



KATEDRA TEORIE OBVODŮ

ČVUT v Praze, Fakulta elektrotechnická

# LETNÍ DOKTORANDSKÉ DNY 2012

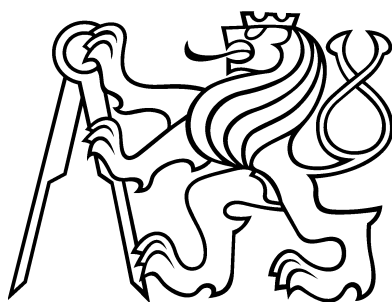
SBORNÍK PŘÍSPĚVKŮ

Editori sborníku

*Doc. Ing. Roman Čmejla, CSc.*

*Ing. Jan Rusz*

*Ing. Daniel Špulák*



---

Květen 2012

### **LETNÍ DOKTORANDSKÉ DNY 2012**

druhý ročník prezentací doktorandů Katedry teorie obvodů

24.-25. května 2012

ČVUT FEL, blok C3, posluchárna T2:C3-54

SBORNÍK PŘÍSPĚVKŮ

<http://sami.fel.cvut.cz/LDD12/>

**České vysoké učení technické v Praze**, Fakulta elektrotechnická

Katedra teorie obvodů

Technická 2

166 27 Praha 6 – Dejvice

<http://obvody.feld.cvut.cz/>

#### **Editoři sborníku:**

Doc. Ing. Roman Čmejla, CSc., Ing. Jan Ruzs, Ing. Daniel Špulák

#### **Poděkování:**

Tento sborník byl vydán za přispění grantu SVK 20/12/F3 „Letní doktorandské dny.“

Výzkum prezentovaný v jednotlivých příspěvcích je podporován mj. těmito granty a výzkumnými záměry: GACR P102/11/0649, GACR P102/11/1795, GACR 102/08/H008, GACR 102/12/2230, IGA NT11460-4/2010, MSM 0021620849, MSM 6840770012, 10--14, NT11443, NT 12288-5/2011, SGS10/179/OHK3/2T/13, SGS10/272/OHK4/3T/13, SGS12/143/OHK3/2T/13, SGS12/185/OHK4/3T/13. Děkujeme Ing. Pavlu Mášovi, PhD., za technickou pomoc při přípravě sborníku.

Vytiskla tiskárna Nakladatelství ČVUT, Zikova 4, 166 36 Praha 6, v roce 2012, nákladem 110 kusů.

**ISBN:** - - - -

# PROGRAM LDD'12 – seznam příspěvků

čtvrtek 24. května 2012 odpoledne, 12:00 - 18:15, přednášky

12:00 Zahájení LDD 2012

## téma: **TEORIE A APLIKACE SIGNÁLŮ A SYSTÉMŮ**

12:05	Ing. Ivana Nová Modelování hystereze magneticky měkkých materiálů pomocí analytických funkcí.....	5
12:30	Ing. Ondřej Kučera Supramolekulární nano-elektromechanika.....	10
12:55	Ing. Jan Janík Multispektrální transformace.....	14
13:20	Ing. Václav Turoň Vybrané adaptivní metody pro spektrální analýzu signálů.....	20
13:45	Ing. Jiří Náhlík Techniky implementace banky filtrů pomocí SC obvodů.....	24
14:10	<i>přestávka</i>	

## téma: **ZPRACOVÁNÍ EEG SIGNALŮ**

14:30	Ing. Milan Kostílek Metody single-trial klasifikace pohybového EEG.....	30
14:55	Ing. Martin Dobiáš Verifikace a úpravy klastrování výsledků ICA při zpracování EEG.....	36
15:20	Ing. Radek Janča Analýza a detekce v intrakraniálním EEG .....	44
15:45	<i>přestávka</i>	

## téma: **BIOLOGICKÉ SIGNÁLY**

16:00	Ing. Daniel Špulák Analýza povrchového EMG.....	49
16:25	Ing. Matouš Pokorný Kompaktní senzorický systém pro sběr biologických signálů.....	53
16:50	Ing. Jakub Parák Klasifikácia život ohrozujúcich stavov .....	57
17:15	Ing. Miroslav Ložek Návrh mechanického modelu kardiovaskulárního systému.....	62

**pátek 25. května 2012 odpoledne, 13:00 - 18:00, přednášky**

téma: **ZPRACOVÁNÍ ŘEČI**

13:00	Ing. Michal Borský Adaptace akustických modelů trifónů pro rozpoznávání v hlučném prostředí.....	68
13:25	Ing. Robert Krejčí Rozpoznávání řeči pomocí signálových procesorů.....	75
13:50	Ing. Jan Bartošek Prozodické modelování.....	78
14:15	<i>přestávka</i>	

téma: **PATOLOGICKÁ ŘEČ**

14:30	Ing. Pavel Grill Odhad formantů v patologické řeči.....	82
14:55	Ing. Lukáš Bauer Odhad FO v patologické řeči.....	86
15:20	Ing. Michal Novotný Detekce změn v patologické řeči.....	92
15:45	<i>přestávka</i>	
16:00	Ing. Jan Janda Objektivizace logopedického věku dítěte.....	98
16:25	Ing. Martina Nejepsová Objektivizace vývojové dysfázie.....	102
16:50	Ing. Tomáš Lustyk Objektivizace koktavosti.....	108
17:15	Ing. Adam Stráník Objektivizace chraptivosti.....	113

# Modelování hystereze magneticky měkkých materiálů pomocí analytických funkcí

Ing. Ivana NOVÁ<sup>1</sup> (školitel: Prof. Ing. Ivan ZEMÁNEK, CSc.<sup>2</sup>)

<sup>1</sup> ČVUT v Praze, Fakulta elektrotechnická, Katedra elektrických pohonů a trakce

<sup>2</sup> ČVUT v Praze, Fakulta elektrotechnická, Katedra teorie obvodů

[nova@fel.cvut.cz](mailto:nova@fel.cvut.cz)

**Abstrakt:** Magneticky měkké materiály patří mezi nejdůležitější konstrukční prvky používané v elektrotechnice. Pro zlepšení procesu měření jejich dynamických vlastností může být přínosem náhrada chování reálného materiálu modelem, který by měl sloužit při realizaci rychlých měřicích algoritmů.

V současnosti existují dva hlavní přístupy – jednak Preisachův dynamický model, a jednak některé modely diferenciální (např. model Hodgdonův, Jiles-Athertonův, Chuův, Wlodarského – všechny v zásadě založené na řešení nelineárních diferenciálních rovnic). Potíže mohou spočívat v nesnadné implementaci modelu, ve zdlouhavosti výpočtu nebo v komplikované identifikaci parametrů modelu.

Jednodušší způsob popisu představuje třída analytických aproximačních modelů, které nahrazují skutečný tvar hysterezních křivek vhodnými matematickými funkcemi a neobsahují diferenciální rovnice. Většina existujících analytických modelů funguje uspokojivě jen při pomalém magnetování. Za předpokladu střídavého magnetování bez stejnosměrné složky lze úlohu zjednodušit, neboť postačí modelovat symetrické hysterezní smyčky. Navíc se lze omezit na 1D modelování, protože i u anizotropních materiálů se předpokládá měření jen ve směru osy snadné magnetizace.

Zásadní je nalezení optimálních parametrů modelů. Protože parametry ve většině případů nebudou konstantní, je třeba zjistit jejich průběhy a prostudovat jejich chování s ohledem na amplitudu a frekvenci budícího signálu. Tato práce řeší jen buzení sinusovým signálem, který se v praxi vyskytuje nejčastěji. Za předpokladu předvídatelného chování parametrů bude jejich průběhy možno vyjádřit vhodnými analytickými funkcemi, příp. vyhledávací tabulkou. V řadě případů bude možné některé složky modelu zanedbat, čímž se situace zjednoduší.

Byly navrženy tři modely s využitím jednoduchých analytických funkcí a s flexibilními parametry – *analytický přímý model*, *analytický inverzní model* a *rozšířený hyperbolický model*. Dynamická hysterezní smyčka je konstruována ze tří až čtyř složek, které s využitím šesti až osmi parametrů nahrazují průběh dynamické hysterezní smyčky od tvaru sigmoidního až po tvar eliptický. Velikost parametrů podle předpokladu není konstantní, ale mění se s frekvencí a amplitudou buzení. Byl hledán způsob analytického vyjádření těchto parametrů. Nejlépe se osvědčily lomené racionální funkce resp. polynomy, a proto konečný model používá odhad parametrů získaný pomocí zmíněných funkcí.

Modely byly ověřeny na vzorcích orientovaného trafoplechu a na materiálu typu Fe-Ni. Kvalita aproximace byla testována se zřetelem na střední kvadratickou odchylku mezi naměřenou a simulovanou hodnotou  $B$  (přímé modely) resp.  $H$  (inverzní model) a dále byly porovnány naměřené a modelované hodnoty měrných ztrát. Porovnání mezi naměřenými a simulovanými průběhy ukazuje poměrně dobrou shodu dynamických hysterezních smyček, odchylky se pohybují převážně v jednotkách procent.

## Modelování hysterese magneticky měkkých materiálů pomocí analytických funkcí

Ing. Ivana Nová



České vysoké učení technické  
Fakulta elektrotechnická  
Katedra teorie obvodů  
Školitel: Prof. Ing. Ivan Zemánek, CSc.

## Záměr práce

- Matematický model složený z analytických funkcí
- Přímý model  $B = f(H)$  + Inverzní model  $H = f(B)$
- Vhodný pro střídavé magnetování (mg. měkké materiály)
  - Běžný rozsah amplitud budícího signálu a frekvencí
  - Symetrické smyčky
  - 1D modelování
  - SIN buzení
- Zlepšení vlastností ferometru KF9a
  - Jednoduchý predikční algoritmus
  - Nastavení optimálních parametrů
  - Zrychlení měřicího procesu

Ing. Ivana Nová



2

## Dynamické hysterese smyčky – modely

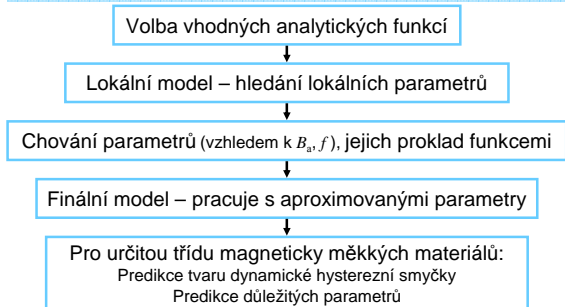
- Dynamický Preisachův model
- Diferenciální modely
  - Jiles-Atherton
  - Chua
  - Wlodarski
  - Hodgdon, Duhem, ...
- Většina uvedených modelů má též inverzní variantu
  - Ale: jsou poměrně komplikované nebo časově náročné
- Modely s analytickou aproximací bez difer. rovnic
  - Různé matematické funkce
  - Ale: obvykle fungují při pomalém magnetování

Ing. Ivana Nová



3

## Schéma návrhu modelu



Ing. Ivana Nová



4

## Inverzní analytický model – volba funkcí

- Tři složky:  $H_{\pm} = h_1(B) + h_2(B) + h_3(B)$

$$1) \quad h_1 = a_1 \cdot \text{sign}(B \mp b_1) \cdot \left( \tan\left(\frac{\pi}{2} \cdot \frac{B \mp b_1}{B_{\max}}\right) \right)^2 \pm a_0$$

$$2) \quad h_2 = \pm a_2 \cdot \left( 1 - \left( \frac{B}{B_a} \right)^2 \right)^k$$

$$3) \quad h_3 = a_3 \cdot B$$

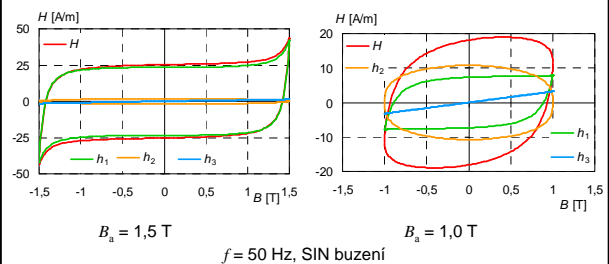
Šest parametrů:  
 $a_1, a_2, a_3, b_1, B_{\max}, k$   
 $a_0 \rightarrow H_{\pm}(B_a) = H_{\mp}(B_a)$

Ing. Ivana Nová



5

## Inverzní model – příklady simulace



Ing. Ivana Nová



6

### Přímý analytický model – volba funkcí

- Čtyři složky:  $B_{\pm} = g_1(H) + g_2(H) + g_3(H) + g_4(H)$

1)  $g_1 = a_1 \cdot \frac{2}{\pi} \text{sign}[\alpha(H \mp H_c)] \cdot \text{atan}\left[\left|\frac{\alpha(H \mp H_c)}{H_c}\right|^k\right]$

2)  $g_2 = \mp a_2 \cdot \left(1 - \left(\frac{H}{H_a}\right)^{k_1}\right)$

Sedm parametrů:  
 $a_1, a_2, a_3, a_4, \alpha, k_1, k_2$

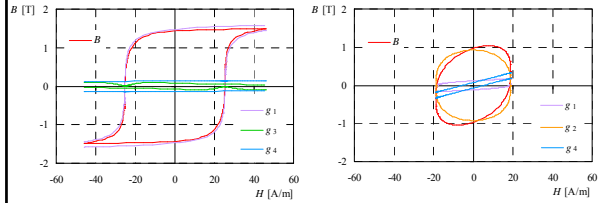
3)  $g_3 = \mp a_3 \cdot k_2 \left(\frac{H}{H_c} \mp 1\right) \cdot \left(1 + \left[k_2 \left(\frac{H}{H_c} \mp 1\right)\right]^{-1}\right)$

4)  $g_4 = a_4 \cdot H + b_0$

$b_0 \rightarrow B_{\pm}(H_a) = B_{\mp}(H_a)$



### Přímý model – příklady simulace



$B_a = 1,5 \text{ T}$

$f = 50 \text{ Hz, SIN buzení}$

$B_a = 1,0 \text{ T}$



### Modifikovaný hyperbolický model

- Tři složky:  $B_{\pm} = g_1(H) + g_2(H) + g_0(H)$

1)  $g_1 = A_1 \cdot \tanh[\alpha_1(H \mp a_{01})] \pm b_1$

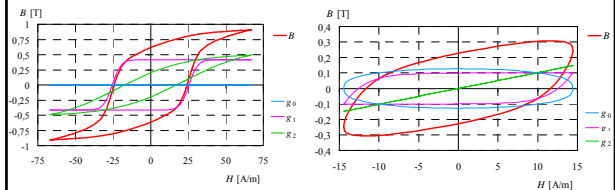
2)  $g_2 = A_2 \cdot \tanh[\alpha_2(H \mp a_{02})] \pm b_2$

3)  $g_3 = \mp A_0 \cdot \left(1 - \left(\frac{H}{H_a}\right)^k\right)$

Osm parametrů:  
 $A_1, A_2, \alpha_1, \alpha_2, a_{01}, a_{02}, A_0, k$   
 $b_1, b_2 \rightarrow B_{\pm}(H_a) = B_{\mp}(H_a)$



