

Automatické hledání významných pozic v Parkinsonických promluvách založených na rychlém opakování slabik /pa/ - /ta/ - /ka/

M. Novotný, J. Rusz, R. Čmejla

České vysoké učení v Praze, Fakulta elektrotechnická, Katedra teorie obvodů

Abstrakt

Práce se zabývá využitím programového prostředí Matlab pro zpracování patologických promluv pacientů trpících Parkinsonovou nemocí (PN). Promluvy založené na rychlém opakování slabik /pa/ - /ta/ - /ka/, byly postupně segmentovány na jednotlivé slabiky, ve kterých byly poté označeny pozice počátku exploze, vokálu a okluze. Výsledky automatické segmentace byly porovnány s ručně označenými pozicemi. Při zpracování 80 (40 PN; 40 kontrolní skupina (KS)) promluv a celkového počtu 1644 slabik (753 PN; 891 KS) bylo s přesností na 5 ms dosahováno úspěšností pro explozi rovnou 79,35 %, pro vokál rovnou 86,4 % a pro okluzi rovnou 53,3 %.

1 Úvod

Parkinsonova nemoc (PN) je druhým nejrozšířenějším neurodegenerativním onemocněním po Alzheimerově nemoci [1]. Toto onemocnění postihuje zejména populaci starší 50 let a tento fakt společně s celosvětovým stárnutím populace naznačuje, že četnost Parkinsonovy nemoci se bude zvyšovat [1].

Jedním z prodromálních, subklinických PN symptomů, který se vyskytuje v 70–90 % případů, je hypokinetická dysartrie, jejíž hodnocení by mohlo napomoci včasné diagnóze a hodnocení vývoje onemocnění [2], [3].

Pro potřeby hodnocení hypokinetické dysartrie se mimo jiné využívá i řečová diadochokinetická (DDK) úloha, založená na rychlém opakování slabik /pa/-/ta/-/ka/. Cílem této práce je ukázat možnost využití programu Matlab pro automatické hodnocení těchto promluv.

Pro hodnocení kvality promluv je třeba v jednotlivých slabikách označit tři základní pozice. První pozicí je počátek exploze (E), druhou pozicí je počátek vokálu (V) a poslední je počátek okluze (O).

2 Metodika

V této sekci bude uvedena informace o databázi a metodách využitých při zpracování jednotlivých signálů. Zároveň, vzhledem k rozdílnému charakteru jednotlivých pozic, bude v každé sekci stručně shrnuta problematika.

2.1 Databáze

Pro návrh algoritmu byla využita část databáze shromážděná pro práci [4], tato část obsahovala pouze nahrávky obsahující promluvy řečové DDK úlohy.

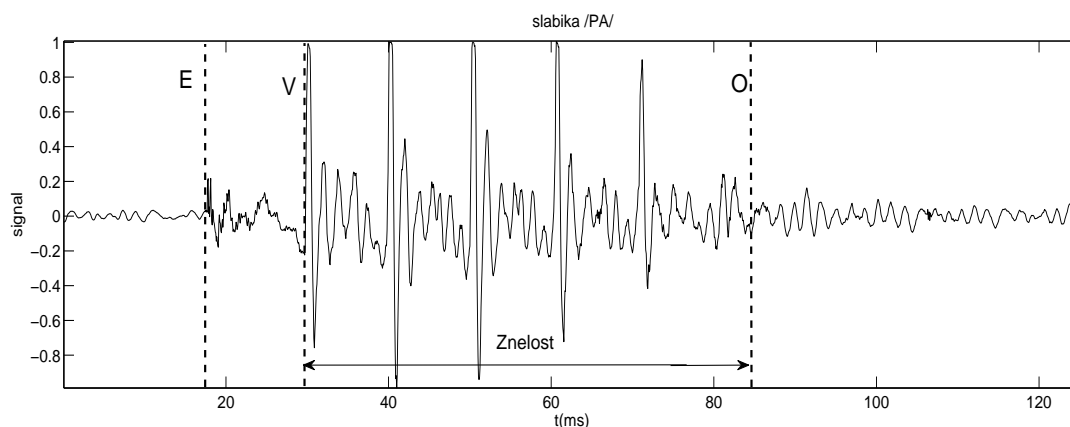
Databáze byla tvořena 80 nahrávkami získanými od 46 česky mluvících rodilých mluvčích (24 PN; 22 KS). Skupina PN byla tvořena lidmi s diagnózou raného stádia Parkinsonovy nemoci, bez předchozí medikace. Kontrolní skupina byla vytvořena tak aby věkově odpovídala skupině PN a obsahovala lidi bez neurologické a řečové patologie.

