STANOVENÍ CHARAKTERU SEGMENTU ŘEČI S VYUŽITÍM REÁLNÉHO KEPSTRA

Oldřich Horák

Univerzita Pardubice, Fakulta ekonomicko-správní, Ústav systémového inženýrství a informatiky

Abstract: *The extraction of the characteristic features of the speech is the important task in the speaker recognition process. One of the basic features is fundamental frequency of speaker's voice, which can be extracted from the voiced segment of the speech signal. This document describes one of the methods providing possibility to distinguish the voiced and surd segments of the voice signal.*
Keywords: *Features Extraction, Fundamental Frequency, Speaker Recognition, Voice Signal*

1. Úvod

Rozpoznávání mluvčího (angl. *speaker recognition*) představuje v dnešní době jednu z možností zdokonalení identifikace uživatele informačního systému. Jedná se o biometrickou metodu, kde je rozpoznávacím znakem subjektu stavba jeho hlasového ústrojí přímo ovlivňující charakteristiku jeho hlasu. Zmíněná charakteristika je tvořena několika skupinami příznaků, které se získávají z různých částí řečové promluvy. Některé z nich lze získávat pouze metodami pracujícími s předem daným textem, který je mluvčím vysloven, a jsou tedy závislé na příslušném textu (angl. *text dependent methods*). To s sebou pochopitelně nese jistá omezení plynoucí z porovnávání charakteru celých slov nebo větných úseků. Metody nezávislé na textu (angl. *text independent methods*) jsou poněkud náročnější a opírají se ve značné míře jen o porovnávání elementárních charakteristik hlasu mluvčího.

Tyto metody obvykle pracují s krátkými řečovými úseky různých typů, které jsou určitým způsobem specifické. Některé charakteristické příznaky lze například získávat z neznělých úseků, které přibližně odpovídají souhláskám. Znělé úseky obvykle odpovídají samohláskám, případně slabikotvorným souhláskám. Na této úrovni jsou získané příznaky využitelné nejen v úlohách rozpoznávání řečníka, ale i pro úlohy zabývající se rozpoznáním řeči ve smyslu jejího významu. [3]

2. Charakter segmentu řeči

2.1 Znělé úseky a základní tón řeči

Výběr znělých úseků řečové promluvy je velmi významným krokem několika postupů při rozpoznávání řeči nebo řečníka [2]. Znělé úseky řeči, jak již bylo uvedeno, obsahují některé charakteristické znaky určující vybrané parametry hlasového traktu

mluvčího. To umožňuje mimo jiné s odpovídající mírou úspěšnosti rozlišit různé řečníky, případně určit identitu řečníka.

Jedním ze znaků, které se vyskytují právě jen ve znělých částech promluvy, je základní tón (základní frekvence, fundamentální frekvence, angl. *fundamental frequency*). Tato frekvence je do jisté míry charakteristická pro konkrétního řečníka a je v krátkém úseku promluvy konstantní. Pozorováním tohoto znaku v delším úseku řeči lze sledovat melodii promluvy (např. stoupání a klesání hlasu v průběhu věty).

Použijeme-li opačný pohled, můžeme z přítomnosti tohoto významného charakteristického znaku usuzovat, zda je daný segment řečové promluvy znělý či neznělý.

2.2 Používané způsoby stanovení charakteru

V současné době jsou publikovány různé způsoby, jak určit charakter řečového segmentu. Obvykle je využito nějaké vlastnosti, která se u znělých a neznělých úseků řeči významně odlišuje.

Jedním z postupů je sledování poměru energií v jednotlivých kmitočtových pásmech. Kmitočtový rozsah je v závislosti na vzorkovací frekvenci rozdělen do čtyř subpásem, pro něž je vypočtena energie signálu. Porovnáním rozložení energií v těchto úsecích kmitočtového rozsahu je odhadnut charakter segmentu řeči. [1]

Jiný postup, který je též částečně založen na rozložení nízkých a vysokých frekvencí v řečovém úseku, využívá vztahu mezi bazální frekvencí (přibližně danou středním počtem průchodů signálu nulovou hodnotou) a krátkodobou energií. [1, 3]

2.3 Přítomnost základního tónu

V literatuře (např. [1, 4]) je uváděn postup, jak ze signálů znělých úseků řečové promluvy stanovovat základní tón. Zároveň jsou diskutovány postupy, jak znělost či neznělost segmentu určit nebo alespoň odhadnout. Zvolme tedy opačný postup a odhadujme znělost segmentu podle přítomnosti základního tónu. Tato práce se bude zabývat odhadnutím znělosti řečového segmentu podle výskytu základního tónu, který bude stanoven kepstrální metodou. Následně budou výsledky porovnány s metodou určení charakteru segmentu řeči pomocí krátkodobé energie a počtu průchodů signálu nulou [1].

