

POROVNÁNÍ METOD PRO MĚŘENÍ ODLIŠNOSTI SPEKTER SIGNÁLŮ

P. Bergl

ČVUT v Praze, Fakulta elektrotechnická

Abstrakt

V příspěvku je porovnáno osm metod pro měření odlišnosti spekter dvou signálů. Je to vzdálenost spektrální, kepstrální, rozdíl *lpc* (Linear Predictor Coefficients), rozdíl kepstrální koeficientů, Itakurova divergence, Kullback-Leiblerova divergence, vzdálenost daná GLR (General Likelihood Ratio) a vzdálenost daná výstupem Bayesovského detektoru. Šest metod je založeno na autoregresním (AR) modelu. Stručně je popsáno generování syntetických AR signálů s požadovanými spektrálními vlastnostmi. Metody jsou porovnány na základě experimentů zaměřených na jejich linearitu, konzistenci aj. V další části jsou míry použity na zpřesnění hranic mezi hláskami v řečovém signálu.

1 Definice měr odlišnosti

Mějme dva signály, $x_1(n)$ a $x_2(n)$. Naším úkolem je zjistit, jak moc jsou si signály (ne)podobné, tj. zjistit míru jejich odlišnosti. Pro jednoduchost předpokládejme, že oba mají stejnou délku N . Nejjednodušší míra odlišnosti signálů může být definována takto:

$$d_x = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} (x_1(n) - x_2(n))^2}. \quad (1)$$

Míra d_x tedy vyjadřuje rozdílnost signálů v časové oblasti. U některých signálů (např. u řeči) nás však více zajímá odlišnost v oblasti spektrální.

1.1 Míry odlišnosti ve spektrální oblasti

Spektra obou signálů vypočteme pomocí diskretní Fourierovy transformace (DFT):

$$X(k) = \mathcal{DFT}\{x(n)\} = \sum_{n=0}^{N-1} x(n) e^{-jnk \frac{2\pi}{N}}, \quad k = 0, \dots, N-1. \quad (2)$$

Míru odlišnosti ve spektrální oblasti pak definujeme:

$$d_X = \sqrt{\frac{1}{N} \sum_{k=0}^{N-1} (X_1(k) - X_2(k))^2}. \quad (3)$$

Se spektrem úzce souvisí tzv. *autoregresní model* signálu, který předpokládá, že každý vzorek signálu je lineární kombinací p předcházejících vzorků signálu, s přičteným šumem $e(n)$. Tedy:

$$d(n) = - \sum_{i=1}^p b_i d(n-i) + G \cdot e(n), \quad n = p, \dots, N-1, \quad (4)$$

kde p je *řád* AR modelu, konstanty b_i se nazývají koeficienty AR modelu (též *lpc* - Linear Predictor Coefficients), G je koeficient zesílení. Pro šum $e(n)$ musí platit následující podmínky:

1. normální rozdělení,

2. nulová střední hodnota,
3. nekorelovanost.

Všimněme si podobnosti rovnice 4 s rovnicí IIR filtrace. Pokud na vstup IIR filtru s přenosovou funkcí:

$$H(z) = \frac{G}{1 + b_1 z^{-1} + \dots + b_p z^{-p}} \quad (5)$$

přivedeme bílý šum, dostaneme na výstupu filtru příslušný AR signál. Ve spektrální oblasti můžeme psát:

$$X(z) = H(z) \cdot S(z), \quad (6)$$

kde $S(z)$ je spektrum bílého šumu, které je (teoreticky) konstantní. Tzn., že přenosová charakteristika filtru $H(z)$ aproximuje spektrum signálu $X(z)$.

Pomocí lpc tak definujeme další míru odlišnosti, danou jejich rozdílem:

$$d_{lpc} = \sqrt{\frac{1}{p} \sum_{i=1}^p (b_{1,i} - b_{2,i})^2}, \quad (7)$$

kde $b_{1,i}$ resp. $b_{2,i}$ jsou koeficienty signálu $x_1(n)$ resp. $x_2(n)$.

Z lpc vychází také poněkud složitější *Itakurova* míra:

$$d_{Itak} = \log \left[\frac{\mathbf{b}_1^T \mathbf{R}_2 \mathbf{b}_1}{\mathbf{b}_2^T \mathbf{R}_2 \mathbf{b}_2} \right], \quad (8)$$

kde:

$$\mathbf{b}_1^T = [1, b_{1,1}, b_{1,2}, \dots, b_{1,p}], \quad (9)$$

$$\mathbf{b}_2^T = [1, b_{2,1}, b_{2,2}, \dots, b_{2,p}]. \quad (10)$$

Prvky autokorelační matice \mathbf{R}_2 jsou:

$$r_{i,j} = R(|i - j|), \quad i, j = 0, 1, \dots, p, \quad (11)$$

přičemž

$$R(i) = \sum_{n=0}^{N-i-1} x_2(n)x_2(n+i). \quad (12)$$

Více o Itakurově míře nalezneme např. v [7].

1.2 Míry odlišnosti dané kepstry signálů

Kepstrum signálu $x(n)$ je:

$$c(n) = \mathcal{DFT}^{-1} \{ \log |\mathcal{DFT} \{ x(n) \}| \}, \quad (13)$$

kde operátor $\mathcal{DFT}^{-1} \{ \}$ představuje inverzní DFT danou vztahem:

$$x(n) = \mathcal{DFT}^{-1} \{ X(k) \} = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{jn k \frac{2\pi}{N}}, \quad n = 0, \dots, N-1. \quad (14)$$

Míra odlišnosti kepster (též kepstrální vzdálenost) je:

$$d_{cep} = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} (c_1(n) - c_2(n))^2}, \quad (15)$$

kde $c_1(n)$ resp. $c_2(n)$ je kepstrum signálu $x_1(n)$ resp. $x_2(n)$.