### Hyperbolický model – příklady simulace



$B_a = 0,9 \text{ T}$

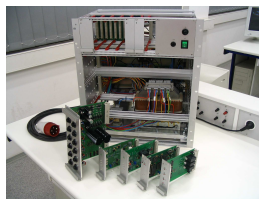
$f = 50 \text{ Hz, SIN buzení}$

$B_a = 0,3 \text{ T}$

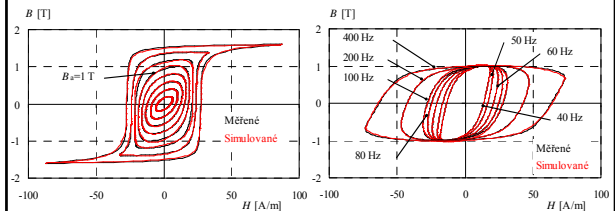


### Testování modelů

- Klasický materiál – orientovaný trafoplech Eo10
  - $f = 50 \text{ Hz, } 0,1 \text{ T} \leq B_a \leq 1,7 \text{ T}$  (krok 0,1 T)
  - $40 \text{ Hz} \leq f \leq 400 \text{ Hz, } B_a = 0,5 \text{ T} - 1,0 \text{ T} - 1,5 \text{ T} - 1,7 \text{ T}$
- Materiál typu FeNi42
  - $f = 50 \text{ Hz, } 0,2 \text{ T} \leq B_a \leq 1,2 \text{ T}$
- Měřicí zařízení: KF9a
  - Kompenzační ferometr
  - Otevřený vzorek (tabule)
  - 200 resp. 128 měřících bodů na smyčku
- SIN magnetování



### Přímý model – příklad lokální simulace

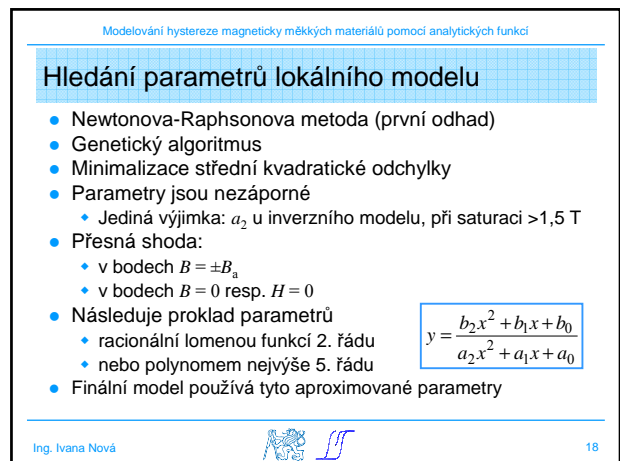
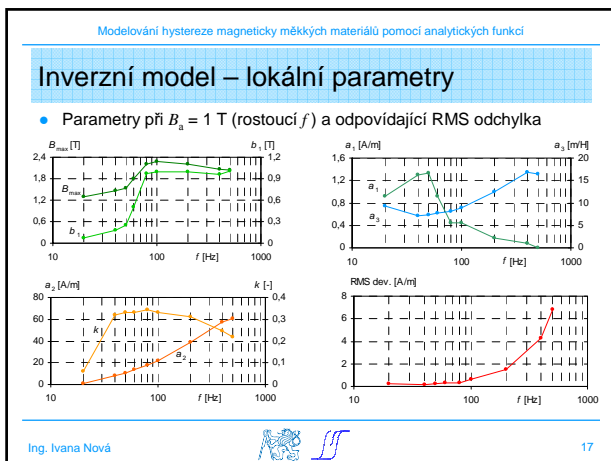
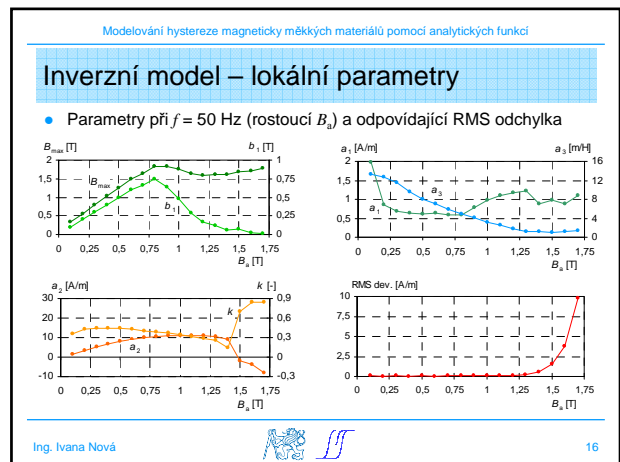
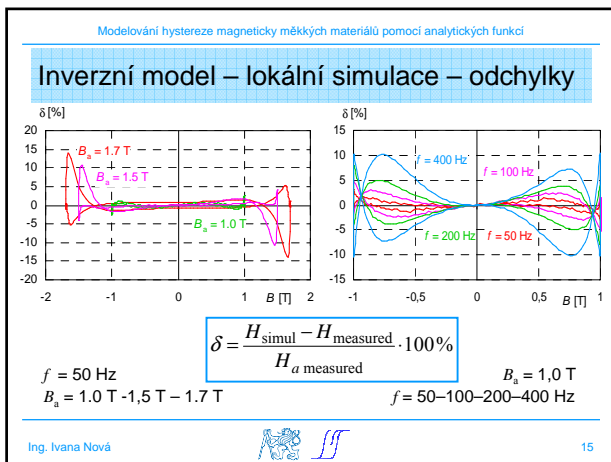
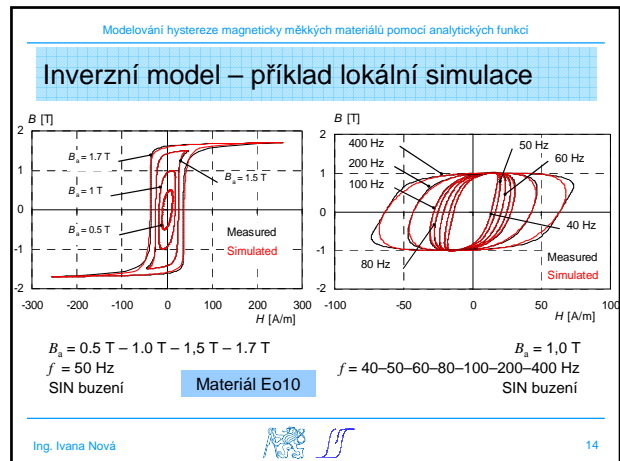
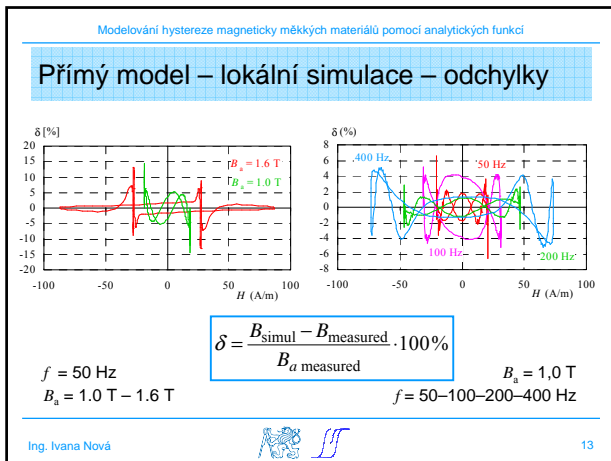


$B_a = 0,2 \text{ T} - 1,6 \text{ T}$  (krok 0,2 T)  
 $f = 50 \text{ Hz}$   
SIN buzení

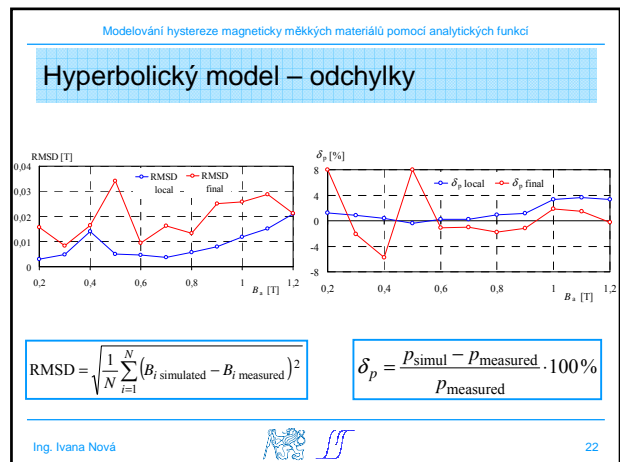
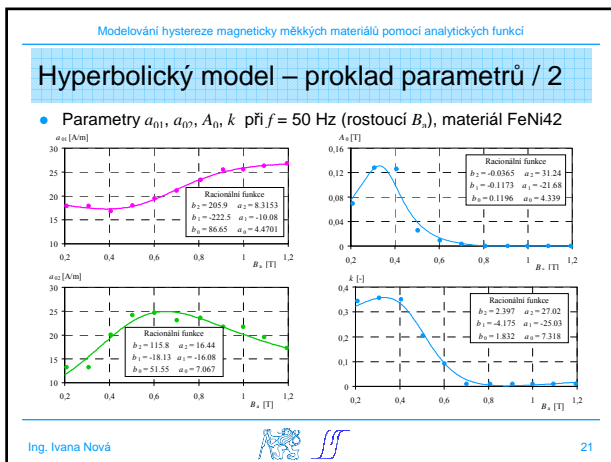
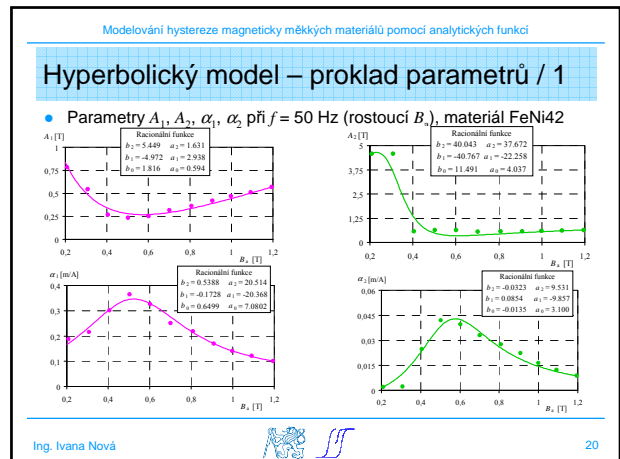
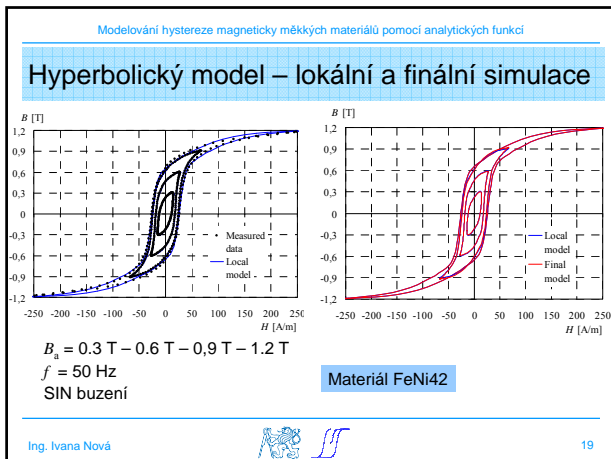
Materiál Eo10

$B_a = 1,0 \text{ T}$   
 $f = 40 - 50 - 60 - 80 - 100 - 200 - 400 \text{ Hz}$   
SIN buzení









Modelování hysterese magneticky měkkých materiálů pomocí analytických funkcí

### Závěry

- Věrohodná simulace dynamických hysteresečních smyček
- Parametry jsou předvídatelné, lze proložit funkcemi
- Přesnost simulace:
  - Inverzní model lokální:  $\text{RMSD} \leq 9,7 \text{ A/m}$ ,  $|\delta_p| \leq 2,8 \%$   
 finální:  $\text{RMSD} \leq 13 \text{ A/m}$ ,  $|\delta_p| \leq 34 \%$
  - Přímý model lokální:  $\text{RMSD} \leq 0,29 \text{ T}$ ,  $|\delta_p| \leq 4,5 \%$   
 finální:  $\text{RMSD} \leq 0,34 \text{ T}$ ,  $|\delta_p| \leq 27 \%$
  - Hyperbolický model lokální:  $\text{RMSD} \leq 0,022 \text{ T}$ ,  $|\delta_p| \leq 4 \%$   
 FeNi42 finální:  $\text{RMSD} \leq 0,035 \text{ T}$ ,  $|\delta_p| \leq 8 \%$
- Citlivost na parametry
- Ve kterých případech je kvalita simulace slabší

Ing. Ivana Nová 23

Modelování hysterese magneticky měkkých materiálů pomocí analytických funkcí

### Literatura

- Iványi, A.: *Hysteresis Models in Electromagnetic Computation*. Akadémiai Kiadó, Budapest, 1997.
- Mayergoyz, I. D.: *Mathematical Models of Hysteresis and Their Applications*. 2<sup>nd</sup> edition. Academic Press, New York, 2003.
- Wlodarski, Z.: Analytical Description of Magnetization Curves. *Physica B*. 2006, vol. 373, no. 2, p. 323–327.
- Takács, J.: *Mathematics of Hysteretic Phenomena*. Wiley-VCH, Berlin, 2003.
- Zemánek, I. – Nová, I. Compensation Ferrometer KF9a – Universal Single Sheet/Strip and On-line Tester for Laboratory and Industrial Use. In *Proceedings of WMM'08*. Ghent, 2008, p. 413–429.
- Zemánek, I. Výzkumná zpráva o měření magnetických parametrů materiálů FeNi42 – č. t. 3986, a FeNi47Cu5 – č. t. 3990, vyrobených v a.s. Kovohutě Rokycany. Praha: ČVUT v Praze, Fakulta elektrotechnická, Katedra teorie obvodů, 2012.
- Nová, I. – Zemánek, I. Analytical model with flexible parameters for dynamic hysteresis loops modelling. *Journal of Electrical Engineering*. 2010, vol. 61, no. 7/s, p. 46–49.
- Nová, I. – Zemánek, I. Analytical inverse model with flexible parameters for dynamic hysteresis loops modelling. *Przegląd Elektrotechniczny*. 2011, vol. 87, no. 9b, p. 89–92.

Ing. Ivana Nová 24

# Supramolekulární nano-elektromechanika

Ondřej KUČERA<sup>1</sup>

<sup>1</sup> České vysoké učení v Praze, Fakulta elektrotechnická, Katedra teorie obvodů

[kucerao@ufe.cz](mailto:kucerao@ufe.cz)

*Abstrakt:* Elektrické a mechanické vlastnosti proteinů hrají důležitou roli v jejich funkčních vlastnostech. V tomto příspěvku diskutujeme význam elektromechanické vazby u vyšších supramolekulárních struktur ve vztahu k jejich endogenním vibračním módům. Tuto myšlenku demonstrujeme zejména na mikrotubulech.

Mikrotubuly jsou elektricky polární struktury cytoskeletu, které splňují předpoklady pro generaci elektromechanických oscilujících v oblasti od kHz do GHz. Přívod energie pro elasto-elektrické vibrace v mikrotubulech může být zajištěn hydrolýzou GTP, interakcí motorových proteinů s mikrotubulem a tokem energie z mitochondrií, nebo také přenosem jiných intracelulárních vibračních signálů, jejichž stručný přehled je v příspěvku zmíněn.

