

AKUSTICKÉ LISTY

České akustické společnosti
www.czakustika.cz

ročník 17, číslo 1–2

srpen 2011

Obsah

Usnesení Valné hromady České akustické společnosti	3
In memoriam – Ing. Pavel Siegl	4
Ing. Pavel Urban, DrSc. – osmdesátníkem	5
Robustní algoritmy detekce špiček pro odhad základní hlasivkové frekvence prodloužených fonací samohlásek u patologických hlasů Robust Peak Detection Algorithms for Fundamental Frequency Estimation of Sustained Vowel Phonation in Pathological Voices <i>Lukáš Bauer, Jan Ruzs a Roman Čmejla</i>	7
Hodnocení vokálních parametrů u patologických hlasů Assessment of Vocal Parameters in Pathological Voices <i>Lukáš Bauer, Jan Ruzs a Roman Čmejla</i>	13
Efekt zvyšování hlasitosti na spektrální charakteristiky hlasu u různých typů použití hlasu a u různých skupin hlasových profesí The Effect of a Voice Loudness Variation on the Spectral Characteristics of the Voice in a Different Manner of a Voice Usage and Different Groups of Voice Professions <i>Marek Frič</i>	19

Usnesení Valné hromady České akustické společnosti, konané dne 20. ledna 2011 v prostorech Fakulty elektrotechnické ČVUT

Valná hromada ČsAS bere na vědomí:

1. zprávu o činnosti Rady ČsAS;
2. zprávu o uspořádání kongresu Euronoise 2012 v Praze;
3. zprávy o činnosti jednotlivých odborných skupin a o jejich dalším zaměření;
4. zprávu o přípravě 82. akustického semináře;
5. zprávu o výsledcích revize hospodaření společnosti;
6. výsledky voleb do Rady společnosti a výsledky voleb předsedů odborných skupin;
7. zprávu o usnášitelnosti Valné hromady, 48/83;
8. informace o plnění úkolů a poslání Akustických listů a vyzývá členy k zasílání příspěvků.

Pro funkční období roku 2011 byli v jednotlivých odborných skupinách zvoleni:

- A. Obecná, lineární a nelineární akustika
předseda – M. BEDNAŘÍK zástupce – M. ČERVENKA
- B. Ultrazvuk a akustické emise
předseda – J. PLOCEK zástupce – R. BÁLEK
- C. Hluk a stavební akustika
předseda – M. MELLER zástupce – V. STRNAD
- D. Zpracování a záznam akustických signálů
předseda – T. SALAVA
- E. Psychoakustika, fyziologická akustika a akustika hudby a řeči
předseda – volba bude provedena dodatečně
- F. Elektroakustika
předseda – Z. KEŠNER zástupce – B. SÝKORA

Valná hromada ČsAS schvaluje:

1. zprávu o činnosti Rady za kalendářní rok 2010 a uděluje Radě absolutorium;
2. zprávu o hospodaření společnosti za kalendářní rok 2010;
3. výši členských příspěvků na rok 2011 (500 Kč pro členy, 150 Kč pro studenty a důchodce);
4. činnost Rady a odborných skupin v roce 2010.

Valná hromada ukládá nově zvolené Radě společnosti na kalendářní rok 2011:

1. pokračovat v odborné a organizační činnosti i v zahraničních kontaktech, v rozvíjení spolupráce s Českou matricí technickou, Slovenskou akustickou společností, společností Elektra, českou sekcí AES, Českým centrem IEE, československou sekcí IEEE, ICA, EAA a I-INCE;
2. věnovat pozornost pořádání odborných akcí a pravidelných seminářů odborných skupin;
3. nadále rozvíjet vydávání Akustických listů;

Valná hromada ukládá nově zvoleným předsedům odborných skupin na kalendářní rok 2011:

1. publikovat informace o připravovaných aktivitách skupin v Akustických listech a na webové stránce s předstihem tak, aby se zájemci mohli včas na akce přihlašovat.

Valná hromada doporučuje Radě ČsAS:

1. pravidelně se zabývat činností a plánem akcí odborných skupin;
2. pravidelně se zabývat plánem a zaměřením konaných akustických konferencí;
3. poskytovat možnost finančních výhod členům společnosti, např. nižšími sazbami vložného na akcích pořádaných společností.

Výsledky voleb do Rady České akustické společnosti:

Počet vydaných platných hlasovacích lístků:		48
Odevzdáno platných hlasovacích lístků:		48
předseda:	O. JIŘÍČEK	48
místopředseda:	V. KUNZL	48
sekretář:	M. BROTHÁNEK	48
hospodář:	O. KUDĚJOVÁ	48
revizní komise:	J. KOZÁK	45
	T. HELLMUTH	48
	D. POTUŽNÍKOVÁ	48

Návrh usnesení sestavila návrhová komise ve složení M. Meller a J. Stěnička.

* * *

In memoriam – Ing. Pavel Siegl



Ing. Pavel Siegl zemřel po těžké nemoci 7. prosince 2010 ve věku 61 let.

Po vystudování elektrotechnické fakulty ČVUT a krátkodobém působení v TESLA Přelouč začal pracovat ve Výzkumném ústavu rozhlasu a televize, kde působil až do jeho zrušení na konci roku 1997. Tehdy spolu se svými kolegy z akustické skupiny přešel do projektového ateliéru Kinotechniky Praha. Od podzimu roku 2001 pak působil ve firmě SONING.

Od počátku své profesní kariéry se zabýval akustikou v celé šíři její problematiky. Ve VÚRTu řešil úkoly zaměřené na problematiku místností pro snímání, poslech a zpracování zvuku, prostorovou akustiku, šíření akustického signálu, akustické obklady, stavební akustiku a přenos hluku a vibrační konstrukcemi. Během své dlouholeté činnosti byl i spoluautorem odborných publikací a s kolegy z VÚRTu se podílel na vydání rozsáhlého katalogu akustických obkladových materiálů. Dlouhodobě se systematicky věnoval měření činitele zvukové pohltivosti obkladových materiálů.

Během své činnosti v akustice si neustále rozšiřoval znalosti shromažďováním nových odborných poznatků, účastí na odborných konferencích a seminářích, studiem literatury i jiných odborných publikací.

Ve volných chvílích rád navštěvoval koncerty, divadelní i filmová představení včetně alternativních kulturních akcí různého zaměření. Po celý svůj život se aktivně věnoval sportu. Měl velmi rád turistiku spojenou s poznáváním pamětihodností, přírody i kuriozit jak u nás doma, tak i v exotičtějších destinacích. Na cestách bylo jeho zálibou fotografování.

Svémi kolegy byl uznáván pro svou čestnost, skromnost, přátelství i ochotu spolupracovat s mladšími kolegy a předávat jim své zkušenosti a vědomosti.

V Ing. Pavlu Sieglovi jsme ztratili nejen zkušeného a všestranného odborníka akustika, ale hlavně neocenitelného člověka, kolegu a kamaráda.

Jana Faitová, Václav Moulík, Petr Novák, Václav Kozel a kol.

Ing. Pavel Urban, DrSc. – osmdesátníkem

Významný člen České akustické společnosti Ing. Pavel Urban, DrSc., se v prosinci loňského roku dožil osmdesáti let. Díky své vitalitě je stále v plném pracovním nasazení a pracovním poměru v TUV SÚD Czech Republic, v divizi Automotive, původně Ústav pro výzkum motorových vozidel. Kromě krátkého období, kdy po skončení studia na elektrotechnické fakultě ČVUT působil jako asistent na katedře sdělovací techniky, zasvětil celý svůj život akustice. Tomuto oboru se začal věnovat po nástupu do Ústavu pro výzkum motorových vozidel, kde z počátku působil jako vedoucí výzkumné skupiny odrušení vozidel a elektroakustiky, pak od r. 1958 samostatné skupiny hluk vozidel a od roku 1982 jako vedoucí odboru hluku a chvění. Od roku 1986 vedl zároveň i odbor příslušenství motorových vozidel. V roce 1990 se stal vědeckým tajemníkem ústavu a vedoucím marketingového oddělení. V letech 1993 až 1996 vedl legislativní divizi, v níž pracuje i v současnosti. V těchto letech se též plně věnoval zajištění akreditace ústavu.

V průběhu času se převážně zajímal o fyzikální problematiku hluku motorových vozidel, ve které dosáhl v roce 1964 vědecké hodnosti kandidáta věd a v roce 1976 vědecké hodnosti doktora technických věd. V této oblasti byl nositelem mnoha výzkumných úkolů řešících zásadní problematiku automobilového průmyslu. Nejinak tomu bylo i s činností normalizační, a to nejen v problematice hluk – automobil, ale i v úzké návaznosti na problematiku hluk – životní prostředí. Významně se podílel na koncepci a znění mezinárodních předpisů v oblasti automobilů v Evropské hospodářské komisi OSN se sídlem v Ženevě, kde působil jako československý expert déle než dvacet let. Obdobně se účastnil práce v Mezinárodním sdružení výrobců automobilů OICA i v Mezinárodní normalizační organizaci ISO.

Během působení v ÚVMV se věnoval i výchově dalších akustiků formou vedení aspirantů, přednáškami v různých předmětech na ČVUT a intenzivní činností v kurzech ČsVTS. Bohatost jeho odborné kariéry dokládá i velký počet publikovaných prací v tuzemsku i v zahraničí. Je autorem nebo spoluautorem několika odborných knižních publikací. Pro ilustraci připomeňme několik konkrétních skutečností: v letech 1963–1969 zavedl v československé akustice jako jeden z prvních principy statisticko-dynamické analýzy signálů mechanického charakteru, zpracoval teoreticky i prakticky a publikoval v té době neobvyklé měřicí postupy, které vyústily patentovou ochranou i v zahraničí. Kniha *Teorie a metody měření přenosu vibrací mechanickými díly* (Academia 1973) byla originálním vkladem do uvedené problematiky. Z pozdější doby je nutno též připomenout, že jako první v Československu zavedl přímé měření energetických veličin zvuku pomocí vlastní metody, založené na měření fázových vztahů signálů ze dvou mikrofónů. Není proto divu, že Ing. Pavla Urbana, DrSc., dnes snad zná každý pracovník, který se věnuje akustice, a to nejen v odvětví motorových vozidel, ale ve všech odvětvích s akustikou souvisejících.

Nad rámec svých pracovních a se zaměstnáním souvisejících povinností byl v letech 1976 až 1988 předsedou odborné skupiny „Hluk a akustika prostředí“ při Československé vědecko-technické společnosti. Na základě všeobecné váženosti mezi akustiky byl v r. 1990 zvolen prvním předsedou nově vzniklé Československé akustické společnosti. Od doby ukončení předsednického mandátu daného stanovami pracuje nepřetržitě v radě ČsAS.

Významně se rovněž podílel na normalizační činnosti na národní úrovni. Byl dlouholetým členem Technické normalizační komise pro akustiku a v letech 1990 až 1996 byl jejím předsedou.

S životní vitalitou Ing. Pavla Urbana, DrSc., a jeho neutuchajícím zájmem o dění v obci akustiků se mohla seznámit i nejmladší akustická generace na loňském 80. akustickém semináři, kde přednesl referát o začátcích a tradicích akustického dění u nás.

U příležitosti životního jubilea se Česká akustická společnost připojuje k řadě gratulantů s přáním pevného zdraví a elánu do dalších let.



Robustní algoritmy detekce špiček pro odhad základní hlasivkové frekvence prodloužených fonací samohlásek u patologických hlasů

Lukáš Bauer^a, Jan Ruzs^{a,b} a Roman Čmejla^a

^aČVUT – FEL, Katedra teorie obvodů, Technická 2, 166 27 Praha 6

^bUniverzita Karlova v Praze, Neurologická klinika 1. LF UK a VFN, Kateřinská 30, 128 21 Praha 2
e-mail: bauerlu3@fel.cvut.cz

This paper presents design of two new methods of speech fundamental frequency (f_0) detection for vowel sustained phonations and the detection method, which use cross-correlation to detect f_0 , is tested. The algorithm consists of certain preprocessing and processing methods. The first method is based on the detection of maxima and the second method is based on band pass filtration. In comparison with the other commonly used f_0 detection methods, our algorithms are designed with respect to speech pathology detection. These methods lead to detection of the other voice parameters such as jitter, shimmer and harmonic-to-noise ratio (HNR). The results of this study are compared with database, which is labeled by the help of Praat algorithm. The results for maximum method succeed at 88.4% and for pass band method at 83.9%. The detection leads to create self-automated method, which detect robustly f_0 .

1. Úvod

Výpočet základní hlasivkové frekvence (f_0) řečového signálu je důležitým krokem v řadě řečových aplikací, jako je syntéza řeči, detekce pohlaví, rozpoznání mluvčího. Ve fonetice se detektory špiček používají k detekci prosodie [1]. Řečové patologie, poruchy hlasového ústrojí, se vyznačují velkou mírou šumu v signálu. Řada metod vedoucích k odhalení řečových patologií je založena na správné detekci základní hlasivkové frekvence signálu.

Běžně používané metody detekce základní hlasivkové frekvence však nejsou využitelné v patologických promluvách, protože nevedou k označení okamžiků potřebných k detekci ostatních vokálních parametrů. Mezi tyto parametry patří jitter (frekvenční nestabilita), shimmer (amplitudová nestabilita) a HNR (odstup harmonické od šumu). Pro detekci těchto parametrů je ale nutné mít dostatečně přesně detekované jednotlivé periody signálu.