Keprum signálu $x(n)$ může být také definováno pomocí jeho koeficientů b_i a zisku G :

$$\begin{aligned} c(0) &= \log(G^2), \\ c(1) &= b_1, \\ c(n) &= -b_n - \sum_{i=1}^{n-1} \frac{i}{n} c(i) b_{n-i}, \quad 2 \leq n \leq p, \\ c(n) &= -\sum_{i=1}^p \frac{n-i}{n} c(n-i) b_i, \quad n = p+1, p+2, \dots \end{aligned} \quad (16)$$

Pro takto vypočtené kepstrum definujeme keprální vzdálenost poněkud jinak (např. viz [8]):

$$d_{cep1} = 4.34 \cdot \sqrt{2 \sum_{i=1}^p (c_1(i) - c_2(i))^2}. \quad (17)$$

Všimněme si, že neuvažujeme ziskové členy $c(0)$.

1.3 Další míry

1.3.1 Kullback-Leiblerova míra

Uvažujme dva signály s normálním rozdělením popsané středními hodnotami μ a rozptyly σ^2 . Kullback-Leiblerova míra je dána:

$$d_{Kull} = \frac{1}{2} \log \frac{\sigma_2^2}{\sigma_1^2} + \frac{\sigma_1^2}{\sigma_2^2} + \frac{(\mu_2 - \mu_1)^2}{\sigma_2^2} - 1. \quad (18)$$

Námi analyzované signály $x_1(n)$ a $x_2(n)$ mají neznámé pravděpodobnostní rozdělení, vztah 18 tak na ně nemůžeme aplikovat. Stačí si však uvědomit, že aditivní šum v AR modelu má normální rozdělení. Pro výpočet d_{Kull} tak použijeme následující postup:

1. Levinson-Durbinovým algoritmem odhadneme koeficienty $b_{1,i}$ signálu $x_1(n)$;
2. Z rovnice 4 vypočteme šum $e_1(n)$, (dosazujeme $b_{1,i}$, $x_1(n)$, $G = 1$);
3. Z rovnice 4 vypočteme šum $e_2(n)$, (dosazujeme $b_{1,i}$, $x_2(n)$, $G = 1$);
4. Odhadneme μ_1 , μ_2 , σ_1^2 , σ_2^2 ze šumů $e_1(n)$, $e_2(n)$, pomocí těchto odhadů vypočteme d_{Kull} .

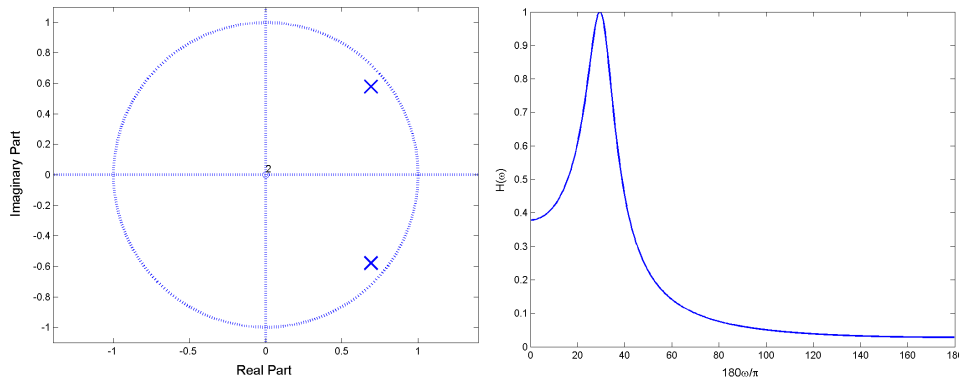
Poznámka: Ve výše uvedeném postupu používáme lpc signálu $x_1(n)$ pro výpočet obou šumů. Je možné totéž učinit pro druhý signál, tj. odhadnout z něj koeficienty a vypočítat oba šumy.

1.3.2 GLR vzdálenost

Výpočet GLR (General Likelihood Ratio) vzdálenosti je netriviální, nebudeme jej zde uvádět (nalezneme jej např. v [2]). Poznamenejme jen, že jeho základem je odhad lpc a zisku G obou signálů pomocí jejich kovariančních matic. GLR vzdálenost je dána:

$$d_{GLR} = -2 \log \frac{l_0}{l_1}, \quad (19)$$

kde l_1 je maximum věrohodnosti za předpokladu libovolných hodnot lpc a zisků G , l_0 je maximum věrohodnosti za předpokladu stejných parametrů, tj. $b_{1,i} = b_{2,i}$ a $G_1 = G_2$.



Obrázek 1: Poloha pólů v z-rovině a frekvenční charakteristika pro $r = 0.9$, $\varphi = 30^\circ$.

1.3.3 Míra určená výstupem RBACDN

Rekurzivní normovaný bayesovský autoregresní detektor změn (RBACDN - Recursive Bayesian Autoregressive Changepoint Detector, viz [8], [9], [4]) je segmentační algoritmus, který analyzuje signál pomocí klouzajícího okna pevné délky. V tomto okně se vypočte pravděpodobnost změny AR modelu pro vzorek signálu ležící uprostřed okna. Jinými slovy, výstup algoritmu je jakousi mírou nepodobnosti mezi signálem v levé a pravé půlce okna.

