SPEECH SEGMENTATION BASED ON VECTOR QUANTIZATION

Petr Andrla
Master Degree Programme (2), FEEC BUT
E-mail: xandrl01@stud.feec.vutbr.cz

Supervised by: Petr Sysel
E-mail: sysel@feec.vutbr.cz

ABSTRACT

Speech segmentation is the process of identifying the boundaries between phonemes in spoken natural languages. This paper describes approach based on vector quantization. In the first step of algorithm, feature extraction is realized. Then speech segments are assigned to calculated centroids. Position where centroid is changed is marked as a boundary of phoneme.

1 ÚVOD

Zpracování řeči se dělí do několika oblastí, jednou z těchto oblastí je předzpracování řeči. Velmi důležitou složkou předzpracování řeči je fonémová segmentace, která se snaží o nalezení hranic mezi jednotlivými řečovými jednotkami. Na přesnosti segmentace závisí úspěšnost dalších kroků při zpracování řečového signálu, ať už je to rozpoznávání, syntéza nebo analýza. Automatické metody segmentace jsou vetšinou založeny na standardních metodách, např. na využití skrytých Markovových řetězců, v menší míře na umělých neuronových sítích. Cílem této práce je ověřit možnost využití vektorové kvantizace pro segmentaci řečového signálu na fonémy.

2 VEKTOROVÁ KVANTIZACE

Kvantizace je aproximací hodnoty vzorku signálu jednou z omezeného počtu číselných hodnot.V případě, že je kvantizován celý blok několika signálů, označuje se tento proces jako vektorová kvantizace. Vektorová kvantizace se často používá při dalším zpracování dat popisující jednotlivé mikrosegmenty řečového signálu, například při kompresi řečového signálu. V tomto případě blok signálů představuje vektor příznaků řečového signálu, získaných pomocí metod krátkodobé analýzy.

2.1 KRÁTKODOBÁ ANALÝZA ŘEČI

Metody krátkodobé analýzy reprezentují řečový signál pomocí příznaků. Z metod krátkodobé analýzy byli použity: krátkodobá energie, krátkodobá střední hodnota průchodů signálu nulovou úrovní, krátkodobá funkce počtu výskytu lokálních extrémů, krátkodobá autokorelační funkce, krátkodobá diskrétní Fourierova transformace, kepstrální analýza. Uvedené metody krátkodobé analýzy jsou podrobněji popsané v literatuře [2].

2.2 KONSTRUKCE KVANTIZÉRU

Jednotlivé vektory Q příznaků vypočítaných pro N zpracovávaných segmentů řečového signálu definují N bodů v Q-rozměrném prostoru. Tyto body jsou v prostoru rozloženy ve shlucích, kdy vektory pro segmenty stejného fonému tvoří vždy jeden shluk X_i. Snahou je ke každému tomuto shluku nalézt centroid, kterým by následně byly nahrazeny všechny body tohoto shluku; to odpovídá vektorové kvantizaci celého signálu. Analyzované segmenty jsou přiděleny k jednotlivým centroidům tak, aby byla minimalizována funkce celkového zkreslení J. K tomu je nutné aby kvantizér splňoval následující kritéria:

1. Vždy pro každý vektor $\mathbf{x}$, charakterizující segment řečového signálu, vybral takový vektor $\mathbf{v}_i$, charakterizující centroid, jehož nahrazením za vektor $\mathbf{x}$ dojde k nejmenšímu zkreslení, musí tedy platit [2]:
$$d(\mathbf{x}, \mathbf{v}_i) \leq d(\mathbf{x}, \mathbf{v}_j) \tag{1}$$
kde $1 \leq i, j \leq L$ a $i \neq j$, a L je počet použitých centroidů.

2. Každý kódový vektor $\mathbf{v}_i$ minimalizoval zkreslení v oblasti X_i ke které je přiřazen. Pak se takový kódový vektor nazývá centroid a určí se pomocí [2]:
$$\mathbf{v}_i = \frac{1}{\mathbf{n}_i} \sum_{x \in X_i} \mathbf{x} \tag{2}$$
kde $\mathbf{n}_i$ je počet vektorů $\mathbf{x}$ v oblasti X_i.

Pro nalezení centroidů a rozdělení segmentů mezi ně byl použit MacQueenův algoritmus [2]. Výsledkem vektorové kvantizace by v ideálním případě byla křivka značící příslušnost segmentu k jednomu centroidu, která je konstantní po dobu trvání jednoho fonému.

Výhoda použití vektorové kvantizace pro segmentaci řeči na fonémy spočívá v tom, že vektorová kvantizace přistupuje ke skupině příznaků jako k jednomu celku a při výpočtu zohledňuje vzájemnou závislost všech příznaků všech mikrosegmentů mezi sebou navzájem. Toho je dosaženo tím, že mikrosegmentu je přidělen jeden bod v Q-rozměrném prostoru.