Z tohoto důvodu je nutné navrhnout nové automatizované a dostatečně robustní metody detekce základní hlasivkové frekvence, sloužící k detekci ostatních řečových parametrů, pomocí nichž jsme schopni detekovat řečové patologie.

2. Detekce základní hlasivkové frekvence

První algoritmy detekce základní hlasivkové frekvence začaly vznikat již v druhé polovině minulého století. Tyto algoritmy jsou běžně založené na metodách, jako je autokorelace, vzájemná korelace, vlnková transformace, kepstrální technika. Dobré porovnání jednotlivých přístupů přináší [2]. V posledních letech se začínají objevovat algoritmy řešící problém pomocí Hilbertovy transformace [3]. Tyto metody jsou však běžně používané v telekomunikač-

ních technologiích a pro detekci řečových patologií jsou nevhodné.

Pro detekci základní hlasivkové frekvence v patologických promluvách jsou nejčastěji používány dva přístupy. Prvním z nich je metoda detekce špiček v signálu (peak-picking), reprezentovaná komerčním programem Multi-Dimensional Voice Program [4]. Druhým přístupem je metoda založená na porovnávání vzájemné podobnosti částí jednotlivých period (waveform-matching) užívaná volně dostupným softwarem Praat [5,6], dostupný na <http://praat.org>. V práci přistupujeme k detekci základní hlasivkové frekvence pomocí metody detekce špiček v signálu.

Typické hodnoty fundamentální frekvence se pro každého jedince liší. Rozsah hodnot f_0 uvádí tabulka 1. Tato tabulka platí pro běžnou populaci. Samozřejmě že se naleznou výjimky hlavně mezi profesionálními zpěváky, u kterých se může pohybovat minimální hodnota f_0 kolem 50 Hz a maximální hodnota kolem 1000 Hz.

	$f_{0,\text{typ}}$ (Hz)	$f_{0,\text{min}}$ (Hz)	$f_{0,\text{max}}$ (Hz)
Muži	125	80	200
Ženy	225	150	350
Děti	300	200	500

Tabulka 1: Frekvenční rozsah hlasivkového tónu [7]

3. Metodika návrhu algoritmů

Každá metoda detekce základní hlasivkové frekvence má své klady a zápory, proto budou v další části textu porovnány výsledky jednotlivých metod.

Metody detekce f_0 pracující ve frekvenční oblasti jsou sice výpočetně méně náročné, nelze však pomocí jejich detekce určit přesné pozice špiček (period), vedoucích k detekci řečových parametrů. Další problém je poměrně špatné rozlišení vzhledem k délce vybraného okna. Pro obdržení dostatečně přesného rozlišení (alespoň 1 Hz) by při vzorkovací frekvenci 48 kHz muselo být vybráno okno délky 48 000 vzorků nebo okno kratší a doplněné nulami. Doplnění signálu nulami je běžně používaná technika pro zlepšení frekvenčního rozlišení, ale i tak okno dosahuje příliš velkých rozměrů a ztrácí tak schopnost zachytit krátkodobé změny. Rovněž problémem detekce ve frekvenční oblasti je problematika možného překrytí formantových frekvencí s fundamentální frekvencí.

3.1. Databáze

Databáze signálů byla pořízena v odrušené místnosti Neurologické kliniky 1. LF UK pomocí kondenzátorového kamerového mikrofonu kamery značky Panasonic NV-GS 180. Mikrofon byl umístěn ve vzdálenosti 15 cm od úst mluvčího. Signál je vzorkován frekvencí 48 kHz a k převodu signálu byl použit 16bitový převodník.

Algoritmy detekce f_0 je nutné testovat a ladit na ručně označované databázi signálů, aby bylo možné porovnávat nalezené pozice špiček. Značkování je ale vzhledem k počtu výskytů špiček u dlouhých signálů poměrně časově náročné, proto pro odladění algoritmu detekce f_0 byly použity zkrácené záznamy vokálů /a/, /i/, /u/ obsahující 63 promluv zdravých jedinců a 70 promluv patologických jedinců. Jejich délka se pohybuje do půl sekundy. Tato délka se pro testování jeví jako dostatečná, na delších signálech se dá předpokládat dosažení stejných, nebo dokonce lepších výsledků.

3.2. Algoritmy detekce základní hlasivkové frekvence

V průběhu práce bylo použito pro detekci fundamentální frekvence řečového signálu různých přístupů. Mezi vybrané metody byla vybrána modifikovaná metoda detekce špiček pomocí vzájemné korelace (A), metoda detekce pomocí pozice maxim (B) a metoda detekce pomocí pásmové propusti (C).

Předzpracování signálu

Před samotnou detekcí základní hlasivkové frekvence je nutné signál předzpracovat. Nejprve je odečtena střední hodnota signálu (SS složky). Následně je pomocí Welchovy metody určena průměrná f_0 signálu. Poté je pro každý algoritmus specifickým způsobem signál segmentován. Detekce řečové aktivity je prováděna pomocí algoritmu založeného na počtu průchodu signálu nulou a energie signálu vypočítávané číslicovým integrátorem.

Výběr okna

Hlavním problémem většiny algoritmů je pevně stanovená délka detekovaného okna. Tento přístup předpokládá linearitu a stacionaritu signálu. Řečový signál tento předpoklad však ve většině případů nespĺňuje, protože se v průběhu promluvy mění. Velká variabilita signálu nastává na začátku, kdy člověk začíná zvuk vytvářet, a na konci prodloužené fonace.

Optimální je navrhnout okno s proměnlivou délkou v závislosti na měnící se frekvenci signálu. Tím dochází k odstranění stacionárního a lineárního pohledu na signál. Ve chvíli, kdy dochází k určení délky oken pomocí průměrování z velkého množství period signálu (pět a více), nereaguje průměrování dostatečně rychle na změny signálu. Optimální délka okna byla stanovena na průměr čtyř předchozích detekovaných period.

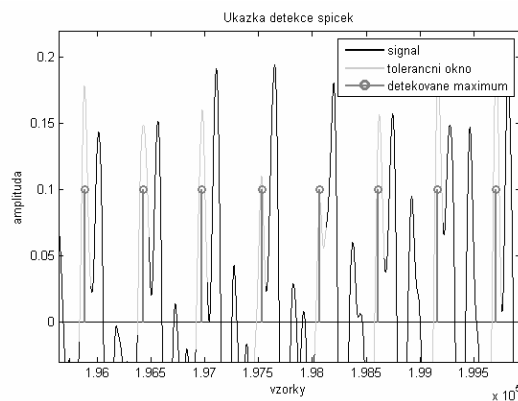
A. Modifikovaná metoda detekce špiček pomocí vzájemné korelace

Metody detekce špiček pracující v časové oblasti jsou zatíženy problémem předpokládané stacionarity délky vybraných segmentů.

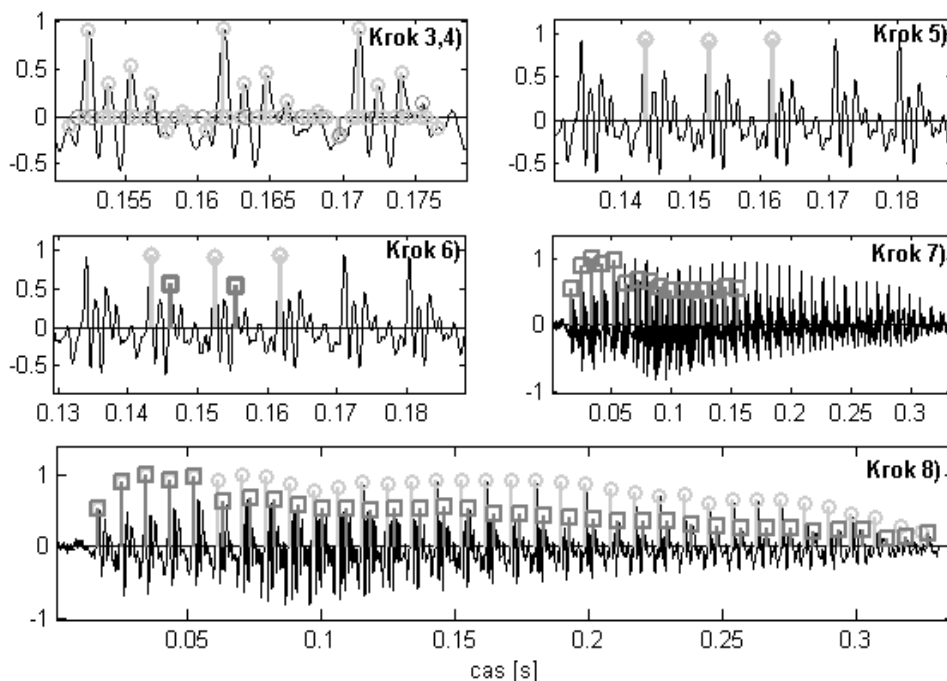
Tato metoda využívá k detekci špiček a následné detekci základní hlasivkové frekvence výběr jedné periody ze signálu. Hlavní problém nastává s výběrem „správné“ periody. Vzhledem ke změnám periodicity v průběhu signálu, hlavně změnám ve velikosti amplitud, je tato metoda značně komplikovaná. Celý postup detekce f_0 pomocí vzájemné korelace je popsán v popisu algoritmu 1.

Popis algoritmu 1:

1. Nalezení průměrné fundamentální frekvence, začátku a konce promluvy
2. Detekce korelační periody
3. Korelace s prvními čtyřmi periodami signálu, pokud nedosahují konce signálu
4. Adaptivní nelineární filtrace a prahování, detekce pozic špiček



Obrázek 1: Úspěšnost nově navržených metod detekce f_0



Obrázek 2: Detekce špiček pomocí maxim, algoritmus 2, ○ iterace z kroku 5, □ iterace z kroku 6

5. Výpočet délek nalezených period a adaptivního posunu
6. Korelace s vybraným úsekem stanoveným adaptivním posunem, pokud není konce signálu, návrat na krok 4

Pokud je nalezena perioda, se kterou bude signál korelován, vybíráme vždy čtyři periody dlouhý segment v signálu ke korelaci. Užitím čtyři periody dlouhého segmentu dochází ke snížení pravděpodobnosti špatného výběru segmentu, se kterým je okno korelováno. Pokud je korelována jedna perioda s jednou periodou, může docházet k špatnému výběru délky periody, a tudíž ke špatné detekci špiček. Segment čtyř period, se kterým dochází ke korelaci, je vybírán adaptivně vzhledem k délce předešlých nalezených period. Na vybraný segment délky čtyř period je poté aplikována nelineární filtrace metodou centrálního klipování, která zajišťuje potlačení formantových frekvencí.

Velkým nedostatkem této metody je nadměrný výskyt zdvojení a půlení fundamentální frekvence (pitch doublingu a pitch halvingu), který nelze úplně odstranit, a možnost automatického užití této metody tedy klesá.

K potlačení výskytu pitch doublingu byla použita metoda kontroly velikosti amplitud sudých a lichých špiček. Pokud je lichá špička menší než sudá nebo naopak a takový výskyt nastává na více než 3/4 pozic, jsou označeny frekvence s nižší amplitudou za chybně detekované.

Ke zlepšení přesnosti algoritmu dochází při segmentaci od středu signálu k jeho krajům, kde f_0 většinou odpovídá průměrné detekované frekvenci a nedochází k takovým proměnám v signálu. Správná pozice špičky je detekovaná s jistou mírou tolerance. Tím se myslí šířka okolí, v kterém hledám novou pozici špičky, viz obr. 1. V pří-

padě, že je zvolená míra tolerance příliš velká, dochází k chybným přeskokům ze „správných“ špiček na formantové špičky.

B. Detekce špiček pomocí maxim

Z testů vyplynulo, že v některých případech je snazší detekce špiček pomocí minim. Tato metoda je používána i v [5]. Volba detekce podle maxim nebo minim je volena na základě výskytu celkové energie nad nulou nebo pod ní. Pokud se nachází výskyt vyšší energie nad nulou, dochází k detekci maxim. Při výskytu vyšší energie pod nulou detekujeme hodnoty minim. Uvedený postup je vysvětlen v popisu algoritmu 2, na obrázku 2 jsou graficky znázorněny kroky 3–8, které vedou k lepšímu pochopení algoritmu.

