výpočet energie těchto dvou částí signálu je pak rutinní záležitost.

**Spektrální páry LSP** je parametrizace, která využívá lineární prediktivní analýzy. Spektrální páry jsou páry frekvencí rozmístěných poblíž vrcholů ve vyhlazeném LPC odhadu spektra. Čím blíže u sebe tyto frekvence jsou, tím výraznější vrchol (tzn. nárůst energie v spektru) se mezi těmito frekvencemi nachází. Podle předpokladu neobsahuje zdravý hlas výraznější nárůst energie na vyšších frekvencích a tím pádem lze nalézt rozdíly mezi spektrálními páry zdravých a patologických hlasů.

**Spektrální deformace SD a spektrální šířka SW** je dvojice parametrů původně vyvinutá pro analýzu elektromyografických signálů (Sinderby et al. [31]). Tyto parametry byly později úspěšně použity pro analýzu frikativních hlásek (Frid and Lavner [10]). Lze předpokládat, že je možné nalézt rozdíly i ve spektrech zdravých a patologických hlasů. Oba výše zmíněné parametry dávají do poměrů spektrální momenty.

**Spektrální rolloff SR** je frekvence ve spektru signálu  $f_r$ , pod kterou je  $p$  procent celkové energie signálu (Frid and Lavner [10]). V práci je použito značení např. SR10, které určuje frekvenci, pod kterou je 10 % celkové energie signálu. Opět lze předpokládat, že průběh frekvencí u zdravého a patologického hlasu bude odlišný a tím pádem se bude lišit i hodnota parametru SR.

Při výpočtu některých z výše uvedených parametrů lze získat odhad hlasivkové frekvence  $f_0$ . Jedná se o odhad v spektrální oblasti CEPSF0, který lze získat při výpočtu parametru CPP, ve spektrální oblasti SPECF0, který lze získat při výpočtu parametru HNR nebo H1A a odhad v časové oblasti ACRF0, který lze získat při výpočtu parametrů HLR nebo SPS. V rámci práce nebyl vyvíjen robustní algoritmus výpočtu hlasivkové frekvence. Pro další práci byly využity i korelace mezi těmito průběhy.

V rámci experimentálního objektivního hodnocení chraptivosti byly analyzovány i nahrávky obsahující čtený text. Takové nahrávky obsahují jak znělé, tak neznělé 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 [33], 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 [8] je na základě experimentů RAPT označen jako nejlepší pro detekci znělých úseků jak spojitě řeči, tak prodloužené fonace. Algoritmus je implementován např. v programu Wavesurfer (Sjölander and Beskow [32]) nebo v balíku skriptů pro MATLAB Voicebox (Brookes [5]).

### 3.4 Popisy průběhů parametrů

Každý z výše popsaných parametrů poskytne časový sled hodnot pro každou nahrávku. V literatuře je běžně používám pouze popis distribuce těchto hodnot pomocí střední hodnoty  $\mu$ , směrodatné odchylky  $\sigma$  nebo mediánu  $M$ . Pro popis distribuce je v této práci použita ještě relativní směrodatná odchylka definovaná vztahem  $\sigma_R = \sigma/\mu$ .

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.

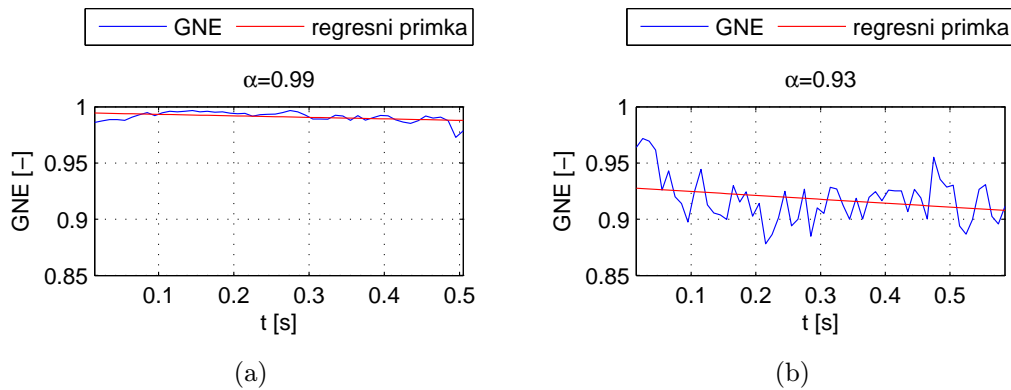
Parametr délka křivky (CL - Curve Length) je definován podle vztahu (Pan et al. [27])

$$\text{CL} = \frac{1}{N} \sum_{n=1}^{N-1} |x[n-1] - x[n]|, [-] \quad (3.1)$$

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 regresní přímky  $\alpha$  byl použit jako jeden z parametrů. Obr. 3.5 ukazuje rozdíly ve směrnici regresní přímky pro zdravý hlas (obr. 3.5(a)) a pro chraptivý hlas (obr. 3.5(b)). Z obrázku je patrné, že zdravý hlas má vyrovnanější průběh, zatímco chraptivý hlas dosahuje v závěru nahrávky nižších hodnot a tím pádem má regresní přímka větší sklon. Tento pokles v závěru může být způsoben větší únavou hlasového traktu dlouhou fonací.