3. Příprava záznamu pro zpracování

K provedení experimentu je využito záznamů řeči, které byly pořízeny pomocí mikrofonu a aplikace pro záznam zvuku osobním počítačem. Jednotlivé záznamy trvají řádově několik sekund a pro účely dalších výpočtů je nutné je rozdělit na segmenty dané délky.

3.1 Značení použitých veličin

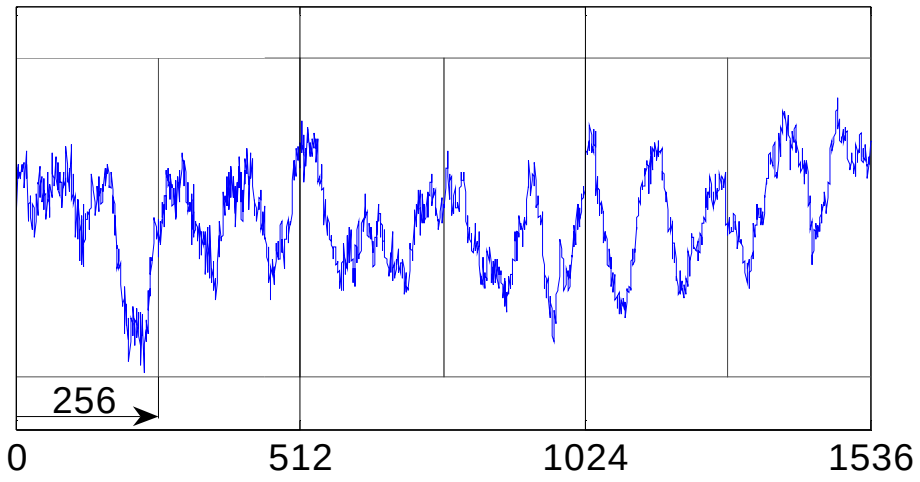
V dalších odstavcích budeme operovat s veličinami popisujícími vlastnosti řečového signálu, je tedy vhodné zde uvést jejich stručný výčet včetně značení, které bude použito:

- f_{vz} – vzorkovací frekvence
- F_0 – frekvence základního tónu
- T_0 – perioda základního tónu
- ZCR – počet průchodů signálu nulovou hodnotou
- E – krátkodobá energie
- c_R – posloupnost koeficientů reálného kepstra

Všechny zpracovávané záznamy byly pořízeny se vzorkovací frekvencí $\boldsymbol{f_{vz} = 22050\ Hz}$.

3.2 Rozdělení záznamu na segmenty

Výpočet reálného kepstra budeme provádět nad segmenty délky ***512 vzorků***, což při použité vzorkovací frekvenci odpovídá přibližně době trvání ***23 ms***. To splňuje požadavek, aby byl signál stacionární, tj. aby se po tuto dobu podstatně neměnily parametry hlasového traktu [4]. Ve výpočtu budeme využívat překrytí segmentů o ***1/2 délky***. Je tedy vhodné rozdělit signál záznamu na ***„půlsegmenty“*** o délce ***256 vzorků***, jak ukazuje Obr. 1.