Popis algoritmu 2:

1. Určení f_0 pomocí Welchovy metody
2. Určení dominantní energie
3. Detekce všech maxim (minim) v segmentech signálu o délce T_0
4. Maxima (minima), která nejsou lokálními extrémy, jsou vynulována
5. Detekce maxim (minim) probíhá od středu signálu k jeho krajům
6. Detekce maxima mezi dvěma prvními detekovanými maximy v kroku 5, tedy posun přibližně o půl periody vede k odstranění špatné počáteční detekce, viz obrázek 2 krok 6, špičky označené čtverečky

7. Detekce maxim (minim) podle nově detekovaných pozic z bodu 6 od středu signálu k jeho krajům
8. Výběr mezi výsledky kroku 5 nebo kroku 7 (špičky s vyšší energií)

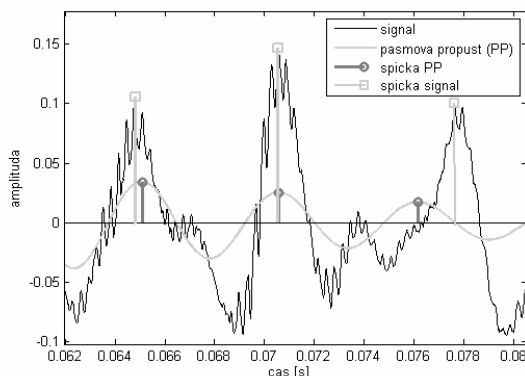
Hlavní problém při detekci špiček nastává s určením dominantní špičky. Ne vždy je dominantní špička správná špička vzhledem k předchozí detekci viz obr. 1.

C. Detekce špiček pomocí pásmové propusti

Následující algoritmus vede k detekci fundamentální frekvence a pozic špiček signálu pomocí pásmové propusti aplikované Butterworthovým filtrem 5. řádu.

Pásmová propust (popis algoritmu 3, obrázek 3) nalezne přibližné pozice maxim a zmenší tak výpočetní náročnost procesu. Správné pozice špiček jsou detekovány v okolí kolem maxim nalezených pásmovou propustí. Užití pásmové propusti je problematické, protože se mění frekvence v průběhu signálu. Proto je nutné signál segmentovat. Nejprve je vybrán úsek, jehož optimální délka je zvolena na 6 000 vzorků s překryvem 2 000 vzorků. Překryv je zvolen, protože začátek a konec pásmové propusti nedetekuje správné pozice maxim (dochází k náběhu a doběhu pásmové propusti).

Ve vybraném úseku je nalezena pomocí Welchovy metody fundamentální frekvence a v okolí této frekvence je pak užitá pásmová filtrace. Filtrace je aplikována s šířkou pásma ± 10 Hz od nalezené fundamentální frekvence. Filtrací dochází k potlačení formantových frekvencí a šumu. Správné pozice špiček jsou detekovány v okolí nalezených maxim pásmové propusti – obrázek 3.



Obrázek 3: Detekce špiček pomocí pásmové propusti na silně patologické promluvě

Popis algoritmu 3:

1. Nalezení průměrné fundamentální frekvence, začátku a konce promluvy
2. Výběr úseku a opětovná detekce f_0 úseku, pokud se liší o více jak 50 Hz od průměrné f_0 celého záznamu, je stanovena na její hodnotu

3. Aplikace pásmové propusti s šířkou pásma 10 Hz f_0 úseku
4. Nalezení maxim pásmovou propustí
5. Nalezení maxim v okolí maxim detekovaných pásmovou propustí, ($2/5 T_0$ před, $1/5 T_0$ za pozicí maxim, kde $T_0 = 1/f_0$)

4. Výsledky a experimenty

Fungujících algoritmů detekce f_0 je velké množství, ale žádný problematiku detekce špiček neřeší dostatečně komplexně a robustně. Algoritmy, které jsou přesné, nejsou automatizované. To znamená, že musíme nastavovat vnější parametry algoritmů a poté kontrolovat, zda byly tyto parametry správně nastaveny.

V této části se budeme zabývat výsledky navržených algoritmů detekce fundamentální frekvence, popsanych v předešlé kapitole.

Vzájemná korelace

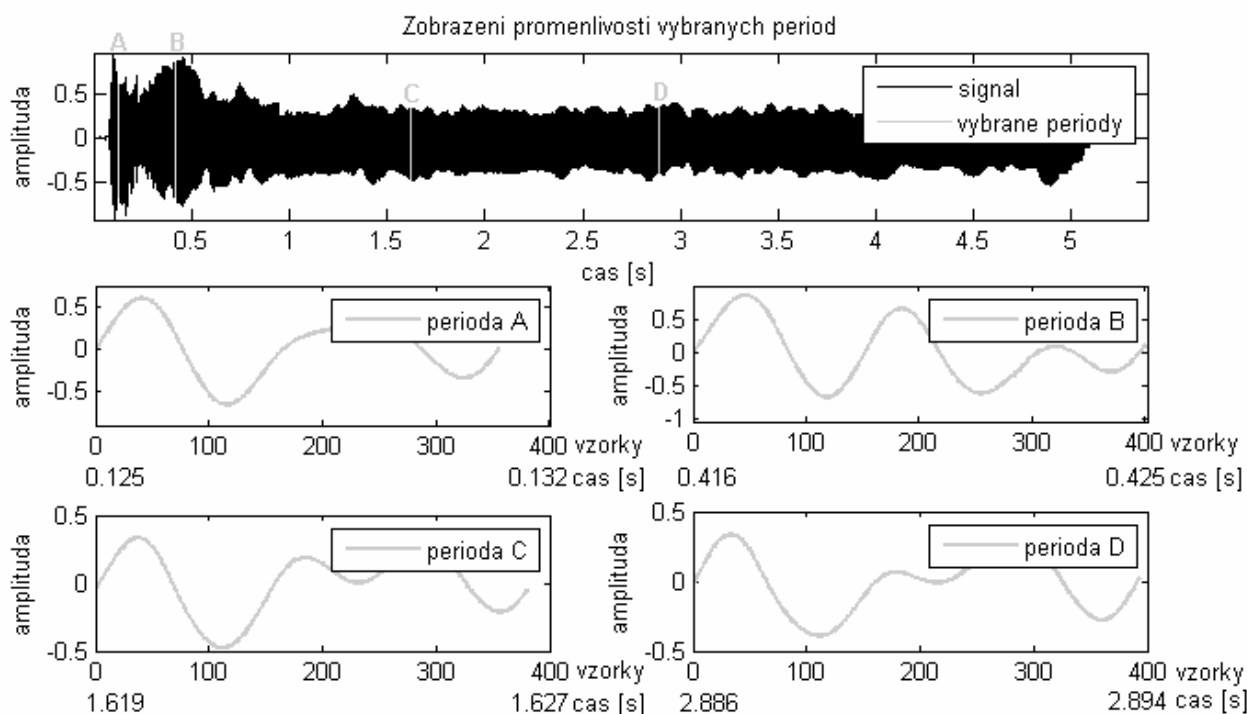
Prvotní záměr bylo užití normalizované vzájemné korelace k detekci f_0 . Detekce špiček pomocí vzájemné korelace ovšem naráží na problematiku výběru okna. V průběhu promluvy se signál mění takovým způsobem, že výběr „správné“ periody, kterou by bylo možné užít ke korelování, je poměrně obtížný. Proměnlivost period v průběhu signálu je zobrazena na obrázku 4. Vzhledem k měnící se délce periody a amplitud jednotlivých špiček dochází k posunutí „správné“ pozice špičky detekované pomocí korelace na chybnou pozici. V závislosti na změně velikosti amplitud pak může v dané periodě korelovat více formantová špička než „správná“ špička. Pokud by došlo k vzájemné korelaci úseku B s úsekem D z obrázku 4, došlo by ke zjevnému posunutí správně detekované pozice.

Vzhledem k chybné detekci se nedá přesná pozice určit a algoritmus může být pouze použit k detekci průměrné fundamentální frekvence signálu. Pro detekci průměrné f_0 je však metoda detekce pomocí vzájemné korelace oproti Welchově metodě zbytečně výpočetně náročná.

B. Detekce špiček pomocí maxim

Detekce špiček pomocí výskytu maxim v signálu funguje na základě nalezení pozic maxim v segmentovaném signálu. Tyto segmenty jsou stanovené jako $1/20$ délky průměrné pitch periody. Délka segmentu $1/20$ byla zvolena s ohledem na přesnost a paměťovou náročnost, neboť $1/10$ délky segmentu se nejevila jako dostatečná, protože docházelo k přeskokům na okolní chybné špičky. Zvolená maxima jsou pak kontrolována na výskyt lokálních extrémů. V případě detekce maxima, které není lokálním extrémem, dojde k vynulování jeho hodnoty.

Aby došlo ke snížení chybné míry detekce na začátcích signálu, jsou pozice špiček detekovány od středu ke krajům. V nalezených maximech je potom hledána optimální cesta po jednotlivých periodách, jejichž délka se opět mění



Obrázek 4: Proměnlivost period v průběhu signálu

adaptivně vzhledem k délce posledních čtyř nalezených period. Snahou je vyrušit chybnou detekci maxim hned od začátku. Toho je dosaženo nalezením maxima mezi dvěma dříve detekovanými maximy, tedy dochází k opětovnému detekování maxim, viz obr 2 dole. Jako správně detekovaná maxima jsou pak vybrány špičky s vyšším součtem amplitud, tedy s vyšší energií.

C. Detekce špiček pomocí pásmové propusti

Metoda detekce špiček pomocí pásmové propusti využívá segmentace signálu a nalezení přibližných pozic špiček pro filtrovaný signál. Tato detekce je pak zpřesněna v detekovaných okolicích nalezených maxim pásmovou propustí.

Metoda by měla vykazovat zrychlení celého procesu detekce, výsledky však teorii vyvracejí. Neustálé využívání filtrace signálu a detekce fundamentální frekvence ve vybraných úsecích je poměrně časově náročné. Při návrhu algoritmu musíme brát ohled i na výpočetní náročnost procesu, a tím jeho možnou využitelnost v klinické praxi.

4.1. Porovnání výsledků detekce vokálních parametrů

Závěrem je uvedeno porovnání výsledků detekce špiček pomocí pásmové propusti a pomocí maxim. Pomocí algoritmu Praatu jsou nalezeny jednotlivé periody signálu a v nich jsou poté označovány pozice špiček. Pozice špiček získané námi navrženými algoritmy jsou poté porovnány

s označovanými pozicemi. Z tabulky 2 vyplývají výsledky jednotlivých algoritmů. Nejlepších výsledků dosahuje metoda detekce špiček pomocí maxim.

Metoda	Maxima	PP	Vzájemná korelace
Přesnost	88,4 %	83,9 %	65,6 %
SD	14,77 %	17,83 %	18,99 %

Tabulka 2: Úspěšnost nově navržených metod detekce f_0

5. Závěr

V článku jsou prezentovány dva nové plně automatické algoritmy detekce základní hlasivkové frekvence a je otestována metoda detekce f_0 pomocí vzájemné korelace. Tyto algoritmy dále slouží k detekci ostatních vokálních parametrů.

Nově navržené algoritmy pracují metodou detekce špiček (v literatuře uváděno jako metoda peak-picking). Algoritmy dosahují více než 83% shody na označované databázi.

Neefektivní přístup algoritmu k signálu spočívá v problematice schopnosti reagovat na proměnlivost amplitud jednotlivých špiček vzhledem k nalezeným pozicím v předchozích periodách. Zde se nabízí řešení ohodnotit jednotlivé pozice špiček vzhledem ke vzdálenosti od předpokládané pozice špičky. Ke zlepšení algoritmu by měla vést

detekce více příznaků v jednotlivých periodách, tak aby pozice period nebyly určovány jenom z polohy špičky, ale v širším měřítku. Například místo průchodu signálu nulou, okamžik uzavírání hlasivek.

Poděkování

Tato práce je podporována z grantů SGS10/180/OHK3/2T/13, GAČR 102/08/H008, NT 11331-6/2010 a NT 12288-5/2011, z výzkumných záměrů MSM 0021620849 a MSM 6840770012.

Reference

- [1] Kotnik, B., Höge, H., Kačič, Z.: Noise robust f_0 determination and epoch-marking algorithms, *Signal-Process*, 2555–2569, 2009
- [2] Hess, W.: Pitch determination of Speech Signals *Springer-Verlag*, Berlin, 1983
- [3] Huang, N. E., Shen, Z., Long, S. R., Wu, M. L., Shin, H. H., Zheng, Q., Yen, N. C., Tung, C. C., Liu, H. H.: The empirical mode decomposition an Hilbert spectrum for nonlinear and nonstationary time series analysis, *Proceedings of Royal Society London A*, 454: 903–995, 1998
- [4] Kay Elemetrics Corp.: Multi-Dimensional Voice Program (MDVP) Model 5106: Software Instruction Manual. Lincoln Park, Kay Elemetrics, 2003
- [5] Boersma, P.: Praat, a system for doing phonetics by computer. *Glott Int* 2001; 5:341–345
- [6] Boersma, P.: Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound, In *Proceedings of the Institute of Phonetics Sciences*, 17: 97–112, 1993
- [7] Uhlíř, J., Sovka, P., Pollák, P., Hanžl, V., Čmejla, R.: *Technologie hlasových komunikací*, Nakladatelství ČVUT, Praha 2007

Hodnocení vokalických parametrů u patologických hlasů

Lukáš Bauer^a, Jan Ruzs^{a,b} a Roman Čmejla^a

^aČVUT – FEL, Katedra teorie obvodů, Technická 2, 166 27 Praha 6

^bUniverzita Karlova v Praze, Neurologická klinika 1. LF UK a VFN, Kateřinská 30, 128 21 Praha 2
e-mail: bauerlu3@fel.cvut.cz

In this paper we present usage of the two novel speech fundamental frequency (f_0) detection methods for sustained phonations to assess other vocal parameter such as jitter, shimmer, and harmonics-to-noise ratio (HNR). These voice parameters are commonly used to detect the speaker's voice characteristics and his voice quality. The aim of the work is to design the algorithm for detecting these vocal parameters, regarding to the robustness and automatization with minimisation of external parameters set-up. The results of particular parameters are compared with commonly used software Praat. The comparison of results between Praat and our designed vocal parameters were 83 % for jitter, 88 % for shimmer and 94 % for HNR.