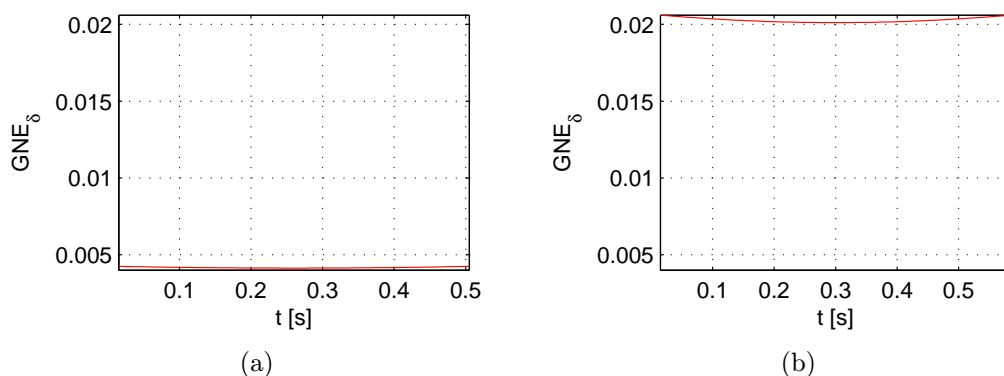
Další parametrizace odpovídá popisu signálu zbaveného lineárního trendu. Tento parametr je nazván LRE (Linear regression error – chyba lineární regrese). Zavedení tohoto parametru je motivováno výše zmíněným faktem: u zdravého hlasu je předpokládán vyrovnanější, více lineární průběh parametrů v průběhu nahrávky.



Obrázek 3.5: Ukázka trendu časového průběhu parametru GNE pro (a) zdravý a (b) chraptivý hlas.

Z popisu pomocí LRE jsou spočteny následující charakteristiky: směrodatná odchylka ( $LRE_{\sigma}$ ), maximální a minimální hodnota ( $LRE_{\max}$  a  $LRE_{\min}$ ), rozsah hodnot ( $LRE_{\text{range}}$ ) a maximální vzdálenost od nulové hodnoty ( $LRE_{\max\text{abs}}$ ). Střední hodnotu a medián LRE není nutné určovat; vzhledem ke způsobu výpočtu parametru LRE budou tyto hodnoty nulové.

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. 3.6 je zobrazen vektor hodnot  $\delta[n]$  pro průběh parametru GNE. Z obrázku je patrné, že pro zdravý hlas (obr. 3.6(a)) je dosaženo výrazně nižších hodnot než pro chraptivý hlas (obr. 3.6(b)). To je způsobeno větší časovou nestabilitou průběhu parametru pro chraptivý hlas. Z parametrizace  $\delta$  jsou určeny následující hodnoty: střední hodnota ( $\delta_{\mu}$ ), směrodatná odchylka ( $\delta_{\sigma}$ ), relativní směrodatná odchylka ( $\delta_{\sigma_R}$ ), minimální a maximální hodnota ( $\delta_{\min}$  a  $\delta_{\max}$ ) a rozsah hodnot ( $\delta_{\text{range}}$ ).



Obrázek 3.6: Ukázka průběhu parametrizace  $\delta$  pro (a) zdravý a (b) chraptivý hlas.



Akustické parametrizace uvedené v části 3.3 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. 3.1. Metodika tvoření výsledného popisu je prováděna podle následujícího klíče:

$$\text{PARAMETRvoiced}_{\text{parametrizace}}$$

kde „PARAMETR“ je jeden z parametrů uvedených v části 3.3, „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. 3.1.

Tabulka 3.1: 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: GNEvoiced<sub>μ</sub>.

zkratka	popis
$\mu$	střední hodnota
$\sigma$	směrodatná odchylka
$M$	medián
$\sigma_R$	relativní směrodatná odchylka
CL	délka křivky
$\alpha$	směrnice lineární regresní přímky
LRE	časový průběh parametru zbavený lineárního trendu
LRE <sub>σ</sub>	směrodatná odchylka hodnot LRE
LRE <sub>max</sub>	maximální hodnota LRE
LRE <sub>min</sub>	minimální hodnota LRE
LRE <sub>range</sub>	rozsah hodnot LRE
LRE <sub>maxabs</sub>	maximální vzdálenost hodnoty LRE od nuly (LRE má nulovou střední hodnotu)
$\delta$	odhad směrodatné odchylky chyb při určení následujícího vzorku $x[n]$ ze vstupních dat
$\delta_\mu$	střední hodnota $\delta$
$\delta_\sigma$	směrodatná odchylka $\delta$
$\delta_{\sigma_R}$	relativní směrodatná odchylka $\delta$
$\delta_{\min}$	minimální hodnota $\delta$
$\delta_{\max}$	maximální hodnota $\delta$
$\delta_{\text{range}}$	rozsah hodnot $\delta$
voiced	parametrizace byla spočtena pouze ve znělých úsecích signálu



## 3.5 Dolování dat

Pro každou nahrávku obsahující prodlouženou fonaci bylo spočteno celkem 480 parametrizací a pro nahrávky obsahující čtený text celkem 201 parametrizací. 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 [13]). 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 [18], Kononenko [19], Robnik-Sikonja and Kononenko [29]).

