VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ
ÚSTAV TELEKOMUNIKACÍ

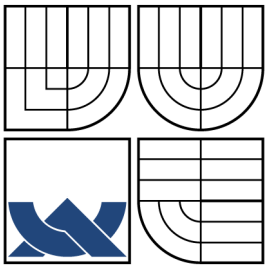
FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

SEGMENTACE ŘEČOVÉHO SIGNÁLU

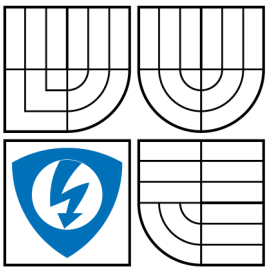
BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE PETR ANDRLA
AUTHOR

BRNO 2008



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

SEGMENTACE ŘEČOVÉHO SIGNÁLU
SPEECH SEGMENTATION INTO PHONEMES

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE — PETR ANDRLA
AUTHOR

VEDOUCÍ PRÁCE — ING. PETR SYSEL
SUPERVISOR

BRNO 2008

ZDE VLOŽIT LIST ZADÁNÍ

Z důvodu správného číslování stránek

ZDE VLOŽIT PRVNÍ LIST LICENČNÍ SMOUVY

Z důvodu správného číslování stránek

ZDE VLOŽIT DRUHÝ LIST LICENČNÍ SMOUVY

Z důvodu správného číslování stránek

ABSTRAKT

V rámci bakalářské práce byl vytvořen program pro segmentaci nahrávek řeči na fonémy. Tento program byl vytvořen v prostředí Matlab a skládá se z několika skriptů. Program umožňuje automatickou i ruční segmentaci. Automatická segmentace je založena na metodě sledování příznaků. Tímto programem byla zpracována skupina nahrávek a vyhodnocena účinnost automatické segmentace. K programu byl vytvořen podrobný návod k obsluze. Dále jsou v práci stručně rozebrány jednotlivé použité metody zpracování řeči s uvedením jejich implementace v programu a odůvodnění nastavení jejich proměnných parametrů.

KLÍČOVÁ SLOVA

Foném, hláska, řeč, segmentace řečového signálu.

ABSTRACT

The programme for the segmentation of a speech into fonems was created as a part of the bachelor´s thesis. This programme was made in the programme Matlab and consists of several scripts. The programme serves for automatic and hand segmentation. Automatic segmentation is based on the method of following symptom. The audiorecords were elaborated by the programme and a operation of the automatic segmentation was analysed. A detailed manual was created to the programme too. Individual used methods of the elaboration of a speech were in the bachelor´s thesis briefly descripted, its implementations in the programme and reasons of set of its parameters.

KEYWORDS

Phoneme, phone, speech, segmentation of speech signal.

ANDRLA P. *Segmentace řečového signálu*. Brno: Vysoké učení technické v Brně. Fakulta elektrotechniky a komunikačních technologií. Ústav telekomunikací, 2008. Počet stran 52 s., Počet stran příloh 8 s. příloh. Vedoucí bakalářské práce Ing. Petr Sysel.

PROHLÁŠENÍ

Prohlašuji, že svou bakalářskou práci na téma „Segmentace řečového signálu“ jsem vypracoval samostatně pod vedením vedoucího bakalářské 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 bakalářské práce dále prohlašuji, že v souvislosti s vytvořením této bakalářské 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 ....

(podpis autora)

Děkuji panu inženýru Petru Syslovi za velmi kvalitní vedení práce a za poskytnutí důležitých informací a materiálů nutných k její realizaci.

OBSAH

SEZNAM OBRÁZKŮ

SEZNAM TABULEK

ÚVOD

Cílem bakalářské práce je vytvořit skripty v prostředí Matlab, které budou provádět automatickou segmentaci, tedy rozpoznají hranice fonémů v řeči, a dále vytvořit grafické rozhraní, ve kterém bude možné provádět ruční korekce rozpoznaných hranic fonémů. K určení hranic mezi fonémy slouží metoda sledování příznaků. Metoda sledování příznaků vychází z toho, že řečový signál má kvazistacionární charakter, to znamená, že má po celou dobu trvání fonému přibližně stacionární průběh.

Charakter signálu se stanovuje metodami krátkodobé analýzy. Metody krátkodobé analýzy se snaží reprezentovat řečový signál pomocí příznaků, získaných z jeho vzorků. Důležité je, aby získané příznaky co nejlépe charakterizovaly průběh signálu a jeho vlastnosti. Existují různé druhy metod krátkodobé analýzy v časové a frekvenční oblasti. Patří mezi ně například krátkodobá energie, krátkodobá střední hodnota průchodů signálu nulou, krátkodobá autokorelační funkce, krátkodobá diskrétní Fourierova transformace nebo kepstrální analýza.

U reálného řečového signálu ale charakteristiky nejsou v rámci jednoho fonému vždy úplně konstantní a na hranicích se sousední fonémy částečně vlivem koartikulace ovlivňují. Navíc některé charakteristiky dvou různých fonémů mohou být velmi podobné, což může vést ke vzniku chyby při segmentaci. Proto je nutné zvolit pro analýzu takové metody a vhodně nastavit jejich parametry, aby se vznik chyb minimalizoval. Ale i přesto není možné docílit stoprocentní spolehlivosti segmentace.

Při praktické realizaci metody sledování příznaků se nejprve provede rozdělení řečového signálu na krátké úseky, mikrosegmenty. Poté se pro každý mikrosegment stanoví příznaky metodami krátkodobé analýzy, pro každý z příznaků se následně vypočte jeho křivka rozdílnosti. Na výsledné křivce rozdílnosti, získané součtem křivek dílčích, se naleznou maxima. Ta by měla odpovídat hranicím fonémů.

1 METODY ANALÝZY ŘEČOVÉHO SIGNÁLU

1.1 Řečové jednotky

1.1.1 Fonetika

Fonetika je věda, která se zabývá komplexním studiem procesu vytváření řeči, a to od vzniku výdechového proudu v plicích, až po modifikace v artikulačním traktu. Jejími zájmy tedy jsou činnost řečových orgánů, způsob tvoření řeči, charakter výsledného zvuku a i jejich sluchové hodnocení [2]. Základní jednotkou ve fonetice je *hláska*. Hláska, neboli fon, reprezentuje soubor foneticky podobných, z fonetického hlediska navzájem neodlišitelných, zvuků. Všechny odlišné hlásky jednoho jazyka tvoří jeho fonetický inventář. Fonetická reprezentace řeči se zapisuje do hranatých závorek. Řeč je tedy z pohledu fonetiky posloupnost jednotlivých hlásek.

1.1.2 Fonologie

Fonologie se zabývá řečovými zvuky z hlediska systémové stavby jazyka. Stojí tedy někde mezi fonetikou a vyšší lingvistikou, zajímá se o postavení, funkci a vztahy mezi zvuky v rámci jazyka [2]. Základní jednotkou ve fonologii je *foném*. Foném je definován jako nejmenší lingvistická jednotka schopná rozlišovat významové jednotky. Fonémy vytváří v daném jazyce celistvou množinu, kde se každý foném liší od všech ostatních. Některé fonémy jsou využívány více, jiné jen velmi málo.

Foném je abstraktní lingvistická jednotka, nepředstavuje přímo žádný zvukový segment, ale lingvistické charakteristiky a konfiguraci hlasového traktu. Až při artikulaci fonému vzniká vlastní zvuk, jeho akustická realizace. Každému fonému odpovídá jisté základní postavení řečových orgánů, avšak při vlastní artikulaci dochází k jeho mírným modifikacím, ty jsou dány částečným stupněm volnosti při vyslovení daného fonému a dále koartikulací, ovlivňováním od okolních fonémů. Důsledkem toho je, že jeden foném může být realizován obrovským množstvím zvuků. Zvuky s podobnými fonetickými vlastnostmi se nazývají hlásky. Jeden foném tedy může být reprezentován více hláskami označovanými jako významové *alofony*. I při opakovaném vyslovení jednoho fonému stejným řečníkem může být akustická realizace tohoto fonému různá. Vlivem koartikulace fonémů mohou vznikat různé alofony, ty jsou nositeli informací, o této koartikulaci. V případě, kdy by došlo k záměně jednoho alofonu za jiný, nemělo by to mít vliv na srozumitelnost řeči, ale pouze na její příjemnost.

1.1.3 Fonetické abecedy

Množina fonetických jednotek jazyka se nazývá fonetický inventář [2]. Fonetický inventář zahrnuje jednotlivé zvuky řeči. Existují různé druhy a definice fonetického inventáře podle toho, jakou mírou podrobnosti se na ně díváme. Nejméně podrobný je inventář fonémů, který například vůbec nezahrnuje koartikulaci. Detailnější jsou alofonické inventáře, nejčastěji se pracuje na úrovni hlásek. Fonetické inventáře se nazývají fonetické abecedy. Pomocí fonetické abecedy lze zachytit a zapsat promluvu. Za tímto účelem vzniklo několik fonetických abeced.

IPA

Mezinárodní fonetická abeceda IPA (International Phonetic Alphabet) je standard pro zápis výslovnosti pro všechny světové jazyky. Tato abeceda je nezávislá na jazyce a poskytuje úplnou soustavu fonetických značek. Pomocí mezinárodní fonetické abecedy lze výslovnost reprezentovat nejen na fonémové úrovni, ale i podrobně na úrovni alofonů. Konkrétní jazyky většinou nevyužívají všechny fonémy použité v této abecedě, je proto běžné pro popis národních jazyků použít jiné abecedy. Mezinárodní fonetická abeceda IPA zahrnuje i prozodické charakteristiky v řeči.

SAMPA

Pro potřebu zápisu promluvy v počítači vznikla fonetická abeceda SAMPA (Speech Assesment Methods Phonetic Alphabet). Hlavním požadavkem je, aby zápis této abecedy byl jednoznačný a nebylo třeba oddělovat symboly mezerou. Abeceda SAMPA pracuje na úrovni fonémů. Symboly z abecedy IPA se kódují v ASCII, přitom ty, které se shodují s malými písmeny latinky zůstávají stejné. Pro prozodické značení vznikla abeceda SAMPROSA. Dále vznikla abeceda X-SAMPA, která obsahuje všechny symboly z abecedy IPA.

ZČFA

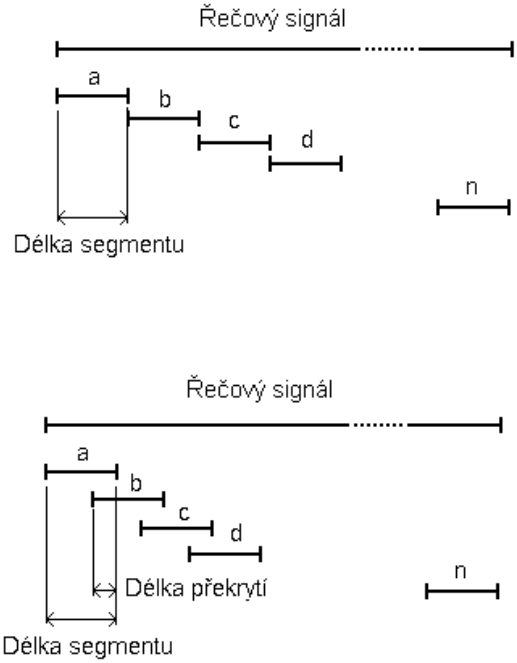
Pro popis českého jazyka lze užít abecedy IPA a SAMPA. Mezinárodní fonetická abeceda IPA však není příliš vhodná. Proto vznikla abeceda ZČFA (zjednodušená česká fonetická abeceda) vycházející ze základního inventáře českého jazyka. Pro symboly, pro které v českém jazyce neodpovídá žádné písmeno převzala ZČFA znaky z IPA. Její výhodou je dobrá čitelnost jejích symbolů, je velmi vhodná pro rozbor české mluvené řeči, ale není příliš vhodná pro zpracování v počítači. Pro počítačové zpracování je vhodnější fonetická abeceda ČFA, její symboly se ovšem obecně skládají z více znaků, proto je nutné je oddělovat mezerami.

1.2 Krátkodobá analýza řeči

Metody krátkodobé analýzy reprezentují řečový signál pomocí příznaků. Důležité je, aby získané příznaky co nejlépe charakterizovaly průběh signálu a jeho vlastnosti. Pro odhalení hranic mezi fonémy je třeba, aby hodnota příznaků sousedních fonémů byla co nejvíc rozdílná, skokově se měnila na hranici fonémů a uvnitř jednoho fonému byla konstantní.

Rozdělení na mikrosegmenty

Společné pro všechny metody krátkodobé analýzy je, jak již z jejího názvu vyplývá, aplikace výpočtů příznaků pouze pro krátký časový úsek. To plyne z předpokladu, že parametry řeči jsou konstantní jen na krátkém úseku. Volba délky těchto mikrosegmentů ovlivňuje výpočty všech metod krátkodobé analýzy a tím i výrazně celý proces segmentace, proto je nutná její vhodná volba. Délka mikrosegmentů musí být dostatečně malá, aby bylo možné postihnout i ty nejkratší fonémy, ale zároveň dostatečně velká, aby bylo možné postřehnout kvaziperiodický charakter řečového signálu. Nejvhodnější délka se může lišit pro výpočet různých příznaků. Většinou se používá délka 10-30 ms, nejčastěji 20 ms. Při segmentaci je možné a vhodné, aby se sousední segmenty vzájemně překrývaly, což vede k lepšímu vyrovnání průběhů počítaných parametrů signálu. Princip segmentace je znázorněn na obrázku 1.1.



Obr. 1.1: Princip segmentace bez překrytí sousedních segmentů a s překrytím sousedních segmentů.

Zpracovávaný řečový signál tedy nejprve rozdělíme na mikrosegmenty. Pro každý z nich pak určíme příznak pomocí některé metody krátkodobé analýzy. Příznaky mikrosegmentů poté zpracujeme metodou sledování příznaků.

Pro rozdělení na mikrosegmenty byl v prostředí Matlab vytvořen skript `Segmentace`. Jeho vstupní parametry jsou:

1. Zvukový signál, zadaný jako sloupcový vektor hodnot vzorků,
2. vzorkovací kmitočet signálu,
3. délka mikrosegmentů, zadává se v mikrosekundách,
4. délka překrytí mikrosegmentů, zadává se jako podíl délky mikrosegmentu (např. 0.5 odpovídá překrytí 50%).

Výstup skriptu je matice, jejíž sloupce jsou jednotlivé mikrosegmenty. Index sloupce tedy odpovídá pořadí mikrosegmentů a index řádku pořadí vzorku v mikrosegmentu. Takto rozdělený signál je připraven k dalšímu zpracování. Parametry délka mikrosegmentů a délka překrytí lze nastavit vlastnosti funkce, tak jak jsou vhodné pro další operace.

