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

ROZPOZNÁVÁNÍ MLUVČÍHO

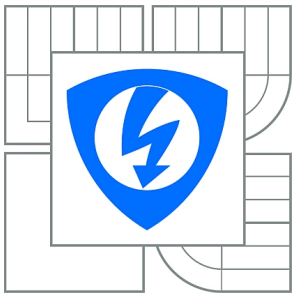
BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE LADISLAV KAŠPAR
AUTHOR

BRNO 2013

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

ROZPOZNÁVÁNÍ MLUVČÍHO

SPEAKER RECOGNITION

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

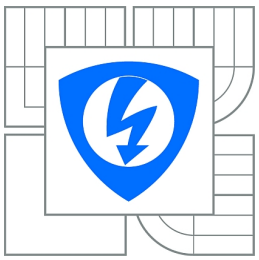
AUTOR PRÁCE — LADISLAV KAŠPAR

AUTHOR

VEDOUCÍ PRÁCE — Ing. PETR SYSEL, Ph.D.

SUPERVISOR

BRNO 2013



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky a komunikačních technologií

Ústav telekomunikací

Bakalářská práce

bakalářský studijní obor
Teleinformatika

Student:	Ladislav Kašpar	***ID:***	125241
Ročník:	3	***Akademický rok:***	2012/2013

NÁZEV TÉMATU:

Rozpoznávání mluvčího

POKYNY PRO VYPRACOVÁNÍ:

Seznamte se s metodami automatického rozpoznávání mluvčího nezávislými na obsahu promluvy a s otevřenou skupinou mluvčích. Vybranou metodu implementujte v prostředí Matlab a otestujte na zkušebních nahrávkách. Zhodnoťte spolehlivost rozpoznávače.

DOPORUČENÁ LITERATURA:

[1] Psutka, J.; Müller, L.; Matoušek, J.; Radová, V. Mluvíme s počítačem česky. 1. vydání. Praha: Academia, 2006. 752 s. ISBN 80-200-1309-1
[2] Deller, J. R.; Hansen, J. H. L.; Proakis, J. G. Discrete-Time Processing of Speech Signals. New York: IEEE Press, 2000. ISBN 0-7803-5386-2

Termín zadání:	11.2.2013	***Termín odevzdání:***	5.6.2013

Vedoucí práce: Ing. Petr Sysel, Ph.D.
Konzultanti bakalářské práce:

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

ABSTRAKT

Ve své bakalářské práci se věnuji problému rozpoznávání mluvčího. Tato práce obsahuje základní teorie k tomuto tématu. Teorie je zaměřena na výpočet parametrů pro rozpoznávání mluvčího a na popis postupu při rozpoznávání mluvčího. Jako hlavní parametry v programu na rozpoznávání mluvčího napsaného v jazyce Matlab využívám kmitočty formantů, kepstrální koeficienty a segmentaci signálu.

KLÍČOVÁ SLOVA

Rozpoznávání mluvčího, verifikace mluvčího, identifikace mluvčího, vytváření řeči, segmentace, analýza řečového signálu, formanty.

ABSTRACT

My bachelor thesis is devoted to the problem of speaker recognition. It includes the basic theory on this topic. The theory focuses on the calculation of parameters for speaker recognition and description of the procedure for speaker recognition. An application for speaker recognition has been written in Matlab. It uses techniques as frequency formants, cepstral coefficients and segmentation of the signal as the main parameters.

KEYWORDS

Speaker recognition, speaker verification, speaker identification, speech production, segmentation, analysis of the speech signal, formants.

KAŠPAR, Ladislav *Rozpoznávání mluvčího*: bakalářská práce. BRNO: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2013. 44 s. Vedoucí práce byl Ing. Petr Sysel, Ph.D.

PROHLÁŠENÍ

Prohlašuji, že svou bakalářskou práci na téma „Rozpoznávání mluvčího“ 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/nebo majetkových a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

BRNO ....

(podpis autora)

PODĚKOVÁNÍ

Rád bych poděkoval vedoucímu bakalářské práce panu Ing. Petru Syslovi, Ph.D. za odborné vedení, konzultace, trpělivost, materiály a podnětné návrhy k práci.

BRNO ....

(podpis autora)

Faculty of Electrical Engineering
and Communication
Brno University of Technology
Purkynova 118, CZ-61200 Brno
Czech Republic

http://www.six.feec.vutbr.cz

PODĚKOVÁNÍ

Výzkum popsaný v této bakalářské práci byl realizován v laboratořích podpořených z projektu SIX; registrační číslo CZ.1.05/2.1.00/03.0072, operační program Výzkum a vývoj pro inovace.

BRNO ....

(podpis autora)

EVROPSKÁ UNIE
EVROPSKÝ FOND PRO REGIONÁLNÍ ROZVOJ
INVESTICE DO VAŠÍ BUDOUCNOSTI

OBSAH

SEZNAM OBRÁZKŮ

SEZNAM TABULEK

ÚVOD

Ve své bakalářské práci se věnuji problematice rozpoznávání mluvčích a tvorbě programu na rozpoznání mluvčího, který bude nezávislý na obsahu promluvy řečníka. V této práci je popsána část teorie k problematice rozpoznávání mluvčích, zpracování řečového signálu, výpočet parametrů pro zjištění mluvčího, návrh a realizace programu pro rozpoznávání mluvčích.

V současné době se řadí do biometrického zabezpečení i identifikace a verifikace mluvčího podle jeho hlasu. Velké množství amerických i evropských firem ve spolupráci s jinými organizacemi (vysoké školy, univerzity, ...) pracují na systémech umožňující identifikaci a verifikaci mluvčího. U biometrických systémů na rozpoznání mluvčího se zadává předem stanovená promluva – přístupové heslo. Systémy umožňující rozpoznávání mluvčího používající se k zabezpečení dat a objektů vyžadují téměř 100% účinnost. Program vytvořený k této práci není určen pro zabezpečení, proto není nutné, aby byl 100% účinný v rozpoznávání. Od tohoto programu se vyžaduje rozpoznávání nezávislé na obsahu promluvy, které je mnohem náročnější. I přes vysokou profesionální úroveň výsledků, jakých tyto programy dosahují nejsou 100% účinné. Aby bylo dosaženo co nejvýšší bezpečnosti objektů či dat jsou systémy na rozpoznávání mluvčího kombinovány s jinými biometrickými systémy (např. doplnění o snímání otisku prstu – přihlášení k počítači nebo biometrický senzor oka – zabezpečení vstupních dveří. Přesnost systémů k ident ifikaci a verifikaci mluvčího závisí na mnoha faktorech řeči.

Protože jeden mluvčí nevysloví stejné slovo identicky dvakrát za sebou, i když si tyto slova budou podobná, tak podoba nebude 100%, jelikož se vytvoří rozdíly. Konkrétní podoba bude záviset na rychlosti promluvy, přízvuku, apod. V této práci se zaměřím na způsob, kterým se dají takto vyslovená slova porovnat.

V části 1.1 jsou stručně popsány biometrické metody. V části 1.2 jsou stručně popsány modely vzniku řečového signálu. V části 1.3 jsou popsány metody analýzy řečového signálu v časové oblasti, především vzorkování signálu a kvantizace kódování. Část 1.4 je zaměřena na zpracování signálu v časové oblasti, především na krátkodobou energii, funkci středního počtu průchodu nulou, lineární prediktivní analýzu a na kepstrální analýzu mluvené řeči. V části 1.6 jsou popsány fáze k určení řečníka. Dále tato část obsahuje stručný popis pracovních režimů a ohodnocení verifikace. Část 2 se věnuje zpracování řečového signálu a to pomocí preemfáze a segmentace signálu. Základní bloky (krátkodobá energie a intenzita, střední hodnota průchodu nulou a kepstrální koeficienty), které jsou implementovány v prostředí Matlab jsou popsány v části 2.3. Následující část 2.4 je zaměřena na první tři kmitočtové formanty českých hlásek. V části 2.5 se zaměřuji na porovnání hlásek od jednoho mluvčího i od více mluvčích. Část 2.6 je zaměřena na práci programu. 2.6.1

je zaměřena na činnost jednotlivých skriptů. V části 2.6.2 je popsána obsluha programu. Výsledky určení identity neznámého mluvčího jsou obsaženy v části 2.6.3.

1 ANALÝZA ŘEČOVÉHO SIGNÁLU

1.1 Biometrické metody

Do skupiny biometrických metod pro identifikaci osob se řadí i rozpoznávání lidí podle charakteristik jejich hlasu. Biometrické metody se vztahují i například na otisky prstů a geometrii ruky, genetickou strukturu, strukturu oční sítnice, rukopis a podobně. Porovnáme–li biometrické metody s metodami pro identifikaci osob, které využívají například magnetické karty, čipy, hesla nebo kódy naučené nazpaměť, dojdeme k závěru, že biometrické metody by měly být spolehlivější, možná dokonce i neomylné. Největší výhodou biometrických charakteristik je, že je nelze ukrást, zapomenout nebo ztratit.

Biometrické charakteristiky lze rozdělit do dvou skupin (tab. 1.1):

- **Fyziologické biometriky** – během lidského života se charakteristiky téměř nemění, pokud člověk nepodstoupí chirurgický zákrok nebo neutrpí nějaké závažné poranění.
- **Behaviorální biometriky** – zachycují chování člověka v dané situaci.

Rozpoznávání osob na základě behaviorálních biometrik je mnohem obtížnější, než rozpoznávání podle fyziologických biometrik. [8],[2]

Tab. 1.1: Rozdělení biometrických metod a jejich příklady.

Fyziologická biometrie	*téměř neměnná*
	otisky prstů
	geometrie ruky
	struktura oční sítnice
	struktura obličeje
	genetická struktura
Behaviorální biometrie	*změna dle situace*
	hlas
	rytmus srdce
	rukopis
	podpis

1.2 Vytváření a vlastnosti mluvené řeči

1.2.1 Úvodní poznámky

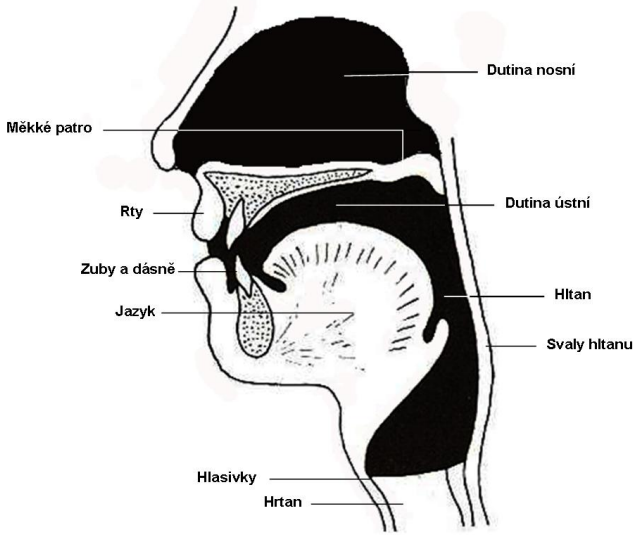
Schopnost vyjádřit myšlenky bývá označována jako jazyk – nejčastěji graficky (např. latinkou, azbukou) nebo akusticky (řeč). Každý jazyk se vyskytuje ve dvou základních podobách: mluvené (řeči) a psané (písmo). Tyto formy lze pokládat si za rovnocenné. Obě mají své výhody, ale i nevýhody. Komunikace je složena z hovoru a naslouchání, což jsme si osvojili již jako děti. Mluvená řeč se přenáší komunikačním kanálem jako akustická vlna. Akustický signál v sobě nese několik druhů informace. Kromě akustické složky (amplitudově-kmitočtového spektra) je z komunikačního hlediska nejdůležitější lingvistická informace, daná například svojí fonetickou, syntaktickou či pragmatickou strukturou, protože vyjadřuje význam myšlenky. V akustickém signálu nadále nalezneme specifické informace o mluvčím. Mezi hlavní můžeme zařadit intonaci, rytmus řeči, barvu hlasu, emocionální stav řečníka a anomálie – vady řeči.