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

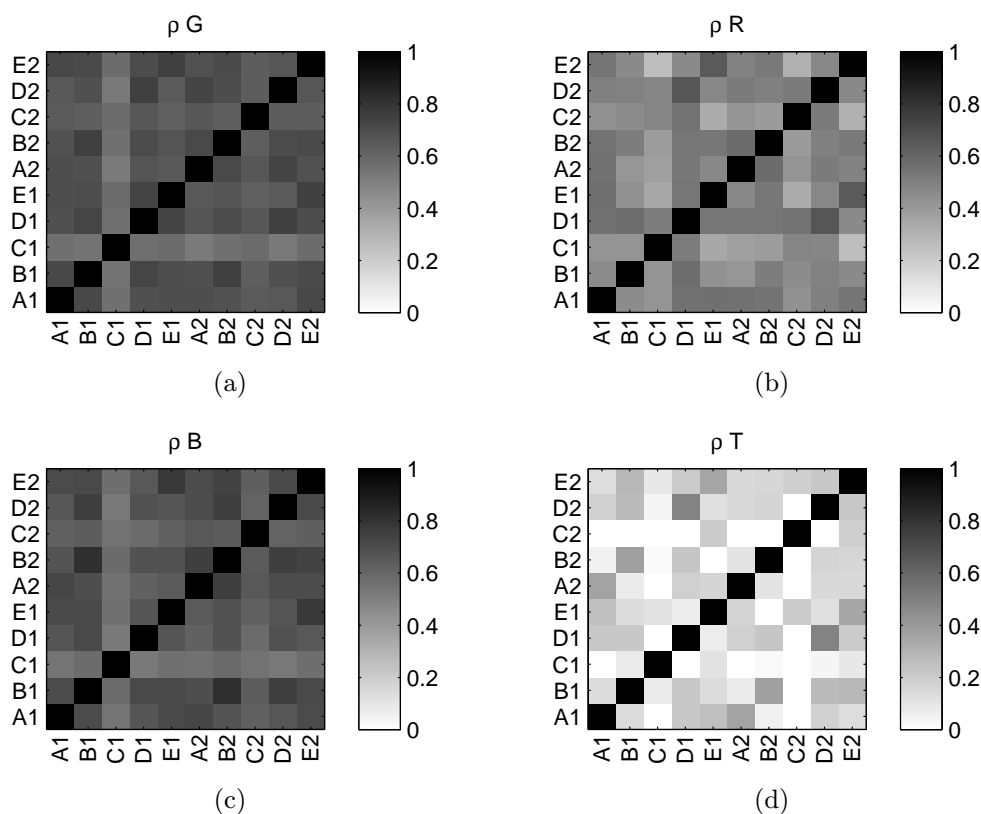
# 4. Výsledky

## 4.1 Subjektivní hodnocení

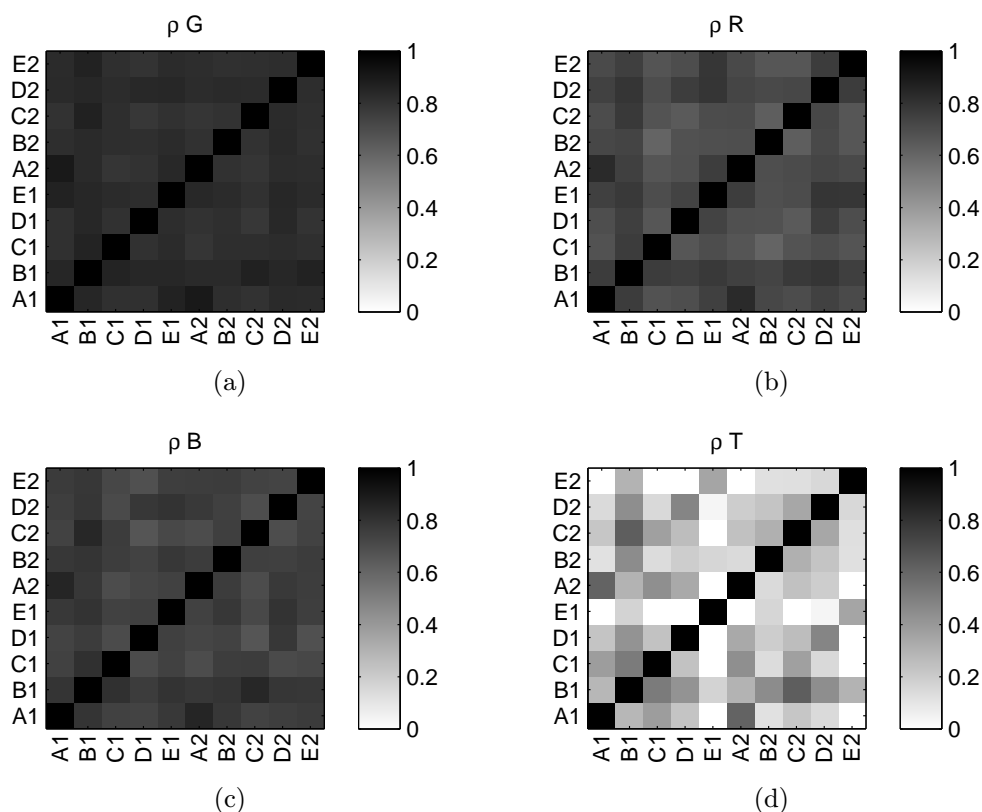
Subjektivní hodnocení bylo podrobena analýze pro ověření vhodnosti, zda je možné toto hodnocení použít jako vstup pro trénování automatických klasifikátorů. Jednotliví hodnotitelé jsou označeni písmeny A, B, C, D a E.

Pro základní náhled nad hodnocením je použito zobrazení pomocí korelačních matic zobrazených na obr. 4.1 a obr. 4.2. Tyto korelační matice popisují jednak vzájemnou shodu hodnotitelů a jednak zobrazují shodu skupiny hodnotitelů – čím je barva tmavší (ať už příslušného binu, nebo celé matice), tím je korelace vyšší.

Detailnější rozbor je pak proveden odpovídajícími statistickými metodami porovnávající tzv. intra- a inter-rater agreement, tzn. shodu v rámci hodnotitele při opakovaném hodnocení (Cronbachova alfa, tab. 4.1) a shodu hodnocení skupiny hodnotitelů (Fleissova kappa, tab. 4.2).



Obrázek 4.1: Korelační matice Spearmanových korelačních koeficientů pro jednotlivé hodnotitele pro parametry G, R, B, T pro prodlouženou fonaci hlásky /a/.



Obrázek 4.2: Korelační matice Spearmanových korelačních koeficientů pro jednotlivé hodnotitele pro parametry G, R, B, T pro čtený text *Podzim na Starém bělidle*.

Tabulka 4.1: 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

## 4.2 Automatické hodnocení chraptivosti

Výsledky jsou prezentovány jako schémata navrhovaného experimentálního systému pro automatické hodnocení chraptivosti, který byl zaveden v cílech práce. Z těchto schémat lze určit parametry použité pro klasifikaci, použité klasifikátory i výsledné

Tabulka 4.2: 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

klasifikační skóre. Klasifikační skóre pro diskretní hodnocení (obr. 4.3) je uvedeno jako přesnost<sup>1</sup> klasifikace, spojitě hodnocení (obr. 4.4) je uvedeno jako Pearsonův lineární korelační koeficient  $r$ .

