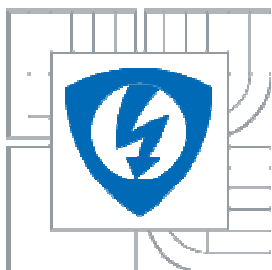


VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH
TECHNOLOGIÍ
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

OBJEKTIVNÍ METODY HODNOCENÍ KVALITY ŘEČI SPEECH QUALITY OBJECTIVE METHODS

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE
AUTHOR

JIŘÍ KOVAŘÍK

VEDOUCÍ PRÁCE
SUPERVISOR

Ing. PETR SYSEL, Ph.D.

BRNO 2010



VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

Ústav telekomunikací

Bakalářská práce

bakalářský studijní obor
Teleinformatika

Student: Jiří Kovařík
Ročník: 3

ID: 106553
Akademický rok: 2009/2010

NÁZEV TÉMATU:

Objektivní metody hodnocení kvality řeči

POKYNY PRO VYPRACOVÁNÍ:

Cílem bakalářské práce je implementovat vybrané metody objektivního hodnocení kvality řeči v prostředí Matlab. Seznamte se s objektivními metodami hodnocení kvality řečového signálu, např. poměr signálu od šumu, PSQM dle ITU-T P.861, Itakurova míra, apod. Zaměřte se především na ne-intrusivní metody, které nevyžadují známý řečový signál bez rušení. Vybrané metody implementujte v prostředí Matlab a ověřte jejich funkci na dostupných nahrávkách z reálných prostředí. Porovnejte výsledky jednotlivých metod a prozkoumejte závislosti mezi metodami.

DOPORUČENÁ LITERATURA:

- [1] Deller, J. R. - Hansen, J. H. L. - Proakis, J. G. Discrete-Time Processing of Speech Signals. Reprint Edition. New York: IEEE, 2000. 908 s. ISBN 0-7803-5386-2
- [2] ITU-T. P.861 Objective quality measurement of telephone-band (300-3400 Hz) speech codecs. 1996.
- [3] ITU-T. Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. 2001.
- [4] Doňar, B., Zaplatílek, K. MATLAB - začínáme se signály. 1. vydání. Praha: BEN, 2006. 272 s. ISBN 80-7300-200-0

Termín zadání: 29.1.2010

Termín odevzdání: 2.6.2010

Vedoucí práce: Ing. Petr Sysel, Ph.D.

prof. Ing. Kamil Vrba, CSc.
Předseda oborové rady

UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské 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 části druhé, hlavy VI, díl 4 Trestního zákoníku č.40/2009 Sb.

Prohlášení

Prohlašuji, že svou bakalářskou práci na téma „Objektivní metody hodnocení kvality řeči“ jsem vypracoval samostatně pod vedením vedoucího semestrálního projektu 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 bakalářské práce dále prohlašuji, že v souvislosti s vytvořením tohoto projektu 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 2.6.2010

.....
podpis autora

Poděkování

Děkuji vedoucímu práce Ing. Petru Syslovi, Ph.D. za ochotu pomoci a za cenné rady při zpracování této bakalářské práce.

V Brně dne 2.6.2010

.....
podpis autora

Abstrakt

V rámci bakalářské práce byl vytvořen program pro hodnocení kvality řeči. Tento program byl vytvořen v programovacím jazyku MATLAB s využitím programovacího prostředí GUI. Pomocí vytvořeného programu byla zpracována skupina nahrávek zarušených různými hluky z reálného prostředí. Jako objektivní metody hodnocení kvality řeči byly vybrány Základní metoda SNR, Segmentální metoda SNR s využitím detektoru řečové aktivity, PESQ, Itakurova míra, Itakurova – Saitova míra. K programu byl vytvořen návod k obsluze. V práci jsou popsány výše zmíněné objektivní metody hodnocení kvality řeči a dále proces vytváření lidské řeči, modelování lidské řeči a lineární prediktivní analýza.

Klíčová slova

Poměr signál od šumu, PESQ, Itakurova míra, Itakurova – Saitova míra, lineární prediktivní analýza, modelování lidské řeči

Abstract

In terms of the bachelor's thesis was created a computer programme for the quality speech estimation. This programme was created in the MATLAB programming language using GUI programming environment. Through the use of created programme was processed group of records disturbed by various noises from the real environment. As a speech quality objective methods was chosen classical SNR, Segmental SNR, PESQ, Itakura distance, Itakura – Saito distance For this programme were created service instructions. In this work are these speech quality objective methods described furthermore the process of the human voice creation, speech modeling and a linear predictive coding.

Keywords

Signal to noise ratio, PESQ, Itakura measure, Itakura – Saito measure, linear predictive coding, speech modeling

KOVAŘÍK, J. *Objektivní metody hodnocení kvality řeči*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2010. 50 s. Vedoucí bakalářské práce Ing. Petr Sysel, Ph.D.

OBSAH

ÚVOD	10
1 LIDSKÁ ŘEČ	11
1.1 Proces vytváření lidské řeči	11
1.1.1 Dechové ústrojí	12
1.1.2 Hlasové ústrojí	12
1.1.3 Artikulační ústrojí	13
1.2 Model vytváření lidské řeči	14
1.2.1 Model hlasivek	15
1.2.2 Model hlasového traktu	15
1.2.3 Model vyzařování zvuku	16
1.2.4 Model buzení řečové produkce	16
1.2.5 Celkový model produkce řeči	16
2 SUBJEKTIVNÍ METODY HODNOCENÍ KVALITY ŘEČI	18
2.1 Metoda relativní preference	18
2.2 MOS (Mean Opinion Scores)	18
3 OBJEKTIVNÍ METODY HODNOCENÍ KVALITY ŘEČI	19
3.1 Poměr signálu od šumu (SNR)	20
3.1.1 Základní metoda SNR	21
3.1.2 Segmentální SNR	21
3.1.3 SNR s rozdělením do kmitočtových pásem	22
3.1.4 Segmentální SNR s využitím VAD	23
3.2 Detektor řečové aktivity	24
3.3 Míry zkreslení založené na lineární prediktivní analýze	25
3.3.1 Lineární prediktivní analýza	25
3.3.2 Itakurova míra	27
3.3.3 Itakurova – Saitova míra	29
3.4 PESQ	30
3.4.1 Doporučené vlastnosti testovaných nahrávek	30
3.4.2 Popis algoritmu PESQ	31
4 VYPRACOVÁNÍ	33
4.1 Grafické závislosti jednotlivých metod	38
5 ZÁVĚR	40
LITERATURA	41
SEZNAM SYMBOLŮ, VELIČIN A ZKRATEK	42

SEZNAM PŘÍLOH	44
A Vypočtené hodnoty programem kvalita_reci	45
A.1 Výsledky pro nahrávku zarušenou Gaussovým šumem	45
A.2 Výsledky pro nahrávku zarušenou šumem o kmitočtu 1 kHz	46
A.3 Výsledky pro nahrávku zarušenou hlukem ledničky	47
A.4 Výsledky pro nahrávku zarušenou škrábáním	48
A.5 Výsledky pro nahrávku zarušenou šumem vysílačky	49
B Přiložené CD	50

Seznam obrázků

OBRÁZEK 1.1: PŘEHLED ORGÁNŮ DÝCHACÍHO ÚSTROJÍ [8].	11
OBRÁZEK 1.2: SCHÉMA POSTAVENÍ HLASIVEK VE FONAČNÍ POZICI (A), V KLIDOVÉ POZICI (B) [9].	12
OBRÁZEK 1.3: CELKOVÝ MODEL PRODUKCE ŘEČI [7].	17
OBRÁZEK 3.1: ZOBRAZENÍ JEDNOKANÁLOVÝCH A DVOUKANÁLOVÝCH METOD [3].	19
OBRÁZEK 3.2: ČASOVÝ PRŮBĚH NEZARUŠENÉHO ŘEČOVÉHO SIGNÁLU.	20
OBRÁZEK 3.3: ČASOVÝ PRŮBĚH ZARUŠENÉHO ŘEČOVÉHO SIGNÁLU GAUSSOVÝM ŠUMEM.	20
OBRÁZEK 3.4: DETEKCE ŘEČI V ZARUŠENÉ NAHRÁVCE.	24
OBRÁZEK 3.5: SCHÉMA METODY PESQ.	32
OBRÁZEK 4.1: ZOBRAZENÍ PROGRAMU KVALITA_REC1 V ZÁKLADNÍ KONFIGURACI IHNEDE PO SPUŠTĚNÍ.	33
OBRÁZEK 4.2: ZOBRAZENÍ GRAFŮ PRO NAHRÁVKU ORIG_LEDNICKA26DB.WAV.	34
OBRÁZEK 4.3: ZOBRAZENÍ NÁPOVĚDY PRO PROGRAM KVALITA_REC1.	35
OBRÁZEK 4.4: ZOBRAZENÍ VÝVOJOVÉHO ALGORITMU PŘI STISKNUTÍ TLAČÍTKA VÝPOČÍTEJ.	37
OBRÁZEK 4.5: POROVNÁNÍ ZÁVISLOSTÍ POMĚRU SIGNÁLU OD ŠUMU S VYUŽITÍM VAD NA HODNOCENÍ KVALITY METODOU PESQ.	38
OBRÁZEK 4.6: POROVNÁNÍ ZÁVISLOSTÍ KLASICKÉHO POMĚRU SIGNÁLU OD ŠUMU NA HODNOCENÍ KVALITY METODOU PESQ.	38
OBRÁZEK 4.7: POROVNÁNÍ ZÁVISLOSTÍ ITAKUROVY MÍRY NA HODNOCENÍ KVALITY METODOU PESQ...	39
OBRÁZEK 4.8: POROVNÁNÍ ZÁVISLOSTÍ ITAKUROVY - SAITOVY MÍRY NA HODNOCENÍ KVALITY METODOU PESQ.	39

Seznam tabulek

TABULKA 2.1: ZOBRAZENÍ STUPNICE MOS [4].	18
TABULKA 3.1: VÁHOVÉ KOEFICIENTY w_j PRO VÝPOČET SNR_{FW-SEG} [1].	22
TABULKA 4.1: NASTAVENÉ PARAMETRY PRO HODNOCENÍ NAHRÁVEK PROGRAMEM KVALITA_REC1.	36

Úvod

Tato bakalářská práce je zaměřena na objektivní metody hodnocení kvality řeči. Je popsán proces vytváření lidské řeči a jeho modelování, abychom věděli, co reprezentují koeficienty lineární prediktivní analýzy.

Výsledná kvalita hlasového přenosu závisí na množství parametrů, které zahrnují okamžité fyzikální vlastnosti všech prvků, které tvoří komunikační řetězec. V minulosti byla kvalita přenosu po analogové lince odhadována převážně z velikostí odstupe signálu od šumu, útlumem trasy a u spojení na velmi dlouhé vzdálenosti také dozvukem či ozvěnou (echo). S nástupem digitální technologie se význam útlumu trasy snížil. U bezdrátových telekomunikačních systémů se však objevily zcela nové příznaky, degradující kvalitu hlasového přenosu, jako je impulsní šum, krátkodobé výpadky, ořezávání signálu a nelineární zkreslení vlivem použití kodeků. Další prvky zpracování signálu v komunikačním řetězci navíc prodlužují dobu šíření signálu a mohou tak mimo jiné přispět ke zvýraznění echa i při komunikaci na krátké vzdálenosti. U paketově orientovaných přenosů se navíc objevuje proměnné zpoždění nazývané jitter, které je způsobeno směrováním paketů odlišnými trasami. Z výše uvedených parametrů vyplývá složitost dokonalé metody pro ohodnocení kvality řeči. Pojem objektivní metody zde neznamena nestranné měření či měření, u kterého hodnoty jsou obecně platné, ale měření bez přítomnosti posluchačů.

V první části jsou stručně uvedeny subjektivní metody hodnocení kvality řeči a to zejména proto, že pomocí objektivních metod bychom měli dosahovat podobných výsledků jako pomocí metod subjektivních. Popsány jsou metody relativní preference a MOS.

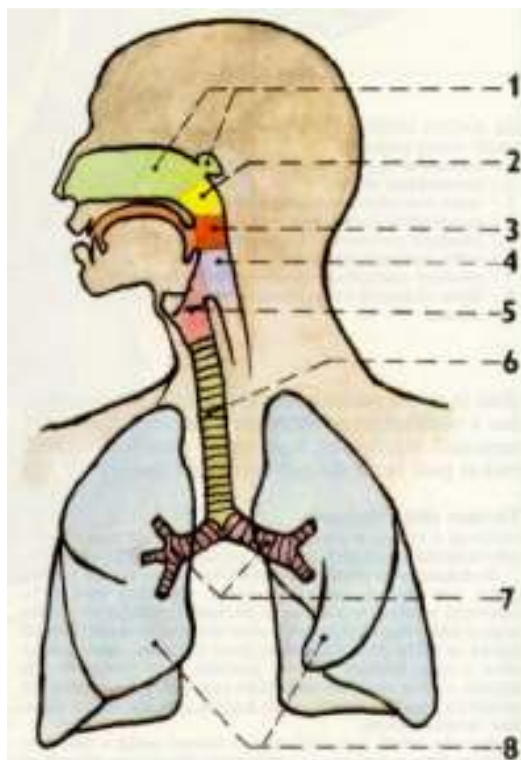
Druhá část semestrálního projektu obsahuje objektivní metody, vysvětlení z jakého důvodu vůbec vznikly. Podrobnější popis metod SNR, měř zkreslení založených na lineární prediktivní analýze, metoda PESQ a okrajové vysvětlení, k čemu se používá detektor řečové aktivity.

1 Lidská řeč

Mluvená řeč je vůbec nejstarším způsobem komunikace mezi lidmi. Až postupem času se člověk naučil své myšlenky zaznamenávat písmem a tak je uchovávat pro další generace. Řeč je pro nás naprosto přirozený automatický proces, u kterého můžeme vykonávat spousty dalších věcí jako např. chůze, psaní, ovládání PC atd.