Uvnitř skriptu `Segmentace` je nejprve zjištěna délka vstupního signálu a počet vzorků, o které se budou mikrosegmenty překrývat. Podle těchto hodnot je vytvořena výstupní matice s příslušným počtem sloupců a řádků. Tato matice je následně naplněna hodnotami tak, aby jednotlivé sloupce odpovídaly mikrosegmentům.

1.2.1 Zpracování v časové oblasti

Při zpracování v časové oblasti se vychází přímo z hodnot vzorků signálu. Příznaky získané z časové oblasti jsou většinou méně náročné na výpočet. Užitím jedné metody by pro mnohé fonémy hodnota příznaků zůstala stejná či se jen velmi málo změnila. Tedy by bylo obtížné stanovit hranice všech fonémů, proto je třeba použít více různých metod zpracování v časové oblasti a společně s nimi metody zpracování ve frekvenční oblasti. Mezi metody využívající zpracování v časové oblasti patří například krátkodobá energie, krátkodobá intenzita, krátkodobá střední hodnota průchodu signálu nulovou úrovní a její modifikace a krátkodobá autokorelační funkce.

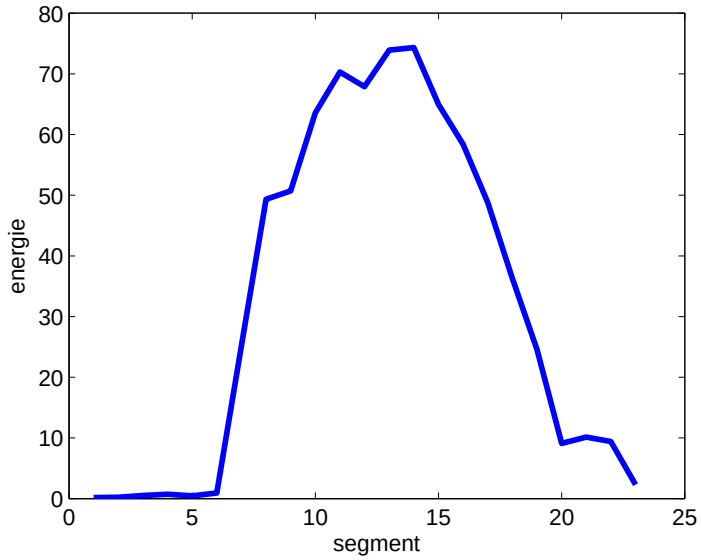
Krátkodobá energie

Slouží k výpočtu energie mikrosegmentů. Touto metodou lze určit hranici mezi znělými a neznělými úseky řeči. Funkci krátkodobé energie lze definovat vztahem [2]:

$$E = \sum_{k=0}^{N-1} x[k]^2, \tag{1.1}$$

kde $x[k]$ je vzorek signálu, k je pořadí vzorku v mikrosegmentu a N je počet vzorků v mikrosegmentu.

Pro výpočet krátkodobá energie byl v prostředí Matlab vytvořen skript `Energie`. Jeho vstupní hodnotou je segmentovaný řečový signál a výstupem řádkový vektor příznaku energie pro každý mikrosegment. Výpočet je uvnitř tohoto skriptu proveden podle vztahu (1.1). Na obrázku 1.2 je graficky znázorněný průběh krátkodobé energie vypočtené tímto skriptem pro slabiku „ce“.



Obr. 1.2: Průběh krátkodobé energie slabiky „ce“.

Krátkodobá intenzity

Krátkodobá intenzita je definována vztahem [2]:

$$I = \sum_{k=0}^{N-1} |x[k]| \,. \tag{1.2}$$

Skript pro výpočet krátkodobé intenzity byl vytvořen také, ale pro výpočet křivky rozdílnosti příznaků využit nebyl. Jeho výsledek by byl téměř shodný s výsledkem krátkodobé energie. Tyto metody by detekovaly pouze stejné hranice fonémů, proto jejich současné použití by bylo neefektivní.

Krátkodobá střední hodnota průchodů signálu nulovou úrovní

Je jedna z metod zpracovávající signál v časové oblasti, přesto frekvenci průchodů nulovou úrovní lze brát jako jednoduchou funkci popisující spektrální charakter signálu. Výhoda střední hodnoty průchodů nulovou úrovní je její úplná nezávislost na energii signálu.

Krátkodobou střední hodnotu průchodů signálu nulou lze definovat vztahem [2]:

$$Z = \sum_{k=0}^{N-1} \left|\mathrm{sgn}(x[k]) - \mathrm{sgn}(x[k-1])\right| / 2, \tag{1.3}$$

kde $x[k]$ je vzorek signálu, k je pořadí vzorku v mikrosegmentu a N je počet vzorků v mikrosegmentu.

Metoda počtu průchodů nulovou úrovní má několik modifikací, například měření okamžitých vzdáleností dvou po sobě následujících průchodů nulou, nebo Krátkodobá funkce středního počtu výskytu lokálních extrémů.

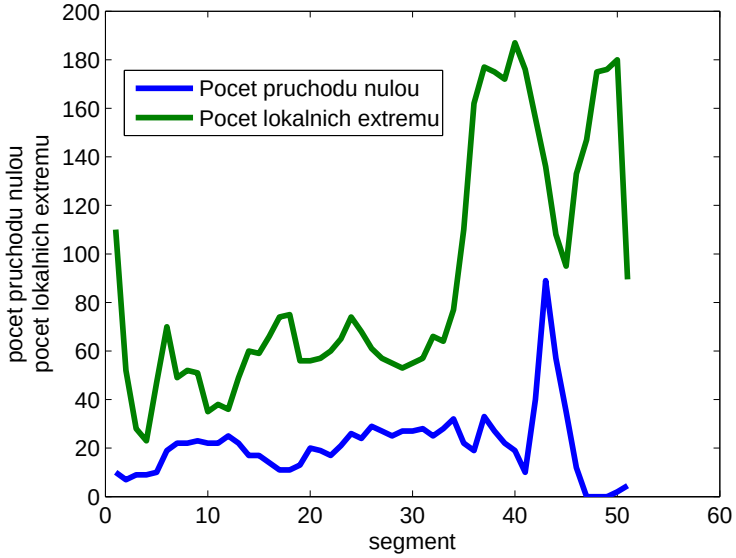
Krátkodobá funkce středního počtu výskytu lokálních extrémů

Tato metoda se pro některé fonémy shoduje s metodou počtu průchodů nulovou úrovní, ale pro jiné je velmi výhodná. Zvláštní význam má pro ty fonémy, které mají stejnosměrnou složku a malý rozkmit signálu. Tyto fonémy nemají téměř žádný počet průchodů nulovou úrovní a zároveň mají velký počet lokálních extrémů. Proto byla tato metoda také zahrnuta do výpočtu křivky rozdílnosti příznaků.

V prostředí Matlab byly vytvořeny skripty pro výpočet krátkodobé střední hodnoty průchodů signálu nulovou úrovní `PruchoduNulou` a pro výpočet krátkodobé funkce středního počtu výskytu lokálních extrémů skript `LokEx`. Výstupem skriptu `PruchoduNulou` je řádkový vektor hodnoty příznaku pro každý mikrosegment, počet průchodů nulovou úrovní je vypočten vztahem (1.3). Skript `LokEx` je oproti `PruchoduNulou` modifikován pouze tím, že před určením počtu nul je každý segment vstupního signálu derivován (respektive je určena jeho diference). Na obrázku 1.3 jsou zobrazeny průběhy funkcí průchodů nulou a lokálních extrémů slova „devět“.

Krátkodobá autokorelační funkce

Autokorelační funkce má některé vlastnosti, díky kterým může být použita pro určení periodicity signálu. Autokorelační funkce v podstatě udává podobnost signálu na sebe samotného, posunutého o k vzorků. Pro výpočet autokorelační funkce by měl mikrosegment obsahovat alespoň dvě periody signálu, musí být tedy dostatečně dlouhý (doporučuje se 20-40 ms). Při dalším zpracování se většinou nepoužívá celý průběh autokorelační funkce, ale jen její část: *autokorelační koeficienty*. Z metod



Obr. 1.3: Krátkodobá střední hodnota průchodů signálu nulou a krátkodobá funkce středního počtu výskytu lokálních extrémů pro slovo „devět“.

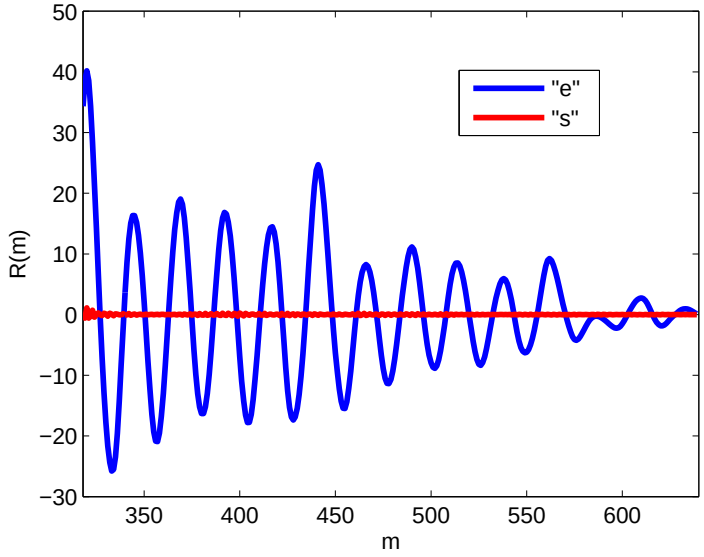
zpracovávající signál v časové oblasti má autokorelační funkce největší výpočetní nároky, zato je v některých případech efektivnější. Krátkodobá autokorelační funkce je definována [2]:

$$R(m) = \sum_{n=0}^{N-1} x(n)x(n+m). \tag{1.4}$$

Autokorelační funkce je poslední z metod zpracovávající signál v časové oblasti využitých při segmentaci řečového signálu na fonémy. Vstupem skriptu `Korelace`, pro výpočet koeficientů autokorelační funkce, je opět segmentovaný řečový signál. Výstupem není skalární hodnota pro každý segment, jak tomu bylo u předchozích metod, ale vektory koeficientů. Kolik prvních koeficientů z vypočtené autokorelační funkce bude použito udává druhý vstupní parametr skriptu. Pro výpočet autokorelace je použita funkce `xcorr`. Pro dva vstupní signály tato funkce, z prostředí Matlab, určí jejich vzájemnou korelaci a při jednom vstupním signálu jeho autokorelaci. Na obrázku 1.4 je znázorněn výsledek autokorelace pro slabiku „es“.

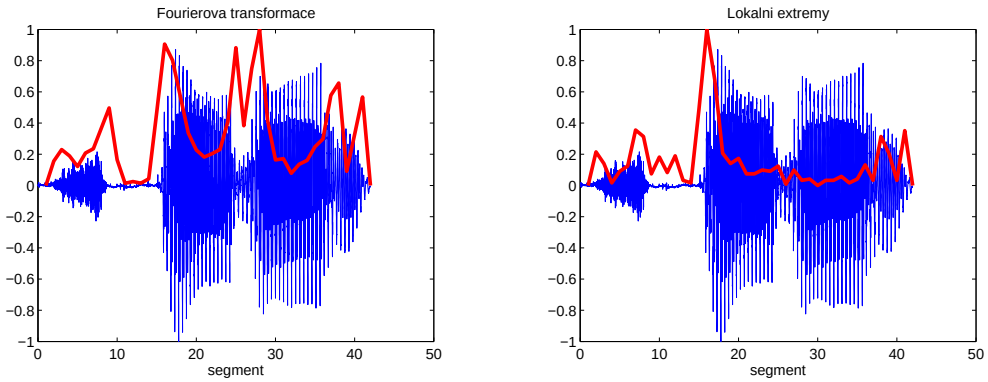
1.2.2 Zpracování ve frekvenční oblasti

Mluvená řeč je ve frekvenční oblasti reprezentována svým spektrem, velikostí frekvenčních složek. Základem většiny metod zpracování ve frekvenční oblasti je diskrétní Fourierova transformace. Výpočetní náročnost metod zpracování ve frekvenční oblasti je často větší než u metod z časové oblasti, většinou však oproti nim



Obr. 1.4: Autokorelační funkce slabiky „es“.

dávají rozdílné výsledky, tedy je díky nim možné detekovat hranice, které by metodami z časové oblasti zůstaly nenalezeny. Na obrázku 1.5 je znázorněno srovnání křivek rozdílnosti příznaků (viz dále) pro slovo „čtyři“ vypočtených diskrétní Fourierovou transformací (zpracování ve frekvenční oblasti) a funkce počtu lokálních extrémů (zpracování v časové oblasti). Kromě diskrétní Fourierovi transformace patří mezi metody využívající zpracování ve frekvenční oblasti ještě například Kepstrální analýza.



a) b)

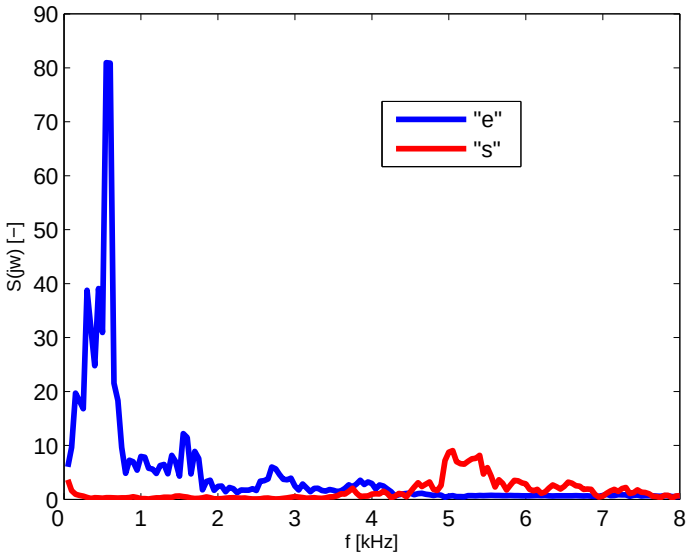
Obr. 1.5: Srovnání detekování hranice fonému pomocí a) diskrétní Fourierovi transformace b) funkce počtu lokálních extrémů.

