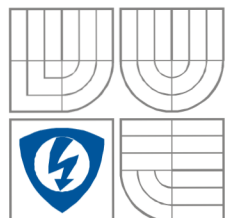




VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



**FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH
TECHNOLOGIÍ**
ÚSTAV RADIOELEKTRONIKY

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF RADIO ELECTRONICS

DIAGNOSTICKÁ ANALÝZA HLASU

DIAGNOSTICAL ANALYSIS OF VOICE

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

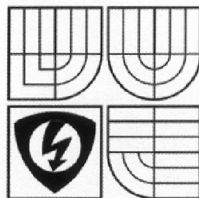
AUTOR PRÁCE
AUTHOR

Bc. Pavel Sala

VEDOUCÍ PRÁCE
SUPERVISOR

prof. Ing. Milan Sigmund, CSc.

BRNO, 2008



VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ
Fakulta elektrotechniky
a komunikačních technologií
Ústav radioelektroniky

Diplomová práce

magisterský navazující studijní obor
Elektronika a sdělovací technika

Student: Sala Pavel, Bc.
Ročník: 2

ID: 89605
Akademický rok: 2007/08

NÁZEV TÉMATU:

Diagnostická analýza hlasu

POKYNY PRO VYPRACOVÁNÍ:

Seznamte se s metodami analýzy řečového signálu pro diagnostické účely. Vypracujte přehledovou studii vhodných informačních zdrojů, dostupných databází patologických hlasů a seznam základních diagnóz, které lze stanovit technickou analýzou hlasu. Naprogramujte a vyzkoušejte algoritmy nejčastěji používané na určování hlasivkových pulsů. Pomocí analýzy hlasivkových pulsů vytvořte a vyhodnoťte vhodná kritéria na určení vybraných diagnóz v závislosti na dostupných databázích (základní řečová data dostanete od vedoucího projektu).

DOPORUČENÁ LITERATURA:

- [1] PSUTKA, J. Komunikace s počítačem mluvenou řečí. Praha: Academia, 1995.
- [2] SIGMUND, M. Analýza řečových signálů. Skriptum. Brno: FEKT VUT v Brně, 2000.

Termín zadání: 5.10.2007

Termín odevzdání: 30.5.2008

Vedoucí projektu: doc. Ing. Milan Sigmund, CSc.

prof. Dr. Ing. Zbyněk Raida
předseda oborové rady



UPOZORNĚNÍ:

Autor diplomové práce nesmí při vytváření diplomové práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

LICENČNÍ SMLOUVA POSKYTOVANÁ K VÝKONU PRÁVA UŽÍT ŠKOLNÍ DÍLO

uzavřená mezi smluvními stranami:

1. Pan/paní

Jméno a příjmení: Bc. Pavel Sala
Bytem: Kobylá nad Vidnavkou 111, Žulová 1, 790 65
Narozen/a (datum a místo): 23. února 1984 v Jeseníku

(dále jen „autor“)

a

2. Vysoké učení technické v Brně

Fakulta elektrotechniky a komunikačních technologií
se sídlem Údolní 53, Brno, 602 00
jejímž jménem jedná na základě písemného pověření děkanem fakulty:
prof. Dr. Ing. Zbyněk Raida, předseda rady oboru Elektronika a sdělovací technika
(dále jen „nabyvatel“)

Čl. 1

Specifikace školního díla

1. Předmětem této smlouvy je vysokoškolská kvalifikační práce (VŠKP):

- ▣ disertační práce
 - ▣ diplomová práce
 - ▣ bakalářská práce
 - ▣ jiná práce, jejíž druh je specifikován jako
- (dále jen VŠKP nebo dílo)

Název VŠKP: Diagnostická analýza řečového signálu

Vedoucí/ školitel VŠKP: prof. Ing. Milan Sigmund, CSc.

Ústav: Ústav radioelektroniky

Datum obhajoby VŠKP: _____

VŠKP odevzdal autor nabyvateli*:

v tištěné formě – počet exemplářů: 2

v elektronické formě – počet exemplářů: 2

2. Autor prohlašuje, že vytvořil samostatnou vlastní tvůrčí činností dílo shora popsané a specifikované. Autor dále prohlašuje, že při zpracovávání díla se sám nedostal do rozporu s autorským zákonem a předpisy souvisejícími a že je dílo dílem původním.
3. Dílo je chráněno jako dílo dle autorského zákona v platném znění.
4. Autor potvrzuje, že listinná a elektronická verze díla je identická.

* hodící se zaškrtněte

Článek 2

Udělení licenčního oprávnění

1. Autor touto smlouvou poskytuje nabyvateli oprávnění (licenci) k výkonu práva uvedené dílo nevýdělečně užit, archivovat a zpřístupnit ke studijním, výukovým a výzkumným účelům včetně pořizování výpisů, opisů a rozmnoženin.
2. Licence je poskytována celosvětově, pro celou dobu trvání autorských a majetkových práv k dílu.
3. Autor souhlasí se zveřejněním díla v databázi přístupné v mezinárodní síti
ihned po uzavření této smlouvy
 - 1 rok po uzavření této smlouvy
 - 3 roky po uzavření této smlouvy
 - 5 let po uzavření této smlouvy
 - 10 let po uzavření této smlouvy
(z důvodu utajení v něm obsažených informací)
4. Nevýdělečné zveřejňování díla nabyvatelem v souladu s ustanovením § 47b zákona č. 111/ 1998 Sb., v platném znění, nevyžaduje licenci a nabyvatel je k němu povinen a oprávněn ze zákona.

Článek 3

Závěrečná ustanovení

1. Smlouva je sepsána ve třech vyhotoveních s platností originálu, přičemž po jednom vyhotovení obdrží autor a nabyvatel, další vyhotovení je vloženo do VŠKP.
2. Vztahy mezi smluvními stranami vzniklé a neupravené touto smlouvou se řídí autorským zákonem, občanským zákoníkem, vysokoškolským zákonem, zákonem o archivnictví, v platném znění a popř. dalšími právními předpisy.
3. Licenční smlouva byla uzavřena na základě svobodné a pravé vůle smluvních stran, s plným porozuměním jejímu textu i důsledkům, nikoliv v tísní a za nápadně nevýhodných podmínek.
4. Licenční smlouva nabývá platnosti a účinnosti dnem jejího podpisu oběma smluvními stranami.

V Brně dne: 30. května 2008

.....
Nabyvatel

.....
Autor

Abstrakt

Cílem práce bylo vytvořit přehledovou studii informačních zdrojů zabývajících se diagnostickou analýzou hlasu. Dalším krokem bylo naprogramování vybraných metod pro odhad hlasivkových pulzů. Posledním krokem bylo vyhodnocení vhodných kritérií pro určení vybraných patologických hlasů a vlivu stresu na průběh hlasivkových pulzů. Byly naprogramovány dva algoritmy pro výpočet hlasivkových pulzů. V závěru práce bylo navrženo několik kritérií pro určení stresového stavu na základě odhadu hlasivkových pulzů.

Klíčová slova: hlasivky, inverzní filtrace, hlasivkové pulzy, patologické hlasy, stres

Abstract

Goal of this work was create survey study of information resources deal with diagnostic analysis of speech signal. Two methods for estimation of glottal flow was programmed. Finally, attention was focused on determination of criterions for description of selected pathological diagnosis and influence of stress on the glottal flow. Outcome of this work is proposal two criterions for describe influence of stress on the glottal flow.

Key words: glottis, inverse filtering, glottal source, pathological voices, stress

Bibliografická citace:

SALA, P. *Diagnostická analýza hlasu*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2008. 56 s. Vedoucí diplomové práce prof. Ing. Milan Sigmund, CSc.

Prohlášení

Prohlašuji, že svou diplomovou práci na téma Diagnostická analýza hlasu jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené diplomové práce dále prohlašuji, že v souvislosti s vytvořením této diplomové práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

V Brně dne 30. května 2008

.....
podpis autora

Poděkování

Děkuji vedoucímu diplomové práce Prof. Ing. Milanu Sigmundovi, CSc. za účinnou metodickou, pedagogickou a odbornou pomoc a další cenné rady při zpracování mé diplomové práce. Děkuji také své rodině a přítelkyni, bez jejichž pochopení a podpory bych tuto práci nikdy nedokončil. Nakonec děkuji všem, kteří se jakýmkoliv způsobem přičinili na tvorbě této práce (buť i nepřímo) a které jsem zde výslovně nezmínil.

V Brně dne 30. května 2008

.....
podpis autora

Obsah

OBSAH	1
SEZNAM TABULEK	3
SEZNAM OBRÁZKŮ	4
1. ÚVOD	5
1.1 ZADÁNÍ	5
1.2 CÍL PRÁCE	5
2. HLASOVÉ ÚSTROJÍ	5
2.1 MODEL VYTVÁŘENÍ ŘEČI	6
2.1.1 Popis funkce modelu pro vytváření řeči	7
3. ANALÝZA ŘEČOVÉHO SIGNÁLU	7
3.1 ÚVOD	7
3.2 DRUHY ZOBRAZENÍ ŘEČOVÉHO SIGNÁLU	7
3.2.1 Oscilogram	7
3.2.2 Fundamentální (základní) frekvence	8
3.2.3 Spektrum	9
3.2.4 Spektrogram	9
3.2.5 Kaskádový spektrogram (Waterfall spectrogram)	9
3.3 POPIS ŘEČI	10
3.4 ZPŮSOB ZÍSKÁNÍ DIGITALIZOVANÉHO ŘEČOVÉHO SIGNÁLU	10
4. ZDROJE INFORMACÍ	11
4.1 ÚVOD	11
4.2 ODBORNÉ ČLÁNKY	11
4.2.1 Články týkající se rakoviny hlasivek	12
4.2.2 Články týkající se Parkinsonovy choroby	13
4.2.3 Články týkající se analýzy řeči a hlasu obecně	14
4.3 KNIHY ZABÝVAJÍCÍ SE ANALÝZOU HLASU	14
4.4 DATABÁZE PATOLOGICKÝCH HLASŮ	15
4.5 ORGANIZACE ZABÝVAJÍCÍ SE VÝZKUMEM V OBLASTI HLASU	15
4.6 ARCHIVY ODBORNÝCH ČLÁNKŮ	17
5. DIAGNÓZY STANOVOVANÉ POMOCÍ ANALÝZY HLASU	18
5.1 ÚVOD	18
5.2 ZPŮSOB STANOVENÍ DIAGNÓZY	18
5.3 HLASIVKOVÉ PULZY	18
5.3.1 Typický řečový signál (modal voice)	19
5.3.2 Pískavý (falsetto)	19
5.3.3 Skřípavý (creak)	20
5.3.4 Šepot (whisper)	20
5.3.5 Chraptivý (harshness)	20
5.3.6 Dýchavičný (breathiness)	20
6. ODHAD TVARU HLASIVKOVÉHO PULZU Z ŘEČOVÉHO SIGNÁLU	21
6.1 METODA ZALOŽENÁ NA OPAKOVANÉM INVERZNÍM FILTROVÁNÍ ŘEČOVÉHO SIGNÁLU - ITERATIVE ADAPTIVE INVERSE FILTERING (IAIF) [4]	21
6.1.1 Princip metody	21
6.1.2 Blokové schéma	22
6.1.2.1 Horní propust	22
6.1.2.2 První opakování	23
6.1.2.3 Druhé opakování	24
6.2 METODA ZALOŽENÁ NA PRŮMĚROVÁNÍ CHARAKTERISTIK HLASOVÉHO TRAKTU [9]	25

6.2.1 Hlasivkové pulzy a hlasový trakt	25
6.2.2 Odhad fází hlasivkového pulzu	26
6.2.3 Odhad koeficientů filtru hlasového traktu	28
6.2.4 Určení částí řečového signálu odpovídajícího fázi uzavření hlasivek	29
6.2.5 Odhad hlasivkového pulzu.....	31
6.3 VÝSTUPY VYBRANÝCH METOD PRO STANOVENÍ HLASIVKOVÝCH PULZŮ	31
6.3.1 Výstup metody IAIF [4]	31
6.3.2 Výstup metody založené na průměrování parametrů hlasového traktu [9].....	31
7. POMŮCKA PRO ODHAD HLASIVKOVÝCH PULZŮ Z ŘEČOVÉHO SIGNÁLU.....	35
7.1 NAČTENÍ ZVUKOVÝCH SOUBORŮ A NASTAVENÍ METOD	35
7.1.1 Nastavení IAIF metody	35
7.1.2 Nastavení MATICE metody	36
7.2 ULOŽENÍ A POROVNÁNÍ HLASIVKOVÝCH PULZŮ.....	36
8. STANOVENÍ VHODNÝCH KRITÉRIÍ PRO URČENÍ VYBRANÝCH DIAGNÓZ	37
8.1 ONEMOCNĚNÍ HLASIVEK	37
8.1.1 Reinek's edema (Reinekův edém)	38
8.1.2 Vocal fold edema (Otok hlasivkových svalů)	39
8.1.3 Vocal tremor (Třes hlasu)	40
8.2 VLIV STRESU	42
8.2.1 Monotónnost pulzů.....	43
8.2.2 Porovnání velikosti ploch.....	45
9. ZÁVĚR	47
PŘÍLOHA A: POPIŠ ALGORITMU DAP[5],[7]	49
PŘÍLOHA B: URČENÍ OKAMŽÍKŮ GCI Z ŘEČOVÉHO SIGNÁLU	49
B.1 METODA PRO ODEČTENÍ POZICE GCI Z MULTISCALE PRODUKTU	50
B.2 URČENÍ POZICE GCI Z EGG.....	51
B.3 POROVNÁNÍ POZIC GCI STANOVENÝCH OBĚMA METODAMI.....	52
10. POUŽITÁ LITERATURA	54
11. SEZNAM ZKRATEK A SYMBOLŮ.....	56

Seznam tabulek

TABULKA 3.1: HODNOTY F_0 PRO RŮZNÁ POHLAVÍ	8
TABULKA A: ALGORITMUS DAP.....	49

Seznam obrázků

OBRÁZEK 2.1:	HRTAN A HLASIVKY (POHLED SHORA)	5
OBRÁZEK 2.2:	SCHÉMA KLIDOVÉHO (A) A FONAČNÍHO (B) POSTAVENÍ HLASIVEK.....	6
OBRÁZEK 2.3:	MODEL VYTVÁŘENÍ ŘEČI	6
OBRÁZEK 3.1:	OSCILOGRAM SLOVA „PHONETICIAN“	8
OBRÁZEK 3.2:	ZÁKLADNÍ FREKVENCE (FORMANT) DANÉHO ŘEČNÍKA.....	8
OBRÁZEK 3.3:	SPEKTRUM SAMOHLÁSKY „O“	9
OBRÁZEK 3.4:	SPEKTROGRAM SLOVA „PHONETICIAN“	9
OBRÁZEK 3.5:	KASKÁDOVÝ SPEKTROGRAM SLOVA „PHONETICIAN“	10
OBRÁZEK 4.1:	PROSTŘEDÍ PROGRAMU DISORDERED VOICE DATABASE AND PROGRAM	15
OBRÁZEK 5.1:	PŘÍKLAD HLASIVKOVÝCH PULZŮ PRO TYPICKÝ ŘEČOVÝ SIGNÁL.....	19
OBRÁZEK 5.2:	PŘÍKLAD HLASIVKOVÝCH PULZŮ PRO PÍSKAVÝ SIGNÁL	19
OBRÁZEK 5.3:	PŘÍKLAD HLASIVKOVÝCH PULZŮ SKŘÍPAVÉHO SIGNÁLU.....	20
OBRÁZEK 5.4:	PŘÍKLAD HLASIVKOVÝCH PULZŮ ŠEPTU.....	20
OBRÁZEK 5.5.:	PŘÍKLAD HLASIVKOVÝCH PULZŮ DÝCHAVIČNÉHO SIGNÁLU	21
OBRÁZEK 6.1.:	MODEL TVORBY ŘEČI	21
OBRÁZEK 6.2.:	BLOKOVÉ SCHÉMA METODY IAIF.....	22
OBRÁZEK 6.3.:	MODEL TVORBY ŘEČOVÉHO SIGNÁLU	25
OBRÁZEK 6.4.:	VÁLCOVÝ MODEL HLASOVÉHO TRAKTU.....	25
OBRÁZEK 6.5.:	VÝSTUP METODY IAIF.....	31
OBRÁZEK 6.6A.:	URČENÍ GCI A ROZDĚLENÍ NA SEGMENTY	32
OBRÁZEK 6.6B.:	ÚSEKY VYBRANÉ PRO VÝPOČET A.....	32
OBRÁZEK 6.7.:	VYPOČÍTANÝ SIGNÁL DUSC.....	33
OBRÁZEK 6.8.:	DETEKCE FÁZE UZAVŘENÍ Z DUSC.....	33
OBRÁZEK 6.9.:	ODHAD HLASIVKOVÝCH PULZŮ	34
OBRÁZEK 7.1.:	UTILITA PRO ANALÝZU HLASIVKOVÝCH PULZŮ.....	35
OBRÁZEK 8.1.:	REINEK'S EDEMA, SKUPINA A (ČTYŘI MLUVČÍ)	38
OBRÁZEK 8.2.:	REINEK'S EDEMA, SKUPINA B (DVA MLUVČÍ)	38
OBRÁZEK 8.3.:	REINEK'S EDEMA, SKUPINA C (DVA MLUVČÍ)	39
OBRÁZEK 8.4.:	REINEK'S EDEMA, SKUPINA D (DVA MLUVČÍ)	39
OBRÁZEK 8.5.:	VOCAL FOLD EDEMA, SKUPINA A (TŘI MLUVČÍ)	39
OBRÁZEK 8.6.:	VOCAL FOLD EDEMA, SKUPINA B (DVA MLUVČÍ)	40
OBRÁZEK 8.7.:	VOCAL FOLD EDEMA, SKUPINA C (DVA MLUVČÍ)	40
OBRÁZEK 8.8.:	VOCAL FOLD EDEMA, SKUPINA D (DVA MLUVČÍ)	40
OBRÁZEK 8.9.:	VOCAL FOLD EDEMA, SKUPINA E (DVA MLUVČÍ)	40
OBRÁZEK 8.10.:	VOCAL TREMOR, SKUPINA A (TŘI MLUVČÍ)	41
OBRÁZEK 8.11.:	VOCAL TREMOR, SKUPINA B (TŘI MLUVČÍ)	41
OBRÁZEK 8.12.:	VOCAL TREMOR, SKUPINA C (DVA MLUVČÍ)	41
OBRÁZEK 8.13.:	VLIV STRESU NA TVAR HLASIVKOVÝCH PULZŮ 1	42
OBRÁZEK 8.14.:	VLIV STRESU NA TVAR HLASIVKOVÝCH PULZŮ 2	43
OBRÁZEK 8.15.:	VLIV STRESU NA TVAR HLASIVKOVÝCH PULZŮ 3	43
OBRÁZEK 8.16.:	VÝBĚR ČÁSTI PULZU POMOCÍ PRAHU	44
OBRÁZEK 8.17.:	CHYBA URČENÍ STRESU PODLE NÁBĚŽNÉ HRANY	44
OBRÁZEK 8.18.:	CHYBA URČENÍ STRESU PODLE SESTUPNÉ HRANY	45
OBRÁZEK 8.19.:	CHYBA URČENÍ STRESU PODLE NÁBĚŽNÉ A SESTUPNÉ HRANY.....	45
OBRÁZEK 8.20.:	PLOCHÝ HLASIVKOVÝCH PULZŮ	46
OBRÁZEK B.1.:	ÚSEK HLÁSKY „A“ ZPRACOVANÝ VLNKOVOU TRANSFORMACÍ	50
OBRÁZEK B.2.:	ÚSEK HLÁSKY „A“ S VYZNAČENÝMI GCI	51
OBRÁZEK B.3.:	ÚSEK HLÁSKY „A“ S VYZNAČENÝMI GCI PODLE EGG.....	52
OBRÁZEK B.4.:	SROVNÁNÍ OBOU METOD PRO URČENÍ OKAMŽIKŮ GCI	52

1. Úvod

1.1 Zadání

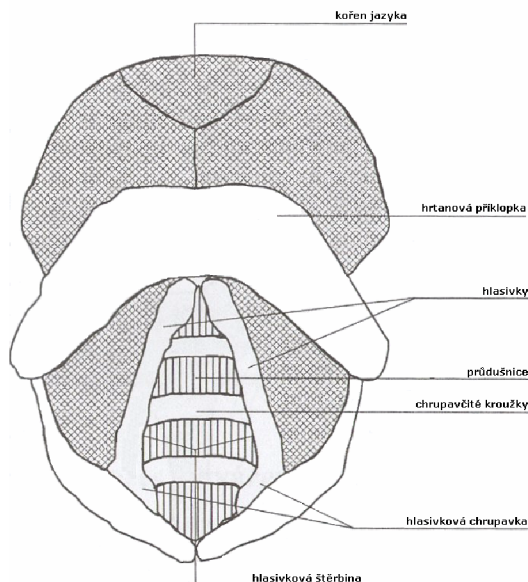
Seznamte se s metodami analýzy řečového signálu pro diagnostické účely. Vypracujte přehledovou studii vhodných informačních zdrojů, dostupných databází patologických hlasů a seznam základních diagnóz, které lze stanovit technickou analýzou hlasu. Naprogramujte a vyzkoušejte algoritmy nejčastěji používané na určování hlasivkových pulzů. Pomocí analýzy hlasivkových pulzů vytvořte a vyhodnoťte vhodná kritéria na určení vybraných diagnóz v závislosti na dostupných databázích (základní řečová data dostanete od vedoucího projektu).

1.2 Cíl práce

Jak říká zadání, cílem práce bude vypracovat seznam dostupných informačních zdrojů, zabývajících se tématem analýzy řečového signálu. V další části práce se budu zabývat naprogramováním konkrétních metod pro určování odhadu průběhu hlasivkových pulzů a v poslední části se budu zabývat nalezením vhodných parametrů ze získaných hlasivkových pulzů tak, aby bylo možné těmito parametry popsat určitý druh diagnózy (např. různá onemocnění hlasivek či psychický stav mluvčího).

2. Hlasové ústrojí

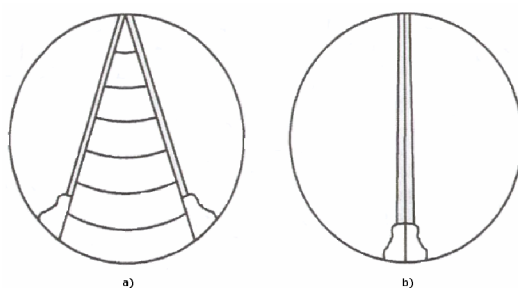
Pojmem hlasové ústrojí (vocal tract) se také často označuje celý systém pro vytváření řeči [3]. My jej však budeme chápat jako označení pouze té části, kde dochází k samotnému vzniku hlasu. Na **Obrázek 2.1** [3] je vidět, z čeho se hlasové ústrojí skládá.