Pro naše účely si tak definujeme pomocný signál $y(n)$:

$$y(n) = \begin{cases} x_1(n), & 0 \leq n \leq N - 1; \\ x_2(n - N), & N \leq n \leq 2N - 1, \end{cases} \quad (20)$$

ten pak analyzujeme pomocí RBACDN s oknem délky $2N$. Míra d_{Bay} je dána výstupem algoritmu.

2 Porovnání měř na syntetických AR signálech

Autoregresní model může být použit nejen pro analýzu signálů, ale také pro generování syntetických signálů s požadovanými spektrálními charakteristikami. Stačí si „vymyslet“ přenosovou funkci filtru $H(z)$ (ta bude shodná s požadovaným spektrem) a na vstup filtru přivést bílý šum.

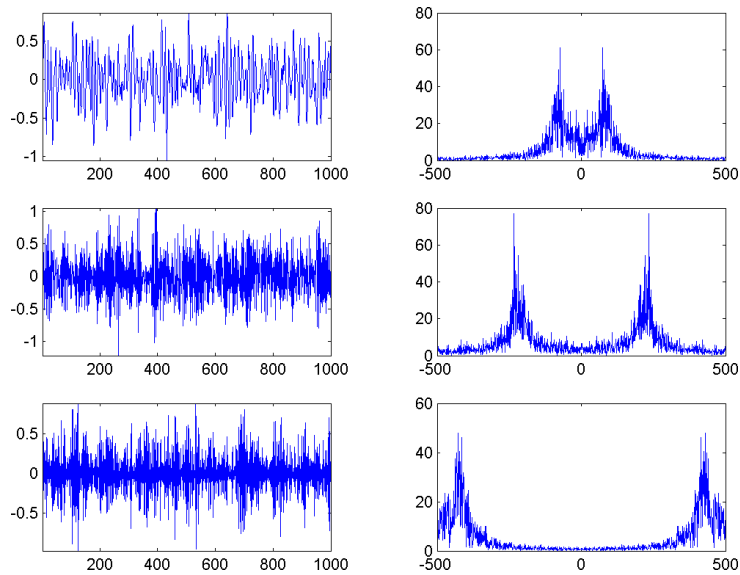
Nejjednodušší příklad je generování signálů druhého řádu, přenosová charakteristika je určena dvojicí komplexně sdružených pólů, jejichž polohu v z-rovině popíšeme pomocí polárních souřadnic r a φ . Ukázka viz obr 1.

Pro naše účely vygenerujeme vždy dvojici signálů s rozdílnými parametry (parametry prvního signálu např. $r_1 = 0.9$, $\varphi_1 = 30^\circ$, parametry druhého např. $r_2 = 0.9$, $\varphi_2 = 50^\circ$). Dostaneme tak signály lišící se zejména ve spektrální oblasti. Míru této odlišnosti proměříme pomocí měř definovaných v předchozích kapitolkách. Provedeme dva jednoduché pokusy, viz dále.

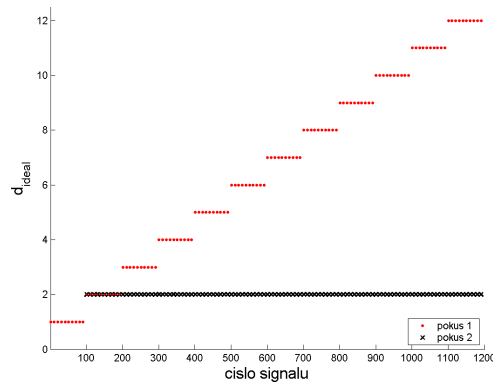
2.1 Pokus 1 - Porovnání měř při rostoucí odlišnosti signálů

Generujeme dvojice signálů, uvažujeme tedy dva filtry. Souřadnice pólů prvního držíme *konstantní*, konkrétně $r_1 = 0.9$, $\varphi_1 = 30^\circ$. Poloměr r_2 pólů druhého filtru bude také neměnný, ale úhel φ_2 postupně *měníme*. Konkrétně začneme s $\varphi_2 = 40^\circ$, poté jej zvětšíme o 10° . Takto pokračujeme až po $\varphi_2 = 150^\circ$. Dostaneme tak celkem dvanáct dvojic signálů (dvanáct „typů“ signálu), jejichž odlišnost ve spektru se postupně zvětšuje. Ukázky signálů viz obr. 2.

Jelikož pracujeme s náhodnými signály, nestačí proměřit pouze jeden signál daného typu (mohli bychom obdržet náhodné výsledky). Proto od každého typu vygenerujeme sto signálů.



Obrázek 2: Ukázka signálů a jejich spekter pro různé polohy pólů. *Nahoře:* $r = 0.9$, $\varphi = 30^\circ$. *Uprostřed:* $r = 0.9$, $\varphi = 80^\circ$. *Dole:* $r = 0.9$, $\varphi = 150^\circ$.



Obrázek 3: Výsledky pokusů pro ideální míru.

Získáme tak celkem 1200 signálů, přičemž v každé stovce mají signály velmi podobná (teoreticky stejná) spektra.