Krátkodobá diskrétní Fourierova transformace

Diskrétní Fourierovou transformací získáme spektrum signálu, ze kterého je možné určovat jeho kmitočtové vlastnosti. DFT je definována [2]:

$$S(e^{j\omega}) = \sum_{n=-\infty}^{\infty} s[n]e^{-j\omega n}. \tag{1.5}$$

Pro výpočet diskrétní Fourierovy transformace byl vytvořen skript `Fourierova`. Jeho základem je algoritmus FFT (fast Fourier transform), v Matlabu realizován funkcí `fft`. Po aplikaci této funkce na vstupní segmentovaný signál získáme komplexní spektrum každého mikrosegmentu. Z něj se vypočte modulové spektrum jako absolutní hodnota komplexního spektra (funkce `abs`). Spektrum reálného signálu získané pomocí `fft` je symetrické kolem poloviny vzorkovacího kmitočtu, proto je pro další výpočty použita jen první polovina vypočteného modulového spektra. V rámci jednoho fonému, je pro dva sousední mikrosegmenty charakter spektra stejný, přesto se jeho hodnota na konkrétních frekvencích může lišit. Z tohoto důvodu je výhodnější při výpočtu křivky rozdílnosti neporovnávat frekvence jednotlivě, ale pouze celkovou energii ve frekvenčních pásmech. V tomto skriptu je spektrum rozděleno do obdelníkových pásem, jejichž šířka se nastavuje druhým vstupním parametrem. Příklad modulového spektra některých fonémů je na obrázku 1.6.



Obr. 1.6: Modulové spektrum fonémů „e“ a „s“.

Kepstrální analýza

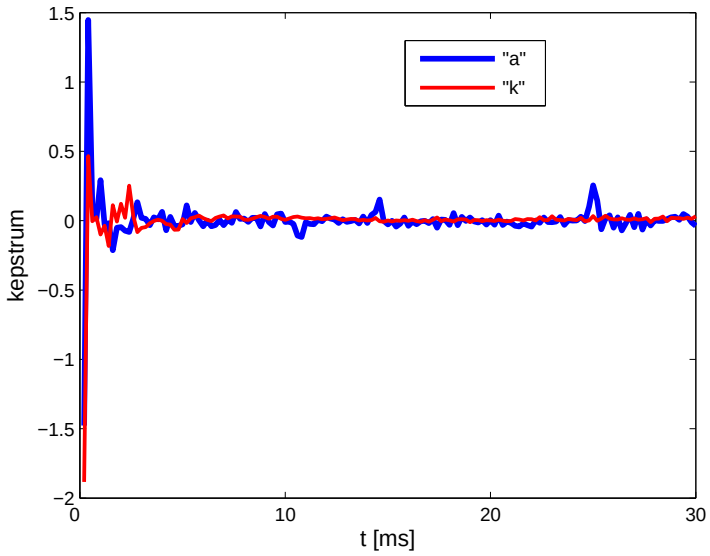
Kepstrální analýza je metoda umožňující ze signálu řeči oddělit parametry buzení a hlasového ústrojí. Každá složka spektra řeči je dána konvolucí buzení a impulzní

odezvy hlasového ústrojí. Pro analýzu řeči je vhodné tyto dva parametry oddělit. Ze signálu získaného konvolucí dvou vstupních signálů je ale obtížné dostat tyto signály zpět. Lze se o to pokusit tak, že zavedeme vhodnou nelineární operaci, například logaritmování, která převede součin na součet ($\ln(ab) = \ln(a) + \ln(b)$). Jednotlivé složky součtu pak lze od sebe oddělit. Proces kepstrální analýzy se skládá z následujících kroků:

1. diskrétní Fourierovy transformace (výpočtu spektra mikrosegmentů řeči),
2. logaritmování spekter mikrosegmentů,
3. zpětné Fourierovy transformace.

Takto získáme kepstrum signálu.

Skript `Kepstrum` sloužící k provedení kepstrální analýzy používá v programu Matlab definovanou funkci `rceps`. Uvnitř funkce `rceps` je implementován algoritmus $y = \text{real}(\text{ifft}(\log|\text{fft}(x)|))$. Pro hledání hranic fonémů metodou sledování příznaků není vhodné použít celé získané kepstrum, ale jen několik prvních koeficientů. Kolik koeficientů bude použito udává druhý vstupní parametr skriptu. Na obrázku 1.7 je znázorněna část kepstra pro fonémy „a“ a „k“.



Obr. 1.7: Část kepstra pro fonémy „a“ a „k“.

1.3 Metoda sledování rozdílnosti příznaků

Metoda sledování rozdílnosti příznaků se používá pro stanovení hranic mezi fonémy uvnitř slova. Tato metoda předpokládá, že řečový signál má v průběhu jednoho fonému stacionární charakter (nemění se hodnoty příznaků). Při změně vysloveného fonému se charakteristika signálu skokem změní a po dobu trvání tohoto fonému bude opět stacionární. Vlastnosti reálného řečového signálu ale mohou způsobit chybu při segmentaci. Charakteristiky signálu se určují metodami krátkodobé analýzy.

Křivka rozdílnosti příznaků

Křivku rozdílnosti příznaků lze zapsat [3]:

$$B(j) = |P(j + x_1) - P(j - x_2)| \qquad pro\ \ 1 - x_2 \leq j \leq J - x_1, \tag{1.6}$$

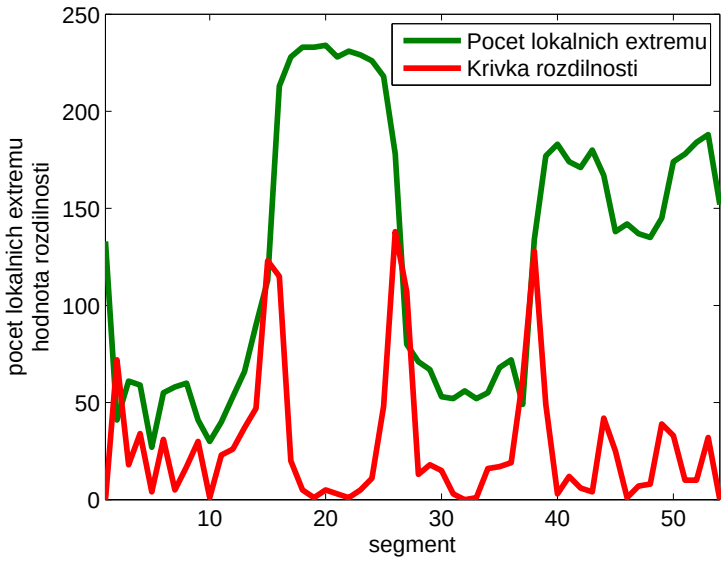
kde $P(j)$ je hodnota příznaku v segmentu j, J je počet segmentů, x_1, x_2 je posun vpřed, vzad.

Její úkol je rozpoznat změny průběhů jednotlivých příznaků v rámci celého řečového signálu. Křivka rozdílnosti příznaků tedy sleduje křivku danou hodnotami příznaků. Uprostřed jednoho fonému je příznak konstantní a křivka rozdílnosti nabývá malých hodnot a na hranicích fonémů se příznak mění a křivka rozdílnosti nabývá velké hodnoty. V ideálním případě by křivka rozdílnosti byla ve všech bodech nulová kromě bodů, ve kterých dochází ke změně fonémů. Vzhledem k tomu je lepší počítat křivku rozdílnosti ne pro dva přímo sousedící segmenty, ale pro segmenty od sebe více vzdálené. K tomuto účelu slouží proměnné x_1 a x_2, které určují posun segmentů, ze kterých se křivka rozdílnosti příznaků stanovuje vůči segmentu, pro který je počítána.

Pro výpočet křivky rozdílnosti slouží skript `VyhodnotData`. Jeho vstupy jsou:

1. řádkový vektor či matice obsahující hodnotu příznaku pro každý mikrosegment, získané některou z metod krátkodobé analýzy,
2. x_2 posun vzad oproti počítanému segmentu,
3. x_1 posun vpřed oproti počítanému segmentu.

Výstup je křivka rozdílnosti pro jeden příznak, získaná podle vztahu (1.6). Na obrázku 1.8 je zobrazen průběh křivky rozdílnosti jednoho příznaku (konkrétně krátkodobé funkce středního počtu výskytu lokálních extrémů) společně s průběhem tohoto příznaku.



Obr. 1.8: Křivka rozdílnosti počtu lokálních extrémů slova „deset“.

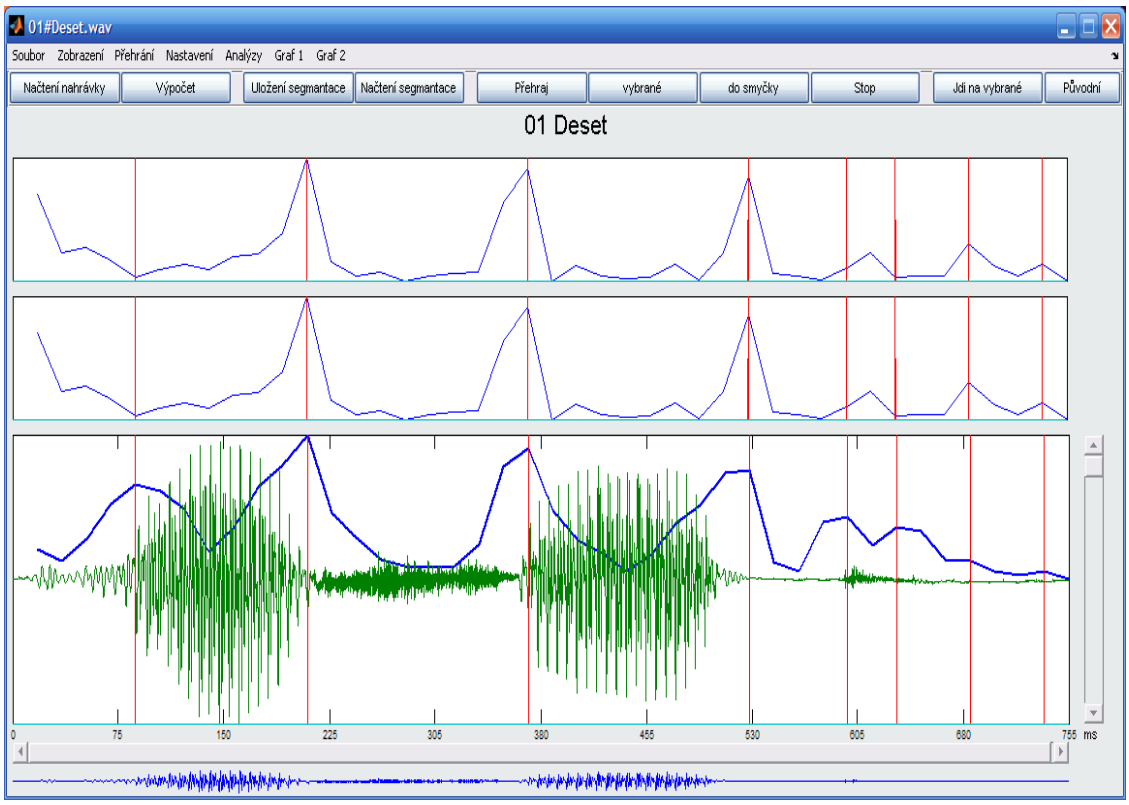
Stanovení hranic fonémů

Jak bylo uvedeno výše, hranice fonémů získáme nalezením maxim na křivce rozdílnosti příznaků. Za tímto účelem byl vytvořen skript `Maxima`. Stejně jako ve skriptu na hledání lokálních extrémů se i zde nejprve provede derivace vstupních dat, poté se určí průchody nulou, což pro data derivovaná značí výskyt lokálních extrémů. Po těchto operacích získáme vektor, který má hodnoty 1 pro maximum, -1 pro minimum a 0 pro segment, kde je křivka rozdílnosti monotónní. Odstraněním minim zůstane nenulová hodnota pouze u segmentů ležící na hranicích fonémů (pokud tedy byly hranice fonému správně detekovány na křivce rozdílnosti).

Program `Prace` slouží pro automatickou ale i ruční segmentaci. Ke stanovení hranic automaticky je v programu `Prace` využito všech výše zmíněných skriptů, v algoritmu pro výpočet segmentace se nejprve načte řečový signál, který zvolíme, ve formátu .wav. V dalším kroku jsou metodami krátkodobé analýzy vypočteny charakteristiky signálu (příznaky). Pro výstup každé z metod se pomocí skriptu `VyhodnotData` určí vlastní křivka rozdílnosti. Tyto křivky se poté sečtou v jednu. Přičemž každé může být nastavena různá váha. Na této výsledné křivce rozdílnosti příznaků se určí maxima. Křivka rozdílnosti s rozpoznanými maximy a společně se řečovým signálem je zobrazena do hlavního grafu programu. Ruční segmentace se provádí s využitím grafického rozhraní programu.

2 POPIS UŽIVATELSKÉHO ROZHRANÍ

Program k bakalářské práci byl vytvořen v programu Matlab ve verzi 7.1.0.246. Vytvořený program je určený pro segmentaci řečového signálu na fonémy, jeho grafické prostředí slouží především pro ruční segmentaci.



Obr. 2.1: Vytvořené grafické uživatelské rozhraní pro automatickou i ruční segmentaci.

Uživatelské rozhraní se skládá ze tří hlavních částí:

1. Ovládací části (nahoře) zahrnující menu programu a panel s tlačítky,
2. grafů pro pomocné průběhy (uprostřed),
3. hlavního grafu s posuvníky (dole).

Popis jejich funkcí a ovládání je podrobně uveden níže.

Okno programu má konstantní rozměr, jeho velikost nelze měnit (mění se pouze přidáním či odebráním pomocných průběhů). Posouvat program po obrazovce lze standardním způsobem.

Titulek v záhlaví okna je shodný s názvem aktuálně načtené nahrávky.

Pod panelem tlačítek je v případě načtené nahrávky vypsán text nahrávky, který slouží pro orientaci, s kterou nahrávkou se pracuje a pro usnadnění ruční segmentace.

Tento text je převzat z názvu nahrávky, nikoliv z jejího obsahu. Nemusí tedy přesně odpovídat jejímu obsahu. V případě, kdy žádná nahrávka nebyla načtena, je na tomto místě vypsán text: „Název nahrávky“.

2.1 Menu programu

Pomocí menu programu se lze dostat k většině funkcí programu. Ovládání z menu by ovšem bylo neefektivní, proto nejčastěji používané funkce jsou přístupné pomocí panelu s tlačítky a změna rozlišení grafů a posun v nich se děje pomocí posuvníků. Navíc k volání některých funkcí lze užít klávesových zkratek. Klávesové zkratky jsou k dispozici vedle názvu funkcí v menu. Dále následuje podrobný popis funkcí tlačítek z menu.