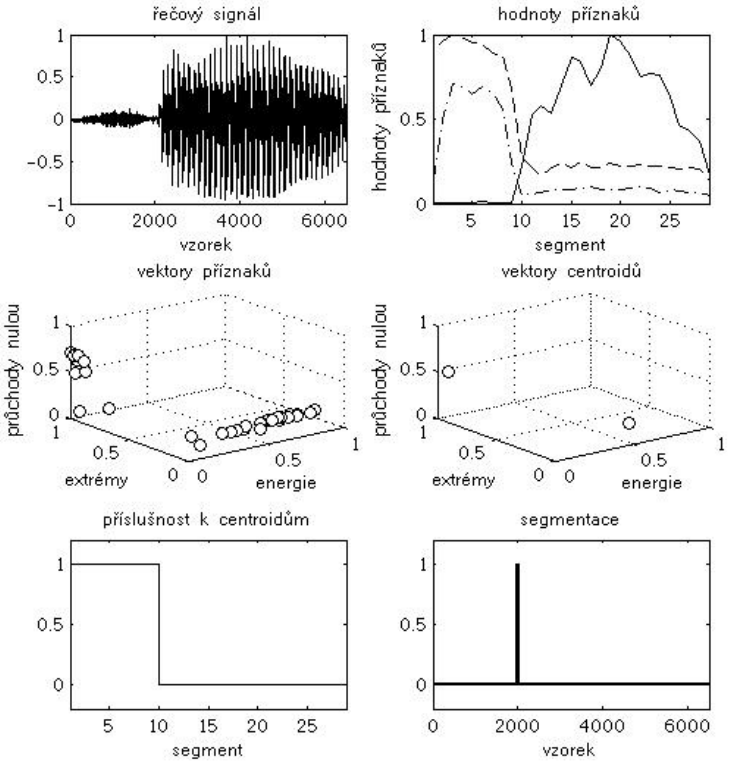
2.3 VEKTOROVÁ KVANTIZACE S POSTUPNÝM PŘIDÁVÁNÍM CENTROIDŮ

V prvním kroku MacQueenova algoritmu je třeba zvolit L počátečních centroidů. Stanovení vhodných počátečních poloh centroidů je důležité k tomu, aby celý algoritmus rychle konvergoval a dále aby dospěl do některého "vhodného" lokálního minima kriteriální funkce J.

Výhodný způsob, jak zajistit volbu počáteční polohy centroidů, je jejich postupné přidávání. A to například přidáním nového centroidu na základě určení nejvzdálenějšího segmentu od stávajících centroidů.

2.4 VÝSLEDKY

Celkem bylo testováno 29 nahrávek obsahující 186 fonémů a 157 hranic fonémů, které byly ručně označeny. Při použití uvedeného algoritmu nebylo rozeznáno 8 hranic. Důvodem nerozpoznání hranic byla např. vzájemná podobnost fonémů, koartikulace či krátké a nevýrazné vyslovení hlásky. Správně rozpoznáno bylo tedy 149 hranic mezi fonémy. Dále byly uprostřed některých fonémů hranice detekovány nesprávně, celkem bylo 15 takto chybně detekovaných hranic.



Obrázek 1: Průběh segmentace pomocí vektorové kvantizace. V pravém horním grafu jsou vykresleny hodnoty příznaků: krátkodobé energie (plnou čarou), počtu průchodů signálu nulovou úrovní (čerchovaně) a počtu lokálních extrémů (čárkovanou čarou).

Na obrázku 1 je znázorněn příklad průběhu segmentace pomocí uvedeného algoritmu. Jedná se o segmentaci difonu "če". Pro možnost názornějšího zobrazení nebyly v tomto případě užity všechny metody krátkodobé analýzy uvedené v kapitole 2.1, ale pouze tři z nich: krátkodobá energie, krátkodobá střední hodnota průchodů signálu nulovou úrovní, krátkodobá funkce počtu lokálních extrémů.

3 ZÁVĚR

Úspěšnost segmentace pohybující se okolo 80% je srovnatelná s výsledky segmentace založené na jiných metodách. Ve výsledcích je také nutno zohlednit, že některé hranice fonémů jsou obtížně detekovatelné i při ruční segmentaci. Další zlepšení úspěšnosti segmentace by mohlo být realizováno výběrem příznaků, které nejvýhodněji rozdělí vektory segmentů ve vektorovém prostoru. Celý projekt je realizován v programovém prostředí MATLAB.

REFERENCE

[1] DELLER, J. R. - HANSEN, J. H. L. - PROAKIS, J. G. *Descrete-Time Processing of Speech Signals.* Reprint Edition. New York: IEEE, 2000. 908 s. ISBN 0-7803-5386-2

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