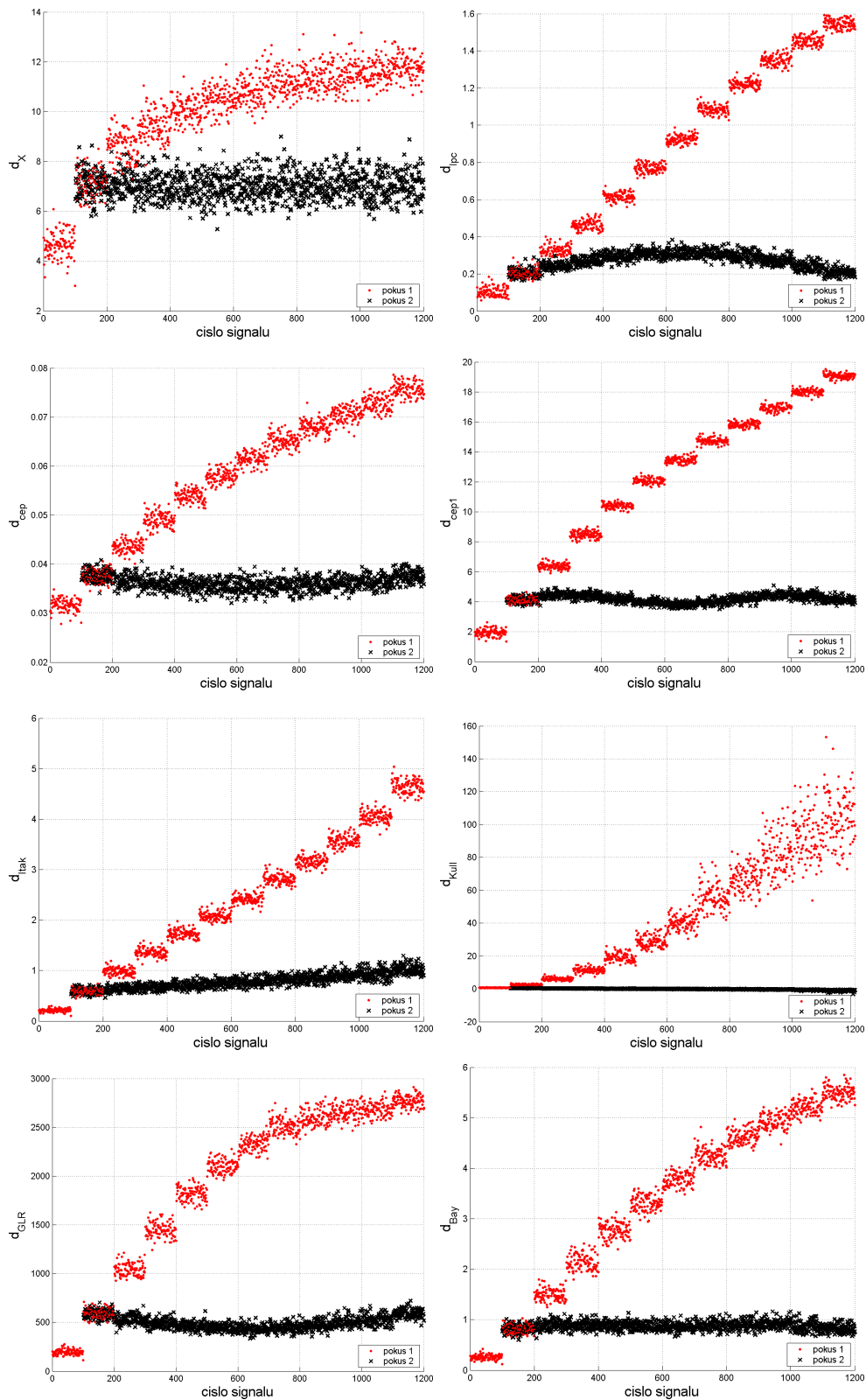
Těchto 1200 signálů proměříme našimi mírami. Pro každý signál vyneseme výsledek do grafu. V ideálním případě (ideální míra) by měl graf vypadat jako „schody“ na obrázku 3 (zatím si všimějme jen červených teček¹, význam křížků bude vysvětlen později).

Výsledky osmi měř vidíme na obrázku 4 (opět si všimějme jen červených puntíků). Zaměřme se zejména na dvě charakteristiky:

- linearitu, tzn. zda „schody“ stoupají rovnoměrně;
- rozptyl míry pro jeden typ signálu, tzn. jak moc jsou jednotlivé „schody rozmazané.“

Tyto vlastnosti oznámujeme jako ve škole (1–nejlepší, 5–nejhorší) a vyneseme do tabulky 1. Nejlepší výsledky vykazují d_{lpc} , d_{cep1} a d_{Itak} , naopak nejhůře se jeví d_X a d_{Kull} .

¹Počet teček by měl být teoreticky 1200, pro přehlednost je však snížen.



Obrázek 4: Výsledky pokusu 1 (tečky) a 2 (křížky) pro míry:

Rozdíl spekter d_X .

Rozdíl kepster d_{cep} .

Itakurova míra d_{Itak} .

GLR vzdálenost d_{GLR} .

Rozdíl lpc koeficientů d_{lpc} .

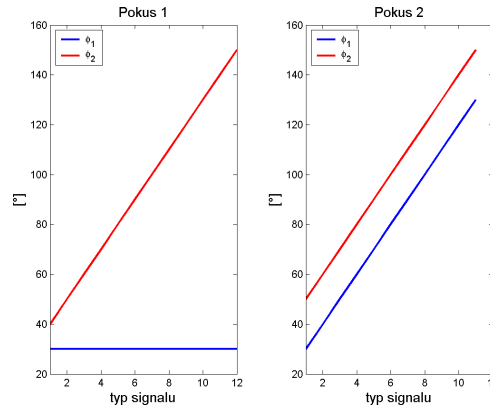
Rozdíl kepstrálních koef. d_{cep1} .