2.1.1 Soubor

Načtení nahrávky

Po aktivaci tlačítka „Načtení nahrávky“ je vyvoláno dialogové okno pro výběr nahrávky. Pomocí něj lze vybrat požadovanou nahrávku a tento výběr potvrdit tlačítkem „Otevřít“. Lze načíst pouze nahrávky ve formátu .wav! Případně lze dialogové okno deaktivovat tlačítkem „Storno“. Pokud je nahrávka načtena, změní se podle ní titulek v záhlaví okna a pole „Text nahrávky“ a do hlavního grafu se vykreslí její průběh. Nyní lze celou nahrávku přehrát, ostatní funkce jsou nepřístupné.

Výpočet

Tlačítko „Výpočet“ slouží k provedení automatické segmentace a vykreslení průběhů do grafů.

Uložení segmentace

Po výpočtu automatické segmentace, případně po úpravě hranic fonémů ruční segmentací, lze rozdělení nahrávky uložit. Tlačítkem „Uložení segmentace“ se vyvolá dialogové okno pro ukládání. Název ukládaného souboru je již předdefinovaný (podle názvu nahrávky), ale lze jej upravit. Potvrzením pomocí tlačítka „Uložit“ se zapíše do souboru zvoleného názvu proměnná obsahující číslo vzorku kde dochází ke změně fonému. Soubor je uložen ve formátu .txt.

Načtení segmentace

Načtení segmentace je možné, pokud je načtená nahrávka a pokud je provedena automatická segmentace. Načtení opět probíhá pomocí dialogového okna, je nutné načíst pouze soubor odpovídající načtené nahrávce.

Konec

Tlačítkem „Konec“ se ukončí program, zavře se okno programu a vymažou se všechny proměnné. Program je vhodnější ukončit tlačítkem „Konec“ než křížkem v rámu okna.

2.1.2 Zobrazení

Přiblížení

Slouží k zvětšení rozlišení na x-ových osách grafů.

Oddálení

Slouží ke snížení rozlišení na x-ových osách grafů.

Původní

Aktivací tlačítka „Původní“ dojde k obnovení původních hodnot posuvníků, tedy grafy budou zobrazeny bez zvětšení.

Vpravo

Pokud je na x-ové souřadnici zvětšeno rozlišení, lze se po ní posouvat. Tlačítkem „Vpravo“ se zobrazení posune vpravo.

Vlevo

Tlačítkem „Vlevo“ se zobrazení posune vlevo.

Jdi na vybrané

V případě, že je na hlavním grafu označen nějaký úsek nahrávky (viz dále), lze tuto část zobrazit přes celý graf tlačítkem „Jdi na vybrané“. Zpět se vrátí kliknutím levého tlačítka myši na libovolném místě v hlavním grafu. To lze využít k detailnějšímu přiblížení části průběhů bez změny nastavení posuvníků.

2.1.3 Přehrání

Přehraj cele

Přehraje celou nahrávku.

Přehraj zobrazené

Přehraje, tu část nahrávky, která je zobrazena v hlavním grafu.

Přehraj vybrané

V případě, že je na hlavním grafu vybrán nějaký úsek nahrávky (viz dále), přehraje tento úsek. (Přehraje tu část, která je mezi dvěma ukazateli.)

Přehraj do smyčky vybrané

Přehraje stejný úsek jako v případě „Přehraj vybrané“, ale přehrává ho do smyčky, tedy po ukončení přehrávání přehrává stejný úsek znovu (maximálně však 50x po sobě). Přehrávání lze ukončit tlačítkem „Stop“.

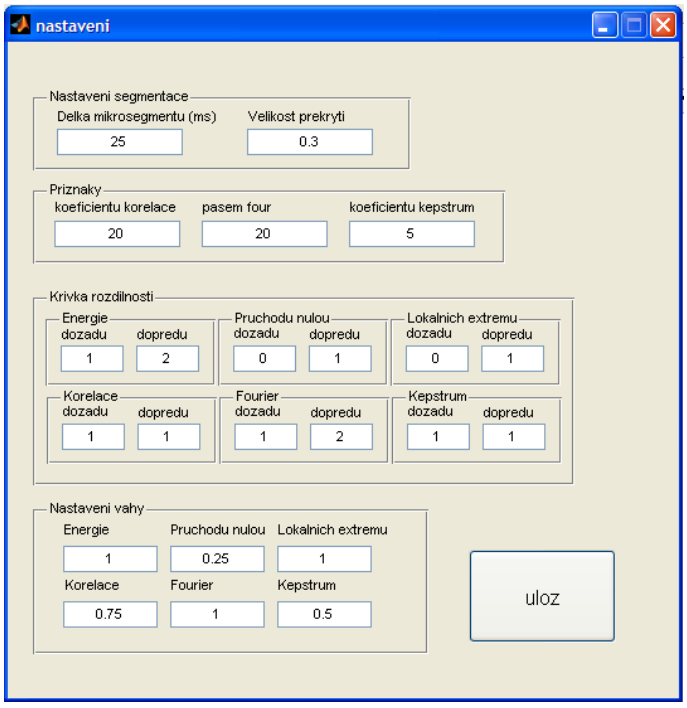
Stop

Umožňuje ukončit přehrávání, (nejen při přehrávání do smyčky).

2.1.4 Nastavení

Po stisku tlačítka „Nastavení“ se otevře okno umožňující změnu parametrů automatické segmentace.

- Délka mikrosegmentu: zadává se v mikrosekundách, udává délku rozdělení řečového signálu na mikrosegmenty.
- Velikost překrytí: zadává se jako zlomek z celkové délky mikrosegmentu, udává část mikrosegmentu o kolik se sousední mikrosegmenty překrývají.
- Koeficientu korelace: zadává se jako celé kladné číslo, udává kolik prvních autokorelačních koeficientů bude použito při dalším zpracování.
- Koeficientu kepstra: zadává se jako celé kladné číslo, udává kolik prvních koeficientů kepstra bude použito při dalším zpracování.
- Pasem four: zadává se jako celé kladné číslo, udává na kolik frekvenčních pásem bude rozděleno spektrum vypočtené Furierovou transformací.



Obr. 2.2: Grafické nastavení jednotlivých parametrů.

- Křivka rozdílnosti (dozadu, dopředu): zadává se jako celé číslo větší nebo rovno nule, udává koeficienty pro výpočet křivky rozdílnosti podle rovnice (1.6), dozadu odpovídá koeficientu x2, dopředu odpovídá koeficientu x1 v rovnici (1.6), pro každou metodu se zadává zvlášť.
- Nastavení váhy: zadává se jako kladné číslo, udává váhu s jakou se bude výpočet jednotlivých metod podílet na celkovém výsledku segmentace.

2.1.5 Analýzy

Po spuštění programu nejsou otevřeny grafy pro vykreslení pomocných funkcí, ty je možné aktivovat v záložce „Analýzy“ vybráním „on“ u prvního nebo druhého průběhu. Může být zobrazen jeden, dva či žádný pomocný průběh. Po zvolení zobrazení pomocného průběhu nad hlavním grafem zobrazí pomocný graf a v ovládacím panelu přibude záložka Graf 1 či Graf 2. Při přidání (odebrání) pomocného grafu se mění velikost okna programu.

2.1.6 Graf 1, Graf 2

Nabídky „Graf 1“, „Graf 2“ umožňují určit, jaké pomocné průběhy budou zobrazeny v grafech. Jedná se o průběhy křivek rozdílnosti všech použitých metod krátkodobé analýzy. Vybraná metoda je označena zatržítkem před jejím názvem.

2.2 Panel tlačítek

Panel tlačítek je k dispozici pod hlavním menu programu. Slouží k rychlé práci s programem. Jeho funkce je totožná s tlačítky v rámci menu programu. Jejich stručnější popis je uveden v předchozí kapitole.

Načtení nahrávky

Načte vybranou nahrávku řeči a vykreslí její časový průběh do hlavního grafu.

Výpočet

Slouží k výpočtu automatické segmentace a vykreslení průběhů do grafů.

Uložení rozdělení

Uloží výsledek ruční segmentace.

Načtení rozdělení

Načte výsledek ruční segmentace.

Přehraj

Přehraje celou nahrávku.

vybrané

Přehraje jednou vybranou část nahrávky (mezi dvěma ukazateli).

do smyčky

Přehraje opakovaně vybranou část nahrávky (mezi dvěma ukazateli).

Stop

Zastaví přehrávání.

Jdi na vybrané

Zobrazí vybranou část nahrávky (mezi dvěma ukazateli).

Původní

Zobrazí do okna celý časový průběh nahrávky.

2.3 Grafy pro vykreslení pomocných průběhů

Pro ruční segmentaci je výhodné zobrazit i křivky rozdílnosti jednotlivých použitých metod krátkodobé analýzy. K tomu slouží grafy pomocných průběhů. Může být zobrazen jeden, dva či žádný pomocný průběh. Průběhy mohou být kdykoliv v průběhu práce programu aktivovány (deaktivovány) v záložce „Analýzy“ v hlavním menu. Při přidání (odebrání) pomocného grafu se mění velikost okna programu. Při vykreslení pomocného průběhu se v menu přidá další položka („Graf 1“ či „Graf 2“), ve které je možné zvolit, který pomocný průběh má být zobrazen. Který průběh je zobrazen aktuálně, je vypsáno nad levou částí grafu. V grafu jsou kromě zvoleného průběhu křivky rozdílnosti zobrazeny i detekované hranice (znázorněny červenou barvou) a polohy ukazatelů (znázorněny světle modrou barvou). Hodnota souřadnice osy x je stejná jako v hlavním grafu.

2.4 Hlavní graf

Hlavní graf slouží k zobrazení průběhu řečového signálu (zelená), výsledné křivky rozdílnosti (tmavě modrá), určených hranic fonémů (červená) a případně poloh ukazatelů (světle modrá). A zároveň slouží i pro ruční segmentaci. Časová osa grafu je v milisekundách.

Stiskem levého tlačítka myši na grafu se na místo stisku přidá ukazatel, znázorněný světle modrou čarou přes graf. Po vytvoření dvou ukazatelů na grafu je možné obsah mezi nimi přehrát (například tlačítkem „Přehraj vybrané“) nebo přiblížit (tlačítkem „Jdi na vybrané“). Pokud jsou na grafu vytvořeny dva ukazatele, dalším stiskem levého tlačítka myši se tyto ukazatele smažou a znovu je možné vytvořit ukazatele na jiném místě.

Stiskem pravého tlačítka myši na grafu se vyvolá kontextové menu. Zde jsou čtyři funkce vztahující se k práci s hranicemi. Tlačítkem „Nová“ se na místě stisku přidá nová hranice, tlačítkem „Smazání“ se smaže hranice pokud se nachází v blízkém okolí stisku tlačítka. Tlačítkem „Vyjmout“ se hranice vyjme a přesune tím, že se vloží na vybrané místo tlačítkem „Vložit“. Pokud není žádná hranice vyjmuta, je aktivní tlačítko „Vyjmout“ a tlačítko „Vložit“ je neaktivní, pokud je hranice vyjmuta, je „Vyjmout“ neaktivní a „Vložit“ aktivní.

Posuvník pro zvětšení

Posuvník napravo od hlavního grafu slouží pro změnu rozlišení grafu na x-ové ose. Posunem posuvníku dolů se rozlišení zvětšuje.

Posuvník pro posuv

Posuvník pod hlavním grafem slouží pro posuv po x-ové ose pokud je rozlišení grafu zvětšeno. Velikost posuvníku udává velikost rozlišení a jeho poloha odpovídá poloze zobrazeného úseku v celé nahrávce. Pod tímto posuvníkem je vykreslený časový průběh celé nahrávky. Tvar posuvníku promítnutý v nahrávce udává zobrazenou část nahrávky na grafu. To umožňuje snadnější orientaci při zvětšeném rozlišení.

2.5 Práce s programem

2.5.1 Spuštění programu

Program je vytvořen v prostředí Matlab (verze 7.1.0.246), nejprve je tedy nutné spustit Matlab, jako aktuální adresář (Current Directory) nastavit „BakalarskaPrace“ a do příkazového okna zapsat „prace“ a potvrdit.

2.5.2 Načtení nahrávky a uloženého rozdělení

Po startu programu je nejprve nutné načíst nahrávku, se kterou chceme pracovat, tlačítkem „Načtení nahrávky“. Ihned po načtení nahrávky je vhodné provést výpočet automatické segmentace tlačítkem „Výpočet“. Pokud je již vytvořen soubor s výsledky ruční segmentace, lze jej načíst tlačítkem „Načtení segmentace“.

2.5.3 Ruční segmentace

Ruční segmentace se provádí na hlavním grafu. Pomocí pravého tlačítka myši se vyvolá kontextové menu. V něm zvolíme, jakou operaci na místě kliknutí chceme provést, zda vložit novou hranici či hranici smazat nebo přesunout.

Jako základ pro ruční segmentaci slouží hranice fonémů určené automatickou segmentací. Správné hranice potom určujeme z časového průběhu nahrávky z průběhů křivek rozdílnosti v pomocných grafech a především přehráním vybrané části nahrávky.

2.5.4 Výstup segmentace

Výsledky ruční segmentace lze uložit tlačítkem „Uložení segmentace“. Výsledky se uloží pod zvoleným názvem do zvolené složky jako textový soubor, který obsahuje sloupcový vektor čísel vzorků, ve kterých byly označeny hranice. Práce s programem se ukončí pomocí tlačítka „Konec“ v menu programu.

3 VÝSLEDKY PRÁCE

3.1 Hodnocení výsledků segmentace

Cílem bakalářské práce je vytvořit program, který by mohl být využit pro automatickou a ruční segmentaci řečového signálu. V rámci programu bylo vytvořeno několik skriptů, provádějících výpočet určité části segmentace.

Vytvořené skripty

1. `prace` - grafické uživatelské rozhraní,
2. `nastaveni` - nastavení vstupních parametrů segmentace,
3. `Segmentace` - rozdělení na mikrosegmenty,
4. `Energie` - krátkodobá energie,
5. `PruchoduNulou` - krátkodobá funkce střední hodnoty průchodů signálu nulou,
6. `LokEx` - krátkodobá funkce středního počtu výskytu lokálních extrémů,
7. `Korelace` - krátkodobá autokorelační funkce,
8. `Fourierova` - diskrétní Fourierova transformace,
9. `Kepstrum` - kepstrální analýza,
10. `VyhodnotData` - křivka rozdílnosti příznaků,
11. `Maxima` - stanovení hranic fonémů.

Jejich podrobnější popis se nachází v kap. 1.