2.2 Hrubá segmentace

Tento krok řeší problém s počáteční neznalostí celkového počtu slabik v jedné promluvě tak, že promluvu rozdělí na úseky obsahující jednotlivé slabiky. Tím problém s neznámým počtem pozic v jednom signálu rozdělí na neznámý počet úloh hledání jedné od každé pozice.

Hrubá segmentace probíhá pomocí energetické obálky a je založena na předpokladu, že každý vokál obsahuje vyšší energii, a proto je možné podle polohy lokálních maxim v energetické obálce hledat přibližné pozice středů slabik.

Jeden segment získaný pomocí hrubé segmentace je zobrazen na obrázku 1. V tomto obrázku jsou také vyznačeny jednotlivé pozice E, V a O.



Obrázek 1: Označená slabika /pa/ zdravého jedince a pacienta s PN

2.3 Detekce exploze

Exploze se nachází v přední části signálu, proto prohledáváme pouze část před lokálním maximem. Zároveň je charakteristická prudkým nárůstem šumové energie, která je sice celkově nižší než energie vokálu, nicméně je vybudena v celé šíři spektra. Z toho důvodu je výhodné přistupovat k řešení ve spektrální oblasti a detekovat E ze spektrogramu.

Ke spektrogramu je přístupováno jako k matici P , která ve sloupcích obsahuje spektra jednotlivých časových úseků. Matice P je filtrována pomocí matice T která má předpis

$$\mathbf{T}_{\text{mez}}(i, j) = w_e \frac{1}{n} \sum_{j=1}^N \mathbf{P}(i, j). \quad (1)$$

Filtrace matice P pak probíhá podle přepisu

$$\mathbf{P}_f(i, j) = \begin{cases} \mathbf{P}(i, j) & \mathbf{P}(i, j) \geq \mathbf{T}_{\text{mez}} \\ 0 & \mathbf{P}(i, j) < \mathbf{T}_{\text{mez}} \end{cases}. \quad (2)$$

Z takto filtrované matice je poté vysčítáním jednotlivých sloupců získána energetická obálka signálu a vysčítáním částí sloupců obsahujících frekvence nad 1500Hz je získána energetická obálka vysokofrekvenční části signálu. Vzájemné a absolutní polohy těžiště jednotlivých energetických obálek jsou použity pro vyřazení falešných detekcí. Zároveň je použita celková energetická obálka pro přibližný odhad polohy V, pomocí které je korigována poloha prohledávané oblasti.

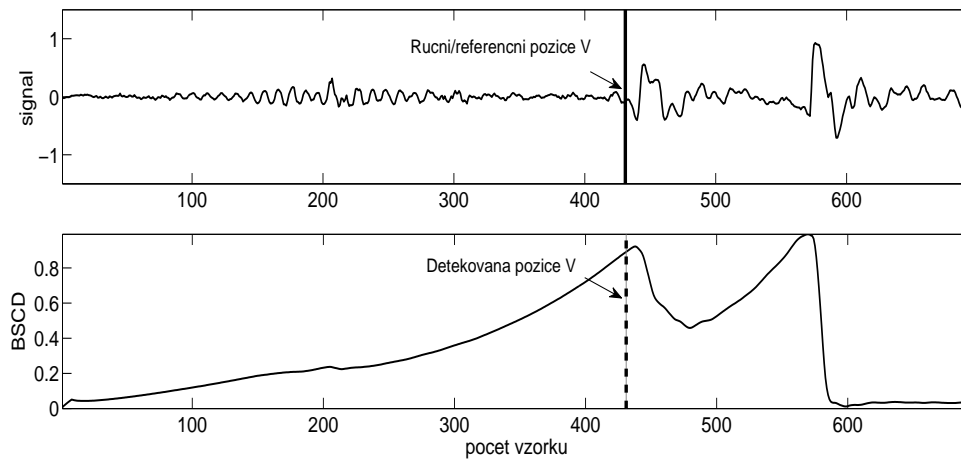
Pro detekci E je poté využita obálka signálu filtrovaného podle předpisu