1.1 Proces vytváření lidské řeči

Mluvená řeč se šíří pomocí mechanického vlnění akustickým kanálem. Akustickým kanálem se rozumí prostředí od zdroje signálu (ústa řečníka), až k příjemci (sluchový orgán). Toto prostředí může být plynné, kapalné či pevné. V plynném a kapalném prostředí se akustický signál šíří pomocí podélného vlnění ve směru šíření a v prostředí pevném pomocí vlnění podélného a příčného. Vytváření lidské řeči se účastní několik orgánů, kterým se souhrnně říká *řečové orgány*. Primárním úkolem těchto orgánů sice není vytvářet lidskou řeč a jejich funkce spolu často nesouvisí, ale výsledná akustická podoba je výrazně ovlivněna funkcí každého z nich. Při tvorbě řeči tyto orgány tvoří *hlasový trakt*. Jeho délka je u dospělého člověka asi 17 cm a jeho příčný průřez se mění v rozmezí 0 až zhruba do 20 cm². Hlasový trakt je rozdělen do tří částí: *dechové, hlasové a artikulační ústrojí*.



1. Dutina nosní
2. Nosohltan
3. Ústní část hltanu
4. Hrtanová část hltanu
5. Hrtan
6. Průdušnice
7. Průdušky
8. Plíce

Obrázek 1.1: Přehled orgánů dýchacího ústrojí [8].

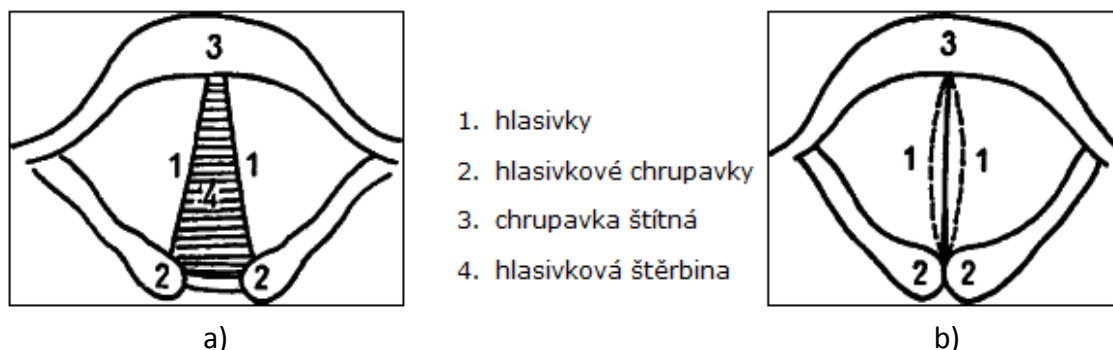
1.1.1 Dechové ústrojí

Dechové ústrojí zastává funkci zdroje energie pro vytváření lidské řeči. Tato energie je ve formě vydechovaného vzduchu, který opouští plíce přibližně konstantní rychlostí $0,2 \text{ l} \cdot \text{s}^{-1}$. Kapacita plic je u dospělého člověka v rozmezí 4 až 5 l, přičemž 1 až 2 l tvoří tzv. zbytkovou kapacitu plic, která musí být vždy zachována. Výdechový proud vzduchu je odváděn z plic průdušnicí, pak prochází hrtanem a nadhrtanovými dutinami, kde se modifikuje a jako akustický signál je vyzařován rty do okolního prostředí. Výdechový proud má vliv na sílu a barvu hlasu. Při normální řeči se spotřebuje až polovina vitální kapacity plic a u hlasité řeči až 80 %. Doplnění nového materiálu pro vytváření řeči se provádí pomocí nádechu a při promluvě se tak děje při pauze v jinak souvislé řeči.

1.1.2 Hlasové ústrojí

Hlasové ústrojí je uloženo v hrtanu a jeho funkcí je vytvářet základní tón, jehož dalšími úpravami vzniká řeč. Základem hlasového ústrojí jsou hlasivky, které se nacházejí v hrtanové dutině přímo za ohryzkem. Jsou to dva valy pokryté sliznicí, které vedou napříč hrtanem v místě jeho nejužšího průchodu. U mužů je jejich průměrná délka 15 mm a u žen 13 mm. Hlasivky jsou napjaty mezi chrupavkou štítnou, kde se navzájem dotýkají a chrupavkami hlasivkovými, které jsou částečně pohyblivé, mohou se různým způsobem otáčet, sblížovat a měnit sklon. Jejich způsob svalového a nervového ovládání je mimořádně složitý a doposud nebyl v úplnosti popsán. Mezi hlasivkami se vytváří hlasivková štěrbina trojúhelníkového tvaru. Maximální velikost plochy štěrby je u mužů 20 mm^2 a u žen 14 mm^2 (při kmitání hlasivek).

V klidové pozici, kdy není vytvářena řeč, je tato štěrbina veliká asi 8 mm^2 a nijak nebrání průchodu vzduchu při dýchání.



Obrázek 1.2: Schéma postavení hlasivek ve fonační pozici (a), v klidové pozici (b) [9].

Ve fonační pozici se hlasivky nastaví do hlasového postavení. Proud vzduchu z plic putuje průdušnicí až k hrtanu bez jakéhokoliv odporu. Zde se mu do cesty postaví hlasivky, které cestu vzduchu úplně uzavřou. Jelikož jsou hlasivky pružné, tak pod tlakem vzduchu začínají kmitat. Vlivem kmitání hlasivek se do této doby konstantní proud vzduchu rozštěpí a vznikají téměř pravidelně hustší a řidší intervaly vzduchu. Tyto intervaly tvoří vzduchovou vlnu, která tvoří základ lidského hlasu a bývá označována jako *základní tón*. Frekvence kmitání hlasivek se označuje jako frekvence základního hlasivkového tónu F_0 . Tato frekvence je jednou ze základních charakteristik akustického signálu a představuje výšku hlasu. Pohybuje se v rozmezí 60 Hz až 400 Hz.

Specifické postavení hlasivek nastává při šepotu, kdy hlasivky nekmitají. Jsou pootevřené a zvuk vzniká turbulencí vzduchu způsobenou třením proudu vzduchu o hlasivky. V některých afrických jazycích toto postavení hlasivek používá i pro normální řeč.

Při tvorbě řeči se používá klidového i fonačního postavení hlasivek. U fonačního postavení se vytvářejí znělé zvuky, které obsahují základní hlasivkový tón. Neznělé zvuky jsou tvořeny při klidovém postavení a vznikají až změnou výdechového proudu v nadhrtanových dutinách a neobsahují základní hlasivkový tón.

1.1.3 Artikulační ústrojí

Artikulační ústrojí je uloženo nad hrtanem a skládá se ze tří dutin: *dutiny hrdelní*, *dutiny nosní* a *dutiny ústní*. Dohromady tyto dutiny bývají označovány jako *nadhrtanové dutiny*. Je to poslední ústrojí, které se podílí na tvorbě řeči a jeho úkolem je vytvářet velké množství zvuků, které charakterizují daný jazyk. Člověk je schopen vyprodukovat velké množství zvuků a to zejména zásluhou artikulačních orgánů. Tyto orgány se na tvorbě řeči podílejí aktivně a tvoří pohyblivé součásti artikulačního ústrojí, které mění velikost nadhrtanových dutin. Nejvýznamnější artikulační orgány, které se podílejí na největším počtu zvuků, jsou jazyk, rty a měkké patro.

Jazyk je vůbec nejdůležitějším a nejsložitějším artikulátorem. Je velmi pružný a schopný vytvářet spoustu tvarů, které rychle mění. Právě tato změna vytváří různé tvary dutiny ústní a dutiny hrdelní.

Dalšími artikulátory jsou potom zuby, tvrdé patro, čelisti nebo hrtan, který kromě podílu na tvorbě znělosti se může pohybovat směrem nahoru a dolů a tak měnit délku hlasového traktu.

Průchodem základního hlasivkového tónu nadhrtanovými dutinami vzniká rezonance. V jaké dutině vznikne závisí na postavení artikulačního ústrojí. Tyto rezonanční frekvence jsou označovány jako *formantové kmitočty* a oblasti, kde

dochází k zesílení akustické energie jako *formanty*. Označují se čísla počínaje formantem s nejnižším kmitočtem $F_1, F_2, \dots, F_n$. Pokud se do procesu vytváření řeči zapojí i dutina nosní dochází k potlačení některých frekvenčních oblastí. Tyto oblasti jsou označovány jako *antiformanty* ($A_1, A_2, \dots, A_n$). Tímto způsobem vzniká znělá část lidské řeči, kde změnou tvaru artikulátorů ovlivňujeme její frekvenční vlastnosti.

U neznělé části řeči se do nadhrtanových dutin dostává přímo výdechový proud, nijak nezměněný hlasivkami a teprve až artikulátory ovlivňují výsledný zvuk. Výdechovému proudu se do cesty postaví překážky ve formě artikulátorů a tím vzniká šum různého druhu, který tvoří neznělou řeč. Různých druhů zvuků se v tomto případě dosahuje různým postavením artikulátorů tzn. různým postavením překážek na různých místech hlasového traktu.

Během řeči dochází k plynulému pohybu všech artikulátorů a dohromady tvoří jistou konfiguraci hlasového traktu. K této konfiguraci nedochází najednou, protože se nejedná o diskrétní systém a přechod z jedné konfigurace do druhé zabere určitý čas. Tento čas je závislý na hmotnosti přemísťovaných orgánů a na použitých svalových silách. Tím že se artikulátory během vytváření řeči plynule pohybují, tak výsledný charakter řeči nezávisí pouze na konfiguraci hlasového traktu, ale i na předchozí promluvě. Tento jev se nazývá *koartikulace* a má za následek, že akustické realizace stejného zvuku jsou závislé na zvucích, které jim předcházely.

1.2 Model vytváření lidské řeči

Při modelování procesu vytváření a vnímání řeči je řeč vyjádřena parametry daného modelu. Cílem modelování produkce lidské řeči je nalézt matematické vztahy, které reprezentují akustické a fyzikální děje při vytváření řeči. Tyto matematické vztahy by měli být výpočetně jednoduché při dostatečné přesnosti. Ideální by bylo, kdyby model byl lineární a časově invariantní. Lidská řeč je ale časově proměnný a nelineární proces. Je to způsobeno tím, že hlasivky nejsou během produkce řeči zcela odpojeny od hlasového traktu. Kvůli neúměrné složitosti zatím nebyl zpracován univerzální model, který by respektoval všechny požadavky a použitý model při zpracování řeči se volí v závislosti na požadovaném použití.

Dále se budeme zabývat modelem produkce řeči, který se používá u LPC kódování a u LPC syntetizátorů. Jedná se o krátkodobý model, který je lineární a časově invariantní. U mnoha zvuků lidské řeči je typ buzení a konfigurace hlasového traktu téměř konstantní po dobu 10 – 30 ms. Proto se model může skládat z lineárního modelu hlasového traktu s pomalu se měnícími parametry a vhodným zdrojem budícího signálu. Je ale důležité si uvědomit, že parametry

modelu jsou v čase proměnné ale v době jednoho rámce (10 – 30 ms) jsou konstantní. Takový model je označován jako *model krátkodobý*. Pro znělou řeč je budící signál periodická posloupnost pulsů odpovídajících frekvenci základního hlasivkového tónu F_0 . U neznělé řeči je budící signál náhodný šum. Číslicový model produkce řeči by měl respektovat chování hlasivek, nadhrtanových dutin, artikulačních orgánů a také proces vyzařování zvuku.

1.2.1 Model hlasivek

Frekvenci kmitání hlasivek ovlivňuje síla výdechového proudu, hmotnost, pružnost hlasivek a velikost hlasivkové štěrbiny ve volném stavu. Model hlasivek reprezentuje nízkopásmový filtr druhého řádu. Parametry tohoto filtru jsou různé pro každého řečníka a zlomová frekvence se pohybuje okolo 100 Hz. Přenosová funkce $G(z)$ modelu hlasivek má tvar:

$$G(z) = \frac{1}{[1 - e^{(-cT)}z^{-1}]^2} \quad (1.1)$$

Kde: c neznámý parametr filtru,
 T perioda vzorkování.

1.2.2 Model hlasového traktu

Model hlasového traktu simuluje především vlastnosti dutiny hrdelní, ústní a dutiny nosní. Je závislý na aktuální konfiguraci dutin a artikulátorů v časovém úseku jednoho rámce (10 – 30 ms). Parametry tohoto modelu jsou velmi proměnné a při vyslovování téhož zvuku nemusí být stejné, i když se jedná o stejného řečníka. V číslicové oblasti se model hlasového traktu skládá z kaskády dvojpolových rezonátorů, kde rezonanční kmitočty reprezentují jednotlivé formantové kmitočty F_n . Vzhledem k akustické teorii by měl model respektovat i antiformantové kmitočty. Z této teorie vyplývá, že by přenosová funkce modelu hlasového traktu měla obsahovat póly i nulové body. Bylo ale dokázáno, že nutnost nul lze kompenzovat zvětšením počtu pólů přenosové funkce. Celopólová funkce $V(z)$ modelu hlasového traktu je:

$$V(z) = \frac{1}{\prod_{i=1}^K [1 - 2e^{(-\alpha_i T)} \cos(\beta_i T) z^{-1} + e^{(-2\alpha_i T)} z^{-2}]} \quad (1.2)$$

Kde: α neznámý parametr,
 β neznámý parametr,
 K počet formantů,
 T perioda vzorkování.

1.2.3 Model vyzařování zvuku

Bližší informace lze najít v literatuře [11]. Přenosová funkce $L(z)$ byla odvozena ve tvaru:

$$L(z) = 1 - z^{-1} \quad (1.3)$$

1.2.4 Model buzení řečové produkce

Pro znělé části řeči je buzení ve formě periodicky se opakujících pulsů s periodou základního hlasivkového tónu T_0 . Budící funkce s periodou vzorkování T je tedy nenulová pouze v časových intervalech odpovídajících periodě základního hlasivkového tónu T_0 , která lze vyjádřit jako:

$$T_0 = b \cdot T \quad (1.4)$$

Kde: T_0 perioda základního hlasivkového tónu,
 b koeficient, který udává kolikrát je T_0 delší než T ,
 T perioda vzorkování.