1. Úvod

Řada řečových patologií je spojena s prolínáním normálních a nepravidelných vibrací hlasivek. Tento jev snižuje kvalitu hlasu a je jedním z prvotních příznaků vyskytujících se poruchy řečového ústrojí. Vzhledem k vysoké časové výpočetní náročnosti bylo v minulosti možné posuzovat kvalitu hlasu jenom rozdíly vnímatelnými sluchem. Proto potřeba vzniku automatického objektivního hodnocení kvality hlasu, poskytujícího robustní a spolehlivé výsledky řečových parametrů, je nyní vysoce aktuální.

Podmínkou pro posouzení kvality hlasu je nahrávka stabilní prodloužené fonace samohlásky po co nejdélnější možnou dobu. Menší změny ve frekvenci a amplitudě hlasového signálu jsou vždy přítomné, i když se snažíme vyslovit prodlouženou fonaci samohlásky perfektně stabilně [1].

U pacientů s hlasovými problémy se mohou tyto amplitudové a frekvenční hlasové výkyvy významně zhoršit. Poslechově pak mohou být tyto příznaky interpretovány jako dysfonie charakterizovaná chrapotem, neznělostí a nezvuchností hlasu. Používanými měřeními pro vyhodnocení příznaků dysfonie jsou jitter (frekvenční nestabilita) a shimmer (amplitudová nestabilita), které jsou založeny na krátkodobé proměnlivosti základní hlasivkové frekvence f_0 a amplitudě zvukového signálu. Rozsáhlý rozbor tohoto tématu je možné najít v literatuře [2]. Dalším možným měřením dysfonie je parametr HNR (odstup harmonických složek od šumu – harmonics-to-noise ratio), který zaznamenává poměr harmonické a šumové složky signálu [3].

Motivací měření těchto tří vokalických parametrů je např. zpětná vazba při terapii a léčbě řady poruch [4].

2. Metody

2.1. Databáze

Databáze signálů byla pořízena v odrušené místnosti Neurologické kliniky 1. LF UK pomocí kondenzátorového kamerového mikrofону kamery značky Panasonic NV-GS 180. Mikrofon byl umístěn ve vzdálenosti 15 cm

od úst mluvčího. Signál je vzorkován frekvencí 48 kHz a k převodu signálu byl použit 16bitový převodník.

Protože detekce jitteru a shimmeru se provádí hlavně na nahrávkách prodloužených fonací [5], skládá se databáze signálů z nahrávek prodloužených fonací vokálů /a/, /e/, /i/ a /o/. Tím odpadá nutnost detekce znělých a neznělých úseků.

Délka záznamů databáze se pohybuje od 5 do 40 sekund (mean = 13,4; SD = 9,5). Databáze obsahuje 107 (90 mužských a 17 ženských) fonací pacientů s ranou Parkinsonovou nemocí (PN) a 57 (42 mužských a 15 ženských) fonací zdravých mluvčích kontrolní skupiny (KS) [6].

Fonace	KS		PN	
	Muži	Ženy	Muži	Ženy
a	8	3	18	3
e	2	0	19	2
i	30	12	50	12
o	2	0	3	0

Tabulka 1: Databáze záznamů [6]

2.2. Algoritmy výpočtu jitteru, shimmeru a HNR

Základem správné detekce řady vokalických parametrů je robustní detekce základní hlasivkové frekvence f_0 . Aktuálně je v klinické praxi pro výpočty hlasových charakteristik využíván komerční program Multi-Dimensional Voice Program [7]. V akademickém světě je pak běžně používán volně šiřitelný program Praat [8]. Metody diagnostiky řečových poruch pomocí kategorií fonace, prosodie a artikulace, které jsou spojeny se závažnými poruchami centrální nervové soustavy, však nejsou volně dostupné. Vývoj těchto metod povede k objektivizaci a detekci patologické řeči.

Jako vybrané metody detekce základní hlasivkové frekvence jsou použity metody detekce špiček pomocí „maxim“ a pomocí „pásmové propusti (PP)“ [9]. Metoda detekce maxim vyhledává optimální cestu v pozicích špiček

nalezených po 1/20 průměrné pitch periody signálu. Metoda detekce pomocí pásmové propusti vyhledává pozice špiček v okolí maxim nalezených filtrací signálu pomocí filtrace pásmovou propustí.

Pomocí algoritmů detekce špiček jsme nyní schopni dále detekovat jitter, shimmer a HNR. Relevantní hodnoty však dostaneme pouze pomocí řádně detekovaných pozic špiček.

Jitter Hovoříme-li v řečovém signálu o detekci jitteru, jedná se o frekvenční nestabilitu signálu, tedy o změny v rychlosti kmitání hlasivek. Detekce správných hodnot jitteru závisí na robustní detekci f_0 . Pokud dojde k chybné detekci jednotlivých period v signálu, dojde k velkým nepřesnostem v detekci hodnot jitteru.

Při detekci jitteru musíme také hledět na dostatečně vysokou hodnotu vzorkovací frekvence. Pokud se vzorkovací frekvence pohybuje pod 20 kHz, může docházet k chybám při detekci jitteru [10].

V praxi je běžně uváděno pět druhů měření jitteru [8]. V této práci je vybrán jako reprezentativní relativní jitter (Jitter:rel), který je vyjádřen jako absolutní rozdíl následujících period T_i a T_{i+1} podělený průměrnou periodou

$$\text{Jitter:rel} = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i+1}|}{\frac{1}{N} \sum_{i=1}^N T_i} \cdot 100\%, \quad (1)$$

kde N je celkový počet period. Jitter je vypočten z detekovaných špiček pomocí metody maxim a pásmové propusti.

Shimmer Vyjadřuje amplitudovou nestabilitu signálu, která určuje míru rozdílnosti energie jednotlivých period v signálu. Velkým problémem při detekci shimmeru je šum obsažený v signálu, který může změnit hodnotu amplitudy. Pokud bychom tento signál chtěli filtrovat, dochází k posunutí a změně amplitudy.

V praxi jsou běžně uváděny čtyři druhy měření shimmeru [8]. V práci je vybrán jako reprezentativní shimmer v dB (ShdB), který je představován dekadickým logaritmem podílu dvou po sobě jdoucích amplitud A_i a A_{i+1}

$$\text{ShdB} = \frac{1}{N-1} \sum_{i=1}^{N-1} |20 \log(A_{i+1}/A_i)| \quad (\text{dB}). \quad (2)$$

Hodnoty shimmeru jsou určovány z poměrů detekovaných špiček vzhledem k nule [11]. Shimmer je vypočten z detekovaných špiček pomocí metody maxim a pásmové propusti.

HNR Fonace zdravého jedince vykazuje vysokou periodičnost, zatímco v patologickém signálu je vždy větší přítomnost neharmonických složek. Z toho důvodu je jako další vokální parametr měřen odstup harmonické složky od šumu (HNR). HNR se vypočítá jako podíl energie periodické E_p části signálu a energie aperiodické E_{ap} části

$$\text{HNR} = 10 \log(E_p/E_{ap}) \quad (\text{dB}). \quad (3)$$

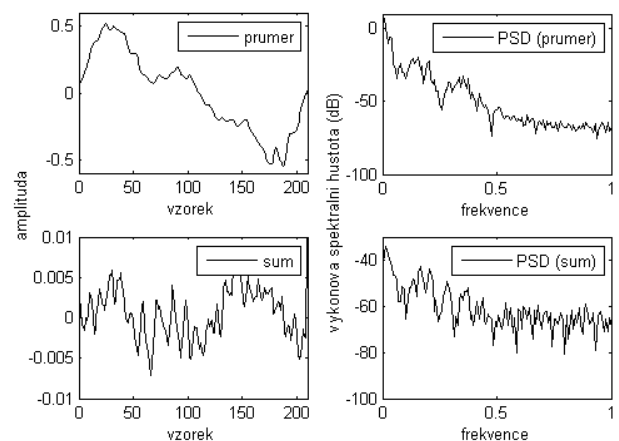
Algoritmy detekce HNR pracují jak v časové, tak ve frekvenční oblasti. Algoritmy pracující ve frekvenční oblasti mají výhodu ve zvolení délky detekovaného okna. Zatímco u algoritmů pracujících v časové oblasti je nutné zvolit okno přesné délky, ve frekvenční oblasti je možné vybrat okno o určité délce, která není násobkem počtu period.

Základním problémem při výpočtu HNR je nalezení šumové složky v signálu. Dominantní část řečového signálu se vyskytuje do 2500 Hz, u některých pacientů až do 3500 Hz, zatímco dominantní část šumové složky řečového signálu se vyskytuje na frekvencích nad 2500 Hz [10]. Algoritmy vypočítávající HNR ve frekvenční oblasti používají často k výpočtu kepstrální techniky zvýrazňující harmonické frekvence [12], kde sedla mezi harmonickými jsou brána jako šumová složka. Problém nastává při výskytu nízké fundamentální frekvence řečového signálu, v důsledku čehož dochází k zúžení sedlového pásma a zvýšení hodnoty HNR.

Určování spektrální obálky v okolí harmonické složky je poměrně komplikované, protože některé harmonické části jsou široké 20 Hz, některé 40 Hz. Občas dochází také k superpozici formantových frekvencí, které nepatří ani do harmonické části, ani do šumové části.

Protože je algoritmus pro detekci HNR komplexnější, nelze vyjít pouze z čistě detekovaných pozic špiček jako v případě jitteru a shimmeru. Námí používaný algoritmus detekce HNR pracuje v časové oblasti. Algoritmem detekce špiček jsou nejdříve detekována maxima jednotlivých period. Poté jsou určeny délky jednotlivých period jako vzdálenost dvou průchodů signálu nulou před pozicemi nalezených maxim. Délky jednotlivých period jsou poté doplněny nulami na délku nejdelší detekované periody v signálu.

Jako harmonický signál je určen průměr šesti po sobě jdoucích period. Od každé periody je poté tento průměr odečten a je z něj spočítána šumová složka (obrázek 1). Hodnota HNR je určována s překryvem tří period, aby



Obrázek 1: Zobrazení detekované periodické a aperiodické části jedné periody

byla eliminována chyba zanášená změnami v signálu. Největší chybu do této metody zanáší chybná detekce špiček, z které vyplývá špatné určení period pomocí počtu průchodu nulou.

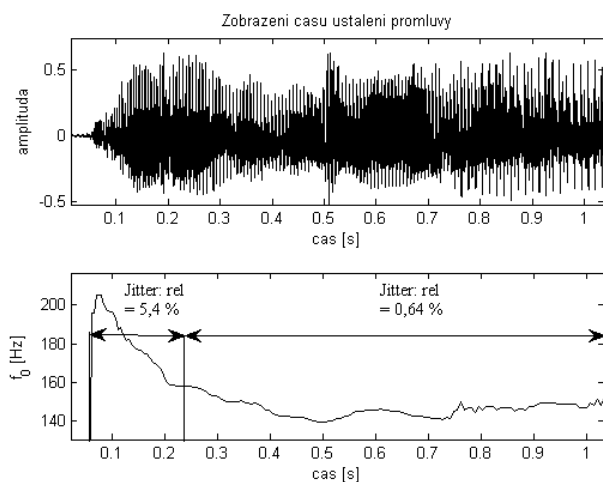
Hodnoty námi detekovaných reprezentativních řečových parametrů, upozorňující na možný výskyt řečové patologie, jsou zobrazeny v tabulce 2.

Parametr	Práh
Jitter:rel	$\leq 1,04\%$
ShdB	$\leq 0,35\text{ dB}$
HNR	$> 20\text{ dB}$

Tabulka 2: Patologické prahy řečových parametrů [7]

2.3. Testování přesnosti algoritmů

Jako referenční program, s kterým jsou porovnávány výsledky, je vybrán volně šiřitelný Praat [8]. Je však nutné brát v úvahu, že hodnoty získané Praatem jsou určeny v závislosti na manuálním nastavení parametrů Praatu, tak aby tyto hodnoty vycházely co nejpřesněji. Detekce pomocí Praatu tedy není plně automatická.



Obrázek 2: Zobrazení nestacionarity signálu na začátku fonace

2.4. Testování začátku, konce a ustálené části fonace

Řečový signál se v průběhu fonace mění. Velká variabilita nastává na začátku fonace při hlasovém nasazení a na konci fonace, kdy už člověku může docházet dech a u pacientů s PN dochází ke zjevné únavě hlasivek. Na obrázku 2 je zobrazen začátek řečové fonace, na kterém je vidět, že ustálení fonace a výskyt periodicity signálu nastává zhruba kolem 250 ms.

Pro detekci vokálních parametrů je nutné mít známou delší než čtyři sekundy, a to zejména z důvodu nutného ustálení prodloužené fonace. Fonace u jedinců

s výskytem poruch hlasu se projeví až následně poklesem amplitud v záznamu.