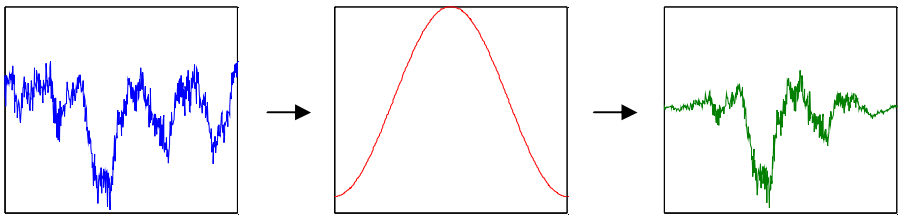


Obr. 1: Rozdělení signálu na segmenty a půlsegmenty

Zdroj: vlastní

Každý vyhodnocovaný segment se tedy bude skládat ze dvou sousedních *půlsegmentů* a postupnou iterací přes tyto *půlsegmenty* zajistíme požadované překrytí segmentů při výpočtu. Před vlastním výpočtem reálného kepstra aplikujeme na data

segmentu *Hammingovo okno* příslušné délky [4]; Obr. 2 ukazuje původní signál, okno a výsledný signál.



Obr. 2: Aplikace Hammingova okna na segment

Zdroj: vlastní

4. Výpočty parametrů

Před vlastním stanovením přítomnosti základního tónu v segmentu řečové promluvy vypočítáme pomocí vztahu (1) koeficienty reálného kepstra ze vzorků signálu daného segmentu.

$$c_{R[i]} = \mathrm{Re}\left\{IFFT\left(\ln\left|FFT\left(S_{[i]}\right)\right|\right)\right\} \qquad (1)$$

Koeficienty $\boldsymbol{c_{R[i]}}$ jsou tedy reálnou složkou *inverzní Fourierovy transformace* ***IFFT*** přirozeného logaritmu z hodnoty modulu *Fourierovy transformace* ***FFT*** vzorků vstupního signálu $\boldsymbol{S_{[i]}}$.

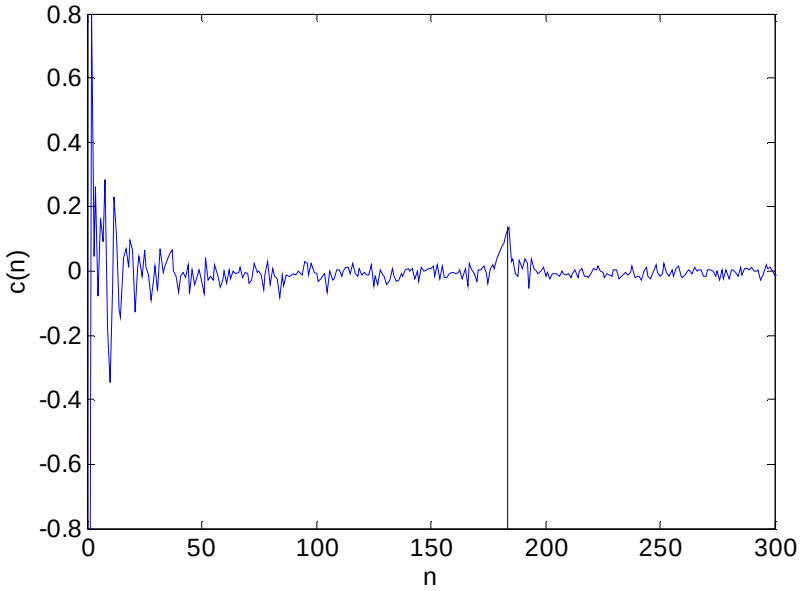
Z indexu kepstrálního koeficientu snadno stanovíme hodnotu odpovídající frekvence a periody. Pro ***k-tý*** koeficient platí vztahy (2).

$$F_k = \frac{f_{vz}}{k}\ , \qquad T_k = \frac{1}{F_k} \qquad (2)$$

Je tedy zřejmé, že výpočet je závislý na vzorkovací frekvenci $\boldsymbol{f_{vz}}$ a pořadí $\boldsymbol{k}$ kepstrálního koeficientu.