Mluvená podoba řeči má pro automatické zpracování počítačem velký význam. Zde je nutné co nejpřesněji zaznamenat co a jak bylo řečeno. Pro syntézu řeči potřebujeme daný výraz podrobně popsat ve formě výslovnosti. V případě rozpoznávání mluvčího je zapotřebí získaný zvukový signál co nejlépe vyjádřit pomocí fonetických a akustických reprezentací a poté využít mluveného projevu k identifikaci mluvčího. [8]

1.2.2 Vytváření řeči

K vytváření řeči jsou v lidském těle skupiny orgánů, které se nazývají řečové (artikulační) orgány. Tyto orgány jsou primárně určeny k jiným funkcím (dýchání, cítění, příjem potravy). U dospělého muže je celková délka hlasového traktu od hrtanu až ke rtům přibližně 17 cm a plocha příčného průřezu se mění od nuly k 20 cm^2. Hlasový trakt je složen ze tří základních částí: dechové, hlasové a artikulační. Hlasový trakt člověka je vyobrazen na obr. 1.1.

Hlasové ústrojí se nachází v hrtanu, který je pomocí průdušnice spojen s plícemi. Nejdůležitější částí jsou hlasivky, které jsou v hrtanové dutině. Nacházejí se mezi chrupavkami hlasivkovými a chrupavkou štítnou. Mezi hlasivkami najdeme hlasivkovou štěrbinu, která při mlčení je odkryta, takže jí prochází vzduch. Při mluvení se proud vzduchu dostává k hlasivkám, které začínají kmitat, vzniká základ lidského hlasu. Frekvence kmitání je různá pro muže, ženu, dítě v různém věku. [8], [3]



Obr. 1.1: Hlasový trakt člověka

1.3 Analýza řečového signálu

Pulsní kódová modulace (PCM) je proces, při kterém se ze signálu získaného mikrofonem v analogovém tvaru stane digitální tvar. Občas se PCM označuje také jako digitalizace, tento proces se skládá ze dvou částí – vzorkování a kvantizace.[8]

1.3.1 Vzorkování signálu

Vzorkování signálu je transformace signálu $s(t)$ spojitého v čase, na posloupnost vzorků $s_a = s(nT)$ diskrétních v čase. Na frekvenci vzorkování $F_v = 1/T$, kde T je perioda vzorkování, jsou kladena omezení. Když je analogový signál $s(t)$ kmitočtově omezen na pásmo 0 až F_m [Hz], z hodnot $s(nT)$ lze opět vytvořit $s(t)$ podle vztahu

$$s(t) = \sum_{n=-\infty}^{\infty} s(nT) \left[\frac{\sin \pi(t/T - n)}{\pi(t/T - n)} \right], \tag{1.1}$$

Při vzorkování dojde k periodizaci původního spektra s periodou F_v.[8]

1.3.2 Kvantizace kódování

Nejčastěji je prováděna A/D převodníkem, který ze vstupního analogového napětí vytvoří odpovídající kódovou reprezentaci. K návrhu kvantizéru postačí kvantizační krok q a počet úrovní kvantování, který je ve tvaru 2^B(B je počet bitů). Během kvantizačního procesu dojde k jisté ztrátě dat vlivem „zaokrouhlováním“ měřených velikostí signálu na nižší celočíselnou hodnotu. Tento děj se označuje jako kvantizační šum nebo zkreslení. Běžný řečový signál je v rozsahu asi 60 dB, což pro nás

znamená, že pro kvalitní záznam je zapotřebí použít 11 až 12 bitový převod. Telefonní přenosové pásmo má $F_v = 8\,\text{kHz}$ a přenosovou rychlost 88 kbit/s. Pro vyšší kvalitu se používá 16bitový převod s $F_v = 16 - 22\,\text{kHz}$ a rychlostí přenosu 256 až 352 kbit/s. Předpokladem je, že získaná vstupní data jsou vložena pomocí PCM a je konstantní perioda vzorkování T.[8]

1.3.3 Kódování tvaru vlny

Snaha o snížení přenosové rychlosti. Hodnotu F_v nelze zmenšovat do nekonečna, tak je úsilí soustředěno na metody snižující počet bitů na vzorek.

μ-law a ***A*-law**

Lineární kódování nemusí být vždy optimální. Velikost kroku q kvantizéru je odvozena ze znělých segmentů řeči a to z jejich amplitud, ale pro neznělé segmenty by se muselo použít menší velikosti kroku. Řešení je v nelineárních charakteristikách kvantizéru. Užití μ-law, A-law kvantizéru vede k užití takové nelineární operace, která nově vzniklou posloupnost řečových vzorků vytvoří tak, aby vyhovovala funkci hustoty pravděpodobnosti blížící se rozdělení, které je rovnoměrné. Nová funkce je založena na aplikaci funkce logaritmus. Transformace funkce μ-law je vyjádřena vztahem

$$\hat{s}_{\mu}(k) = \frac{s_{max}}{\log(1+\mu)}\text{sign}(s[k])\log\left(1 + \frac{\mu|s[k]|}{s_{max}}\right), \tag{1.2}$$

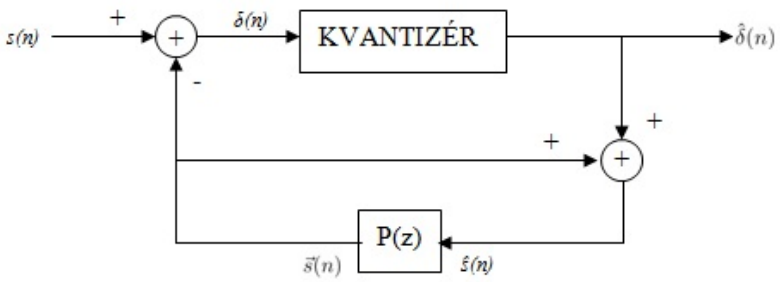
kde pro velké hodnoty $s[k]$ je téměř logaritmický a pro malé hodnoty $s[k]$ je téměř lineární. Transformace funkce A-law je podobná

$$\hat{s}_A[k] = \frac{s_{max}}{1+\log A}\text{sign}(s[k]). \tag{1.3}$$

Zde se používá 8bitové kódování se vzorkovacím kmitočtem 8 kHz. V České republice se používá A-law, kde $A = 87,56$.

Diferenční pulsní kódová modulace (DPCM).

Po PCM jsou změny signálu pomalé a rozdíly mezi vzorky mají menší disperzi než vstupní signál. To umožní zavést obecnou metodu diferenčního kvantování vyobrazenou na obr. 1.2. V případě, že $\tilde{s}[n]$ (odhad řečového vzorku $s[n]$ je dobrý, pak by odchylka rozdílu $\delta[n] = s([n] - \tilde{s}[n]$ měla být menší než odchylka $s[n]$. Řečové vzorky $s[n]$ potřebují větší počet bitů než $\delta[n]$. Signál $\hat{s}[n]$ se liší od $s[n]$ jen kvantizační chybou $e[n]$.[8]



Obr. 1.2: Blokové schéma diferenčního kvantizéru

1.4 Zpracování v časové oblasti

V metodách krátkodobé analýzy v časové oblasti se vyskytuje váhová posloupnost neboli tzv. okénko $w[n]$. Tímto okénkem se „váží“ vzorky $s[k]$. Okénkem je určena priorita při zpracovávání signálu. Nejčastěji se využívá pravoúhlého a Hammingova okénka.[8]

1.4.1 Krátkodobá energie

Krátkodobou energii signálu definujeme vztahem

$$E_n = \sum_{k=-\infty}^{\infty} [s[k]w[n-k]]^2, \tag{1.4}$$

kde $s[k]$ je vzorek signálu v čase k a $w[n]$ zastupuje zvolené okénko. Délku mikrosegmentu je vhodné použít 10–20 ms. Každý mikrosegment má informaci o průměrné hodnotě krátkodobé energie. Tato metoda projevuje značnou citlivost na velké úrovňové změny signálu, což je nedostatek. Dynamika řečového signálu je ještě více zvýšena vlivem kvadrátu v rovnici (1.4). Zmíněný nedostatek se neobjevuje u krátkodobé intenzity, proto se také velmi často používá

$$M_n = \sum_{k=-\infty}^{\infty} |s[k]|w[n-k]. \tag{1.5}$$

Krátkodobé intenzity se využívá hlavně k oddělování tichých segmentů od řečových segmentů, ale také lze intenzitu využít k rozpoznání znělých a neznělých částí promluvy.[8]

1.4.2 Funkce středního počtu průchodu nulou

Spektrální vlastnosti signálu lze přirovnat k frekvenci průchodů signálu nulou. Pokud daný průběh je sinusový s kmitočtem f, tak je průměrný počet průchodů nulou roven $2f$ [průchodů/s].

$$Z_n = \sum_{k=-\infty}^{\infty} |\mathrm{sgn}[s[k]] - \mathrm{sgn}[s[k-1]]| w[n-k], \tag{1.6}$$

kde $w[n]$ je pravoúhlé okénko.[8]

1.5 Zpracování v kmitočtové oblasti

1.5.1 Lineární prediktivní analýza

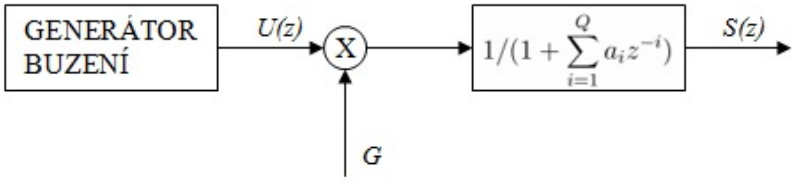
Lineární prediktivní kódování (LPC) patří mezi nejefektivnější metody pro analýzu akustických signálů. Půvab této metody najdeme v možnosti zabezpečit odhad parametrů. Model řeči je složen z generátoru budících funkcí a ze systému s časově proměnným přenosem. Generátor budí systém náhodným šumem při vytváření neznělých zvuků a posloupností impulsů při tvorbě znělých zvuků.

$$s[n] = \sum_{i=1}^{Q} a_i s[k-i] + Gu[k], \tag{1.7}$$

kde G je koeficientem zesílení a Q je řádem modelu. Přenosová funkce modelu se pak zapíše jako

$$H[z] = \frac{S[z]}{U[z]} = \frac{G}{A[z]} = \frac{G}{1 + \sum_{k=1}^{Q} a_i z^{-i}}. \tag{1.8}$$

Proces modelování je zobrazen na obr. 1.3. Koeficienty a_i číslicového filtru a zesílení G. Metoda nejmenších čtverců se využívá pro stanovení a_i a G za předpokladu přibližné stacionarity signálu na daném intervalu. Pokud není znám člen $Gu[k]$ v rovnici



Obr. 1.3: Blokový model vytváření řeči s lineárním číslicovým filtrem.

(1.7) se tím vytvoří chyba predikce $e[k]$ mezi $s[k]$ a $s)\hat{[}k]$. Výpočet spektra signálu je možný pomocí koeficientů a_i. Toto spektrum má podobu vyhlazené spektrální obálky původního diskretizovaného signálu $s[k]$. Zavedeme–li substituci $z = \mathrm{e}^{\mathrm{j}\omega}$

$$H(j\omega) = \frac{G}{1 + \sum_{i=1}^{Q} a_i e^{-\mathrm{j}\omega i}}. \tag{1.9}$$

Výkonové spektrum $P(\omega)$ je dáno vztahem, který má po úpravě tvar

$$P(\omega) = \frac{G^2}{RA(0) + 2\sum_{i=1}^{Q} RA(i)\cos(i\omega)}, \tag{1.10}$$

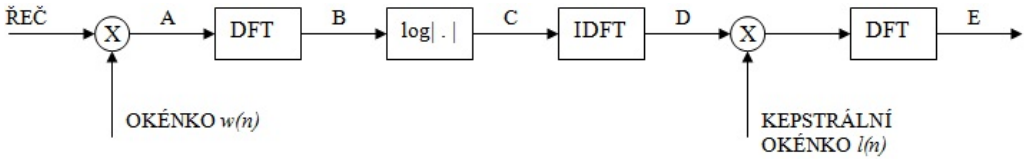
kde

$$RA(i) = \sum_{k=0}^{Q-i} a_k a_{k+1}, \tag{1.11}$$

kde $a_0 = 1$, $0 \leq i \leq Q$. Jako úhlový kmitočet do vztahů (1.9) a (1.10) dosadíme $\omega = 2\pi f/F_v$, kde f je proměnný kmitočet (platí vzorkovací teorém, že $f \leq 0,5F_v$).[8]

1.5.2 Kepstrální analýza mluvené řeči

Jelikož řečové kmity mohou být modelovány na krátkodobém základě ve vztahu na lineární buzení pro znělou řeč a šumu pro neznělou řeč. Pokud signál A byl vytvořen diskrétní konvolucí řečového signálu $s[n]$ a funkce okénka $w[n]$. Po vstupu do bloku DFT je výstupem signál B, který je Furierovou transformací buzení a impulsní odezvy hlasového ústrojí a je přiveden na blok log|.|. Výstupem z bloku log|.| je C, které je sumou transformace odezvy hlasového ústrojí a logaritmů transformace buzení. Změnou C se pozvolna mění i část kepstra D, ale složka logaritmu se mění rychle, což se projevuje velkými špičkami. Spektrální obálka, výstup E lze získat odstraněním složky C lineární filtrací a následně se provede DFT. Úpravy v C se dají provést vynásobením kepstra tzv. kepstrálním okénkem $l[n]$.[10],[8]



Obr. 1.4: Blokové schéma kepstrální analýzy.