Ve většině případů dochází k vynechání začátku a konce fonace, abychom obdrželi stabilní úsek prodloužené fonace. To je ve většině případů prováděno ručně se zobrazením časového průběhu, založeného na přítomnosti znaku nestability (např. neznělé segmenty, hlasové mezery, příliš vysoké nebo nízké nasazení hlasu apod.) [13].

Vzhledem k tomu, že tyto nestabilní úseky vznikají většinou těsně na začátku nebo konci fonace, vedla v této práci snaha o automatickou detekci k experimentálnímu výběru časového úseku 250 ms, po kterém ve většině případů dojde k potřebnému ustálení.

Abychom objasnili, zda je vhodné vynechat prvních a posledních 250 ms záznamu, provedli jsme testy závislosti patologických a zdravých fonací jednotlivých vokálních parametrů v daných časových úsecích (začátek, ustálená část, konec fonace). Tímto testem lze také odhalit výskyt důležité informace v těchto částech fonace.

2.5. Statistika

Pro porovnání výsledků jednotlivých měření vokálních parametrů s nalezenými hodnotami pomocí Praatu je použito Pearsonova korelačního koeficientu. Test je prováděn na základě parametrické korelace.

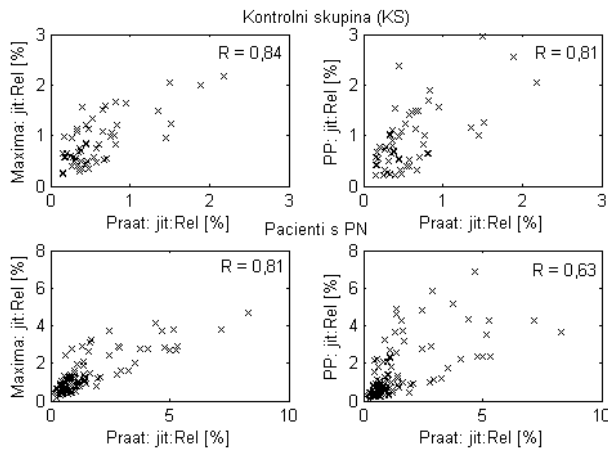
Pro porovnání rozdílů mezi jednotlivými úseky záznamů pro PN a kontrolní skupiny je použit Wilcoxonův neparametrický test. Oba testy jsou provedeny na 5% hladině významnosti.

3. Výsledky

Ačkoliv fungujících algoritmů detekujících f_0 a vokálních parametrů je velké množství, žádný z nich problematiku neřeší dostatečně komplexně a robustně. Obzvláště detekce parametru HNR je ve světě poměrně málo dokumentovaná, tudíž je navrženo malé množství funkčních metod, se kterými by bylo možné porovnávat výsledky. U většiny těchto metod nejsou dostupné zdrojové kódy a popsání metodiky není dostatečně přesné, a proto je časově náročné reprodukovat jejich principy.

3.1. Detekce jitteru, shimmeru a HNR

Jitter Z výsledků algoritmu detekce jitteru porovnávaného s hodnotami získanými pomocí programu Praat vyplývá, že detekce hodnot jitteru velice závisí na detekovaných pozicích špiček. Výsledné hodnoty detekce jitteru pomocí metody maxim dosahují ($R = 0,83$; $p < 0,001$) a pomocí metody pásmové propusti dosahují ($R = 0,66$; $p < 0,001$). Na obrázku 3 jsou zobrazeny výsledné hodnoty detekovaného jitteru pomocí programu Praat a metod maxim a pásmové propusti pro kontrolní skupinu a pacienty s PN. Z obrázku je patrné, že při vyšších hodnotách jitteru se algoritmy začínají významně rozcházet s výsledky

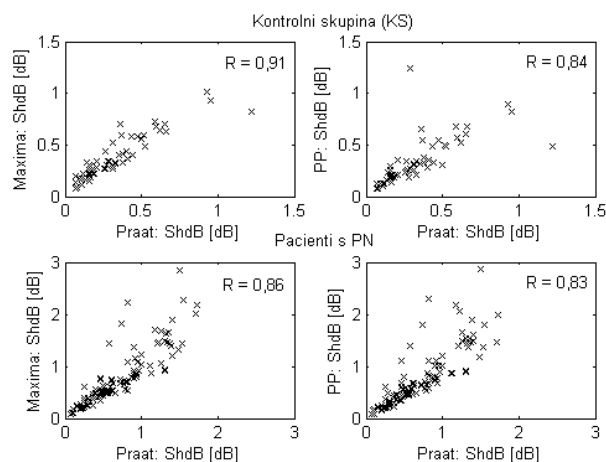


Obrázek 3: Graf porovnání naměřených hodnot jitteru s programem Praat

Praatu. U takto vysokých hodnot, je však výskyt patologie nesporný. Toto je s největší pravděpodobností dáno detekcí zašuměných míst fonací Praatem. V těchto místech pak není jitter vyhodnocován, čímž dochází ke snížení hodnot jitteru. Samotný Praat podhodnocuje hodnoty jitteru [14].

Shimmer Výsledky algoritmu detekce shimmeru dosahují vyšší procentuální úspěšnosti než výsledky detekce jitteru. Z toho vyplývá, že detekce shimmeru není tolik závislá na přesné detekci pozic špiček.

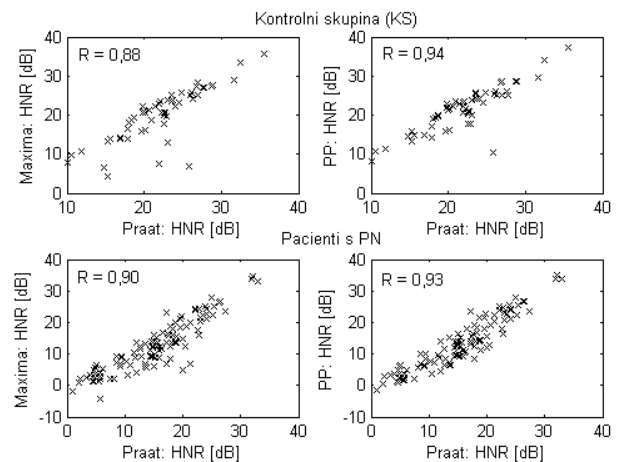
Výsledné hodnoty detekce shimmeru pomocí metody maxim dosahují ($R = 0,88$) a pomocí metody pásmové propusti dosahují ($R = 0,86$). V obou případech je hladina významnosti $p < 0,001$. Na obrázku 4 jsou zobrazeny výsledné hodnoty detekce shimmeru. Z obrázku vyplývá, že hodnoty detekce dosahují srovnatelných výsledků pro porovnávané algoritmy detekce špiček pomocí maxim a



Obrázek 4: Graf porovnání naměřených hodnot shimmeru s programem Praat

pásmové propusti. Pro algoritmus pásmové propusti pak dosahují horších výsledků pro vyšší hodnoty shimmeru.

HNR Výsledné hodnoty detekce HNR pomocí metody maxim dosahují ($R = 0,91$; $p < 0,001$) a pomocí metody pásmové propusti dosahují ($R = 0,94$; $p < 0,001$). Na obrázku 5 jsou k porovnání výsledky detekce HNR. Výsledky ukazují mírné podhodnocení hodnot detekovaného HNR oběma námi navrženými metodami oproti Praatu. Míra korelace detekovaných hodnot dosahuje nejlepších výsledků ze všech detekovaných vokálních parametrů. Oba algoritmy dosahují u toho parametru nejmenších rozdílů detekovaných hodnot oproti ostatním vokálním parametrům.



Obrázek 5: Graf porovnání naměřených hodnot HNR s programem Praat

Porovnání výsledků detekce vokálních parametrů Závěrem je uvedeno porovnání výsledků detekce špiček pomocí pásmové propusti a pomocí maxim s výsledky získanými programem Praat. V obou případech je dosaženo středně vysokých až vysokých hodnot korelace. Míra korelace shimmeru a jitteru vychází vyšší pro detektor špiček. Míra korelace pro HNR vychází vyšší pro detekci pásmovou propustí. Z výsledků vyplývá, že metoda detekce špiček pomocí pásmové propusti je vhodnější na detekci parametru HNR. Detekce špiček pomocí maxim ale dosahuje lepších a stabilnějších výsledků u ostatních parametrů. Každá metoda má tak své přednosti a své nedostatky.

V tabulce 3 jsou zobrazeny míry korelace jednotlivých parametrů pro kontrolní skupinu a pacienty s PN metodami detekce maxim a pásmové propusti. Z porovnání tabulky 3 a 4 vyplývá, že celková detekce jednotlivých parametrů není závislá na pohlaví pacienta.

Parametr	Maxima			Pásmová propust		
	KS	PN	Celk	KS	PN	Celk
Jitter (%)	0,84*	0,81*	0,83*	0,81*	0,63*	0,66*
Shimmer (dB)	0,91*	0,86*	0,88*	0,84*	0,83*	0,86*
HNR (dB)	0,88*	0,90*	0,91*	0,94*	0,93*	0,94*

* $p < 0,001$

Tabulka 3: Míry korelace jednotlivých parametrů pro zdravé a nemocné jedince metodami detekce maxim a pásmové propusti

Parametr	Maxima	Pásmová propust
Muži		
Jitt:Rel (%)	0,81*	0,67*
ShdB (dB)	0,88*	0,85*
HNR (dB)	0,91*	0,93*
Ženy		
Jitt:Rel (%)	0,82*	0,42*
ShdB (dB)	0,85*	0,84*
HNR (dB)	0,83*	0,93*

* $p < 0,001$

Tabulka 4: Zobrazení úspěšnosti algoritmů v závislosti na pohlaví

3.2. Testování začátku, konce a ustálené části fonace

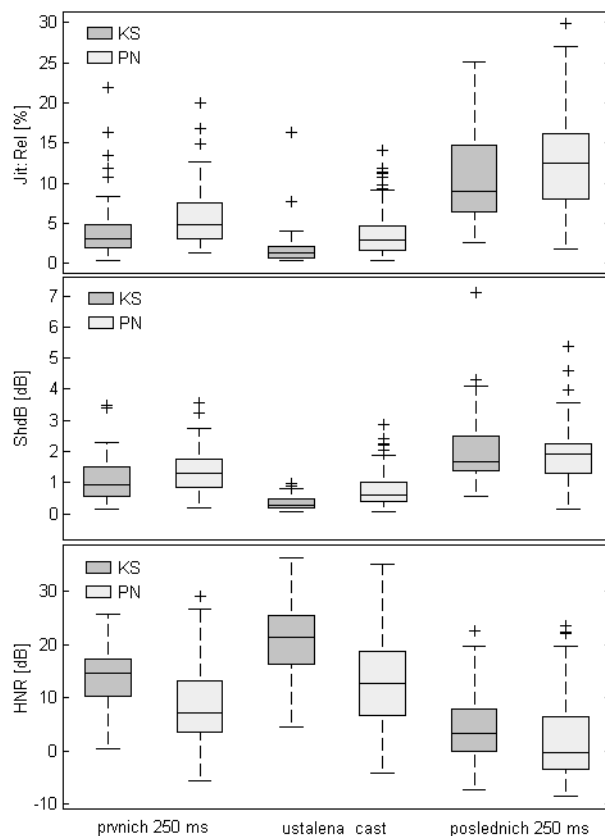
Z obrázku 6 vyplývá větší závislost středních úseků než začátků fonace. Naopak konce fonace jsou velice závislé jak pro patologické, tak pro zdravé fonace. To je pravděpodobně způsobeno snahou pacienta udržet fonaci co nejdéle, což zanáší značné změny do nahrávky. Tyto změny jsou však pro oba typy fonace shodné.

4. Diskuze

Je vůbec sporné, zda se dá jitter v patologických fonacích detekovat podle pozic maxim. Z porovnání dvou metod detekce špiček MDVP (metoda peak-picking) a Praat (metoda waveform-matching) [14] vyplývá, že detekce jitteru pomocí špiček je přesná a dosahuje podobných výsledků pouze do určitých hodnot jitteru a poté nastává značný rozdíl ve výsledcích. V [15] dochází k porovnání detekovaných hodnot jitteru pomocí MDVP a Praat. Výsledky porovnání vycházejí téměř shodně s našimi, tedy u vyšších hodnot jitteru dochází k velkému rozdílu MDVP a Praat.

Z tohoto důvodu je vhodnější využívat algoritmus detekce pomocí podobnosti signálu (waveform-matchingu). Zatímco parametr HNR tolik netrpí nepřesnou detekcí pozic špiček, u jitteru dochází k značnému zkreslení. Z tohoto důvodu vyplývá, že detekce jitteru by měla probíhat pomocí detekce dalších pozic než pouze maxim v signálu.

Je důležité si ovšem uvědomit, že neexistují žádné stanovené postupy detekce a záleží pouze na deterministicky zvolené funkci algoritmu, tudíž ani hodnoty Praatu či MDVP nemusí přesně určovat správnou hodnotu jednotli-



Obrázek 6: Porovnání závislosti detekce vokálních parametrů na různých úsecích zdravých a patologických fonací. V horní části pro parametr jitter, ve střední části pro parametr shimmer a ve spodní části pro parametr HNR

vých měření. Cílem daného algoritmu tedy je, aby dokázal správně interpretovat fyziologii dané fonace a dosáhl dostatečné přesnosti, a to s použitím co nejmenšího počtu nutně definovaných parametrů, které musí být uživatelem definovány (tzv. automatizace algoritmu).