4.1 Stanovení přítomnosti základního tónu

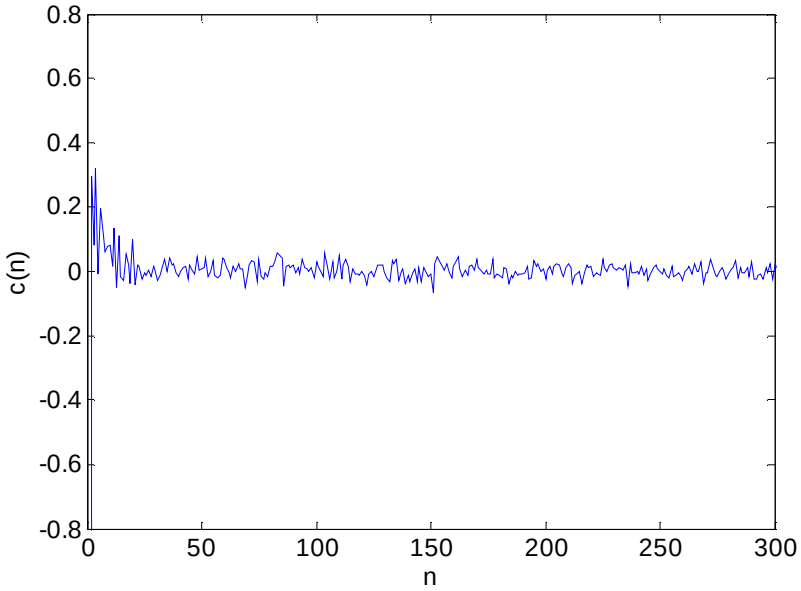
Pro rozlišení charakteru řečového segmentu a stanovení znělosti budeme zkoumat přítomnost základního tónu. Pokud se v reálném kepstru objeví koeficient se znaky základní frekvence, budeme segment považovat za znělý. Kepstrum s vyznačeným maximem je na *Obr. 3*, přičemž $\boldsymbol{n}$ je pořadí kepstrálního koeficientu a $\boldsymbol{c(n)}$ je jeho hodnota. Maximum odpovídající základnímu tónu je vyznačeno svislou čarou.



Obr. 3: Reálné kepstrum znělého segmentu

Zdroj: vlastní

V případě, že přítomnost základního tónu neprokážeme, klasifikujeme segment jako neznělý, O*br. 4*.



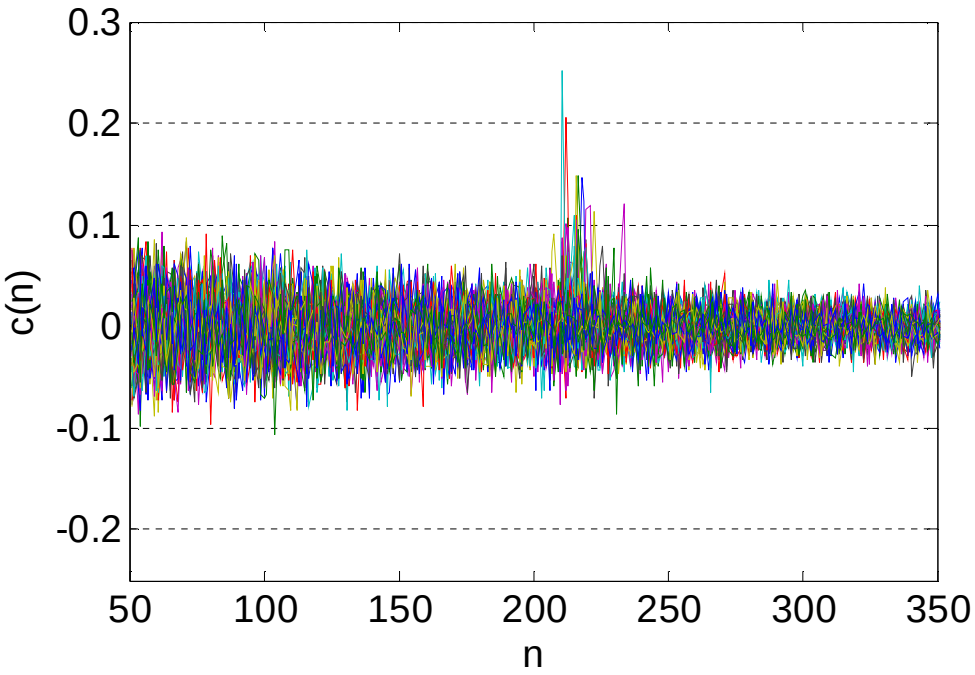
Obr. 4: Reálné kepstrum neznělého segmentu