OBRÁZEK 2.1: HRTAN A HLASIVKY (POHLED SHORA)

Hlasové ústrojí je uloženo v hrtanu (larynx), který je s plícemi spojen průdušnicí. Nejdůležitější část tvoří hlasivky (vocal forde), nacházející se v hrtanové dutině přímo za ohryzkem (Adamovo jablko). Jsou to dvě ostré slizniční řasy, které vedou napříč hrtanem

v místě jeho nejužšího průchodu. Z jedné strany jsou napojeny na chrupavky hlasivkové a z druhé strany na chrupavku štítnou. Jsou pokryty sliznicí a jejich základ tvoří hlasový vaz a hlasivkový sval. Prostor mezi hlasivkami tvoří hlasivková štěrbina trojúhelníkového tvaru (glottis). Pokud člověk mlčí, jsou hlasivky neaktivní a štěrbina je odkrytá, takže jí může vzduch z plic bez problémů proudit (Obrázek 2.2a) [3]. Při mluvení (fonaci) jsou hlasivky v takzvaném hlasovém (fonačním) postavení (Obrázek 2.2b) [3]. Proud vydechovaného vzduchu je pak hlasivkami „blokován“. Hlasivky jsou staženy a působením tlaku vzduchu přicházejícího z plic začínají kmitat. Při tom se střídavě postupně otevírají a prudce uzavírají. Tak vzniká tzv. vzduchová vlna, kterou vnímáme jako zvuk. Takto vzniklý periodický proud vzduchových pulsů tvoří základ lidského hlasu. Bývá označován termínem „základní (hlasivkový) tón“ a představuje nosný zvuk řeči. Frekvence kmitání se označuje F_0 a nazývá se fundamentální frekvence nebo také frekvence základního hlasivkového tónu. F_0 ve skutečnosti není konstantní, ale kolísá. Tomuto kolísání periody se říká jitter a závisí na duševním (hlavně emocionálním) stavu mluvčího. Kolísání amplitudy hlasivkových pulsů se označuje jako shimmer. Při normální řeči kolísá jitter v rozmezí 0,5 – 1% a shimmer mezi 0,04 – 0,21 dB. Takto malé změny posluchač vůbec nepostřehne.

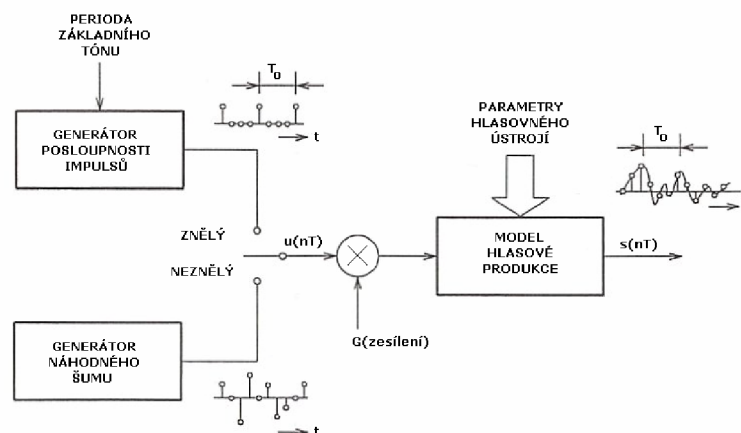


OBRÁZEK 2.2: SCHÉMA KLIDOVÉHO (a) A FONAČNÍHO (b) POSTAVENÍ HLASIVEK

Klidové postavení hlasivek a fonační postavení představují dvě základní protikladné varianty činnosti hlasového ústrojí. Při fonačním postavení jsou vytvářeny ZNĚLÉ zvuky a při klidovém naopak zvuky NEZNĚLÉ.

2.1 Model vytváření řeči

Hlasové ústrojí je jen jednou z částí, která se podílí na vytváření řeči. Dalšími jsou dechové ústrojí (základní zdroj energie pro řeč) a artikulační ústrojí. Model popisující způsob vytváření řeči je na **Obrázek 2.3** [3].



OBRÁZEK 2.3: MODEL VYTVÁŘENÍ ŘEČI

2.1.1 Popis funkce modelu pro vytváření řeči

Z pohledu na **Obrázek 2.3** je vidět, že se model skládá ze tří bloků. Bloky „GENERÁTOR POSLOUPNOSTI IMPULZŮ“ a „GENERÁTOR NÁHODNÉHO ŠUMU“ slouží k vytvoření buzení pro „MODEL HLASOVÉ PRODUKCE“. Pokud je přepínač přepnut nahoru, přivádí se do modelu hlasové produkce periodický signál z horního generátoru. Bylo zjištěno, že řečové úseky, které jsou periodické, jsou také ZNĚLÉ (samohlásky a některé znělé souhlásky). Naopak pokud je přepínač přepnut dolů, na vstup modelu hlasové produkce je přiváděn náhodný signál (šum). Takto jsou tvořeny všechny NEZNĚLÉ úseky řeči. Zesílení G umožňuje měnit intenzitu vytvořené řeči. Model hlasové produkce postihuje všechny další parametry hlasového ústrojí (nosní dutina, ústní dutina, rty, jazyk, zuby, hrdelní dutina, ...). Tyto parametry se v průběhu mluvení neustále mění. Bylo však zjištěno, že v úsecích 10 – 30 ms zůstávají tyto parametry téměř konstantní, hovoříme o tzv. krátkodobých modelech.

3. Analýza řečového signálu

3.1 Úvod

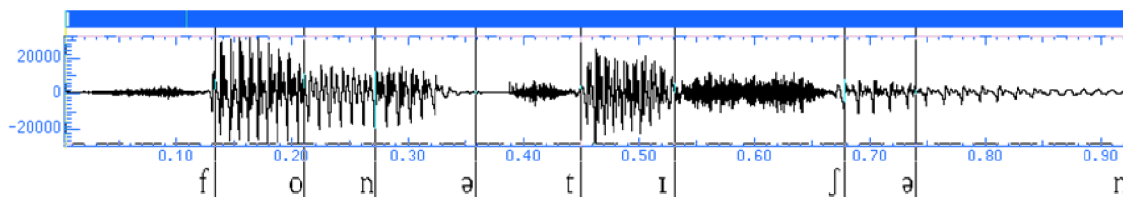
Fonetika je částí jazykových věd. Zabývá se zvuky, které produkují lidské hlasové orgány (hlasivky) a více přesněji zvuky, které jsou používány v lidské řeči. Důležitým aspektem výzkumu fonetiky (hlasu) je pomocná analýza řeči. Ta je často nazývána jako experimentální fonetika (experimental phonetics) nebo „počítačová“ fonetika (machine phonetics).

Pomocná analýza může být provedena užitím jednoho z mnoha dostupných prostředků. Může to být například X-ray (využívá zdroj záření X) fotografie a film, air-flow tubes (proudění vzduchu v rourách), elektromyografie (EMG), spektrografy, miringografy, elektroglotografy a další. Cílem většiny těchto metod je nějakým způsobem vizualizovat řečový signál a přenést jej na papír nebo v dnešní době spíše na obrazovku počítače. Počítač je nejjednodušším a obvykle nejrychlejším nástrojem. Zobrazený signál potom analyzujeme - zkoumáme určité charakteristické rysy jako například fundamentální frekvenci, formanty atd. .

3.2 Druhy zobrazení řečového signálu

3.2.1 Oscilogram

Vzato fyzikálně je řečový signál (vlastně každý zvukový signál) skupinou změn tlaku v prostředí mezi zdrojem zvuku a příjemcem. Nezákladnější reprezentací řečového signálu je oscilogram, často také nazývaný jako „časový průběh signálu“. Osou času je horizontální osa (zleva doprava). Křivka signálu nám dává informaci o změnách tlaku (zvyšování, snižování). Na ukázkou bylo vybráno slovo „**phonetician**“ (česky fonetik neboli odborník v hláskosloví), namluveno dospělým mužem americkou angličtinou. Výraz byl přepsán užitím IPA hláskové abecedy, která je nejčastěji používána. Signál je rozdělen na jednotlivé fonémy. Na **Obrázek 3.1** [18] je vidět průběh signálu a také, která část signálu odpovídá danému fonému. Na ose X je zobrazen čas [s] a na ose Y amplituda signálu [mV].

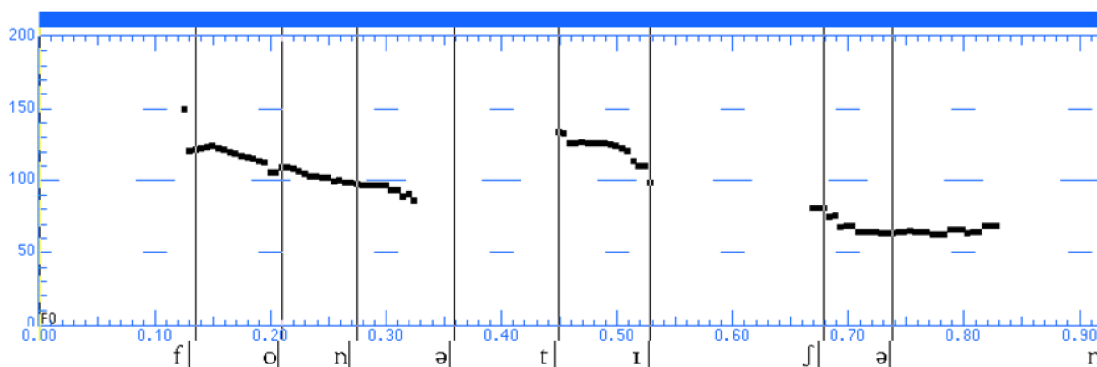


OBRÁZEK 3.1: OSCILOGRAM SLOVA „PHONETICIAN“

3.2.2 Fundamentální (základní) frekvence

Další reprezentaci řečového signálu můžeme získat analýzou jeho hlavní (základní) frekvence. Na řečový signál je možné nahlížet jako na fyzikální proces skládající se ze dvou částí: ze zdroje zvukového signálu (hlasivek) a z filtru, kterým tento signál prochází a kde se také pozmění (jazyk, rty, zuby atd.). Tato analýza se snaží o odhalení základní (fundamentální) frekvence zdroje zvukového signálu pomocí analýzy výsledného řečového signálu. Fundamentální frekvence je hlavní frekvence produkovaná hlasivkami. Analýza řečového signálu je poměrně komplikovaná. Vyskytuje se zde několik problémů, ztěžujících rozhodnutí o tom, zda je daná část signálu znělá či ne. Je tedy obtížné určit, které oscilace (přísluší znělým hláskám) jsou opravdu oscilacemi zdroje signálu (tedy hlasivek) a které oscilace vznikly při průchodu původního harmonického signálu „filtrem“ (ten tvoří jazyk, rty atd.). Pro určení těchto frekvencí bylo vyvinuto několik algoritmů. Bohužel tyto algoritmy nepracují zcela správně pro všechny druhy řečového signálu. Základní frekvence je nejvíce ovlivněna tím, jak posluchač vnímá intonaci a přízvuk řečníka.

Na **Obrázek 3.2** [18] je vidět základní frekvence (F_0 = základní formant) daného řečníka. Osa X zobrazuje čas[s] a osa Y frekvenci [Hz].



OBRÁZEK 3.2: ZÁKLADNÍ FREKVENCE (FORMANT) DANÉHO ŘEČNÍKA

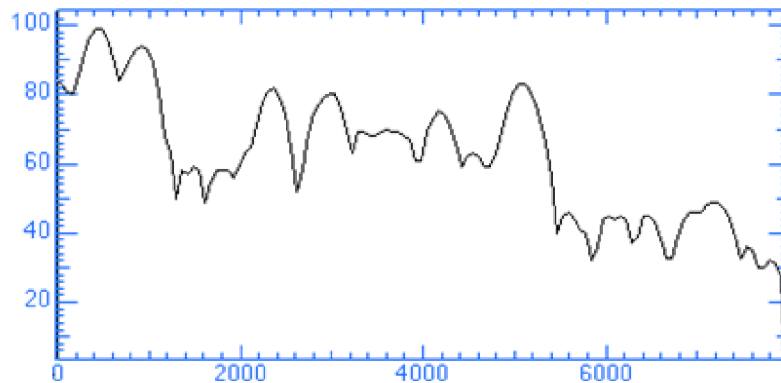
Křivka F_0 je zobrazena jako množina bodů. Body se objeví pouze tam, kde hlasivky řečníka rezonují (u znělých hlásek). Z **Obrázek 3.2** je vidět, že základní frekvence leží mezi 100 - 150 Hz. To odpovídá mužskému mluvčímu. Typické rozsahy F_0 jsou uvedeny v tabulce **Tabulka 3.1**.

Tabulka 3.1: Hodnoty F_0 pro různá pohlaví

Muž	80 - 200 Hz
Žena	150 - 350 Hz

3.2.3 Spektrum

Souhlasně se základní teorií může být každý periodický signál popsán součtem jednotlivých základních signálů (sinusovek), přičemž každý základní signál (sinusovka) může mít různou amplitudu, frekvenci a fázi. Spektrum potom zobrazuje rozložení energie signálu mezi jednotlivé frekvence, ze kterých se signál skládá. Na **Obrázek 3.3** [18] je zobrazeno spektrum samohlásky „o“. Na ose X je zobrazena frekvence [Hz] a na ose Y je zobrazena intenzita [%].

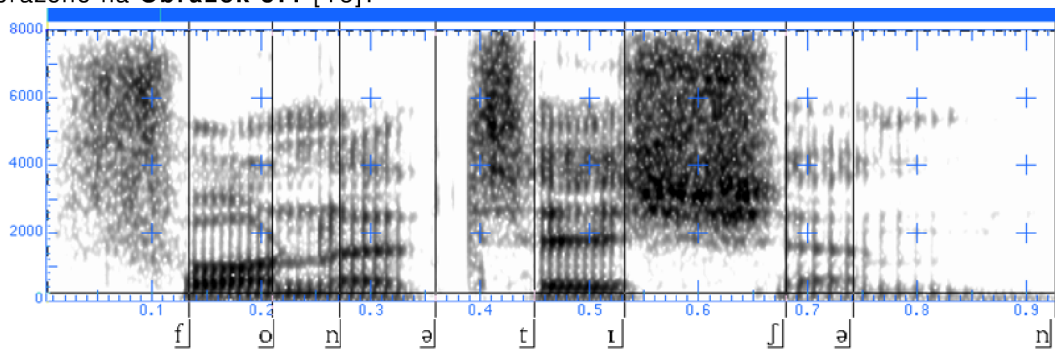


OBRÁZEK 3.3: SPEKTRUM SAMOHLÁSKY „O“

Pokud bychom chtěli zobrazit spektrum řečového signálu jako funkci času, dostali bychom trojrozměrnou reprezentaci signálu. Jednou z možností je použít spektrogram.

3.2.4 Spektrogram

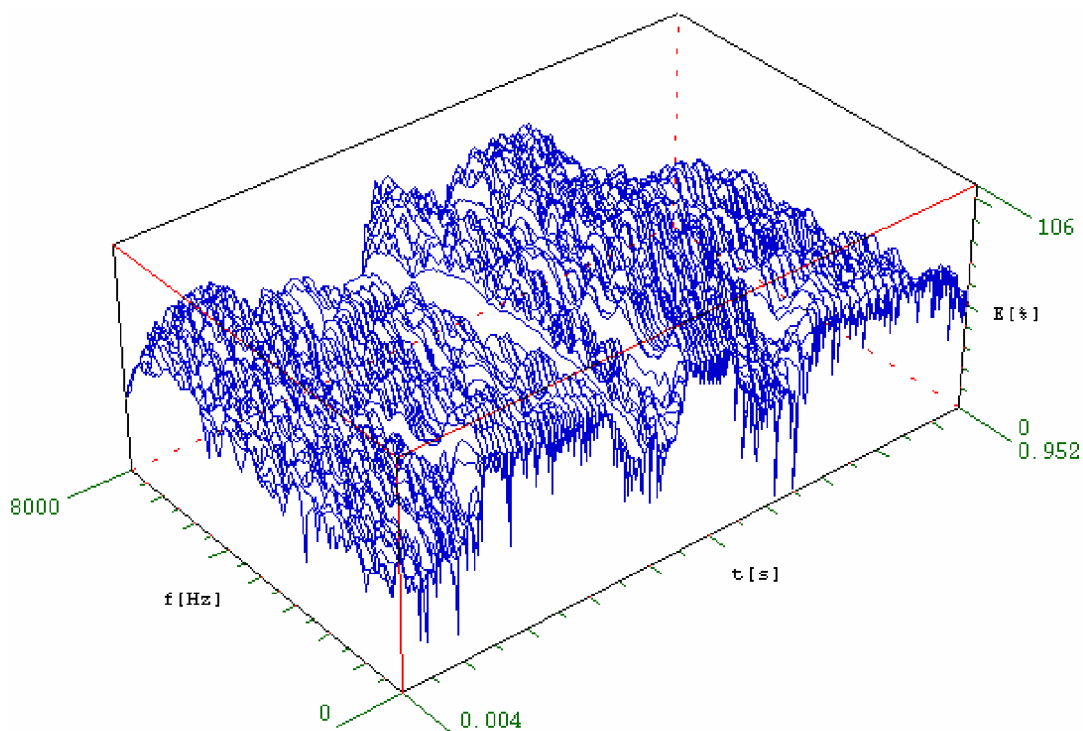
Ve spektrogramu je osa času[s] horizontální osou a osa frekvence[Hz] je vertikální osou. Třetí dimenzi (osa Z) tvoří amplituda. Ta je reprezentována odstíny šedi. Vysoká intenzita je reprezentována tmavými body (nejvyšší body černými) a nízká intenzita je reprezentována světlými body (nulová body bílými). Spektrogram slova „phonetician“ je zobrazeno na **Obrázek 3.4** [18].



OBRÁZEK 3.4: SPEKTRGRAM SLOVA „PHONETICIAN“

3.2.5 Kaskádový spektrogram (Waterfall spectrogram)

Kaskádový spektrogram je další způsob, jak lze „trojrozměrně“ zobrazit řečový signál. Na ose X je zobrazen čas [s] na ose Y frekvence [Hz] a na ose Z intenzita (většinou normovaná vůči nejvyšší hodnotě intenzity) signálu [%] (dané frekvenční složky signálu). Kaskádový spektrogram slova „phonetician“ je vidět na **Obrázek 3.5** [18].



OBRÁZEK 3.5: KASKÁDOVÝ SPEKTROGRAM SLOVA „PHONETICIAN“

Z **Obrázek 3.5** můžeme odhadnout jednotlivé formanty (nejvýznamnější frekvence). Formanty je zde možné vidět v podobě po sobě následujících nejvýše položených bodů v každém (dílkém) spektru.

3.3 Popis řeči

Nejčastěji se řeč zapisuje pomocí abecedy (s využitím pravidel pravopisu). Osoby zabývající se fonetikou však již delší dobu používají pro zápis řeči jinou abecedu. Tato abeceda se snaží udržovat co nejbližší vztahy mezi znaky a psanými a vyslovovanými. Nejpoužívanější abecedou je IPA (International Phonetic Association). Vznikla v 19. století a její poslední revize pochází z roku 1989. IPA se snaží zahrnout fonémy (základní prvky řeči) všech světových jazyků.

Popis řeči se může více či méně zužovat. Můžeme se například snažit rozpoznat jaké slovo bylo v určitém jazyku vysloveno, nebo podrobněji rozpoznávat varianty tohoto slova v různých situacích (radost, stres atd.).

3.4 Způsob získání digitalizovaného řečového signálu

Abychom mohli řečový signál zkoumat, je vhodné jej zaznamenat. Nejvhodnější bude řečový signál zdigitalizovat (zaznamenat například do počítače). Existuje několik způsobů, jak získat z řečového signálu digitální signál.

Záznam řeči pomocí mikrofonu

Jedná se bezesporu o nejčastější metodu získání digitalizovaného řečového signálu. K počítači jenž je vybaven vhodnou zvukovou kartou je připojen mikrofon. Za pomoci vhodného nahrávacího programu je potom zaznamenán úsek řeči (například několik slov). Pro další analýzu je důležité, jak je signál kvalitní (vzorkovací frekvence). Proto je důležité použít co nejkvalitnější zvukovou kartu. Na kvalitu záznamu má také vliv typ

použitého mikrofону. Vhodnější bude použít mikrofon jehož charakteristika je úzce směrová, aby při záznamu zvuku přijímal co nejméně okolního hluku (rušení).

Záznam řeči pomocí měření elektrické aktivity svalů

Nejedná se o záznam řeči jako takové, ale o záznam elektrické aktivity hlasivkových svalů. Z takto zaznamenaného signálu tedy nelze určit obsah mluveného projevu. Ostatně toto pro náš účel není důležité. Ani v prvním případě (při záznamu pomocí mikrofону) nás nezajímá obsahová stránka mluveného signálu, ale jeho významné znaky (např. formanty).

Metoda měření je nazývána zkratkou LEMG (Laryngeal Electromyography) a jejím výstupem je elektroglograf. Dává informaci zda je signál na vstupu nervu měřeného svalu v pořádku nebo ne. V průběhu měření je elektrická aktivita zaznamenána, zesílena a zobrazena monitoru. Získaný signál může být také klasicky přehrán v reproduktoru.

Vstupní signál nervu je nezbytný pro pohyb svalu. Je to vlastně elektrický impuls spouštějící elektrickou aktivitu ve svalových vláknech, které jsou na něj napojené. Ta se potom šíří svalem, způsobuje chemické změny a nakonec vyvolá jeho pohyb.

Při měření je sledována aktivita čtyř svalů (dva typy na obou stranách hlasivkových chlopní). Tím dostaneme informace o dvou nervech spojených s hlasivkovým ústrojím, označených jako: RLN (Recurrent Laryngeal Nerve) který vytváří základní tón a SLN (Superior Laryngeal Nerve) díky kterému je možné měnit intenzitu základního tónu.

Snímání elektrické aktivity bylo dříve zajišťováno pomocí tenkých jehliček, které se „zapíchny“ do měřených svalů a které tedy plnily funkci měřících elektrod. Po „zapíchnutí“ měřících elektrod musí měřený objekt zůstat v klidu a pokud možno nemluvit a nepolykat. Po nějaké době je vyzván aby vyslovil „eee“. Nejdříve potichu a potom hlasitěji. Podle změřených hodnot je pak stanovena diagnóza. Nyní se místo jehliček používají měřící elektrody, které se přiloží ke krku (asi jako obojek) a vyhodnocuje se jimi impedance hlasivek. Získáme takzvaný EGG (Electro Glotto Graph) signál.

4. Zdroje informací

4.1 Úvod

Při získávání informací jsem využíval sítě internet. Většinou se jednalo o odborné články případně příspěvky z konferencí. Volně přístupné byly tedy pouze abstrakty. U většiny takto nalezených informací ale existuje možnost daný článek zakoupit.

4.2 Odborné články

Články a příspěvky jsou v angličtině. Budu zde tedy uvádět anglické názvy. Ke každému bude stručně uvedeno, čím se zabývá, kde a kdy byl publikován a kdo je jeho autorem.