Kullback-Leiblerova míra d_{Kull} .

Míra daná výstupem RBACDN d_{Bay} .

míra	d_X	d_{lpc}	d_{cep}	d_{cep1}	d_{Itak}	d_{Kull}	d_{GLR}	d_{Bay}
linearita	3-	1	2	2	1	3	2-	2
rozptyl	3-	2	2-	1	2	3	2-	2
konstantnost	2	3	2	2	3	1	2	1-
přesahy	2	3	2	1	3	3	1	1
průměr	2,75	2,25	2,125	1,5	2,25	2,5	2	1,625

Tabulka 1: Známkování měř pro oba pokusy.



Obrázek 5: Průběhy úhlů φ_1 a φ_2 pro oba pokusy.

2.2 Pokus 2 - Porovnání měř pro přibližně konstantní odlišnost signálů

Opět uvažujme dva filtry. Parametry pólů prvního budou na počátku pokusu $r_1 = 0.9$, $\varphi_1 = 30^\circ$, parametry druhého filtru $r_2 = 0.9$, $\varphi_2 = 50^\circ$. Poté *oba* úhly φ_1 , φ_2 postupně zvětšujeme o 10° . Takto pokračujeme až po $\varphi_1 = 130^\circ$ a $\varphi_2 = 150^\circ$. Průběhy úhlů φ_1 a φ_2 pro oba pokusy jsou na obr. 5.

Dostaneme tak celkem jedenáct „typů“ signálu, jejichž odlišnost ve spektru je přibližně konstantní a odpovídá druhému schodu z pokusu 1. Pro každý typ vygenerujeme 100 signálů. Dostaneme 1100 signálů, které proměříme našimi mírami. Pro každý signál vyneseme výsledek do grafu. Pro ideální míru by měl graf vypadat jako *konstanta* ve výši druhého schodu, viz obr. 3 (tentokrát si všimějme zejména černých křížků²).

Výsledky osmi měř vidíme na obrázku 4 (opět si všimějme černých křížků). Zaměřme se zejména na dvě charakteristiky:

- konstantnost, tzn. zda míra nekolísá pro různé typy signálu (různé části spektra);
- přesahy do jiných „schodů“.

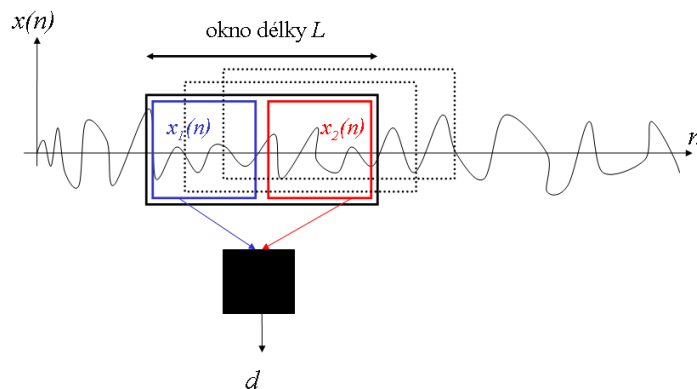
Tyto vlastnosti oznámujeme jako v pokusu 1 a vyneseme do tabulky 1. Nejlepší výsledky vykazuje míra d_{Bay} , naopak nejhůře se chová d_{Itak} a d_{lpc} .

3 Aplikace měř pro segmentaci řečového signál

3.1 Princip segmentace signálu pomocí měř odlišnosti

Představme si, že máme reálný řečový signál. Naším úkolem je nalézt hranice mezi jednotlivými hláskami, tj. signál segmentovat. Pokud by se nám to povedlo, velmi by se tím usnadnila úloha

²Počet křížků by měl být teoreticky 1100, pro přehlednost je však snížen.



Obrázek 6: Výpočet míry d pro segmentaci signálu $x(n)$.

rozpoznání řeči. Přechody mezi hláskami by měly být doprovázeny patrnými změnami ve spektrální oblasti. Otázkou je, zda by se míry odlišnosti zkoumané v této práci nedaly použít pro hledání těchto změn.

Signál budeme analyzovat klouzajícím oknem pevné délky L . Okno nejprve nasadíme na začátek signálu. Levá půlka okna definuje signál $x_1(n)$, pravá půlka definuje $x_2(n)$. Z těchto signálů vypočteme míru odlišnosti d . Poté okno posuneme. Situaci popisuje obrázek 6. Velikosti d vypočtené pro jednotlivé posuny si zapamatujeme a vyneseme je pro časové okamžiky odpovídající středu okna. Výsledkem je křivka C , ve které by měla být místa změn hlásek označena lokálním maximem.

Tento postup použijeme u sedmi měř, u d_{Bay} pro vyčíslení použijeme přímo algoritmus RBACDN.

3.2 Zpřesňování pozic hranice mezi hláskami

Předpokládejme, že řečový signál je nejprve analyzován pomocí HMM (skrytých Markovovských modelů) a jsou odhadnuty pozice hranic mezi hláskami b_i^{HMM} . Signál poté proměříme pomocí jedné z měř, výsledkem je křivka C zmiňovaná v předchozím odstavci. Pozice lokálních maxim v C bychom rádi zužitkovali pro zpřesnění hranic b_i^{HMM} . Použijeme následující „aktualizační“ algoritmus:

For $i = 1, \dots, M$

1. Najdi polohu lokálního maxima t_i^M v blízkosti b_i^{HMM} (poloměr prohledávaného okolí je ε_i).
2. Porovnej hodnotu maxima $C(t_i^M)$ s prahem h .
Když $C(t_i^M) > h$ je nová poloha hranice dána pozicí maxima t_i^M , t.j. $b_i^{new} = t_i^M$;
jinak se pozice nemění, t.j. $b_i^{new} = b_i^{HMM}$.