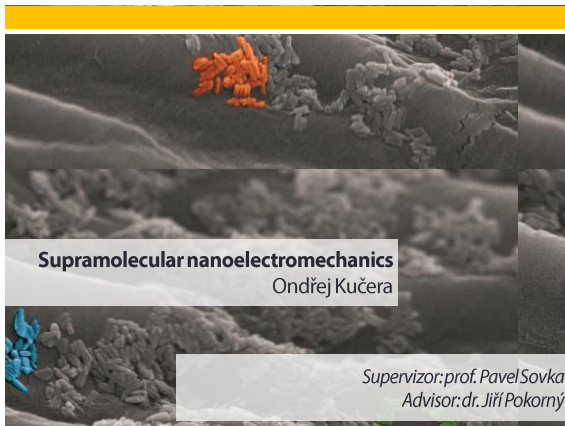
Elektrické pole generované axiálními longitudinálními vibračními módy mikrotubulu simulujeme pomocí MRDNA metody pro různé konformace mikrotubulárních sítí, ať už rostoucí mikrotubuly, pravidelnou síť, nebo mitotické vřeténko. Výsledkem simulací jsou mapy intenzity elektrického pole a odhad vyzářeného výkonu.

Fyziologický význam elektromechanických oscilací, resp. jejich elektrické složky, může spočívat v silových účincích tohoto pole. Ty mohou působit jednak pohybem hmoty, nebo také přenosem informace. Z hlediska chemických procesů v buňce uvažujeme zejména přenos reakčních komponent na místo reakce, lokální pohyb pro dosažení vhodné polohy komponent a pohyb náboje nutný pro vytvoření chemické vazby. Výsledky modelu naznačují, že tyto efekty jsou významné v případě rezonanční interakce.

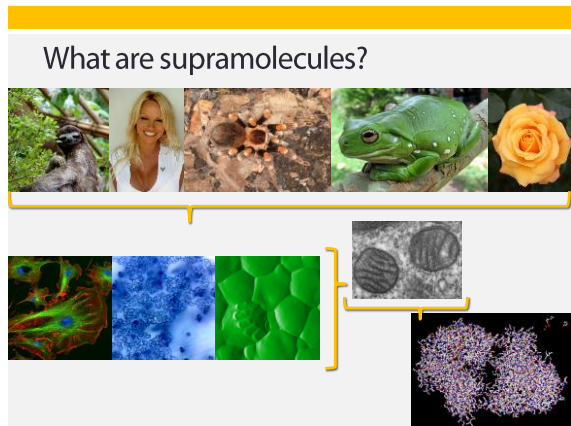
## Supramolecular nanoelectromechanics

Ondřej Kučera

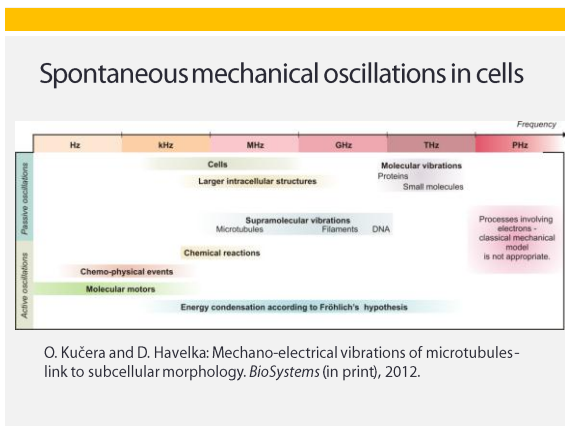
Supervisor: prof. Pavel Sovka  
Advisor: dr. Jiří Pokorný



## What are supramolecules?

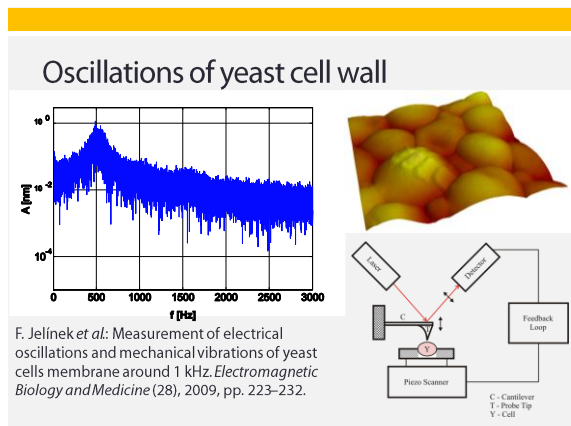


## Spontaneous mechanical oscillations in cells



O. Kučera and D. Havelka: Mechano-electrical vibrations of microtubules-link to subcellular morphology. *BioSystems* (in print), 2012.

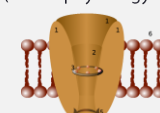
## Oscillations of yeast cell wall



F. Jelínek *et al.*: Measurement of electrical oscillations and mechanical vibrations of yeast cells membrane around 1 kHz. *Electromagnetic Biology and Medicine* (28), 2009, pp. 223–232.

## Mechanisms of generation of EM field

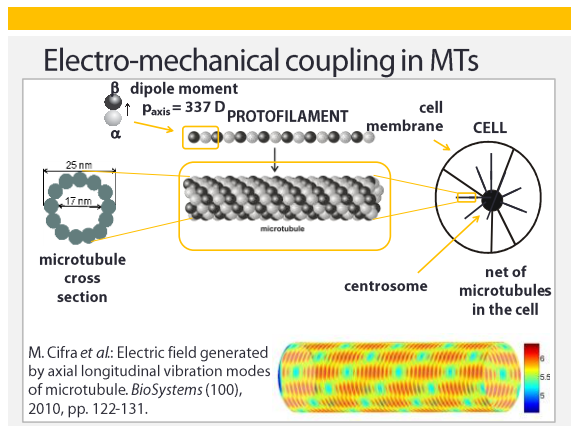
- Flow of ions – ion channels (electrophysiology – ECG, EEG...)
- Frequency up to 200 Hz



- Chemoluminescence – the product of chemical reaction is in excited state and emits light

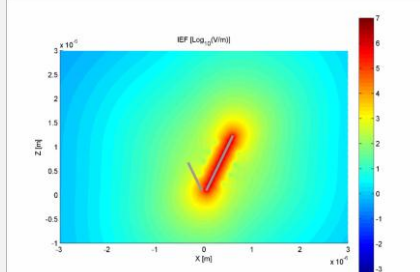
$$[A] + [B] \rightarrow [C] + hf$$

## Electro-mechanical coupling in MTs



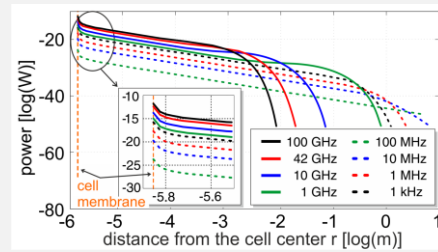
M. Cifra *et al.*: Electric field generated by axial longitudinal vibration modes of microtubule. *BioSystems* (100), 2010, pp. 122-131.

## Oscillating electric field generated by MTs



O. Kučera and D. Havelka: Mechano-electrical vibrations of microtubules - link to subcellular morphology. *BioSystems* (in print), 2012.

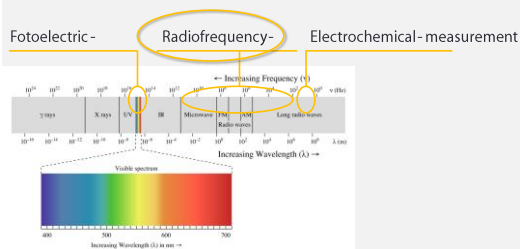
## EM power radiated from a cell



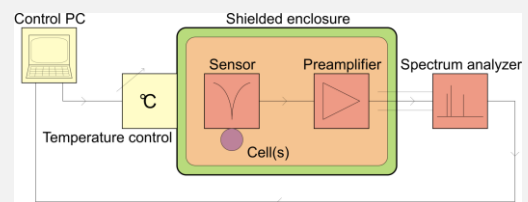
D. Havelka *et al.*: High-frequency electric field and radiation characteristics of cellular microtubule network. *Journal of Theoretical Biology* (286) 2011, pp 31-40.

## Measurement of EM field of living cells

- The principle of detection depends upon frequency.



## Measurement of RF - EM field of living cells



O. Kučera *et al.*: Technical aspects of measurement of cellular electromagnetic activity. *European Biophysics Journal* (39), 2010, pp. 1465-1470.

## Implications for morphogenesis

- Force acting on dipole

$$F_d = \nabla(\vec{p} \cdot \vec{E})$$

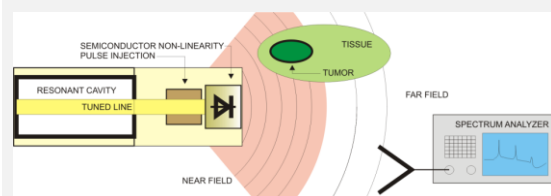
- Dielectrophoretic force

$$F_\epsilon \approx sV(\epsilon_p - \epsilon_m)\nabla E^2$$

- Electrorotational torque

$$\vec{\tau} = \vec{p} \times \vec{E}$$

## Implications for cancer diagnostics



J. Pokorný *et al.*: Cancer physics: diagnostics based on damped cellular elastoelectrical vibrations in microtubules. *European Biophysics Journal* (40), 2011, pp. 747-759.

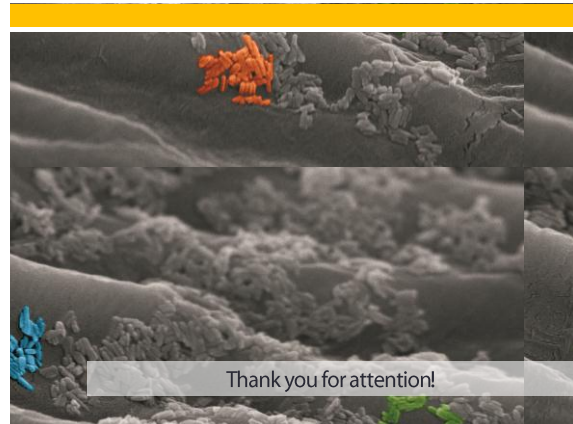
## Acknowledgements

### Funding

- Czech Science Foundation, GA CR, grants nos. P102/11/0649 and 102/08/H008.
- Grant Agency of the Czech Technical University in Prague, grant no. SGS10/179/OHK3/2T/13

### Collaboration

- Institute of Photonics and Electronics, AS CR (dr. M. Cifra)
- Institute of Microbiology, AS CR (dr. J. Hašek, prof. Binarová)
- Czech Technical University in Prague, Dept. of Electromagnetic Field (D. Havelka)



# Cesta k Aproximovanej Diskrétnej Zolotarev Kosínusovej Transformácii

Ján JANÍK<sup>1</sup>, Pavel SOVKA<sup>1</sup>, Miroslav VLČEK<sup>2</sup>

<sup>1</sup> České vysoké učení v Praze, Fakulta elektrotechnická, Katedra teorie obvodů

<sup>2</sup> České vysoké učení v Praze, Fakulta dopravní, Ústav aplikované matematiky

[janikjan@fel.cvut.cz](mailto:janikjan@fel.cvut.cz), [sovka@fel.cvut.cz](mailto:sovka@fel.cvut.cz), [vlcek@fd.cvut.cz](mailto:vlcek@fd.cvut.cz)

**Abstrakt:** V tomto článku predstavujeme novú Aproximovanú Diskrétnu Zolotarev Kosínusovú Transformáciu (ADZCT), ktorej zmyslom je vylepšiť vlastnosti Diskrétnej Kosínusovej Transformácie (DCT). Zaoberáme sa vývojom transformácie a problémami, na ktoré sme počas vývoja narazili.

## 1. Úvod

Naša výskumná skupina sa zaoberá vývojom nových časovo-frekvenčných selektívnych transformácií založených na Zolotarevových polynómoch už niekoľko rokov. Zolotarevove polynómy zobecňujú Čebyševove polynómy a účelom použitia týchto polynómov je vylepšiť vlastnosti transformácií založených práve na Čebyševových polynómoch. Pred nedávnou dobou sme predstavili novú Diskrétnu Zolotarevovu Transformáciu (DZT) a jej aproximovanú formu (ADZT) [1], [2], [3]. Táto nová transformácia zlepšila vlastnosti Diskrétnej Fourierovej Transformácii (DFT) hlavne v oblasti detekcii nestacionárnych signálov.

Výsledky ADZT transformácii nás priviedli k myšlienke vytvoriť obdobným spôsobom Diskrétnu Zolotarev Kosínusovú Transformáciu (DZCT). DZCT by rozšírila schopnosti Diskrétnej Kosínusovej Transformácie výmenou jej bazových funkcií za selektívne kosínusy, resp. Zolotarevove polynómy.

Prvý krát sa myšlienka o DZCT objavila v publikácii [4]. V tomto článku sa budeme zaoberať jednotlivými krokmi vývoja a o prechode k aproximovanej forme DZCT. Spomenieme jednotlivé problémy, s ktorými sme sa počas vývoja stretli.

## 2. Diskrétna Kosínusová Transformácia

V prvom kroku je vhodné predstaviť samotnú Diskrétnu Kosínusovú Transformáciu [5], ktorá tvorí základ pre Diskrétnu Zolotarev Kosínusovú Transformáciu. DCT je známa hlavne výbornými výsledkami v oblasti spracovania a komprimácie obrázkov [6] a v oblasti spracovania rečových signálov [7].

Základom DCT je expanzia signálu na párný signál. Podľa zvoleného typu expanzie poznáme štyri základné definície DCT. Pre naše účely sme vybrali najznámejšiu a najpoužívanejšiu definíciu označovanú DCT-II (1), ktorá bola prvý krát publikovaná v roku 1974 [8].

DCT-II expanduje analyzovanú postupnosť  $x[n]$  dĺžky  $N$  na párnú postupnosť dĺžky  $2N$  a rozkladá ju do takzvaných párných a nepárných kosínusov. Termín párný a nepárny kosínus sa zrejme ešte v literatúre neobjavili, pretože pre samotnú DCT je irelevantný. Prvý krát sme túto definíciu použili v [4] a poukázali na ich význam pre multi-spektrálny rozklad signálu. Vďaka nepárnym kosínusom dokáže DCT rozkladať nepárny signál rovnako dobre

ako páry. Pre urýchlenie výpočtu DCT spektra bol navrhnutý rýchly algoritmus využívajúci rýchlu Fourierovú transformáciu [9], [10].

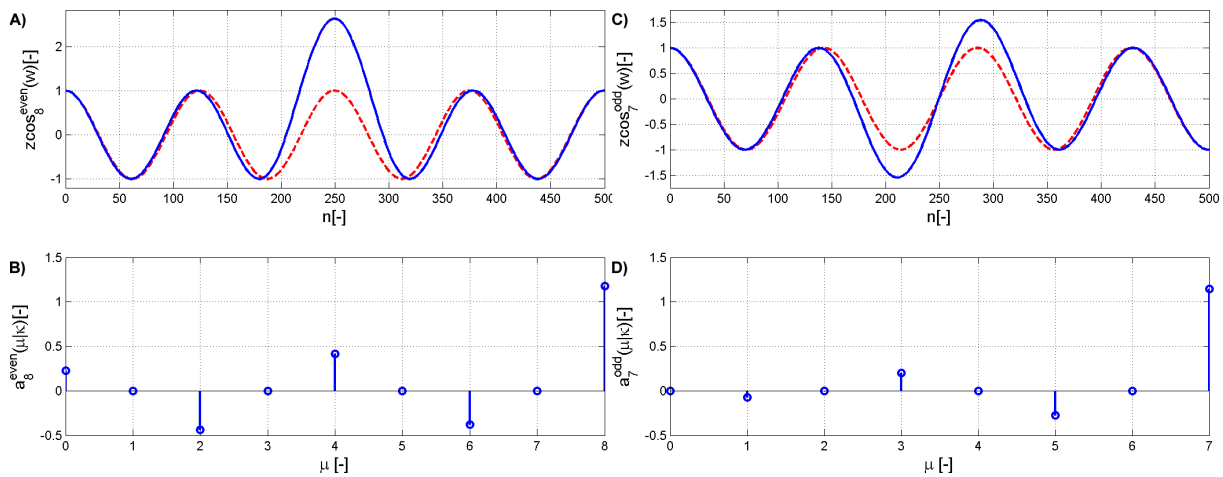
$$C(k) = \varepsilon(k) \sum_{n=0}^{N-1} x[n] \cos\left(\frac{2n+1}{2N} k\pi\right)$$

$$\varepsilon(k) = \begin{cases} \sqrt{\frac{1}{N}} & \dots k = 0 \\ \sqrt{\frac{2}{N}} & \dots k \neq 0 \end{cases} \quad (1)$$

$$k = 1, 2, \dots, N-1$$

### 3. Selektívne kosínusy a báza

Rozdiely medzi párnymi a nepárnymi kosínusmi sa prejavujú hlavne pri ich selektívnych formách. Selektívny kosínus (zcos) disponuje špeciálnou vlastnosťou zväčšovať centrálny lalok, alebo laloky v prípade nepárnej formy, a odsunúť nuly k okrajom intervalu. Z toho dôvodu sú obe zcos určené stupňom  $p$  a novým parametrom  $\kappa$  nazývaným selektivita a definujúcim veľkosť centrálnych lalokov.