Zdroj: vlastní

Jak bylo výše vysvětleno, budeme hledat výrazné maximum, které by se mělo vyskytovat mezi kepstrálními koeficienty s indexy ***50*** až ***350***. To je dáno obvyklou frekvencí základního tónu lidské řeči v rozsahu ***60*** až ***400 Hz*** [1, 3], se vzorkovací

frekvencí f_{vz} = ***22050 Hz*** pak podle (2) dostáváme uvedený rozsah sledovaných kepstrálních koeficientů. Vzhledem k tomu, že se počet koeficientů výsledného kepstra podle vztahu (1) rovná počtu vzorků v segmentu a z nich se využívá pouze první (levá) polovina, je třeba segment před výpočtem doplnit zprava nulovými hodnotami na délku ***1024***, čímž dostáváme ***512*** použitelných koeficientů. Kdybychom tuto úpravu neprovedli, dostali bychom pouze poloviční počet koeficientů a ty by nepokryly celý sledovaný interval.

Je zřejmé, že vždy nalezneme nějakou maximální hodnotu. Abychom příslušný koeficient mohli považovat za příznak základního tónu, musí být toto maximum výrazně vyšší, než ostatní hodnoty v příslušném rozsahu koeficientů. Musíme tedy stanovit prahovou hodnotu závislou na průměru ostatních koeficientů v prohledávaném intervalu. Na *Obr. 5* jsou složena kepstra segmentů řečového signálu pro sledovaný interval koeficientů.



Obr. 5: Maxima koeficientů ve sledovaném intervalu

Zdroj: vlastní

Patrné je především to, že koeficient odpovídající základnímu tónu (pokud takový existuje) má vždy znatelně vyšší hodnotu než koeficienty ostatní. Experimentálně byl práh stanoven na *pětinásobek průměrné hodnoty* nezáporných koeficientů.

4.2 Výpočet dalších parametrů segmentů

Dalšími parametry, které budeme při posouzení charakteru řečových segmentů využívat, jsou počet průchodů signálu nulou ***ZCR*** a krátkodobá energie ***E***. [1]

Tyto veličiny určíme pomocí vzorců (3) a (4) pro ***i-tý*** segment, kde ***N*** je počet vzorků v příslušném segmentu a x_n je hodnota ***n-tého*** vzorku.

$$ZCR_i = \frac{1}{2}\sum_{n=0}^{N-2}\left|\mathrm{sgn}(x_n) - \mathrm{sgn}(x_{n+1})\right| \tag{3}$$

$$E_i = \frac{1}{N}\sum_{n=0}^{N-1}(x_n)^2 \tag{4}$$

Obě veličiny jsou vzhledem k dělení signálu na segmenty aditivní, výpočet tedy můžeme zrychlit tak, že hodnoty veličin stanovíme pro jednotlivé *půlsegmenty* a výsledky pro příslušný segment určíme jako součet hodnot *půlsegmentů*, z nichž je složen. To lze vyjádřit pomocí vztahů (5) a (6), kde hodnoty označené symbolem vlnovky odpovídají příslušným *půlsegmentům*.

$$ZCR_i = \widetilde{ZCR}_{2i} + \widetilde{ZCR}_{2i+1} \tag{5}$$

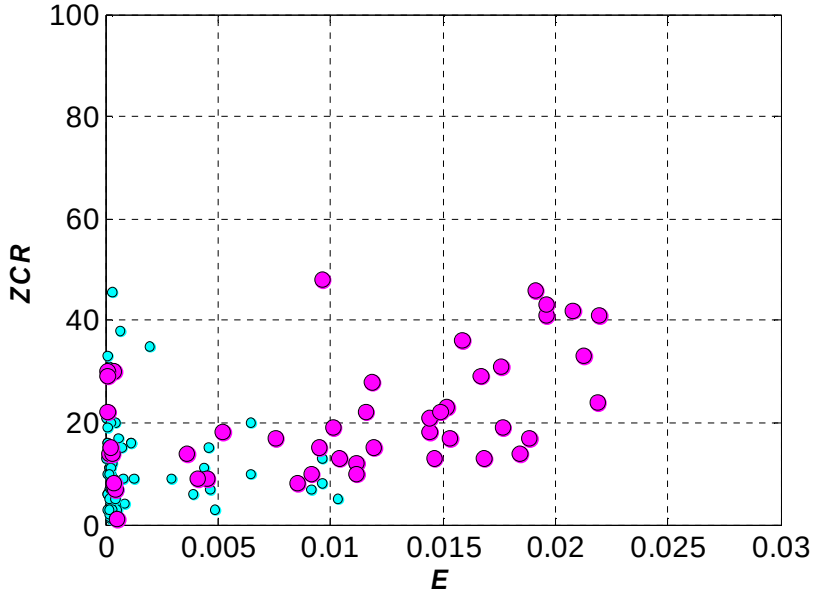
$$E_i = \widetilde{E}_{2i} + \widetilde{E}_{2i+1} \tag{6}$$