M je počet hranic, práh h se nastavuje vzhledem k typu míry, např. se h nastaví na střední hodnotu křivky C . Poloměr hledání ε_i se mění v závislosti na rozdílné délce hlásek.

Předpokládejme následující:

- Máme k dispozici databázi řečových signálů.
- Pro každý signál byly ručně vytvořeny kontrolní hranice b_i^{con} .
- Každý signál byl analyzován pomocí HMM $\Rightarrow b_i^{HMM}$ jsou k dispozici.

- Databázi náhodně rozdělíme na trénovací a testovací část.

Nyní každý signál z trénovací části analyzujeme všemi zkoumanými měrami. Pro každou míru použijeme řadu nastavení (L je délka klouzajícího okna ve vzorcích, p je řád AR modelu):

- d_X, d_{cep} : $L \in \{200, 400, \dots, 1200\}$;
- d_{cep1}, d_{Bay} : $L \in \{800, 1000\}, p \in \{2, 4, 8\}$;
- $d_{GLR}, d_{Kull}, d_{Itak}, d_{lpc}$: $L \in \{400, 800\}, p \in \{2, 4, 8\}$.

Dohromady tak obdržíme 48 křivek pro každý signál. Pro každý signál a každou křivku provedeme „aktualizační“ algoritmus popsany výše. Obdržíme tak množiny nových hranic b_i^{new} . Pro rozhodnutí, zda jsou nové hranice lepší či horší než ty původní, použijeme ruční labely b_i^{con} .

Skóre pro odhadnutou pozici můžeme definovat takto:

$$s(b_i) = \begin{cases} 1, & |b_i - b_i^{con}| \leq \tau; \\ 0, & |b_i - b_i^{con}| > \tau; \end{cases} \quad (21)$$

čili $s(b_i)$ je 1, jestliže odhadnutá pozice leží dostatečně blízko k pozici skutečné (kontrolní), τ je zvolená tolerance, např. 15 ms. Toto skóre můžeme vypočítat pro oba odhady, tedy jak pro původní b_i^{HMM} tak pro aktualizované b_i^{new} (nezapomeňme, že b_i^{new} se váže pouze na jednu křivku). Index zlepšení pak můžeme definovat jako

$$I_i = s(b_i^{new}) - s(b_i^{HMM}), \quad (22)$$

tj. I_i se rovná 1, když se pozice hranice zlepšila (aktualizovaná pozice je v toleranci dané rovnicí 21, zatímco původní nebyla);

I_i je 0, když nedošlo k žádné změně (tj. stav původní i nové hranice je totožný; buď jsou obě v toleranci, nebo v ní není ani jedna);

I_i je -1 , když aktualizací došlo ke zhoršení (nová pozice není v toleranci, zatímco původní byla).

Index I_i spočteme pro všechny hranice v signálu, zvlášť pro každou z křivek (každá křivka definuje odlišnou množinu b_i^{new}). Nyní pro jednu konkrétní hranici ve všech signálech a pro jednu konkrétní křivku (jednu množinu b_i^{new}) vyčíslíme následující statistiky:

- S_i^0 je počet signálů, kde $I_i = 0$;
- S_i^1 je počet signálů, kde $I_i = 1$;
- S_i^{-1} je počet signálů, kde $I_i = -1$.

Příklad těchto hodnot pro míru d_{Bay} , okno délky $L = 1000$ a AR řád $p = 8$ nalezneme v tabulce 2 (vlevo). Příklad je pro databázi čítající 18 signálů, trénovací množinu tvoří 10 z nich. Každý signál je pouze jedno slovo „mateřídouška“ vyslovené dítětem, vzorkovací frekvence je 8 kHz. Tabulka 2 popisuje schopnost míry d_{Bay} (s konkrétním nastavením) zlepšit pozici hranic v tomto slově. Podívejme se např. na pátý řádek tabulky (hranice mezi 't' a 'e'). Je zjevné, že ke změně nedošlo pro 2 signály ($S_i^0 = 2$), ke zlepšení došlo u 8 signálů ($S_i^1 = 8$), ke zhoršení nedošlo ani u jednoho ($S_i^{-1} = 0$). Z toho vyplývá, že míra d_{Bay} pro $L = 1000$ a $p = 8$ je schopna výrazně upřesnit pozici hranice mezi 't' a 'e'.

Na druhou stranu, výsledky pro některé jiné hranice jsou dosti špatné, např. pro hranici 'š-k' je 8 signálů beze změny, pozice se zhoršila pro 2 signály. V tabulce 2 (vpravo) jsou uvedeny výsledky pro kepstrální vzdálenost d_{cep} , $L = 1000$. Výsledky pro hranici 'š-k' jsou dobré ($S_i^0 = 6$, $S_i^1 = 4$, $S_i^{-1} = 0$). Zdá se tedy, že kepstrální vzdálenost je pro detekci 'š-k' vhodnější než míra d_{Bay} .

hranice	d_{Bay}			d_{cep}		
	S_i^0	S_i^1	S_i^{-1}	S_i^0	S_i^1	S_i^{-1}
X-m	9	0	1	8	1	1
m-a	8	1	1	9	0	1
a-t	6	4	0	4	5	1
t-e	2	8	0	8	2	0
e-ř	5	4	1	7	2	1
ř-í	5	4	1	6	4	0
i-d	6	2	2	5	2	3
d-ou	3	6	1	5	3	2
ou-š	8	2	0	7	3	0
š-k	8	0	2	6	4	0
k-a	8	1	1	6	3	1
a-X	10	0	0	9	1	0

Tabulka 2: Hodnoty S -statistik pro míru d_{Bay} , $L = 1000$, $p = 8$ a míru d_{cep} , $L = 1000$. Výsledky pro všechny hranice ('X' je pauza), pro 10 signálů z trénovací množiny.