4.2.1 Články týkající se rakoviny hlasivek

NADPIS:	Prospective computer-assisted voice analysis for patients with early stage glottic cancer: a preliminary report of the functional result of laryngeal irradiation
AUTOR:	Harrison LB, Solomon B, Miller S, Fass DE, Armstrong J, Sessions RB Department of Radiation Oncology, Memorial Sloan-Kettering Cancer Center, New York
VYDÁNO:	International journal of radiation oncology, biology, physics. 1990 Jul;19(1):123-7
ODKAZ:	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=2380077&dopt=Abstract
OBSAH:	Bylo zkoumáno 25 pacientů s rakovinou hlasivek v ranném stádiu. Před radioterapií (působení ionizujícího záření na buňky) byl jejich hlas analyzován pomocí počítače. Po skončení radioterapie bylo měření provedeno znovu. Z naměřených dat byly určeny 3 odlišné vzory řeči.

NADPIS:	Multidimensional assessment of voice characteristics after radiotherapy for early glottic cancer
AUTOR:	Verdonck-de Leeuw IM, Hilgers FJ, Keus RB, Koopmans-van Beinum FJ, Greven AJ, de Jong JM, Vreeburg G, Bartelink H Institute of Phonetic Sciences/IFOTT, University of Amsterdam, The Netherlands
VYDÁNO:	American Otological Society.; American Laryngological, Rhinological, and Otological Society. 1999 Feb;109 (2 Pt 1):241-8
ODKAZ:	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?itool=abstractplus&db=pubmed&cmd=Retrieve&dopt=abstractplus&list_uids=10890774
OBSAH:	Byla zkoumána kvalita hlasu u 60 pacientů, kteří dříve podstoupili radioterapii. Výsledkem bylo zjištění, že charakteristiky hlasu u pacientů, kteří podstoupili radioterapii se zlepšili (v porovnání s charakteristikami hlasu zdravých jedinců).

NADPIS:	Quantitative analysis of voice quality in early glottic laryngeal carcinomas treated with radiotherapy
AUTOR:	Dagli AS, Mahieu HF, Festen JM University Hospital VU, Department of Otorhinolaryngology, Head and Neck Surgery, Amsterdam, The Netherlands
VYDÁNO:	European archives of oto-rhino-laryngology : official journal of the European Federation of Oto-Rhino-Laryngological Societies (EUFOS) : affiliated with the German Society for Oto-Rhino-Laryngology - Head and Neck Surgery. 1997;254 (2):78-80
ODKAZ:	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?itool=abstractplus&db=pubmed&cmd=Retrieve&dopt=abstractplus&list_uids=9065660
OBSAH:	Byly porovnávány parametry hlasu u pacientů s vyléčenou rakovinou hlasivek v ranném stádiu. Výsledkem bylo zjištění, že kvalitu hlasu u pacientů, kteří podstoupili radioterapii nelze považovat za normální.

NADPIS:	Vocal changes in patients undergoing radiation therapy for glottic carcinoma
AUTOR:	Miller S, Harrison LB, Solomon B, Sessions RB Speech and Hearing Center, Memorial Sloan-Kettering Cancer Center, New York
VYDÁNO:	American Otological Society.; American Laryngological, Rhinological, and Otological Society. 1990 Jun;100 (6):603-6
ODKAZ:	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?itool=abstractplus&db=pubmed&cmd=Retrieve&dopt=abstractplus&list_uids=2348739
OBSAH:	Příspěvek se zabývá zkoumáním změn charakteru hlasu u pacientů s rakovinou hlasivek během radioterapie. Předběžný výzkum našel tři různé způsoby chování: snížení, zvýšení a kolísání poměru hlasu (voice ratio) v průběhu radioterapie.

4.2.2 Články týkající se Parkinsonovy choroby

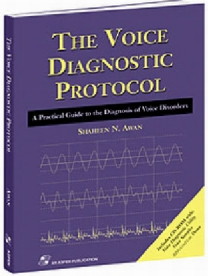
NADPIS:	Parkinsonian Voice Acoustic Analysis in Real-Time after Stereotactic Thalamotomy
AUTOR:	Miryana Nagulica, Jovan Davidovicb, Iliya Nagulica Institute of Neurosurgery, Faculty of Medicine, Belgrade University, and Aviation Medicine, Belgrade, Serbia & Montenegro
VYDÁNO:	Stereotactic and Functional Neurosurgery 2005;83:115-121 (DOI: 10.1159/000087308)
ODKAZ:	http://content.karger.com/ProdukteDB/produkte.asp?Aktion=ShowAbstract&ProduktNr=224132&Ausgabe=231222&ArtikelNr=87308&ContentOnly=false
OBSAH:	U sedmi pacientů s parkinsonovou chorobou byl analyzován hlas týden před a týden po thalamotomii. Po operaci bylo u všech pacientů zaznamenáno dvojnásobné zvýšení intenzity zvuku (při vyslovení „306“) oproti stavu před operací. Zlepšila se také plynulost výslovnosti.

NADPIS:	Processing and Analysis of Voice Anomalies in Course of Parkinson's Diseases
AUTOR:	A. Izworski, P. Augustyniak, and T. Orzechowski (Poland)
VYDÁNO:	From Proceeding (534) Signal and Image Processing - 2006
ODKAZ:	http://www.actapress.com/Content_Of_Proceeding.aspx?ProceedingID=399
OBSAH:	Autoři se zabývají analýzou hlasu osob postižených parkinsonovou chorobou. Snaží se najít odchylky od normálního charakteru hlasu, které na základě tohoto onemocnění vznikají. Z výsledků potom určili důležité parametry takto ovlivněného hlasu.

4.2.3 Články týkající se analýzy řeči a hlasu obecně

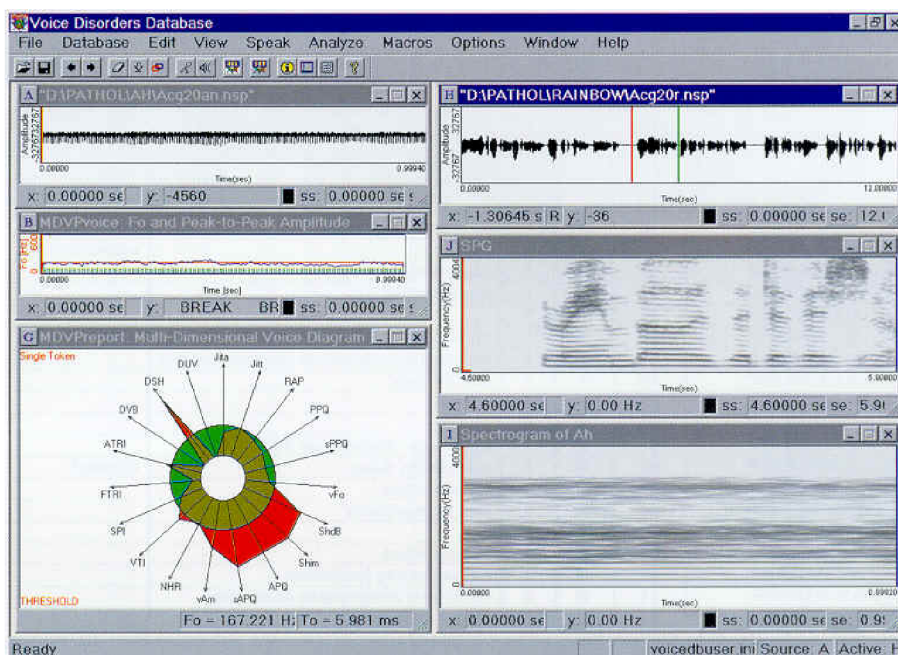
NADPIS:	Computer-Assisted Voice Analysis
AUTOR:	Paolo Campisi, MD, MSc; Ted L. Tewfik, MD, FRCSC; John J. Manoukian, MD, FRCSC; Melvin D. Schloss, MD, FRCSC; Elaine Pelland-Blais, MOA, SLP(C); Nader Sadeghi, MD, FRCSC
VYDÁNO:	Archives of Otolaryngology–Head & Neck Surgery Reader's Choice: Continuing Medical Education
ODKAZ:	http://archotol.ama-assn.org/cgi/content/full/128/2/156
OBSAH:	Autoři se zabývají vytvořením základní databáze pro MDVP (Multi-Dimensional Voice Program). Takto vytvořenou databázi potom porovnali s databází hlasových profilů pacientů s uzlinami na hlasivkách. Závěr tvoří zjištění, že pacienti s uzlinami na hlasivkách vykazovali shodné zvukové profily a že pomocí základní („zdravé“) databáze je tedy možné odhalit různé patologické abnormality hlasivek.

4.3 Knihy zabývající se analýzou hlasu

NÁZEV:	The Voice Diagnostic Protocol
AUTOR:	Shaheen N. Awan
ODKAZ:	http://www.alimed.com/ProductDetail.asp?style=88872&fprd=The+Voice+Diagnostic+Protocol&oid1=&oid2=
OBSAH:	 <p>Kniha se zabývá procentuální analýzou, objektivní analýzou frekvence hlasu, intenzity hlasu a jeho kvality. Cena této knihy je 66 dolarů.</p>

4.4 Databáze patologických hlasů

Nalezl jsem pouze jednu databázi patologických hlasů. Jedná se o databázi obsahující 1400 zvukových vzorků. Jsou v ní zastoupeny vzorky hlasu od pacientů s neurologickými, psychogenetickými, organickými a mnoha dalšími poruchami. Program se jmenuje **Disordered Voice Database and Program (Model 4337)** a je vyvíjen společností KayPENTAX v laboratořích *Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab* nacházejících se v Bostonu. Adresa internetových stránek laboratoří je <http://www.meei.harvard.edu/>. Na **Obrázek 4.1** je ukázáno prostředí programu.



OBRAZEK 4.1: PROSTŘEDÍ PROGRAMU DISORDERED VOICE DATABASE AND PROGRAM


Adresa stránek společnosti je <http://www.kayelemetrics.com/>.


4.5 Organizace zabývající se výzkumem v oblasti hlasu

NÁZEV:	Center For Voice Disorders of Wake Forest University
ODKAZ:	http://www.wfubmc.edu/
POPIS ČINNOSTI:	Centrum bylo založeno roku 1981, na Wake Forest University a stalo se tak jedním z prvních komplexních, multidisciplionárních hlasových center ve Spojených státech. Sídlí na oddělení otolaryngologie (ušní a krční lékařství) univerzitní zdravotní školy Wake Forest. Díky svému špičkovému vybavení poskytuje pacientům tu nejlepší péči a současně podporuje další výzkum a vzdělání v oblasti hlasových, řečových a polykacích poruch.

NÁZEV:	Departments of Otolaryngology (Drs Campisi, Tewfik, Manoukian, Schloss, and Sadeghi) and Speech-Language Pathology (Ms Pelland-Blais), The Montreal Children's Hospital, McGill University Health Centre, Montreal, Quebec
ODKAZ:	http://www.muhc.ca/pfv/mch/ , http://www.muhc.ca/
POPIS ČINNOSTI:	Výzkumný institut MUHC je mezinárodně uznávané biomedicínské a zdravotnické nemocniční výzkumné centrum. Institut podporuje přes 500 výzkumníků stejně jako 1000 graduálních a postgraduálních studentů. Operuje s více jak 300 laboratořemi vybavenými pro široké spektrum základních a klinických výzkumů, publikujících ročně přibližně 1400 příspěvků. Hlavními obory výzkumů jsou: dětské lékařství (The Montreal Children's Hospital), lékařství, chirurgie, neurologie, gynekologie a psychiatrie.

NÁZEV:	Massachusetts Eye and Ear Infirmary
ODKAZ:	http://www.meei.harvard.edu/
POPIS ČINNOSTI:	<p>Institut se zabývá výzkumem v mnoha odvětvích. Pro nás je zajímavé odvětví otolaryngologie (ušní a krční lékařství). Konkrétně výzkum nazvaný <i>Objective Assessment of Vocal Hyperfunction (objektivní odhad hlasových hyperfunkcí)</i>. Výzkum se prováděl v <i>Voice and Speech Laboratory Harris Pevton Mosher Larvngological Research Laboratory</i>.</p> <p>Z názvu je patrné, že jde o zkoumání hlasových hyperfunkcí (to jsou nejčastěji typy hlasových poruch). Výsledky výzkumu měli později pomoci při odhalování poruch hlasu.</p> <p>Odkaz na abstrakt toho výzkumu je: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=2739390&dopt=Abstract.</p> <p>Tato instituce se také podílí na vývoji výše zmíněného programu <i>Disordered Voice Database and Program</i>.</p>

NÁZEV:	Laboratoř zpracování řečového signálu – katedra teorie obvodů, FEL ČVUT
ODKAZ:	http://noel.feld.cvut.cz/speechlab/start.php?page=history&lang=cz
POPIS ČINNOSTI:	<p>Skupina zpracování řeči a analýzy signálů vznikla v roce 1981. V počáteční fázi byl výzkum zaměřen na analýzu a syntézu řečového signálu. V následující době byl výzkum v oblasti rozpoznávání řeči zaměřen na zlepšení a návrh nových možností parametrizace pro účely rozpoznávání s cílem zvýšit množství rozpoznávaných slov až k rozpoznávání spojitě řeči. V návaznosti na výzkum v oblasti rozpoznávání je od roku 1992 věnována velká pozornost chování systémů zpracování řeči narušené šumem. Od roku 1997 spolupracuje na problematice rozpoznávání řeči v reálných podmínkách. Tzv. robustní rozpoznávání řeči se stalo hlavní výzkumnou aktivitou skupiny. Hlavní pozornost je věnována rozpoznávání v prostředí jedoucího automobilu a rozpoznávání po telefonní lince.</p> 

NÁZEV:	Acoustics Research Institute, Rakousko	
ODKAZ:	http://www.kfs.oeaw.ac.at/	
POPIS ČINNOSTI:		<p>Institut se zabývá výzkumem ve čtyřech hlavních odvětvích:</p> <ul style="list-style-type: none"> - počítačová akustika - počítačové slyšení, psychoakustika - experimentální audiologie - digitální zpracování signálu a matematika

NÁZEV:	Laboratoř inteligentních komunikačních systémů (LİKS). Katedra informatiky a výpočetní techniky - FAV ZČU (Západočeská univerzita v Plzni)	
ODKAZ:	http://www.kiv.zcu.cz/	
POPIS ČINNOSTI:		<p>Laboratoř je zaměřena na získání teoretických i aplikovaných výsledků v oblasti rozpoznávání a syntézy mluvené řeči (češtiny), reprezentace významu výrazů přirozeného jazyka a využití metod strojového učení pro zpracování přirozeného jazyka. Zabývá se problematikou základních počítačových řečových technologií a výzkumem v oboru dialogových systémů (modely a implementace dialogových systémů, dialogové strategie, jazyky pro jejich popis, apod.) a nabízí vybraným studentům možnost začlenění se do výzkumu a práce laboratoře počínaje programováním a konče teoretickým výzkumem. Podstatným cílem pedagogickým je na základě zmíněného výzkumu vyškolit řadu postgraduálních i pregraduálních studentů v nově vznikajícím pomezím oboru, pro který se ujal název "jazykové inženýrství" (language engineering).</p>

4.6 Archivy odborných článků

Spoustu informací lze také nalézt v různých archivech konferencí a odborných článků. Nejrozsáhlejší jsou asi následující dva archivy.

NÁZEV:	Institute of Electrical and Electronics Engineers
ODKAZ:	http://www.ieee.org/portal/site
POPIS:	Organizace sdružující 365,000 členů z 150 zemí světa.

NÁZEV:	ISCA (International Speech Communication Association)
ODKAZ:	http://www.isca-speech.org/archive/index.html
POPIS:	Archiv obsahuje 14336 příspěvků z 90 konferencí a seminářů. 10992 příspěvků pochází z 18 konferencí (INTERSPEECH, dříve EUROSPEECH a ICSLP); 2576 příspěvků z 63 ISCA (ESCA) výukového a výzkumného semináře a 768 příspěvků pochází z 9 akcí podporovaných organizací ISCA.

NÁZEV:	The Journal of the Acoustical Society of America
ODKAZ:	http://scitation.aip.org/jasa/
POPIS:	Organizace vznikla roku 1929 a sdružuje 7000 členů z celého světa.

5. Diagnózy stanovované pomocí analýzy hlasu

5.1 Úvod

Nejčastěji se analýza hlasu používá ke stanovení raného stádia rakoviny hlasivek. Jedná se totiž o neinvazivní a poměrně nenákladnou metodu. Pomocí analýzy hlasu lze však stanovit také spoustu dalších poruch hlasu spojené většinou s onemocněním hlasivek. Byly vytvořeny databáze patologických hlasů - to jsou hlasy lidí, kteří trpí nějakou poruchou. Pomocí porovnání vzorků v databázi se vzorky od nového (vyšetřovaného) pacienta je pak možné určit zda pacient není nemocen.

5.2 Způsob stanovení diagnózy

Pro stanovování diagnóz se využívá znělých úseků řeči. Při vytváření znělých úseků totiž hlasivky vykonávají periodický pohyb. Při tomto pohybu se opakovaně otevírají (povolna) a uzavírají (rychle). Pohyb hlasivek při mluvení je řízen tzv. hlasivkovými pulzy. Průběh hlasivkových pulzů lze získat pomocí elektroglotografu (LEMG). Jednodušší je zaznamenávat až výsledný řečový signál (časový průběh amplitudy). Ten v sobě informaci o tvaru hlasivkových pulzů nese také, je ovšem v signálu „skryta“. Proud vzduchu modulovaný hlasivkami se totiž průchodem přes hlasové ústrojí (myslím teď celý systém pro vytváření řeči) dále mění. Hlasivkový pulz může být také ovlivněn organickými nesymetriemi (například nádor) hlasivek. Pro zjištění tvaru hlasivkových pulzů z řečového signálu, je možné využít některý z početných postupů, které budou popsány dále.

5.3 Hlasivkové pulzy

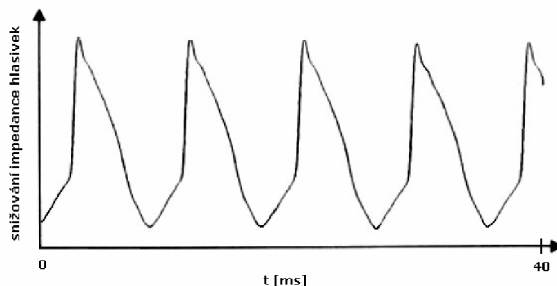
Zkoumáním hlasivkových pulzů se zabývala a zabývá celá řada odborníků. Jak říká [4] „roku 1971 bylo Ladefogedem stanoveno 9 rozdílných řečových signálů a k nim příslušející tvary hlasivkových impulzů. Při stanovování těchto signálů se nejdříve stanovil základní řečový signál („modal voice“ - proud vzduchu procházející hlasivkami je modulován pouze jejich periodickým otevíráním a zavíráním a není ovlivněn žádnými poruchami), tak aby ostatní typy mohli být v porovnání s ním jednoznačně odlišeny. Některé z později stanovených signálů však nebylo možné v porovnání se základním jednoznačně odlišit. Podrobnější definici zbývajících 8 typů tedy doplnil v tomtéž roku Hollien.

Výše zmíněnými devíti druhy řečových signálů jsou:

- 1) typický řečový signál (modal voice)
- 2) pisklavý (falsetto)
- 3) skřípavý (creak)
- 4) šepot (whisper)
- 5) chraptivý (harshness)
- 6) dýchavičný (breathiness)
- 7) pisklavý s šumem (whispery falsetto)
- 8) pisklavý s praskáním (creaky falsetto)
- 9) pisklavý s praskáním a šumem (whispery creaky falsetto)

5.3.1 Typický řečový signál (modal voice)

Základní frekvence tohoto signálu se nachází v nižší části frekvenčního spektra, užívaného při normální konverzaci. Tato frekvence se liší u mužů (80-160Hz), žen (150-300Hz) a dětí (200-420Hz). Pohyb hlasivek je pravidelný a výsledný řečový signál neobsahuje slyšitelné vady (ty vznikají pokud nejsou hlasivky při svém pohybu zcela uzavírány). Příklad odpovídajícího hlasivkového pulsu je na **Obrázek 5.1**.



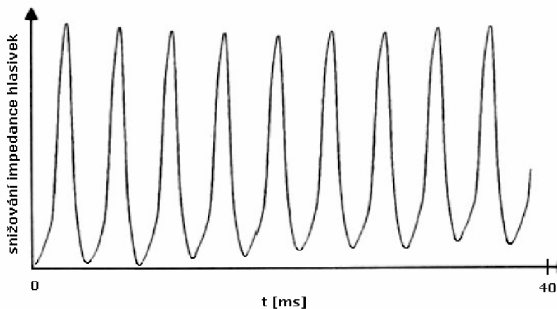
OBRÁZEK 5.1: PŘÍKLAD HLASIVKOVÝCH PULZŮ PRO TYPICKÝ ŘEČOVÝ SIGNÁL

5.3.2 Pískavý (falsetto)

Tento stav je charakterizován několika faktory:

- základní frekvence je značně vyšší než u typického řečového signálu
- díky vysoké základní frekvenci jsou vyšší harmonické rozprostřeny po celém spektru a proto v jakémkoli vybraném frekvenčním rozsahu najdeme méně zvukových komponent než je v signálu nízkou základní frekvencí
- hrana hlasivkového impulsu je strmější než u typického řečového signálu (klesá s 20dB na oktávu)

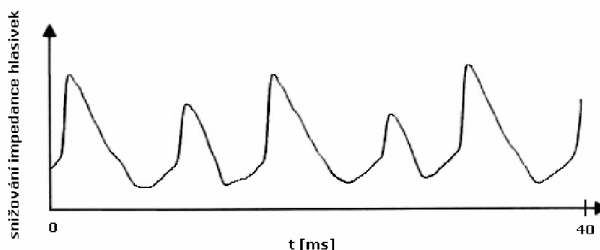
Typický tvar odpovídajícího hlasivkového impulsu je na **Obrázek 5.2**.



OBRÁZEK 5.2: PŘÍKLAD HLASIVKOVÝCH PULZŮ PRO PÍSKAVÝ SIGNÁL

5.3.3 Skřípavý (creak)

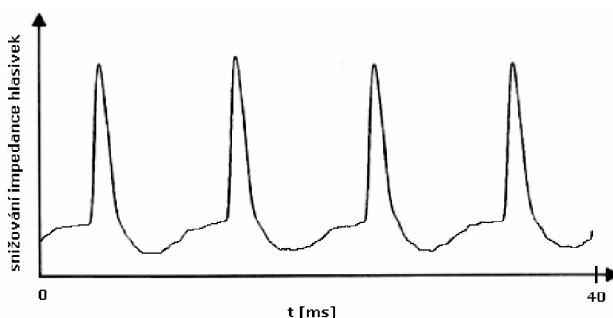
Při poslechu se jeví jako série po sobě rychle jdoucích klepnutí. Tento signál je velmi podobný s chraptivým (harshness). Signál se posluchači jeví, jako by byl tlumen (zanikal). Frekvence tohoto tlumení se pohybuje v rozmezí 20 až 90Hz. Útlum je v signálu vnímán pokud hlasivkový pulz klesá z jeho maximální hodnoty se strmostí 42 až 44 dB na oktávu. Průběh takového hlasivkového pulsu je na **Obrázek 5.3**.