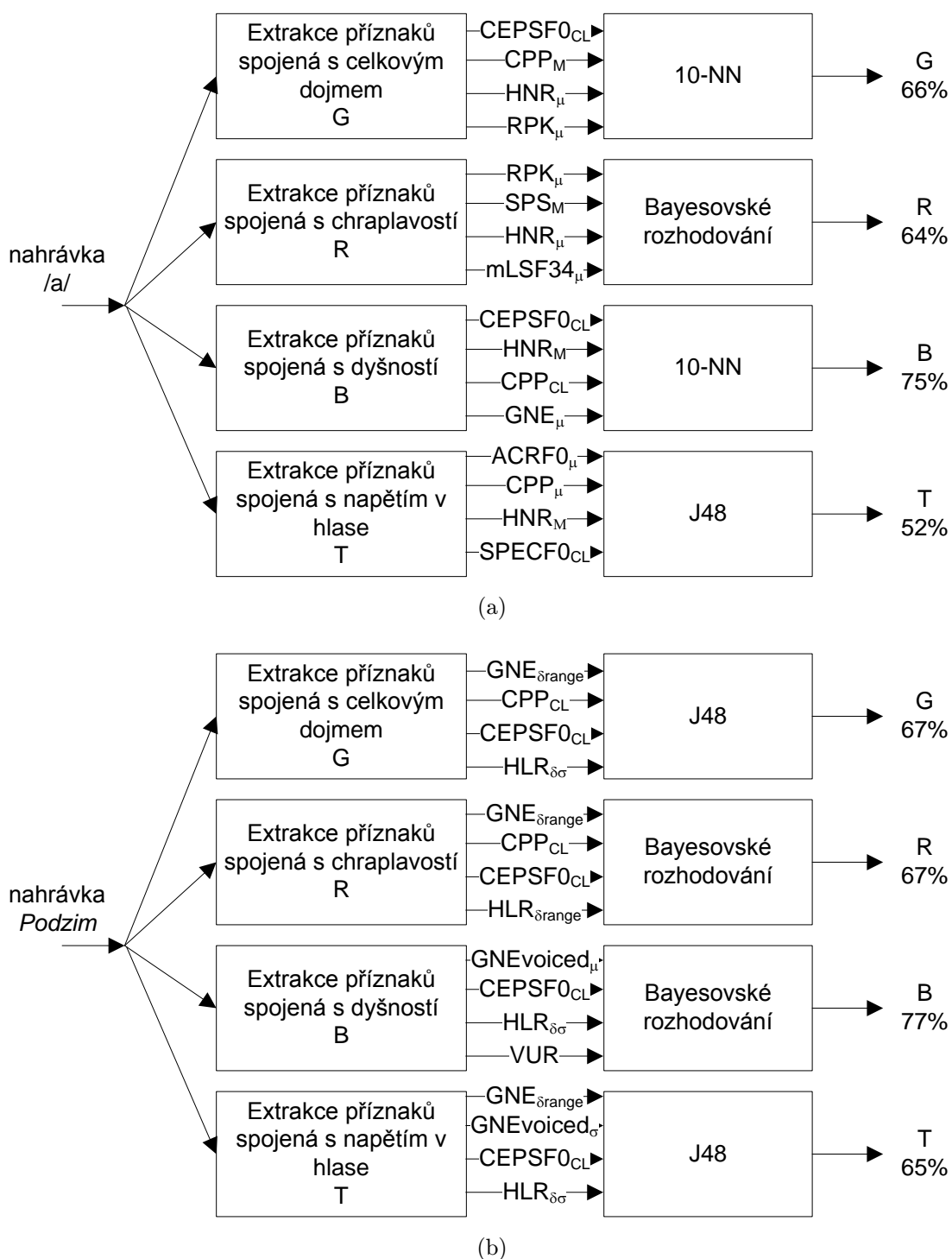
Pro porovnání subjektivního a objektivního hodnocení podle diskretní stupnice viz tab. 4.3, která porovnává dosažené skóre automatické klasifikace a shody hodnotitelů při opakovaném hodnocení.

Pro porovnání subjektivního a objektivního hodnocení podle spojitě stupnice viz obr. 4.5 a obr. 4.6, které ukazují závislost subjektivního a objektivního hodnocení.