5. Závěr

V článku jsou prezentovány nové algoritmy pro hodnocení vokálních parametrů prodloužené fonace s využitím databáze zdravých mluvčích a pacientů s výskytem rané fáze PN. Zvolené algoritmy vycházejí z výsledků detekce f_0 pomocí špiček [9]. Výsledky algoritmů dosáhly středně silných až vysokých korelací ve srovnání s jedním ze světově uznávaných algoritmů programu Praat.

Algoritmy vykazují podobné výsledky detekce řečových parametrů. Algoritmus detekce špiček pomocí pásmové propusti vykazuje lepší výsledky detekce parametru HNR, kde dosahuje téměř 94% úspěšnosti jak pro zdravé, tak pro patologické pacienty. Algoritmus detekce špiček pomocí maxim pak vykazuje stabilní výsledky pro všechny parametry převyšující 80% úspěšnost shody s programem Praat.

Z testování začátků, ustálených částí a konců fonací pro zdravé a patologické fonace vyplývá, že v ustálených částech fonací je nesena největší informace o vokálních parametrech. Začátky fonací pacientů s PN a kontrolní skupiny jsou méně závislé, proto je vhodné je v tomto případě vynechat. Závěrem lze tedy říci, že pokud nedochází k testování poruch projevujících se hlavně v začátcích fonace, je vhodné vynechat prvních i posledních 250 ms fonace, v kterých nastává velká změna signálu. Tyto části fonace neobsahují pro nás dostatečně důležitou informaci.

Ke zlepšení metod detekce vokálních parametrů by mohla být použita obálka signálu, podobně jako je využito v [11], kde tento postup potlačuje saturované a zašuměné špičky. K dalšímu zpřesnění by došlo použitím detekce parametrů z více příznaků jedné periody.

Poděkování

Tato práce je podporována z grantů SGS10/180/OHK3/2T/13, GAČR 102/08/H008, NT 11331-6/2010 a NT 12288-5/2011 a z vybraných výzkumných záměrů MSM 0021620849 a MSM 6840770012.

Reference

- [1] Titze I. R.: Principles of Voice Production. *Englewood Cliffs*, Prentice-Hall, 1994
- [2] Baken R. J., Orlikoff R. F.: Clinical Measurement of Speech and Voice. *Singular Publishing Group*, San Diego, 2000
- [3] Yumoto E., Gould W. J., Baer T.: Harmonics to noise ratio as an index of the degree of hoarseness, *The Journal of the Acoustical Society of America*, vol. 71, no. 6, pp. 1544–1550, 1982
- [4] Novák A.: Foniatrie a pedaudiologie II. Poruchy hlasu u dětí a dospělých – základy anatomie a fyziologie hlasu, diagnostika, léčba, reedukace a rehabilitace poruch hlasu. *Unitisk*, Praha, 2000
- [5] Li X., Tao J., Johnson M. T., Soltis J., Savage A., Leong K. M., Newman J. D.: Stress and emotion classification using jitter and shimmer features, In *International Conference on Acoustics Speech and Signal Processing 2007 (ICASSP07)*, Honolulu, Hawaii, 2007
- [6] Ruzs J., Čmejla R., Ružičková H., Klempír J., Majeřová V., Picmausová J., Roth J., Ružička E.: Acoustic assessment of voice and speech disorders in Parkinson's disease through quick vocal test, *Movement Disorders*, in press, 2011
- [7] Kay Elemetrics Corp: Multi-Dimensional Voice Program (MDVP) Model 5106: *Software Instruction Manual Lincoln Park*, Kay Elemetrics, 2003
- [8] Boersma P.: Praat, a system for doing phonetics by computer. *Glott Int*, 5:341–345, 2001
- [9] Bauer L.: Hodnocení vokálních parametrů u patologických hlasů, ČVUT–FEL, *Diplomová práce*, 2011
- [10] Hillebrand J.: A methodological study of perturbation and additive noise in synthetically generated voice signals, *Journal of Speech Hearing Research*, 30:448–461, 1987
- [11] Boersma P.: Stemmen meten met Praat, *Stem-, Spraak- en Taalpathologie*, 12:237–251, 2004
- [12] Severin F., Boykurt B., Dutoit T.: HNR extraction in voiced speech, oriented towards voice quality analysis, In *Proc. EUSIPCO, European Signal Processing Conference, Antalya (Turkey)*, 2005
- [13] Marin Y., Corthals P., Bodt De M., Cauwenberge, von P.: Perturbation Measures of Voice: A Comparative Study between Multi-Dimensional Voice Program and Praat, *Folia Phoniatrica et Logopedica*; 61:217–226, 2009
- [14] Boersma P.: Should Jitter Be Measured by Peak Picking or by Waveform Matching?, *Folia Phoniatrica et Logopaedica*, 61:5, 2009
- [15] Amir O., Wolf M., Amir N.: A clinical comparison between two acoustic analysis softwares: MDVP and Praat, *Biomedical Signal Processing and Control*, Volume 4, Issue 3, Pages 202–205, July 2009

Efekt zvyšování hlasitosti na spektrální charakteristiky hlasu u různých typů použití hlasu a u různých skupin hlasových profesí

Marek Frič

Výzkumné centrum hudební akustiky (MARC)

Zvukové studio Hudební fakulty Akademie múzických umění v Praze, Malostranské nám. 13, 118 00 Praha 1
e-mail: marek.fric@hamu.cz

A gain factor (GF) has previously been determined as an amount of gain per frequency band related to an increase in overall sound pressure level. There has been published the dependence of GF on the vocal loudness, subglottal pressure or vocal effort variation, acoustic spaces, gender and age differences. This study examines changes of GF in 3 different voice tasks (habitual speaking voice, supported speaking voice and singing) and in 3 groups of voice users, separately for both genders: professional operatic singers, professional actors and a nonprofessional control group of 20 students of 1st class of acting. GF was calculated from third-octave spectral characteristics of manually extracted long vowels /a:/ (approximately 300 ms duration) when gradually increasing loudness of the word /ma:ma/. Mean overall GF values were compared by Student's *t*-test for the 3 tasks and the 3 groups. Generally, no statistical difference was found in GF between habitual and supported speaking in all groups of subjects, significant difference was found between speaking and singing. Results suggest that different groups use different spectral gain strategies in different voice tasks. From reported results could be presumed that different strategies in gradual increasing of the voice loudness divided spectrum in 5 specific bands.

1. Úvod

Jedním ze základních požadavků hlasových profesionálů (nejenom herců a zpěváků, ale i učitelů, manažerů a prodáváčů) je přizpůsobení se zvýšené hlasové zátěži v jejich každodenní praxi. Pracovní podmínky při profesním používání hlasu, nutnost vystupování před širším publikem a ve větších prostorech, vyžadují hlavně zvýšení hlasitosti.

Při změně hlasitosti hlasu se obecně uplatňují tři vázané systémy: ovládání výdechu, fonační nastavení a nastavení vokálního traktu. Ovládání výdechového tlaku a průtoku vzduchu z plic v konečném důsledku mění subglotický (podhlasivkový) tlak, který je základním faktorem změny hlasitosti [1]. Fonační nastavení na úrovni hrtanu, v závislosti na způsobu nastavení napětí hlasivek, ovlivňuje proudění vzduchu přes hlasivkovou štěrbinu (glottal flow). Nastavení rezonančních vlastností vokálního traktu ovlivňuje způsob zesílení jednotlivých spektrálních oblastí.

Na úrovni hlasivek je za hlavní faktor podílející se na ovládání hlasitosti považována míra jejich dovržení (addukce). S nárůstem addukce hlasivek se zvyšuje aerodynamický odpor hlasivek vůči proudění vzduchu z plic, a tím se zvětšuje subglotický tlak. To znamená, že při rozevření hlasivkové štěrbinou proudí mezi hlasivkami vzduch větší rychlostí. Hlasivkový puls má tedy větší amplitudu. V modálním (hrudním) rejstříku se s nárůstem hlasitosti zkracuje doba otevření hlasivek (open quotient – OQ) [2]. Vyšší míra addukce zvyšuje rychlost uzavírání hlasivek, čímž se prohlubuje záporný vrchol první derivace „glottal flow“ [3, 4] – označovaný jako „Maximum flow declination rate“ (MFDR).

Z akustického hlediska se zvyšování hlasitosti projevuje nárůstem hladin harmonických zejména vyšších spektrálních oblastí [1, 5, 6].

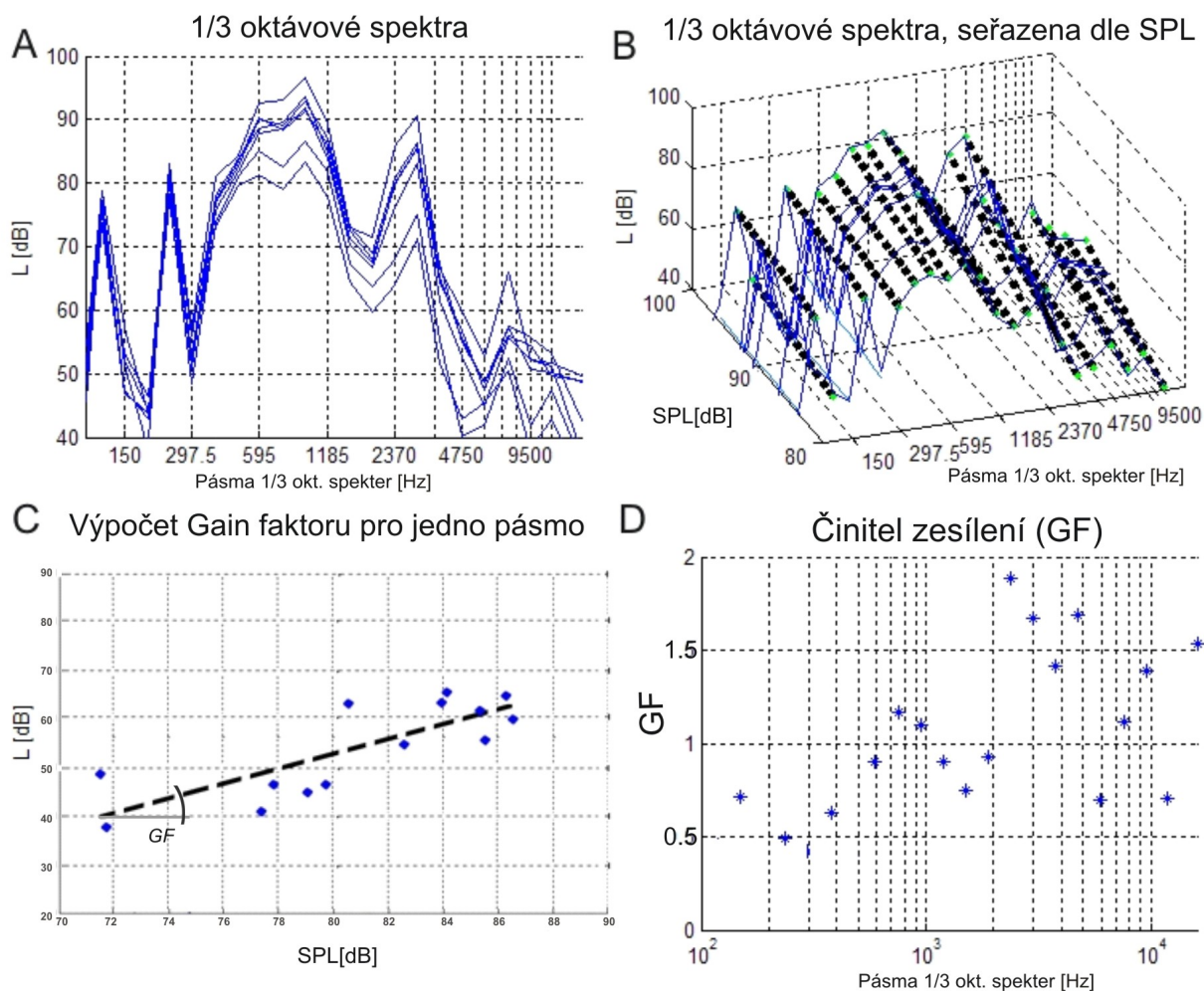
Při analýze snižování celkové hlasitosti (decrecendo vokálů /ae, a/) byla u profesionálních barytonů prokázána existence lineárního vztahu mezi rozdílem hladin akustického tlaku (SPL) spektrálních oblastí prvního formantu a pěveckého formantu (L1–LSF) a logaritmem subglotického tlaku [1].

Ternström prokázal, že při přirozeném zvyšování hlasitosti (při volání nebo při zvyšování námahy hlasu (voice effort)) narůstá základní frekvence, narůstají střední frekvence formantů F_1 , F_3 a prodlužuje se délka vokálů [6].