OBRÁZEK 5.3: PŘÍKLAD HLASIVKOVÝCH PULZŮ SKŘÍPAVÉHO SIGNÁLU

5.3.4 Šepot (whisper)

Tento signál vzniká pokud jsou hlasivky otevřeny tak, že otvor, kterým prochází proud vzduchu z plic připomíná trojúhelník. Pokud se šum kombinuje s jiným druhem signálu jako například typickým řečovým signálem nebo falsetem vytváří, vzniká jejich směs. Průběh hlasivkového pulzu je na **Obrázek 5.4**.



OBRÁZEK 5.4: PŘÍKLAD HLASIVKOVÝCH PULZŮ ŠEPTU

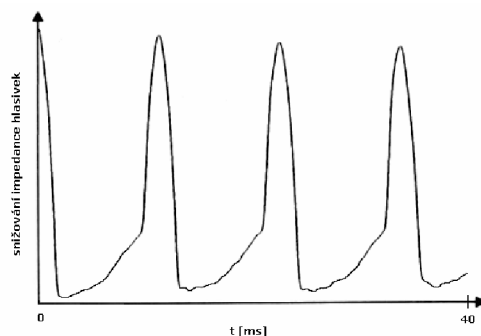
5.3.5 Chraptivý (harshness)

Takovýto řečový signál se vyznačuje, jak řekl Michel, neperiodickým nebo šumovým spektrem s úrovní základní frekvence stejnou jako u typického řečového signálu a s odchylkou frekvence větší než střední hodnota základní frekvence signálu. Tato odchylka je nazývána jitter. Člověk je velmi vnímavý i malé změny jitteru.

5.3.6 Dýchavičný (breathiness)

V porovnání s typickým řečovým signálem jsou vibrace hlasivek nedostatečné a jsou doprovázeny mírným slyšitelným třením. Námaha hlasivkových svalů je malá a proto jsou hlasivky podél většiny svojí délky poněkud otevřeny. Hlasivky se nikdy úplně neuzavřou. Každý pokus o jejich uzavření má snahu být vyrušen, čímž se snižuje odpor hlasivek vůči proudu vzduchu přicházejícího z plic a jeho rychlost se zvětšuje. Hlasivky

jsou tedy zúženy (přibližně na 25% z jejich maximálního otevření), ovšem ne tak moc, aby mohly produkovat šepot. Průběh hlasivkového pulzu je na **Obrázek 5.5**.



OBRÁZEK 5.5.: PŘÍKLAD HLASIVKOVÝCH PULZŮ DÝCHAVÍČNÉHO SIGNÁLU

6. Odhad tvaru hlasivkového pulzu z řečového signálu

Tvar hlasivkových pulzů lze určit přímo pomocí měření elektrické aktivity svalů (LEMG). Tento způsob ovšem nemusí být pro pacienta příjemný. Proto se častěji snažíme odhadnout tvar hlasivkového pulzu neinvazivní metodou a to přímo z řečového signálu.

Nejčastější přístupy pro získání hlasivkových pulzů z řečového signálu jsou dva. První z nich se snaží určit **parametry hlasového traktu** a pomocí nich pak, s použitím inverzní filtrace řečového signálu, určit hlasivkové pulzy. Druhý přístup používá některý ze známých **modelů hlasového traktu** (Liljencrants-Fantův nebo Rosenbergův model) a snaží se z řečového signálu určit parametry modelu tak, aby co nejdříveji popisoval původní hlasivkový pulz.

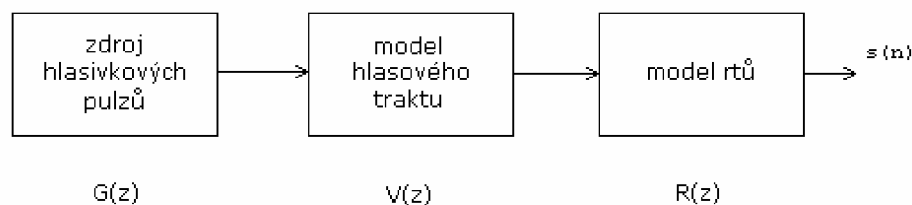
Některé z metod určování tvaru hlasivkových pulzů z řečového signálu budou popsány v následujících kapitolách. Při popisování jednotlivých metod uvádím zdroj informací v hlavním nadpisu dané kapitoly. Pokud není výslovně uvedeno jinak, všechny vztahy použité při popisu jednotlivých metody pocházejí z těchto zdrojů.

6.1 Metoda založená na opakovaném inverzním filtrování řečového signálu - Iterative Adaptive Inverse Filtering (IAIF) [4]

Metodu vytvořil Paavo Alku a publikoval ji roku 1992.

6.1.1 Princip metody

Jak je z názvu metody patrné, pro odhad hlasivkových pulzů se používá inverzní filtrace. Přitom se vychází z modelu tvorby řeči viz **Obrázek 6.1**.



OBRÁZEK 6.1.: MODEL TVORBY ŘEČI

$G(z)$ představuje hlasivkové pulzy, které z řečového signálu odhadujeme. Model hlasového traktu se popisuje all-pole filtrem. Obecný zápis v rovině Z je

$$V(z) = \frac{1}{1 + \sum_{i=1}^K c_i \cdot z^{-i}} \quad (1)$$

, kde c_i jsou koeficienty modelu hlasového traktu a K je počet koeficientů.

Funkci $rt_{\hat{u}}$ je možné chápat jako derivátor a popsat pomocí vztahu (2)

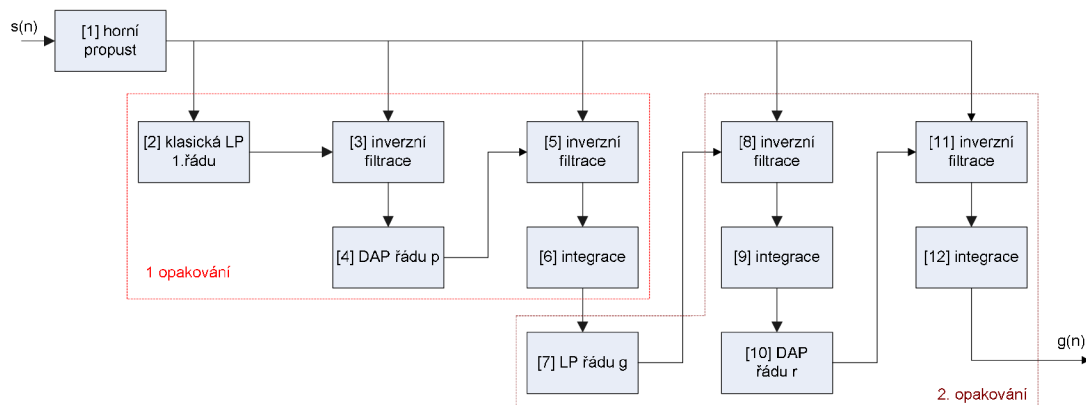
$$R(z) = 1 - \mu z^{-1} \quad , \text{kde } \mu \in \langle 0,98; 1 \rangle. \quad (2)$$

Aby byl výsledek správný, **musí** být splněny následující předpoklady:

- v průběhu odhadování parametrů hlasového traktu se jeho parametry nesmí měnit (toto lze splnit na krátkém časovém úseku řečového signálu – běžně kolem 30 – 100 ms)
- vliv $rt_{\hat{u}}$ lze modelovat jako derivátor a tudíž kompenzovat pomocí integrace

6.1.2 Blokové schéma

Na **Obrázek 6.2** je vidět blokové schéma postupu výpočtu.



OBRÁZEK 6.2.: BLOKOVÉ SCHÉMA METODY IAF

6.1.2.1 Horní propust

Na začátku musí být řečový signál filtrován horní propustí pro odstranění rušivých nízkých frekvencí. Mezní frekvence horní propusti musí být nižší než základní (fundamentální) frekvence řečového signálu. Frekvenci pro horní propust jsem zvolil 90% ze základního tónu řeči [5]. Pro určení základního tónu lze využít různé metody. Například jednoduchou metodu AMDF (Average Magnitude Difference Function), kde periodu určíme podle vzdálenosti minim[8]. V tomto případě byla použita přesnější metoda YIN, popsána v [6] a [5], která je založena na modelu hlasu (periodickém modelu). Model pak může být rozšířen tak, aby popisoval i určité druhy neperiodických signálů[6].

6.1.2.2 První opakování

Pomocí lineární predikce (LP) prvního řádu (**blok 2**) určíme koeficienty pro následující inverzní filtr. Lineární predikce je založena na předpokladu, že pomocí lineární kombinace předchozích vzorků řečového signálu lze predikovat další (neznámé) vzorky řečového signálu viz následující vztahy.

$$s[n] = -\sum_{i=1}^p a_i \cdot s[n-i] \quad (3)$$

kde a jsou koeficienty lineární predikce a p je stupeň (řád) predikce.

Pro výpočet koeficientů se nejčastěji používá autokorelační metoda [3]. Nejprve určíme autokorelaci zkoumaného signálu pomocí vzorce

$$R(m) = \sum_{n=0}^{N_0-1-m} s(n) \cdot s(n+m), \quad (4)$$

pak sestavíme matice A a v podle

$$A = \begin{bmatrix} R(0) & R(1) & R(2) & \dots & R(p-1) \\ R(1) & R(0) & R(1) & \dots & R(p-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R(p-1) & R(p-2) & R(p-3) & \dots & R(0) \end{bmatrix} \quad (5) \text{ a}$$

$$v = [R(1) \ R(2) \ \dots \ R(p)]^T. \quad (6)$$

Koeficienty potom vypočítáme pomocí

$$A \cdot a = v. \quad (7)$$

Koeficienty vypočítané pomocí LP pak použijeme pro inverzní filtraci (**blok 3**) řečového signálu. Touto filtrací odstraníme náklon spektra řečového signálu, vyvolaný spektrem budícího signálu a vyzařováním rtů.

Signál vzniklý inverzní filtrací je potom použit pro odhad parametrů hlasového traktu. Řád modelu je určen podle vzorkovací frekvence [5] a sice

$$p = \text{round}\left(\frac{fvz}{1000}\right) + 2. \quad (8)$$

Tak (**blok 4**) získáme odhad parametrů hlasového traktu. Pro odhad je použita metoda DAP (Discrete All-Pole Modeling) [5],[7] viz **Příloha A**. Důvodem je velká chyba LP při aplikaci na signály s vyšší frekvencí. Ta vzniká kvůli tomu, že LP se snaží minimalizovat chybu predikce porovnáváním autokorelace spojitého all-pole modelu s autokorelací zpracovávaného signálu a nebere přitom v potaz aliasing, který vznikl v disktrétním spektru[7].

Vstupní řečový signál je filtrován inverzním filtrem (**blok 5**) s využitím parametrů z bloku 4.

Integrací (**blok 6**) výsledku inverzního filtru z bloku 5 se potlačí vliv rtů a výsledkem je první (hrubý) odhad hlasivkových pulzů.

6.1.2.3 Druhé opakování

Vstupním signálem pro LP (**blok 7**) je předchozí hrubý odhad hlasivkových pulzů. Řád predikce je volen $g = 2$ [5]. Získáme tak model spektra popisující vliv hlasivkových pulzů na řečové spektrum.

Inverzní filtrací (**blok 8**) odstraníme vliv hlasivek na řečový signál a integrací (**blok 9**) vliv rtů.

Opět pomocí DAP (**blok 10**) získáme nový model hlasového traktu. Řád modelu r volíme stejný jako v bloku 4 tedy $r = p$.

Konečný odhad hlasivkových pulzů získáme inverzní filtrací řečového signálu (**blok 11**) pomocí koeficientů získaných v bloku 10, čímž odstraníme vliv hlasového traktu na hlasivkové pulzy. Nakonec pomocí integrace (**blok 12**) odstraníme vliv rtů.

6.2 Metoda založená na průměrování charakteristik hlasového traktu [9]

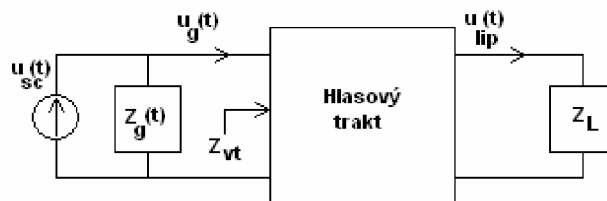
Podle autorů této metody je předpoklad, že vliv hlasivek na řečový signál je v okamžiku GCI (Glottal Closure Instant – okamžik uzavření hlasivek) nulový, špatný. I v okamžicích GCI se můžou v signálu vyskytnout vzruchy vyvolané pohybem hlasivkových úponů (pohybují hlasivkami). Tyto vzruchy jsou sice malé v porovnání s velikostí hlasivkového pulzu, ale v derivovaném hlasivkovém pulzu, se tyto malé ale rychlé změny nezanedbatelně projeví.

Metoda předpokládá, že hlasivkový pulz je periodicky stacionární signál. Aby toto tvrzení bylo správné, je potřeba dostatečně dlouhý řečový signál (kolem 3s) s téměř neměnnou hlasitostí a samozřejmě po celou dobu s neměnnou znělou hláskou.

Pomocí vztahů uvedených níže jsou pak z řečového signálu odstraněny periodické komponenty a tím je řečový signál zbaven vlivu impulzní odezvy hlasivek. Z takto získaného signálu je pak možné stanovit charakteristiku hlasového traktu. Stanovením více charakteristik hlasového traktu a jejich průměrováním je potom odhad zpřesněn.

6.2.1 Hlasivkové pulzy a hlasový trakt

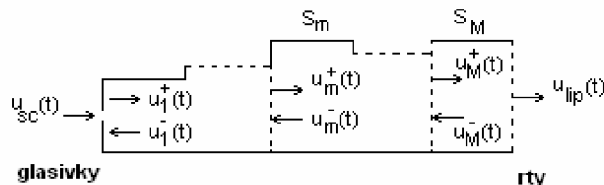
V okamžiku otevření hlasivek prochází proud vzduchu z plic přes hlasivky a je popsán jako podélný (transverzální) tlak v oblasti hlasivek. Když jsou hlasivky uzavřené, nemusí být proud vzduchu nulový, protože pohyb hlasivek může vyvolat turbulence. Model tvorby řečového signálu může být popsán modelem na **Obrázek 6.3**.



OBRÁZEK 6.3.: MODEL TVORBY ŘEČOVÉHO SIGNÁLU

Hlasivkové pulzy $u_g(t)$ jsou modelovány zdrojem vlnění $u_{sc}(t)$ a jeho impedancí $Z_g(t)$. $Z_g(t)$ se blíží k nekonečnu, pokud jsou hlasivky uzavřené. Z_{VT} vyjadřuje impedanci na vstupu hlasového traktu. Z_{LIP} představuje impedanci rtů.

Hlasový trakt funguje jako filtr, který vstupní signál $u_g(t)$ převádí na výstupní $u_{LIP}(t)$ a často bývá popsán válcovým modelem viz **Obrázek 6.4**. Je tvořen M sekcemi se stejnou šířkou ale různou plochou.



OBRÁZEK 6.4.: VÁLCOVÝ MODEL HLASOVÉHO TRAKTU

Signál $u_g(t)$ lze vyjádřit jako $u_g(t) = u_1^+(t) + u_1^-(t)$. Počet sekcí válcového modelu M je vztahen k vzorkovací frekvenci F_{VZ} a je popsán takto

$$M = \frac{2 \cdot L_{VT} \cdot F_{VZ}}{c} \quad (9)$$

kde c je rychlost šíření zvuku a L_{VT} je délka hlasového traktu.

Přenosová funkce hlasového traktu by neměla obsahovat vliv $Z_g(t)$. Můžeme ji vyjádřit vztahem

$$H_{VTF}(z) \equiv \frac{U_{lip}(z)}{U_g(z)} = \frac{U_{lip}(z)}{(U_1^+(z) + U_1^-(z))}. \quad (10)$$

Velká písmena značí Fourierovu transformaci. Celkovou přenosovou funkci zahrnující i vliv časové proměnné impedance hlasivek Z_g lze vyjádřit vztahem

$$H_{GVTF}(z) \equiv \frac{U_{lip}(z)}{U_{sc}(z)}. \quad (11)$$

6.2.2 Odhad fází hlasivkového pulzu

Pokud jsou hlasivky otevřeny, je $u_{sc}(t)$ lineární funkcí časově proměnné oblasti hlasivek. Okamžik uzavření hlasivek GCI proto může být detekován v okamžicích kdy derivace $u_{sc}(t)$ nabývá záporných špiček. Fáze otevření hlasivek je detekována pro tu část $u_{sc}(t)$, kdy jeho derivace zůstává kladná. Fáze uzavření hlasivek je naopak detekována v okamžicích kdy hodnoty derivace $u_{sc}(t)$ zůstávají záporné. Část mezi oběma fázemi (otevření a uzavření) se nazývá fáze uzavření.

Řečový signál můžeme v časové oblasti popsat následujícím vztahem

$$p_{mic}(n) = u'_{scf} \left(n - \frac{M}{2} - \Delta \right) + b_1 p_{mic}(n-1) + L + b_{M+2} p_{mic}(n-M-2) \quad (12)$$

kde M je řád válcového modelu hlasového traktu (9) a $\Delta = \frac{r \cdot F_{VZ}}{c}$, kde r je vzdálenost rtů od mikrofonu, F_{VZ} je vzorkovací frekvence signálu a $c = 350 [m/s]$ je rychlost šíření zvuku.

Předpokládejme, že signál $u'_{scf}(n - M/2 - \Delta)$ získáme inverzní filtrací řečového signálu pomocí průměrovaného GVTF a že koeficienty $\bar{b}_1, K, \bar{b}_{M+2}$ jsou průměrovanou verzí b_1, K, b_{M+2} . Dále předpokládejme, že l -tý hlasivkový cyklus koresponduje s intervalem $[n_l, n_l + 1, K, n_l + T_l - 1]$, kde T_l označuje počet vzorků v jedné periodě hlasivkového pulzu. S použitím (12) lze $\bar{b}_1, K, \bar{b}_{M+2}$ vyjádřit pomocí

$$U'_{scf} + P_l \bar{B} = p_l + E_l \quad (13)$$

a po rozepsání jako

$$\begin{bmatrix} u'_{scf} \left(n_l - \frac{M}{2} - \Delta \right) \\ \vdots \\ u'_{scf} \left(n_l - \frac{M}{2} - \Delta + (T_l - 1) \right) \end{bmatrix} + \begin{bmatrix} p_{mic}(n_l - 1) & \cdots & p_{mic}(n_l - 1 - (M + 1)) \\ \vdots & \ddots & \vdots \\ p_{mic}(n_l - 1 + (T_l - 1)) & \cdots & p_{mic}(n_l - 1 + (T_l - 1) - (M + 1)) \end{bmatrix} \cdot \begin{bmatrix} \bar{b}_1 \\ \vdots \\ \bar{b}_{M+2} \end{bmatrix} = \begin{bmatrix} p_{mic}(n_l) \\ \vdots \\ p_{mic}(n_l + (T_l - 1)) \end{bmatrix} + \begin{bmatrix} e(1) \\ \vdots \\ e(T_l) \end{bmatrix} \quad (14)$$

Vektory \bar{B} a U'_{scf} neznáme, p_l je měřen v intervalu $[n_l, n_l + 1, K, n_l + T_l - 1]$, P_l v intervalu $[n_l - M - 2, n_l - M - 1, K, n_l + T_l - 2]$ a E_l je chybový vektor daný vztahem $E_l = P_l(\bar{B} - B)$, kde B je aktuální vektor hodnot b . Vztah (13) nelze vyřešit, protože nutný počet hodnot $M + 2 + T_l$ je větší než počet dostupných hodnot v intervalu T_l . Pro vyřešení tohoto problému sestojíme stejný vztah

$$U'_{scf} + P_J \bar{B} = p_J + E_J \quad (15)$$

ovšem z dalšího nejbližšího intervalu řečového signálu označeného J. Je lepší volit další J cyklus hlasivkového pulzu co nejbližše předchozímu I-tému cyklu, aby změna hlasivkového pulzu byla co nejmenší. První vzorek n_j v J-tém cyklu musí být na stejné relativní pozici jako vzorek n_l v I-tém intervalu. Výraz (15) sestavíme podle následujících pravidel p_j je měřen v intervalu $[n_j, n_j + 1, K, n_j + T_l - 1]$, P_j v intervalu $[n_j - M - 2, n_j - M - 1, K, n_j + T_l - 2]$ a E_j je chybový vektor daný vztahem $E_j = P_j(\bar{B} - B) + (U'_{scfI} - U'_{scfJ})$ viz vzorec (14).

Kombinací (13) a (15) dostaneme

$$\begin{pmatrix} P_l \\ \vdots \\ P_J \end{pmatrix} \bar{B} = \begin{pmatrix} p_l \\ \vdots \\ p_J \end{pmatrix} + \begin{pmatrix} E_l \\ \vdots \\ E_J \end{pmatrix} \quad (16)$$

Matici koeficientů \bar{B} pak můžeme vypočítat podle

$$\bar{B} = (X^T X)^{-1} X^T Y \quad (17)$$

Odhad signálu $u'_{scf}(n - M/2 - \Delta)$ potom můžeme získat filtrací $p_{mic}(n)$ filtrem $\sum_1^{M+2} \left(1 - \bar{B}(m)z^{-m} \right)$. Získaný přenos $H_{GVTF}(z)$ může být nestabilní. Toto lze vyřešit volbou dalších 2 hlasových cyklů a použitím jiných segmentů hlasového signálu pro vytvoření X a Y.

Z $u'_{scf}(n - M/2 - \Delta)$ nakonec určíme (způsobem, který je uveden na začátku této kapitoly) jednotlivé fáze hlasivkového pulzu.