Testované nahrávky

Program byl využit při segmentaci nahrávek pocházející od deseti různých mluvčích. Část nahrávek jsou kratší promluvy, jedná se o samostatná slova, konkrétně číslovky od nuly po deset: „nula“, „jedna“, „dvě“, „tři“, „čtyři“, „pět“, „šest“, „sedum“, „osum“, „devět“, „deset“.

U druhé části, delších nahrávek, se jedná o celé věty:

- „Hoří hospoda, zavolejte hasiče.“
- „Musíme přežít zimu.“

- „Potřebuji vyřešit tuto rovnici.“
- „Pozor na úraz elektrickým proudem.“
- „Přišel jsem včera večer pozdě.“
- „Severní vítr je krutý.“
- „Tatínek našel pěkné kotě.“
- „Vlak z Lelechovic do Brna dnes nejede.“
- „Vysyp všechny pytle s pšenicí.“

Chyby v segmentaci

Jak již bylo zmiňováno v předchozí kapitole při automatické segmentaci nelze zamezit vzniku chyb, ty mohou vzniknout z několika příčin. První z nich je, že průběh charakteristik (tedy následně i vypočtených příznaků) reálné řeči není často v rámci fonému úplně konstantní. Dále některé dvojice fonémů jsou oproti sobě velmi rozdílné, ale jiné jsou si částečně podobné, hranice mezi nimi jsou pak různě obtížně detekovatelné. Sousední fonémy se vlivem koartikulace ovlivňují, tudíž jednotlivé charakteristiky přecházejí postupně z jedné úrovně na druhou. Další vlastností řeči je velká rozdílnost délky segmentů, což může znamenat problém především u velmi krátkých segmentů. V neposlední řadě je pak možnost vzniku chyby závislá na řečníkovi a na kvalitě výslovnosti, některé hlásky mohou být vysloveny méně zřetelně či polknuty.

Při segmentaci mohou vzniknout chyby tří druhů:

1. hranice fonémů není detekována, přestože se v řeči nachází,
2. hranice je detekována chybně (tzn. uvnitř fonému),
3. rozpoznaná hranice je mírně posunutá oproti skutečné.

V případě, kdy hranice není rozpoznána či je automatickou segmentací označena v místě, kde skutečná hranice fonémů není, lze toto jednoznačně určit a označit jako chybu. Složitější případ nastává, kdy je hranice rozpoznána správně, jen posunuta oproti správné poloze. Vzhledem k vlivu koartikulace přesnou hranici mezi sousedními fonémy určit nelze, a to ani ruční segmentací. Při automatické segmentaci jsou hranice určovány ve vzdálenostech odpovídajících velikosti jednoho mikrosegmentu, proto nelze požadovat jejich nalezení s větší přesností. Hranice posunuté o méně než 25 ms oproti hodnotě určené ruční segmentací tedy nebudou považovány za chybně detekované.

3.2 Optimální nastavení parametrů segmentace

Většina skriptů obsahuje jako vstupní hodnotu nastavitelný parametr, kterým je možné ovlivnit výpočet uvnitř skriptu a tím i jeho výstup. Snaha je pak přiřadit všem těmto parametrům takové hodnoty, aby celý proces segmentace byl co nejúčinnější. Vzhledem k tomu že jsou zpracovávány nahrávky, jejichž charakter je různý, nelze nalézt optimální nastavení vstupních parametrů. Je nutné tedy hledat nejlepší možné řešení, které bude vyhovovat všem zpracovávaným nahrávkám.

Nastavení vstupních parametrů proběhlo pro prozkoumání vlivu na několika nahrávkách různého charakteru. Byly použity kratší i delší nahrávky a vždy od více mluvčích. U každé nahrávky byly určeny chyby v segmentaci, ve srovnání s ruční segmentací, pro jednu hodnotu parametru se chyby ve všech nahrávkách sečetly. Jako nejvhodnější pak bylo vybráno takové nastavení skriptu, při kterém byla suma chyb nejmenší. V některých skriptech se přihlíželo i k dalším důsledkům nastavení parametrů.

3.2.1 Nastavení vstupních parametrů skriptu `Segmentace`

Volba délky mikrosegmentů a jejich vzájemného překrytí je velmi důležitá, neboť ovlivňuje celý proces segmentace. Délka mikrosegmentů by měla být malá, aby bylo možné postihnout krátké fonémy, ale zároveň dostatečně velká, aby bylo možné zaznamenat kvaziperiodický charakter řečového signálu.

Ve skriptu `Segmentace` se délka mikrosegmentů zadává v mikrosekundách jako třetí vstup skriptu. Při malé délce mikrosegmentu jsou rozpoznány i ty nejkratší fonémy, chyba nenalezení hranice je zde tedy velmi malá, ale zároveň je mnoho hranic detekováno i uvnitř fonémů. S rostoucí délkou segmentu se snižuje počet chybně detekovaných hranic ale zvyšuje počet nedetekovaných. Musí se tedy hledat řešení, kdy počet těchto dvou druhů chyb je vyrovnaný a tím celkově chyb nejméně.

Vliv měnící se délky mikrosegmentů na velikost chyb v procesu segmentace byl testován na 6 nahrávkách:

a, „čtyři“,

b, „deset“,

c, „Hoří hospoda, zavolejte hasiče“,

d, „Musíme přežít zimu“,

e, „Potřebuji vyřešit tuto rovnici“,

f, „Severní vítr je krutý“.

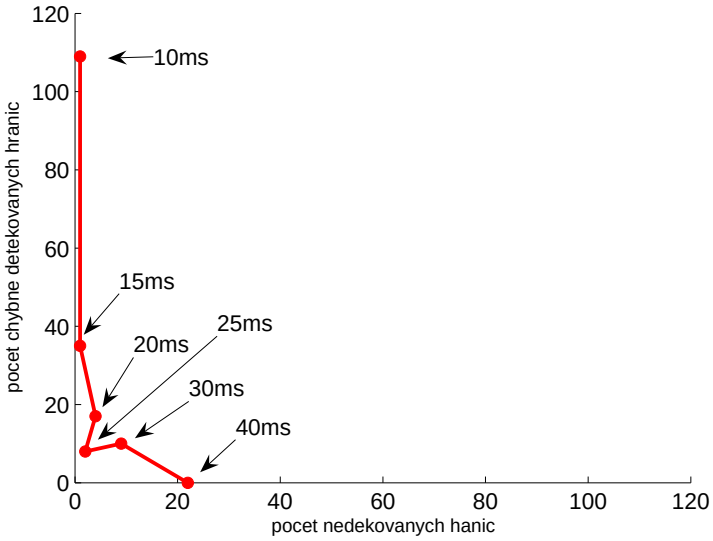
Tab. 3.1: Vliv délky mikrosegmentů na chyby v segmentaci.

	10ms		15ms		20ms	
nahravka	nenalezené	chybně	nenalezené	chybně	nenalezené	chybně
a,	0	8	0	2	0	2
b,	0	12	0	5	0	1
c,	1	25	1	12	3	5
d,	0	15	0	4	1	3
e,	0	29	0	6	0	4
f,	0	20	0	6	0	3
celkem	**1**	**109**	**1**	**35**	**4**	**17**
	25ms		30ms		40ms	
nahravka	nenalezené	chybně	nenalezené	chybně	nenalezené	chybně
a,	0	1	0	0	1	0
b,	0	1	0	1	0	0
c,	1	1	3	5	9	0
d,	0	0	1	0	4	0
e,	1	2	2	2	4	0
f,	0	3	3	2	4	0
celkem	**2**	**8**	**9**	**10**	**22**	**0**

V tab. 3.1 jsou uvedeny počty chyb automatické segmentace oproti ruční, při různé délce mikrosegmentu (10 ms, 15 ms, 20 ms, 25 ms, 30 ms a 40 ms). Jedná se o chyby, kdy hranice fonémů není detekována, přestože se v řeči nachází, a kdy je hranice detekována chybně (tzn. uvnitř fonému). Na obr. 3.1 je rozložení počtu chyb graficky znázorněno. Rozložení chyb na tomto obrázku má tvar podobný hyperboly, pokles při 25 ms je způsoben malým počtem testovaných nahrávek. Počet chyb byl určován pouze při měnící se délce mikrosegmentů, ostatní parametry byly nastaveny ve všech případech stejně. Pro změnu jiného parametru skriptů by vznikla jiná závislost. Z tab. 3.1, a ještě lépe z obr. 3.1, je patrné, že pro malé délky, konkrétně 15 ms, segmentace nalezne prakticky všechny hranice fonémů, ale kromě nich mnoho hranic nesprávně. Oproti tomu při užití dlouhých mikrosegmentů, 40 ms, nedojde k určení hranice tam, kde ve skutečnosti není, ale nenajde všechny hranice, které v promluvě jsou. Jako nejvhodnější délka mikrosegmentu bylo zvoleno 25 milisekund. Při této délce je účinnost segmentace nejvyšší.

Čtvrtá vstupní hodnota skriptu `Segmentace` udává, jaké bude překrytí sousedních mikrosegmentů. Testováním vlivu nastavení velikosti překrytí na nahrávkách, podobně jako v případě délky mikrosegmentů, bylo hledáno nejvhodnější nastavení

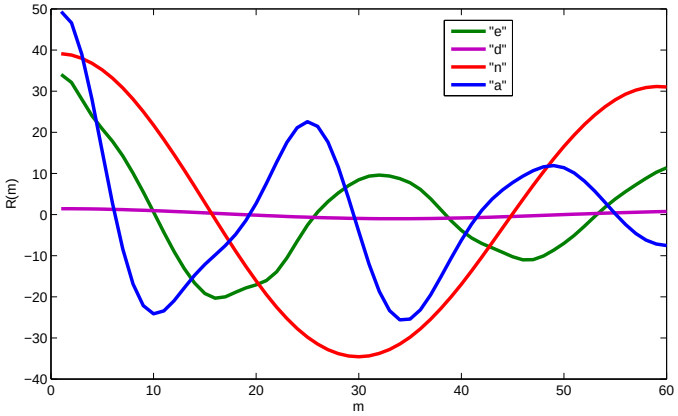
překrytí. Velikost překrytí nemá na segmentaci tak velký vliv jako délka mikrosegmentů. Jako nejvhodnější byla určena velikost překrývání sousedních mikrosegmentů 30% jejich délky.



Obr. 3.1: Závislost vzniku chyb na délce mikrosegmentu.

3.2.2 Nastavení vstupních parametrů skriptů `Korelace`, `Kepstrum` a `Fourierova`

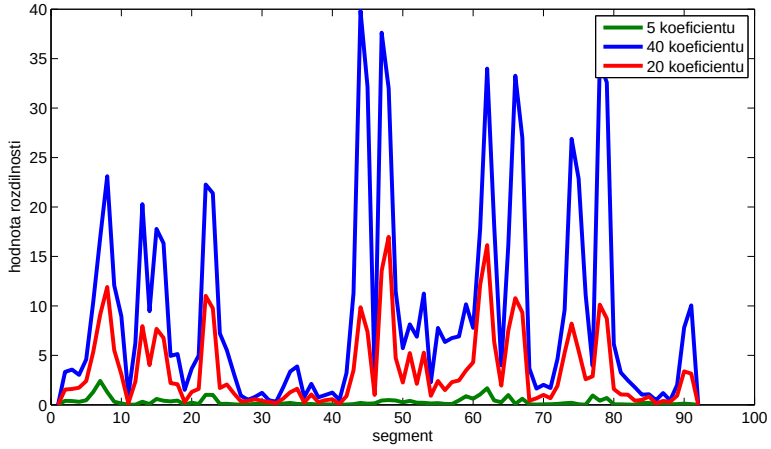
U skriptů `Korelace` a `Kepstrum` parametr udává, kolik prvních koeficientů bude použito k vyhodnocení křivky rozdílnosti. U skriptu `Fourierova` do kolika frekvenčních pásem bude spektrum rozděleno.



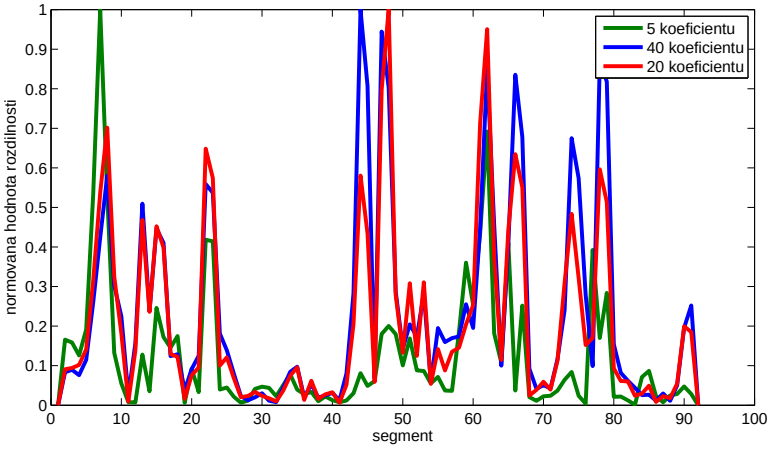
Obr. 3.2: Autokorelační koeficienty hásek „e“, „d“, „n“, „a“.

Na obrázku 3.2 je pro názornost zobrazeno 60 autokorelačních koeficientů čtyř hlásek ze slova „jedna“.

Užití většího počtu autokorelačních koeficientů vede ke zvýšení výpočetní náročnosti výpočtu křivky rozdílnosti pro krátkodobou autokorelační funkci. Není účinné tedy použít velký počet koeficientů, pokud k obdobnému výsledku lze dojít užitím i méně koeficientů.



a)



b)

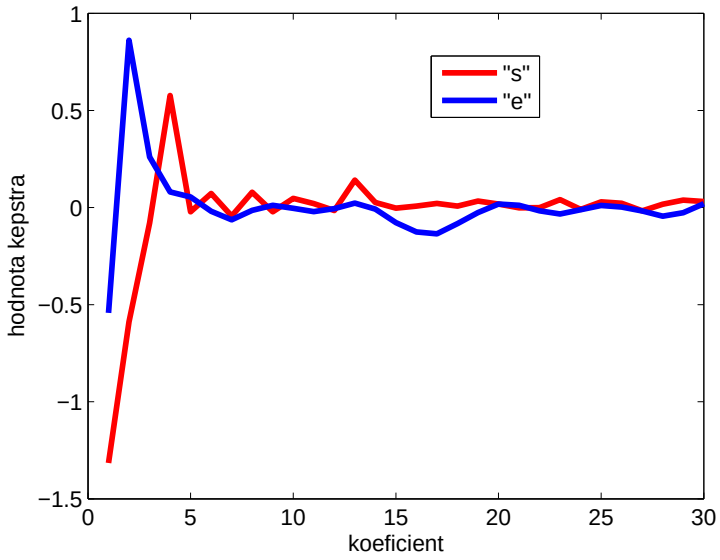
Obr. 3.3: Průběh rozdílnosti autokorelační funkce věty „Severní vítr je krutý“ a) vypočtené hodnoty a b) normované hodnoty.