Pro účely hodnocení závislosti nárůstu hladin jednotlivých spektrálních oblastí vzhledem k celkové hladině akustického tlaku dB (SPL) byl zaveden **činitel zesílení GF** (gain factor) [6]. GF odpovídá směrnici přímkou proložené přes hodnoty akustických hladin sledovaného spektrálního pásma vztahených k celkové hladině akustického tlaku v dB (SPL) – viz rovnice (1), grafická interpretace – viz obrázek 1C

$$GF_{\text{pásma}} = \frac{\Delta L_{\text{pásma}}}{\Delta \text{SPL}} . \quad (1)$$

Cílem této studie je zjištění základních strategií použití hlasu při zvyšování hlasitosti a jejich akustická charakterizace pomocí sledování hodnot činitelů zesílení u různých hlasových profesionálů. Herci a zpěváci reprezentují specificky trénovanou skupinu subjektů, naproti tomu začínající studenti herectví reprezentují běžnou populaci.



Obrázek 1: Zobrazení výpočtu činitele zesílení (GF). A) Třetiooktávová spektra vybraných vokálů, které splňují podmínku základní frekvence v rozmezí púltónů od sledované výšky hlasu. B) Tato spektra jsou seřazena dle vypočítané celkové hladiny dB (SPL). C) GF byl vypočten jako směrnice přímky (vznačena přerušovanou čarou) tvořené hladinami sledovaného spektrálního pásma 1/3-oktávových spekter (zobrazeno na ose y) vzhledem k celkové hladině akustického tlaku dB (SPL) vokálů (zobrazeno na ose x). Na D) jsou zobrazeny výsledné hodnoty GF (bezrozměrné číslo) v 1/3-oktávových pásmech, jejichž hladina statistické významnosti korelačního koeficientu lineární regrese byla $p < 0,05$

2. Materiál a metody

Pro účely studia zvyšování hlasitosti byly pořízeny nahrávky 20 hlasových profesionálů (herců a zpěváků) a 20 hlasových neprofesionálů (studentů prvního ročníku herectví na DAMU Praha). Počty subjektů v jednotlivých skupinách jsou popsány v tabulce 1.

Všichni (N = 40)	Profesionálové (N = 20)		Neprofesionálové (N = 20)
	herci	zpěváci	
			studenti 1. ročníku herectví
muži 21	7	3	11
ženy 19	4	6	9

Tabulka 1: Rozložení subjektů při nahrávkách hlasu

2.1. Nahrávky

Se všemi subjekty byla uskutečněna nahrávka tří hlasových úkonů dle metodiky [7]: 1) **habituální (neopřemý) mluvní hlas**; 2) **opřemý (jevištní) mluvní hlas** (v případě studentů šlo o jejich představu opřemého hlasu); 3) **zpěvní hlas**. Pro všechny hlasové úkony bylo nahráno postupné zvyšování hlasitosti slova „máma“ v předem určené výšce hlasu, postupně pro všechny výšky (s celotónovým rozestupem) z celkového tónového rozsahu hlasu daného hlasového úkonu u sledovaného subjektu v jeho modálním hlasovém rejstříku. Rejstřík a hlasový úkon byly kontrolovány percepčně při nahrávce.

Nahrávací sestava Nahrávky byly provedeny ve Zvukovém studiu DAMU pomocí dvou kondenzátorových mi-

krofonů typu Sennheiser K6P s vložkou ME 62 s kulovou směrovou charakteristikou. Mikrofony byly umístěny ve vzdálenosti 30 a 100 cm v přímém směru od úst subjektu, před bližším mikrofonem byl umístěn „pop-filtr“ pro eliminaci turbulencí vzduchu. Nahrávky byly pořízeny v bezeztrátovém formátu wav, s 24bitovou hloubkou a vzorkovací frekvencí 48 kHz.

Separace vokálů Ze všech nahrávek byly interaktivně při poslechu separovány vokály /a:/ pomocí vlastního software vyvinutého v prostředí MATLAB. Na podkladě širokopásmového spektrogramu byly určeny hranice vokálů a konkrétní výřez vokálu o délce 100–300 ms byl prováděn v jeho střední části. Současně byl vokál kontrolován poslechem. Z důvodu odstranění lupnutí při poslechu byl výsledný pseudostacionární segment vokálu upraven konstantním lineárním náběhem 20 ms na začátku a konstantním lineárním ztišením 30 ms na konci výřezu.

2.2. Výpočet GF

Pro všechny hlasové úkony separátně byly postupně vypočteny činitele zesílení (GF) pro všechny výšky hlasového rozsahu subjektů v daném úkonu s celotónovým rozestupem. Výpočet byl uskutečněn následovně (viz obrázek 1): A) ze skupiny všech vokálů separovaných pro konkrétní hlasový úkon, byly postupně vybrány vokály, jejichž průměrná výška byla v rozmezí plus/minus jednoho půltónu od sledované výšky. B) Z akustických třetino-oktávových spekter vybraných vokálů byly vypočteny lineární regrese hladin jednotlivých pásem spekter vůči celkové hladině akustického tlaku (SPL) vokálů. Hodnoty GF odpovídají směrnici vypočtených přímkou lineárních regresí.

U konkrétní sledované výšky byly vypočteny GF jenom pro spektrální pásma se střední frekvencí vyšší než daná základní frekvence. Z výpočtu tak byla eliminována spektrální pásma pod základními frekvencemi sledovaných vokálů. Do dalšího zpracování byly zahrnuty jenom ty hodnoty GF, u kterých lineární regrese dosáhla hladiny statistické významnosti $p < 0,05$.

Následně byly hodnoty GF jednotlivých spektrálních pásem zprůměrovány pro všechny hodnoty sledovaných výšek hlasu konkrétního subjektu a konkrétního hlasového úkonu.

2.3. Statistické zpracování

Pro všechny typy hlasových úkonů byly porovnány vypočtené hodnoty GF ve všech pásmech 1/3-oktávových spekter. Pomocí Studentova t -testu byly porovnány tyto hodnoty pro jednotlivé skupiny subjektů, zvlášť pro muže a zvlášť pro ženy.

3. Výsledky

Hodnoty činitelů zesílení (GF) spektrálních oblastí se pohybovaly pro všechny úkony a všechny subjekty v rozmezí

hodnot 0 až 2,5. V pásmech pod 500 Hz v zásadě GF dosahovalo hodnoty v rozmezí 0 až 1. Nejvyšší hodnoty činitelů zesílení dosahovaly hodnot 1,5 až 2,5 (v závislosti na skupině a hlasovém úkonu). Nejvyšší hodnoty GF byly zásadně v oblastech 2,0–3,5 kHz, což ve spektrální oblasti odpovídá oblastem pěveckého, respektive řečnického formantu. Druhé nejvyšší lokální maximum GF bylo v oblasti cca 1200 Hz, což odpovídá pásmu druhého formantu českého vokálu /a:/.

Porovnání GF pro konkrétní skupinu subjektů mezi jednotlivými hlasovými úkony je uvedeno v tabulce 2.

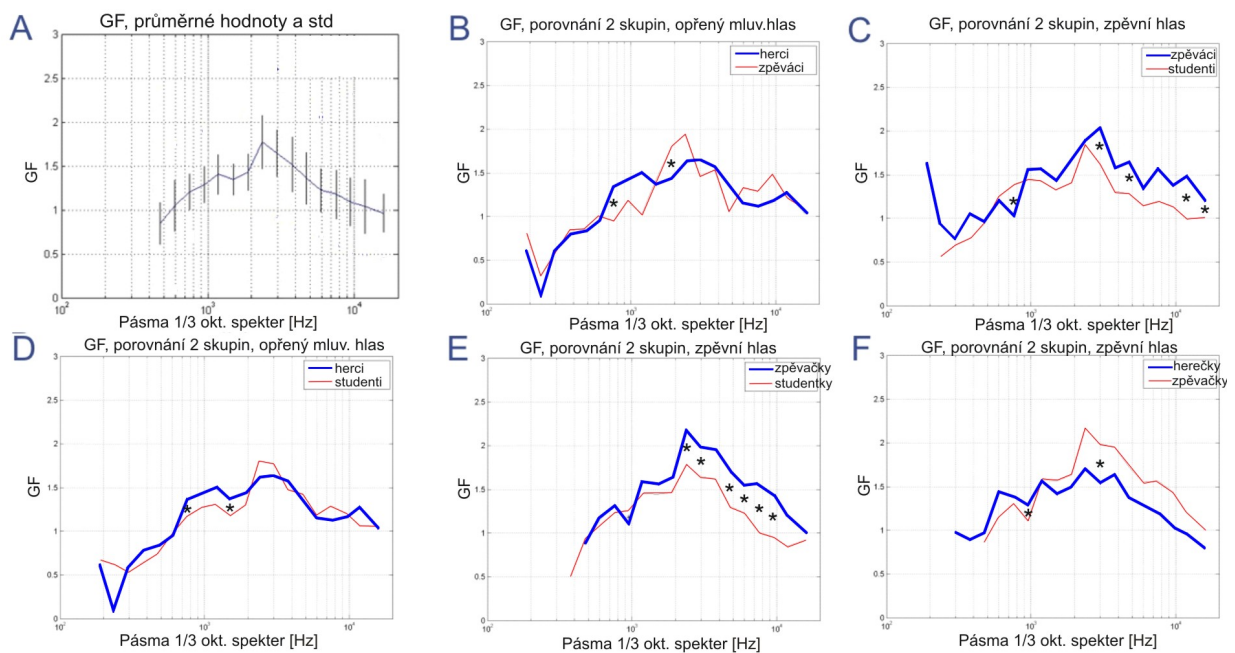
Ve všeobecnosti nebyly nalezeny rozdíly mezi GF habituálního a opřeného mluvního hlasu ve všech skupinách, kromě profesionálních herců mužů a profesionálních hereček, kde v obou případech byly hodnoty GF vyšší pro opřenou řeč v pásmech 190 Hz respektive 4750 Hz.

Největší rozdíly v hodnotách GF byly pozorovány mezi mluvním hlasem a zpěvem. Porovnání výsledků mezi GF pro habituální řeč a pro zpěv ukazuje různé rozdíly zesilování jednotlivých pásem třetino-oktávového spektra pro jednotlivé skupiny: muži-zpěváci nevykazují žádný rozdíl v GF, muži-herci zesilují při zpěvu více oblast prvního, druhého a pěveckého formantu (0,6; 1,9 a 2,9 kHz); muži-studenti více zesilují při zpěvu oblast prvního a druhého formantu v porovnání s habituální řečí. U žen bylo pozorováno při zpěvu vzhledem k habituální řeči větší zesilování: prvního formantu u zpěvaček; druhého formantu a vysoké oblasti 9,5 kHz u hereček; a větší zesilování prvního i druhého formantu u studentek.

Obecně větší hodnoty GF byly pozorovány při zpěvu vzhledem k opřené mluvě u mužů: u zpěváků v oblastech prvního i druhého formantu a oblasti 4,7 kHz; u herců v oblasti prvního a pěveckého formantu; u studentů pro celou hlubší oblast spektra (0,4–1,2 kHz). U žen byly podobně pozorovány vyšší hodnoty GF pro zpěv vůči opřené řeči: u zpěvaček v oblasti prvního a pěveckého formantu (0,6–0,8 kHz a 1,9–3 kHz), u hereček v oblasti druhého formantu a druhého antiformantu (1,2–1,9 kHz); a u studentek v téměř celé spektrální oblasti (0,5–3,0 kHz) kromě pásma druhého antiformantu (1,5 kHz).

Rozdíly v GF mezi jednotlivými skupinami pro konkrétní hlasové úkony jsou uvedeny v tabulce 3. Nejvýznamnější rozdíly v hodnotách GF mezi skupinami subjektů pro konkrétní použití hlasu jsou znázorněny na obrázku 2, pásma se statisticky významnými rozdíly GF jsou označena hvězdičkou.

Pro habituální hlas je u mužů typické, že zpěváci více zesilují oblast druhého antiformantu jak vzhledem k hercům, tak vzhledem ke studentům. Podobně se tímto způsobem odlišují zpěváci od herců (viz obrázek 2B) a studentů (viz obrázek 2D) při opřené mluvě. Zpěváci se dále při opřené řeči významně odlišují i v oblasti prvního formantu, který naopak proti hercům i studentům zesilují méně. U zpěvu mužů nebyly nalezeny statisticky významné odlišnosti mezi herci a zpěváky, obě tyto skupiny se ale odlišovaly od studentů: studenti dosahovali vyšších hodnot GF v oblasti 1. formantu a nižších v oblasti pě-



Obrázek 2: A) Průměrné hodnoty GF vypočtené pro všechny skupiny a všechny úkony (plná čára), standardní odchylky naměřených GF (vertikální čáry). B–F) Výsledky statistického porovnání GF v pásmech třetino-oktávových spekter mezi různými skupinami subjektů při různých hlasových úkonech. Hvězdičkami jsou vyznačena pásma statisticky signifikantních rozdílů hodnot GF ($p < 0,05$) dle t -testu. První skupina je vyznačena modře a tlustě, druhá červeně a tenče. Porovnání průměrných hodnot GF: B) mezi muži herci a zpěváky při opřené hlasu; C) mezi muži zpěváky a studenty při zpěvu; D) mezi muži herci a studenty při opřené mluvním hlasu; E) mezi ženami zpěvačkami a studentkami při zpěvu; F) mezi ženami herečkami a zpěvačkami při zpěvu

veckého formantu vzhledem k zpěvákům (viz obrázek 2C), vůči hercům se při zpěvu odlišovali nižšími hodnotami GF v oblasti jak druhého, tak pěveckého formantu.