6.2.3 Odhad koeficientů filtru hlasového traktu

V této kapitole získáme nezaujatý odhad VTF (Vocal Tract Filter). Řečový signál lze vyjádřit vzorcem

$$p(n) = p_{mic}(n + \Delta). \quad (18)$$

Můžeme také psát

$$p(n) = u'_g \left(n - \frac{M}{2} \right) + a_1 p(n-1) + \dots + a_{M+1} p(n-M-1). \quad (19)$$

Při výpočtu koeficientů a_m můžeme použít pouze ty vzorky $p(n)$, které nejsou ovlivněny otevřenými hlasivkami. Interval, kde se tyto vzorky nachází popíšeme jako $[n_{ci} - M - 1, n_{ci} - M, K, n_{ci} + L - 1]$ v i -tém cyklu, kde $L = N_c - M - 1$ a N_c je délka uzavření hlasivek. Na základě (19) můžeme psát

$$U'_{gfc_i} + P_{ci} A = p_{ci} \quad (20)$$

kde U'_{gfc_i} a A neznáme. Vztah (20) můžeme rozepsat takto

$$\begin{aligned} & \begin{bmatrix} u'_{gf} \left(n_{ci} - \frac{M}{2} \right) \\ \vdots \\ u'_{gf} \left(n_{ci} - \frac{M}{2} + (L-1) \right) \end{bmatrix} + \begin{bmatrix} p(n_{ci} - 1) & \dots & p(n_{ci} - (M+1)) \\ M & \dots & M \\ p(n_{ci} - 1 + (L-1)) & \dots & p(n_{ci} - 1 + (L-1) - (M+1)) \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ \vdots \\ a_{M+1} \end{bmatrix} = \\ & = \begin{bmatrix} p(n_{ci}) \\ M \\ p(n_{ci} + (L-1)) \end{bmatrix}. \end{aligned} \quad (21)$$

Počet nutných vzorků $M+1+L$ je větší než dostupný počet vzorků L . Abychom mohli koeficienty a_m vypočítat, musíme použít víc vztahů z následujících j -tých cyklů hlasivkového signálu. Využíváme předpokladu, že řečový signál je tvořen souvislým signálem a tím pádem jej můžeme považovat za stacionární náhodný proces. Interval $[n_{cj} - M - 1, n_{cj} - M, K, n_{cj} + L - 1]$ v j -tém cyklu bude na stejné relativní pozici jako interval $[n_{ci} - M - 1, n_{ci} - M, K, n_{ci} + L - 1]$. Můžeme potom vytvořit další výrazy pomocí

$$U'_{gfc_j} + P_{cj} A = p_{cj} + \varepsilon, \quad (22)$$

Kde P_{cj} je odečítán v intervalu $[n_{cj} - M - 1, n_{cj} - M, K, n_{cj} + L - 2]$, p_{cj} v intervalu $[n_{cj}, K, n_{cj} + L - 1]$ a ε je rozdíl mezi derivovaným průběhem hlasového pulzu mezi intervaly i a j . Vztah (22) bude po rozepsání vypadat takto

$$\begin{aligned}
& \begin{bmatrix} u'_{gf} \left(n_{ci} - \frac{M}{2} \right) \\ \vdots \\ u'_{gf} \left(n_{ci} - \frac{M}{2} + (L-1) \right) \end{bmatrix} + \begin{bmatrix} p(n_{cj}-1) & \cdots & p(n_{cj}-(M+1)) \\ \vdots & \ddots & \vdots \\ p(n_{cj}-1+(L-1)) & \cdots & p(n_{cj}-1+(L-1)-(M+1)) \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ \vdots \\ a_{M+1} \end{bmatrix} = \\
& = \begin{bmatrix} p(n_{cj}) \\ \vdots \\ p(n_{cj}+(L-1)) \end{bmatrix} + \begin{bmatrix} u'_{gf} \left(n_{ci} - \frac{M}{2} \right) - u'_{gf} \left(n_{cj} - \frac{M}{2} \right) \\ \vdots \\ u'_{gf} \left(n_{ci} - \frac{M}{2} + (L-1) \right) - u'_{gf} \left(n_{cj} - \frac{M}{2} + (L-1) \right) \end{bmatrix}. \quad (23)
\end{aligned}$$

Ze vztahů (20) a (22) dostaneme

$$\left(\underset{Q}{P_{4^j 2} - P_{4^i}} \right) A = \left(\underset{S}{p_{4^j 2} - p_{4^i}} \right) + \varepsilon. \quad (24)$$

Koeficienty VTF pak konečně vypočítáme pomocí

$$A = \left(Q^T Q \right)^{-1} Q^T S. \quad (25)$$

Z (24) a (25) je vidět že koeficienty A jsou ovlivněny chybou ε . Přesnost koeficientů A může být zvýšena průměrováním odhadů z různých částí řečového signálu.

U některých řečových signálů může dojít k tomu, že doba uzavření hlasivek je příliš krátká. Ve vztahu (24) pak nemusí být dostatečný počet vzorků signálu na vytvoření vektorů Q a S . V takovém případě může být vytvořeno více výrazů formálně shodných s (24), ale zkonstruovaných pro jiné intervaly hlasivkových pulzů. Postup ukazují vztahy (26), (27) a (28).

$$\left(\underset{R}{P_{4^j 2} - P_{4^i}} \right) A = \left(\underset{V}{p_{4^j 2} - p_{4^i}} \right) + \delta \quad (26)$$

$$\begin{bmatrix} Q \\ R \end{bmatrix} A = \begin{bmatrix} S \\ V \end{bmatrix} + \begin{bmatrix} \varepsilon \\ \delta \end{bmatrix} \quad (27)$$

$$A = \left(\Omega^T \Omega \right)^{-1} \Omega^T \Phi. \quad (28)$$

6.2.4 Určení částí řečového signálu odpovídajícího fázi uzavření hlasivek

V kapitole 6.2.2 byl uveden způsob jak získat signál $u'_{scf}(n - M/2 - \Delta)$. Předpokládáme, že v i -tém cyklu nabývá signál $u'_{scf}(n - M/2 - \Delta)$ maximální záporné hodnoty v okamžiku n_{0i} , následně se pozvolna nebo s kolísáním vrací k nule a po průchodu nulou v okamžiku $n_{0i} + N_c - 1$ již zůstává kladný. N_c označuje délku uzavření hlasivek. Stejný

okamžik lze v $u'_{scf}(n)$ vymeziť interval $[n_{0i} - M/2 - \Delta, K, n_{0i} - M/2 - \Delta + (N_c - 1)]$. Pro přehlednější zápis interval přepíšeme jako $[n_{close}, K, n_{close} + (N_c - 1)]$, kde $n_{close} = n_{0i} - M/2 - \Delta$ a označuje okamžik uzavření hlasivek.

Jako další si uvedeme, které hodnoty $p(n)$ lze použít pro konstrukci výrazů (22) - (25). Podle (19) v okamžiku $n = n_{close} + M/2 + (M + 1)$ je tlak vzduchu na rtech

$$p(n_{close} + M/2 + (M + 1)) = U'_{gf}(n_{close} + (M + 1)) + a_1 p(n_{close} + M/2 + (M + 1)) + L + a_{M+1} p(n_{close} + M/2).$$

V předchozím výrazu $p(n_{close} + M/2)$ vyjadřuje zvuk, který prochází hlasivkami v okamžiku n_{close} a ke rtům dospěje v okamžiku $n_{close} + M/2$. V okamžiku $n_{close} + M/2 + (N_c - 1)$ je tlak vzduchu na rtech

$$p(n_{close} + M/2 + (N_c - 1)) = u'_{gf}(n_{close} + (N_c - 1)) + a_1 p(n_{close} + M/2 + (N_c - 1) - 1) + L + a_{M+1} p(n_{close} + (-M/2 - 1) + (N_c - 1)).$$

Předchozí výraz $p(n_{close} + M/2 + (N_c - 1))$ obsahuje poslední okamžik kdy jsou hlasivky uzavřené $n_{close} + (N_c - 1)$. Pro vytvoření P_{ci} a p_{ci} se tedy použijí vzorky v intervalu $[n_{close} + M/2, K, n_{close} + M/2 + M, K, n_{close} + M/2 + (N_c - 1)]$.

Konkrétně pro $p_{ci} = [p(n_{close} + M/2 + (M + 1)), K, p(n_{close} + M/2 + (N_c - 1))]^T$. Délka vektoru p_{ci} je $L = N_c - M - 1$.

Pro přímé odečítání hodnot z $p_{mic}(n)$ přepočítáme pozice z $p(n)$ do $p_{mic}(n)$. Mezi těmito dvěma signály je vztah $p(n - \Delta) = p_{mic}(n)$. Protože $n_{close} = n_{0i} - M/2 - \Delta$, bude interval pro vytvoření P_{ci} a p_{ci} roven $[n_{0i}, K, n_{0i} + (N_c - 1)]$.

Konkrétně pak

$$p_{ci} = [p_{mic}(n_{0i} + (M + 1)), K, p(n_{0i} + (N_c - 1))]^T \quad (29)$$

a

$$P_{ci} = \begin{bmatrix} p_{mic}(n_{0i} - 1 + (M + 1)) & L & p_{mic}(n_{0i} - 1 + (M + 1) - (M + 1)) \\ & M & M \\ p_{mic}(n_{0i} - 1 + (N_c - 1)) & L & p_{mic}(n_{0i} - 1 + (N_c - 1) - (M + 1)) \end{bmatrix}. \quad (30)$$

Podobně pro vytvoření P_{cj} a p_{cj} použijeme interval $[n_{0j}, K, n_{0j} + (N_c - 1)]$, kde n_{0j} je okamžik, kdy $u'_{scf}(n - M/2 - \Delta)$ nabývá v j-tém cyklu největší záporné hodnoty.

Jak bylo ukázáno, lze segmenty v $p_{mic}(n)$ potřebné pro konstrukci jednotlivých vektorů určit podle průběhu $u'_{scf}(n - M/2 - \Delta)$ bez nutnosti znát vzdálenost mikrofonu od úst Δ .

6.2.5 Odhad hlasivkového pulzu

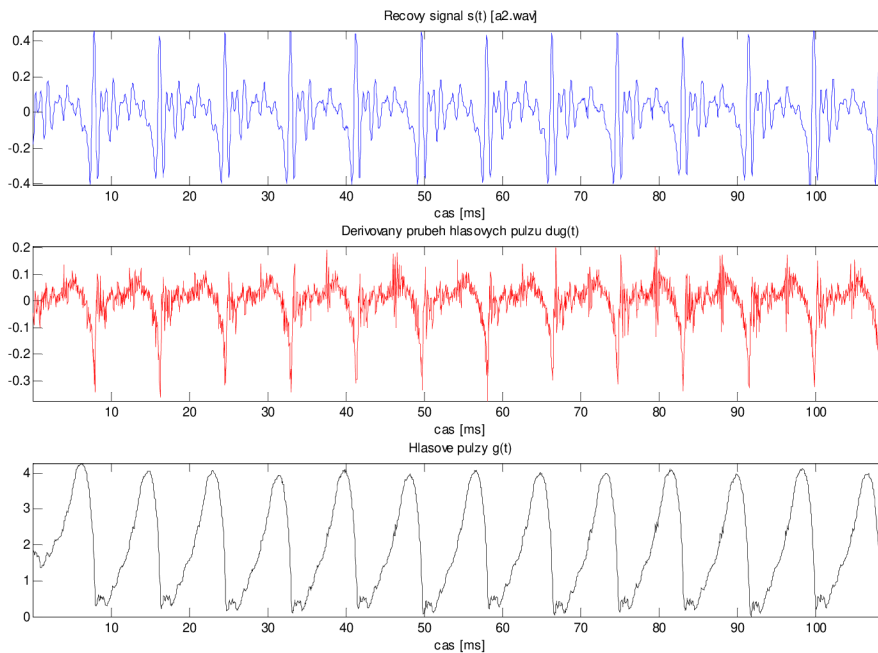
V kapitole 6.2.3 jsme získali koeficienty hlasového traktu. Použijeme je nyní pro inverzní filtraci signálu $p_{mic}(n)$ podle vzorce $\sum_{m=1}^{M+1} (1 - A^*(m)z^{-m})$. Tím získáme derivované průběhy hlasového pulzu $u'_g(n - M/2 - \Delta)$. Integrací pomocí $1/(1-z)^{-1}$ konečně získáme průběh hlasových pulzů $u_g(n - M/2 - \Delta)$.

6.3 Výstupy vybraných metod pro stanovení hlasivkových pulzů

Vybrané metody jsem použil pro získání odhadů průběhů hlasových pulzů pro hlásku „a“. Výstupy metod jsou uvedeny v dalších dvou kapitolách.

6.3.1 Výstup metody I A I F [4]

Na **Obrázek 6.5** je v prvním grafu vidět část řečového signálu „a“. Druhý graf zobrazuje derivovaný průběh hlasivkových pulzů a poslední graf přímo odhadnuté hlasové pulzy.



OBRÁZEK 6.5.: VÝSTUP METODY I A I F

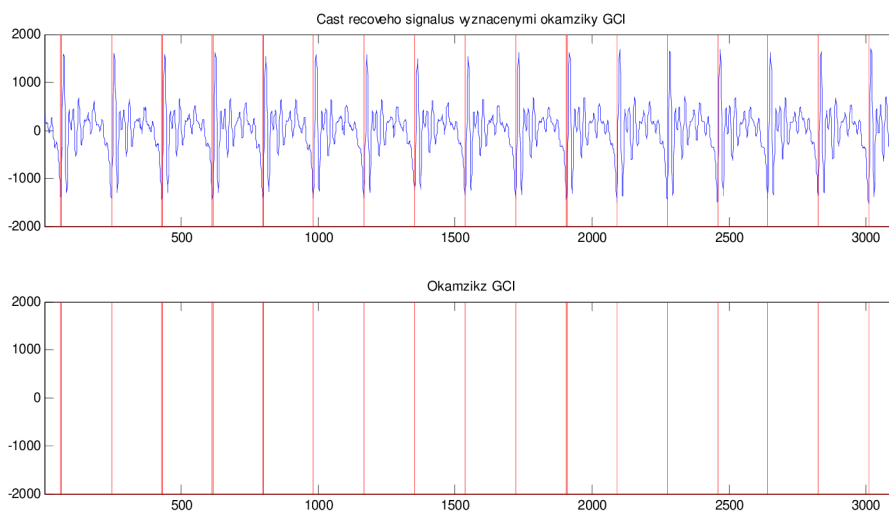
6.3.2 Výstup metody založené na průměrování parametrů hlasového traktu [9]

Při výpočtech se používá délka hlasového traktu L_{VT} . Každý člověk má přirozeně hlasový trakt jinak dlouhý. V práci [9] byl algoritmus předveden na hláске „a“. Pro tuto hlásku jsou také v dané práci uvedeny průměrné hodnoty délky hlasového traktu, které lze pro odhad hlasivkového impulsu také použít. Pro ostatní hlásky (e,i,o,u) jsem použil závěry práce [10]. V práci jsou použita anglická slova (např.: „head“ „e“) a nejedná se tedy o „čistokrevné“ hlásky. Výsledné délky VTF by ovšem měly být dostačující. Pokud bychom chtěli použít přesnější hodnoty, lze použít výpočet založený na lineární predikci a

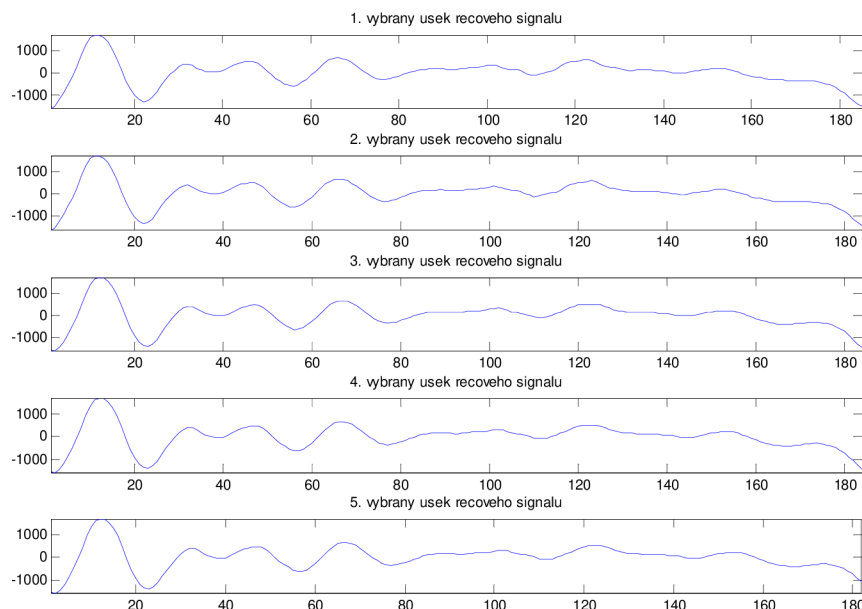
následném zjišťování poloh a šířek formantů, díky nimž lze délku hlasového traktu vypočítat. I tato metoda je v [9] a [10] popsána.

Protože je metoda založena na průměrování velkého množství odhadů parametrů VTF, je potřeba pro výpočet použít úsek znělého řečového signálu s téměř stálou amplitudou o délce alespoň 3 sekund (60 – 200 period signálu). Použil jsem řečový signál „a“.

Při výpočtu koeficientů A s jejichž pomocí se pak počítá d'_{scf} se v [9] hovoří o cyklech. Není však řečeno, kde jednotlivé cykly začínají a končí. Pro výběr cyklů (period) jsem použil vlnkovou transformaci popsanou v **Příloha B**, abych určil okamžiky GCI viz **Obrázek 6.6a** a **Obrázek 6.6b**

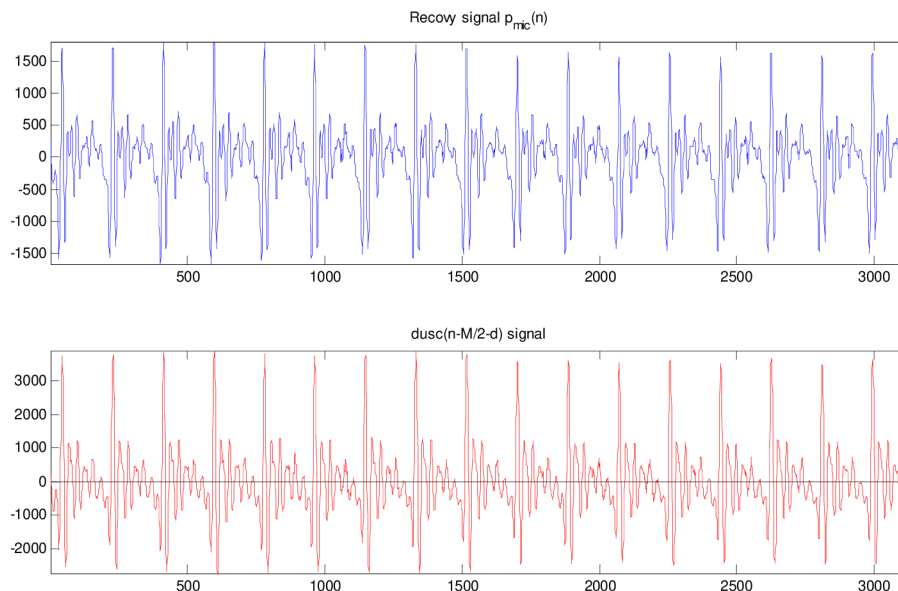


OBRÁZEK 6.6A.: URČENÍ GCI A ROZDĚLENÍ NA SEGMENTY



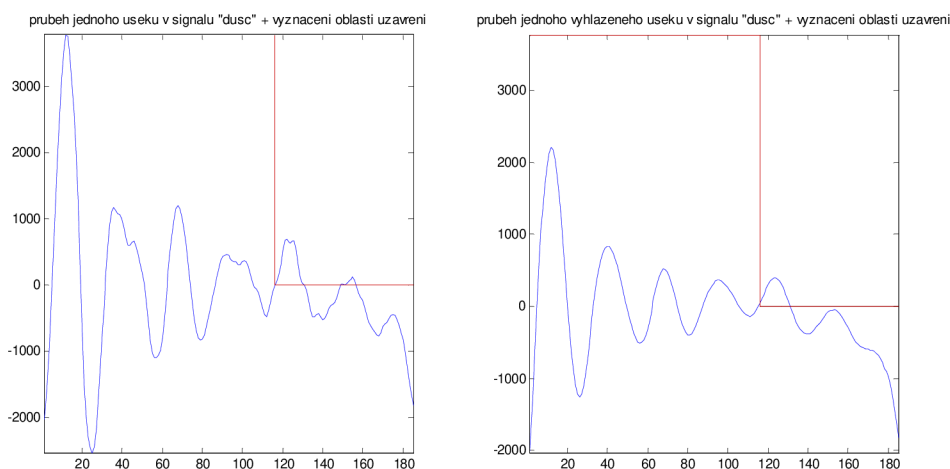
OBRÁZEK 6.6B.: ÚSEKY VYBRANÉ PRO VÝPOČET A

Takto určené GCI (okamžiky uzavření hlasivek) jsem pak považoval za začátky jednotlivých hlasivkových cyklů. Signál d'_{scf} se počítá pro určení GCI a uzavřené fáze hlasivek jak bylo uvedeno v kapitole (6.2.2). Výsledek výpočtu je na **Obrázek 6.7.**



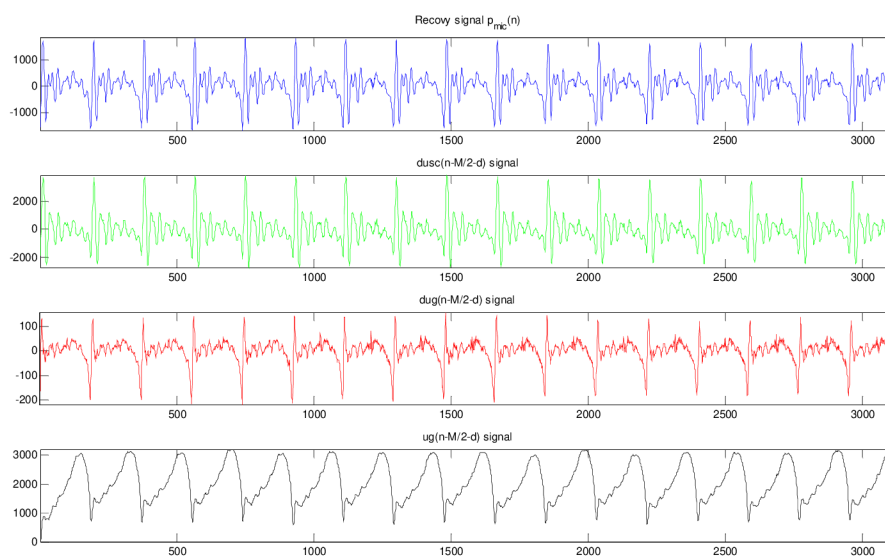
OBRÁZEK 6.7.: VYPOČITANÝ SIGNÁL DUSC

Často se fáze uzavření určí jako konstanta (např. 30% z délky jedné periody). Určil jsem dobu uzavření hlasivek právě pomocí této konstanty (30%). Na porovnání jsem také určil dobu uzavření metodou popsanou v [9]. Na **Obrázek 6.8.** je vidět pouze jedna (červená) křivka. Ta vyznačuje dobu uzavření určenou pomocí metody popsané v [9]. Tato křivka překryla původní (černou) křivku, která vyznačovala 30% z délky periody. Je tedy vidět, že můžeme použít jak metodu z [9] tak jednoduché určení doby uzavření jako 30% délky periody. Úsek signálu může být zarušený, proto je před určováním podle [9] vyhlazen (pravý graf).