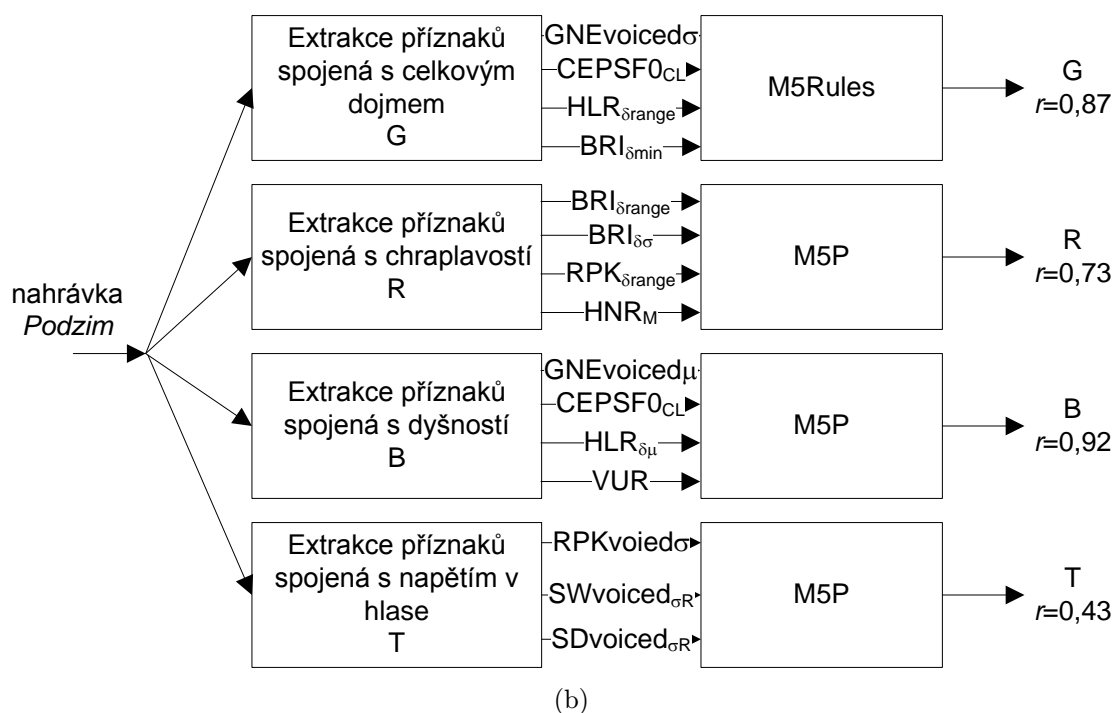
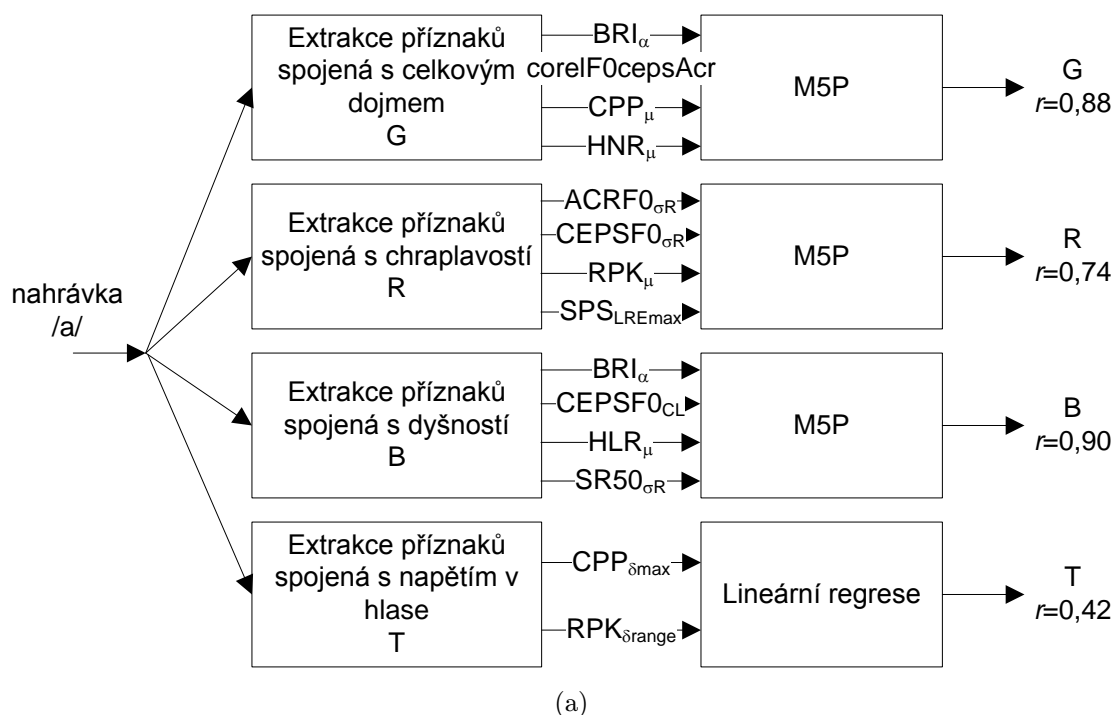
Tabulka 4.3: Vážený průměr hodnot popisujících diskretní klasifikaci. Porovnání výsledků subjektivního a objektivního hodnocení pomocí senzitivity (SEN) a specificity (SPE) a metrik TP, FP, TN, FN.

		objektivní						subjektivní					
		TP	FN	FP	TN	SEN	SPE	TP	FN	FP	TN	SEN	SPE
/a/	<b>G</b>	0,66	0,34	0,21	0,79	0,66	0,79	0,58	0,42	0,19	0,81	0,58	0,81
	<b>R</b>	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>B</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
	<b>T</b>	0,52	0,48	0,21	0,79	0,52	0,79	0,55	0,45	0,17	0,83	0,55	0,83
text	<b>G</b>	0,67	0,33	0,17	0,83	0,69	0,83	0,59	0,41	0,17	0,83	0,59	0,83
	<b>R</b>	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>B</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
	<b>T</b>	0,65	0,35	0,18	0,82	0,65	0,82	0,63	0,37	0,16	0,84	0,63	0,84