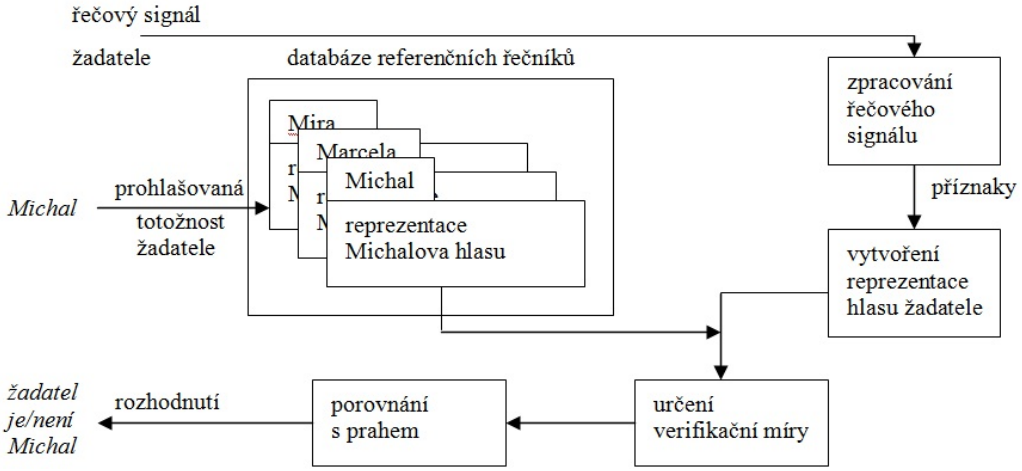
1.6 Určení řečníka

Základní dělení kroků pro určení mluvčího jsou identifikace mluvčího a verifikace mluvčího.

Verifikace řečníka

V této fázi máme k dispozici zvukový záznam neznámého člověka, který se za někoho vydává. Nahrávka se nyní musí porovnat se záznamem osoby, za kterou se náš neznámý vydává a určit míru podobnosti. Do objektu zabezpečeného uzavíráním na hlasový zámek mají právo vstoupit jen tři osoby (v našem případě Michal, Mira

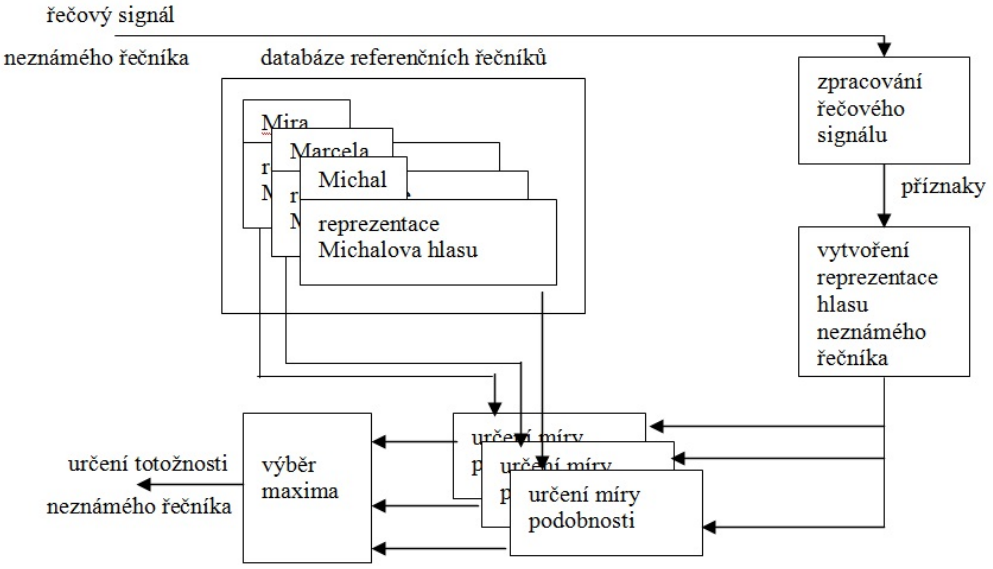
a Marcela). Tyto tři budeme označovat jako referenční řečníky. Ke vstupu mají tito jedinci vlastní heslo, které je uloženo v databázi. Pokud někdo chce vstoupit do objektu musí nejprve zvolit jako kdo chce být do objektu vpuštěn–např. stisknutím příslušného tlačítka. Tato osoba se označuje jako neznámý řečník nebo žadatel a totožnost, která byla stisknuta na panelu se považuje za prohlašovanou totožnost. Následně se vloží heslo žadatele a to je porovnáno s heslem uloženým v databázi. V případě rozdílu je žadatel přeznačen na podvodníka. Pokud se shodují hesla, tak je žadatel přeznačen jako správný referenční řečník. Žadatelem může být i někdo mimo skupinu referenčních řečníků (např. Libor nebo Lucka).



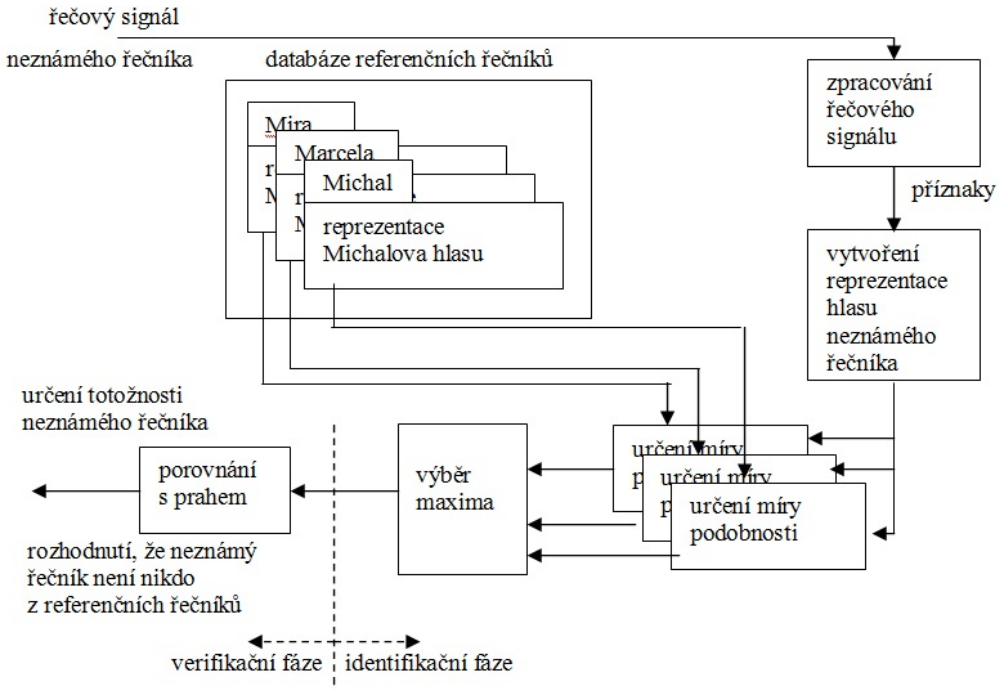
Obr. 1.5: Blokové schéma verifikace mluvčího.

Identifikace řečníka

Platí tu stejné označení jako u verifikace řečníka. V této části se porovnává kdo ze skupiny známých řečníků má nejvíce podobný hlas neznámé osoby. Nejprve se předpokládá, že hlas neznámého patří do skupiny referenčních řečníků, což znamená že se jedná o identifikaci v uzavřené množině. Porovnáním se pokusíme najít nejvíce podobnou osobu. Náznak postupu je zobrazen na obr. 1.6. Pokud je předpokladem, že hlas neznámého nepatří do skupiny referenčních řečníků, pak se jedná o identifikaci v otevřené množině na obr. 1.7. Nejprve se neznámý porovnává v uzavřené skupině. Výsledek z této metody se následně považuje za prohlášenou totožnost. Verifikací se porovná podobnost hlasu neznámého řečníka s hlasem prohlašované osoby. V případě dostatečné podobnosti je neznámý označený jako řečník, pro kterého je totožnost daná výsledkem identifikačního procesu. Pokud není dostatečná podoba, tak je neznámy označen jako mluvčí, který nepatří do referenční skupiny řečníků. Jinými slovy identifikace v otevřené množině je kombinací verifikace a identifikace v uzavřené množině.[8]



Obr. 1.6: Blokové schéma procesu identifikace mluvčího v uzavřené množině.

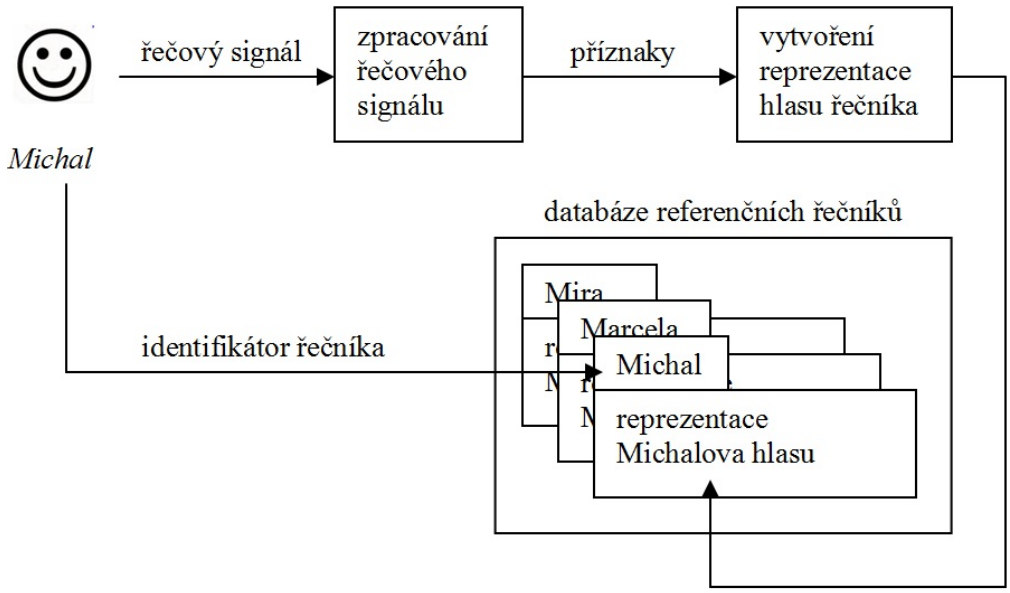


Obr. 1.7: Blokové schéma procesu identifikace mluvčího v otevřené množině při využití míry podobnosti mezi reprezentacemi hlasů řečníků.