OBRÁZEK 6.8.: DETEKCE FÁZE UZAVŘENÍ Z DUSC

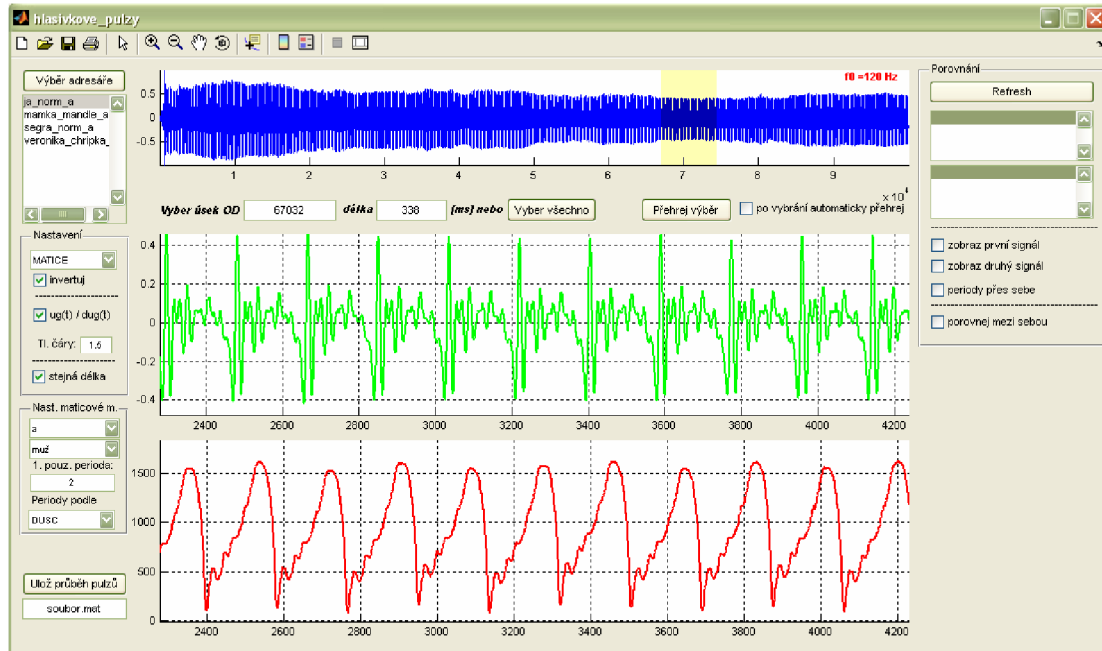
Pro každou periodu řečového signálu p_{mic} jsem vypočítal parametry inverzního filtru B. Nakonec jsem všechny parametry B zprůměroval a dostal tak zpřesněný odhad inverzního filtru. Tímto filtrem jsem filtroval řečový signál p_{mic} a dostal derivovaný průběh hlasivkového pulzu. Pomocí integrace jsem potom z derivovaného průběhu u'_g získal odhad hlasivkového pulzu jak je zobrazeno na **Obrázek 6.9**.



OBRÁZEK 6.9.: ODHAD HLASIVKOVÝCH PULZŮ

7. Pomůcka pro odhad hlasivkových pulzů z řečového signálu

Pro jednodušší manipulaci se skripty v matlabu jsem vytvořil grafické rozhraní, které usnadňuje základní úkony při výpočtu hlasivkových pulzů (hlasivkove_pulzy.m). Okno programu je vidět na **Obrázek 7.1**.



OBRAZEK 7.1.: UTILITA PRO ANALÝZU HLASIVKOVÝCH PULZŮ

7.1 Načtení zvukových souborů a nastavení metod

Nejdříve je potřeba načíst zvukový soubor do programu. Pomocí tlačítka „Výběr adresáře“ se zvolí adresář, který obsahuje zvuková data (wav soubory). Po jeho zvolení, se v seznamu pod tlačítkem objeví nalezené soubory. Po kliknutí na název souboru, se soubor zobrazí v prvním (nejvyšším) grafu. Pomocí tažení myši je možné z grafu vybrat kratší část, která se po dokončení výběru zobrazí v druhém grafu (zelený průběh). Nyní už jen zbývá zvolit (pod seznamem souborů) metodu výpočtu pulzů (IAIF[4] nebo MATICE[9]).

Výběr signálu z hlavního grafu lze „dostavovat“ pomocí dvou prvků („Vyber úsek od“ a „délka“). Vybraný úsek je také možno přehrát pomocí tlačítka „Přehrej výběr“.

Pro metodu IAIF není dobré vybírat příliš dlouhé úseky signálu, protože pak není splněn předpoklad, že parametry hlasového traktu se nemění. Úsek by neměl být delší než 100 ms [4].

Metoda MATICE naopak pro dobrý odhad potřebuje delší průběhy. Délka by měla být volena tak, aby zahrnovala alespoň 60 – 200 period řečového signálu [9] (metoda je založena na průměrování parametrů hlasového traktu). Testováním jsem zjistil, že je možné vybrat i kratší úseky k dosažení stejného výsledku. Je však nutné volit vhodnou část řečového signálu (podle předpokládaného výsledku, tvaru impulzů).

7.1.1. Nastavení IAIF metody

U této metody je nastavení poměrně snadné. Vysvětlím význam jednotlivých nastavovacích prvků.

- invertuj* – po vybrání (zatržení) je řečový signál před výpočtem invertován
- ug(t)/ dug(t)* – po vybrání (zatržení) se jako výstup zobrazí signál hlasivkových pulzů $ug(t)$. Když není volba vybrána, zobrazí se derivovaný průběh $dug(t)$
- Tl.čáry* – umožňuje změnit tloušťku čáry ve výsledných grafech (je třeba zadat číselnou hodnotu a případné desetinné čísla psát s „.“)
- stejná délka* – po vybrání (zatržení) bude vybraný úsek řečového signálu zkrácen, aby délkově odpovídal hlasivkovým pulzům (ty jsou „kratší“ v důsledku filtrování signálu při výpočtu)

7.1.2 Nastavení MATICE metody

Nastavení popsané u předchozí metody zůstává. Navíc se po vybrání metody MATICE zobrazí další blok nastavení, který teď vysvětlím.

- volba hlásky* – podle volby hlásky se použijí příslušné rozměry hlasového traktu [10]
- volba pohlaví* – délky hlasových traktů pro zvolenou hlásku jsou jiné u mužů a žen, proto je nutné ještě vybrat pohlaví
- 1 pouz.perioda* – lze zvolit začátek zpracovávání signálu (vynechá předchozí periody), není třeba nastavení měnit
- periody podle* – volba způsobu, kterým budou zjištěny počátky period (podle metody GCI (Příloha B) nebo DUSC [9])

7.2 Uložení a porovnání hlasivkových pulzů

Pokud chceme porovnat průběhy hlasivkových pulzů, například dvě stejná slova (jedno vyslovené ve stresové situaci, druhé „v klidu“), vybereme daný wav soubor, část signálu (pro lepší porovnání můžeme délku vybrané oblasti u dvou porovnávaných slov nastavit ručně viz prvek „délka“) a zobrazíme hlasivkové pulzy. Zobrazený pulz lze potom uložit pomocí tlačítka „Ulož průběh pulzů“.

Pokud máme průběhy pulzů uloženy a chceme je porovnávat, stiskneme nejdříve tlačítko „Refresh“. V seznamech pod tlačítkem se zobrazí nalezené průběhy hlasových pulzů. Vybereme si jaké pulzy chceme zkoumat (označíme je kliknutím myši) a zatržením „*zobraz první/ druhý signál*“ si zobrazíme průběhy pulzů. Průběhy jsou před zobrazením normovány na hodnotu 1 a „usazeny“ na nulovou hodnotu.

Zatržením volby „*periody přes sebe*“ se zobrazí hlasivkové pulzy vykreslené přes sebe. Před vykreslením se pulzy pomocí metody borcení časové osy (Dynamic Time Warping) upraví na stejnou délku periody [16].

Zatržením volby „*porovnej mezi sebou*“ se oba signály vykreslí do jednoho grafu.

8. Stanovení vhodných kritérií pro určení vybraných diagnóz

Utilita popsaná v kapitole 7 má spíše demonstrativní charakter. To znamená, že s její pomocí lze lehce získat odhad hlasivkových pulzů pomocí dvou naprogramovaných metod. Výsledné pulzy je možné uložit do MAT souborů. Při analyzování většího počtu dat (řečových souborů), by práce s tímto programem nebyla příliš efektivní. Vytvořil jsem tedy v prostředí Matlab další 3 programy, tak aby vybírání úseků signálu, vytváření vzorků a analýza byly co nejméně namáhavé na „manuální“ práci. Tyto programy se nevyznačují složitým ovládáním a nebudu je zde tedy podrobně popisovat. Jsou přiloženy na doprovodném CD a jejich zdrojový kód je komentován.

8.1 Onemocnění hlasivek

Vybral jsem tři onemocnění hlasivek a to „Reinek's edema“, „Vocal fold edema“ a „vocal tremor“. Pro každý druh onemocnění jsem měl k dispozici řečové signály od alespoň desíti mluvčích o délce jedné sekundy.

Pro analýzu impulzů pomocí metody MATICE[9] jsem zvolil následující postup:

- u všech zkoumaných signálů jsem určil základní tón a vypočítal jeho průměrnou hodnotu, všechny signály jsem převzorkoval na vzorkovací frekvenci 22050 Hz a invertoval
- ze zkoumaného signálu jsem vybral 5 různých úseků (délky 30 period, kvůli příliš krátkému signálu)
- z každého úseku jsem vypočítal odhad hlasivkových pulzů
- z každého odhadu jsem vybral jeden pulz a tím získal pro každý řečový signál skupinu 5-ti pulzů
- z průměrné základní frekvence jsem vypočítal periodu a první z pěti hlasivkových pulzů upravil na stejnou délku
- takto upravenou první periodu jsem potom použil jako referenční průběh pro metodu borcení časové osy, pomocí které jsem upravil délku zbývajících 4 pulzů
- výslednou pěti pulzů jsem potom vykreslil do jednoho grafu

Pro analýzu impulzů pomocí metody IAIF[4] jsem zvolil následující postup:

- u všech zkoumaných signálů jsem určil základní tón a vypočítal jeho průměrnou hodnotu, všechny signály jsem převzorkoval na vzorkovací frekvenci 22050 Hz a invertoval
- ze zkoumaného signálu jsem vybral 5 různých úseků (o délce 90 ms)
- z každého úseku jsem vypočítal odhad hlasivkových pulzů
- z každého odhadu jsem vybral jeden pulz a tím získal pro každý řečový signál skupinu 5-ti pulzů
- z průměrné základní frekvence jsem vypočítal periodu a první z pěti hlasivkových pulzů upravil na stejnou délku
- takto upravenou první periodu jsem potom použil jako referenční průběh pro metodu borcení časové osy, pomocí které jsem upravil délku zbývajících 4 pulzů
- výslednou pěti pulzů jsem potom vykreslil do jednoho grafu

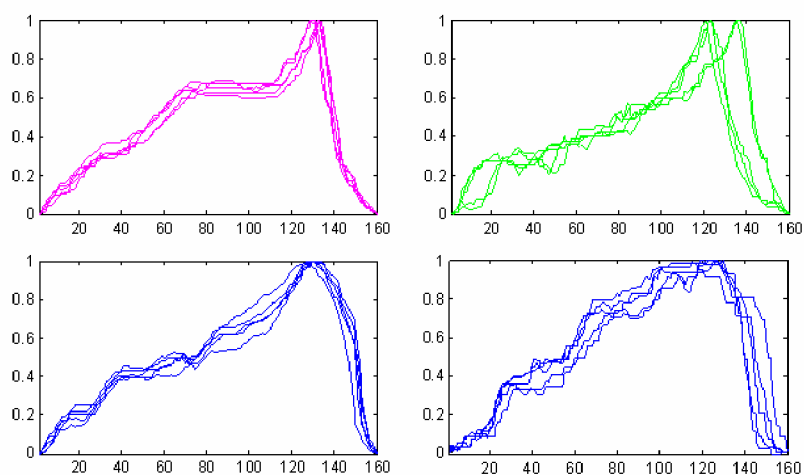
Průběhy pulzů, které jsem získal jednotlivými metodami se od sebe tvarově příliš nelišily. Protože signál byl dlouhý pouze 1s, nebylo možné vybrat pro metodu MATICE pět různých úseků, ale většinou jen dva dostatečně dlouhé úseky. Metoda IAIF se spokojí s kratšími úseky vstupních dat, proto nebyl problém získat vždy 5 různých úseků. Jak už jsem uvedl, tvarově se pulzy získané různými metodami téměř nelišily. Při analýze jsem proto použil pulzy získané metodou IAIF, protože bylo vždy k dispozici 5 různých úseků signálu.

Jak bude dále ukázáno, průběhy od jednotlivých mluvčích se dosti liší. I když si všechny průběhy nejsou podobné, dají se rozdělit do skupin podle podobnosti. Pro „rozumnou“

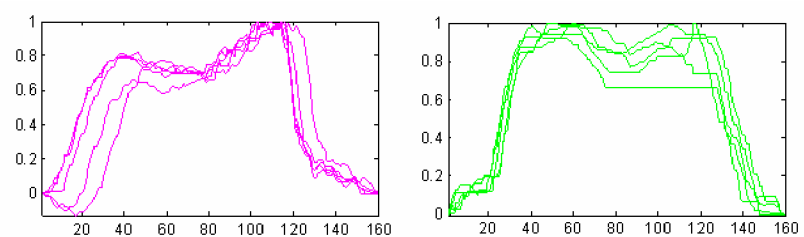
analýzu by bylo nutné získat rozsáhlejší databázi patologických hlasů, aby byly jednotlivé skupiny zastoupeny větším počtem signálů a bylo možné je nějakým způsobem statisticky vyhodnotit. Protože jsem měl k dispozici pouze omezený počet patologických signálů, rozčlenil jsem jednotlivé mluvčí do skupin a další analýzu jsem neprováděl.

8.1.1 Reinek's edema (Reinekův edém)

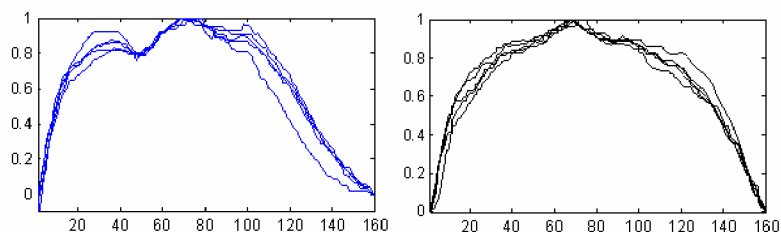
Měl jsem k dispozici řečové signály od 15 různých mluvčích ženského pohlaví. Pomocí postupu popsaného výše jsem vytvořil odhad hlasivkových pulzů. Zjistil jsem, že pulzy různých mluvčích se mezi sebou tvarově dost odlišují a nebude tedy možné mezi všemi najít společnou vazbu či parametr, který by je popisoval. Nabízelo se pulzy rozdělit do skupin podle podobnosti. Po rozdělení vznikly 4 různé skupiny. Pět průběhů se nedalo zařadit ani do jedné ze skupin a proto jsem je vyřadil. Protože v každé skupině se nacházelo jen několik průběhů, nemá smysl zabývat se hledáním společných parametrů, protože výsledek by byl pravděpodobně nerepresentativní. Na **Obrázek 8.1**, **Obrázek 8.2**, **Obrázek 8.3** a **Obrázek 8.4** jsou ukázány průběhy čtyř vytvořených skupin.



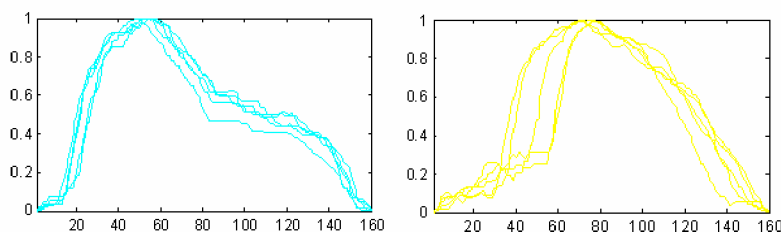
OBRÁZEK 8.1.: REINEK'S EDEMA, SKUPINA A (ČTYŘI MLUVČÍ)



OBRÁZEK 8.2.: REINEK'S EDEMA, SKUPINA B (DVA MLUVČÍ)



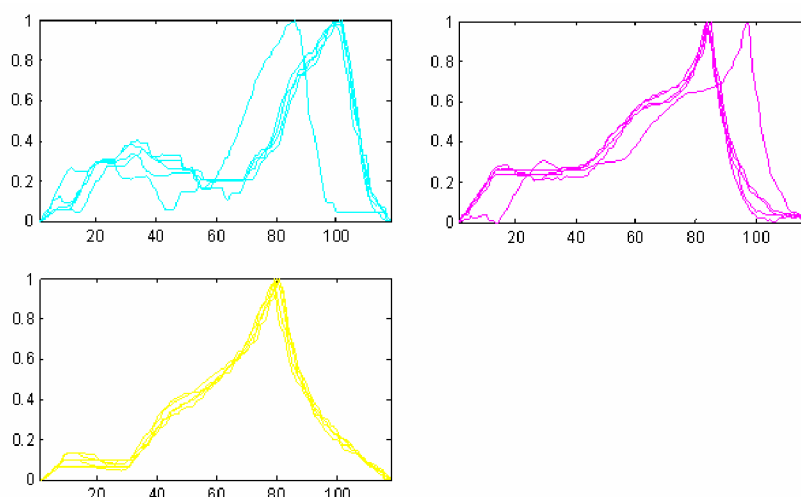
OBRÁZEK 8.3.: REINEK'S EDEMA, SKUPI NA C (DVA MLUVČÍ)



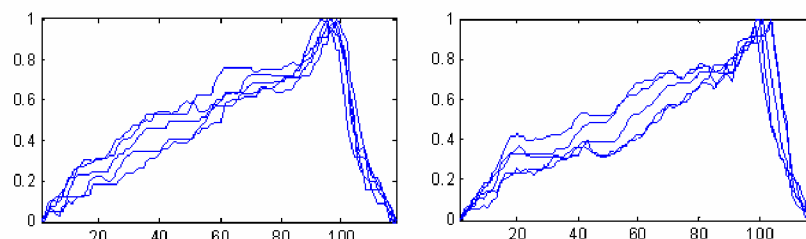
OBRÁZEK 8.4.: REINEK'S EDEMA, SKUPI NA D (DVA MLUVČÍ)

8.1.2 Vocal fold edema (Otok hlasivkových svalů)

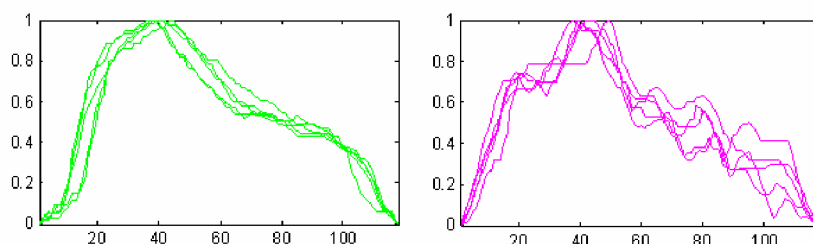
Měl jsem k dispozici řečové signály od 15 různých mluvčích ženského pohlaví. Pomocí postupu popsaného výše jsem vytvořil odhad hlasivkových pulzů. Opět jsem zjistil, že pulzy různých mluvčích se mezi sebou tvarově dost odlišují a nebude tedy možné mezi všemi najít společnou vazbu či parametr, který by je popisoval. Rozdělil jsem je tedy do skupin podle podobnosti. Po rozdělení vzniklo 5 různých skupin. Čtyři průběhy se nedaly zařadit ani do jedné ze skupin a proto jsem je vyřadil. Protože v každé skupině se nacházelo jen několik průběhů, nemá smysl zabývat se hledáním společných parametrů, protože výsledek by byl pravděpodobně nereprezentativní. Na **Obrázek 8.5**, **Obrázek 8.6**, **Obrázek 8.7**, **Obrázek 8.8** a **Obrázek 8.9** jsou ukázány průběhy pěti vytvořených skupin.



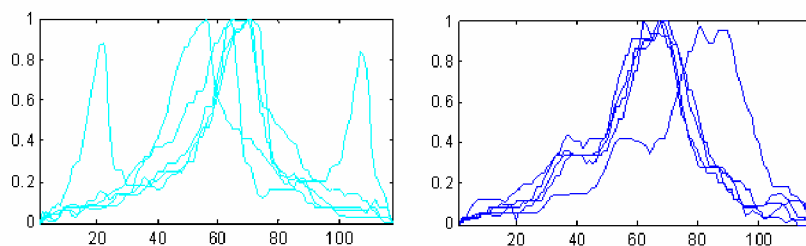
OBRÁZEK 8.5.: VOCAL FOLD EDEMA, SKUPI NA A (TŘI MLUVČÍ)



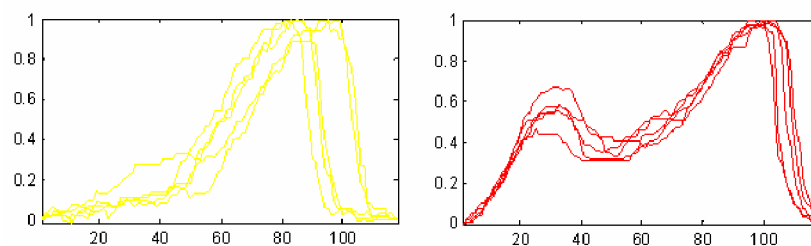
OBRÁZEK 8.6.: VOCAL FOLD EDEMA, SKUPINA B (DVA MLUVČÍ)



OBRÁZEK 8.7.: VOCAL FOLD EDEMA, SKUPINA C (DVA MLUVČÍ)



OBRÁZEK 8.8.: VOCAL FOLD EDEMA, SKUPINA D (DVA MLUVČÍ)

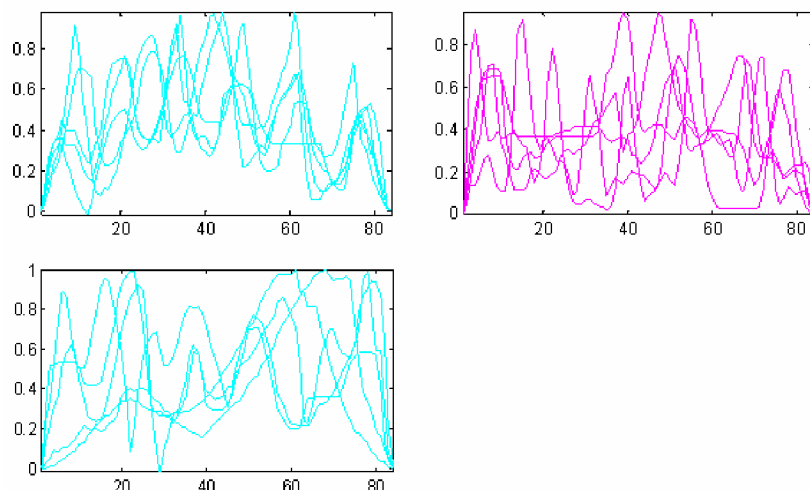


OBRÁZEK 8.9.: VOCAL FOLD EDEMA, SKUPINA E (DVA MLUVČÍ)

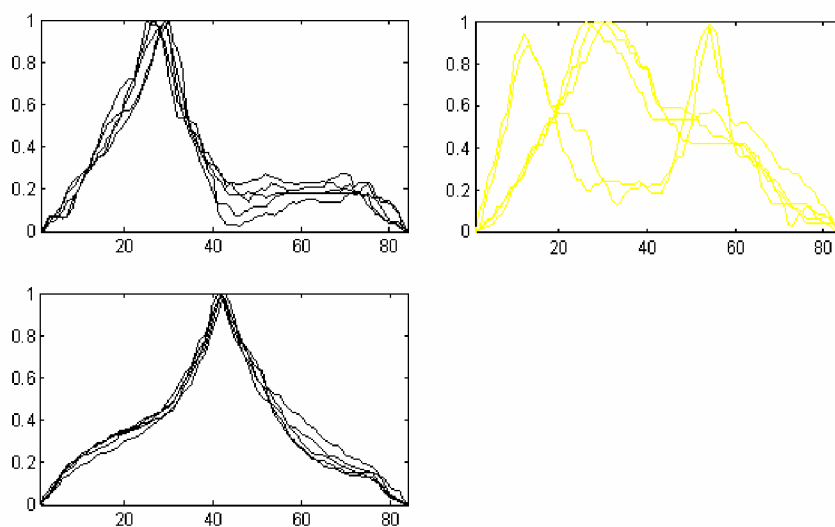
8.1.3 Vocal tremor (Třes hlasu)

Měl jsem k dispozici řečové signály od 11 různých mluvčích ženského pohlaví. Pomocí postupu popsaného výše jsem vytvořil odhad hlasivkových pulzů. Opět jsem zjistil, že pulzy různých mluvčích se mezi sebou tvarově dost odlišují a nebude tedy možné mezi všemi najít společnou vazbu či parametr, který by je popisoval. Rozděлил jsem je tedy do skupin podle podobnosti. Po rozdělení vznikly 3 různé skupiny. Tři průběhy se nedaly

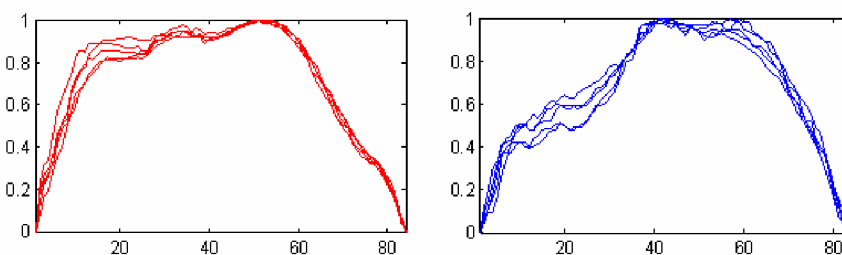
zařadit ani do jedné ze skupin a proto jsem je vyřadil. Protože v každé skupině se nacházelo jen několik průběhů, nemá smysl zabývat se hledáním společných parametrů, protože výsledek by byl pravděpodobně nerepresentativní. Na **Obrázek 8.10**, **Obrázek 8.11** a **Obrázek 8.12** jsou ukázány průběhy tří vytvořených skupin.