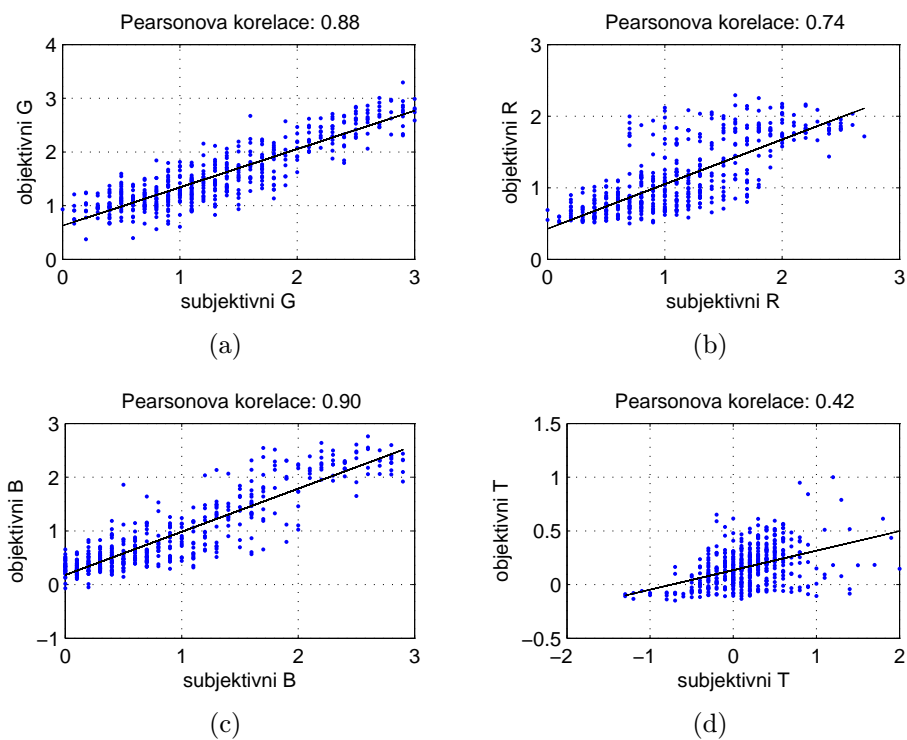
<sup>1</sup>Jako přesnost je zde myšlen anglický termín accuracy, nikoliv precision.



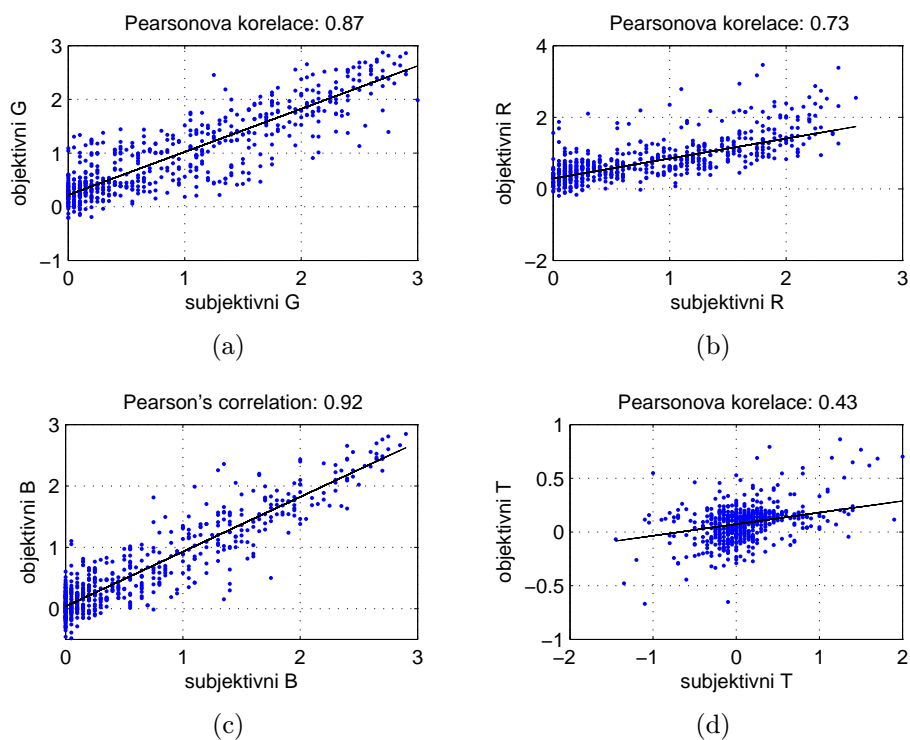
Obrázek 4.3: Schémata experimentálních systému pro automatické hodnocení chraptivosti podle diskretní stupnice u (a) prodloužené fonace a (b) čteného textu.



Obrázek 4.4: Schémata experimentálních systému pro automatické hodnocení chraplavosti podle spojitě stupnice u (a) prodloužené fonace a (b) čteného textu.



Obrázek 4.5: Výsledky klasifikace prodloužené fonace hlásky /a/ podle spojité stupnice hodnocení.



Obrázek 4.6: Výsledky klasifikace čteného textu podle spojité stupnice hodnocení.

## 5. Závěry a další cíle

V rámci práce byla provedena analýza možností automatického hodnocení chraptivosti, byly vytvořeny dvě databáze nahrávek s chraptivými hlasy (469 nahrávek prodloužené fonace hlásky /a/ a 593 nahrávek standardizovaného čteného textu), bylo provedeno statistické vyhodnocení subjektivního hodnocení nahrávek, bylo implementováno množství algoritmů pro akustickou analýzu signálu, byly zavedeny původní popisy časového průběhu parametrů a byl navržen systém pro automatické hodnocení chraptivosti na základě analýzy akustického signálu.