**Obrázok 1:** A) párný kosínus (čiarkovaný) a selektívny kosínus (plný). B) selektívne koeficienty  $a_{2\mu}(\kappa)$  párneho zcos. C) nepárny kosínus (čiarkovaný) a jeho selektívny kosínus (plný). D) selektívne koeficienty  $a_{2\mu+1}(\kappa)$  nepárneho zcos.

#### 3.1. Párný selektívny kosínus

Párný selektívny kosínus, zobrazený na obrázku 1A, je párný, rovnomerne zvlnený, Zolotarevov polynóm. Lokálne maximum tohto zcos sa nachádza v strede symetrie a preto obsahuje jeden centrálny lalok. Každý párný selektívny kosínus stupňa  $p$  sa skladá zo všetkých štandardných párných kosínusov nižších stupňov a stupňa  $p$ . Preto párný zcos môžeme vyjadriť ako váženú sumu (2), kde  $a$  sú váhové koeficienty vypočítané pomocou rekurzívneho algoritmu vyvinutého M. Vlčkom [11]. Koeficient  $a$  s najvyšším indexom reprezentuje stacionárnu časť, zatiaľ čo zvyšok koeficientov reprezentuje adaptívnu, nestacionárnu časť, ako je zobrazené na obrázku 1B.

$$z\cos_p^{even}(n|\kappa) = \sum_{\mu=0}^{p/2} a_{2\mu}(\kappa) \cos\left(\frac{2n+1}{2N} 2\mu\pi\right) \quad (2)$$

### 3.2. Nepárny selektívny kosínus

Nepárny selektívny kosínus je zobrazený na obrázku 1C. Je to taktiež rovnomerne zvlhnený Zolotarevov polynóm, ale nepárneho typu. To znamená, že v strede symetrie sa nachádza nulový bod a preto obsahuje dva centrálné laloky. Každý nepárny selektívny kosínus stupňa  $p$  je opäť zložený z nepárnych štandardných kosínusov všetkých nižších stupňov a stupňa  $p$  a preto môže byť identicky vyjadrený ako vážená suma (3). Stacionárna a nestacionárna časť je taktiež definovaná analogicky ako u párneho kosínusu.

$$z\cos_p^{odd}(n|\kappa) = \sum_{\mu=0}^{(p-1)/2} a_{2\mu+1}(\kappa) \cos\left(\frac{2n+1}{2N} (2\mu+1)\pi\right) \quad (3)$$

### 3.3. Selektívna báza

Oba typy selektívnych kosínusov tvoria selektívnu bázu DZCT transformácie. DZCT báza bola vytvorená zámennou štandardných kosínusov DCT bázy za selektívne s využitím vlastností selektívnych kosínusov, ktoré môžeme vyjadriť ako váženú sumu štandardných kosínusov. Ak maticu  $\mathbf{L}$  je spodnú trojuholníkovú maticu všetkých selektívnych koeficientov  $a$  a  $\mathbf{W}$  je riadková matica obsahujúca DCT bázu, tak matica obsahujúca selektívnu bázu  $\mathbf{W}_Z$  stupňa 3 môže byť vyjadrená nasledovne

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ a_{2,0} & 0 & a_{2,2} & 0 \\ 0 & a_{3,1} & 0 & a_{3,3} \end{bmatrix}, \quad (4)$$

$$\mathbf{W}_Z = \mathbf{L} \cdot \mathbf{W}.$$

## 4. Diskrétna Zolotarev Kosínusová Transformácia

Aplikovaním selektívnej bázy  $\mathbf{W}_Z$  na signál získame DZCT spektrum  $\mathbf{C}_Z$ . DZCT spektrum  $\mathbf{C}_Z$  môže byť vypočítané buď priamo pomocou matice  $\mathbf{W}_Z$  alebo aplikovaním matice  $\mathbf{L}$  na DCT spektrum  $\mathbf{C}$

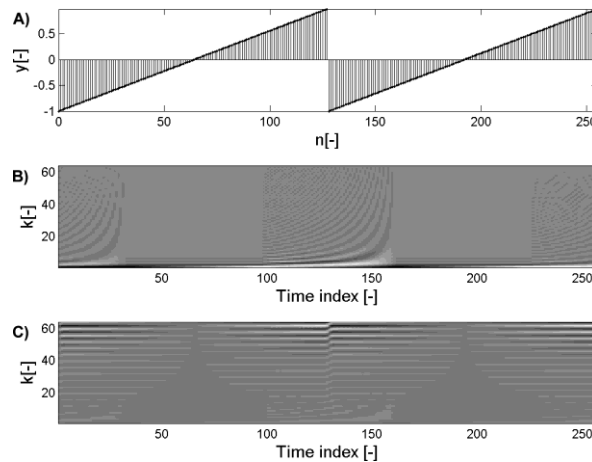
$$\mathbf{C}_Z = \mathbf{W}_Z \cdot \mathbf{S} = \mathbf{L} \cdot \mathbf{W} \cdot \mathbf{S} = \mathbf{L} \cdot \mathbf{C}. \quad (5)$$

Stupeň  $p$  narastá rovnako ako v prípade DCT, od 0 po  $N-1$ . Priamy výpočet z DCT spektra  $\mathbf{C}$  môže byť chápaný ako reorganizácia alebo filtrácia DCT spektra. Otázkou a zároveň problémom v tomto kroku je možnosť zostavenia matice  $\mathbf{L}$  bez predošlej znalosti parametru selektivity  $\kappa$ .

Príklad výpočtu takéhoto spektra je znázornený na obrázku 2 vo forme spektrogramu. DZCT spektrogram bol vypočítaný s konštantnou selektivitou  $\kappa$  počas celej realizácie spektrogramu, ktorá bola vybraná experimentálne pre daný signál. Je evidentné, že tento spektrogram nám dáva menej informácie o signály než DCT spektrogram. Strata informácie

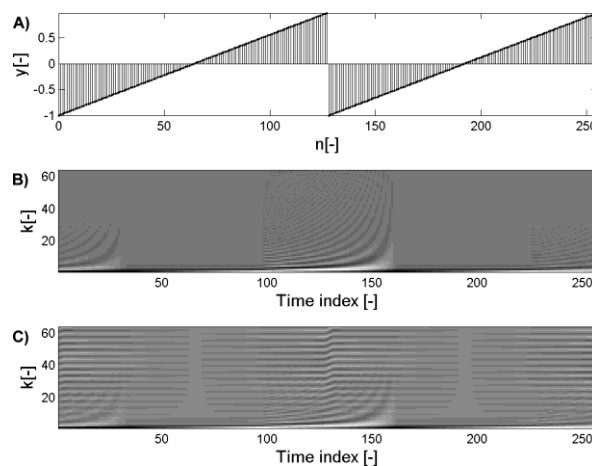


je zapríčinená narastaním energie selektívnych kosínusov úmerne s nárastom ich selektivity  $\kappa$  a stupňa  $p$ . Pretože  $z\cos$  sa skladá zo všetkých kosínusov nižších stupňov, obrázok 1, tak  $z\cos$  akumuluje energiu všetkých nižších spektrálnych zložiek do centrálneho laloku a tým narastá jeho celková energia.



**Obrázok 2:** A) vstupný signál  $y[n]$ . B) DCT spektrogram signálu  $y[n]$ . C) Nenormovaný DZCT spektrogram signálu  $y[n]$ .

Obrázok 3 zobrazuje rovnakú realizáciu spektrogramu, ale DZCT spektrálne zložky  $C_z$  sú normované energiou príslušných  $z\cos$ . Táto normalizácia výrazne zlepšila výsledok DZCT spektrogramu, ktorý lokalizuje nestacionárne oblasti signálu (okolo indexu 128). Účelom normovania je ponechanie konštantnej energie pre všetky bázové funkcie, rovnako ako u DCT alebo DFT transformácií. Pre toto normovanie nemáme zatiaľ žiaden matematický podklad, nakoľko je potrebné ešte vyriešiť energetické pomery a ortogonalitu selektívnych kosínusov, čo nie je triviálna úloha. DZCT spektrogram z obrázku 3 by mohol byť ďalej vylepšený adaptívnym nastavením selektivity  $\kappa$ .



**Obrázok 3:** A) vstupný signál  $y[n]$ . B) DCT spektrogram signálu  $y[n]$ . C) Normovaný DZCT spektrogram signálu  $y[n]$ .

## 5. Aproximovaná DZCT

Nedokonalé výsledky z predošlej kapitoly nás priviedli k myšlienke aplikovania ADZT aproximačného algoritmu na DCT transformáciu, čím by sme vytvorili novú Aproximovanú

DZCT [1], [2], [3]. Koeficienty ADZT spektra  $S_z$  sú výsledkom skalárneho násobenia vstupného signálu a selektívnych exponenciál. Algoritmus ADZT minimalizuje spektrálnu funkciu  $|S_z|^2$ , čo vedie k pomerom medzi koeficientmi Fourierovho a Zolotarevovho spektra.

ADZCT aproximuje selektívnu bázu hľadaním optimálnej šírky pásma  $m$  odrezaním nestacionárnej časti Zolotarevových polynómov.

### 5.1. Algoritmus Aproximovanej ADZCT

Aproximačný algoritmus ADZCT je adaptovaná verzia algoritmu ADZT vyvinutým R. Špetíkom [1], [2]. Základnou myšlienkou je reorganizovať Fourierové spektrum obdobným spôsobom ako (4), ale zatiaľ čo ADZT algoritmus rozlišuje reálnu a imaginárnu časť, ADZCT algoritmus oddeľuje párnú a nepárnú časť DCT spektra a vytvára  $\mathbf{Z}$  maticu. ADZCT spektrum je počítané nasledovne

$$\hat{\mathbf{C}}_z = \mathbf{Z} \cdot \mathbf{C}, \quad (6)$$

kde matice  $\hat{\mathbf{C}}_z$  a  $\mathbf{C}$  obsahujú ADZCT a DCT spektrálne koeficienty. Matica  $\mathbf{Z}$  je aproximovaná dolná trojuholníková matica  $\mathbf{L}$  a je vytvorená aproximačným algoritmom.

Aproximačný algoritmus môžeme sumarizovať nasledovne:

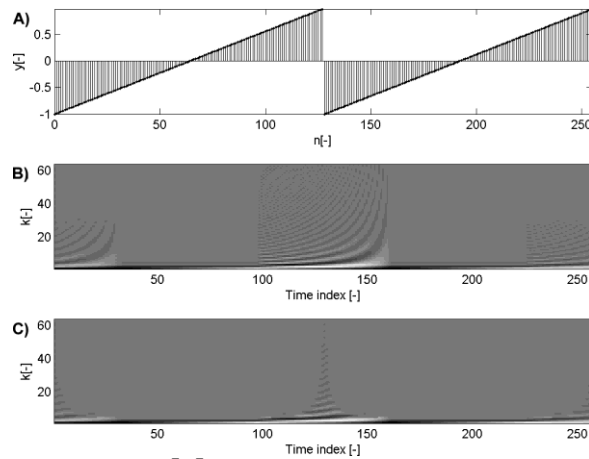
1. Výpočet DCT spektrálnych koeficientov analyzovaného signálu  $C(k) = DCT\{s(n)\}$ .
2. Rozdelenie spektrálnych koeficientov na párne a nepárne zložky  $C^{even}(k)$ ,  $C^{odd}(k)$ .
3. Separovanie stacionárnej a nestacionárnej časti aktuálneho koeficientu  $C(k)$ .  
Nájdienie optimálnej šírky pásma  $m$  a selektivity  $\kappa$ .
4. Skonstruovanie matice  $\mathbf{Z}$  použitím optimálnej šírky pásma  $m$  a selektivity  $\kappa$ .

Hľadanie optimálnej šírky pásma skracovaním nestacionárnej časti chápeme ako frekvenčnú moduláciu zcos nižšieho stupňa na stupeň vyšší. Aplikáciou tejto metódy, ADZCT spektrálna zložka nemusí obsahovať energiu všetkých nižších zložiek. Ďalším veľmi dôležitým dôsledkom skrátenia nestacionárnej časti je rozšírenie selektívnej bázy. Podľa vzťahov (2) a (3), každý zcos stupňa  $p$  môže byť lineárnou kombináciou štandardných kosínusov. Ak  $\mu \geq 0$ , tak zcos môže obsahovať od 0 do  $M - 1$  štandardných kosínusov rovnakej parity v nestacionárnej časti. Pričom pre  $M$  platí

$$M_{odd} = \frac{p-1}{2} + 1, \quad M_{even} = \frac{p}{2} + 1. \quad (7)$$

## 6. Zhrnutie a záver

Aby boli výsledky kompletne, musíme ukázať ADZCT spektrogram testovaného signálu z predošlých príkladov. Obrázok 4 porovnáva ADZCT spektrogram a DCT spektrogram. Ako si môžeme všimnúť, ADZCT filtruje DCT spektrum, vymazáva spektrálne presakovanie a napomáha k presnejšej lokalizácii nestacionárnej oblasti – zlomu v amplitúde (v okolí indexu 128). Výsledky sú dokonca lepšie ako výsledky normovaného DZCT spektrogramu z obrázku 3.



**Obrázok 4: A) vstupný signál  $y[n]$ . B) DCT spektrogram signálu  $y[n]$ . C) ADZCT spektrogram signálu  $y[n]$ .**

## Pod'akovanie

Autori ďakujú Grantovej Agentúre Českej Republiky za podporu nášho výskumného projektu P102/11/1795: Novel Selective Transforms For Non-Stationary Signal Processing.