Na obrázku 3.3 je vykreslen průběh křivky rozdílnosti autokorelační funkce pro větu „Severní vítr je krutý“. Je patrné, že při změně počtu autokorelačních koeficientů, ze kterých se tato křivka stanovuje, se mění její velikost více než tvar. Protože pro rozpoznání rozdílnosti je důležitý jen její tvar a protože před výpočtem výsledné

křivky rozdílnosti jsou všechny křivky rozdílnosti normovány (poděleny svou maximální hodnotou), je na obrázku 3.3 b, zobrazen normovaný průběh. Při užití malého počtu koeficientů není příliš využito tvaru autokorelačních křivek, proto křivka rozdílnosti vypočtena z méně koeficientů je málo kvalitní, detekuje méně hranic fonémů. Zato při užití středního a velkého počtu koeficientů dává výsledky obdobné. Optimální je užít 20 až 30 koeficientů. Jako vstupní parametr skriptu `Korelace` tedy bylo zvoleno 20 autokorelačních koeficientů.

Stejně jako u skriptu `Korelace` je výstup skriptu `Kepstrum` jen několika prvních koeficientů.

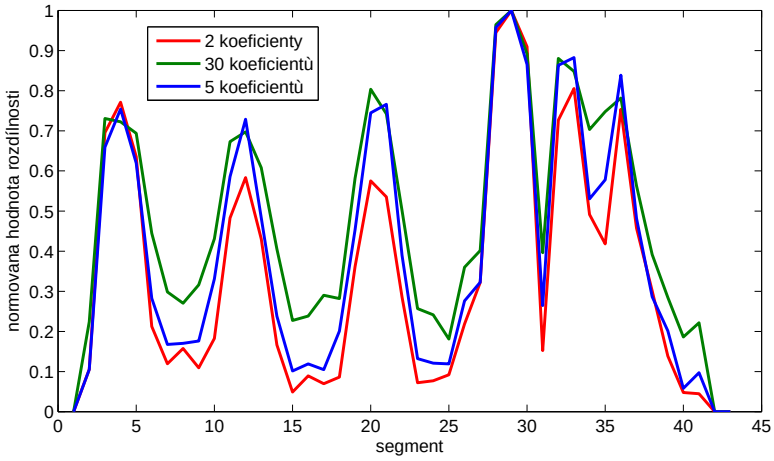
Na obrázku 3.4 jsou znázorněny průběhy velikostí prvních třiceti kepstrálních koeficientů hlásek „s“ a „e“. Lze na něm pozorovat, že nejvíce rozdílných je pouze několik prvních koeficientů, ostatní jsou si relativně podobné.



Obr. 3.4: Kepstrální koeficienty hlásek „s“ a „e“.

Na obrázku 3.5 jsou vykresleny normované průběhy křivky rozdílnosti kepstra při použití dvou, pěti a třiceti koeficientů. Všechny tyto tři průběhy se nijak zásadně neliší. Přesto křivka rozdílnosti pro 2 koeficienty dosahuje v některých maximech menších hodnot, což je dáno tím, že neobsahuje všechny nejvíce rozdílné koeficienty. A křivka rozdílnosti pro 30 koeficientů má minima položena níže, tedy je menší rozdíl pro maximum a minimum, což je způsobeno, že tato křivka byla určována i z podobnějších koeficientů. Jako počet koeficientů kepstra a tedy i vstupní parametr skriptu `Kepstrum` bylo zvoleno 5 koeficientů.

U skriptu `Fourierova` udává vstupní parametr do kolika frekvenční pásem bude



Obr. 3.5: Křivky rozdílnosti kepstra při použití dvou, pěti a třiceti koeficientů.

spektrum rozděleno. Pro nahrávky s vzorkovacím kmitočtem 16000 Hz a při rozdělení na mikrosegmenty 25 ms, je ve spektru 201 vzorků. Pásma by měla být natolik široká, aby se v rámci něj vyrovnaly odchylky na jednotlivých frekvencích a zároveň úzký, aby nebyl ovlivněn spektrální charakter mikrosegmentu. Jako nejvhodnější šířka pásma pro testované nahrávky bylo zvoleno kolem 10 vzorků spektra, tedy parametr skriptu `Fourierova` byl nastaven na 20.

Nastavení vstupních parametrů skriptů `Korelace`, `Kepstrum` a `Fourierova` má na kvalitu procesu segmentace jen částečný vliv (v porovnání například s nastavení délky segmentů). Přesto jejich velmi nevhodné nastavení by mělo za následek vznik chyb, proto je třeba jim věnovat pozornost.

3.2.3 Nastavení vstupních parametrů skriptu `VyhodnotData`

Jak již bylo uvedeno v části 1.3 vstupem skriptu `VyhodnotData` je vektor či matice obsahující hodnotu příznaku pro každý mikrosegment, a parametry x_2 a x_1. Parametr x_2 udává z kolik posunutého segmentu vzad se bude určovat hodnota rozdílnosti oproti počítanému segmentu, x_1 má stejný význam pouze udává posun vpřed. Pro správnou segmentaci by křivka rozdílnosti každého příznaku měla mít tyto vlastnosti:

1. velký rozdíl maxim a minim (maxima co nejvyšší, minima ideálně nulová),
2. maxima by měla být úzká (maximum by nemělo být detekováno na velkém počtu sousedních segmentů),
3. maxima by měla být přesně na rozhraní fonémů, neměla by být posunuta.

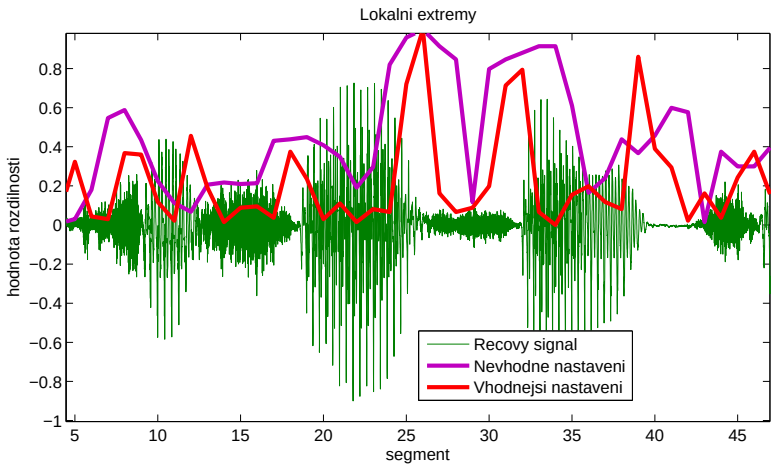
K docílení těchto požadavků je nutné vhodně nastavit parametry x_2 a x_1. Nejdůležitější z požadavků je, aby hodnota detekovaného maxima nebyla posunuta. V případě jeho nedodržení se může stát, že pomocí dvou příznaků jsou detekována maxima obě různě posunutá, po jejich sečtení by výsledná křivka rozdílnosti určila dvě chybné hranice fonémů, ale tu skutečnou ne.

Energie v jednom fonému často není úplně konstantní. Zvláště u hlásek s větší energií, které jsou vysloveny za a před hláskami s energií nízkou, dochází k tomu, že energie této hlásky postupně narůstá. V prostředním mikrosegmetu je největší a opět postupně klesá. Z tohoto důvodu je vhodné křivku rozdílnosti určovat z více vzdálených mikrosegmentů.

V případě krátkodobé funkce počtu průchodů nulou a funkce počtu lokálních extrémů je přechod těchto charakteristik na hranicích fonémů relativně strmý. Proto se může hodnota rozdílnosti těchto charakteristik určit přímo z hodnot sousedních mikrosegmentů. Při výpočtu z více vzdálených mikrosegmentů dochází k nežádoucímu posunu a rozšíření maxim na křivce rozdílnosti.

U autokorelační funkce a Fourierovy transformace a kepstrální analýzy bylo nejvhodnějšího průběhu křivky rozdílnosti docíleno při výpočtu z málo vzdálených mikrosegmentů.

Vliv vhodnosti nastavení vstupních parametrů skriptu `VyhodnotData` je zobrazen na obrázku 3.6. V tomto případě se jedná o výpočet křivky rozdílnosti počtu lokálních extrémů a lze pozorovat, že v případě správného nastavení maximum na křivce odpovídá hranici fonému, zatímco při nevhodném nastavení by nebylo možné všechny hranice správně detekovat.



Obr. 3.6: Vliv nastavení parametrů při výpočtu křivky rozdílnosti.

Nejvhodnější nastavení parametrů skriptu `VyhodnotData` je pro:

- krátkodobou energii $x_2 = 1$ $x_1 = 2$,
- krátkodobou funkci střední hodnota průchodů signálu nulou $x_2 = 0$ $x_1 = 1$,
- krátkodobou funkci středního počtu výskytu lokálních extrémů $x_2 = 0$ $x_1 = 1$,
- krátkodobou autokorelační funkci $x_2 = 1$ $x_1 = 1$,
- diskrétní Fourierovu transformaci $x_2 = 1$ $x_1 = 2$,
- kepstrální analýzu $x_2 = 1$ $x_1 = 1$.

3.2.4 Nastavení váhy metod krátkodobé analýzy pro stanovení křivky rozdílnosti

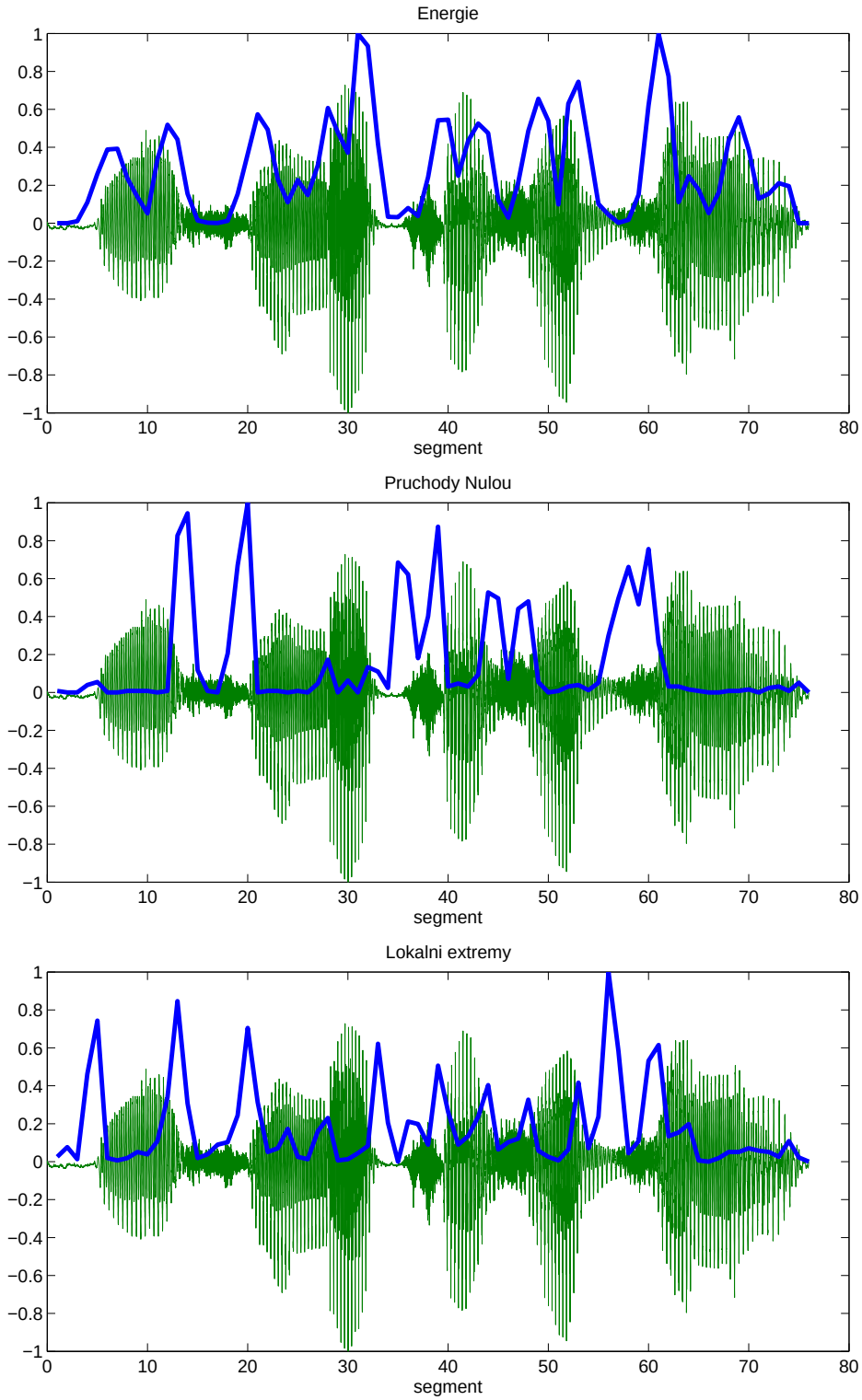
Po výpočtu křivek rozdílnosti každého příznaku, je každá tato křivka normována (podělena svou maximální hodnotou). V tomto místě mají tedy křivky rozdílnosti všech příznaků stejnou váhu. Neplatí ale, že by všechny metody byly stejně účinné, proto při stanovení výsledné křivky rozdílnosti lze přidělit každé metodě jinou váhu. Je tedy nutné prozkoumat účinnosti jednotlivých metod. Účinnost metody se stanoví na základě poměru správně detekovaných hranic fonémů vůči chybně detekovaným. Ještě je nutné brát ohled na to, že správně určené maximum (na rozhraní fonémů) musí mít velkou hodnotu. Dále je nutno brát ohled i na to, že nesprávně detekovaná maxima s nízkou hodnotou způsobí menší chybu, něž veliká nesprávná maxima. Ještě je nutné ověřit, jestli metoda nedává pouze ty výsledky, které jsou již přesvědčivě získány pomocí jiných metod. Nastavení vah metod předcházelo testování jejich účinností na řadě nahrávek.

Funkce počtu průchodů nulou nedokáže výrazně rozpoznat všechny hranice, ale jen část. Proto její váha byla nastavena na 0,25.