1.6.1 Pracovní režimy

Systém určený k rozpoznávání mluvčího pracují v režimech, kde hlavním režimem je rozpoznávání. Dochází zde k verifikaci a identifikaci. Databáze referenčních řečníku je zde velice důležitou částí. Je tvořena dvojicí identifikátor řečníka a reprezen-

tace hlasu řečníka. Tyto dva záznamy jsou vytvořeny v režimu trénování. Nejprve si každý mluvčí zadá identitu (např. *Michal*) a následně zadá vzorek řeči. Signál se zpracuje stejným způsobem, kterým se ve fázi rozpoznávání bude zpracovávat neznámý signál. Získané příznaky se uloží do databáze, jak je vidět na obr. 1.8. Druhá část systému na rozpoznávání mluvčího je po trénování část testování, kde



Obr. 1.8: Blokové schéma procesu trénování systémů identifikace a verifikace mluvčího.

se systém chová stejně jako v režimu rozpoznávání s rozdílem v tom, že známe skutečnou totožnost neznámého. Potřeba dostat systém do stavů, do kterých se může dostat během reálného provozu, tzn. použít i hlasy osob, které nebyly nijak použity v trénovací části. Poslední částí rozpoznávacího systému je vyhodnocení. Je třeba vyhodnocování provádět na odlišných datech od dat použitých pro trénování a testování s převáděním systému do všech možných stavů, které je možné v reálném užívání systému zažít.[8]

1.6.2 Ohodnocení verifikace

Mohou nastat čtyři stavy systému (tab. 1.2). Při verifikaci řečníka se používá míra související s počtem nesprávného odmítnutí a počtem nesprávného přijetí. Průběh verifikace pak jde ohodnotit dvojicí chyb:

- poměrný počet chyb nesprávného odmítnutí $R_{FR}(\theta)$,
- poměrný počet chyb nesprávného přijetí $R_{FA}(\theta)$.

Tab. 1.2: Možné výsledky činnosti verifikace za předpokladu, že v databázi referenčních řečníků jsou uloženy reprezentace hlasu Michala a Marcely.

Skutečná totožnost	**Prohlašovaná totožnost**	**Rozhodnutí verifikačního systému**	**Výsledná situace**
Michal	Michal	žadatel je Michal	správné přijetí
Michal	Michal	žadatel není Michal	nesprávné odmítnutí
Lucka	Marcela	žadatel je Marcela	nesprávné přijetí
Lucka	Marcela	žadatel není Marcela	správné odmítnutí

Tyto chyby se vztahují na daný verifikační práh θ. $R_{FA}(\theta)$ je odhad pravděpodobnosti, že systém příjme podvodníka. Výpočet je pak dán vztahem

$$R_{FA}(\theta) = \frac{n_{FA}\theta}{n_{podv}}, \tag{1.12}$$

kde n_{podv} je zastoupení počtu verifikačních pokusů, kde žadatelem byl podvodník, $n_{FA}(\theta)$ je počtem pokusů, kdy systém při verifikačním prahu θ přijal podvodníka jako osobu, která má vstup povolen. $R_{FR}(\theta)$ je odhad pravděpodobnosti, že systém odmítne referenčního řečníka. Výpočet je pak dán vztahem

$$R_{FR}(\theta) = \frac{n_{FR}\theta}{n_{sp-ref}}, \tag{1.13}$$

kde n_{sp-ref} je vyjádření počtu pokusů, kdy žadatelem byl referenční řečník, $n_{FR}(\theta)$ udává počet pokusů, kdy systém při verifikačním prahu θ referenčního řečníka odmítl. Pro možnost vyjádřit činnost systému na verifikaci mluvčího jedním číslem se používá míra rovnosti chyb R_{EER}, kde platí

$$R_{EER} = R_{FR}(\theta_{EER}) = R_{FA}(\theta_{EER}). \tag{1.14}$$

Jinými slovy musíme najít takovou hodnotu prahu θ_{EER}, kde se poměrný počet chyb nesprávného přijetí rovná poměrnému počtu chyb nesprávného odmítnutí, abychom mohli určit míru R_{EER}. Jelikož θ_{EER} je poměrně těžké určit, tak se místo R_{EER} k ohodnocení systému na verifikaci mluvčího používá přibližná hodnota R'_{EER}, která je dána vztahem

$$R'_{EER} = \frac{R_{FR}(\theta'_{EER}) + R_{FA}(\theta'_{EER})}{2}, \tag{1.15}$$

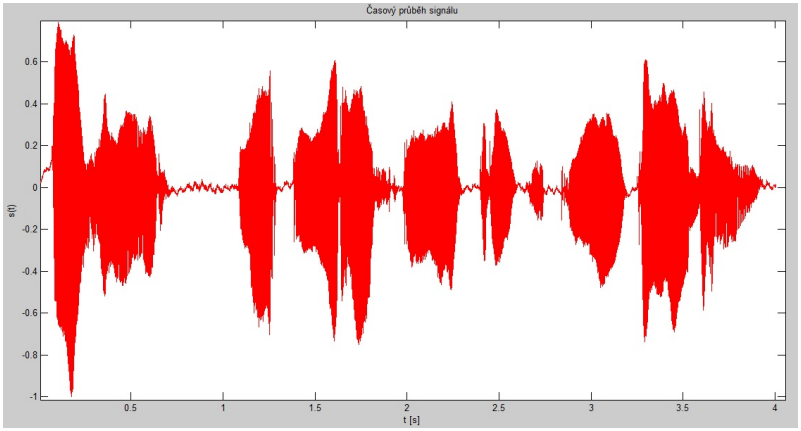
kde [8]

$$\theta'_{EER} = argmin_{\theta}|R_{FR}(\theta) - R_{FA}(\theta)|. \tag{1.16}$$

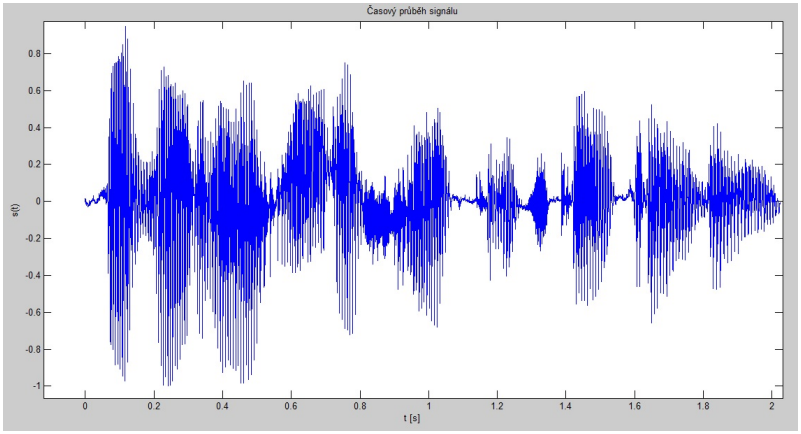
2 ZPRACOVÁNÍ ŘEČOVÉHO SIGNÁLU

Základní bloky vybrané metody pro rozpoznávání mluvčího byly implementovány v prostředí Matlab. Těmito bloky byly otestovány zkušební nahrávky `pozor_001.wav` až `pozor_020.wav`. Text nahrávek je „Pozor na úraz elektrickým proudem“.

Pro demonstraci nám postačí dvě nahrávky a to například `pozor_017.wav`, zachycující mladou dívku a `pozor_004.wav`, zachycující muže. Jejich časový průběh je zobrazen na obr. 2.1 a obr. 2.2.



Obr. 2.1: Časový průběh signálu – `pozor_017.wav`.



Obr. 2.2: Časový průběh signálu – `pozor_004.wav`.

2.1 Preemfáze

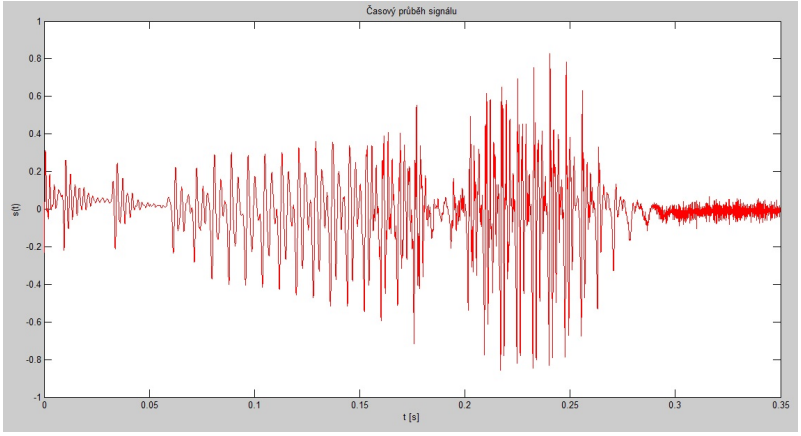
Pod tímto názvem se skrývá označení metody, která se užívá ke zvýraznění vyšších kmitočtů v signálu jelikož mají nižší úroveň. V kmitočtovém pásmu v nižších

hodnotách se nachází podstatná část celkové energie řečového signálu. Nad těmito hodnotami kmitočtu se nacházejí užitečné informace signálu. Preemfáze bývá nejčastěji realizována jednoduchým filtrem číslicového charakteru:

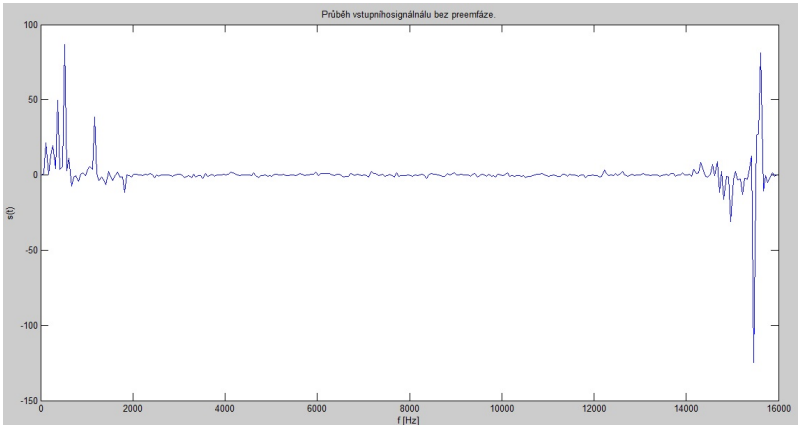
$$y(n) = x(n) - a * x(n-1), \tag{2.1}$$

kde $y(n)$ je n–tým vzorkem signálu po preemfázi a $x(n)$ je n–tým vzorkem originálního signálu. Toto zesílení můžeme vidět na obr. 2.5 a porovnat jej s původním signálem, který nebyl upraven filtrem preemfáze na obr. 2.4.

Konstanta a se volí přibližně 0,95–0,98. [9]



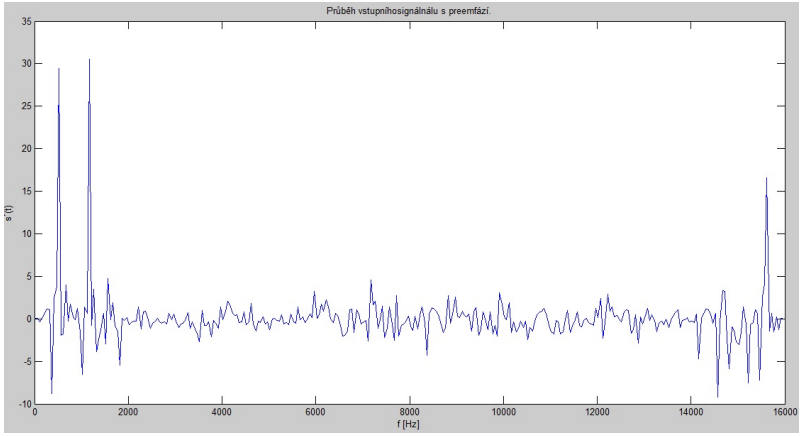
Obr. 2.3: Časový průběh slova „pozor“.



Obr. 2.4: Průběh vstupního signálu bez preemfáze slova „pozor“.