V případě, že se hodnoty počtu průchodů nulou a krátkodobé energie v sousedních *půlsegmentech* významně liší, můžeme předpokládat, že do daného segmentu spadá řečový předěl, tedy změna povahy signálu ze znělého na neznělý nebo naopak. Pokud chceme pro další experiment vybírat např. pouze čistě znělé části signálu, je vhodné považovat předělový segment automaticky za neznělý.

5. Ověření teorie

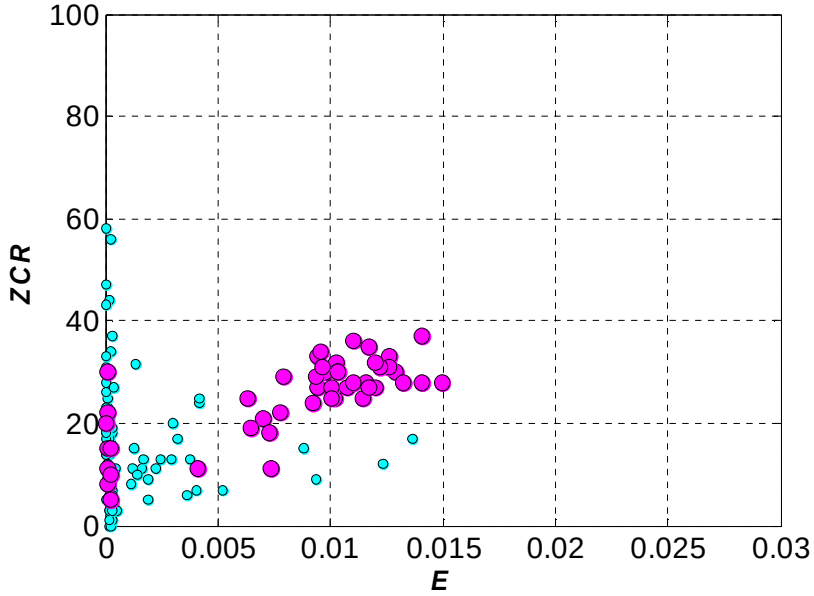
Nyní ověříme teorii uvedenou v [1], kde se popisuje určení charakteru segmentu pomocí hodnot počtu průchodů signálu nulou ***ZCR*** a krátkodobé energie segmentu ***E***. Podle předpokladu se znělé úseky řečového signálu vyznačují větší krátkodobou energií a menším počtem průchodů nulou, než je tomu u neznělých úseků. Vypočítané hodnoty můžeme zobrazit do grafu a zároveň použít odlišné symboly pro segmenty, v nichž jsme našli nebo nenašli příznak přítomnosti základního tónu. Na *Obr. 6 až 10* jsou vyneseny hodnoty počtu průchodů signálu nulou a krátkodobé energie pro segmenty slov „**jedna**“ až „**pět**“ (základní číslovky, které se například vyžadují jako součásti hesla a především se jako slova od sebe poměrně liší).

Segmenty, v nichž byla nalezena frekvence základního tónu (znělé segmenty), jsou znázorněny většími body (v barevné verzi fialově). Neznělé segmenty jsou vyznačeny menšími body (v barevné verzi světle modře).