Za předpokladu, že perioda vzorkování $T = 1$ bude z-obraz budící funkce $U(z)$ ve tvaru:

$$U(z) = G \sum_{n=0}^{\infty} (z^{-b})^n = \frac{G}{1 - z^{-b}} \quad (1.5)$$

Kde: G zesílení zdroje buzení.

Pro neznělé zvuky řeči je zdrojem buzení náhodný šum s rozprostřeným spektrem. V číslicové oblasti může být tento zdroj vytvořen pomocí generátoru náhodných čísel, kde posloupnost čísel bude mít rozprostřené spektrum.

1.2.5 Celkový model produkce řeči

Celkový model je složen z modelu hlasivek, hlasového traktu, modelu procesu vyzařování a příslušného zdroje signálu podle toho jestli se jedná o znělý segment či o segment neznělý. Výsledný model je znázorněn na obrázku (1.3). Přenosová funkce $H(z)$ celkového modelu a má tvar:

$$H(z) = \frac{1 - z^{-1}}{[1 - e^{(-cT)}z^{-1}]^2 \{\prod_{i=1}^K [1 - 2e^{(-\alpha_i T)} \cos(\beta_i T)z^{-1} + e^{(-2\alpha_i T)}z^{-2}]\}} \quad (1.6)$$

Výsledná přenosová funkce $H(z)$ je dána sloučením vzorců (1.1), (1.2), (1.3) a význam jednotlivých členů byl popsán v předchozích kapitolách.

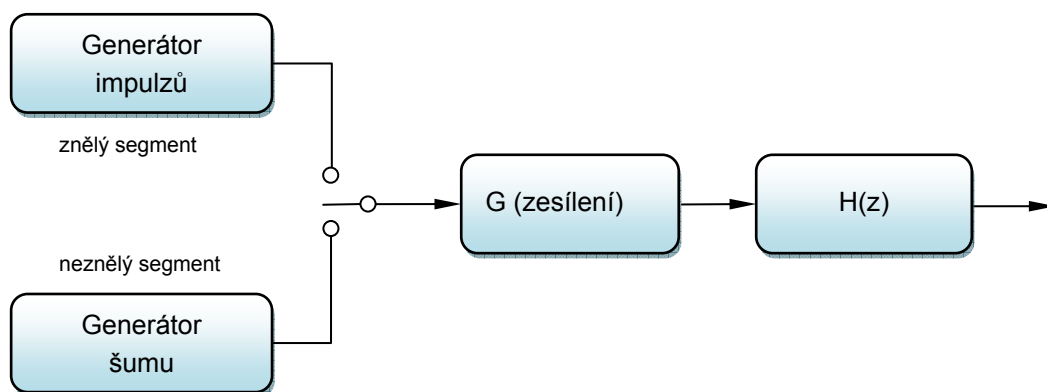
Protože člen cT je daleko menší než jedna potom bude člen $e^{(-cT)}$ roven přibližně jedné a je možné vykrátit členy $1 - z^{-1}$ v čitateli i jmenovateli. Výsledná přenosová funkce $H(z)$ modelu produkce řeči obsahuje pouze nulové body, označuje se jako celopólový filtr a je ve tvaru:

$$H(z) = \frac{1}{1 + \sum_{i=1}^Q a_i z^{-i}} \quad (1.7)$$

Kde: Q řád filtru,
 a neznámý parametr.

Řád filtru $Q = 2K + 1$ kde K značí počet formantů, které chceme modelem postihnout. Chceme-li navíc kompenzovat vliv chybějících nulových bodů v přenosové funkci poté by měl řád filtru nabývat hodnot $Q > 2K + 1$. Všechny póly přenosové funkce vždy leží uvnitř jednotkové kružnice a z toho vyplývá že model produkce řeči je stabilní systém. Známe-li z-obraz budící funkce $U(z)$ a přenosovou funkci modelu produkce řeči $H(z)$ bude z-obraz výstupu $S(z)$ roven:

$$S(z) = H(z) \cdot U(z) \quad (1.8)$$



Obrázek 1.3: Celkový model produkce řeči [7].

2 Subjektivní metody hodnocení kvality řeči

Subjektivní metody hodnocení kvality řeči jsou založeny na názoru posluchače či skupiny posluchačů. Ti buď porovnávají, která nahrávka je podle jejich mínění kvalitnější, nebo určují kvalitu podle předem stanovené stupnice.

2.1 Metoda relativní preference

Jedná se o jednu z prvních subjektivních metod. Posluchačům jsou přehrány dva záznamy. První záznam je *referenční* a druhý *testovací*. Referenční úsek řeči je určitým způsobem degradován. Posluchači mají určit který z úseků testovaný, či referenční má podle jejich subjektivního mínění lepší kvalitu. Referenční záznamy bývají degradovány aditivním šumem, filtrací nebo často používanou metodou NMRU definovanou ve standartu ITU-P.810. Nevýhodou této metody je, že nám často nestačí hodnocení, která z nahrávek je kvalitnější, ale potřebujeme tuto kvalitu zařadit na nějakou stupnici.

2.2 MOS (Mean Opinion Scores)

V této metodě hodnotí posluchači kvalitu řeči podle pětibodové stupnice viz tabulka 1.1. Měřená kvalita je určena jako průměr výsledků získaných od jednotlivých posluchačů. Aby byl rozdíl mezi názory posluchačů, na pojem dobré kvality sjednocen, tak se testy metody MOS provádějí ve dvou fázích. První fáze je tzv. *trénování*, kdy jsou posluchačům přehrávány postupně nahrávky s výbornou, střední a špatnou kvalitou. Po té následuje druhá fáze, kdy jsou posluchačům přehrávány testovací nahrávky a ti hodnotí kvalitu řeči. Výhodou metody MOS je, že posluchači přidělí přehrávané nahrávce hodnocení podle vlastního názoru, co je stále ještě dobrá kvalita a co ne. Tímto je test použitelný pro širokou škálu zkreslení.

Tabulka 2.1: Zobrazení stupnice MOS [4].

hodnocení	kvalita řeči	úroveň zkreslení
1	vynikající	nerozpoznatelné
2	dobrá	roznosnatelné, ale ne nepřijemné
3	přijatelná	roznosnatelné a trochu nepřijemné
4	špatná	nepřijemné, ale ne protivné
5	nepřijatelná	velmi nepřijemné a protivné

3 Objektivní metody hodnocení kvality řeči

Subjektivní metody měření patří mezi nejspolehlivější metody, protože samotný posluchač hodnotí kvalitu a srozumitelnost řeči. Nicméně tyto subjektivní testy jsou časově náročné a vyžadují ve většině případů testované a trénované posluchače. Z těchto důvodů byly postupně vytvořeny objektivní metody měření, které mají jednodušším způsobem odhadnout subjektivní vlastnosti řeči.

Kritérium pro objektivní metody je korelace s metodami subjektivními. Pro získání korelačních koeficientů pro subjektivní a objektivní metody musí být použita stejná databáze testovaných signálů. Korelační analýza je využita pro zjištění, jestli objektivní metody předpovídají kvalitu řeči tak, jako by ji hodnotili posluchači pomocí testů subjektivních.

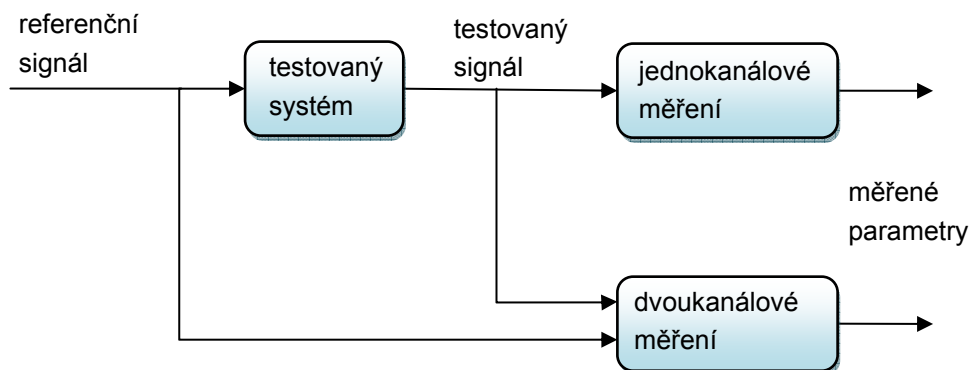
Objektivní metody lze rozdělit na dvě skupiny:

- Jednokanálové měření (non - intrusive)

U těchto měření máme k dispozici pouze degradovaný nebo zvýrazněný řečový signál. Řečový signál rozdělíme na segmenty pevné délky (většinou 15 – 30 ms), poté provedeme výpočty pro každý rámec a výsledek se vytvoří z hodnot vypočtených v jednotlivých rámcích.

- Dvoukanálové měření (intrusive)

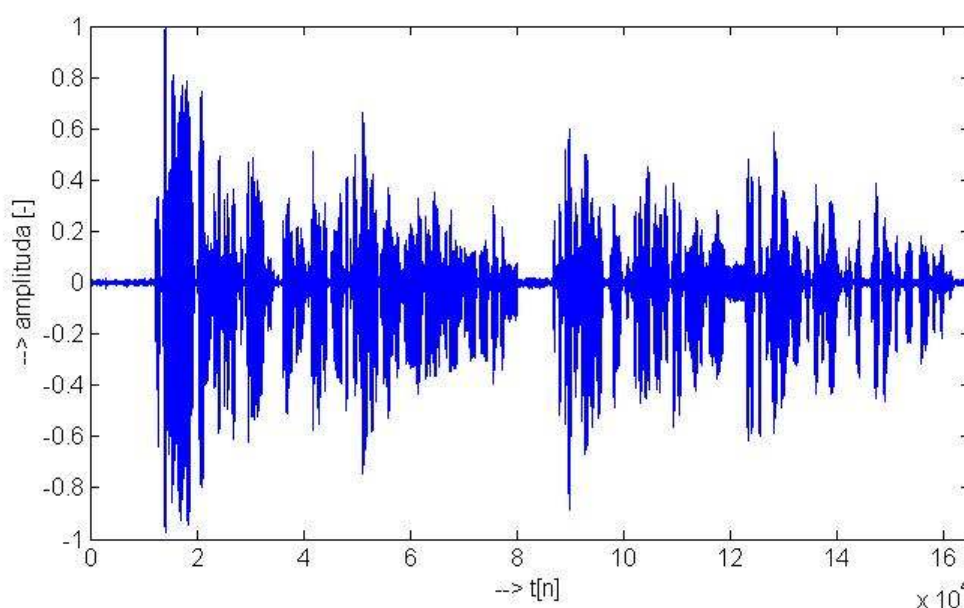
Porovnávají originální signál a degradovaný, který prošel testovaným systémem.



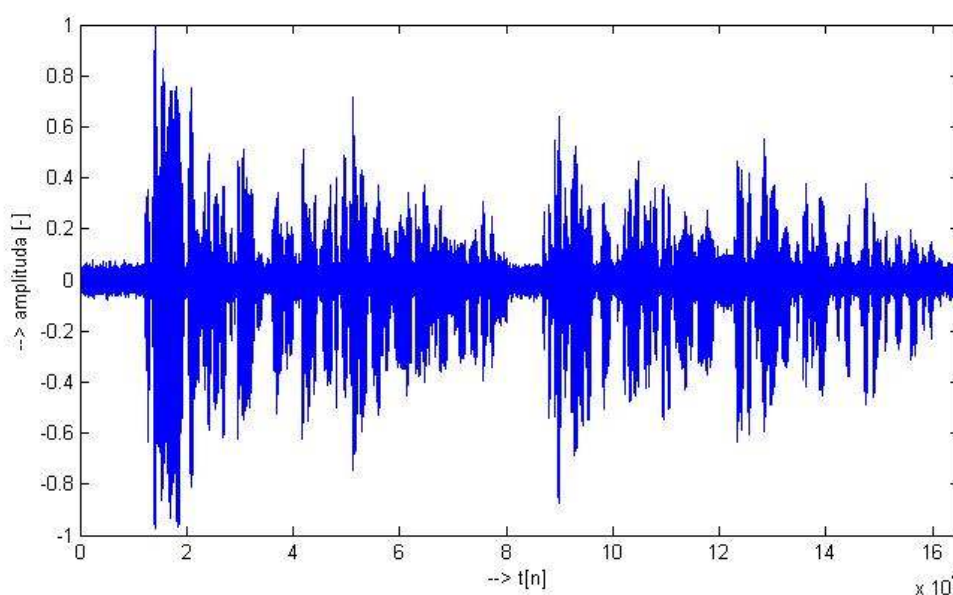
Obrázek 3.1: Zobrazení jednokanálových a dvoukanálových metod [3].

3.1 Poměr signálu od šumu (SNR)

Odstup signálu od šumu (označován jako SNR popřípadě jako S/N) je velmi často používaná veličina v elektrotechnice. SNR je definován jako poměr výkonu užitečného signálu (řečového) a výkonu neužitečného signálu (šum). Čím je odstup vyšší, tím je signál kvalitnější (obsahuje méně šumu). Jednotkou poměru signálu od šumu je decibel [dB]. Hlavní výhodou metod SNR je, že jsou výpočetně jednoduché.



Obrázek 3.2: Časový průběh nezarušeného řečového signálu.



Obrázek 3.3: Časový průběh zarušeného řečového signálu Gaussovým šumem.

3.1.1 Základní metoda SNR

SNR není příliš vhodné pro určení kvality řečového signálu. Výsledky se příliš neshodují s výsledky dosaženými subjektivním měřením. Hlavním důvodem je, že řečový signál je nestacionární.

Chybový signál je definován jako:

$$\varepsilon[n] = s[n] - \hat{s}[n] \quad (3.1)$$

Kde: $\varepsilon[n]$ chybový signál,
 $s[n]$ řečový signál bez šumu,
 $\hat{s}[n]$ řečový signál s šumem.

Pro výpočet SNR je poté použit vzorec [2]:

$$SNR = 10 \log_{10} \frac{E_s}{E_\varepsilon} = 10 \log_{10} \frac{\sum_n s^2[n]}{\sum_n [s[n] - \hat{s}[n]]^2} \quad [\text{dB}] \quad (3.2)$$

Kde: E_s energie řečového signálu,
 E_ε energie chybového signálu,
 $s[n]$ řečový signál bez šumu,
 $\hat{s}[n]$ řečový signál s šumem.