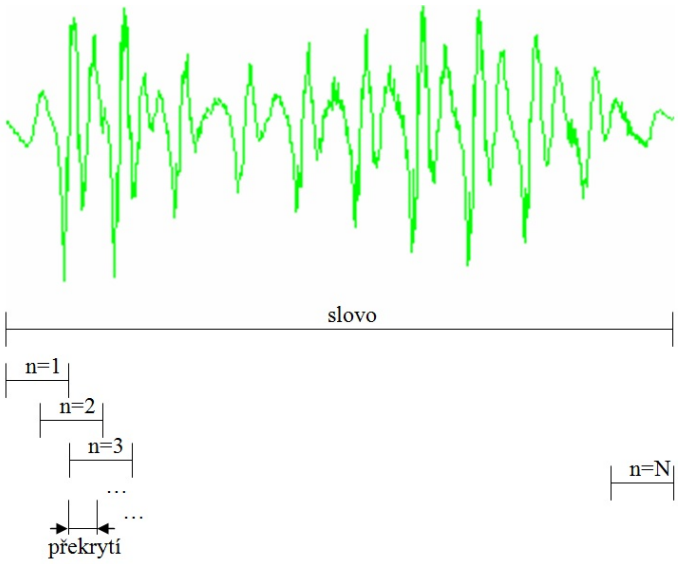
2.2 Segmentace signálu

Nejčastěji se užívá charakteristik z frekvenční oblasti. Signál při stejné kvalitě lze popsat menším počtem složek. Při frekvenční oblasti lze předpokládat, že v průběhu

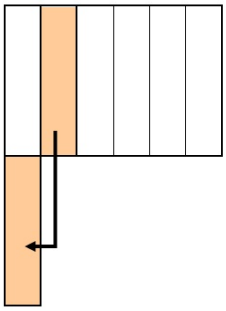


Obr. 2.5: Průběh vstupního signálu s preemfází slova „pozor“.

několika milisekund jsou parametry řečového signálu téměř konstantní. Tyto části se volí v intervalu od 10 ms do 35 ms. V této práci používám segmentaci dvakrát. Pro výpočet spektra hustoty, krátkodobé energie a intenzity, střední hodnoty průchodu nulou, LPC koeficientů a formantů používám segmentaci 20 ms – vytvořeno skriptem `prekryti.m`, pro výpočet kepstrálních koeficientů a základní periody řeči (T_0) využívám segmentace 50 ms – vytvořeno skriptem `prekryti2.m`. Řečový signál je tedy rozdělen do úseků o velikosti N vzorků, jak je vidět na obr. 2.6. Každý následující úsek je překryt o M vzorků, kde $M < N$, v této práci jsou zvolena překrytí 50 %. Začátek 2. segmentu je přikopírován ke konci 1. segmentu, což je naznačeno na obr. 2.7.



Obr. 2.6: Segmentace signálu.

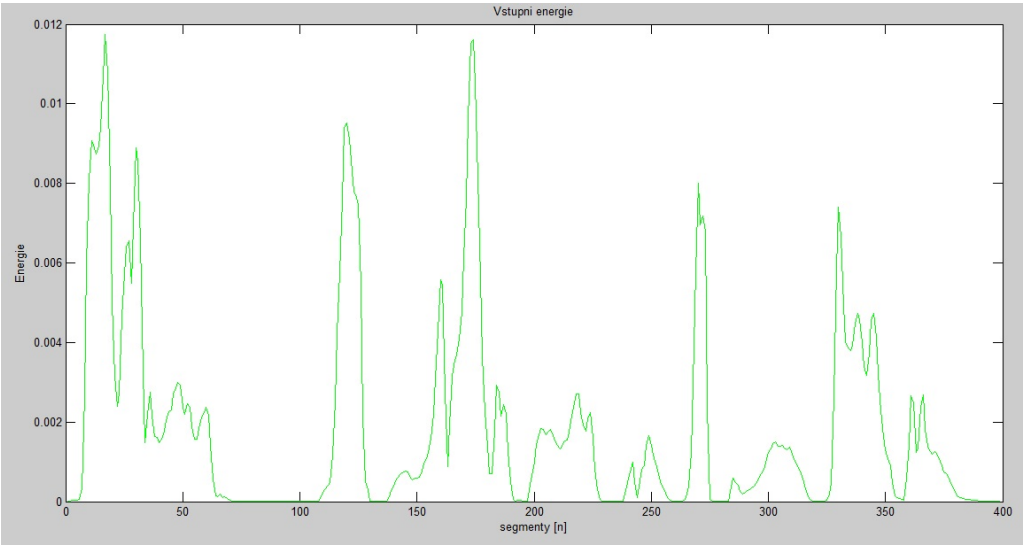


Obr. 2.7: Nákres rozdělených úseků a 50 % překrytí.

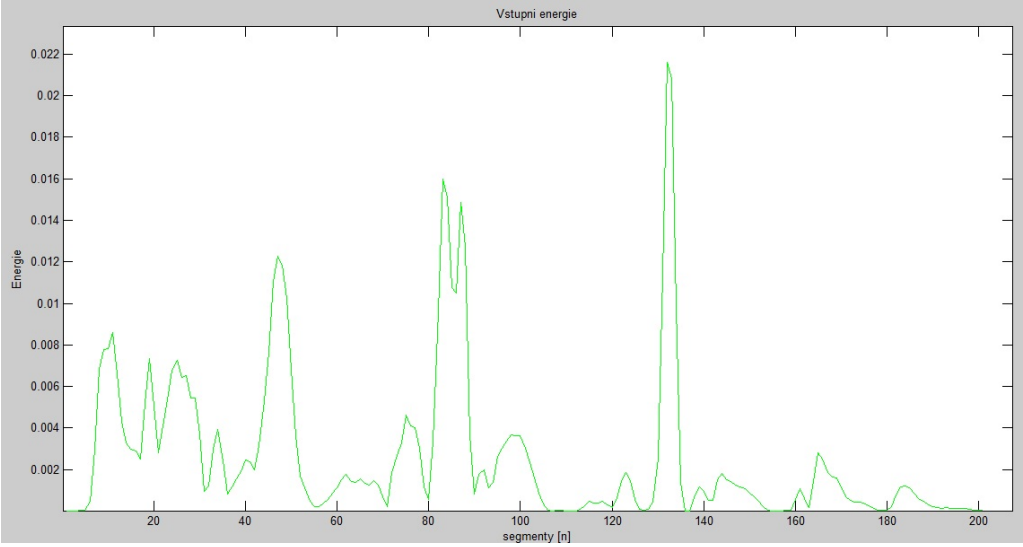
2.3 Základní bloky implementovány v prostředí Matlab

2.3.1 Krátkodobá energie

Vstupní signál byl rozdělen na segmenty po 20 ms a v každém segmentu je vypočtena krátkodobá energie podle rovnice (1.4). Místa, kde se v záznamu objevil šum z okolí jsou ve výsledném grafickém zobrazení nízké energetické hodnoty, jak lze vidět na obr. 2.8 a obr. 2.9. Toto grafické vyjádření je výsledkem skriptu napsaném v Matlabu, pod názvem `energie.m`.



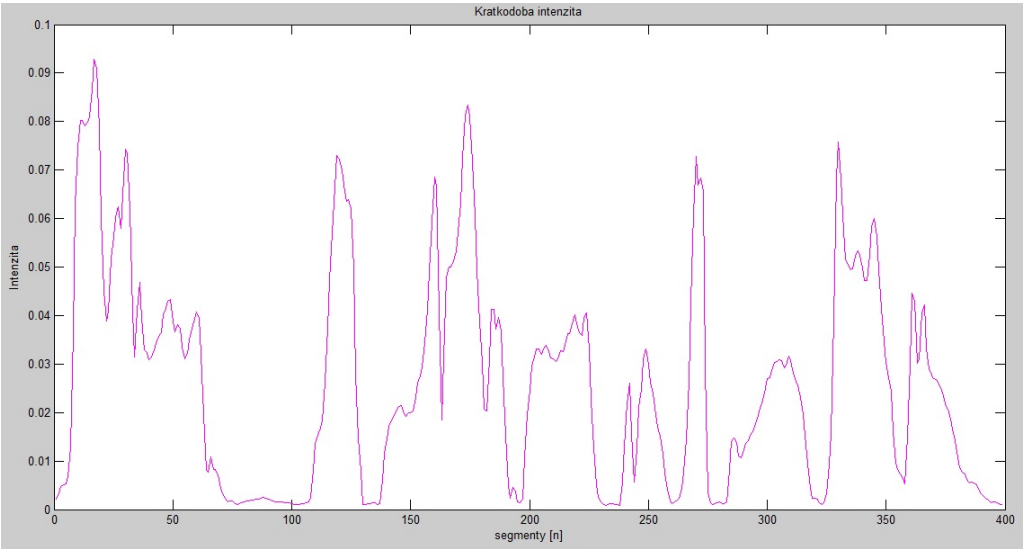
Obr. 2.8: Vstupní energie záznamu, na kterém je zachycena mladá dívka – `pozor_017.wav`.



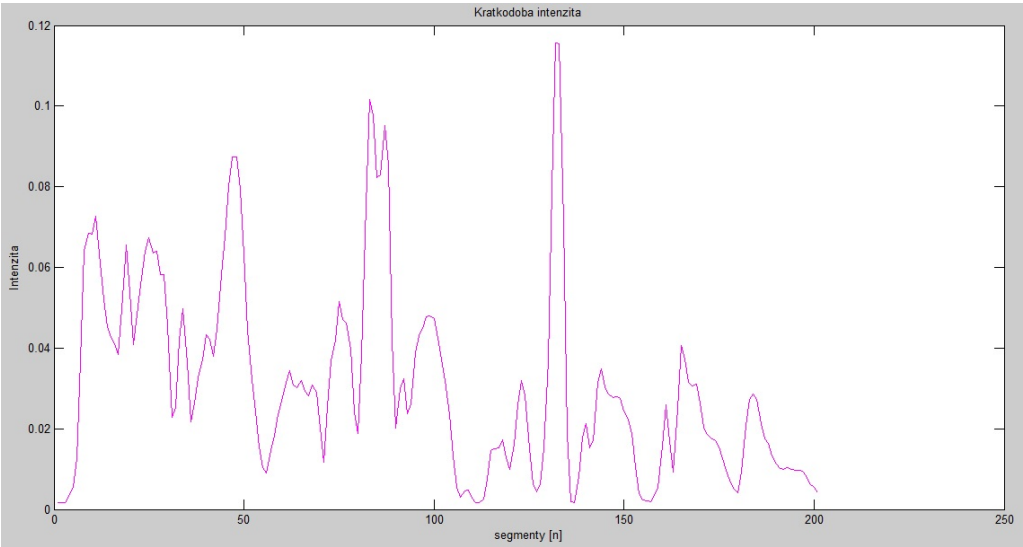
Obr. 2.9: Vstupní energie záznamu, na kterém je zachycen muž – `pozor_004.wav`.

2.3.2 Krátkodobá intenzita

Hlavní využití krátkodobé intenzity v praxi se nachází v rozpoznání znělých a neznělých částí daného slova, ale také v možnosti oddělení tichého segmentu od řečového segmentu. Krátkodobá intenzita je vypočtena v každém segmentu podle rovnice (1.5). Tento výpočet v Matlabu provádí napsaný skript `intenzita.m`. Grafické zobrazení intenzity můžeme vidět na obr. 2.10 a obr. 2.11.



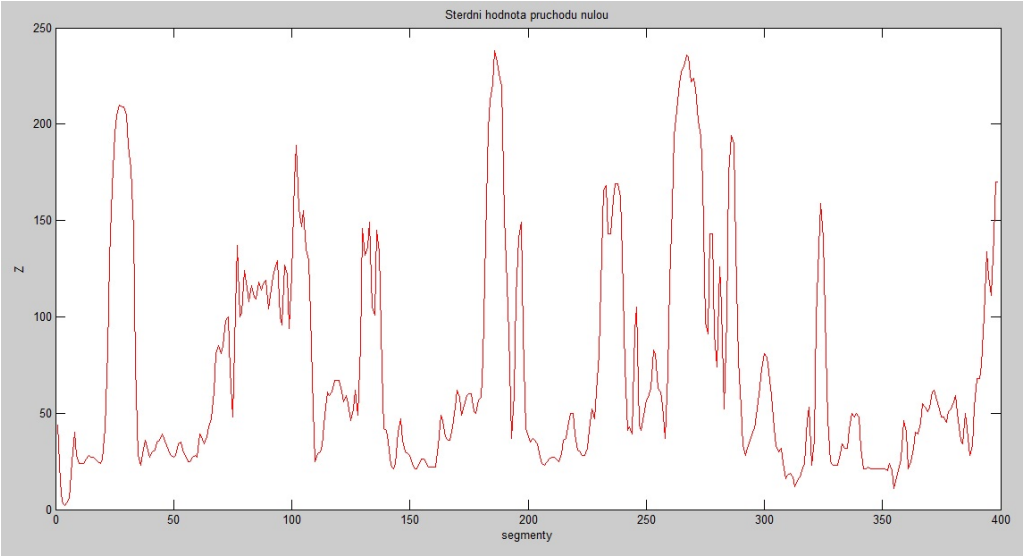
Obr. 2.10: Krátkodobá intenzita záznamu `pozor_017.wav` – mladá dívka.