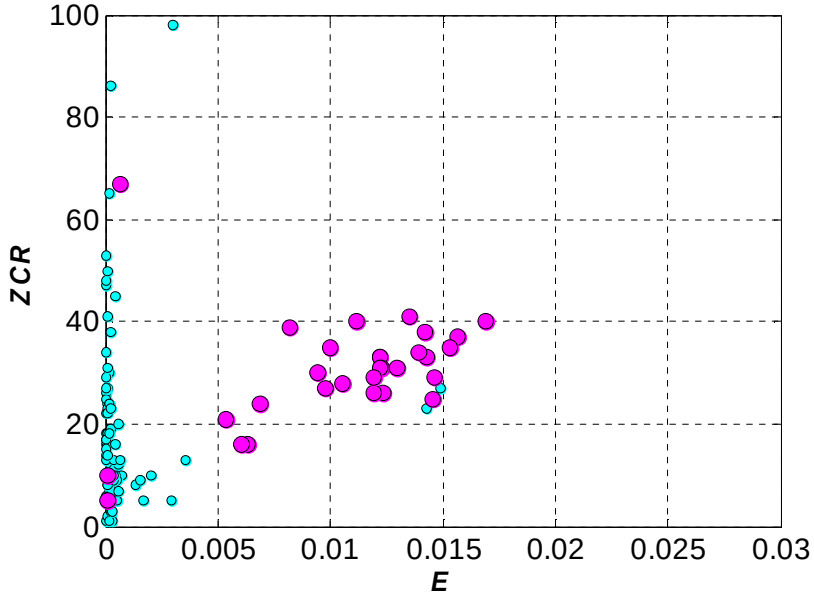
Obr. 6: Znělé a neznělé segmenty slova „jedna“

Zdroj: vlastní



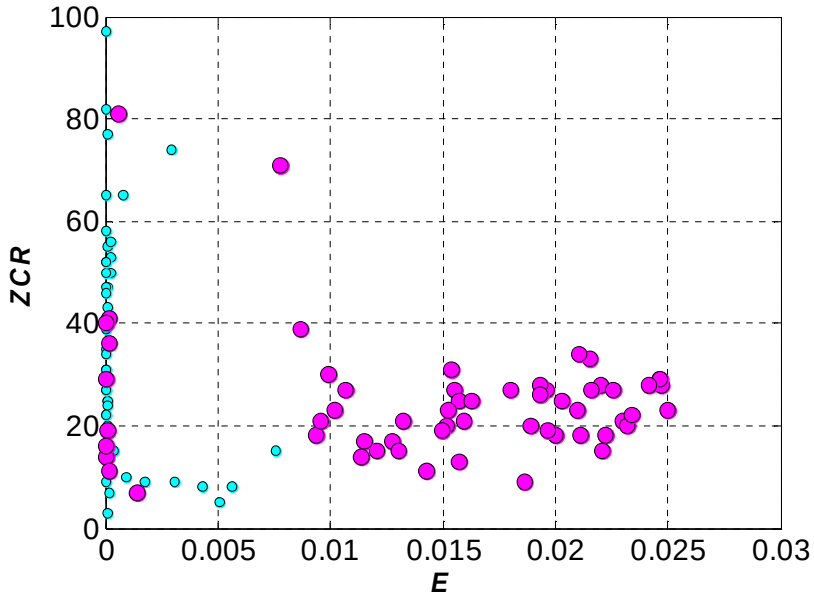
Obr. 7: Znělé a neznělé segmenty slova „dvě“

Zdroj: vlastní



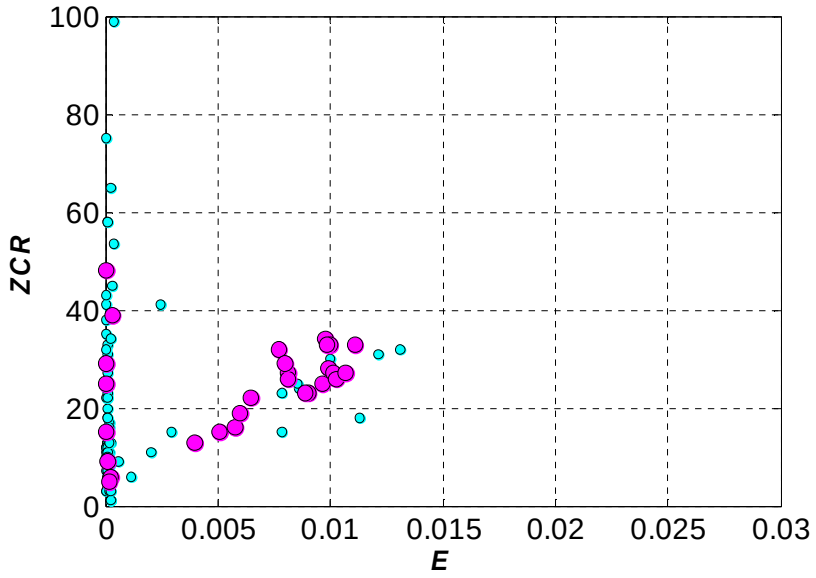
Obr. 8: Znělé a neznělé segmenty slova „tři“