Subjektivní hodnocení nahrávek bylo provedeno metodou GRBT, která popisuje jednotlivé příznaky chraptivosti: G (celkový dojem z hlasu), R (chraplavost, vjem spjatý s nepravidelností kmitání hlasivek), B (dyšnost, vjem spojený s mírou šumové příměsi v hlase) a T (hlasová slabost nebo hlasové napětí).

Subjektivní hodnocení nahrávek s chraptivostí bylo na základě statistických analýz shledáno dostatečně přesné a lze jej použít pro trénování modelů automatických klasifikátorů.

Množství parametrizací bylo vhodnými metodami dolování dat zredukováno na maximálně čtyři parametry pro každý sledovaný aspekt chraptivosti.

Výsledky automatického hodnocení podle diskrétní stupnice a schémata experimentálního systému pro hodnocení chraptivosti jsou zobrazeny na obr. 4.3. Z obrázku je patrné, že nejlepších výsledků je dosažováno při automatickém hodnocení dyšnosti B (75 % u prodloužené fonace a 77 % u čteného textu). Výsledky hodnocení celkového dojmu G a chraplavosti R jsou v obou případech srovnatelné a pohybují se okolo 66 %. Při hodnocení napětí v hlase T je dosažováno 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 zobrazeny na obr. 4.4. 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 chraplavosti 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.

**Použití popisu časových průběhů parametrů se ukázalo jako klíčové**, jak je vidět z obr. 4.3 a obr. 4.4. Tyto parametrizace jsou ve výsledných klasifikátorech často používány a i samostatně dosahují dobrých výsledků z pohledu schopnosti odlišit jednotlivé stupně chraptivosti.



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;
- V prostředí MATLAB bylo implementováno velké množství algoritmů pro akustickou analýzu zvukového signálu
- 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: diskrétní: 66 %, spojité 0,88
    - \* R: diskrétní: 64 %, spojité 0,74
    - \* B: diskrétní: 75 %, spojité 0,90
    - \* T: diskrétní: 52 %, spojité 0,42
  - čtený text
    - \* G: diskrétní: 67 %, spojité 0,87
    - \* R: diskrétní: 67 %, spojité 0,73
    - \* B: diskrétní: 77 %, spojité 0,92
    - \* T: diskrétní: 65 %, spojité 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

## 5.1 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 keprstrálních koeficientů

pro analýzu chraptivosti. Další zajímavou parametrizací je bezesporu tzv. modulační spektrum (Markaki and Stylianou [22]), které by mohlo přinést další možný pohled na patologickou řeč. Dále by bylo vhodné otestovat vhodnost detektoru náhlých změn ve spektru (Čmejla et al. [25]) 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. [16].

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] 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.
- [2] 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>.
- [3] O. Amir and R. Levine-Yundof. Listeners’ attitude toward people with dysphonia. *Journal of Voice*, 2013. In press.
- [4] 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.
- [5] Mike Brookes. Voicebox: Speech processing toolbox for matlab. online, 2011. URL <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html>.
- [6] 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 quaiity. *Journal of Voice*, 11(1):74–80, 1997.
- [7] 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.
- [8] Thomas Drugman and Abeer Alwan. Joint robust voicing detection and pitch estimation based on residual harmonics. In *INTERSPEECH*, pages 1973–1976. ISCA, 2011.
- [9] 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.
- [10] 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/EEEI.2010.5662110.

- [11] 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.
- [12] 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.
- [13] M. A. Hall. *Correlation-based Feature Subset Selection for Machine Learning*. PhD thesis, University of Waikato, Hamilton, New Zealand, 1988.
- [14] 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.
- [15] Minoru Hirano. *Clinical Examination of Voice*. Springer London, 1981.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] E.P.-M. Ma and E.M.-L. Yiu. Multiparametric evaluation of dysphonic severity. *Journal of Voice*, 20(3):380–390, 2006.
- [22] 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.

- 
- [23] 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.
- [24] MATLAB. *version 8.0.0.783 (R2012b)*. The MathWorks Inc., 2012.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] R. Riley. *Cool Edit 2000 a Cool Edit Pro 2: Strih, efekty, procesory*. Computer Press, Brno., 2004.
- [29] 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.
- [30] 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.
- [31] C. Sinderby, L. Lindström, and A. E. Grassino. Automatic assessment of electromyogram quality. *J Appl Physiol*, 75(5):1803–1815, 1995.
- [32] K. Sjölander and J. Beskow. *Wavesurfer [Computer program] (Version 1.8.5)*. URL <http://www.speech.kth.se/wavesurfer>.
- [33] 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.
- [34] Jan Vokřál. *Akustické parametry chraptivosti*. PhD thesis, České vysoké učení technické v Praze, Fakulta elektrotechnická, 1998.

- [35] 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.
- [36] 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>.

# 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 *Králíky 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

# Summary

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.

# Resumé

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 Foniatické 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 Foniatické 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.