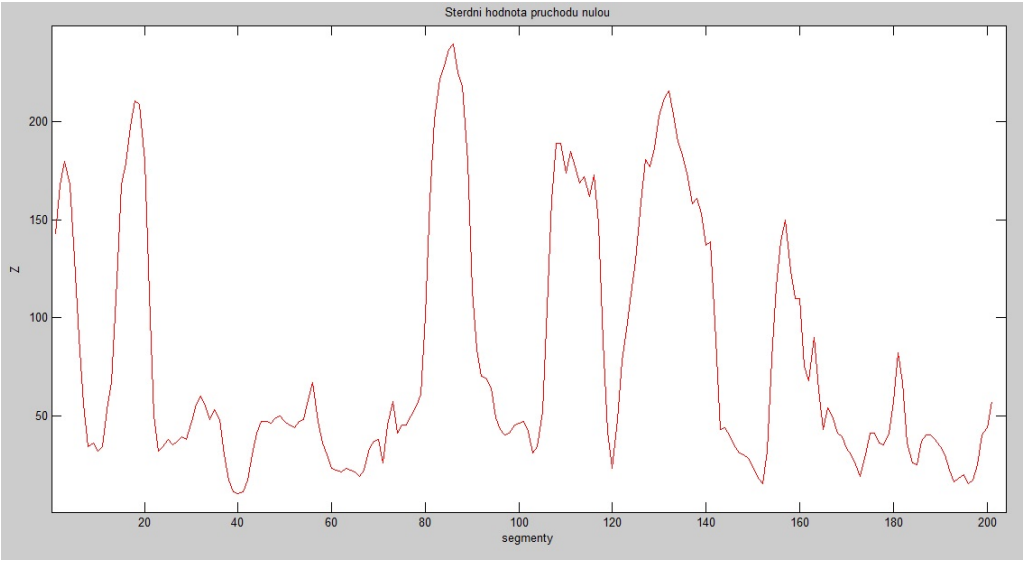
Obr. 2.11: Krátkodobá intenzita záznamu `pozor_004.wav` – muž.

2.3.3 Průchod nulou

Funkce středního počtu průchodu nulou je vypočtena v každém segmentu podle rovnice (1.6). Tento výpočet v Matlabu provádí napsaný skript `shpn.m`. Grafické zobrazení intenzity můžeme vidět na obr. 2.12 a obr. 2.13.



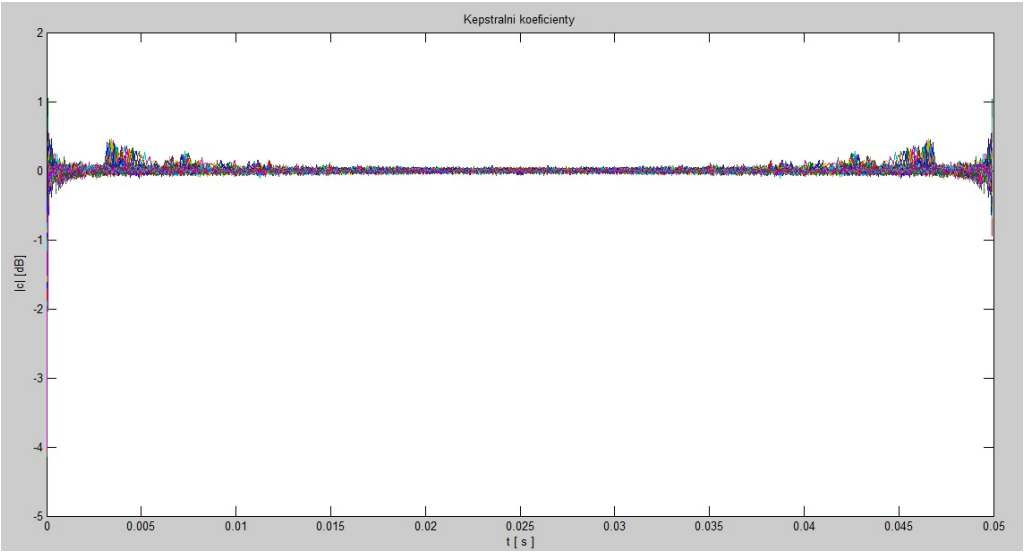
Obr. 2.12: Funkce středního počtu průchodu nulou záznamu, na kterém je zachycena mladá dívka – `pozor_017.wav`.



Obr. 2.13: Funkce středního počtu průchodu nulou záznamu, na kterém je zachycen muž – `pozor_004.wav`.

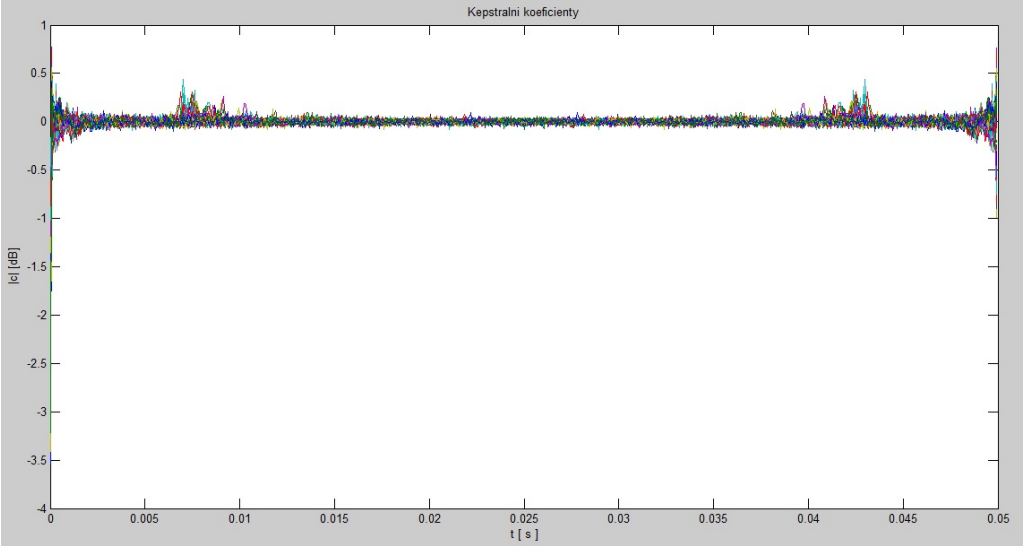
2.3.4 Kepstrální koeficienty

Kepstrální koeficienty jsou vypočteny v každém segmentu (viz. 1.5.2). Tento výpočet v Matlabu provádí napsaný skript `kepstral.m`. Grafické zobrazení koeficientů můžeme vidět na obr. 2.14 a obr. 2.15. Zároveň si na těchto vyobrazeních lze všim-



Obr. 2.14: Kepstrální koeficienty pro záznamu, na kterém je zachycena mladá dívka – `pozor_017.wav`.

nout toho, že daný průběh je symetrický podle středu $(n/2)$, kde n je počet vzorků



Obr. 2.15: Kepstrální koeficienty pro záznamu, na kterém je zachycen muž – `pozor_004.wav`.

v segmentu. V segmentech obsahujících znělé hlásky se projeví špička, která v čase odpovídá základní hlasivkové periodě (T_0).

Na grafickém zobrazení kepstrálních koeficientů na obr. 2.14 a obr. 2.15 je vidět více špiček. Vynecháme prvních 10 – 15 vzorků, které jsou určeny pro kmitočty formantů promluvy. Po vypuštění těchto vzorků prohledáme spektrum a nejvyšší špička bude hodnota základního tónu. Tímto postupem dosáhneme hodnoty T_0, který je realizován ve skriptu `T-nula.m`.

2.4 Formanty

Jsou to lokální maxima (špičky) ve spektru znělých hlásek. Díky rezonanci v dutinách hlasového ústrojí (nosní, ústní, hltanová), ale také v dutinách hudebních nástrojů atd., vznikají formanty.

Lidský hlas je vytvářen pomocí tónu se základním kmitočtem F_0, který vzniká chvěním hlasivek. Tento tón se dále pohybuje hlasovým ústrojím a vzniká rezonance v dutinách. První tři základní kmitočty formantů jsou důležité pro rozpoznávání mluvčích, ale i jednotlivých samohlásek tzv. vokálů. [1]

Výše jsou uvedeny v tabulce 2.1 typické kmitočtové rozsahy prvních tří formantů F_1, F_2 a F_3. [8]

Tab. 2.1: Hodnoty pásem prvních tří formantů pro české souhlásky.

Formanty/ Samohlásky	F_1 [Hz] pásmo	F_2 [Hz] pásmo	F_3 [Hz] pásmo
/i/,/í/	300–500	2000–2800	2600–3500
/e/,/é/	480–700	1560–2100	2500–3000
/a/,/á/	700–1100	1100–1500	2500–3000
/o/,/ó/	500–700	850–1200	2500–3000
/u/,/ú/	300–500	600–1000	2400–2900

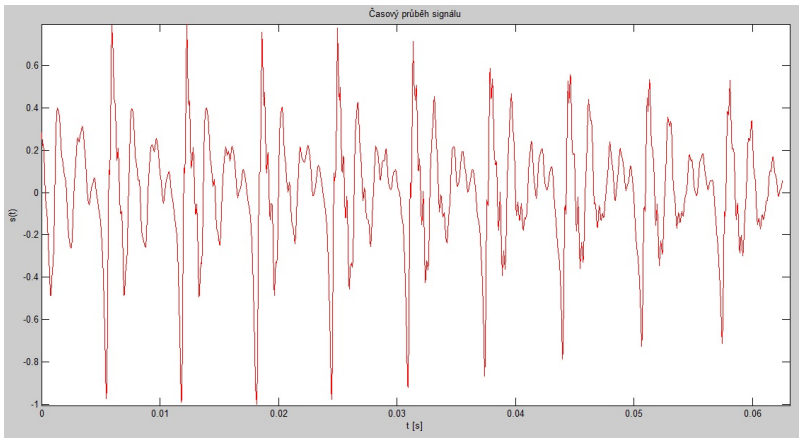
2.5 Porovnání hlásek

2.5.1 Porovnání pro jednoho mluvčího

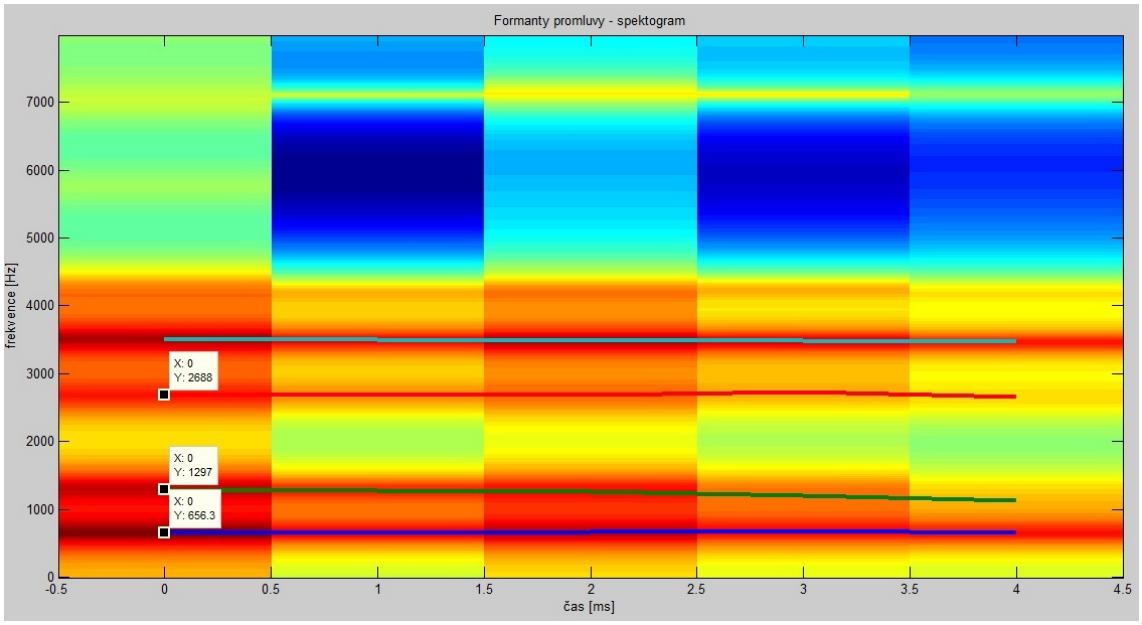
Z nahrávky `pozor_001`, `pozor_002`, `pozor_003`, `pozor_004`, `pozor_012`, `pozor_017`, `pozor_018` a `pozor_019` jsem vytvořil samostatné nahrávky pro jednotlivé hlásky „a“, „e“, „i“, „o“, „u“ a uložil je pro pozdější zpracování. Časový průběh hlásky „a“ je možno vidět na obr. 2.16 a spektogram této hlásky od téhož mluvčího je možné vidět na obr. 2.17. Časový průběh hlásky „o“ je možno vidět na obr. 2.18 a spektogram této hlásky od téhož mluvčího je možné vidět na obr. 2.19. Časový průběh hlásky „y“ je možno vidět na obr. 2.20 a spektogram této hlásky od téhož mluvčího je možné vidět na obr. 2.21. V případě hlásky „a“ je hodnota prvního formantu přibližně o 50 Hz nižší než rozsah udaný v tab. 2.1, hodnoty ostatních formantů tabulce odpovídají. Pro hlásky „o“ (tab. 2.3) a „y“ (tab. 2.4) se hodnoty shodují v celé stupnici.