## Referencie

- [1] R. Špetík, "The Discrete Zolotarev Transform", Doctoral Thesis, Czech Technical University in Prague, FEE, February 2009.
- [2] J. Janík, V. Turon, P. Sovka, R. Spetik and M. Vlcek, "A way to a new multi-spectral transform," The 11th WSEAS International Conference on Signal Processing, Computational Geometry and Systems Theory, Florence, August 2011.
- [3] V. Turon, J. Janík, R. Spetik, P. Sovka and M. Vlcek, "Study of ADZT properties for spectral analysis," The 11th WSEAS International Conference on Signal Processing, Computational Geometry and Systems Theory, Florence, August 2011.
- [4] J. Janík, M. Vlček and P. Sovka, "Selective Zolotarev-Cosines for Selective Spectral Transforms," Digital Technologies 2010, Žilina, November 2010.
- [5] K. R. Rao and P. YIP, "Discrete Cosine Transform, Algorithms, Advantages, Applications", Academic Press 1990.
- [6] G. K. Wallace, "The JPEG Still Picture Compression Standard", IEEE Transactions on Computer Electronics, Vol. 38, No. 1, February 1992.
- [7] J. Huang and Y. Zhao, "A DCT-Based Fast Signal Subspace Technique for Robust Speech Recognition", IEEE Transactions on Speech and Audio Processing, Vol. 8, No. 6, November 2000, pp 747-751.
- [8] N. Ahmed, T. Natarajan and K. R. Rao, "Discrete Cosine Transform," IEEE Transactions on Computers, January 1974, pp. 90-93.
- [9] J. Makhoul, "A Fast Cosine Transform in One and Two Dimensions", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-28, No. 1, February 1980, pp. 27-34.
- [10] E. Feig and S. Winograd, "Fast Algorithms for the Discrete Cosine Transform", IEEE Transactions on Signal Processing, Vol. 40, No. 9, September 1992, pp. 2174-2193.
- [11] M. Vlček and R. Unbehauen, "Zolotarev Polynomials and Optimal FIR Filters", IEEE Transaction on Signal Processing, Vol. 47, No. 3, March 1999, pp. 717-730.

# Vybrané adaptivní metody pro spektrální analýzu signálů

Václav TUROŇ<sup>1</sup>

<sup>1</sup> České vysoké učení v Praze, Fakulta elektrotechnická, Katedra teorie obvodů

[turonvac@fel.cvut.cz](mailto:turonvac@fel.cvut.cz)

*Abstrakt:* Jedním z hlavních cílů spektrální analýzy signálu je sledovat informaci, která je obsažena nebo přenášena signálem a nelze ji snadno analyzovat v časové oblasti. K tomuto účelu existuje mnoho metod, které se dají rozdělit do dvou skupin - na klasické a adaptivní. Mezi klasické metody patří například: Krátkodobá Fourierova transformace (STFT), Vlnková transformace (WT), Wigner-Villova transformace (WVT) nebo relativně nová Hilbert-Huangova transformace (HHT). Výběr vhodné metody zaleží především na typu analyzovaného signálu nebo na požadovaném výstupu. Jednou z metod, která se v praxi běžně používá, je STFT, jejíž hlavní výhodou je intuitivní reprezentace výsledků v podobě spektrogramu. Ten zobrazuje energii analyzovaného signálu rozloženou v čase a frekvenci. Nevýhodou této transformace je to, že časové a frekvenční rozlišení je ovlivněno mnoha faktory, jako například délkou nebo tvarem segmentačního okna. Z těchto důvodů je použití STFT při analýze nestacionárních signálů velmi omezené - musí být zvolen kompromis mezi časovým a frekvenčním rozlišením. Jednou z dalších metod, která se v praxi často používá k analýze nestacionárních signálů, je WT, která je založena na korelaci analyzovaného signálu s vlnkovými funkcemi odvozenými od mateřské vlnkové funkce. Časové a frekvenční rozlišení WT není konstantní, a proto při analýze nestacionárních signálů dosahuje lepšího časového a frekvenčního rozlišení než STFT. Nevýhoda WT je v nutnosti správné volby mateřské vlnky a neintuitivní reprezentaci jejich výsledku ve formě škálogramů.

Druhou skupinu tvoří adaptivní metody, které často využívají ke zpracování signálů klasické metody jen s tím rozdílem, že jejich parametry jsou nastavovány podle aktuálních vlastností analyzovaného signálu. Mnoho těchto metod je založeno na STFT, kde je nastavována aktuální délka segmentu pomocí různých kritérií, jako jsou: velikost spektrálního prosakování (MESP), špičatost spektra (CURT) nebo na velikost aktuální frekvence signálu (KATK). Jednou z dalších adaptivních metod je Krátkodobá aproximovaná diskretní Zolotarevova transformace (STADZT). Tato transformace je Fourierova typu a její báze je tvořena selektivními Zolotarevovými polynomy (ZP) prvního a druhého druhu. Díky jedinečným spektrálním vlastnostem těchto polynomů a výběrem jejich řádu dle aktuálních vlastností signálu dosahuje STADZT lepšího časového a frekvenčního rozlišení než klasické metody a některé odvozené adaptivní metody.

Ve své práci se zabývám především popisem a analýzou této nové časově-frekvenční transformace STADZT a jejím srovnáním s ostatními metodami spektrální analýzy, jako například: STFT a její adaptivní modifikace, WT, WVT nebo HHT.



## Vybrané adaptivní metody spektrální analýzy signálů

Letní doktorandské dny  
24. 5. 2011

Doktorand: Václav Turoň  
Školitel: Pavel Sovka  
Školitel specialista: Miroslav Vlček

## Obsah

1. Přehled spektrálních metod
  - klasické
  - adaptivní
2. Aproximovaná diskretní Zolotarevova transformace (ADZT)
  - Zolotarevovy polynomy
  - odvození
  - vlastnosti
3. Spektrální analýza – testovací nestacionární signál
4. Spektrální analýza – řečový signál
5. Spektrální analýza – elektrokardiogram
6. Závěr

24.5.2011

Letní doktorandské dny

2

## 1 Přehled spektrálních metod

- Krátkodobá Fourierova transformace (STFT)
  - nutná volba délky a tvaru okna
  - kompromis mezi frekvenčním a časovým rozlišením
- Vlnková transformace (WT)
  - nutná volba mateřské vlnky
  - časové a frekvenční rozlišení není konstantní
- Hilbert-Huangova transformace (HHT)
  - rozklad signálu do instrinstických funkcí
  - neintuitivní reprezentace výsledků pomocí Hilbertova spektra
- Wiegner-Villova transformace (WVD)
  - využití nestacionární autokorelační funkce
  - vznik spektrálních křížových interferencí

24.5.2011

Letní doktorandské dny

3

## 1 Přehled spektrálních metod

- Minimální spektrální prosakování (MESP)
  - optimální délka okna je volena pomocí kritéria vybírající spektrum s nejmenší mírou spektrálního prosakování
- Špičatost spektra (CURT)
  - optimální délka okna je volena pomocí kritéria posuzující časově-frekvenční koncentraci vypočtenou pomocí čtvrtého centrálního momentu
- Aktuální frekvence (KATK)
  - optimální délka okna je volena pomocí kritéria posuzující odhad aktuální frekvence signálu statistickými metodami
- Krátkodobá aproximovaná diskretní Zolotarevova tr. (STADZT)
  - optimální báze je vybírána tak, aby energie stacionární a nestacionární části Zolotarevova spektra byla minimální

24.5.2011

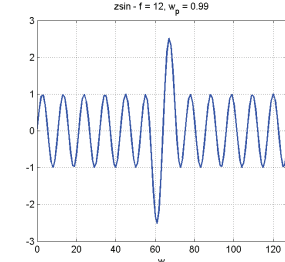
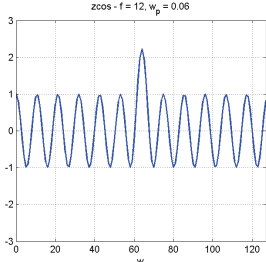
Letní doktorandské dny

4

## 2 Zolotarevovy polynomy

- Symetrické Zolotarevovy polynomy prvního a druhého typu

$$z\cos(N, w_p) = \sum_{\mu=0}^N a_{2\mu}(w_p) \cos(2\pi\mu n) \quad z\sin(N, w_p) = \sum_{\mu=1}^N b_{2\mu-1}(w_p) \sin(2\pi\mu n)$$



- Báze Zolotarevovy transformace:

$$z\exp\left(l, i, 2\pi \frac{n}{N}\right) = z\cos\left(l, 2\pi \frac{n}{N}\right) + i z\sin\left(l, i, 2\pi \frac{n}{N}\right) = \sum_{\mu=-l}^l c'_{2\mu} \exp\left(2\pi \mu \frac{n}{N}\right)$$

24.5.2011

Letní doktorandské dny

5

## 2 Aproximovaná diskretní Zolotarevova transformace

- Přímý zápis diskretní Zolotarevovy transformace

$$s(n) = \sum_{k=-N}^N S_Z(k) c\exp\left(i 2\pi k \frac{n}{N}\right)$$

$$S_Z(k) = \left\langle z\exp\left(i 2\pi k \frac{n}{N}\right), s(n) \right\rangle = 0, \quad \forall |k| > N$$

- Maticový zápis výpočtu koeficientu  $S_Z$

$$s(n) = W(n) \cdot Z^{-1} \cdot S_Z = W_C(t) \cdot S_F \quad \longrightarrow \quad S_Z = Z \cdot S_F$$

- Matice Z je regulární a obsahuje koeficienty Zolotarevových polynomů  $c_{2n}$
- Minimalizace energie stacionární a nestacionární části Zolotarevova spektra

$$S_Z(k) = \sum_{m=-k}^k c_m S_F(m)$$

24.5.2011

Letní doktorandské dny

6

## 2 Vlastnosti ADZT

- ADZT zobecňuje diskrétní Fourierovu transformaci (DFT)
- Při vhodném nastavení parametrů báze ADZT odpovídá bázi DFT
- Dobré frekvenční i časové rozlišení
- Není potřeba používat váhovací okna
- Vyšší výpočetní nároky
- Pracuje se spektrem signálu
  - rozdělení na stacionární a nestacionární část
  - minimalizace vzájemné energie
- Amplitudové spektrum je citlivé na fázi signálu
  - potlačení spektrálního prosakování
  - zachycení pulsování energie mezi jednotlivými harmonickými složkami signálu