$$\mathbf{P}_f(i, j) = \begin{cases} 1 & \mathbf{P}(\mathbf{i}, \mathbf{j}) \geq \mathbf{T}_{mez} \\ 0 & \mathbf{P}(\mathbf{i}, \mathbf{j}) < \mathbf{T}_{mez} \end{cases} \quad (3)$$

Takto filtrovaný signál lépe zohledňuje to, že při explozi je vybuzeno celé spektrum a snižuje vliv vyšších energií na nižších frekvencích obsažených u vokálu.

2.4 Detekce vokálu

Jak již bylo uvedeno výše, vokál obsahuje největší energii, která se projevuje zejména na nižších frekvencích a má harmonický charakter. Využitou metodou pro detekci V je Bayesovský skokový detektor (BSCD) [5], který využívá charakteru přechodu mezi šumovou explozí a znělým vokálem. Výstup z BSCD je zobrazen na obrázku 2.



Obrázek 2: Postup detekce počátku vokálu

V horní části je zobrazen signál v časové oblasti s vyznačenou pozicí V, ve spodní části je pak zobrazen výstup z BSCD. Po získání BSCD výstupu je nutné určit které lokální maximum odpovídá pozici V. Zde vycházíme z předpokladu, že základní perioda vokálu je kratší než celá explozíva. Proto vybíráme maximum následující za nejdelší mezerou. Následující periodicky se opakující peaky odpovídají jednotlivým periodám hlasivkového tónu.

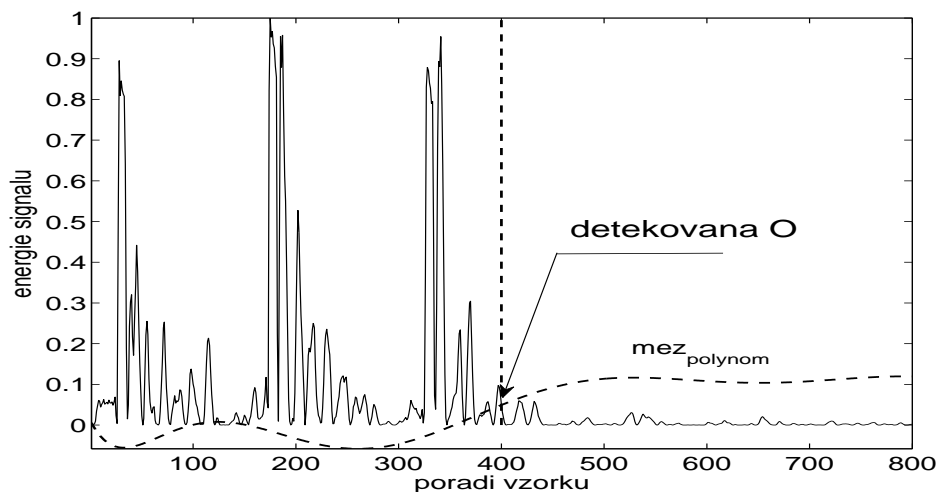
2.5 Detekce okluze

Pro detekci se jako nejúčinnější ukázala invertovaná polynomiální mez, která aproximuje energetickou obálku signálu a má tvar Výběr stupně polynomu byl prováděn pomocí automatického algoritmu, který vyhodnocuje chybu aproximace. Výsledná detekce je ilustrována na obrázku 3.