Tab. 2.2: Získané hodnoty pásem prvních tří formantů pro hlásku „a“ od mluvčího 001.

Formanty/ Zdroj	F_1 [Hz] pásmo	F_2 [Hz] pásmo	F_3 [Hz] pásmo
spektogram	656,3	1297	2688
konzole	656,3	1296,9	2687,5
shoda s tbl. 2.1	**x**	√	√



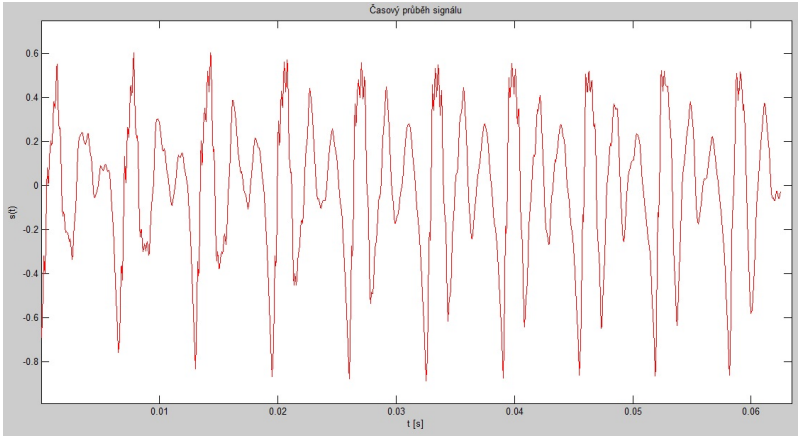
Obr. 2.16: Časový průběh hlásky „a“ od mluvčího 001.



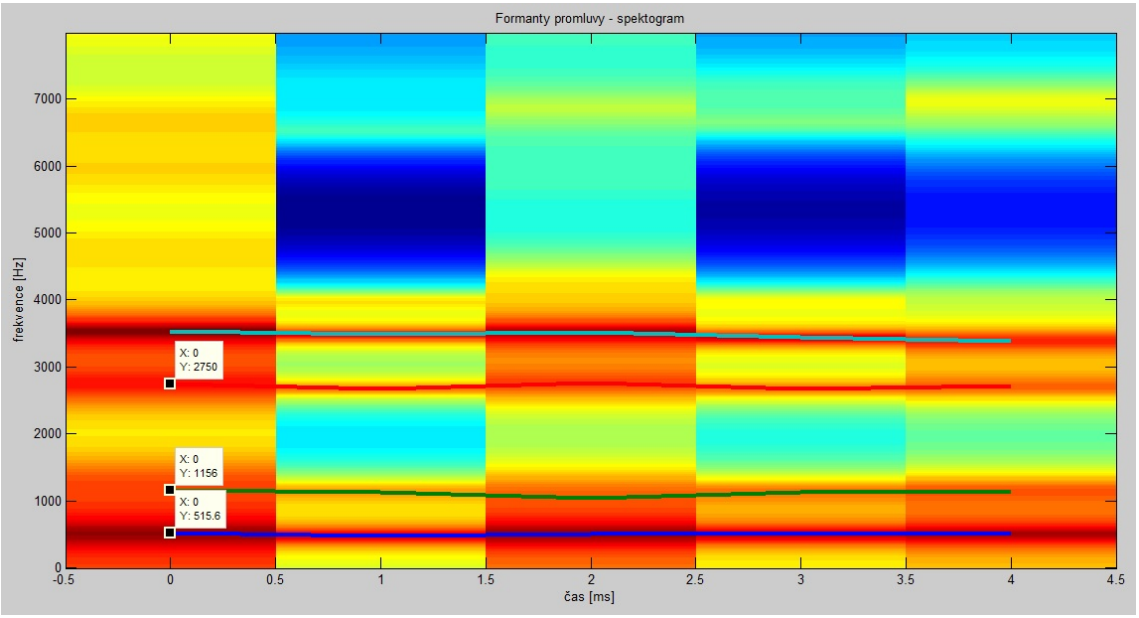
Obr. 2.17: Spektogram hlásky „a“ od mluvčího 001.

Tab. 2.3: Získané hodnoty pásem prvních tří formantů pro hlásku „o“ od mluvčího 001.

Formanty/ Zdroj	F_1 [Hz] pásmo	F_2 [Hz] pásmo	F_3 [Hz] pásmo
spektogram	515,6	1156	2750
konzole	515,6	1156,3	2750,0
shoda s tbl. 2.1	√	√	√



Obr. 2.18: Časový průběh hlásky „o“ od mluvčího 001.



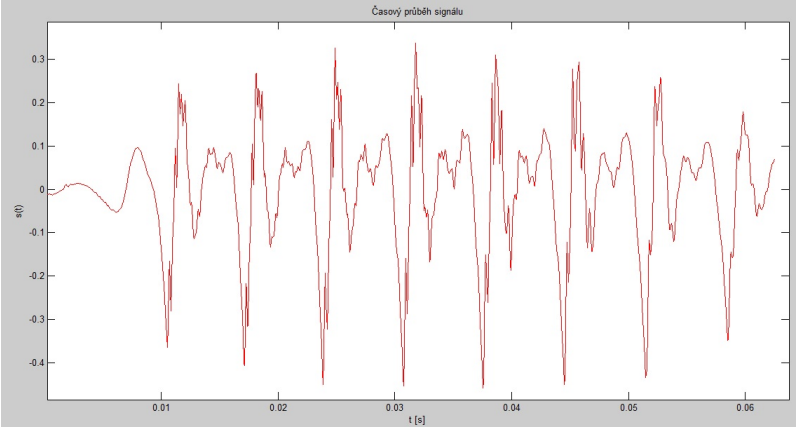
Obr. 2.19: Spektogram hlásky „o“ od mluvčího 001.

Tab. 2.4: Získané hodnoty pásem prvních tří formantů pro hlásku „y“ od mluvčího 001.

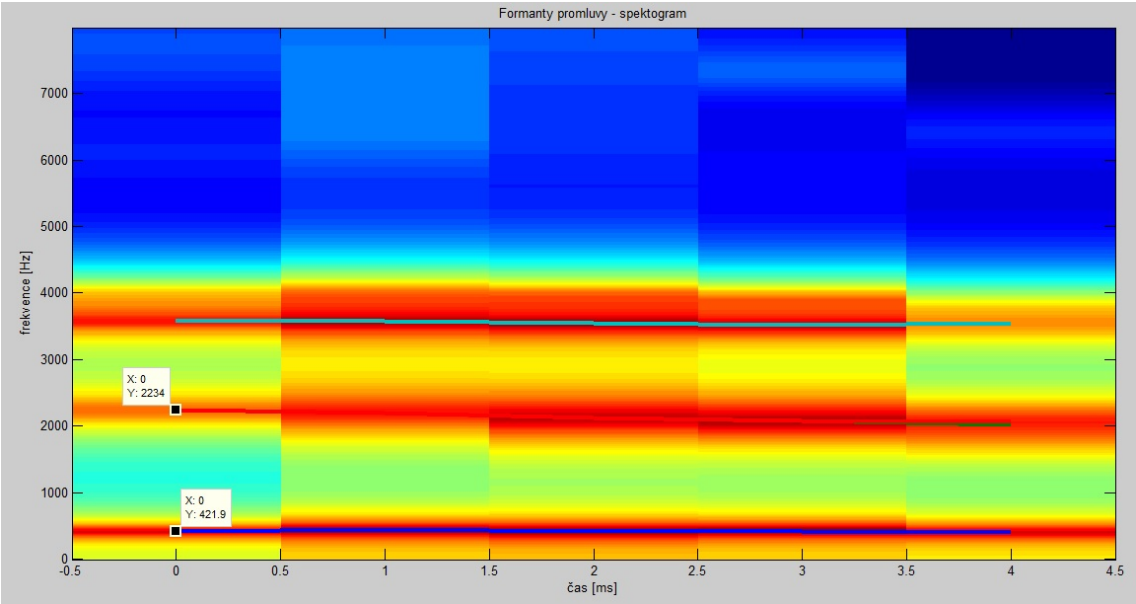
Formanty/ Zdroj	F_1 [Hz] pásmo	F_2 [Hz] pásmo	F_3 [Hz] pásmo
spektogram	421,9	2234	2234
konzole	421,9	2234,4	2234,4
shoda s tbl. 2.1	√	√	√

2.5.2 Porovnání pro více mluvčích

Porovnal jsem vytvořené nahrávky hlásky „a“ pro různé mluvčí a hodnoty formantů jsem zaznamenal do tabulky 2.5, kde získané hodnoty byly ručně odečteny, ale také



Obr. 2.20: Časový průběh hlásky „y“ od mluvčího 001.



Obr. 2.21: Spektogram hlásky „y“ od mluvčího 001.

do tabulky 2.6, kde jsou hodnoty získány skriptem `formant.m`. Jak lze vidět v tabulkách 2.5 a 2.6, tak testované nahrávky hlásky „a“, které jsou od žen (017 – 019) mají vyšší hodnoty formantů než nahrávky od mužů (001 – 004, 012). Hodnoty formantů jsou zde posunuty asi o 200 Hz pro F_1 a F_2, ale pro hodnoty F_3 jsou již téměř srovnatelné

2.6 Práce programu

Program na rozpoznávání mluvčího je napsán v jazyce Matlab (z anglického označení **MAT**rix **LAB**oratory), který je interaktivní programové prostředí a skriptovací

Tab. 2.5: Získané hodnoty pásem prvních tří formantů pro „a“ od více mluvčích – 001, 002, 003, 004, 012 (muži) a 017, 018, 019 (ženy). Hodnoty ručně odečtené ze spektogramu.

Formanty/ Zdroj	F_1 [Hz] pásmo	F_2 [Hz] pásmo	F_3 [Hz] pásmo
001	656,3	1297	2688
002	546,9	1281	2625
003	609,4	1250	2344
004	484,4	1297	2359
012	578,1	1203	2063
017	906,3	1781	2281
018	937,5	1375	2750
019	843,8	1578	2922
z tab. 2.1	700–1100	1100–1500	2500–3000

Tab. 2.6: Získané hodnoty pásem prvních tří formantů pro „a“ od více mluvčích – 001, 002, 003, 004, 012 (muži) a 017, 018, 019 (ženy). Hodnoty získané výpočtem ze skriptu `formant.m`.