3.1.2 Segmentální SNR

Řečový signál rozdělíme do M segmentů stejné délky. Délka segmentů je volena na 15 – 20 ms [1]. Potom je pro každý rámec vypočítáno SNR a výsledná hodnota obsahuje průměr všech hodnot SNR z rámců celého signálu. V některých případech může nastat problém s výsledkem SNR a to pokud se v řeči objeví tichá pauza. Potom výsledek nabývá záporných hodnot. Řešením je vyloučení tiché pauzy pomocí krátkodobého měření energie a porovnání hodnot s nastaveným prahem [1] [2].

Vzorec pro výpočet SNR_{seg} je následující:

$$SNR_{\text{seg}} = \frac{1}{M} \sum_{j=0}^{M-1} 10 \log_{10} \left[\frac{\sum_{n=m_j-N+1}^{m_j} s^2[n]}{\sum_{n=m_j-N+1}^{m_j} [s[n] - \hat{s}[n]]^2} \right] \quad [\text{dB}] \quad (3.3)$$

Kde: M počet segmentů,
 m konec předešlého segmentu.

3.1.3 SNR s rozdělením do kmitočtových pásem

Další snahou přiblížit objektivní parametr poměru signálu k šumu ke skutečnému procesu slyšení člověka, bylo vložit do výpočtu kritická pásma. Protože k vnímání hlasitosti přispívají všechna spektrální maxima jinou mírou, tak rozdělením do kritických pásem můžeme tuhle vlastnost lidského ucha simulovat. To se děje pomocí váhových koeficientů viz tabulka 3.1.

Studie ukazují, že výsledky získané pomocí této metody více korelují se subjektivními testy než výsledky získané pomocí SNR či SNR_{seg} .

SNR_{fw-seg} je formulován následujícím vztahem:

$$SNR_{fw-seg} = \frac{1}{M} \sum_{j=0}^{M-1} 10 \log_{10} \left[\frac{\sum_{k=1}^K w_j 10 \log_{10} \frac{F^2[j,k]}{(F[j,k] - \hat{F}[j,k])^2}}{\sum_{j=1}^K w_j} \right] \quad [\text{dB}] \quad (3.4)$$

Kde: K počet kritických pásem,
 M počet segmentů,
 $F[j, k]$ součet modulů v kritickém pásmu čisté řeči,
 $\hat{F}[j, k]$ součet modulů v kritickém pásmu řeči s šumem,
 w_j váhový koeficient.

Tabulka 3.1: váhové koeficienty w_j pro výpočet SNR_{fw-seg} [1].

číslo pásma	střední kmitočet pásma [Hz]	váha w_j
1	50	0,003
2	120	0,003
3	190	0,003
4	260	0,007
5	330	0,010
6	400	0,016
7	470	0,016
8	540	0,017
9	617	0,017
10	703	0,022
11	798	0,027
12	904	0,028
13	1020	0,030
14	1148	0,032
15	1288	0,034
16	1442	0,035
17	1610	0,037
18	1794	0,036
19	1994	0,036
20	2221	0,033
21	2446	0,030
22	2701	0,029
23	2978	0,027
24	3276	0,026
25	3597	0,026

3.1.4 Segmentální SNR s využitím VAD

Tato metoda patří do jednokanálových objektivních metod oproti výše uvedeným metodám výpočtů SNR, které patřily do metod dvoukanálových. K dispozici máme pouze degradovaný či zvýrazněný řečový signál, tudíž musíme použít detektor řečové aktivity (VAD). Ten odhadne úseky, v kterém segmentu se vyskytuje řečový signál a v kterém ne. Pokud se v daném segmentu nachází řečová aktivita, potom výstupní hodnota z VAD je 1. Segmentaci volíme stejně jako u SNR_{seg} .

Energie řečového signálu v m -tém rámci:

$$E_y[m] = \sum_{n=mN}^{mN+N-1} y^2[n] \quad (3.5)$$

Průměrná energie šumu:

$$E_n = \frac{\sum_{m=0}^{M-1} (1 - f_{vad}[m]) \cdot E_y[m]}{\sum_{m=0}^{M-1} (1 - f_{vad}[m])} \quad (3.6)$$

Energie řečového signálu se šumem:

$$E_x = \frac{\sum_{m=0}^{M-1} f_{vad}[m] \cdot E_y[m]}{\sum_{m=0}^{M-1} f_{vad}[m]} \quad (3.7)$$

Odhad energie zvýrazněné řeči:

$$E_s = E_x - E_n \quad (3.8)$$

Výpočet SNR:

$$SNR_{vad} = 10 \log_{10} \frac{E_s}{E_n} \quad [\text{dB}] \quad (3.9)$$

Kde: M počet segmentů,
 N délka segmentu,
 m pořadové číslo segmentu,
 $y[n]$ řečový signál,
 $f_{vad}[m]$ hodnota z VAD, která udává, jestli se jedná o segment s řečovou aktivitou či nikoliv.

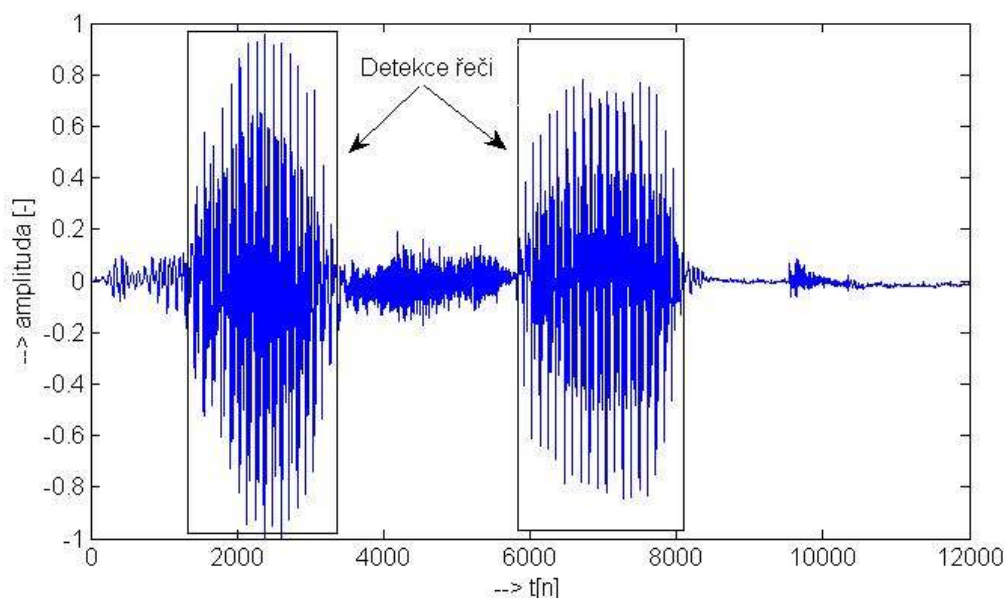
3.2 Detektor řečové aktivity

Detekce řeči je významnou součástí mnoha aplikací pro zpracování řeči. Nachází využití v systémech pro zvýrazňování řeči, ve vokodéru pro přenos pouze řečového signálu a také v řečových rozpoznávačích pro detekci začátku a konce promluvy. Pro přenos pouze řečových částí v paketových sítích (VoIP). Řečový signál rozdělíme do segmentů a v každém segmentu vypočítáme parametr a podle nastaveného prahu rozhodneme, jestli se v daném segmentu nachází řečová aktivita či nikoliv. Rozhodovací práh může být pevný nebo adaptivní podle toho, jaké vlastnosti má pozadí řeči.

Nejjednodušší výpočetní algoritmy jsou založené na rozhodování podle počtu průchodu nulou a na rozhodování podle energie obsažené v segmentu. Jejich nevýhodou je větší chybovost při detekci řeči v šumovém prostředí a při rozhodování podle energie potlačení hlásek s nízkou energií. Na druhou stranu je výhodou malá výpočetní náročnost.

Spolehlivější algoritmy pro detekci řečového signálu vycházejí ze spektrální vzdálenosti mezi řečovým signálem a pozadím řeči.

Pro detekci řečové aktivity ve velmi zarušeném prostředí se využívají vícekanálové metody. S jejich nasazením, ale zároveň vzrůstá výpočetní náročnost detekčního algoritmu a zvyšují se také hardwarové nároky na systém, protože je nutné mít nahrávku nahranou vícekanálově.



Obrázek 3.4: Detekce řeči v zarušené nahrávce.

3.3 Míry zkreslení založené na lineární prediktivní analýze

Míry zkreslení vyjadřují odlišnost dvou mikrosegmentů signálu řeči. Jeden mikrosegment je originální (referenční) a druhý testovaný. Označíme-li t a r jako vektory příznaků dvou mikrosegmentů, pak lze míru zkreslení označit výrazem $d(t, r)$. Jsou-li splněny podmínky symetrie a pozitivní definitnosti, poté $d(t, r)$ označujeme jako lokální vzdálenost nebo míra vzdálenosti mezi mikrosegmentem t a r . Pokud výše uvedené podmínky nejsou splněny, tak o $d(t, r)$ hovoříme jako o míře zkreslení. Abychom mohli míry vzdálenosti nebo míry zkreslení využívat i v reálném čase je nutné zabezpečit, aby jejich vyčíslení bylo dosaženo v co možná nejmenším množství operací.

3.3.1 Lineární prediktivní analýza

Pomocí této metody odhadujeme parametry modelu vytváření řeči, který byl popsán v kapitole 1.2. Tyto parametry jsou platné pro jeden mikrosegment délky 10 – 30 ms a získávají se přímo z akustického signálu. Lineární prediktivní analýza se též označuje jako lineární prediktivní kódování z anglické zkratky LPC (Linear Predictive Coding). Při řešení soustavy rovnic (3.15) lze aplikovat dva přístupy a to autokorelační a kovarianční. Dále se budeme zabývat pouze autokorelační metodou, která je převážně využívána při zpracování řeči. U autokorelační metody předpokládáme, že signál je roven nule mimo příslušný mikrosegment. Bližší informace ke kovariantnímu přístupu lze najít v literatuře [2]. Metoda LPC je založena na předpokladu, že n -tý vzorek signálu $s[n]$ lze popsat lineární kombinací Q předchozích vzorků a buzení $u[n]$. Tento předpoklad je reprezentován vzorcem:

$$s[n] = -\sum_{i=1}^Q a_i s[n-i] + Gu[n] \quad (3.10)$$

Kde: Q řád modelu (většinou se volí 7 až 20),
 G koeficient zesílení.

Pokud v rovnici (3.10) neznáme člen $Gu[n]$ vzniká chyba predikce mezi skutečnou hodnotou $s[n]$ a predikovanou hodnotou $\hat{s}[n]$. Potom lze krátkodobou energii chyby signálu zapsat:

$$E = \sum_k \{s[n] - \hat{s}[n]\}^2 = \sum_k \{s[n] + \sum_{i=1}^Q a_i s[n-i]\}^2 \quad (3.11)$$

Krátkodobá energie chyby signálu má minimum v bodě:

$$\frac{\partial E}{\partial a_j} = 0, \quad 1 \leq j \leq Q. \quad (3.12)$$

Řešením rovnice (3.12) dostaneme soustavu rovnic:

$$\sum_{i=1}^Q a_i \sum_k s[n-j]s[n-i] = -\sum_k s[n-j]s[n] \quad (3.13)$$

Kde: $1 \leq j \leq Q$.

Jak bylo uvedeno výše u autokorelačního přístupu předpokládáme, že mimo mikrosegment je akustický signál roven nule. Když označíme délku mikrosegmentu jako N , potom bude akustický signál $s[n]$ nenulový na intervalu $0 \leq n \leq N-1$ a vzorec (3.13) lze vyjádřit jako:

$$\sum_k s_n[n-j]s_n[n-i] = \sum_{k=0}^{N-1-(j-i)} s_n[n]s_n[n+j-i] \quad (3.14)$$

Kde: $1 \leq j \leq Q$,
 $0 \leq i \leq Q$.

V rovnici (3.14) lze nalézt krátkodobou autokorelační funkci R_n a lze ji přepsat do následujícího tvaru:

$$\sum_k s_n[n-j]s_n[n-i] = R_n[j-i] \quad (3.15)$$

Jestliže bude krátkodobá autokorelační funkce $R_n[i]$ sudá funkce, poté lze rovnici (3.13) přepsat do tvaru:

$$\sum_{i=1}^Q a_i R_n[|j-i|] = -R_n[j] \quad (3.16)$$

Kde: $1 \leq j \leq Q$.

Rovnici (3.16) lze přepsat též do maticového tvaru, kde matice bude symetrická a v Töplitzově tvaru, pak lze pro výpočet koeficientů a_i použít vysoce účinný Levinsonův – Durbinův algoritmus [7]:

$$\begin{aligned} E_n^{(0)} &= R_n(0) \\ k_i &= -\left\{R_n[i] + \sum_{j=1}^{i-1} a_j^{(i-1)} R_n[i-j]\right\} / E_n^{(i-1)} \\ a_i^{(i)} &= k_i \\ a_j^{(i)} &= a_j^{(i-1)} + k_i a_{i-j}^{(i-1)} \\ E_n^{(i)} &= (1 - k_i^2) E_n^{(i-1)} \end{aligned} \quad (3.17)$$

Kde: $1 \leq j \leq i-1$

$i = 1, 2, \dots, Q$

$a_j^{(i)}$ j -tý parametr predátoru řádu i ,

k_i koeficient odrazu (PARCOR).

Pro výpočet zesílení G lze za předpokladu, že budící funkce má tvar jednotkového impulsu nebo se jedná o bílý šum odvodit vztah:

$$G^2 = R_n[0] + \sum_{i=1}^Q a_i R_n[i] = E_n \quad (3.18)$$

Kde: G parametr zesílení,
 R_n krátkodobá autokorelační funkce,
 E_n chyba predikce.

Pokud provedeme v přenosové funkci $H(z)$ substituci $z \rightarrow e^{j\omega}$ můžeme vypočítat též spektrum signálu, které bude mít charakter vyhlazené spektrální obálky oproti spektru získanému pomocí FFT, které obsahuje informace o hlasovém traktu i o budící funkci. Úhlovou frekvenci dostaneme z následujícího vztahu:

$$\omega = \frac{2\pi f}{F_v} \quad (3.19)$$

Kde: f dosazovaná frekvence, která musí respektovat vzorkovací teorém,
 F_v vzorkovací frekvence.