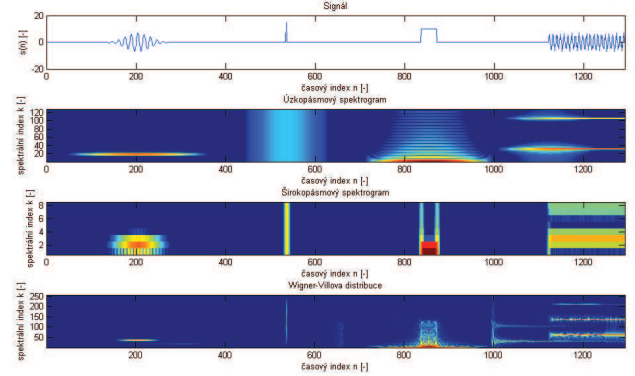
24.5.2011

Letní doktorandské dny

7

## 3 Spektrální analýza

- testovací nestacionární signál



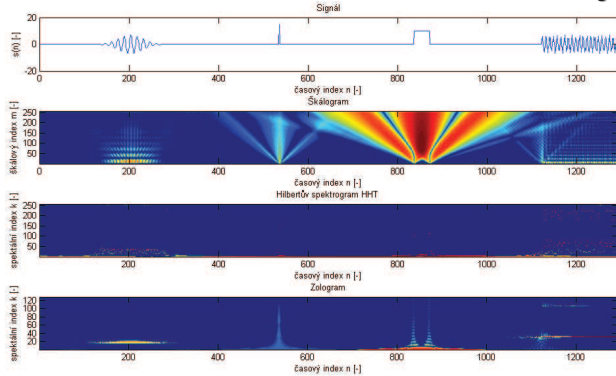
24.5.2011

Letní doktorandské dny

8

## 3 Spektrální analýza

- testovací nestacionární signál



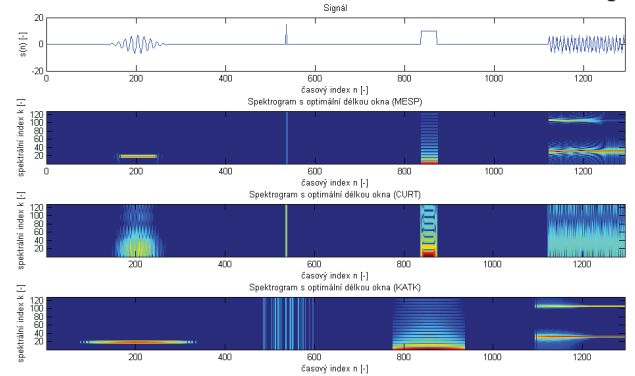
24.5.2011

Letní doktorandské dny

9

## 3 Spektrální analýza

- testovací nestacionární signál



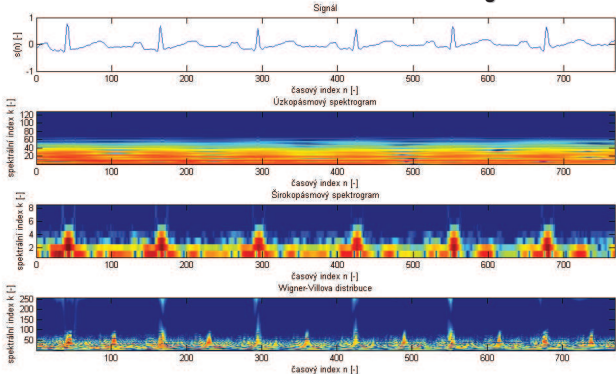
24.5.2011

Letní doktorandské dny

10

## 4 Spektrální analýza

- elektrokardiogram



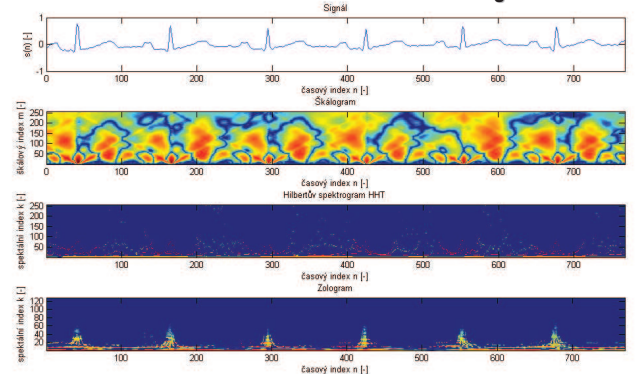
24.5.2011

Letní doktorandské dny

11

## 4 Spektrální analýza

- elektrokardiogram



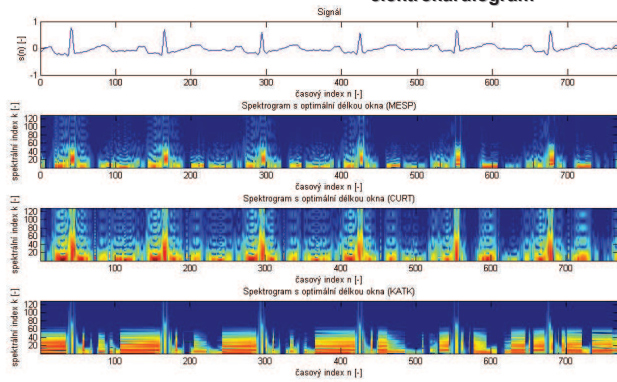
24.5.2011

Letní doktorandské dny

12

## 4 Spektrální analýza

- elektrokardiogram



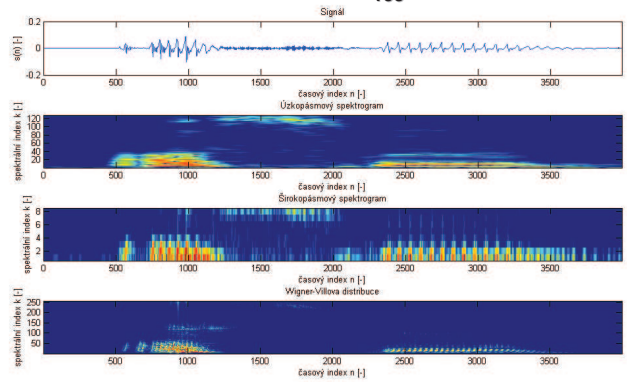
24.5.2011

Letní doktorandské dny

13

## 5 Spektrální analýza

- řeč



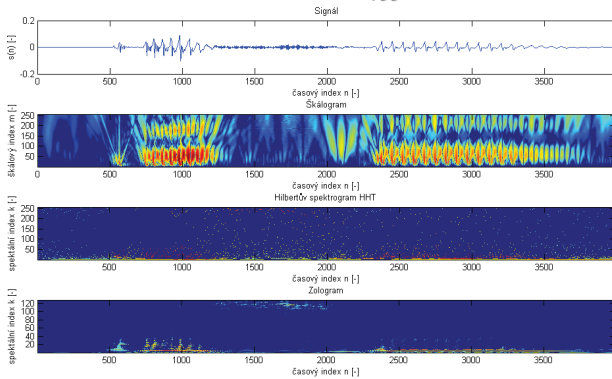
24.5.2011

Letní doktorandské dny

14

## 5 Spektrální analýza

- řeč



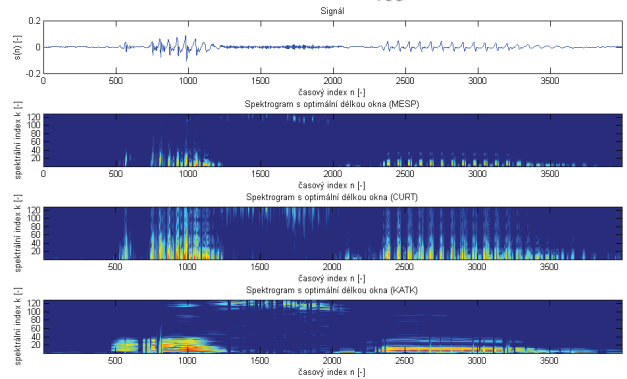
24.5.2011

Letní doktorandské dny

15

## 5 Spektrální analýza

- řeč



24.5.2011

Letní doktorandské dny

16

## 6 Závěr

- Výběr spektrální metody záleží na typu zpracovávaného signálu a na požadovaném výstupu
- Omezení použití klasických metod pro analýzu nestacionárních signálů
- Adaptivní metody dosahují dobrých výsledků při analýze nestacionárních signálů
- ADZT dosahuje dobrého časového i frekvenčního rozlišení díky spektrálním vlastnostem Zolotarevových polynomů

### ➤ Plánované cíle:

- Důkladný popis a analýza ADZT
- Ověření spojitosti s teoretickými předpoklady
- Pokračování v práci na ADZT katalogu
- Srovnání s dalšími metodami spektrální analýzy Music, Pisarenko, PCA

24.5.2011

Letní doktorandské dny

17

## Děkuji Vám za pozornost

Tato práce byla podpořena:

- Grantová agentura ČR P102/11/1795 „Nové selektivní transformace pro číslicové zpracování nestacionárních signálů“

24.5.2011

Letní doktorandské dny

18

# Techniky implementace banky filtrů pomocí SC obvodů

Jiří NÁHLÍK<sup>1</sup>, Jiří HOSPODKA<sup>1</sup>

<sup>1</sup> České vysoké učení v Praze, Fakulta elektrotechnická, Katedra teorie obvodů

[nahlijir@fel.cvut.cz](mailto:nahlijir@fel.cvut.cz), [hospodka@fel.cvut.cz](mailto:hospodka@fel.cvut.cz),

**Abstrakt:** Technika spínaných kapacitorů je velmi často používána pro realizaci filtrů na čipu. Analýza těchto obvodů je však oproti spojitě pracujícím obvodům složitější a lze ji nejlépe provést pomocí simulace. Střídavou analýzu takový obvodů pak provádíme pomocí analýzy v časové oblasti a výslednou přenosovou charakteristiku získáme pomocí diskrétní Fourierovy transformace. Touto metodou byla v programu WinSpice simulována dvoukanálová banka filtrů pracující se spínacím kmitočtem 16 kHz resp. 8 kHz. V simulaci byly uvažovány reálné vlastnosti obvodových součástí tak, aby byl ukázán jejich vliv na funkci obvodu. Dle výsledků simulace byla následně banka filtrů upravena.

## 1. Úvod

Střídavá analýza obvodů se spínanými kapacitami je oproti analýze spojitě pracujících obvodů složitější. Tyto obvody se ale s oblibou používají ve filtrech realizovaných na čipu, neboť kapacity lze oproti rezistorům vyrobit přesněji a filtry se dají velmi snadno doladovat pomocí změny spínacího kmitočtu.

Pro střídavou analýzu obvodů se spínanými kapacitami lze použít tři následující metody:

1. Analýza pomocí nábojových rovnic [4,5]
2. Řešení obecných lineárních diferenciálních rovnic pomocí metod založených na Fourierově nebo Laplaceově transformaci [6].
3. Obecné řešení obvodu pomocí simulace v časové oblasti. Přenosová charakteristika je poté získána pomocí diskrétní Fourierovy transformace. [7,8]

První metodou lze získat přenosovou funkci ideálního obvodu v rovině Z. Druhou metodou lze získat přenosovou funkci pouze lineárních spínaných obvodů. Poslední zmíněnou metodou lze však postihnout všechny parazitní jevy a nelinearity obvodu. Bohužel ale potřebuje vysoký výpočetní výkon.

Tento článek se věnuje analýze dvoukanálové banky filtrů třetí metodou pomocí programu WinSpice.

## 2. Nastavení simulačního programu WinSpice

Pro svou dostupnost a jednoduchost nastavení byl zvolen simulační program WinSpice. Přenosová charakteristika banky filtrů je získávána pomocí časové analýzy, jejíž výsledek je do spektrální oblasti přepočten diskrétní Fourierovou transformací (DFT).



## 2.1. Budící zdroj banky filtrů

Jako zdroj vstupního signálu banky filtrů lze použít sadu harmonických zdrojů spojených v sérii. Kmitočty těchto zdrojů by měly být voleny v ekvidistantních rozestupech a jejich periody se musí rovnat celistvým násobkům periody hodinového signálu  $1/f_c$  [2].

Vstupním signálem může být i úzký puls. Šířka pulzu ale musí být větší nebo rovna polovině periody hodinového signálu ( $1/2f_c$ ). V případech, kdy spínaný obvod zpracovává vstupní signál i ve druhé fázi hodinového signálu, je nutné, aby šířka pulsu byla minimálně rovna celé periodě hodinového signálu ( $1/f_c$ ).

## 2.2. Nastavení časové analýzy

Pro simulaci spínaných obvodů je dle [1] doporučeno použít metodu Gear a nastavit parametr RELTOL v intervalu  $<0.01; 1>$ . Výsledek simulace pro neperiodické buzení lze zlepšit použitím váhovacího okna.

Krok analýzy je vhodné nastavit tak, aby simulátor analyzoval alespoň 10 vzorků za periodu hodinového signálu.

## 2.3. Nastavení pro výpočet diskretní Fourierovy transformace

Pro výpočet spektra odezvy obvodu na vstupní signál pomocí DFT je nutné, aby obvod byl v ustáleném stavu. Dále je nutné, aby byl obvod analyzován v ekvidistantních časových krocích. Toho docílíme použitím příkazu LINEARIZE. Výpočet DFT pak realizuje příkaz SPEC dle vztahu:

$$X(k) = \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{-ikn(2\pi/N)}. \quad (1)$$

## 3. Banka filtrů

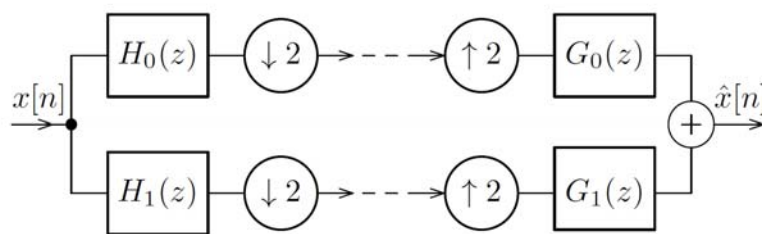
Simulovaná banka filtrů rozkládá diskretní signál vzorkovaný kmitočtem 16 kHz na dva subpásmové signály s vzorkovacím kmitočtem 8 kHz pomocí filtrů realizovaných technikou spínaných kapacitorů. Tyto dva signály jsou přenášeny na druhý konec sdělovacího řetězce, kde jsou opět složeny na signál s původním vzorkovacím kmitočtem 16 kHz. Principiální blokové schéma banky je uvedeno na obr. 1.

Aby byl výstupní signál banky filtrů  $\hat{x}[n]$  co nejvíce podobný vstupnímu signálu  $x[n]$ , musí přenosové funkce filtrů splňovat následující podmínky:

$$H_1(z) = H_0(-z), \quad (2)$$

$$G_0(z) = 2H_1(-z) = 2H_0(z), \quad (3)$$

$$G_1(z) = -2H_0(-z) = -2H_1(z). \quad (4)$$



Obr. 1: Principiální blokové schéma dvoukanálové banky filtrů

Tyto podmínky splňují i přenosové funkce simulované banky filtrů. Ty lze zapsat následujícími vztahy:

$$H_0(z) = \frac{1}{2}G_0(z) = \frac{1}{2} \frac{0.5z^3 + z^2 + z + 0.5}{z(z^2 + 0.5)}, \quad (5)$$

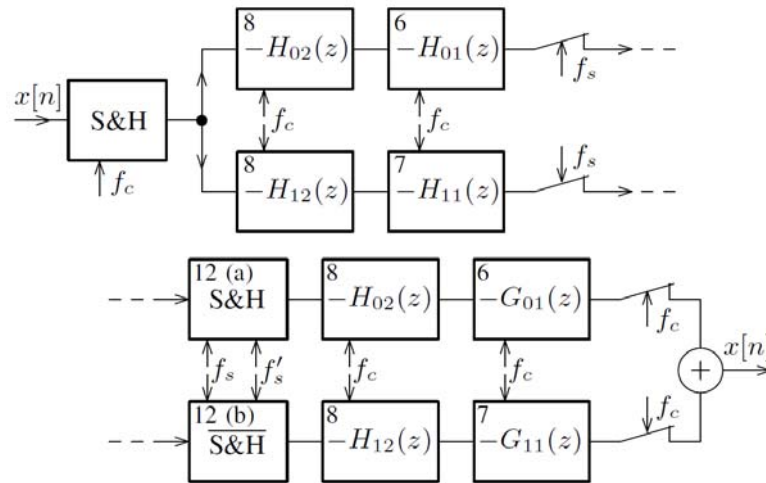
$$H_1(z) = -\frac{1}{2}G_1(z) = -\frac{1}{2} \frac{1 - 0.5z^3 + z^2 - z + 0.5}{z(z^2 + 0.5)}. \quad (6)$$

Funkce  $H_0$  resp.  $H_1$  realizují přenosové funkce typu dolní resp. horní propust. Tyto funkce jsou realizovány kaskádou filtru s bikvadratickou přenosovou funkcí a filtru prvního řádu (integrátoru, popř. derivátoru).

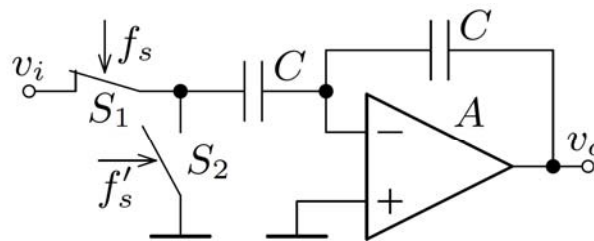
Blokové schéma simulované banky filtrů je uvedeno na obr. 2. Bloky  $H_{02}$  a  $H_{12}$  realizují bikvadratické přenosové funkce, bloky  $H_{01}$  a  $G_{01}$  realizují přenosovou funkci integrátoru a bloky  $H_{11}$  a  $G_{11}$  realizují přenosové funkce derivátoru. Pro správnou funkci filtrů obsahuje banka i obvody typu sample and hold (S&H). Navíc sample and hold obvod označený v obrázku symbolem  $\overline{\text{S\&H}}$  realizuje inverzi vstupního signálu tak, aby byla splněna podmínka daná rovnicí (4). Schéma obvodu  $\overline{\text{S\&H}}$  je uvedeno na obr. 3. Kompletní obvodové řešení banky filtrů lze nalézt v [3].

#### 4. Simulace banky filtrů

Realizace obvodu se spínanými kapacitami vyžaduje použití rychlého operační zesilovače s velkou rychlostí přeběhu a malou hodnotou vstupní napěťové nesymetrie. Pro simulaci byl tedy použit model operačního zesilovače LT1055.



Obr. 2: Blokové schéma banky filtrů



Obr. 3: Schéma realizace  $\overline{\text{S\&H}}$  obvodu

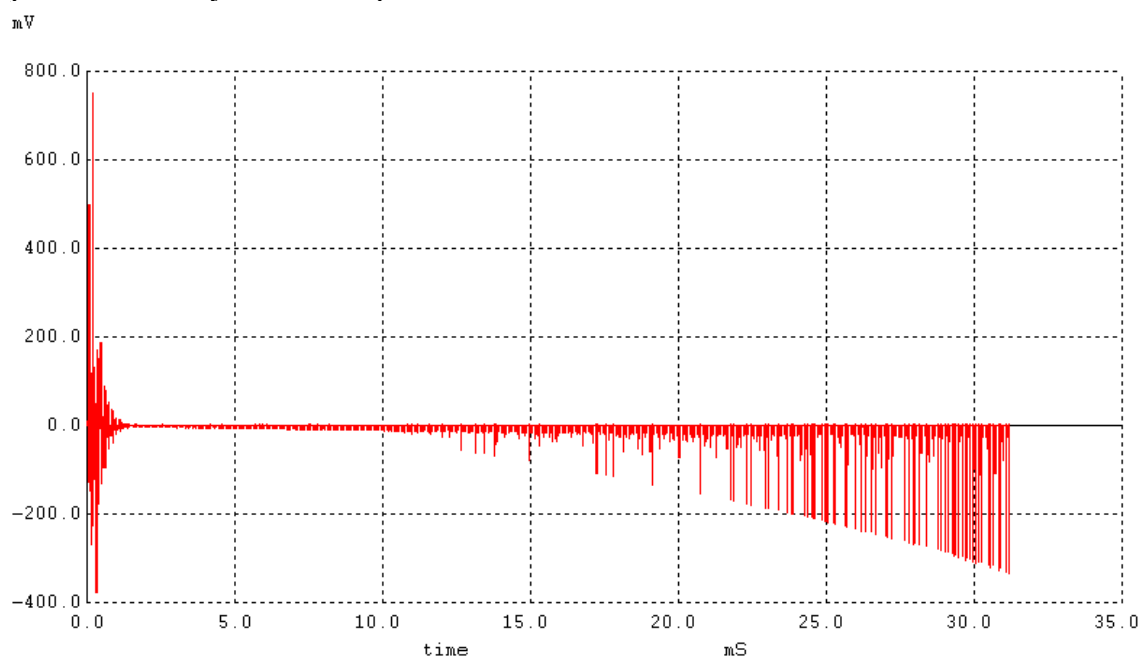
Jako budící signál byl zvolen puls s amplitudou 1 V a šířkou rovnou polovině periody hodinového signálu.