OBRÁZEK 8.10.: VOCAL TREMOR, SKUPINA A (TŘI MLUVČÍ)



OBRÁZEK 8.11.: VOCAL TREMOR, SKUPINA B (TŘI MLUVČÍ)



OBRÁZEK 8.12.: VOCAL TREMOR, SKUPINA C (DVA MLUVČÍ)

8.2 Vliv stresu

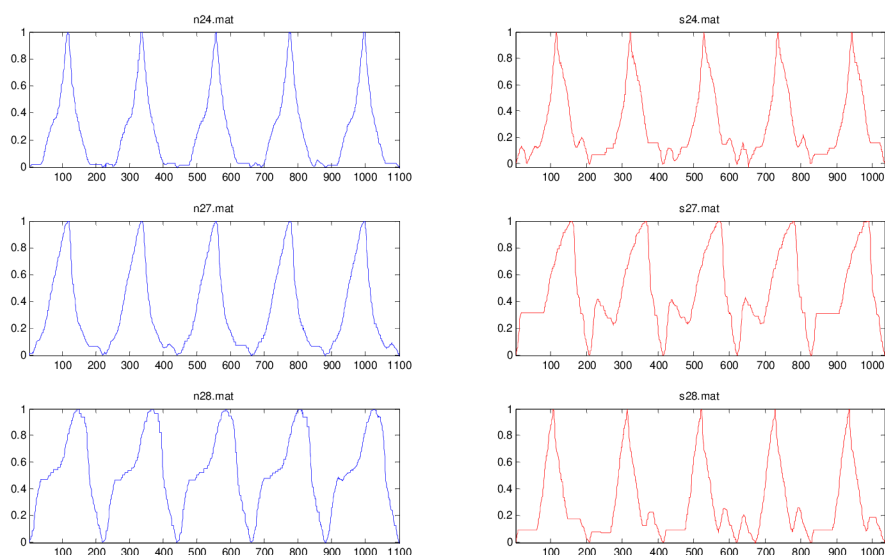
Pro analýzu vlivu stresu na tvar hlasivkových pulzů jsem měl k dispozici 9 mluvčích mužského pohlaví z databáze „ExamStress“ [17], která vzniká postupně na VUT v Brně a obsahuje záznamy obhajob diplomových a disertačních prací některých studentů.

Pro každého mluvčího jsou k dispozici dva řečové signály – stresový a normální. Ze stresového řečového signálu jsem vybral 5 různých úseků hlásky „a“ a k nim odpovídající úseky hlásky „a“ ze normálního řečového signálu. Protože vybrané úseky byli velmi krátké, nebylo by vhodné použít metodu MATICE. Pro analýzu jsem tedy použil jen pulzy získané metodou IAIF.

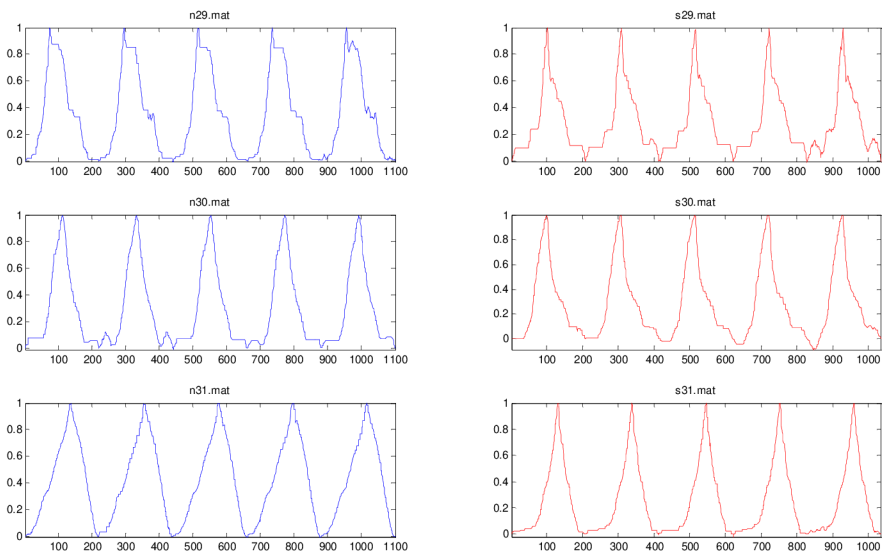
Pro analýzu impulzů jsem zvolil následující postup:

- u všech zkoumaných signálů (řečníků) jsem určil základní tón a vypočítal jeho průměrnou hodnotu, všechny signály jsem převzorkoval na vzorkovací frekvenci 22050 Hz a invertoval
- z každého úseku jsem vypočítal odhad hlasivkových pulzů
- z každého odhadu jsem vybral jeden pulz a tím získal pro každý řečový signál skupinu 5-ti pulzů
- z průměrné základní frekvence jsem vypočítal periodu a první z pětice (pro každého mluvčího jsem měl k dispozici pět různých pulzů) hlasivkových pulzů upravil na stejnou délku
- takto upravenou první periodu jsem potom použil jako referenční průběh pro metodu borcení časové osy, pomocí které jsem upravil délku zbývajících 4 pulzů

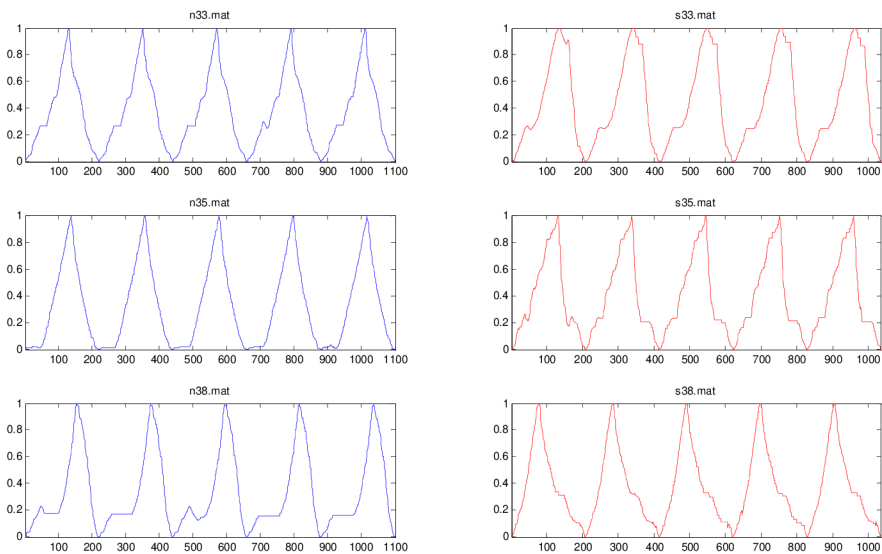
Na **Obrázek 8.13**, **Obrázek 8.14** a **Obrázek 8.15** jsou vidět průběhy hlasivkových pulzů pro všech 9 mluvčích. Modré průběhy představují normální hlasivkové pulzy a červené hlasivkové pulzy ovlivněné stresem.



OBRÁZEK 8.13.: Vliv stresu na tvar hlasivkových pulzů 1



OBRAZEK 8.14.: VLIV STRESU NA TVAR HLASIVKOVÝCH PULZŮ 2



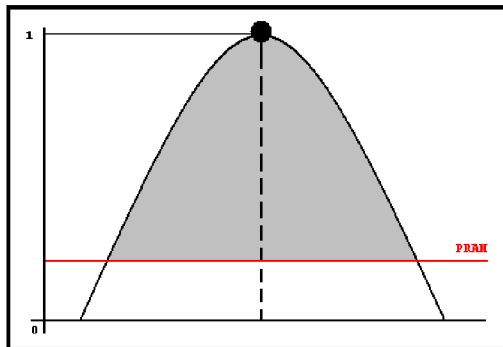
OBRAZEK 8.15.: VLIV STRESU NA TVAR HLASIVKOVÝCH PULZŮ 3

8.2.1 Monotónnost pulzů

Pokusil jsem se identifikovat stresový stav pomocí jednoduché analýzy monotónnosti hlasivkových pulzů. Parametrem, který jsem sledoval, byl počet průchodů nulou derivovaného průběhu pulzů.

Postupoval jsem takto:

- zvolil jsem si prah (například 10% z maximální úrovně signálu) podle kterého jsem vybral část pulzu (šedá oblast) viz **Obrázek 8.16**



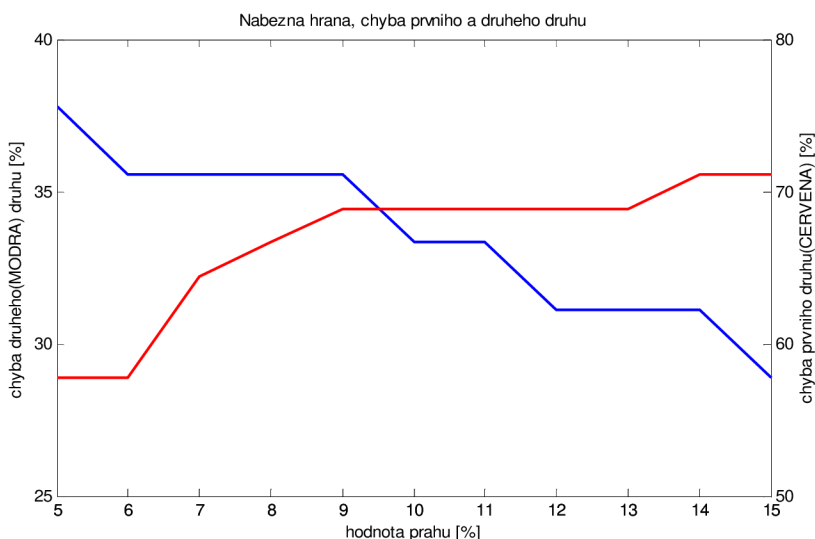
OBRÁZEK 8.16.: VÝBĚR ČÁSTI PULZU POMOCÍ PRAHU

- vybral náběžnou a sestupnou hranu
- každou hranu derivoval a určil počet průchodů nulou (počet změn směru)

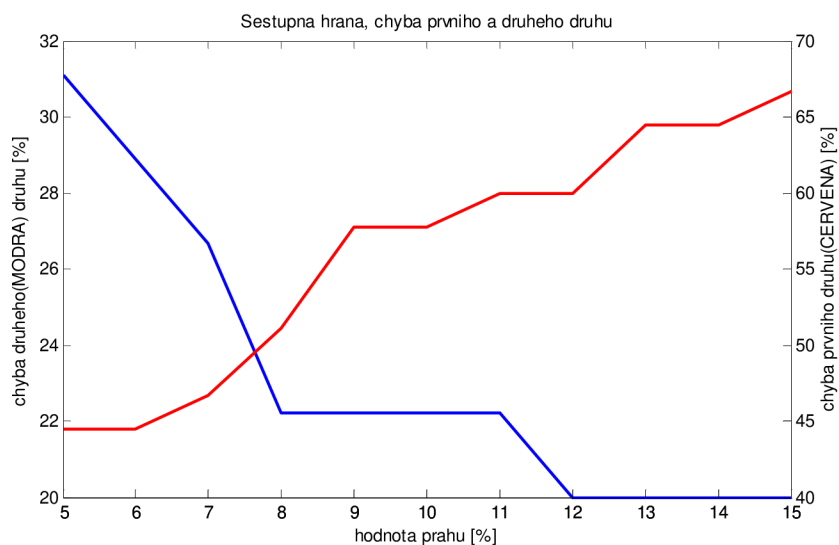
Pro normální signály (neovlivněné stresem) by měla být změna směru nulová – funkce (pulz) buď monotónně stoupá nebo klesá. V pulzu ovlivněném stresem dochází v průběhu impulsu ke změnám směru – počet průchodů nulou nebude nulový.

Pro nalezení prahu jsem vypočítal chybu prvního (stresový pulz je identifikován jako normální) a druhého (normální pulz je identifikován jako stresový) druhu a vybral prah tak, aby obě chyby byly co nejmenší.

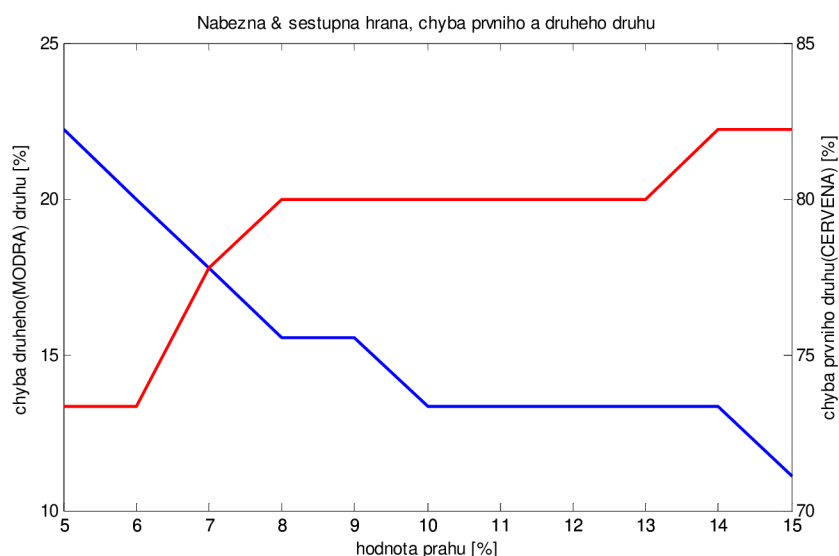
Zkoumal jsem zvlášť náběžnou a sestupnou hranu zvlášť a poté náběžnou a sestupnou hranu současně. Grafy chybovosti jsou uvedeny na obrázcích **Obrázek 8.17**, **Obrázek 8.18** a **Obrázek 8.19**.



OBRÁZEK 8.17.: CHYBA URČENÍ STRESU PODLE NÁBĚŽNÉ HRANY



OBRÁZEK 8.18.: CHYBA URČENÍ STRESU PODLE SESTUPNÉ HRANY

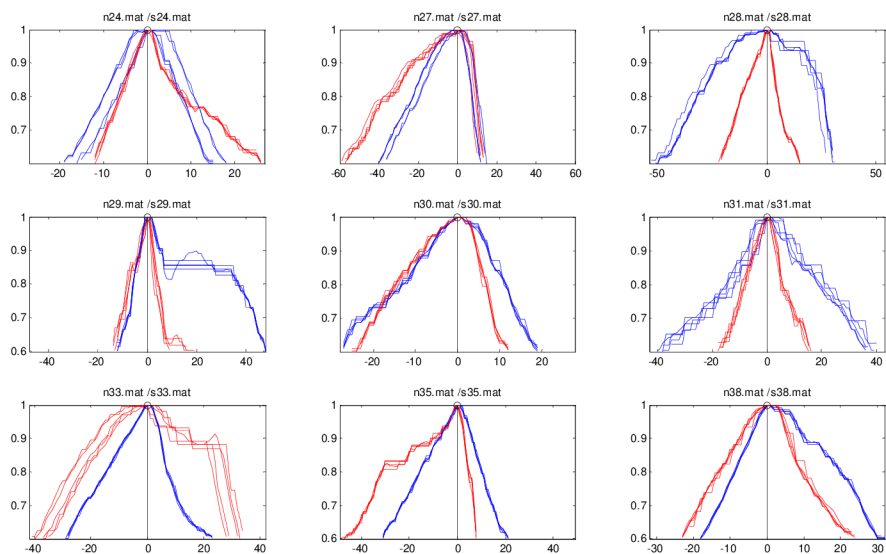


OBRÁZEK 8.19.: CHYBA URČENÍ STRESU PODLE NÁBĚŽNÉ A SESTUPNÉ HRANY

Z obrázků je vidět, že nejmenší chyby lze dosáhnout při sledování sestupné hrany a při prahu 8%. Chyba prvního druhu je potom 50% a chyba druhého druhu asi 21%.

8.2.2 Porovnání velikosti ploch

Dalším způsobem, jak určovat, zda je člověk ve stresu by mohlo být porovnání plochy stresového a normálního pulzu. Systém by se samozřejmě musel nejprve natrénovat na daného člověka. Po natrénování by už byl schopen stresový stav od normálního rozeznat. Na **Obrázek 8.20** jsou zobrazeny normální a stresové pulzy (červená barva) pro všech 9 mluvčích. Pulzy se protínají ve společném bodě (maximu) a jsou z intervalu < 60% max, max > postup je stejný jako postup vysvětlovaný na **Obrázek 8.16**.



OBRAZEK 8.20.: PLOCHY HLASIVKOVÝCH PULZŮ

Pokud si budeme všimát jen pravé poloviny plochy (od 0 hodnoty vpravo) zjistíme, že téměř ve všech případech je plocha pulzu ve stresu menší než plocha normálního pulzu. Ve třech případech tomu tak není. Pak je nutné porovnávat levou plochu. Pokud levá plocha zkoumaného pulzu bude větší než levá plocha normálního impulzu, bylo by rozhodnuto, že zkoumaný pulz je stresový. U jednoho z devíti mluvčích (mluvčí 24, první graf) by však tato metoda (jak je vidět z grafů) selhala.

9. Závěr

Seznámil jsem se s metodami analýzy řečového signálu pro diagnostické účely. V současné době probíhají výzkumy zaměřující se na získávání odhadu hlasivkových pulzů z řečového signálu. Ukázalo se totiž, že se mnoho diagnóz (fyzických i psychických) projevuje v hlase. Lidským uchem tyto změny nejsou rozpoznatelné, pomocí výpočetní techniky (matematicky) však je možné v řečovém signálu tyto jevy sledovat. Výzkumy se například zaměřují na diagnostikování rakoviny hlasivek (v ranném stádiu), Parkinsonovu chorobu nebo určování emočních stavů mluvčích (stres, strážlivost, únava atd.). Seznam informačních zdrojů, které se mi podařilo objevit je uveden v kapitole 4.

Vybral jsem dvě metody[4][9], kterými lze z řečového signálu získat odhad průběhu hlasivkových pulzů. Obě programované metody mají za cíl získat co nejlepší odhad parametrů hlasového traktu, který je v modelu tvorby řeči popisován all-pole filtrem, a následnou inverzní filtrací řečového signálu získat průběhy hlasivkových pulzů. Hlasivkové pulzy představují zdroj periodického budícího signálu, proto je možné je získat pouze ze znělých úseků řečového signálu viz kapitola 2. Obě metody jsem realizoval v systému Matlab a pro snadnější použití jsem vytvořil jednoduché grafické rozhraní. Ukázky odhadů hlasivkových pulzů pro hlásku „a“ jsou uvedeny v kapitole 6.3.

Z hlasivkových pulzů, které jsem naprogramovanými metodami obdržel, jsem se pokusil získat vhodná kritéria, kterými by bylo možné vybrané diagnózy identifikovat. Nejdříve jsem se zaměřil na tři onemocnění a to „Reinek's edema“, „Vocal fold edema“ a „Vocal tremor“. Ke každému druhu diagnózy jsem měl k dispozici 11 až 15 řečových signálů od různých mluvčích (ženy). Po vypočítání hlasivkových pulzů, jejich normování a přizpůsobení na stejnou délku periody jsem zjistil, že průběhy pulzů (pro jeden druh diagnózy) různorodé a nebude tedy možné je společně popsat jedním pravidlem. Ve skupině pulzů pro každou diagnózu bylo možné objevit 3 až 4 podskupiny, kam se potom pulzy od jednotlivých mluvčích dali zařadit. Hlasivkové pulzy, které se do žádné skupiny zařadit nedaly, byly vyřazeny. Po roztřídění do skupin se v každé skupině vyskytovali většinou signály od dvou různých mluvčích. Aby případná parametrizace těchto průběhů byla objektivní, bylo by potřeba mnohem více představitelů (mluvčích) v každé podskupině. Neměl jsem bohužel k dispozici více řečových signálů se stejnou diagnózou a proto jsem signály v rámci daných diagnóz pouze roztřídil do podskupin viz kapitola 8.1.