3.3.2 Itakurova míra

Itakurova míra je nesymetrická, tudíž nepředstavuje vzdálenost, ale míru zkreslení mezi testovaným a referenčním mikrosegmentem. Tyto mikrosegmenty jsou vyjádřeny reprezentacemi LPC. Itakurovu míru lze zapsat jako [7]:

$$d_{LLR}(\mathbf{t}, \mathbf{r}) = \log \left(\frac{\mathbf{a}_r^T \mathbf{R}_t \mathbf{a}_r}{\mathbf{a}_t^T \mathbf{R}_t \mathbf{a}_t} \right) \quad (3.20)$$

Kde: $\mathbf{a}_r$ vektor koeficientů LPC referenčního mikrosegmentu,
 $\mathbf{a}_t$ vektor koeficientů LPC testovaného mikrosegmentu,
 $\mathbf{R}_t$ autokorelační matice v Töplitzově tvaru,
 $\mathbf{a}_r^T$ transponovaný vektor koeficientů LPC referenčního mikrosegmentu,
 $\mathbf{a}_t^T$ transponovaný vektor koeficientů LPC testovaného mikrosegmentu.

LPC koeficienty $\mathbf{a}_r$ a $\mathbf{a}_t$ jsou sloupcové vektory, potom transponované vektory budou řádkového tvaru a autokorelační matice má rozměry $(Q + 1) \times (Q + 1)$.

Itakurova míra je založená na využití zbytkové chyby predikce. Úplná kvadratická chyba je dána vztahem:

$$E = \sum_{k=-\infty}^{\infty} \left\{ \sum_{i=0}^Q a_i s[n-i] \right\}^2 \quad (3.21)$$

Protože předpokládáme, že mimo aktuální mikrosegment je signál nulový, můžeme rovnici (3.21) upravit pro výpočet jednoho mikrosegmentu na následující vztah:

$$E = \sum_{k=0}^{N+Q-1} \left\{ \sum_{i=0}^Q a_i s[n-i] \right\}^2 = \mathbf{a}^T \mathbf{R} \mathbf{a} \quad (3.22)$$

Kvadratickou chybu E můžeme také uvažovat jako výstup filtru $A(z)$, kde:

$$A(z) = 1 + \sum_{i=1}^Q a_i z^{-i} \quad (3.23)$$

Pokud tímto filtrem projde signál $s_t[n]$ na jehož základě byl navržen, potom chyba predikce bude minimální a lze ji vyčíslit pomocí vztahu:

$$E_t = \mathbf{a}_t^T \mathbf{R}_t \mathbf{a}_t \quad (3.24)$$

Naopak pokud filtrem $A_r(z)$, který byl navržen na základě vzorků $s_r[n]$ projde posloupnost vzorků $s_t[n]$ bude zbytková energie β větší nebo rovna zbytkové energii E_t a lze ji vyjádřit vztahem:

$$\beta = \sum_{k=0}^{N+Q-1} \left[\sum_{i=0}^Q a_{ri} s(k-i) \right]^2 = \mathbf{a}_r^T \mathbf{R}_t \mathbf{a}_r \geq E_t \quad (3.25)$$

Pokud se nyní podíváme na základní vztah (3.20) zjistíme, že Itakurova míra je poměr β/E_t , který vyjadřuje míru zkreslení mezi referenčním a testovaným modelem. Pokud jsou oba mikrosegmenty stejné, pak poměr $\beta/E_t = 1$ a logaritmus z jedné je nula. Pro výpočet LPC koeficientů referenčního a testovaného mikrosegmentu, autokorelační matice $\mathbf{R}_t$ se nejčastěji využívá Durbinova algoritmu popřípadě Levinsonova – Durbinova algoritmu. Výpočet Itakurovy míry podle vzorce (3.20) je poměrně výpočetně náročný pro aplikace, které kladou důraz na výpočty v reálném čase. V literatuře [7] lze najít vzorec, který vyžaduje méně výpočetních kroků a je pro tyto aplikace vhodnější.

Na Itakurovu míru d_{LLR} se lze také dívat jako na míru spektrální a bylo odvozeno, že ji lze zapsat i v následujícím tvaru:

$$d_{LLR}(\mathbf{t}, \mathbf{r}) = \log \left(\int_{-\pi}^{\pi} \left| \frac{A_r(e^{j\omega})}{A_t(e^{j\omega})} \right|^2 \frac{d\omega}{2\pi} \right) \quad (3.26)$$

3.3.3 Itakurova – Saitova míra

Itakurova – Saitova míra je spektrální nesymetrická míra tudíž platí, že $d_{IS}(\mathbf{t}, \mathbf{r}) \neq d_{IS}(\mathbf{r}, \mathbf{t})$, a je definována vztahem [7]:

$$d_{IS}(\mathbf{t}, \mathbf{r}) = \int_{-\pi}^{\pi} |e^{V(\omega)} - V(\omega) - 1| \frac{d\omega}{2\pi} \quad (3.27)$$

A funkci $V(\omega)$ lze definovat ve tvaru:

$$V(\omega) = \log \left(\frac{G_t^2}{|A_t(e^{j\omega})|^2} \right) - \log \left(\frac{G_r^2}{|A_r(e^{j\omega})|^2} \right) \quad (3.28)$$

Kde: G_t koeficient zesílení pro testovaný mikrosegment,
 G_r koeficient zesílení pro referenční mikrosegment,
 $A_t(e^{j\omega})$ spektrální koeficient filtru pro testovaný mikrosegment,
 $A_r(e^{j\omega})$ spektrální koeficient filtru pro referenční mikrosegment.

Bylo odvozeno, že vzorec (3.27) lze převést do tvaru, kde jsou využity LPC koeficienty, autokorelační matice $\mathbf{R}_t$ a koeficienty zesílení. Pro výpočet všech potřebných proměnných je vhodné využít Levinsonův – Durbinův algoritmus stejně jako u Itakurovy míry. Itakurova – Saitova míra lze vyčíslit pomocí následujícího vztahu:

$$d_{IS}(\mathbf{t}, \mathbf{r}) = \frac{G_t^2 \mathbf{a}_r^T \mathbf{R}_t \mathbf{a}_r}{G_r^2 \mathbf{a}_t^T \mathbf{R}_t \mathbf{a}_t} + \log \frac{G_r^2}{G_t^2} - 1 \quad (3.29)$$

Kde: G_t koeficient zesílení pro testovaný mikrosegment,
 G_r koeficient zesílení pro referenční mikrosegment,
 $\mathbf{a}_r$ vektor koeficientů LPC referenčního mikrosegmentu,
 $\mathbf{a}_t$ vektor koeficientů LPC testovaného mikrosegmentu ,
 $\mathbf{R}_t$ autokorelační matice v Töplitzově tvaru,
 $\mathbf{a}_r^T$ transponovaný vektor koeficientů LPC referenčního mikrosegmentu ,
 $\mathbf{a}_t^T$ transponovaný vektor koeficientů LPC testovaného mikrosegmentu.

3.4 PESQ

Tento standard se řadí do metod intrusivních a je znám pod pojmem Perceptual Evaluation of Speech Quality (PESQ). Je to výsledek několika let výzkumu ITU-T Study Group 12 a je použitelný nejenom pro řečové kodeky, ale také pro end-to-end měření. PESQ je vylepšení předešlé metody PSQM, která je použitelná pouze pro ohodnocení řečových kodeků a nepočítá s proměnným zpožděním, filtrací a ztrátou paketů. Metoda PESQ výše uvedené jevy uvažuje a proto je vhodná pro měření reálných přenosových systémů.

3.4.1 Doporučené vlastnosti testovaných nahrávek

Pro určení kvality kodeku je důležité, aby testovací signály použité pro PESQ byly reprezentovány reálnými signály, které prošly komunikační sítí. Síť se může chovat k řeči a k tichu rozdílně a kódovací algoritmus je většinou optimalizován pro řeč a to může dávat méněcenné výsledky, když testovací nahrávky neobsahují časové a kmitočtové vlastnosti řeči. Použití umělých řečových nahrávek je doporučeno jen tehdy, když reprezentují časovou a fonetickou strukturu reálného řečového signálu.

- Časové vlastnosti řečových nahrávek
Řečová nahrávka musí obsahovat řečovou aktivitu i tiché pauzy aby reprezentovala přírodní řeč. Řeč by měla být aktivní mezi 40 % až 80 % času. Samozřejmě tahle doba závisí na jazyku nahrávky. Pokud chceme testovat delší nahrávky je vhodné je rozdělit na více dílů po 8 s až 20 s.
- Filtrování a kalibrace úrovně
Signály by měli projít přes filtr s příslušnou frekvenční charakteristikou pro simulaci stejné kmitočtové charakteristiky jako telefonního zařízení. Tím dojde k přizpůsobení úrovně na různých kmitočtech jako při běžném telefonním hovoru. Způsob filtrace je popsán v doporučení [6].
- Přidávání šumu
Je možné použít PESQ pro určení vlivu šumu pozadí na kvalitu nahrávky. Přidaný šum by měl projít přes podobný filtr definovaný v [6]. To je důležité zejména pro nízkofrekvenční šum, který by byl utlumen filtrem v telefonním zařízení. Šum by měl být přidán k degradované nahrávce ještě než projde přenosovým systémem. Originální nahrávka zůstává nazarušená.

3.4.2 Popis algoritmu PESQ

PESQ porovnává originální signál se signálem degradovaným, který je výsledkem průchodu originálního signálu přes komunikační systém. Výstup PESQ je odhad kvality která by byla přiřazená degradované nahrávce pomocí subjektivních testů. Celý algoritmus je složen z několika kroků podrobně popsanych v [3]:

- Vyrovnání hlasitosti
Pro další výpočty je důležité, aby originální a degradovaný signál měl stejnou úroveň. Tato úroveň je 79 dB_{SPL}
- IRS filtrace
IRS filtrace zohledňuje frekvenční vlastnosti telefonního zařízení, které jsou blíže definované v [5].
- Časové zarovnání
Provádí se proto, aby pro další výpočty si odpovídaly části originálního a degradovaného signálu. Při hrubém zarovnání je vypočítaná obálka výkonu originálního a degradovaného signálu.

$$X_{ES}(t)_k = \log_{10} \frac{\text{MAX}\{E(k)\}}{E} \quad (2.10)$$

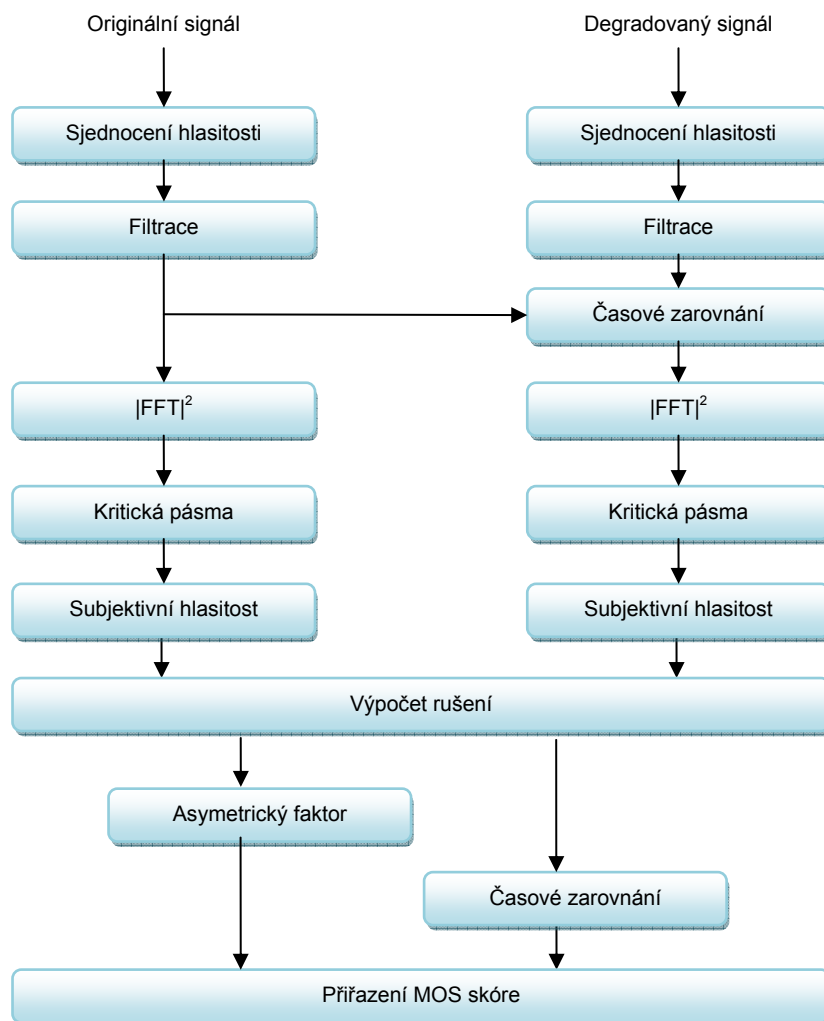
Kde: k číslo rámce
 $E(k)$ energie v rámci
 E prahová energie daná VAD

Po té je provedena korelace originálního a degradovaného signálu. Rozlišení hrubého zarovnání jsou 4 ms, to je dáno délkou rámce.

U jemného zarovnání se změni délka rámce z 4 ms na 64 ms a vynásobí se Hannovým oknem s přesahem 75 %. Po té je znovu provedena korelace. Histogram hodnot korelace je vyhlazen konvolucí s trojúhelníkovým signálem šířky 1 ms.

- FFT
Přechod z časové oblasti se děje pomocí krátkodobé FFT s použitím Hannova okna o délce 32 ms a přesahem 50 %. Pro vzorkovací frekvenci 8 kHz odpovídá 256 vzorků na okno a pro 16 kHz 512 vzorků. Výkonové spektrum je potom vypočteno jako kvadrát modulu hodnot získaných z FFT
- Přizpůsobení vnímání lidského ucha
V prvním kroku implementujeme kritická pásma. Po této implementaci se změni jednotka z Hertzů na Barky. V dalším kroku je spektrální hustota výkonu převedena na hustotu subjektivní hlasitosti.