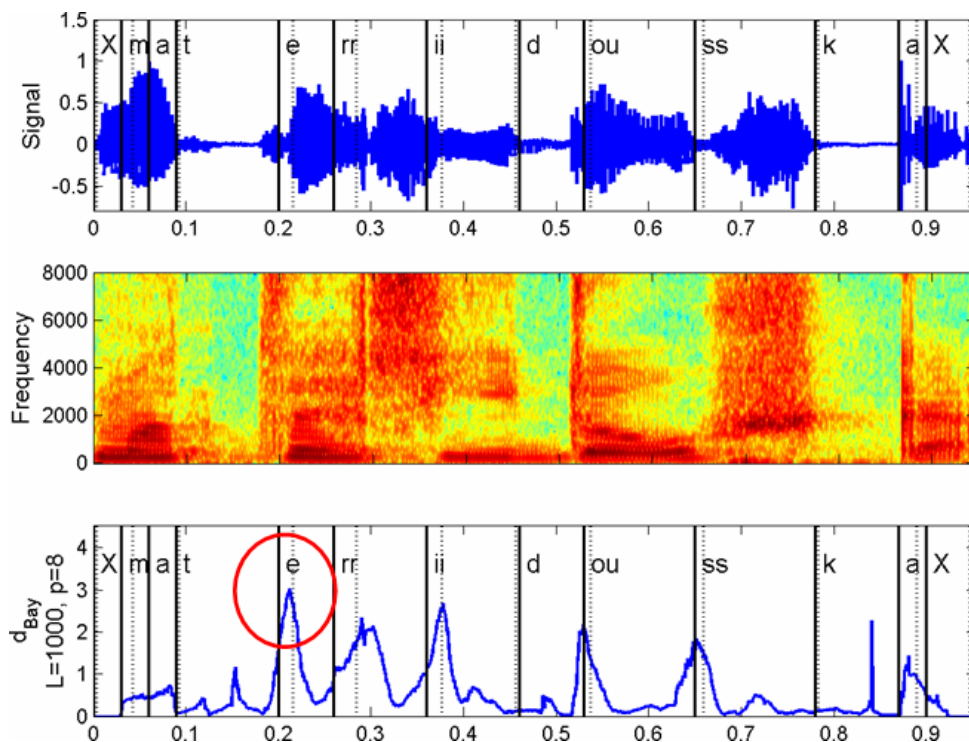
hranice	nej. metody			ověření		
	míra	L	p	S_i^0	S_i^1	S_i^{-1}
X-m	-	-	-	8	0	0
m-a	d_{Kull}	800	8	7	1	0
a-t	d_{GLR}	800	8	4	3	1
t-e	d_{Bay}	1000	8	3	4	1
e-ř	d_{dC}	600	-	3	4	1
ř-í	d_{Itak}	400	4	4	3	1
i-d	-	-	-	8	0	0
d-ou	d_{cep1}	1000	2	6	2	0
ou-š	d_{Bay}	800	2	2	6	0
š-k	d_{cep}	1000	-	6	2	0
k-a	d_{cep}	400	-	5	0	3
a-X	-	-	-	8	0	0

Tabulka 3: Vlevo: Nejvhodnější metody a jejich nastavení pro každou hranici (pro 'X-m', 'i-d' a 'a-X' žádná vhodná metoda nalezena nebyla). Vpravo: Hodnoty S -statistik pro osm testovacích signálů. Hodnoty se vztahují k metodě uvedené na stejném řádku v levé části tabulky.

Uvědomme si, že každá z tabulek reprezentuje hodnoty pro jednu křivku (jedna metoda a nastavení); celkem tak sestavíme stejný počet tabulek jako je křivek, tj. 48. Naším úkolem je teď najít nejvhodnější metodu a její nastavení pro každou z hranic. jinými slovy musíme najít metodu, pro kterou je hodnota S_i^1 maximální a hodnota S_i^{-1} naopak minimální. Výsledky (nejlepší míry a nastavení) nalezneme v tabulce 3 (vlevo).

Tyto metody použijeme pro "aktualizační" algoritmus pro osm signálů z testovací části databáze. Tj. hranice 'm-a' bude aktualizována v každém signálu pomocí Kullback-Leiblerovy divergence; hranice 'a-t' pomocí GLR vzdálenosti, hranice 't-e' pomocí Bayesovského detektoru (jeden z testovacích signálů, křivku pro d_{Bay} , HMM a kontrolní labely nalezneme na obr. 7). Pro další pozice postupujeme obdobně.

Nyní je naším úkolem ověřit, zda výše popsany princip (selekce metod plus "aktualizační" algoritmus) vedl ke zpřesnění hranic v testovacích signálech. Pro aktualizované pozice spočteme S -statistiky, výsledky viz tabulka 3 (vpravo). Je zjevné, že pozice většiny hranic byla zpřesněna.