U žen, podobně jako u mužů, více zesilují oblast pěveckého formantu zpěvačky při habituálním hlasu jak vůči herečkám, tak i vůči studentkám. U opřené mluvy byly nalezeny statisticky významné rozdíly jenom mezi zpěvačkami a studentkami v oblastech pěveckého formantu a vyšších spektrálních oblastech (7,5–10 kHz), kde zpěvačky dosahovaly vyšších hodnot GF. U zpěvu se vyčleňovaly zpěvačky vůči herečkám (viz obrázek 2F) i vůči studentkám (viz obrázek 2E) vyššími hodnotami GF v oblasti pěveckého formantu a nižšími hodnotami v oblastech 2. formantu. Porovnání GF hodnot hereček a studentek u zpěvu ukazuje zvýšení hodnot GF v oblasti prvního formantu a oblasti 7,5 kHz u hereček.

4. Diskuze

V článku byly prezentovány výsledky měření činitele zesílení (gain factor) pro gradaci hlasitosti vokálu „a“. Pomocí statistického testu byly porovnány hodnoty GF globálně pro tři způsoby užití hlasu (habituální, opřené mluvním hlas a zpěvním hlas) a tři skupiny hlasových uživatelů: profesionálních zpěváků a herců a začínajících studentů herectví na DAMU Praha. Pomocí uvedené kategorizace byl proveden

odhad a charakterizace různých strategií použití hlasu uvedených skupin subjektů.

Podobně jako v [5] byly změřeny nejvyšší hodnoty GF pro všechny skupiny uživatelů a všechny hlasové úkony v oblastech 1,5–3 kHz (viz obrázek 1D a 2A). Průměrné hodnoty GF pro celý tónový rozsah hlasu mají podobný tvar třetino-oktávových spekter (viz obrázek 2B–F), tedy jejich globální maxima odpovídají formantům vokálu „a“ (zejména prvnímu v oblasti 0,6–0,8 kHz, druhému v oblasti 0,95–1,2 kHz a pěveckému formantu v oblasti 2,4–3,0 kHz) a minima odpovídají antiformantům (zejména v oblasti 1,5–2 kHz antiformantu mezi 2. a pěveckým formantem). Z uvedených údajů je možno odhadnout strategii zesilování jednotlivých spektrálních oblastí v závislosti na použití hlasu a konkrétní skupině uživatelů hlasu.

Z akustického hlediska byl popsán rozdíl mezi mluvním a klasickým pěveckým hlasem (operní zpěv) v spektrální oblasti tzv. „pěveckého formantu“ (singer’s formant) [8]. Dle uvedeného vzniká pěvecký formant jako spojení formantů F_3 – F_5 s výrazným posílením akustické energie v rozmezí 2,3–3,8 kHz. Pěvecký formant umožňuje, aby byl zpěvák odlišen od orchestrálního doprovodu, a je spojován se zvonivou (ringing) barvou hlasu.

Podobně i uvedená měření ukazují největší rozdíly mezi zpěváky a nezpěváky (herci a studenty) v naměřených hodnotách GF v oblasti pěveckého formantu u mužů

zejména při zpěvu a u žen i při habituálním hlase, proto je možno předpokládat, že tato oblast je zodpovědná za odlišení pěveckého hlasu od nepěveckého.

Profesionální mluvní technika je podobně spojována s řečnickým formantem (speaker's, actor's formant) vznikajícím přiblížením 3. a 4. formantu [9]. Vyznačuje se posunutou centrální frekvencí formantu k vyšším frekvencím (3 kHz) a nižšími akustickými hladinami v porovnání s pěveckým formantem. Další charakteristikou profesionálního mluvního hlasu je další rezonanční vrchol (formant) v okolí 8 kHz [10]. V této studii byly také pozorovány rozdíly v GF ve vysokých spektrálních oblastech (cca 7,5 kHz) zejména pro zpěvačky v porovnání s ostatními skupinami žen pro všechny typy použití hlasu, proto je možno předpokládat, že toto pásmo je charakteristické hlavně pro zpěvní ženský hlas.

Uvedené výsledky, vzhledem k malým počtům jednotlivých subjektů v jednotlivých skupinách hlasových profesionálů, není možno zobecňovat. Předběžné výsledky ale naznačují, že 1/3-oktávová spektra na základě hodnocení činitelů zesílení je možno rozdělit do pěti zásadních oblastí:

- První významnou spektrální oblastí se ukazuje **oblast 1. formantu** (500–800 Hz), která je obecně při zpěvu víc zesilována než v mluvním projevu. Také je více zesilována při opřené řeči u profesionálních herců vůči zpěvákům i studentům a v habituální řeči u hereček vůči studentkám.
- Druhou významnou oblastí je frekvenční **oblast 2. formantu** (0,8–1,2 kHz). V této oblasti je možno odlišit pravděpodobně neškolenou strategii zpěvu, protože její vyšší hodnoty GF jsou typické pro zpěv studentů na rozdíl od zpěváků a profesionálních herců.
- Třetí důležité pásmo je **oblast mezi 2. formantem a pěveckým formantem** (1,5–2 kHz) – tj. oblast 2. antiformantu, která je pravděpodobně typická větším zesilováním u profesionálního zpěvu. Zesilují ji více zpěváci při všech hlasových úkonech, ale i ostatní skupiny hlavně při zpěvu v porovnání k mluvním úkonům.
- Další specifickou oblastí je **oblast pěveckého nebo řečnického formantu** (2,1–3,2 kHz). Je více zesilována při zpěvu než při opřené mluvě nebo habituálním hlase.
- Interpretace vysoké **spektrální oblasti nad 3,7 kHz** je diskutabilní, protože odpovídá jak za odlišení habituální mluvy hereček od jejich opřené řeči a zpěvu, tak za odlišení profesionálních zpěváků (mužů i žen) od studentů, kdy profesionálové dosahují vyšší GF v této oblasti při zpěvu.

Výsledky této studie jsou jenom orientační, protože uvedené skupiny subjektů byly rozděleny jenom na základě jejich profese, a tím se do jednotlivých skupin dostalo málo

subjektů. Pro zvýšení relevance výsledků, by bylo vhodné rozdělení subjektů na skupiny lepších a horších řečníků nebo zpěváků na základě poslechových testů.

Vzhledem k uvedenému postupu statistického zpracování - porovnání hodnot GF pro celkový tónový rozsah, nebylo možné sledovat hodnoty GF separátně pro pásma jednotlivých nízkých harmonických složek. Pro možnost porovnání GF faktorů typických pro první a druhou harmonickou je nutno separátně porovnávat GF faktory pro konkrétní výšku hlasu, ale pro statistické ohodnocení by bylo nutno porovnávat větší počet subjektů.

Popisované výsledky charakterizují způsob zvyšování hlasitosti u jednotlivých skupin subjektů jenom z akustického – spektrálního hlediska. Pro jejich interpretaci z hlediska fyziologie a funkce hlasového systému (tedy charakterizaci použití dechu, funkcí kmitání hlasivek nebo nastavení vokálního traktu) je nutno akustická měření spojit s jinými metodami sledování hlasových funkcí (např. elektrolottografií).

Specifickou akustickou interpretaci akustických výsledků bude možno provést při doplnění uvedených dat o jiná akustická měření (např. poměry energie vyšších spektrálních oblastí vzhledem k nižším oblastem – definovaných jako parametry – **vyváženosti spektra** (spectral balance) [11], nebo separátní pozorování poměru hladin prvních dvou harmonických, či sledování jiných parametrů hlasových signálů). Této problematice se budou věnovat navazující studie.

5. Závěr

Činitel zesílení (gain factor) je parametr, který popisuje míru zesílení jednotlivých spektrálních oblastí při změně celkové hlasitosti hlasu. Naměřené průměrné hodnoty ukazují, že lokální maxima GF odpovídají druhému „vokalic-kému“ formantu a vyššímu formantu - pravděpodobně odpovídajícímu pěveckému nebo řečnickému formantu. Statistické porovnání hodnot GF mezi různými hlasovými úkony jednotlivých skupin subjektů naznačuje na různé strategie zesilování hlasu. Porovnání jednotlivých skupin subjektů (s rozdílným způsobem hlasového tréninku) mezi sebou pro konkrétní hlasový úkon naznačuje, že dle měření GF je možno 1/3-oktávové spektrum vokálu „a“ rozdělit do **5 zásadních oblastí** – pravděpodobně charakteristických pro konkrétní použití hlasu.

Poděkování

Při vytváření podkladů práce spolupracovaly kolegyně Mgr. Klára Kadlecová a MgA. Regina Szymiková. Práce vznikla za podpory projektu Ministerstva školství, mládeže a tělovýchovy, ČR Pr. No. 1M0531.

muži	typ hlasu 1	typ hlasu 2	118	150	190	237	297	377	475	595	755	950	1185	1500	1900	2370	2975	3775	4750	5950	7550	9500	11850	16000			
zpěváci	habit. řeč	opřen.řeč																									
herci					+																						
studenti																											
zpěváci	habit. řeč	zpěv																									
herci										+						+		+									
studenti												+	+														
zpěváci	opřen.řeč	zpěv																									
herci																	+	+									
studenti									+	+	+	+	+	+							+						
ženy							F1		F2		AF2		FS														
zpěvačky	habit. řeč	opřen.řeč																									
herečky																											
studentky																											
zpěvačky	habit. řeč	zpěv																									
herečky																											
studentky										+	+	+		+													
zpěvačky	opřen.řeč	zpěv																									
herečky																											
studentky										+	+	+	+	+	+	+	+	+	+								

Tabulka 2: Schematické znázornění statisticky významných rozdílů mezi jednotlivými typy použití hlasu u jednotlivých skupin subjektů. Znaménko plus označuje statisticky významnou větší hodnotu GF pro typ hlasu 2 vůči typu hlasu 1

muži	skupina 1	skupina 2	118	150	190	237	297	377	475	595	755	950	1185	1500	1900	2370	2975	3775	4750	5950	7550	9500	11850	16000	
habit. řeč	herci	zpěváci													+	+									
	herci	studenti																							
	zpěváci	studenti				-																			
opřená řeč	herci	zpěváci														+									
	herci	studenti																							
	zpěváci	studenti									+						+								
zpěv	herci	zpěváci																							
	herci	studenti																							
	zpěváci	studenti									+														
ženy							F1		F2		AF2		FS												
habit. řeč	herečky	zpěvačky																							
	herečky	studentky																							
	zpěvačky	studentky																							
opřená řeč	herečky	zpěvačky																							
	herečky	studentky																							
	zpěvačky	studentky																							
zpěv	herečky	zpěvačky																							
	herečky	studentky																							
	zpěvačky	studentky																							

Tabulka 3: Schematické znázornění statisticky významných rozdílů mezi jednotlivými skupinami subjektů při použití různých typů hlasů. Znaménko plus (podbarveno šedě) označuje statisticky významnou větší hodnotu GF, znaménko minus označuje statisticky významnou menší hodnotu GF pro skupinu uvedenou v sloupci skupina 2 vůči skupině uvedené v sloupci skupina 1

Reference

- [1] Sjolander P., Sundberg J.: Spectrum effects of subglottal pressure variation in professional baritone singers. *J. Acoust. Soc. Am.* 2004; 115:1270–1273.
- [2] Henrich N., d'Alessandro Ch., Doval B., Castellengo M.: Glottal open quotient in singing: Measurements and correlation with laryngeal mechanisms, vocal intensity, and fundamental frequency. *J. Acoust. Soc. Am.* 2005; 117:1417–1430.
- [3] Holmberg E. B., Hillman R. E., Perkell J. S.: Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice, *J. Acoust. Soc. Am.* 1988; 84:511–529.
- [4] Sundberg J., Fahlstedt E., Morell A.: Effects on the glottal voice source of vocal loudness variation in untrained female and male voices, *J. Acoust. Soc. Am.* 2005; 117:879–885.
- [5] Nordenberg M., Sundberg J.: Effect on LTAS of vocal loudness variation. *Logopedics Phoniatrics Vocology* 2004; 29:183–191.
- [6] Ternström S.: Long-time average spectrum characteristics of different choirs in different rooms. *Speech, Music and Hearing – Quarterly Progress and Status Report* 1989; 30:15–031.
- [7] Frič M., Kadlecová K., Szymiková R.: Záznam mluvního a zpěvního hlasu pro psychoakustické analýzy; Praha, *MARC-Technologický list číslo 13*, Zvukové studio HAMU, 2007.
- [8] Sundberg J.: Articulatory interpretation of the “singing formant”. *J. Acoust. Soc. Am.* 1974; 55:838–844.
- [9] Bele I. V.: The speaker’s formant. *J. Voice* 2006; 20:555–578.
- [10] Lee S. H., Kwon H. J., Choi H. J., Lee N. H., Lee S. J., Jin S. M.: The Singer’s Formant and Speaker’s Ring Resonance: A Long-Term Average Spectrum Analysis. *Clin Exp Otorhinolaryngol* 2008; 1:92–96.
- [11] Collyer S., Davis P. J., Thorpe C. W., Callaghan J.: Fundamental frequency influences the relationship between sound pressure level and spectral balance in female classically trained singers. *J. Acoust. Soc. Am.* 2009; 126:396–406.