$$mez_{polynom} = \prod_{i=1}^n (a_i \mathbf{x} + b(i)) + 2\bar{x}. \quad (4)$$

2.6 Zpětná vazba

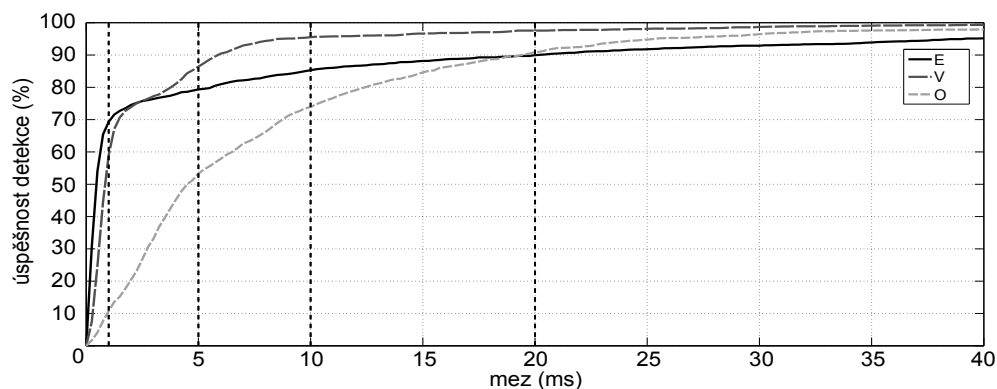
Pro zlepšení výsledků detekce byla realizována zpětná vazba, kdy byla porovnána vzdálenost mezi E a V a v případě, že nespadá do fyziologických mezí je nejprve porovnána pozice V s odhadem V z části 2.3.



Obrázek 3: Postup detekce počátku vokálu

2.7 Hodnocení úspěšnosti

Pro hodnocení úspěšnosti byly jako referenční hodnoty použity ručně označené pozice E, V a O. Na obrázku 4 je vidět ROC křivka.



Obrázek 4: ROC křivky pro detekce jednotlivých pozic E, V a O

3 Závěr

V práci byl prezentován algoritmus pro automatickou segmentaci řečové DDK úlohy. Tento algoritmus dosahoval při 5 ms mezi těchto hodnot $E_{5\text{ms}} = 79,4\%$, $V_{5\text{ms}} = 86,4\%$ a $O_{5\text{ms}} = 53,3\%$. Ovšem pro detekci O je vzhledem k jeho charakteru vypovídající spíše 10 ms mez kde $O_{10\text{ms}} = 73,9\%$. Zároveň je pro PN skupinu dosahováno výsledků $E_{5\text{ms}} = 73,6\%$, $V_{5\text{ms}} = 82,6\%$, $O_{5\text{ms}} = 46,6\%$ a $O_{10\text{ms}} = 66,6\%$. Z uvedených hodnot je vidět, že program funguje robustně i pro detekci u PN.

4 Poděkování

Tato práce je podporována z grantů SGS 12/185/OHK4/3T/13 a GAČR 102/12/2230.

Reference

- [1] Van Den Eeden, S., K., Tanner, C., M., Bernstein, A., L., Fross, R., D., Leimpeter, A., Bloch, D., A., Nelson, L., M.: Incidence of Parkinson's disease: Variation by age, gender, and race/ethnicity, *Am. J. Epidemiol.*, 157, 1015 – 1022, 2003.
- [2] Logemann, A., J., Fisher, H., B., Boshes, B., Blonsky, E., R.: Frequency and cocurrence of vocal tract disfunction in the speech of a large sample of Parkinson patients, *J. Speech Hear. Disord.*, 43, 47 – 57, 1978.
- [3] Duffy, J., R.: Motor Speech Disorders: Substrates, Differential Diagnosis and Management, 2nd ed. *Mosby, New York, NY, 2005*, pp. 1 – 592.
- [4] Rusz, J., Čmejla, R., Růžičková, H., Klempř, J., Majerová, V., Picmausová, J., Roth, J., Růžička, E.: Acoustic assesment of voice and speech disorders in Parkinson's disease through quick vocal test, *Mov. Disord.*, 26(10), 1951-1952, 2011.
- [5] Čmejla, R., Sovka, P.: Recursive Bayesian Autoregressive Changepoint Detector for Sequential Signal Segmentation, *EUSipco Proceedings, Wien(2004)*, 245 – 248.

M. Novotný
novotm26@fel.cvut.cz

J. Rusz
ruszjan@fel.cvut.cz

R. Čmejla
cmejla@fel.cvut.cz