Zdroj: vlastní



Obr. 9: Znělé a neznělé segmenty slova „čtyři“

Zdroj: vlastní



Obr. č. 10: Znělé a neznělé segmenty slova „pět“

Zdroj: vlastní

Z Obr. 6 až 10 je patrné, že segmenty, v nichž byla prokázána přítomnost frekvence, kterou lze považovat za základní tón řeči, jsou ve velké míře shromážděny v jiné oblasti grafu, než ostatní segmenty. Rozložení je velmi podobné, přestože jako příklad byla vybrána slova, která si nejsou příliš podobná.

Za povšimnutí stojí i poměrně významná skupina segmentů se ***ZCR*** v intervalu *od nuly do cca 40* a téměř *nulovou* krátkodobou energií ***E***, které se vyskytují podél svislé osy. Jedná se o segmenty signálu s vyššími frekvencemi a nízkou změnou amplitudy signálu, které odpovídají šumovým složkám řeči [3]. V těchto segmentech se základní tón nevyskytuje.

Podobný graf uvedený v [1] jako obr. 7 zobrazuje rozložení většího počtu segmentů, jejichž charakter byl stanoven experimentálně (patrně prostým poslechem). Rozložení znělých a neznělých segmentů odpovídá výsledkům, které byly stanoveny pomocí přítomnosti základního tónu v segmentu.

6. Závěr

V úvodu jsme si položili otázku, zda je možné stanovit charakter segmentu řečové promluvy pomocí přítomnosti základního tónu řeči. Jednotlivé kapitoly se zabývaly přípravou signálu a příslušnými výpočty. Pro stanovení základního tónu byla použita kepstrální metoda, která je rychlá a výpočetně nenáročná v porovnání s jinými postupy, což je vyváženo nižší přesností a menší odolnosti vůči šumu [1].

Pokud by účelem bylo vyloučit neznělé segmenty, případně segmenty, jejichž charakter je obtížné určit, pak je možno tento způsob stanovení charakteru segmentu považovat za dostačující, protože přítomnost základního tónu tímto způsobem lze prokázat, nikoliv vyloučit.

Z výsledných grafů je zřejmé, že vhodným omezením oblasti lze vybrat téměř výhradně jen znělé segmenty i za cenu toho, že nebudou vybrány všechny, jak je pomocí lineárního klasifikátoru ukázáno v [1]. Postup, který byl popsán v tomto článku, umožňuje přesnější výběr, je však výpočetně náročnější.

Použité zdroje:

[1] ATASSI, H. Metody detekce základního tónu řeči. Elektrorevue [online]. 2008, 4,[cit. 2010-06-08]. Dostupný z WWW: <http://www.elektrorevue.cz/cz/clanky/zpracovani-signalu/0/metody-detekce-zakladniho-tonu-reci/>. ISSN 1213-1539.

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

[3] PSUTKA, J., et al. *Mluvíme s počítačem česky*. 1. Praha : Academia, 2006. Prozodické vlastnosti řeči, s. 752. ISBN 80-200-1309-1.

[4] VONDRA, M. Kepstrální analýza řečového signálu. Elektrorevue. 2001, 48. ISSN 1213-1539.

Kontaktní adresa:

Ing. Oldřich Horák
Univerzita Pardubice
Fakulta ekonomicko-správní
Ústav systémového inženýrství a informatiky
Studentská 84
532 10 Pardubice
e-mail: oldrich.horak@upce.cz
tel. č.: +420 466 036 038