Formanty/ Zdroj	F_1 [Hz] pásmo	F_2 [Hz] pásmo	F_3 [Hz] pásmo
001	656,3	1296,9	2687,5
002	546,9	1281,3	2625,0
003	609,4	1250,0	2342,8
004	484,4	1296,9	2359,4
012	578,1	1203,1	2062,5
017	906,3	1781,3	2281,3
018	937,5	1375,0	2750,0
019	843,8	1578,1	2921,9
z tab. 2.1	700–1100	1100–1500	2500–3000

programovací jazyk novější generace. Matlab je výtvorem společnosti MathWorks. Nejprve byl Matlab vyvíjen pro matematické účely, ale postupem času byl doplněn mnoha funkcemi a v současné době je využívaný v široké paletě aplikací. Lze také Matlab využít i k tvorbě uživatelských rozhraní a propojení s programy napsanými v různých jazycích, včetně C, C++, Java a Fortran. [6],[5]

Program se skládá z 12 skriptů (podprogramů), z nichž 11 implementuje algo-

ritmy pro výpočet jednotlivých parametrů. Spouštěcím (hlavním) podprogramem je `worker.m`, který spustí zbylé podprogramy.

2.6.1 Činnost programu

Program pracuje po jistých úsecích, jak již bylo popsáno výše. Nejprve se provede volba zdroje nahrávky mluvčího (`*.wav`), která je analyzována a jsou uložena i druhotná data (kmitočet a počet bitů nahrávky). Dále je provedena preemfáze zvoleného signálu. Následně je provedena segmentace s překrytím 50 %, kde je signál s preemfází upraven do matice. Je použit podprogram `prekryti.m` a obdobný podprogram `prekryti2.m`, kde je nastavena odlišná velikost segmentace. Takto upravený signál je dále zpracováván podprogramy.

- **`prekryti.m`** – je využíván šesti podprogramy (`Spektrum.m`, `intenzita.m`, `energie.m`, `myLPC.m`, `shpn.m` a `formant.m`).
- **`prekryti2.m`** – je použit jen ve dvou podprogramech (`kepstral.m` a `Tnula.m`).

Podprogramy `Spektrum.m`, `intenzita.m` a `energie.m` vypočítají hodnoty, vyplývající z názvů. Následně jsou vypočteny lineární prediktivní koeficienty, střední hodnota průchodu nulou a kepstrální koeficienty, kde je využito Hammingovo okno. Nakonec se určí hodnota T_0 a první při hodnoty formantů (F_1, F_2 a F_3).

Pro jednotlivé hlásky „a“, „e“, „o“, „i“, „u“ a základního tónu od každého mluvčího byl stanoven koeficient, který reprezentuje daného mluvčího jediným číslem. Tyto koeficienty jsou uloženy pro jednotlivé mluvčí v adresáři `data` a to jako samostatné soubory typu `*.dat` (např. *001.dat*). Z tohoto adresáře jsou koeficienty všech mluvčích porovnány s koeficientem pro neznámého mluvčího pomocí skriptu `hledej.m`. Neznámý mluvčí je identifikován jako řečník v adresáři `data` s nejbližším koeficientem promluvy.[7],[4]

2.6.2 Práce s programem

- Nejprve si spustíme Matlab.
- Nastavíme si pracovní adresář (Current Folder), kde jsou uloženy podprogramy.
- Zavoláme `worker.m` napsáním do konzole: **worker** a potvrdíme.
- Objeví se okno, kde nastavíme cestu k testované nahrávce a tu si vybereme. (Testovací nahrávky musí být ve formátu `*.wav`.)
- Zbytek už program dokončí sám. Výsledek komu se neznámý mluvčí podobá nejvíce se zobrazí jako poslední informace.
- Ukončení Matlabu pomocí „x“ nebo zadáním z klávesnice **quit** nebo **exit**.

2.6.3 Výstup programu

Programový výstup pro uživatele je buď grafické vyjádření vypočtených hodnot nebo textový výpis na konzoli.

Nejvíce očekávaný výsledek je označení neznámého mluvčího, které se vypíše do konzole. Program v testovacích podmínkách dosáhl 70 % uspěšnosti. Z poskytnutých nahrávek `pozor_001` až `pozor_020` identifikoval program uspěšně mluvčího ve 14 případech z dvaceti (2.7).

Tab. 2.7: Pokusy o identifikaci mluvčího.

Zadávaná identita	Určená identita	Skutečná identita
001	001	√
002	002	√
003	009	**X**
004	004	√
005	005	√
006	005	**X**
007	007	√
008	008	√
009	009	√
010	010	√
011	011	√
012	012	√
013	020	**X**
014	019	**X**
015	015	√
016	016	√
017	015	**X**
018	018	√
019	005	**X**
020	020	√

3 ZÁVĚR

Během řešení semestrálního projektu byla provedena implementace základních bloků metody pro rozpoznávání řečníka v programu Matlab. K testování daných bloků byly použity vzorky nahrávek `pozor_001.wav` až `pozor_020.wav`. Pro názornou ukázku byly vybrány dvě nahrávky `pozor_017.wav` a `pozor_004.wav`, kde jednotlivé výstupy bloků byly prezentovány formou grafů. Implementované bloky jsou pro zpracování vstupního signálu řečníka.

V bakalářské práci byl doplněn blok pro výpočet základního tónu řeči (T_0) a také blok pro zjištění prvních třech hodnot formantů pro české hlásky. Z promluvy pro každého mluvčího byly vytvořeny jednotlivé nahrávky pro hlásky „a“, „e“, „i“, „o“, „u“, kde byly vypočteny $F1$, $F2$, $F3$ a T_0. Pomocí kterých byl stanoven koeficient promluvy, který byl uložen do souboru pro porovnání s koeficientem neznámého mluvčího.

Neznámý řečník z nahrávek `pozor_001.wav` až `pozor_020.wav` zadal svoji promluvu do programu a z ní vyšla identita na základě nejvyšší podobnosti koeficientů. Při vyhodnocení vstupních a výstupních hodnot byla vypočtena 70 % úspěšnost programu.

LITERATURA

[1] **BÁŇA, Josef.** *POROVNÁNÍ ANALÝZY ŘEČOVÉHO SIGNÁLU V ZÁVISLOSTI NA VĚKU A POHLAVÍ MLUVČÍHO* [online]. Brno, 2008, 2008 [cit. 22. 05. 2013]. Dostupné z: <http://www.vutbr.cz/www_base/zav_prace_soubor_verejne.php?file_id=9227>.
Bakalářská práce. VUT Brno. Vedoucí práce Ing. HICHAM ATASSI.

[2] **HINNER, Jiří.** *Biometrické metody v bezpečnostní praxi: (1).* Třetí pól: Magazín plný pozitivní energie [online]. 2006 [cit. 1. 12. 2012]. Dostupné z: <http://3pol.cz/480-biometricke-metody-v-bezpecnostni-praxi-(1)>.

[3] E-learningová podpora mezioborové integrace výuky tématu vědomí na UP Olomouc. **LUNGOVÁ, Vlasta.** KATEDRA ZOOLOGIE A ORNITOLOGIE, PřF UP Olomouc. *Stavba a funkce hlasového ústrojí.* [online]. 2012, 16. 11. 2012 [cit. 7. 12. 2012].
Dostupné z: <http://pfyziollfup.upol.cz/castwiki/?p=2661>.

[4] **MathWorks.** *MathWorks: Documentation Center* [online]. 1994–2013 [cit. 25. 5. 2013]. Dostupné z: <http://www.mathworks.com/help/signal/>.

[5] **MathWorks.** *MathWorks: MATLAB* [online]. 1994–2013 [cit. 25. 5. 2013]. Dostupné z: <http://www.mathworks.com/products/matlab/>.

[6] **MATLAB. In:** *Wikipedia: the free encyclopedia.* [online]. San Francisco (CA): Wikimedia Foundation, 2013, 26. 3. 2013 [cit. 25. 5. 2013].
Dostupné z: <http://cs.wikipedia.org/wiki/MATLAB>.

[7] **MATLAB Primer CZ. SIGMON, Kermit.** *MATLAB Primer CZ* [online]. 2. vyd. 1989, 11. 2. 2012 [cit. 25. 5. 2013]. Dostupné z: <http://artax.karlin.mff.cuni.cz/~beda/cz/matlab/primercz/matlab-primer.html>.

[8] **PSUTKA, J.; MÜLER, L.; MATOUŠEK, J.; RADOVÁ, V.** *Mluvíme s počítačem česky.* 1. vyd. Praha: Academia, 2006, 746 s. ISBN 80–200–1309–1.

[9] **SIGMUND, M.** *Analýza řečových signálů, 2000.* 1. vyd. Brno: Fakulta elektrotechniky, 2000, 86 s. ISBN 80–214–1783–8.

[10] **VONDRA, Martin.** *Kepstrální analýza řečového signálu.* [online]. 2001 [cit. 3. 12. 2012]. Dostupné z: <http://www.elektrorevue.cz/clanky/01048/index.html>.

[11] **ZAPLATÍLEK, Karel; DOŇAR, Bohuslav.** *MATLAB: začínáme se signály.* 1. vyd. Praha: BEN – technická literatura, 2010, 272 s. ISBN 80–730–0200–0.

SEZNAM SYMBOLŮ, VELIČIN A ZKRATEK

a	konstanta pro výpočet preemfáze
c_{n}	kepstrální koeficienty
δ	odchylka rozdílu
$DPCM$	Diferenční pulsní kódová modulace – Differential Pulse Code Modulation
DTF	Diskrétní Fourierova Transformace – Discrete Fourier transform
E_{n}	krátkodobá energie signálu
f	kmitočet signálu
f_{v}	vzorkovací kmitočet
F_0	základní kmitočet lidského hlasu
F_1	první formant lidské promluvy
F_2	druhý formant lidské promluvy
F_3	třetí formant lidské promluvy
H_{z}	přenosová funkce modelu
LPC	Lineární prediktivní analýza – Linear predictive coding
M_{n}	krátkodobá intenzita signálu
P_ω	výkonové spektrum
PCM	Pulsní kódová modulace – Pulse code modulation
q	kvantizační krok
$R_{\mathrm{FA}}(\theta)$	poměrný počet chyb nesprávného přijetí
$R_{\mathrm{FR}}(\theta)$	poměrný počet chyb nesprávného odmítnutí
$s(t)$	signál spojitý v čase
$\tilde{s}(n)$	odhad řečového vzorku signálu $s(n)$
θ	verifikační práh

T	perioda vzorkování
T_0	základní perioda řeči
$w(n)$	váhová posloupnost – okénko
$y(n)$	signál po preemfázi
Z_{n}	střední počet průchodu nulou

A PŘÍLOHA

A.1 Obsah DVD – Rozpoznávání mluvčího.

- Elektronická verze bakalářské práce.
- Vytvořené skripty a soubory v Matlabu (A.1).
- Nahrávky hlásek, ze kterých byly vytvořeny soubory s koeficienty mluvčích v adresáři `data` (A.2).
- Nahrávky promluvy mluvčích s textem: „Pozor na úraz elektrickým proudem.“ (A.2).

Tab. A.1: Přiložené soubory Matlabu.

Název	Formát	Popis
`data`	`dat`	Uložené koeficienty mluvčích (001 – 020).
`energie.m`	`m-file`	Výpočet vstupní energie.
`formant.m`	`m-file`	Výpočet kmitočtových formantů.
`hledej.m`	`m-file`	Vyhledá a porovná výsledky.
`intenzita.m`	`m-file`	Výpočet intenzity.
`kepstral.m`	`m-file`	Výpočet kepstrálních koeficientů.
`myLPC.m`	`m-file`	Výpočet LPC koeficientů.
`prekryti.m`	`m-file`	Tvorba překrytí signálu (320).
`prekryti2.m`	`m-file`	Tvorba překrytí signálu (1024).
`shpn.m`	`m-file`	Výpočet střední hodnoty průchodu nulou.
`Spektrum.m`	`m-file`	Výpočet spektra signálu.
`Tnula.m`	`m-file`	Výpočet základního tónu.
`worker.m`	`m-file`	Spouštěcí soubor.

Tab. A.2: Nahrávky hlásek a celé promluvy od jednotlivých mluvčích.

Mluvčí	Formát	Promluva
`001 - 020`	`wav`	hláska „a“
`001 - 020`	`wav`	hláska „e“
`001 - 020`	`wav`	hláska „i“/y
`001 - 020`	`wav`	hláska „o“
`001 - 020`	`wav`	hláska „u“
`001 - 020`	`wav`	„Pozor na úraz elektrickým proudem.“