Jako součtový člen na konci řetězce banky filtrů byl použit invertující sumátor s operačním zesilovačem. Ve výsledné simulaci bylo tedy nutné, aby si invertující a neinvertující S&H obvod vyměnily navzájem pozice.

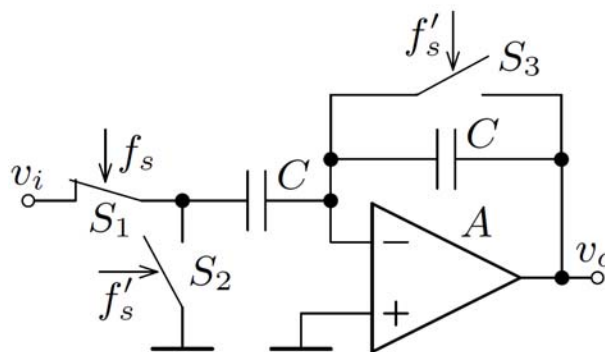
## 5. Výsledky simulací

První simulace obvodu ukázala, že nenulová vstupní napěťová nesymetrie operačního zesilovače způsobuje postupný nárůst odezvy banky filtrů až do saturace (viz obr. 4.). Tento nárůst byl způsoben nevhodnou konstrukcí  $\overline{S\&H}$  obvodu, která umožňovala nabíjení zpětnovazební kapacity i při nulovém vstupním napětí. Jak ukazuje obr. 5, byl tento problém vyřešen přidáním spínače  $S_3$ .

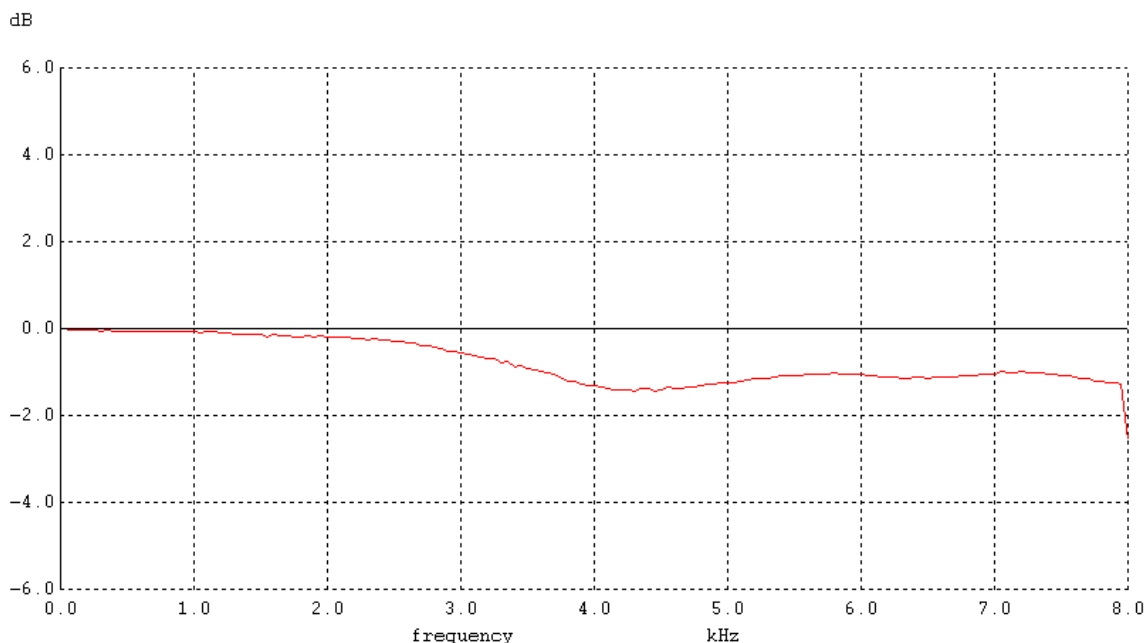
Výsledná přenosová charakteristika banky filtrů je zobrazena na obr. 6. Pokles přenosu signálu v oblasti vyšších kmitočtů než 4 kHz je dán především nastavením parametru  $G_{min} = 1e-9$  (minimální vodivost mezi každými dvěma uzly obvodu). Hodnota tohoto parametru má největší vliv na přenosovou funkci derivátoru.



Obr.4: Odezva banky filtrů na budící puls způsobená vstupní napěťovou nesymetrií operačního zesilovače



Obr.5: Upravená konstrukce  $\overline{S\&H}$  obvodu



**Obr.6: Přenosová charakteristika banky filtrů**

## 6. Závěr

Příspěvek se zabývá simulací dvoukanalové banky filtrů realizované technikou spínaných kapacitorů v programu WinSpice. Banka byla simulována s modely reálných součástek. Výsledná přenosová charakteristika banky vykazuje pokles přenosu o cca 1 dB na kmitočtech vyšších než 4 kHz. Ten je způsoben především poklesem přenosu derivátoru vlivem svodů. Pro konstantní přenosovou charakteristiku je tedy třeba obvod dále optimalizovat.

## Poděkování

Tato práce byla podpořena grantem Studentské grantové soutěže ČVUT číslo SGS12/143/OHK3/2T/13.

## Reference

- [1] Bičák J., Hospodka J.: Frequency Response of Switched Circuits in SPICE , ECCTD' 03, Cracow 2003, Vol. I, pp. 333-336, Poland, ISBN 83-88309-95-1.
- [2] Bičák J., Hospodka J.: Using WinSpice Program for Switched Capacitor Filter Analysis, WSEAS Transactions on Circuits and Systems. 2005, vol. 8, no. 4, p. 992-1001. ISSN 1109-2734.
- [3] Hospodka J., Sovka P., Pšenička B.: Design and Realization of a Filter Bank by Switched Capacitor Technique. In 20th European Conference on Circuit Theory and Design (ECCTD 2011) [CD-ROM]. Linköping: IEEE, 2011, ISBN 978-1-4577-0616-5.
- [4] Kurth C. F., Moschytz, G. S.: Nodal analysis of switched-capacitor networks, IEEE Transaction on Circuits and Systems, Vol. CAS/26, No. 2, February 1979, pp. 93-104.
- [5] Bičák J., Hospodka J., Martinek P.: Analysis of SI Circuits in Maple Program, ECCTD' 01, Espoo 2001, Vol. 1, pp. 121-124, Finland.

- [6] Vlach J., Opal A.: Analysis and sensitivity of periodically switched linear networks, IEEE Trans. on Circuits and Systems, Vol. 36, No. 4, April 1989, pp. 522-532.
- [7] Haigh D.G., Toumazou C.: On Computer Simulation of integrated Switched Capacitor Circuits, Proceeding of the ECCTD'89 London, pp. 199-204, UK 1989.
- [8] Scott J., Parker A.: Modern Guide to Spectral analysis with SPICE, IEEE Circuit and Device Magazine, Vol. 11, pp. 10-16, ISSN: 8755-3996, 1995.
- [9] Smith M.: WinSpice User's Manual, <http://www.winspice.com>, 2003.
- [10] Fliege N. J., Multirate Digital Signal Processing, John-Wiley & Sons, New York, 1994.

# Využití EEG signálu pro identifikaci osob

Milan KOSTÍLEK<sup>1</sup>

<sup>1</sup> České vysoké učení v Praze, Fakulta elektrotechnická, Katedra teorie obvodů

[kostimil@fel.cvut.cz](mailto:kostimil@fel.cvut.cz)

*Abstrakt:* Tento článek se věnuje metodě biometrické identifikace osob pomocí elektroencefalografického signálu (EEG) a jejímu využití v rozhraní mozek-stroj. Ve stručnosti je uvedeno seznámení s problematikou obecné klasifikace pohybového EEG a současný stav výzkumu v této oblasti. Druhá část příspěvku je pak zaměřena na provedené identifikační experimenty. Nejprve je popsán identifikační algoritmus, využívající metodu frekvenčně přiblíženého auto-regresního modelování (FZ-AR) a klasifikátor založený na výpočtu Mahalanobisovy vzdálenosti. Dále jsou uvedeny výsledky identifikace při aplikaci na konkrétní offline EEG databázi. Hlavním cílem experimentů bylo zhodnotit dlouhodobou časovou stabilitu identifikačního algoritmu, proto byla použita databáze, obsahující signály získané během dvou nezávislých měření oddělených přibližně roční pauzou. Na závěr je uvedeno několik možných vylepšení tohoto identifikačního algoritmu.

## 1. Úvod

Frekvenční spektrum EEG signálu se běžně rozděluje do několika frekvenčních pásem. Pásmo od 8 do 13 Hz přitom bývá nazýváno alfa, jeho součástí jsou však také tzv.  $\mu$ -rytmy. Ty jsou generovány sensorimotorickými neurony, a jejich amplituda je největší když není vykonáván žádný pohyb a naopak je utlumována během pohybu. S poměrně velkou přesností lze na základě těchto  $\mu$ -rytmů klasifikovat úseky EEG signálu do dvou skupin, a to EEG pohybové a klidové. Tuto klasifikaci lze využít v rozhraní mozek-stroj (BCI), což je takový systém, který dokáže signály lidského mozku převést na signály jiného typu, např. signály srozumitelné pro software běžného počítače. Pro efektivní využití BCI systému je však potřeba co nejvyšší datová propustnost, které lze dosáhnout například zvýšením počtu tříd pro klasifikaci. Při klasifikaci pohybového EEG lze tedy dále rozlišovat mezi pohyby na pravé a levé straně těla, mezi pohyby jednotlivých končetin, nebo také mezi konkrétními typy pohybu jako jsou například extenzní či flexní pohyby.

Výzkum týkající se klasifikace pohybového EEG a především vylepšování jejich výsledků se v současnosti ubírá mnoha různými směry. Ve studii [1] bylo ukázáno, že lze pomocí EEG signálu rozlišovat mezi poměrně blízkými pohyby zápěstí a prstů s přesností přibližně 70 %. V práci [2] byla představena metoda dekompozice  $\mu$ -rytmů, která slouží ke zjištění užšího  $\mu$ -pásma specifického pro daný subjekt, využití tohoto nalezeného frekvenčního pásma při klasifikaci pohybového EEG pak vede ke zlepšení klasifikačních výsledků. Specifická metoda klasifikace představovaného pohybového EEG sloužící k ovládní kursoru ve třech dimenzích byla navržena ve studii [3]. To že jde zjistit z pohybového EEG signálu nejen druh vykonávaného pohybu, ale i jeho směr a kinematika, pak ukazuje souhrn výsledků v práci [4].

Další část tohoto příspěvku se zabývá identifikací subjektu pouze na základě jeho naměřeného EEG signálu. Tato metoda je potencionálním předstupněm při vytváření BCI systému s vysokou datovou propustností. Identifikace je prováděna pomocí klasifikace EEG signálu dle fundamentálních frekvencí v  $\mu$ -pásmu. V tomto případě se tedy nejedná o

standardní uplatnění klasifikace pohybového EEG, ale naopak výsledky této identifikace lze využít k usnadnění práce s BCI systémy a usnadnění klasifikace pohybového EEG obecně.

## 2. Identifikace osob pomocí EEG signálu

Biometrie je automatické rozpoznávání osob na základě některých biologických znaků. Biometrický systém lze využít pro identifikaci či autentizaci osob, přičemž tyto dva termíny nelze zaměňovat. Během identifikačního procesu je prohledána celá databáze uložených údajů a subjektu je přiřazena nejpravděpodobnější identita. Naopak při autentizaci jsou pouze porovnány charakteristiky naměřeného subjektu s charakteristikami identity, za kterou se měřený subjekt prohlašuje. Výsledkem autentizace je tedy pouze potvrzení či zamítnutí nárokované identity, kdežto výsledkem identifikace je určení předem neznámého subjektu. Biometrické systémy pracují s biologickými charakteristikami, které mohou být fyziologické nebo behaviorální. Mezi běžně využívané fyziologické charakteristiky patří otisky prstů, vzory duhovky či sítnice, obličejové rysy, DNA a další. Behaviorálními charakteristikami pak mohou být hlasové vzory, dynamické parametry podpisu a další.

EEG signál je biologický znak stojící na pomezí fyziologické a behaviorální charakteristiky, z hlediska fyziologického je pro daný subjekt jedinečný a založený na genetické informaci, z hlediska behaviorálního je pak ovlivnitelný současným stavem mysli subjektu. Využití EEG signálu v biometrickém systému již bylo ukázáno v několika studiích, a to jak při identifikaci tak i při autentizaci [5] [6] [7]. Identifikační experimenty, které jsou zmíněny dále v tomto článku, byly provedeny pomocí algoritmu [8] využívajícího metody frekvenčně přiblíženého auto-regresního modelování (FZ-AR) a klasifikátoru založeného na Mahalanobisově vzdálenosti.

### 2.1. Identifikační algoritmus

Metoda FZ-AR modelování sestává z několika kroků. Nejprve je provedena modulace kolem střední frekvence požadovaného frekvenčního pásma, dle rovnice:

$$x_m(n) = e^{j2\pi(f_m/f_s)n}x(n) \quad (1)$$

kde  $x_m(n)$  je výsledný modulovaný signál,  $f_m$  je požadovaná modulační frekvence,  $f_s$  je vzorkovací frekvence signálu a  $x(n)$  je původní nedomulovaný EEG signál. V dalším kroku je provedena decimace modulovaného signálu, čímž je docíleno zvýšení frekvenčního rozlišení. Následně jsou vypočteny parametry FZ-AR modelu pomocí standardních metod autoregresního modelování. Pro modulaci byla zvolena modulační frekvence  $f_m = 10.5$  Hz jako střed požadovaného frekvenčního pásma 8-13 Hz, decimace byla provedena s decimačním faktorem  $K_{ZOOM} = 28$ .

Pro klasifikaci byly zvoleny dva typy parametrů, nejprve čisté koeficienty FZ-AR modelu a poté fundamentální frekvence  $\mu$ -rytmů získané z pozice pólů v AR modelu. Použitý klasifikátor využívá regularizovanou Mahalanobisovu vzdálenost:

$$D_{RM}(\mathbf{x}_i, s) = (\mathbf{x}_i - \mathbf{m}[s])^T [(1-\lambda)(\Sigma[s] + \epsilon \mathbf{I})^{-1} + \lambda \mathbf{I}] (\mathbf{x}_i - \mathbf{m}[s]) \quad (2)$$

kde  $\mathbf{x}_i$  je klasifikovaný vektor,  $\mathbf{m}[s]$  je centroid trénovacích dat pro subjekt  $s$ ,  $\lambda$  je parametr kontrolující poměr mezi hypersférickou a hyperelipsoidální složkou vzdálenosti,  $\Sigma[s]$  je kovarianční matice trénovacích dat pro subjekt  $s$ ,  $\epsilon$  je parametr kontrolující regularizaci vzdálenosti, a  $\mathbf{I}$  je jednotková matice. Nejvhodnější hodnoty parametrů  $\epsilon$  a  $\lambda$  byly zvoleny experimentálně jako  $\epsilon = 0.2$  a  $\lambda = 0.4$ .