Obrázek 7: Jeden z testovacích signálů, jeho spektrogram, křivka pro d_{Bay} . Odhady získané pomocí HMM reprezentují plné čáry, čárkovaně jsou vykresleny kontrolní labely. Zobrazená křivka je použita pro aktualizaci hranice 't-e'. Všimněme si, že pozice maxima (tj. i nové hranice) je mnohem blíže správné pozici.

Jinou statistikou popisující přesnost odhadů je průměrná hodnota skóre z rovnice 21. Ta je definována následovně:

$$\bar{s}^{HMM} = \frac{100}{R} \sum_{r=1}^R \sum_{i=1}^N s(b_{i,r}^{HMM}); \quad (23)$$

$$\bar{s}^{new} = \frac{100}{R} \sum_{r=1}^R \sum_{i=1}^N s(b_{i,r}^{new}); \quad (24)$$

kde R je počet signálů; $b_{i,r}$ je odhad i -té pozice v r -tém signálu; \bar{s}^{HMM} tedy reprezentuje průměrný počet případů, ve kterých je HMM odhad dostatečně blízko ke kontrolním labelům (tj. „počet zásahů“), \bar{s}^{new} je „počet zásahů“ po aktualizaci. Pro naši databázi $\bar{s}^{HMM} = 43\%$ a $\bar{s}^{new} = 65\%$. Zřetelné navýšení počtu zásahů demonstruje užitečnost výše popsaného postupu.

4 Závěr

Výkonnost jednotlivých měř v části zaměřené na syntetické signály asi nejlépe vystihuje tabulka 1. Celkově si nejlépe vedly míry d_{cep1} a d_{Bay} , nejhůře d_X a d_{Kull} . Je celkem zajímavé, že míra d_X nevykázala dobrou linearitu v pokusu 1, zatímco míra d_{lpc} ano. Přitom d_{lpc} vychází z lpc , jejichž prostřednictvím se aproximuje spektrum signálu. Tento jev, stejně jako velmi dobré výsledky většiny měř založených na AR modelu, je pravděpodobně způsoben tím, že pracujeme se syntetickými signály. Pro některé reálné signály nemusí být použití AR modelu vhodné, pro ty by pak výkonnost metod s AR modelem mohla výrazně poklesnout.

V části změřené na segmentaci řečových signálů jsme ukázali, že „aktualizační algoritmus“ spolu s precizním výběrem metod může vést k podstatnému zpřesnění hranic mezi hláskami získaných pomocí HMM.

Kullback-Leiblerova divergence se jako jediná zdá být přijatelná pro hledání hranic typu

nazál–samohláska ('m-a'). Pro hranici typu samohláska–okluze ('a-t') byla zvolena GLR vzdálenost. Bayesovský detektor se ukázal být velmi vhodný pro typ explozíva–samohláska ('t-e'). Zatímco většina ostatních detektorů reaguje na explozi 't', Bayesovský detektor spolehlivě reaguje až na hranici mezi hláskami. Kepstrální detektory se zdají být vhodné pro frikativní kontext.

Je nutné však také zmínit některé nevýhody zmíněného přístupu. Pokud jsou HMM odhady naprosto špatné, aktualizace s tím nic nezmuže, poněvadž hledání lokálního maxima ve křivce je omezeno pouze na blízké okolí HMM labelu. Další nevýhodou je výpočetní náročnost algoritmu. Zvláště trénovací část (výpočet většího počtu měř) je značně časově náročná. Algoritmus však rozhodně není určen pro on–line použití jeho hlavním úkolem je usnadnit zpracování rozsáhlých databází.

Reference

- [1] Omar, K.M., Hasegawa-Johnson, M. Levinson, S.: Gaussian Mixture Models of Phonetic Boundaries For Speech Recognition. *IEEE Workshop on Automatic Speech Recognition and Understanding Workshop*. Trento, Italy, December 2001.
- [2] Appel, U., Brandt, A.: Adaptive Segmentation of Piecewise Stationary Time Series, *Information Sciences*, 1983, no. 29, p. 27 – 56
- [3] Ruanaidh, J.O., Fitzgerald, W.J.: Numerical Bayesian Methods Applied to Signal Processing, *Statistics and Computing*, Springer, Berlin, 1996.
- [4] Čmejla, R., Sovka, P.: Recursive Bayesian Autoregressive Change-point Detector for Sequential Signal Segmentation. In *EUSIPCO-2004 – Proceedings [CD-ROM]*. Wien: Technische Universität, 2004.
- [5] Sooful, J.J., Botha, J.C., An acoustic distance measure for automatic cross-language phoneme mapping, *PRASA '01*, pp. 99 – 102, South Africa, November 2001.
- [6] Couvreur, L., Boite, J.-M., Speaker Tracking in Broadcast Audio Material in the Framework of the THISL Project, *Proc. of ESCA ETRW Workshop on Accessing Information in Spoken Audio*, Cambridge (UK), pp. 84 – 89, April 1999.
- [7] J.Psutka, *Komunikace s počítačem mluvenou řečí*, Academia, Praha, 1995
- [8] Čmejla, R.: *Bayesovská detekce náhlých změn*. [Docentská habilitační práce]. Praha: ČVUT, Fakulta elektrotechnická, 2002. 114 s.
- [9] Čmejla, R. – Sovka, P.: Audio Signal Segmentation Using Recursive Bayesian Change-point Detectors. In *Proceedings of the WSEAS International Conferences [CD-ROM]*. New York: WSEAS Press, 2004, vol. 1, s. 1087 – 1091.