O hledání společných parametrů jsem se také pokusil pro stresový signál viz kapitola 8.2. K dispozici jsem měl 9 řečových signálů (muži) stresových a k nim odpovídajících signálů normálních. Tyto řečové signály pocházely z databáze ExamStress[17]. V řečových signálech jsem vyhledal 5 úseků hlásky „a“, které byly dostatečně dlouhé pro výpočet hlasivkových pulzů. Získané průběhy jsem nejdříve popsal z hlediska monotónnosti. Konkrétně jsem se zajímal o počet změn směru náběžné a sestupné hrany jednotlivých hlasivkových pulzů. Neanalyzoval jsem však celou délku hran, ale zvolil jsem vhodný práh (minimum), kterého jsem analýzu prováděl. Předpokládal jsem, že hrany signálu, který není ovlivněn stresem, stoupají/klesají monotónně a tedy, že nedochází k žádným výrazným změnám směru. S tímto předpokladem jsem provedl výpočet chyb prvního a druhého typu, které můžou při určování stresu vzniknout, pro různé hodnoty prahu. Nejvhodnějším prahem se jeví být hodnota 8% z maximální hodnoty. Při tomto prahu při zkoumání počtu změn sestupné hrany je chyba prvního druhu 50% a chyba druhého druhu asi 21% **Obrázek 8.18**.

U stejných pulzů jsem potom provedl porovnání na základě velikosti ploch zabraných jednotlivými pulzy **Obrázek 8.20**. Z obrázku je patrné, že při porovnávání ploch vpravo od „0“ je plocha stresového hlasivkového pulzu téměř vždy menší. Ve třech případech, kdy je plocha stresového pulzu větší nebo srovnatelná s plochou normálního impulzu je možné porovnáním levých polovin dva z nich opět spolehlivě označit za stresové (stresový pulz na levé půlce zabírá totiž větší plochu než normální pulz). U jednoho průběhu však rozpoznání selže.

Obě metody popsané v kapitole 8.2 (test monotónnosti hran a porovnání velikosti ploch pulzů) jsou i když s docela velkou chybovostí funkční. Pro zvýšení přesnosti metod by bylo zapotřebí analyzovat řečové signály od více mluvčích a vytvořit další zpřesňující kritéria díky nimž by rozpoznání stresového signálu bylo přesnější.

Porovnání velikostí ploch by se samozřejmě dalo použít pouze v takovém systému, který by se nejdříve natrénovával na daného mluvčího. Tento přístup by se dal využít například v aplikacích pro mobilní telefony, kdy by si telefon samovolně zaznamenával parametry normálních pulzů (například v průběhu hovoru). Při požadavku na určení stresu by potom díky těmto parametrům stres mohl rozpoznat.

Pro určení parametrů popisujících vybrané diagnózy (kapitola 8.1) by bylo opět nutné větší množství řečových signálů od různých mluvčích. Jednotlivé diagnózy by se potom daly popsat statistickými veličinami.

Příloha A: Popis algoritmu DAP[5],[7]

Tabulka A popisuje algoritmus výpočtu DAP metody pro odhad parametru hlasového traktu.

Tabulka A: Algoritmus DAP

<p>1) výpočet spektra řečového signálu pomocí FFT, zjištění formantů a k nim odpovídajících hodnot výkonového spektra $P(\omega)$</p> <p>2) výpočet $R(i) = \frac{1}{N} \sum_{m=1}^N P(\omega_m) e^{j\omega_m i}$, kde N je počet nalezených formantů</p> <p>3) pomocí LP určit první odhad koeficientů a, stupeň predikce viz Obrázek 6.2</p> <p>4) výpočet $A(\omega_m) = \sum_{k=0}^p a_k e^{-j\omega_m k}$ pro $m \in \langle 1, N \rangle$</p> <p>5) výpočet $\hat{h}(-i) = \frac{1}{N} \sum_{m=1}^N \frac{e^{-j\omega_m i}}{A(\omega_m)}$ pro $i \in \langle 0, p \rangle$</p> <p>6) výpočet nových parametrů a_{m+1} pomocí</p> $\hat{h}(-i) = \frac{1}{N} \sum_{m=1}^N \frac{e^{-j\omega_m i}}{A(\omega_m)} \quad a_{m+1} = a_m (1 - \alpha) + \alpha R^{-1} \hat{h}_m \quad \text{pro } \alpha = 0.5$ <p>7) výpočet chyby $E_{IS} = \frac{1}{N} \sum_{m=1}^N \frac{P(\omega_m)}{\hat{P}(\omega_m)} - \ln \frac{P(\omega_m)}{\hat{P}(\omega_m)} - 1$ s $\hat{P}(\omega_m) = \frac{1}{ A(\omega) ^2} = \frac{1}{\left \sum_{k=0}^p a_k e^{-j\omega_m k} \right ^2}$,</p> <p>za koeficienty a při výpočtu $\hat{P}(\omega_m)$ se dosadí výsledek z bodu 6). Pokud je chyba E_{IS} větší než zvolený práh (např. 0.001) pokračuje se s výpočtem od bodu 4).</p>

Příloha B: Určení okamžiků GCI z řečového signálu

Tímto tématem se zabývá [11] a [12]. Podstata spočívá v rozložení řečového signálu pomocí speciální vlnky. Kombinací výsledků potom získáme tzv. multiscale signál, z něhož je už možné rozpoznat okamžiky GCI (Glottal Closure Instants – okamžiky uzavření hlasivek). Mateřská vlnka je popsána následujícím vztahem (gausíán)

$$g(t) = -\cos(2 \cdot \pi \cdot f_0 \cdot t) \cdot e^{\left(\frac{-t^2}{2\tau^2}\right)} \quad (31)$$

kde $f_0 = \frac{f_e}{2}$, $\tau = \frac{1}{2 \cdot f_0}$ a f_e je frekvence, jakou je řečový signál vzorkován.

Transformovaný signál se potom získá použitím vzorce

$$y(t) = s(t) * g\left(\frac{t}{s_i}\right) \quad (32)$$

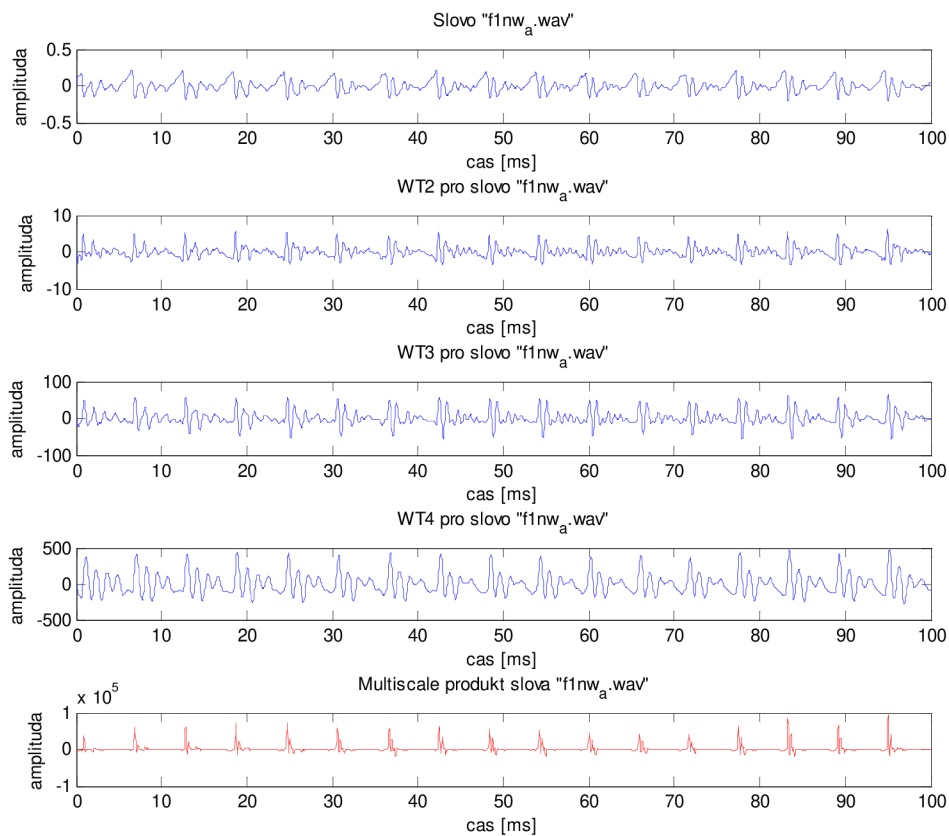
kde $s_i = 2^i$ určuje kompresi/expenzi mateřské vlnky. Pro detekci GCI nejvíce poslouží vlnky 2,3 a 4. Hodnota i tedy bude v intervalu $i = \{2,3,4\}$. Pozn.: * ve vzorci (32) značí konvoluci.

Pro získání multiscale signálu nakonec použijeme tento vzorec

$$p(n) = \prod_i w_{s_i} * s(n). \quad (33)$$

Pozici GCI v řečovém signálu určují lokální minima po levé straně maxima v $p(n)$ viz **Obrázek B.2**.

Na **Obrázek B.1** je úsek hlásky „a“, k tomuto úseku vypočítány jednotlivé transformace a nakonec multiscale produkt.

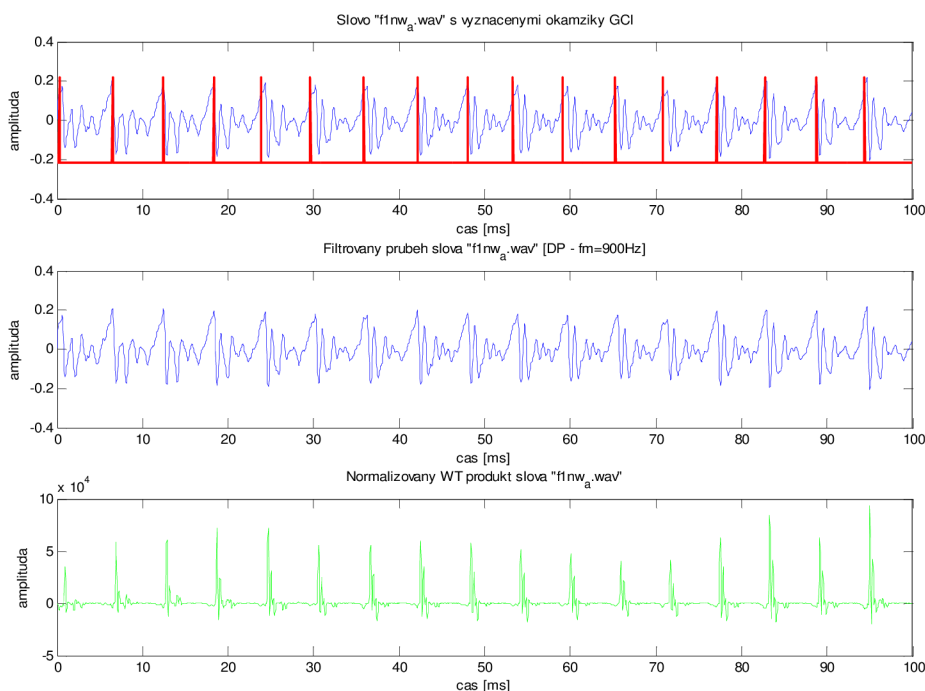


OBRÁZEK B.1.: ÚSEK HLÁSKY „A“ ZPRACOVANÝ VLNKOVOU TRANSFORMACÍ

B.1 Metoda pro odečtení pozice GCI z multiscale produktu

Jak bylo řečeno, okamžiky GCI nastávají v okamžicích lokálních minim vlevo od lokální maximální hodnoty. Pro každou periodu řečového signálu tedy není problém okem rozpoznat, kde se GCI nachází. Počítač toto bohužel „nevidí“. Abych správně určil

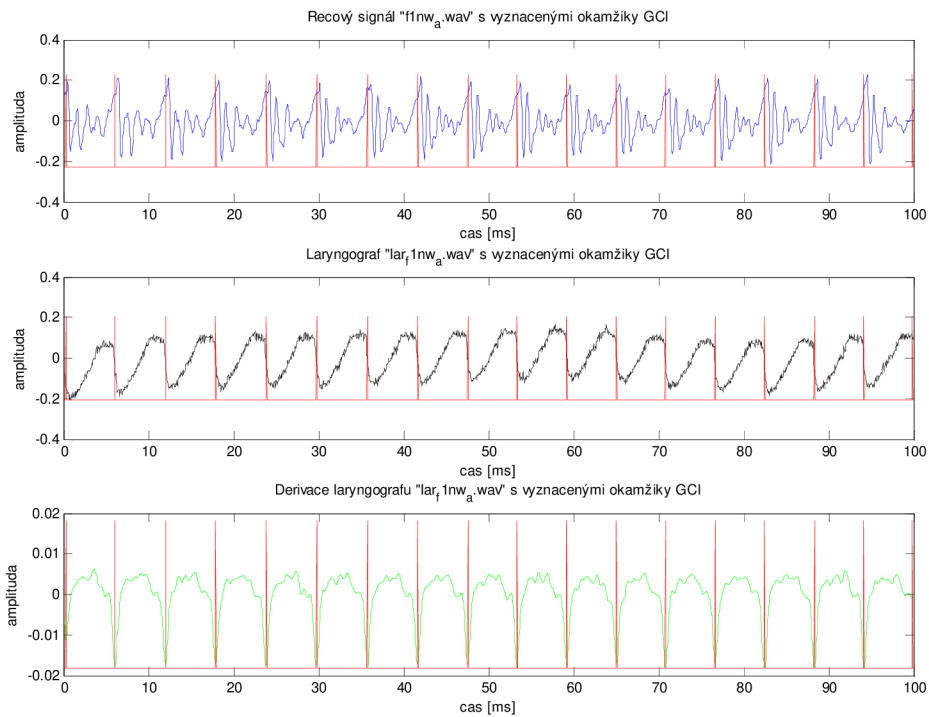
okamžiky GCI postupují takto: Na začátku vybereme z multiscale produktu úsek o průměrné délce periody základního tónu. V tomto úseku najdu maximum a potom postupuji zpět v čase a hledám minimum. Pozici minima d_1 uložím jako pozici GCI. Pro nalezení dalších GCI teď vybereme ze signálu úsek $s(d_1 + k, 2d_1 + k)$ kde $k = \bar{d} \cdot 0,8$ je posuv, který zajistí, že nebudeme žádný GCI identifikovat dvakrát a \bar{d} je průměrná délka periody základního tónu řeči. Na **Obrázek B.2** je ukázka úseku hlásky „a“ se zakreslenými okamžiky GCI.



OBRAZEK B.2.: ÚSEK HLÁSKY „A“ S VYZNAČENÝMI GCI

B.2 Určení pozice GCI z EGG

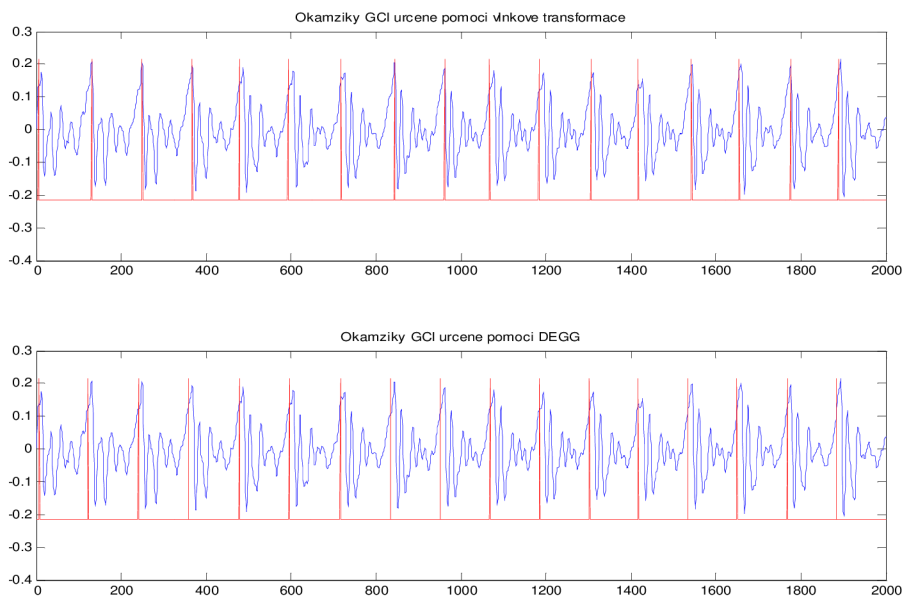
Protože okamžiky GCI podle [14] korespondují s okamžiky záporných špiček v DEGG použil jsem tuto metodu jako referenční. Část signálu EGG, která korespondovala k části řečového signálu v [15] jsem před derivací filtroval pomocí algoritmu v [13]. Pro filtraci se opět díky svým vlastnostem detekovat rychlé změny využívá vlnková transformace s mateřskou vlnkou tvořenou kubickým splajnem. Takto jsem EGG zbavil šumu. Minima z DEGG jsou potom lépe čitelná. Na **Obrázek B.3** jsou zobrazeny okamžiky GCI určené na základě DEGG.



OBRÁZEK B.3.: ÚSEK HLÁSKY „A“ S VYZNAČENÝMI GCI PODLE EGG

B.3 Porovnání pozic GCI stanovených oběma metodami

Na Obrázek B.4 je vidět srovnání výsledků dosažených oběma metodami.



OBRÁZEK B.4.: SROVNÁNÍ OBOU METOD PRO URČENÍ OKAMŽIKŮ GCI

Největší rozdíl v určení GCI byl 10 vzorků, což při vzorkovací frekvenci 22kHz činí 0,45 ms. Průměrná odchylka v celém zkoumaném úseku byla 5 vzorků, což činí 0,23 ms. Odchyly jsou velmi malé a proto lze okamžiky GCI výsledky stanovené na základě vlnkové transformace považovat za správné.

10. Použitá literatura

- [1] FILIPSSON, M.: Speech Analysis Tutorial.
Dostupné z < <http://www.ling.lu.se/research/speechtutorial/tutorial.html> >
- [2] The Voice Fundation, LEMG: Highlights.
Dostupné z < <http://www.voicefoundation.org/>>
- [3] PSUTKA, J. - MÜLLER, L. - MATOUŠEK, J. - RADOVÁ, V.: Mluvíme s počítačem česky. 1. vyd. Praha: Academia 2006. ISBN 80-200-1309-1
- [4] PULAKKA, H.: Analysis of Human Voice Production Using Inverse Filtering. High-Speed Imaging, and Electroglottography. Helsinki University of Technology. Master's Thesis. 2005
Dostupné z: < <http://lib.tkk.fi/Dipl/2005/urn007925.pdf>>
- [5] AIRAS, M. – PULAKKA, H. – BÄCKSTRÖM, T. – ALKU, P.: A Toolkit for Voice Inverse Filtering and Parametrisation. Laboratory of Acoustics and Audio Signal Processing. Helsinki University of Technology, Espoo, Finland. 2005.
Dostupné z: < <https://sourceforge.net/projects/aparat/>>
- [6] KAWAHARA, H.: YIN – Fundamental Frequency Estimator for Speech and Music. Wakayama University. 2002
Dostupné z: < http://www.ircam.fr/pcm/cheveign/pss/2002_JASA_YIN.pdf>
- [7] EL-JAROUDI, A. – MAKHOUL, J.: Discrete All-pole Modeling. IEEE Transactions on signal processing, Vol. 39, No. 2, February 1991
Dostupné z:
< <http://ieeexplore.ieee.org/iel4/78/2656/00080824.pdf?tp=&isnumber=2656&arnumber=80824>>
- [8] ČERNOCKÝ, J.: Zpracování řečových signálu – studijní opora. Vysoké učení technické v Brně, Fakulta informačních technologií, 2006
Dostupné z: < http://www.fit.vutbr.cz/~cernocky/speech/opora/zre_opora.pdf>
- [9] DENG, H. – WARD, R. K. – BEDDOES, M. P. – HODGSON, M.: A New Method for Obtaining Accurate Estimates of Vocal-Tract Filters and Glottal Waves From Vowel Sound. IEEE Transaction on audio, speech, and language processing. Vol. 14, No. 2. p. 445-455 . March 2006
- [10] WAKITA, H.: Normalization of Vowels by Vocal-Tract Length and Its Application to Vowel Identification. IEEE Transactions on Acoustics, Speech, and Signal Processing. Vol. ASSP-25. No. 2. April 1977
Dostupné z:
< <http://ieeexplore.ieee.org/Xplore/login.jsp?url=/iel6/29/26124/01162929.pdf?arnumber=1162929>>
- [11] BOUZID, A. – NOUREDDINE, E.: Glottal opening instant detection from speech signal. National School of Engineers of Tunis, Le Belvédère, Tunis
Dostupné z:< <http://cat.inist.fr/?aModele=afficheN&cpsidt=17633371>>
- [12] RAHMOUNI, A. B. S. – BOUZID, A. – ELLOUZE, N.: Wavelet decomposition of voiced speech and mathematical morphology analysis for glottal closure instants detection.
Dostupné z:< <http://www.eurasip.org/Proceedings/Eusipco/2002/articles/paper225.pdf>>

- [13] SADLER, B., M. – SWAMI, A.: Analysis of Wavelet Transform Multiscale Products for Step Detection and Estimation. U.S. Army Research Laboratory. 1998
Dostupné z
< <http://ieeexplore.ieee.org/Xplore/login.jsp?url=/iel4/18/16471/00761341.pdf?arnumber=761341>>
- [14] BOUZID, A. – NOUREDDINE, E.: Singularity Detection of Electroglottogram Signal by Multiscale Product Method. le Belvedere 1002. Tunis. Tunisia. 2006
Dostupné z: < <http://www.eurasip.org/Proceedings/Eusipco/Eusipco2006/papers/1568982250.pdf>>
- [15] Databáze obsahující společně zaznamenané průběhy řečového signálu a elektroglotograf 10 mluvčích.
Dostupné z < <ftp://ftp.cs.keele.ac.uk/pub/pitch>>
- [16] Internetová encyklopedie Wikipedie. Metoda borcení časové osy (DTW).
Dostupné z < http://en.wikipedia.org/wiki/Dynamic_time_warping>
- [17] M. Sigmund, "Introducing the database ExamStress for speech under stress," Proceedings of 7th IEEE Nordic Signal Processing Symposium (NORSIG 2006). Reykjavik, strana. 290-293, 2006.
- [18] N. N. VU. Speech Analysis Tutorial.
Dostupné z < <http://khoaanh.hcmup.edu.vn/index.php?name=News&file=article&sid=58>>

11. Seznam zkratk a symbolů

EGG	Electro Glotto Graph (záznam elektrické aktivity hlasivkových svalů)
DEGG	derivovaný průběh EGG
LP	lineární predikce
IAIF	Iterative Adaptive Inverse Filtering
AMDF	Average Magnitude Difference Function
DAP	Discrete All-Pole Modeling
GCI	Glottal Closure Instant (okamžik uzavření hlasivek)
$s(n)$	diskretizovaný řečový signál
f_0	základní frekvence řečového signálu
T_0	základní perioda řečového signálu
$u_g(t)$	hlasivkový pulz
$v_g(t)$	derivovaný hlasivkový pulz
$p_{mic}(n)$	signál z mikrofону
$G(z)$	zdroj periodického signálu pro model tvorby řeči v rovině „z“
$V(z)$	přenosová funkce hlasového traktu v rovině „z“
$R(z)$	impedance rtů v rovině „z“
$R(m)$	autokorelace
$H_{VTF}(z)$	přenosová funkce hlasového traktu v rovině „z“
$U_{LIP}(z)$	řečový signál přicházející na rty
$U_G(z)$	hlasivkový pulz v rovině „z“
Z_{VTF}	impedance hlasového traktu
F_{VZ}	vzorkovací frekvence