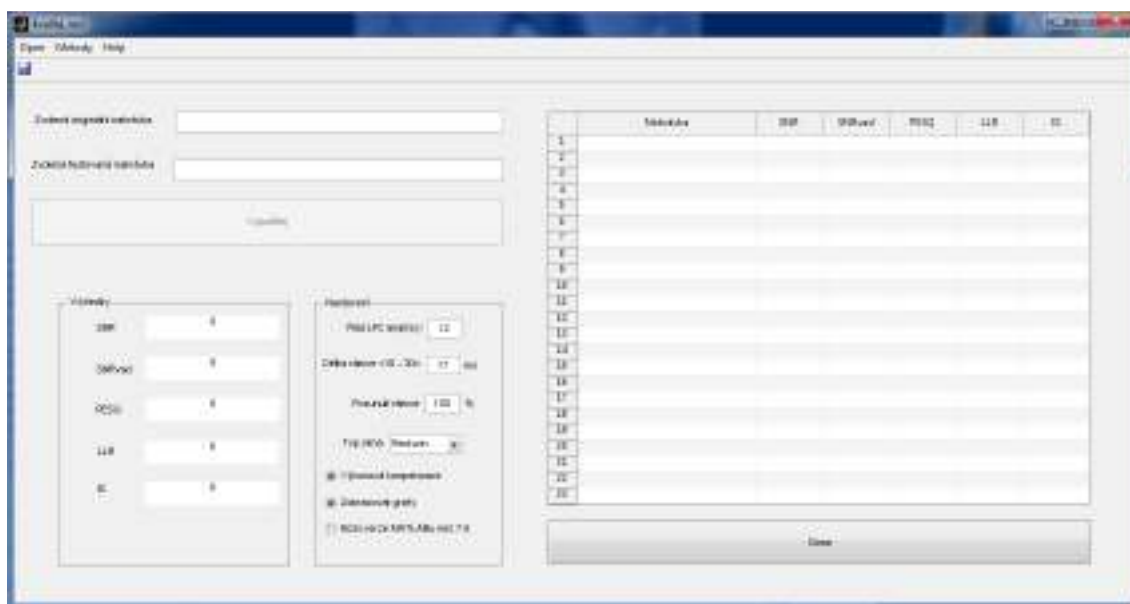
- **Výpočet rušení**
Je vypočítán rozdíl hlasitostí v každém rámci mezi degradovaným a originálním signálem. Pokud je výsledek kladný, tak se k originálnímu signálu během přenosu přidal šum. Záporný výsledek zase znamená, že degradovaný signál je zeslaben. Tento výsledek je označen jako hrubý výpočet rušení. Na tyto hodnoty je poté aplikováno maskovací pole, které simuluje jev maskování.
- **Asymetrický faktor**
Při průchodu kodekem se změní některé frekvenční složky, proto se použije asymetrický faktor, který tyto změny eliminuje
- **Výsledná hodnota**
Výsledná hodnota je vypočítána z průměrných hodnot rušení a rušení s uvážením asymetrického faktoru a je v rozmezí $-0,5 \div 4,5$ ale ve většině případů se výsledek pohybuje mezi $1 \div 4,5$.



Obrázek 3.5: Schéma metody PESQ.

4 Vypracování

Výsledkem bakalářské práce je program `kvalita_reci` vytvořený v programovacím jazyce MATLAB. Program využívá grafického prostředí GUI (Graphical user interface), kde uživatel nemusí vytvářet žádné skripty nebo používat příkazovou řádku. Program ovládá v tomto prostředí a není zapotřebí mít znalosti programovacího jazyka MATLAB. Výsledný program v základní konfiguraci po spuštění je zobrazen na obrázku 4.1.



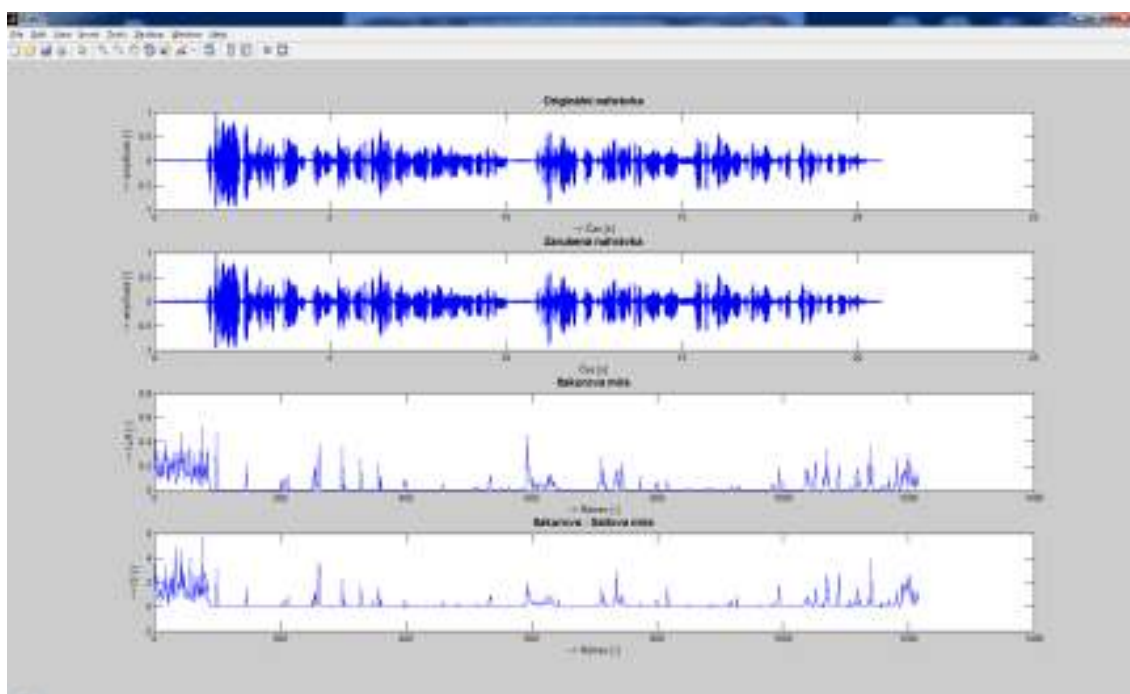
Obrázek 4.1: Zobrazení programu `kvalita_reci` v základní konfiguraci ihned po spuštění.

Podmínkou pro zpřístupnění tlačítka Vypočítej, je otevření originální a testovací nahrávky v menu Open. Druhou a poslední podmínkou je, že nahrávky musí mít stejnou délku. V menu Metody lze vybrat metody, které budou vypočítány, po stisknutí tlačítka Vypočítej. Lze vybrat následujících metod:

- SNR Základní metoda SNR (poměr signálu od šumu)
- SNRvad Segmentální SNR s využitím detektoru řečové aktivity
- PESQ Metoda PESQ
- LLR Itakurova míra
- IS Itakurova – Saitova míra

V panelu Nastavení si může uživatel zvolit řád LPC analýzy Q , která se používá pro výpočty Itakurovy míry a Itakurovy – Saitovy míry. Dále je možné si zvolit délku rámce v milisekundách. Pro krátkodobý model by měla být délka rámce 10 až 30 ms, ale v programu není nijak omezená stejně jako možnost posunutí rámce. Posunutí rámce by se mělo pohybovat mezi 1 až 100 %. Pro

segmentaci dat byla použita funkce `segmentation.m` jejímž autorem je Ing. Petr Sysel, Ph.D. Další položkou v nastavení je typ okna. Uživatel si zde může vybrat obdélníkové, Hammingovo, Bartlettovo, Blackmanovo, Hannovo či Kaiserovo okno. Zvolené okno je aplikováno na každý segment originální i zarušené nahrávky. Při vybrané položce Výkonová kompenzace se volá funkce `výkonová_kompence.m` a jejím výstupem je upravené okénko tak, aby plocha okénka byla stejná jako u okénka obdélníkového. Při vybrané možnosti Zobrazovat grafy se po výpočtu všech zvolených metod zobrazí nové okno, zobrazené na obrázku 4.2, ve kterém je vykreslen časový průběh originální a testované nahrávky, dále Itakurova a Itakurova – Saitova míra v každém rámci. Uživatel si může každý průběh přiblížit nebo oddálit pomocí symbolu lupy.



Obrázek 4.2: Zobrazení grafů pro nahrávku `orig_lednicka26db.wav`.

Položka nižší verze MATLABu než 7.6 je poslední položkou v panelu nastavení. Od verze 7.6 nabízí program MATLAB komponentu `table`, která není viditelná pro nižší verze. Při ukládání vypočtených hodnot se berou data právě z komponenty `table`. Výsledky je možné uložit ve formátu `*.xls` nebo `*.mat`. V prvním sloupci je uložen název testované nahrávky, v dalších sloupcích jsou výsledky hodnot v následujícím pořadí: Základní metoda SNR, Segmentální SNR s využitím VAD, PESQ, Itakurova míra a Itakurova – Saitova míra. Pokud je vybraná položka nižší verze MATLABu než 7.6 ukládání dat se realizuje z edit boxů a data je možné uložit pouze do formátu `*.mat`. První sloupec zde už neobsahuje název testované nahrávky, protože je zbytečný a můžeme si

uložená data pojmenovat po názvu nahrávky. Je možné uložit akorát jednu aktuální nahrávku, u které jsou zobrazeny výsledky v edit boxech. Výsledky jsou uloženy v řádkovém vektoru a pořadí je následující: Základní metoda SNR, Segmentální SNR s využitím VAD, PESQ, Itakurova míra a Itakurova – Saitova míra. Informace pro ukládání dat jsme získali z literatury [10]. Poslední možností v menu je položka Help, kde je uživateli stručně vysvětleno, jak se program ovládá a uveden odkaz bakalářské práce pro nalezení bližších informací.



Obrázek 4.3: Zobrazení nápovědy pro program kvalita_reci.

Všechny výpočty se realizují po stisknutí tlačítka Vypočítej. Pro určení základního poměru signálu od šumu (SNR) byl použit vzorec (3.2) ve funkci `snrzakl.m`.

Metoda Segmentální SNR s využitím VAD je realizovaná pomocí funkce `snr_vad.m`. V prvním kroku je odhadnuta řečová aktivita v každém rámci pomocí funkce `vad_energy.m` jejímž autorem je Ing. Petr Sysel, Ph.D. VAD se rozhoduje na základě krátkodobé řečové energie, která je použita pro další výpočet ve vzorcích (3.8) a (3.9). Druhým krokem je aplikace vzorců (3.8) a (3.9) ve funkci `SNRdata.m`.

Pro výpočet PESQ byl použit program `pesq.m`, který je součástí balíčku `composite.zip`, je volně šiřitelný a autorem je Dr. Philip C. Loizou.

Itakurova míra je vypočítána pomocí funkce `Itakura_measure.m`. V této funkci je aplikován vzorec (3.20). Funkce vrací hodnotu Itakurovy míry pro jednotlivé rámce a průměrnou Itakurovu míru pro celou nahrávku.

Itakurova – Saitova míra je vypočítána pomocí funkce `Itakura_Saito.m`. Výsledků je dosaženo aplikováním vzorce (3.29) s uvážením vzorce (3.18), kde jsme použili rovnou chybu predikce. Stejně jako u Itakurovy míry funkce vrací hodnoty pro jednotlivé rámce a průměrnou hodnotu pro celou nahrávku.

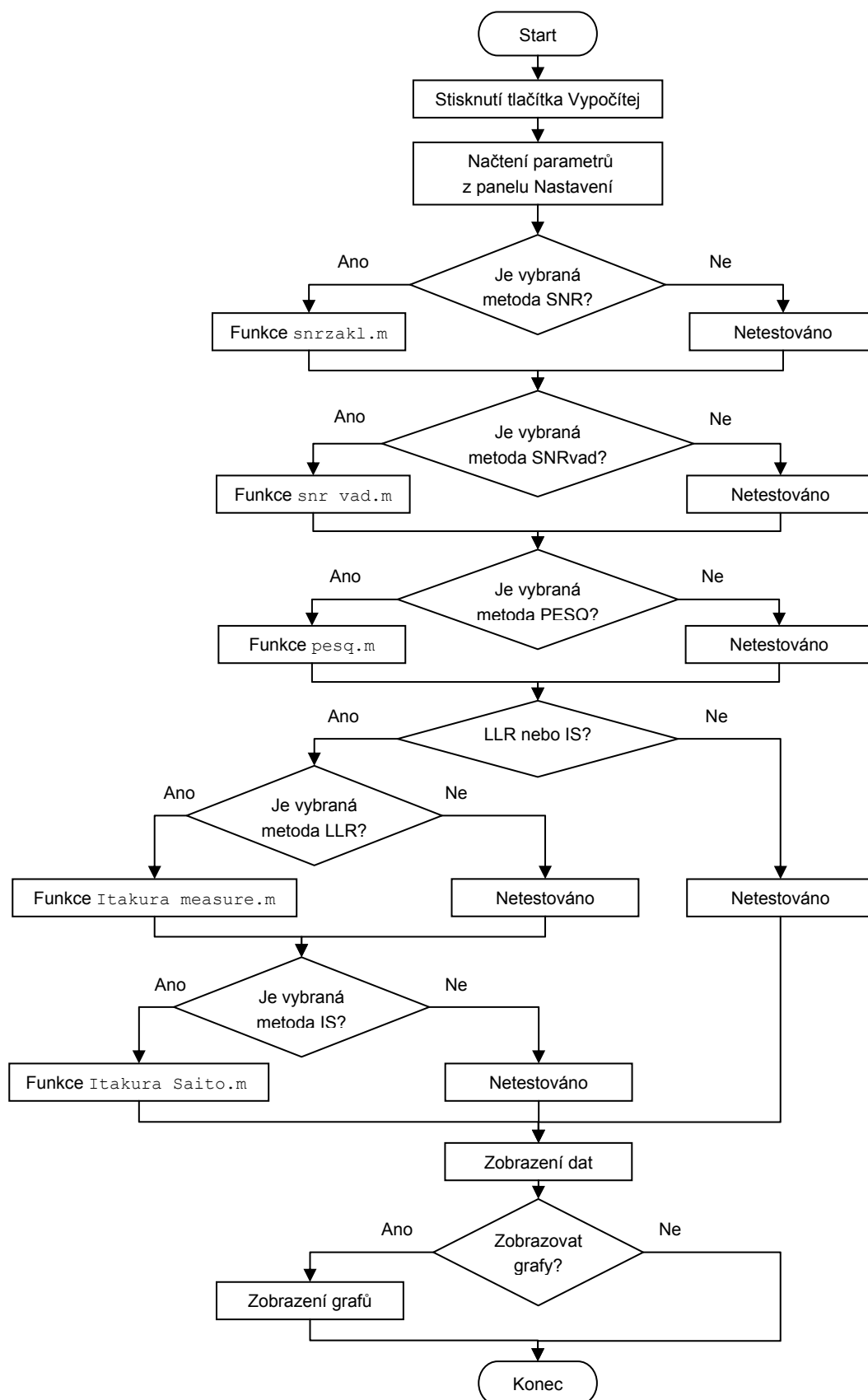
Vstupní proměnné pro Itakurovu a Itakurovu – Saitovu míru jsou získány pomocí funkce `LPC`, která je součástí programu MATLAB. Jedná se o koeficienty a_r , a_t a chyby predikce pro originální a testovací nahrávku. Vývojový diagram výpočtů po stisknutí tlačítka Vypočítej je zobrazen na obrázku 4.4.