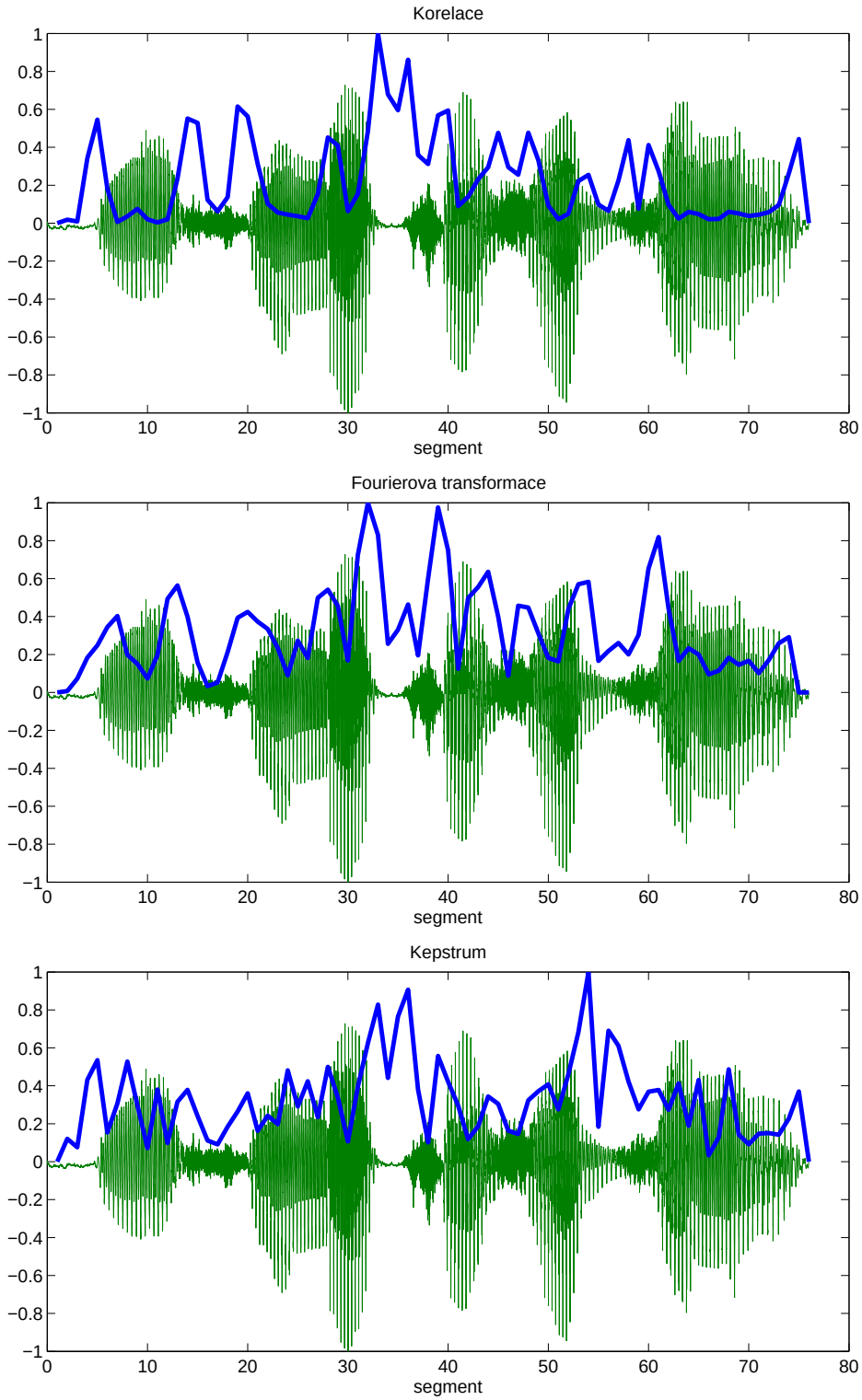
Autokorelační funkce najde většinu hranic, ale ty, které jsou obtížně detekovány ostatními metodami autokorelační funkce nenalezne také. Váha autokokelační funkce byla nastavena na 0,75.

Kepstrální analýzou lze detekovat většinu hranic fonémů a to i těch, které ostatní metody detekují jen nevýrazně nebo vůbec. Ale na její křivce rozdílnosti se nachází i mnoho vysokých maxim v bodech uprostřed fonému, proto její váha byla nastavena na 0,5.

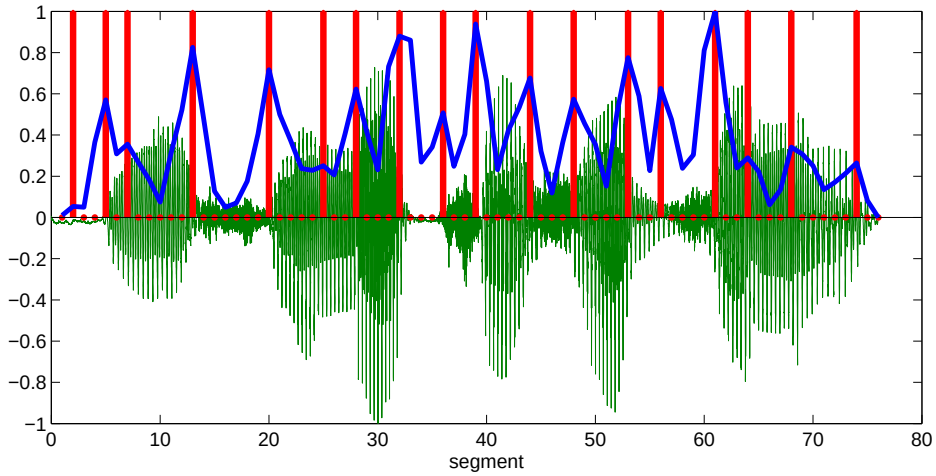
Funkce krátkodobé energie, počtu lokálních extrémů a krátkodobé Fourierovi transformace umožní nalézt velkou část hranic fonémů, aniž by určily hranice nesprávně. Jejich váha byla nastavena na 1.



Obr. 3.7: Křivky rozdílnosti věty „Musíme přežít zimu“ vypočteny z krátkodobé energie, krátkodobé funkce počtu průchodů nulou a krátkodobé funkce počtu lokálních extrémů.



Obr. 3.8: Křivky rozdílnosti věty „Musíme přežít zimu“ vypočteny z krátkodobé autokorelační funkce, krátkodobé Fourierovi transformace a krátkodobé kepstrální analýzy.



Obr. 3.9: Křivka rozdílnosti věty „Musíme přežít zimu“ s vyznačenými maximy.

Na obrázku 3.9 je vykreslen průběh křivky rozdílnosti věty „Musíme přežít zimu“ a naznačeno rozdělení na fonémy. Na obrázcích 3.7, 3.8 jsou zachyceny průběhy křivek rozdílnosti všech použitých metod krátkodobé analýzy.

3.3 Výsledky segmentace pro testované nahrávky

Funkce skriptu byla ověřena testováním nahrávek. Pro každou nahrávku byla vypočtena segmentace pomocí programu a následně v grafickém prostředí provedena ruční segmentace. Rozdílně určené hranice byly určeny jako chyba v automatické segmentaci. Chyba byla započtena při nenalezení hranice fonému, či jejím určení nesprávně (uprostřed fonému). Jako chyba nebylo považováno špatné určení hranic na začátku a konci nahrávek, tedy nalezení nesprávné hranice na konci nahrávky (při doznívání poslední hlásky) a neurčení hranice na začátku prvního fonému. Výsledky testování jsou zaznamenány v tabulkách uvedených v příloze. Kromě toho, že chyby byly rozpoznány, bylo po část z nich určeno, kde k nim dochází.

Výsledky pro krátké nahrávky

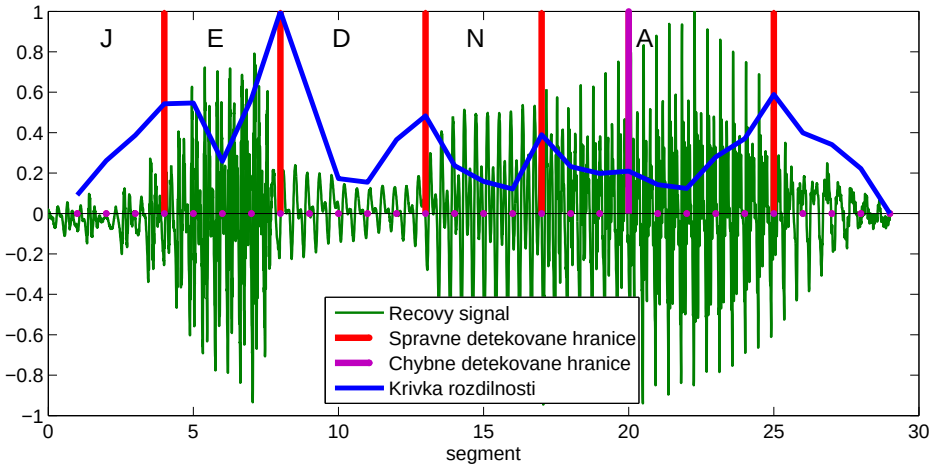
Krátké nahrávky jsou vyslovené číslovky od nuly do desíti: „nula“, „jedna“, „dvě“, „tři“, „čtyři“, „pět“, „šest“, „sedum“, „osum“, „devět“, „deset“.

Při segmentaci kratších nahrávek prakticky nedocházelo k tomu, že by hranice mezi fonémy nebyla rozpoznána. Naproti tomu zde vzniká velký počet chyb, kdy je určena hranice fonému nesprávně. To je dáno charakterem těchto nahrávek. Doba vyslovení jednoho fonému je u těchto nahrávek delší než je tomu pro stejný foném u nahrávek vět, tedy na stejný foném zde připadá více mikrosegmentů. Z tohoto

důvodu by pro kratší nahrávky mohly být nastaveny parametry segmentace jinak, než pro delší nahrávky, především délka mikrosegmentu. Byla snaha parametry segmentace nastavit tak, aby bylo možné v rámci stejného nastavení zpracovat různé typy nahrávek. Dalším znakem nahrávek číslic je to, že naprostá většina hlásek zde byla vyslovena poměrně zřetelně.

Tabulky C.1 a C.2 obsahují vyhodnocení všech nahrávek od prvních čtyř mluvčích. Kromě počtu chyb jsou v tabulkách uvedeny i počty fonémů v každém slově, aby mohla být porovnána relativní účinnost. V tabulce C.3 je vyhodnocení několika vybraných nahrávek od ostatních mluvčí. Výsledky pro stejnou nahrávku od různých mluvčích se mohou značně lišit a to především v závislosti na charakteru a kvalitě promluvy. U každé tabulky je uveden celkový počet chyb a součet počtu všech fonémů ve slovech. Lze tedy vypočítat, kolik chyb průměrně připadá na určitý počet fonémů.

Na obrázku 3.10 je vykreslen výsledek segmentace nahrávky „jedna“ od pátého mluvčího. Jsou zde znázorněny správně detekované hranice (vyznačeny červeně) a chybně detekované (vyznačeny fialově).



Obr. 3.10: Segmentace slova „jedna“.

U krátkých nahrávek byla chyba téměř vždy způsobena určením většího počtu hranic fonémů než je skutečný. Největší část nesprávně určených hranic se nacházela uprostřed dlouze vyslovených fonémů, konkrétně v těchto případech:

- uprostřed samohlásek (20x),
- uprostřed hlásek „m“ a „n“ (12x),
- uprostřed hlásek „s“, „š“, „č“ a „ř“ (9x).

Za každým bodem je v závorce uvedeno, kolikrát právě tato chyba způsobila nenalezení hranice. Druhá velká část nesprávně určených hranic byla u hlásek „k“, „p“ a především hlásky „t“. Tyto hlásky mají velmi malou energii a uprostřed svého průběhu charakter rázu a právě tento ráz je detekován jako samostatný foném, často jsou zde detekovány dvě nesprávné hranice. Chyba v těchto hláskách byla určena dvanáctkrát. Správná hranice nebyla nalezena pouze jedenkrát a to na hranicích hlásek „n“ a „u“.

Celkem bylo testováno 54 nahrávek číslic od 10 různých mluvčí obsahujících 230 fonémů. V těchto nahrávkách 1 hranice fonémů nebyla rozpoznána a 80 hranic fonémů bylo detekováno chybně.

Výsledky pro delší nahrávky

Delšími nahrávkami jsou celé věty.

a: „Hoří hospoda, zavolejte hasiče.“

b: „Musíme přežít zimu.“

c: „Potřebuji vyřešit tuto rovnici.“

d: „Pozor na úraz elektrickým proudem.“

e: „Přišel jsem včera večer pozdě.“

f: „Severní vítr je krutý.“

g: „Tatínek našel pěkné kotě.“

h: „Vlak z Lelechovic do Brna dnes nejede.“

ch: „Vysyp všechny pytle s pšenicí.“

V tabulkách C.4 a C.5 jsou uvedeny počty chyb nalezených v nahrávkách všech vět od čtyř mluvčích v tabulce C.6 pak chyby nalezené v nahrávkách „Přišel jsem včera večer pozdě“ a „Tatínek našel pěkné kotě“, od všech mluvčí.

V testovaných nahrávkách byl přibližně stejný počet chyb způsobených nenalezením hranice jako chybným nalezením. Každý druh z těchto chyb byl charakteristický pro určité hlásky. K chybám, kdy hranice fonémů nebyla detekována, docházelo především v těchto případech. Když:

- „m“ nebo „n“ sousedí se samohláskou (11x),
- „r“ sousedící s hláskou „m“, „n“ a nebo samohláskou (8x),

- sousedí spolu dvě hlásky s velmi malou energií (např. „k“, „t“, „p“) (4x),
- sousedí spolu hlásky „t“ nebo „d“ a „ř“ nebo „z“ (3x),
- sousedí spolu hlásky „v“ a „n“ (3x),
- při výskytu krátké hlásky „d“, „d´“, „j“ a především „l“ (15x).

Za každým bodem je v závorce uvedeno, kolikrát právě tato chyba způsobila nenalezení hranice. Další nalezené chyby se vyskytovaly pouze ojediněle, jednalo se o hranice fonémů „v“ a „š“, „i“ a „š“, „k“ a „n“ a fonem „t“ mezi dvěma samohláskami. Nejvíce chyb bylo způsobeno nenalezením krátkých hlásek, tyto hlásky se většinou vlivem koartikulace váží k některé delší hlásce s větší energií. Často je tato chyba i ovlivněna kvalitou výslovnosti. Foném „l“ ve většině případů, kdy se ve větě nalézal, nebyl rozpoznán. Další častá chyba nastává na hranici fonémů „m“ nebo „n“ a samohlásky, což je dáno jejich podobným energetickým i spektrálním charakterem. Specifickým je foném „r“, který se vyskytuje v rozdílných alofonických realizacích podle toho, kde ve slově se nachází.

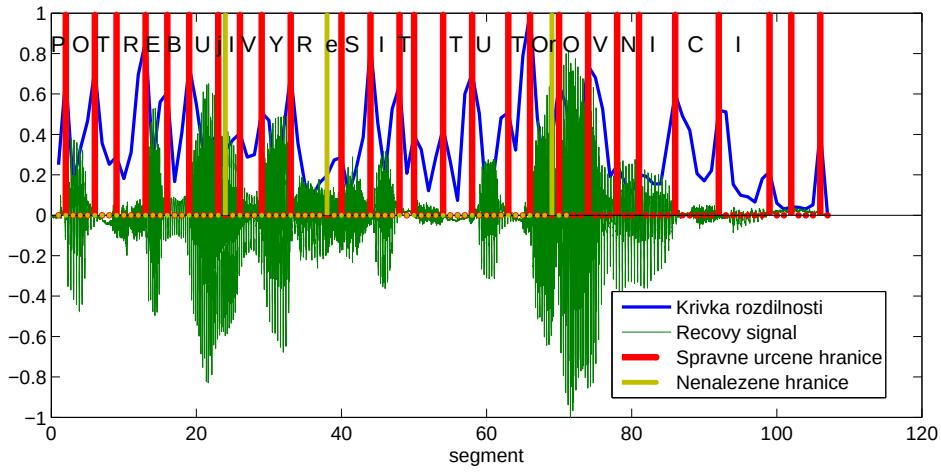
K chybám, kdy hranice fonémů byla detekována nesprávně, docházelo ve stejných případech jako u krátkých nahrávek. Tedy:

- uprostřed samohlásek (17x),
- uprostřed hlásek „s“, „š“, „č“ a „ř“ (17x),
- uprostřed hlásek „t“, „p“ a „k“ (16x),
- uprostřed hlásek „m“ a „n“ (4x).

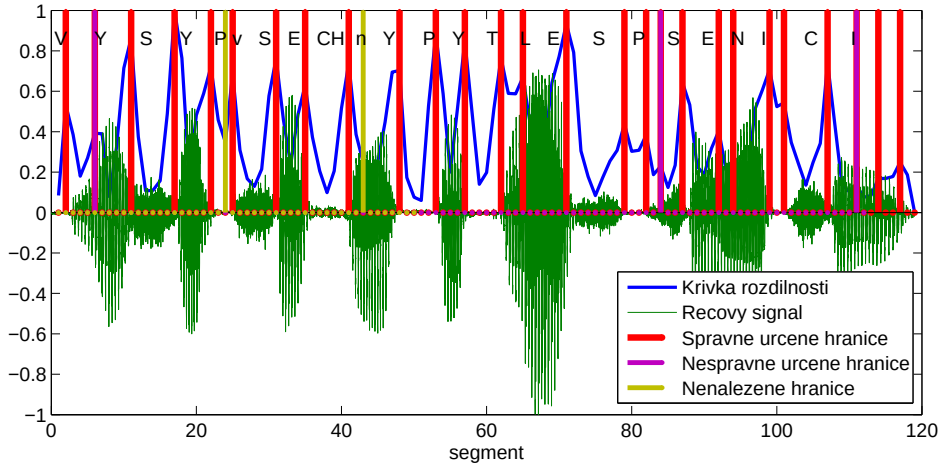
Dále bylo určeno několik hranic při výskytu pomlky mezi slovy, ale i v rámci jednoho slova.

Celkem bylo testováno 46 nahrávek 9 různých vět od 10 mluvčí, které obsahovaly celkem 1111 fonémů. Z nich bylo 1030 detekováno správně, tedy 81 hranic fonémů bylo nenalezeno. Dalších 108 hranic bylo detekováno chybně.

Na obrázku 3.11 jsou zobrazeny výsledné grafy pro dvě testované věty. Je zde vykreslen průběh řečového signálu a křivka rozdílnosti příznaků a dále červeně označené správně rozpoznané hranice, fialově nesprávně určené hranice a žlutě nerozpoznané. Některé nerozpoznané hranice bylo obtížné nalézt i při ruční segmentaci, neboť z průběhu signálu nebyly prakticky určitelné a i při přehrání nebyly příliš zřetelné.



a)



b)

Obr. 3.11: Výsledek segmentace pro věty a) „Potřebuji vyřešit tuto rovnici.“ b) „Vysyp všechny pytle s pšenicí.“

4 ZÁVĚR

Pro automatickou segmentaci byl v prostředí Matlab vytvořen program `prace`, který díky grafickému rozhraní umožňuje automatickou i ruční segmentaci. Tento program k provedení segmentace využívá dalších deset vytvořených skriptů. K rozdělení na mikrosegmenty slouží skript `Segmentace`. Pro určení příznaků krátkodobé analýzy slouží skripty `Energie`, `PruchoduNulou`, `LokEx`, `Korelace`, `Fourierova` a `Kepstrum`. Výpočet křivky rozdílnosti provádí skript `VyhodnotData`. K určení maxim na křivce rozdílnosti, tedy hranic fonémů, slouží skript `Maxima`. A pro nastavení vstupních parametrů segmentace při běhu programu byl vytvořen skript `nastaveni`.

Pro vytvořené skripty byly hledány jejich nejvhodnější vstupní parametry tak, aby bylo docíleno co nejspolehlivější segmentace. Nastavení parametrů všech skriptů je uvedeno v části 2.2. Dále byly porovnány křivky rozdílnosti jednotlivých metod krátkodobé analýzy, aby bylo možné stanovit jejich vhodnost pro segmentaci. Podle jejich účinnosti jim byl přidělen koeficient váhy, který udává jakou měrou se budou podílet na výpočtu výsledné křivky rozdílnosti. Jako nejefektivnější byly vyhodnoceny příznaky určené z krátkodobé energie, krátkodobé funkce středního počtu výskytu lokálních extrémů a krátkodobé diskrétní Fourierovi transformace. Nejméně efektivní se pak jevila krátkodobá funkce střední hodnoty průchodů signálu nulou.

Funkce skriptů a spolehlivost segmentace byla testována na přiložených nahrávkách. Tyto nahrávky pocházejí od deseti různých mluvčích. Část z nich jsou kratší nahrávky, slova, druhá část delší nahrávky, věty. Tyto testované nahrávky byly nejprve segmentovány pomocí vytvořených skriptů. Následně byla v grafickém rozhraní provedena ruční segmentace a výsledky srovnány. Výsledky ruční segmentace jsou uloženy v rámci práce.

Celkem bylo testováno 54 nahrávek číslic, kratších nahrávek, od 10 různých mluvčí obsahujících 230 fonémů. V těchto nahrávkách nebyla 1 hranice fonémů rozpoznána a 80 hranic fonémů bylo detekováno chybně. Delších nahrávek bylo testováno 46, skládajících se z 9 různých vět od 10 mluvčích, které obsahovaly celkem 1111 fonémů. Z nich bylo 1030 detekováno správně, 81 hranic fonémů bylo nenalezeno a dalších 108 hranic bylo detekováno chybně.

Chyby, kdy hranice fonémů nebyla detekována nebo kdy hranice fonémů byla detekována nesprávně, nastávaly především u některý fonémů (např. hláska „l“) a na hranicích dvou fonémů (např. hláska „n“ a samohláska). Další případy, kdy docházelo k opakovaným chybám při segmentaci, jsou vyjmenovány v části 3.3.

LITERATURA

[1] PSUTKA, J. *Komunikace s počítačem mluvenou řečí.* 1. vydání. Praha: Academia, 1995. 287 s. ISBN 80-200-0203-0

[2] PSUTKA, J. et al. *Mluvíme s počítačem česky.* 1. vydání. Praha: Academia, 2006. 752 s. ISBN 80-200-1309-0

[3] BUČEK, J. Segmentace řeči na fonémy metodou sledování rozdílnosti příznaků. In *Sborník prací studentů a doktorandů FEI VUT.* Brno: Akademické nakladatelství CERM , 1997. s. 73-75. ISBN 80-214-0637-2

[4] DOŇAR, B.- ZAPLATÍLEK, K. *Matlab začínáme se signály.* 1. vydání. Praha: Ben, 2006. 272 s. ISBN 80-7300-200-0

[5] DOŇAR, B.- ZAPLATÍLEK, K. *Matlab tvorba uživatelských aplikací.* 1.dotisk 1. vydání. Praha: Ben, 2005. 216 s. ISBN 80-7300-133-0

[6] Rybička, J. *LATEX pro začátečníky.* 3. vydání. Brno: KONVOJ, 2003. 238 s. ISBN 80-7302-049-1

SEZNAM PŘÍLOH

A OBSAH PŘILOŽENÉHO CD

CD: BakalarskaPrace		–vytvořené skripty
	nahrávky	–nahrávky řeči pro testování
	segmentované	–segmentace nahrávek
Segmentace LaTex		–zdrojové soubory práce
segmentace.pdf		–vypracovaná bakalářská práce
metadata.pdf		–metadata k práci
popis.txt		–informace o CD

B FONETICKÁ ABECEDA SAMPA

Tab. B.1: Význam některých speciálních znaků použitých v abecedě SAMPA, převzato z [2]

SAMPA	Definice	SAMPA	Definice
H	vysoká výška	/	stoupající tón
L	nízká výška	\	klesající tón
=	stejný tón	:	značka délky
−	nižší výška	”	hlavní přízvuk
+	vyšší výška	%	vedlejší přízvuk

Tab. B.2: Fonetická abeceda SAMPA pro český jazyk, převzato z [2]

SAMPA	Slovo	Transkripce	SAMPA	Slovo	Transkripce
i	*lis*	lis	p	*pec*	pet_s
e	*pes*	pes	b	*bratr*	bratr=
a	*sad*	sad	t	*tuk*	tuk
o	*kov*	kof	d	*dům*	du:m
u	*sukně*	sukJe	c	*tělo*	celo
i:	*víno*	vi:no	J\	*děda*	J\eda
e:	*lék*	le:k	k	*kost*	kost
a:	*sál*	sa:l	g	*tygr*	tigr=
o:	*kód*	ko:t	m	*muž*	muS
u	*růže*	ru:Ze	n	*nos*	nos
o_u	*bouda*	bo_uda	J	*laňka*	laJka
a_u	*auto*	a_uto	t_s	*cena*	t_sena
e_u	*euro*	e_uro	t_S	*oči*	ot_Si
f	*fík*	fi:k	d_z	*leckdo*	led_zgdo
v	*vítr*	vi:tr=	d_Z	*džbán*	d_Zba:n
s	*sůl*	su:l	N	*tango*	taNgo
z	*koza*	koza	M	*tramvaj*	traMvaj
S	*škola*	Skola	G	*abych byl*	"abiG bil
Z	*žena*	Zena	Q\	*tři*	tQ\i
x	*chata*	xata	r=	*krk*	kr=k
h\	*hůl*	h\u:l	l=	*vlk*	vl=k
l	*vlak*	vlak	m=	*osm*	osm=
r	*rok*	rok	?	*Ano.*	?ano
P\	*moře*	moP\e	@	*www*	v@v@v@
j	*jev*	jev			

C TABULKY VÝSLEDKŮ

Tab. C.1: Chyby vzniklé při segmentaci krátkých nahrávek prvního a druhého mluvčího.

	mluvčí 1		mluvčí 2		
nahrávka	nenalezené	chybně	nenalezené	chybně	fonemů
čtyři	0	0	0	1	5
deset	0	2	0	0	5
devět	0	2	0	2	5
dvě	0	0	0	0	3
jedna	0	0	0	2	5
nula	0	1	0	0	4
osum	0	3	0	1	4
pět	0	2	0	1	3
sedum	0	3	0	2	5
šest	0	2	0	0	4
tři	0	0	0	1	3
celkem	0	15	0	10	46

Tab. C.2: Chyby vzniklé při segmentaci krátkých nahrávek třetího a čtvrtého mluvčího.

	mluvčí 3		mluvčí 4		
nahrávka	nenalezené	chybně	nenalezené	chybně	fonemů
čtyři	0	0	0	1	5
deset	0	1	0	2	5
devět	0	0	0	2	5
dvě	0	2	0	2	3
jedna	0	1	0	3	5
nula	1	1	0	3	4
osum	0	2	0	4	4
pět	0	3	0	1	3
sedum	0	1	0	2	5
šest	0	3	0	3	4
tři	0	1	0	3	3
celkem	1	15	0	26	46

Tab. C.3: Chyby vzniklé při segmentaci krátkých nahrávek ostatních mluvčí.

nahrávka	mluvčí	nenalezené	chybně	fonémů
jedna	5	0	1	5
deset	5	0	0	5
dvě	6	0	2	3
pět	6	0	0	3
sedum	7	0	3	5
devět	7	0	1	5
tři	8	0	2	3
osum	8	0	2	4
čtyři	9	0	1	5
šest	9	0	2	4
celkem		0	14	46

Tab. C.4: Výsledky segmentace delších nahrávek prvních dvou mluvčích.

	mluvčí 1		mluvčí 2		
nahrávka	nenalezené	chybně	nenalezené	chybně	fonémů
Hoří hospoda…	1	4	1	4	26
Musíme přežít…	1	0	1	2	16
Potřebuji vyřešit…	2	2	1	3	27
Pozor na…	2	2	4	3	29
Přišel jsem…	3	1	2	2	25
Severní vítr…	0	3	1	2	18
Tatínek našel…	0	2	2	3	22
Vlak z…	1	5	3	1	30
Vysyp všechny…	1	3	1	3	26
celkem	11	22	16	23	219

Tab. C.5: Výsledky segmentace delších nahrávek od třetího a čtvrtého mluvčího.

	mluvčí 3		mluvčí 4		
nahrávka	nenalezené	chybně	nenalezené	chybně	fonémů
Hoří hospoda…	0	3	3	2	26
Musíme přežít…	1	1	1	1	16
Potřebuji vyřešit…	3	2	2	3	27
Pozor na…	5	3	4	2	29
Přišel jsem…	3	0	3	2	25
Severní vítr…	2	2	3	2	18
Tatínek našel…	2	2	3	3	22
Vlak z…	4	3	2	2	30
Vysyp všechny…	2	1	1	1	26
celkem	22	17	21	18	219

Tab. C.6: Výsledky segmentace nahrávek „Přišel jsem včera večer pozdě“ a „Tatínek našel pěkné kotě“.

	Přišel jsem…		Tatínek našel…	
mluvčí	nenalezené	chybně	nenalezené	chybně
1	3	1	0	2
2	2	2	2	3
3	3	0	2	2
4	3	2	3	3
5	2	1	2	3
6	2	2	1	3
7	0	3	2	1
8	0	5	0	5
9	0	2	2	3