Originální nahrávka je pořízena Mgr. Pavlem Rajmicem, Ph.D. Obsahem nahrávky je „Rock and roll a realita nikdy nebyly velcí kamarádi. Pro lidi určitého věku je také dráha rockové hvězdy možná tou nejoblíbenější únikovou fantazií. Jako by nabízela možnost přemístit se do nějakého alternativního vesmíru, kde neplatí žádná ubíjející pravidla včetně zákonitostí plynutí času“. K této originální nahrávce je postupně přidáván šum pomocí programu `Colea`. Autorem programu je Dr. Philip C. Loizou. Byl použit Gaussův šum, hluk vysílačky, hluk ledničky, škrábání a zvuk o frekvenci 1 kHz. Všechny tyto šумы byly upraveny, tak, aby měly stejnou délku jako originální signál. Je to z důvodu, abychom dostali objektivní výsledek pomocí metody PESQ, protože kdybychom měli nahrávku jinak dlouhou než originální, tak by bylo do výsledku započítáno i časové posunutí. Toto posunutí by ale nebylo započítáno u ostatních metod a výsledky by byly neobjektivní.

Tabulky s výsledky jednotlivých nahrávek ohodnocených programem `kvalita_reci` jsou uvedeny v příloze A. Výsledky jsou platné pro nastavené parametry uvedené v tabulce 4.1.

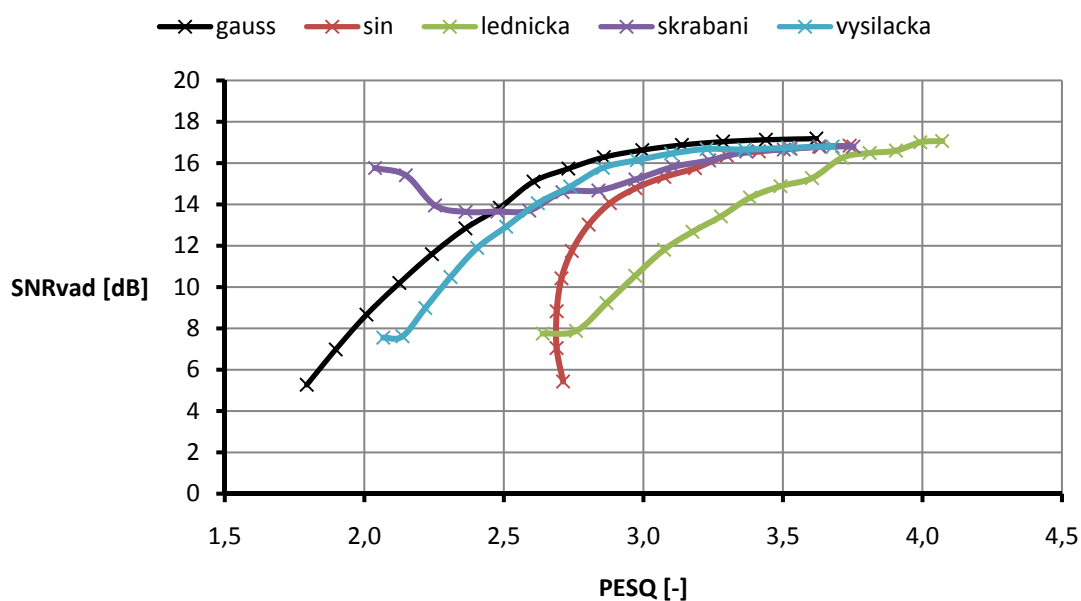
Tabulka 4.1: Nastavené parametry pro hodnocení nahrávek programem `kvalita_reci`.

Řád LPC analýzy Q [-]	12
Délka rámce [ms]	17
Posunutí rámce [%]	100
Typ okna	Rectwin
Výkonová kompenzace	Ano

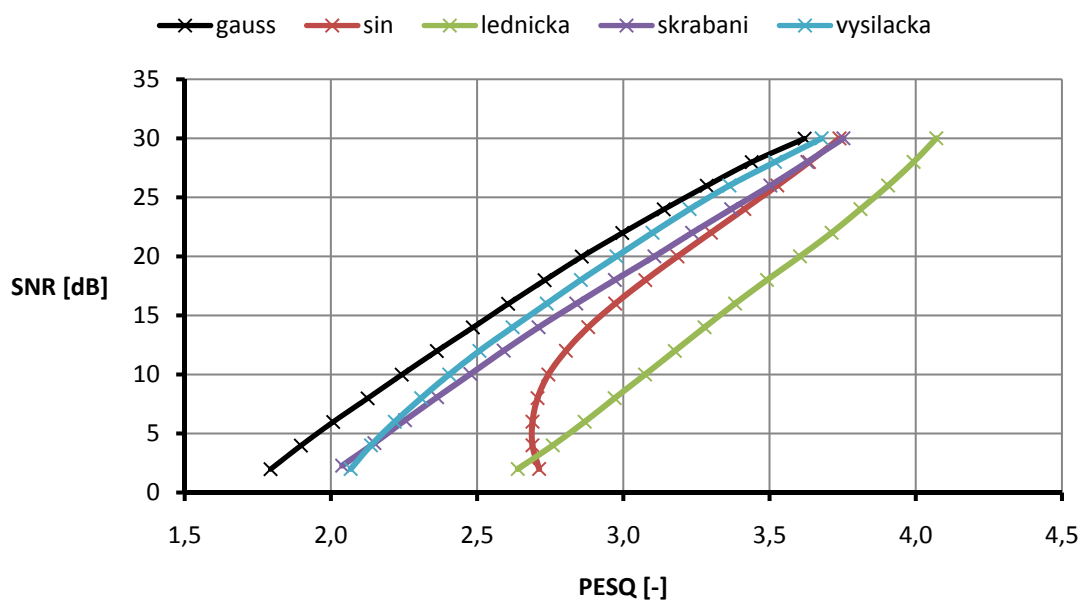


Obrázek 4.4: Zobrazení vývojového algoritmu při stisknutí tlačítka Vypočítej.

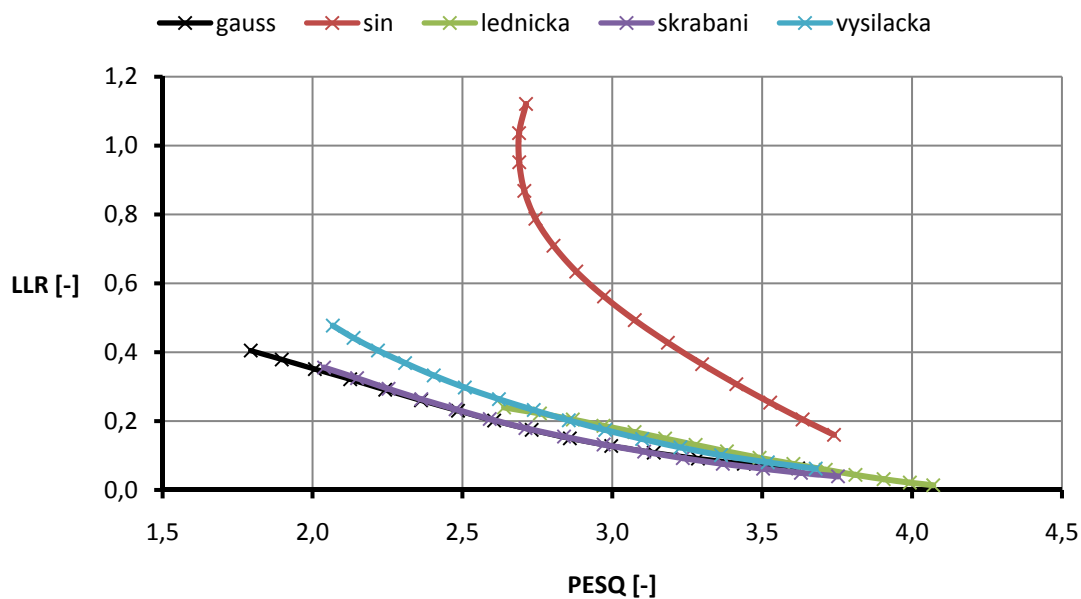
4.1 Grafické závislosti jednotlivých metod



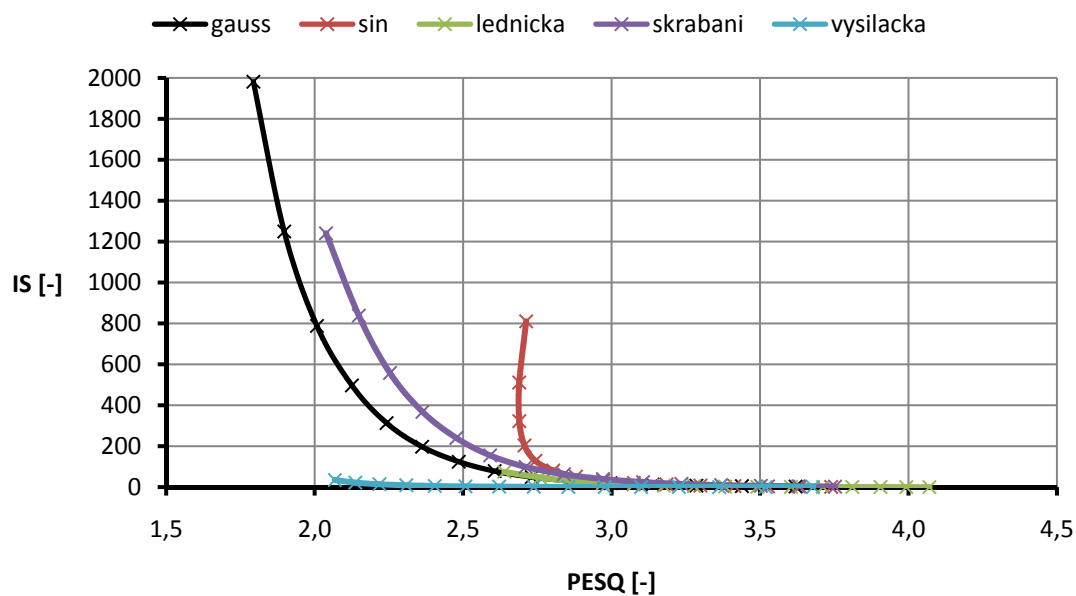
Obrázek 4.5: Porovnání závislosti poměru signálu od šumu s využitím VAD na hodnocení kvality metodou PESQ.



Obrázek 4.6: Porovnání závislosti klasického poměru signálu od šumu na hodnocení kvality metodou PESQ.



Obrázek 4.7: Porovnání závislostí Itakurovy míry na hodnocení kvality metodou PESQ.



Obrázek 4.8: Porovnání závislostí Itakurovy - Saitovy míry na hodnocení kvality metodou PESQ.

5 Závěr

Cílem bakalářské práce bylo implementování vybraných metod objektivního hodnocení kvality řeči v programovacím jazyku MATLAB. Z objektivních metod byla vybrána Základní metoda SNR, Segmentální SNR s využitím VAD, PESQ, Itakurova míra a Itakurova – Saitova míra. V testovaných nahrávkách bylo použito celkem pět druhů šumu, z toho čtyři byly z reálného prostředí a jeden vytvořen pomocí programovacího jazyku MATLAB.

Jako referenční metoda, oproti které se srovnávají další metody, byla zvolena metoda PESQ a to z důvodu, že se jedná o metodu, která nejvíce koreluje s metodami subjektivními. Srovnání výsledků je zobrazeno na obrázcích 4.5, 4.6, 4.7, 4.8. Z těchto závislostí je patrné, že hodnotu PESQ nelze přímo určit ani z jedné metody. Výsledná hodnota PESQ je silně závislá na druhu šumu obsaženého v nahrávce. Je tomu zejména proto, že metoda PESQ se snaží přiblížit vlastnostem lidského ucha. Zvuk se stejnou hladinou akustického tlaku, ale o různých frekvencích je vnímán jako různě hlasitý a přispívá různou mírou do hodnocení kvality řeči. Tento jev lze pozorovat v obrázku 4.5 na průbězích gauss (odpovídá nahrávkám orig_sum) a sin (odpovídá nahrávkám orig_sin). Frekvenční složky šumu nahrávky orig_sum jsou rozloženy téměř stejnoměrně v celém frekvenčním pásmu, za to šum nahrávky orig_sin je reprezentován pouze složkou o frekvenci 1 kHz. Na této frekvenci je lidské ucho nejcitlivější a to je důvod rozdílných průběhů sin ve všech obrázcích. Veliký rozdíl mezi výsledky Základní metody SNR a Segmentální SNR s využitím VAD je způsoben tím, že VAD vyhodnocuje segmenty pomocí krátkodobé energie a pro silněji zarušené nahrávky je krátkodobá energie šumu vyšší než krátkodobá energie neznělých hlásek. Potom se tyto segmenty vyhodnotí jako šumové a tím se snižuje výsledný poměr signálu od šumu. V obrázku 4.5 lze dokonce vidět, že se snižující kvalitou nahrávky poměr signálu od šumu s využitím VAD roste. Při porovnání metod Itakurovy míry a Itakurovy – Saitovy míry, lze vidět, že i když se jedná o metody, které vycházejí ze stejné lineární prediktivní analýzy, dosahují rozdílných výsledků. Itakurova – Saitova míra, která je oproti Itakurově míře závislá na koeficientech zesílení, dosahuje výrazně vyšších hodnot pro silně zarušené nahrávky než Itakurova míra. Závislosti obou metod jsou zobrazeny v obrázcích 4.7 a 4.8. Podle literatury dosahuje přesnějších výsledků Itakurova míra.

Literatura

- [1] SMÉKAL, Zdeněk Objektivní a subjektivní metody posuzování kvality a srozumitelnosti řeči . *Telekomunikace*. 2009, č. 3, s. 16-22. ISSN 0040-2591.
- [2] DELLER, John, HANSEN, John, PROAKIS, John. *Discrete-Time Processing of Speech Signals*. 2nd edition. New York : IEEE, 2000. ISBN 0-7803-5386-2.
- [3] ITU-T Rec. P.862, Series P: *Telephone transmission quality, telephone installations, local line networks. Perceptual Evaluation of Speech Quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band tele-phone networks and speech codecs*, 2001.
- [4] ITU-T Rec. P.800, Series P: *Telephone transmission quality. Methods for subjective determination of transmission quality*, 1996.
- [5] ITU-T Rec. P.48, Series P: *Specification for an intermediate reference system*, 1988.
- [6] ITU-T Rec. P.830, Series P: *Subjective performance assessment of telephone-band and wideband digital codecs*, 1996.
- [7] PSUTKA, Josef, MÜLLER, Luděk, MATOUŠEK, Jindřich, RADOVÁ, Vlasta. *Mluvíme s počítačem česky*. První vydání. Praha: Academia, 2006. ISBN 80-200-1309-1.
- [8] ČIHÁK, Radomír. *Anatomie 2*. Druhé vydání. Praha: Grada, 2002. ISBN 80-247-0143-2.
- [9] KRČMOVÁ, Marie. *Fonetika*. <<http://is.muni.cz/elportal/estud/ff/js07/fonetika/materialy/index.html>>.
- [10] DOŇAR, Bohuslav, ZAPLATÍLEK, Karel. *MATLAB – Začínáme se signály*. První vydání. Praha: BEN, 2006. ISBN 80-7300-200-0.
- [11] FANT, Gunnar. *Acoustic Theory of Speech Production*. Revised edition. Hague: de Gruyter Mouton, 1970. ISBN 90-279-1600-4.

Seznam symbolů, veličin a zkratek

a	neznámý parametr
$\mathbf{a}_r$	vektor koeficientů LPC referenčního segmentu
$\mathbf{a}_t$	vektor koeficientů LPC testovaného segmentu
$\mathbf{a}_r^T$	transponovaný vektor koeficientů LPC referenčního segmentu
$\mathbf{a}_t^T$	transponovaný vektor koeficientů LPC testovaného segmentu
$A_t(e^{j\omega})$	spektrální koeficient filtru pro testovaný segment,
$A_r(e^{j\omega})$	spektrální koeficient filtru pro referenční segment.
α	neznámý parametr
b	koeficient, který udává kolikrát je T_0 delší než T
β	neznámý parametr
c	neznámý parametr filtru
$d_{IS}(\mathbf{t}, \mathbf{r})$	Itakurova – Saitova míra
$d_{LLR}(\mathbf{t}, \mathbf{r})$	Itakurova míra
E_s	energie řečového signálu
E_ε	energie chybového signálu
E_n	chyba predikce
$E(k)$	energie v rámci
E	prahová energie daná VAD
$\varepsilon[n]$	chybový signál
f	kmitočet
$F[j, k]$	součet modulů v kritickém pásmu čisté řeči
$\hat{F}[j, k]$	součet modulů v kritickém pásmu řeči s šumem
F_v	vzorkovací kmitočet
$f_{vad}[m]$	hodnota z VAD
F_0	kmitočet základního hlasivkového tónu
G	zesílení zdroje buzení
$G(z)$	přenosová funkce modelu hlasivek
G_t	koeficient zesílení pro testovaný mikrosegment
G_r	koeficient zesílení pro referenční mikrosegment
$H(z)$	přenosová funkce modelu produkce řeči
$L(z)$	přenosová funkce modelu vyzařování zvuku
LPC	lineární prediktivní analýza – linear predictive coding
M	počet segmentů
N	délka segmentu
$PESQ$	metoda $PESQ$ - Perceptual Evaluation of Speech Quality
Q	řád LPC modelu
R_n	krátkodobá autokorelační funkce

$\mathbf{R}_t$	autokorelační matice v Töplitzově tvaru
$s[n]$	řečový signál bez šumu
$\hat{s}[n]$	řečový signál s šumem
SNR	poměr signálu od šumu – signal to noise ratio
SNR_{fw-seg}	SNR s rozdělením do kmitočtových pásem
SNR_{seg}	segmenrální SNR
SNR_{vad}	segmenrální SNR s využitím VAD
T	perioda vzorkování
T_0	perioda základního hlasivkového tónu
VAD	detektor řečové aktivity – voice activity detection
$V(z)$	přenosová funkce modelu hlasového traktu
w_j	váhový koeficient

Seznam příloh

A	Vypočtené hodnoty programem kvalita_reci	45
A.1	Výsledky pro nahrávku zarušenou Gaussovým šumem	45
A.2	Výsledky pro nahrávku zarušenou šumem o kmitočtu 1 kHz	46
A.3	Výsledky pro nahrávku zarušenou hlukem ledničky	47
A.4	Výsledky pro nahrávku zarušenou škrabáním	48
A.5	Výsledky pro nahrávku zarušenou šumem vysílačky	49
B	Přiložené CD	50

A Vypočtené hodnoty programem kvalita_reci

A.1 Výsledky pro nahrávku zarušenou Gaussovým šumem

Název	SNR [dB]	SNR _{vad} [dB]	PESQ [-]	LLR [-]	IS [-]
orig_sum2db.wav	1,9809	5,2712	1,7939	0,4048	1981,2435
orig_sum4db.wav	3,9809	6,9798	1,8976	0,3789	1249,5737
orig_sum6db.wav	5,9809	8,6624	2,0077	0,3509	787,9686
orig_sum8db.wav	7,9809	10,1963	2,1258	0,3214	496,7688
orig_sum10db.wav	9,9809	11,5975	2,2424	0,2909	313,0763
orig_sum12db.wav	11,9809	12,8382	2,3621	0,2602	197,2206
orig_sum14db.wav	13,9809	13,8444	2,4855	0,2302	124,1622
orig_sum16db.wav	15,9809	15,1054	2,6062	0,2014	78,1034
orig_sum18db.wav	17,9809	15,7322	2,7309	0,1745	49,0783
orig_sum20db.wav	19,9809	16,2864	2,8583	0,1499	30,7958
orig_sum22db.wav	21,9809	16,6279	2,9963	0,1279	19,2885
orig_sum24db.wav	23,9809	16,8782	3,1384	0,1083	12,0544
orig_sum26db.wav	25,9809	17,0404	3,2857	0,0911	7,5114
orig_sum28db.wav	27,9809	17,1308	3,4391	0,0760	4,6650
orig_sum30db.wav	29,9809	17,1885	3,6200	0,0628	2,8856

A.2 Výsledky pro nahrávku zarušenou šumem o kmitočtu 1 kHz

Název	SNR [dB]	SNR _{vad} [dB]	PESQ [-]	LLR [-]	IS [-]
orig_sin2db.wav	2,0000	5,4295	2,7123	1,1209	810,2453
orig_sin4db.wav	3,9992	7,0498	2,6892	1,0363	511,1660
orig_sin6db.wav	5,9993	8,8316	2,6896	0,9518	322,3837
orig_sin8db.wav	8,0009	10,4246	2,7067	0,8685	203,2273
orig_sin10db.wav	9,9995	11,7438	2,7438	0,7876	128,1755
orig_sin12db.wav	12,0007	13,0200	2,8038	0,7093	80,7738
orig_sin14db.wav	14,0003	14,0601	2,8801	0,6342	50,9071
orig_sin16db.wav	16,0024	14,7840	2,9722	0,5620	32,0535
orig_sin18db.wav	17,9976	15,3312	3,0751	0,4931	20,2048
orig_sin20db.wav	19,9969	15,7534	3,1853	0,4276	12,7165
orig_sin22db.wav	22,0052	16,3624	3,2995	0,3654	7,9806
orig_sin24db.wav	24,0052	16,5653	3,4139	0,3073	5,0133
orig_sin26db.wav	26,0058	16,6982	3,5268	0,2534	3,1449
orig_sin28db.wav	27,9862	16,7835	3,6350	0,2047	1,9790
orig_sin30db.wav	29,9943	16,8389	3,7395	0,1602	1,2346

A.3 Výsledky pro nahrávku zarušenou hlukem ledničky

Název	SNR [dB]	SNR _{vad} [dB]	PESQ [-]	LLR [-]	IS [-]
orig_lednicka2db.wav	2,0005	7,7510	2,6392	0,2395	74,3868
orig_lednicka4db.wav	4,0004	7,8783	2,7589	0,2222	46,7877
orig_lednicka6db.wav	6,0003	9,2255	2,8681	0,2045	29,3988
orig_lednicka8db.wav	8,0002	10,5469	2,9707	0,1864	18,4495
orig_lednicka10db.wav	10,0001	11,8109	3,0744	0,1680	11,5604
orig_lednicka12db.wav	12,0001	12,6768	3,1762	0,1495	7,2304
orig_lednicka14db.wav	14,0000	13,4205	3,2780	0,1307	4,5130
orig_lednicka16db.wav	16,0000	14,3251	3,3823	0,1119	2,8103
orig_lednicka18db.wav	18,0000	14,8861	3,4913	0,0934	1,7457
orig_lednicka20db.wav	20,0001	15,2720	3,6043	0,0755	1,0815
orig_lednicka22db.wav	22,0000	16,2574	3,7122	0,0589	0,6682
orig_lednicka24db.wav	24,0000	16,4864	3,8111	0,0440	0,4116
orig_lednicka26db.wav	26,0000	16,6034	3,9055	0,0314	0,2528
orig_lednicka28db.wav	28,0000	17,0072	3,9927	0,0213	0,1547
orig_lednicka30db.wav	30,0000	17,0702	4,0705	0,0138	0,0944

A.4 Výsledky pro nahrávku zarušenou škrabáním

Název	SNR [dB]	SNR _{vad} [dB]	PESQ [-]	LLR [-]	IS [-]
orig_sum_ntb2db.wav	2,2775	15,7658	2,0384	0,3557	1240,5467
orig_sum_ntb4db.wav	4,1891	15,4072	2,1488	0,3246	838,3722
orig_sum_ntb6db.wav	6,1249	13,9494	2,2536	0,2937	557,9963
orig_sum_ntb8db.wav	8,0745	13,6415	2,3632	0,2635	368,1303
orig_sum_ntb10db.wav	10,0376	13,6558	2,4773	0,2343	240,0356
orig_sum_ntb12db.wav	12,0142	13,7044	2,5911	0,2062	154,8885
orig_sum_ntb14db.wav	14,0000	14,5953	2,7105	0,1797	99,0884
orig_sum_ntb16db.wav	16,0000	14,6770	2,8392	0,1549	62,4404
orig_sum_ntb18db.wav	18,0000	15,2052	2,9713	0,1320	39,3306
orig_sum_ntb20db.wav	20,0000	15,8141	3,1063	0,1112	24,7615
orig_sum_ntb22db.wav	22,0000	16,1265	3,2354	0,0925	15,5794
orig_sum_ntb24db.wav	24,0000	16,5578	3,3689	0,0761	9,7952
orig_sum_ntb26db.wav	26,0000	16,6560	3,5030	0,0618	6,1528
orig_sum_ntb28db.wav	28,0000	16,8048	3,6294	0,0496	3,8605
orig_sum_ntb30db.wav	29,9999	16,8038	3,7528	0,0394	2,4192

A.5 Výsledky pro nahrávku zarušenou šumem vysílačky

Název	SNR [dB]	SNR _{vad} [dB]	PESQ [-]	LLR [-]	IS [-]
orig_vysilacka2db.wav	2,0002	7,5546	2,0679	0,4774	34,9266
orig_vysilacka4db.wav	4,0001	7,6178	2,1369	0,4416	22,0557
orig_vysilacka6db.wav	6,0000	8,9844	2,2185	0,4051	13,9326
orig_vysilacka8db.wav	8,0000	10,4842	2,3089	0,3686	8,8045
orig_vysilacka10db.wav	10,0000	11,9001	2,4046	0,3326	5,5644
orig_vysilacka12db.wav	12,0000	12,9215	2,5081	0,2975	3,5151
orig_vysilacka14db.wav	14,0000	14,0560	2,6217	0,2638	2,2177
orig_vysilacka16db.wav	16,0000	14,8656	2,7379	0,2321	1,3954
orig_vysilacka18db.wav	18,0000	15,7735	2,8548	0,2022	0,8759
orig_vysilacka20db.wav	20,0000	16,1385	2,9767	0,1742	0,5490
orig_vysilacka22db.wav	22,0000	16,4595	3,0997	0,1480	0,3453
orig_vysilacka24db.wav	24,0000	16,6831	3,2269	0,1235	0,2199
orig_vysilacka26db.wav	26,0000	16,6584	3,3629	0,1009	0,1444
orig_vysilacka28db.wav	28,0000	16,7427	3,5188	0,0802	0,0992
orig_vysilacka30db.wav	30,0000	16,7968	3,6784	0,0617	0,0722

B Přiložené CD

- Elektronická verze bakalářské práce
- Programy
 - kvalita_reci
 - Colea
 - Composite
- Nahrávky
 - Originální nahrávka
 - Zarušené nahrávky
- Výsledky
 - Hodnoty.xlsx