**Tab. 1: Průměrná úspěšnost identifikace pro všechny testované subjekty**

Část databáze	Délka segmentu	Řád AR modelu	Parametry klas.	Úspěšnost ident.
1. část	60 sekund	1	AR parametry	94.9 %
1. část	60 sekund	1	$\mu$ -frekvence	97.4 %
1. část	60 sekund	7	AR parametry	97.4 %
1. část	60 sekund	7	$\mu$ -frekvence	91.7 %
1. část	30 sekund	7	AR parametry	98.2 %
1. část	30 sekund	7	$\mu$ -frekvence	98.2 %
2. část	60 sekund	1	AR parametry	94.6 %
2. část	60 sekund	1	$\mu$ -frekvence	97.2 %
2. část	60 sekund	7	AR parametry	97.9 %
2. část	60 sekund	7	$\mu$ -frekvence	87.1 %
2. část	30 sekund	7	AR parametry	97.6 %
2. část	30 sekund	7	$\mu$ -frekvence	97.6 %
Sloučená databáze	60 sekund	1	AR parametry	64.7%
Sloučená databáze	60 sekund	1	$\mu$ -frekvence	77.6 %
Sloučená databáze	60 sekund	7	AR parametry	75.1 %
Sloučená databáze	60 sekund	7	$\mu$ -frekvence	77.8 %
Sloučená databáze	30 sekund	7	AR parametry	73.9 %
Sloučená databáze	30 sekund	7	$\mu$ -frekvence	81.8 %

## 2.2. Popis použité databáze

Databáze EEG signálů [9], která byla použita pro všechny následující experimenty, obsahuje data 10 subjektů (všichni mužského pohlaví, průměrný věk 32 let s  $\sigma = 11.8$ ). Databáze se skládá ze dvou oddělených částí, data v první části databáze byla naměřena přibližně o rok dříve než data v druhé části. Jeden ze subjektů se neúčastnil druhého měření, proto byly pro identifikační experimenty použity EEG signály pouze devíti subjektů. EEG signály byly získány z celkem 53 elektrod rozmístěných dle systému 10-10, vzorkovací frekvence byla 1024 Hz, během předzpracování byla decimací zmenšena na 256 Hz. Měření každého subjektu bylo rozděleno do čtyř bloků, během prvních tří bloků vykonával subjekt extenzní a flexní pohyby pravého či levého ukazováčku, během čtvrtého bloku pak bylo naměřeno klidové EEG.

## 2.3. Identifikační experimenty a výsledky

Nejprve byli provedeny experimenty pro nalezení nejvhodnějších parametrů pro klasifikátor, a to jak pro obě části databáze odděleně, tak i pro spojenou databázi. V případě oddělených částí databáze byla z důvodu poměrně malého množství dat použita technika krosvalidace, vždy bylo provedeno 10 nezávislých běhů klasifikátoru, přičemž pro každý z nich bylo 75% dat náhodně zvoleno jako trénovací množina a zbylá data byla použita jako množina testovací. Spojení databáze pak bylo provedeno takovým způsobem, že první část byla použita jako množina trénovací a druhá část jako množina testovací. I v případě spojené databáze bylo technikou krosvalidace provedeno 10 nezávislých běhů klasifikátoru, vždy bylo tedy pro trénování použito pouze 75% náhodně zvolených dat z první části databáze.

V Tabulce č. 1 jsou výsledky těchto experimentů pro různá nastavení vlastností použitého klasifikátoru, lepších výsledků přitom dosahuje klasifikátor využívající fundamentální frekvence  $\mu$ -rytmů. Další vlastnosti klasifikátoru byly s ohledem na nižší výpočetní nároky nastaveny takto: řád FZ-AR modelu 1, délka použitých segmentů 60



**Tab. 2: Úspěšnost identifikace pro jednotlivé subjekty, klasifikátor využívající  $\mu$ -frekvence**

Subjekt	Úspěšnost identifikace		
	1. část databáze	2. část databáze	Sloučená databáze
1	99.0±1.8 %	98.5±2.1 %	43.5±6.8 %
2	94.0±3.1 %	93.1±3.0 %	25.0±0.1 %
4	95.7±1.4 %	98.9±1.9 %	100.0±0.0 %
5	97.7±1.9 %	97.8±1.8 %	81.9±4.9 %
6	97.3±1.3 %	97.6±2.0 %	96.6±0.1 %
7	96.9±1.6 %	97.3±2.2 %	100.0±0.0 %
8	98.2±1.5 %	98.3±1.7 %	60.3±8.0 %
9	100±0.0 %	95.9±4.5 %	100.0±0.0 %
10	97.1±1.9 %	97.1±1.9 %	100.0±0.0 %

sekund. Důležitou částí experimentů bylo ověření dlouhodobé časové stability použitého identifikačního algoritmu, proto byla databáze spojena způsobem zmíněným výše. Již z Tabulky č. 1, kde jsou zobrazeny průměrované výsledky, je patrné, že při použití spojené databáze dochází k zhoršení průměrné úspěšnosti identifikace. Z výsledků identifikace pro jednotlivé subjekty (viz Tabulka č. 2) je však vidět, že ke zhoršení výsledků dochází jen u některých konkrétních subjektů.

Dále byly provedeny experimenty pro různé kombinace bloků použitých EEG signálů, nastavení klasifikátoru bylo ponecháno stejné jako pro předchozí experimenty. Výsledky jsou uvedeny v Tabulce č. 3, je vidět že nejlepších výsledků dosahuje identifikace pro kombinaci druhého a třetího bloku EEG signálů. První blok signálu obsahuje data ze začátku měření, je proto možné že obsahuje některé pohybové artefakty, způsobené počátečním neklidem měřeného subjektu. Čtvrtý blok signálu obsahuje naopak jen klidové EEG, nejsou zde žádné změny související s pohybem, které jsou pro klasifikační postup důležité. Tyto předpoklady souhlasí s výsledky experimentů, z čehož vyplývá, že je nejvýhodnější pro identifikaci používat pouze druhý a třetí blok EEG signálů.

**Tab. 3: Průměrná úspěšnost identifikace pro různé kombinace bloků EEG signálu**

Bloky	Úspěšnost identifikace	Bloky	Úspěšnost identifikace	Bloky	Úspěšnost identifikace
1	71.9 %	1-3	74.5 %	1-2-3	73.8 %
2	78.9 %	1-4	77.5 %	1-2-4	77.5 %
3	71.1 %	2-3	87.1 %	1-3-4	77.6 %
4	65.9 %	2-4	80.9 %	2-3-4	80.9 %
1-2	72.1 %	3-4	74.0 %	1-2-3-4	77.6 %

### 3. Závěr

Identifikační algoritmus použitý pro experimenty v tomto článku dosahuje velice dobrých výsledků pro jednotlivé části použité databáze, úspěšnost takové klasifikace je v průměru přibližně 97 %. Pro ověření časové stability byla databáze sloučena, její první část byla použita pro trénování klasifikátoru, a druhá část, získaná s ročním odstupem, byla použita pro testování. Po sloučení databáze dochází ke zhoršení identifikačních výsledků, celková úspěšnost klesá v průměru na 78%. Toto zhoršení však nastává jen u některých subjektů a celkové výsledky jsou i po sloučení databáze velmi dobré. Lze tedy prohlásit, že použitý identifikační algoritmus je pro většinu testovaných subjektů stabilní i během delšího časového období. Při použití pouze části EEG signálů, bez počátečního a koncového bloku, dochází ke zlepšení celkových výsledků v průměru až na 87%.

Pro další zlepšení identifikačních výsledků je možno upravit algoritmus několika různými způsoby. Jedním takovým vylepšením může být použití ARMA modelu namísto jednoduššího AR modelu. Parametry ARMA modelu vypovídají přesněji o charakteru použitého signálu, ovšem za cenu vyšších výpočetních nároků. Dále lze jako parametry klasifikátoru využít kromě fundamentálních  $\mu$ -frekvencí také informaci o šířce  $\mu$ -pásma, jak bylo ukázáno v práci [10]. K potenciálnímu zlepšení výsledků by také mohla vést náhrada klasifikátoru využívajícího Mahalanobisovu vzdálenost za jiný typ klasifikátoru, například ve studii [11] bylo dosaženo dobrých identifikačních výsledků při využití metody Support Vector Machine (SVM).

### Poděkování

Výzkum popsáný v tomto článku byl podporován z interního grantu ČVUT SGS12/143/OHK3/2T/13 "Algoritmy a hardwarové realizace číslicového zpracování signálů".

### Reference

- [1] Mohamed, A. K.; Marwala T.; John L. R. Single-trial EEG Discrimination between Wrist and Finger Movement Imagery and Execution in a Sensorimotor BCI. *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (2011), 6289–6293.
- [2] Veluvolu K. C.; Wang Y.; Kavuri S. S. Adaptive estimation of EEG-rhythms for optimal band identification in BCI. *Journal of Neuroscience Methods* 203 (2012), 163-172.
- [3] Li T.; Hong J.; Zhang J. Electroencephalographic (EEG) Control of Cursor Movement in Three-dimensional Scene Based on Small-world Neural Network. *2010 IEEE International Conference on Intelligent Computing and Intelligent Systems* (2010), 587-591.
- [4] Jerbi K.; Vidal J. R.; Mattout J.; Maby E.; Lecaigard F.; Ossandon T.; Hamamé C. M.; Dalal S. S.; Bouet R.; Lachaux J. P.; Leahy R. M.; Baillet S.; Garnero L.; Delpuech C.; Bertrand O. Inferring hand movement kinematics from MEG, EEG and intracranial EEG: From brain-machine interfaces to motor rehabilitation. *IRBM* 32 (2011), 8-18.
- [5] Poulos M.; Rangoussi M.; Kafetzopoulos E. Person identification via the EEG using computational geometry algorithms. *Proceedings of the Ninth European Signal Processing* (1998), 2125-2128.

- [6] Tuoyama H. EEG-Based Personal Identification. *Biomedical Engineering* (2009), 415-424.
- [7] Palaniappan R.; Mandic D. P. Biometrics from Brain Electrical Activity: A Machine Learning Approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2007), 738-742.
- [8] Šťastný J. Brain-Computer Interface with an automatic user identification. *Utility Model no. 19972, application no. 2009-21380, Industrial Property Office*, accepted on the 24th of August 2009.
- [9] Doležal J.; Šťastný J; Švadlenka M. EEG database merging for BCI applications. *under review in ElectroScope* (2011).
- [10] Varner J. L.; Potter R. A.; Rohrbaugh J. W. A procedure for automatic classification of EEG genetic variants. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (1991), 451-452.
- [11] Brigham K.; Kumar V. Subject Identification from Electroencephalogram (EEG) Signals During Imagined Speech. *Fourth IEEE International Conference on Biometrics: Theory Applications and Systems* (2010), 1-8.

# Verifikace a úpravy klastrování výsledků ICA při zpracování EEG

Martin DOBIÁŠ<sup>1</sup>

<sup>1</sup> České vysoké učení v Praze, Fakulta elektrotechnická, Katedra teorie obvodů

[dobiama3@fel.cvut.cz](mailto:dobiama3@fel.cvut.cz)

*Abstrakt:* Tento příspěvek popisuje metodu použití analýzy nezávislých komponent (ICA) při číslicovém zpracování pohybového elektroencefalografického (EEG) signálu. Zaměřuje se na postup rozlišování pohybů na jedné straně těla v naději, že vyvinuté metody pomohou zvýšit rychlost rozhraní mozek-stroj. Předchozí výzkum ukázal, že pomocí analýzy nezávislých komponent je možné zlepšit výsledky klasifikace distální vs. proximální pohyb na jedné straně těla. Tato práce se snaží výsledky replikovat na jiné databázi obsahující pohyby palce a malíčku pravé ruky. Vzhledem k tomu, že výsledky ICA jsou závislé na inicializaci iteračního algoritmu, je nutné hodnotit spolehlivost jednotlivých komponent. Tato zpráva se zaměřuje na hodnocení spolehlivosti pomocí metody aglomerativního hierarchického klastrování.

## 1. Úvod

Tento příspěvek si klade za cíl přispět k vylepšení vlastností a schopností rozhraní mozek-stroj (Brain-Computer Interface - BCI). Přístroje vybavené takovýmto rozhraním mají veliký potenciál pro využití v asistivních technologiích, pro vojenské účely, či v zábavním průmyslu. Rychlost takovýchto rozhraní je však momentálně malá a zdaleka nedosahuje rychlostí těch běžně používaných. Naším dlouhodobým cílem je tuto rychlost zvýšit alespoň na 100 bitů za minutu, což je cca desetkrát méně než například u klávesnice.

V našem výzkumu používáme pohybové EEG, jelikož pro lidské bytosti je velice přirozené ovládat své okolí pohybem. Jelikož rozlišení pohyb na levé/pravé straně těla omezuje počet zjišťovaných stavů v EEG na dva, snažíme se pomocí analýzy časového průběhu EEG na elektrodách rozlišit různé pohyby na jedné straně těla. Takovéto rozlišení v budoucnu umožní dále zvyšovat rychlost BCI.

### 1.1. Pohybové EEG

Pohybová aktivita je v EEG doprovázena rytmickými změnami, které jsou zpravidla patrné v  $\mu$  (8-13 Hz) a v  $\beta$  (12-30 Hz) pásmu. Tyto změny se nazývají desynchronizace (Event-Related Desynchronization - ERD), snížení aktivity, a synchronizace (Event-Related Synchronization - ERS), zvýšení aktivity. Podle pásem se pak rozdělují na  $\mu$ ERD,  $\beta$ ERD,  $\mu$ ERS a  $\beta$ ERS. ERD začíná obvykle cca 2 s před pohybem a končí několik málo sekund po něm. ERS je spojeno s nástupem pohybu a je zpravidla silnější v  $\beta$  pásmu, jeho špička je zhruba 1 s po pohybu, kdy ERD stále ještě probíhá. ERD je výraznější v  $\mu$  pásmu.

## 1.2. Databáze

Použitá databáze byla původně nahrána pro potřeby [1] a obsahuje nahrávky pohybů pravého malíčku, pravého palce a klidové aktivity. EEG bylo získáno od 8 subjektů, ze kterých bylo 7 mužů a 1 žena. Průměrný věk byl 24,5 roku.

Bylo použito 41 Ag/AgCl elektrod rozmístěných symetricky a ekvidistantně (2,5 cm) nad centrálním pohybovým kortexem. Impedance elektrod byla držena pod 10 kΩ. Dva kanály electrooculogramu a dva kanály electromyogramu na palci a malíčku byly také nahrány. Použitá vzorkovací frekvence byla 256 Hz, v průměru bylo nahráno 84 epoch každého pohybu. Pro zpracování časově-frekvenčního vývoje bylo použito okno délky 1 s s posuvem 250 ms.

## 1.3. Analýza nezávislých komponent

Analýza nezávislých komponent (Independent Component Analysis - ICA) je velice vyhledávaným algoritmem pro číslicové zpracování signálů. Při zpracování biologických signálů se často používá k odstranění artefaktů či k rozkladu signálů. ICA je jednou z metod slepé separace zdrojů (Blind Source Separation - BSS), z nahraných signálů odhaduje původní, skryté, zdroje.

Jedním z modelů ICA vhodných pro analýzu EEG je takzvaný model s okamžitým směřováním, kdy předpokládáme, že zpoždění signálu od míst vzniku (zdroje) k místům pozorování je zanedbatelné. Tento model velice dobře odpovídá tomu, jak EEG vzniká a šíří se na povrch skalpu. Může být popsán následovně:

$$\mathbf{X} = \mathbf{A} \cdot \mathbf{S},$$

kde řádky  $\mathbf{X} = (x_1 \dots x_P)^T$  reprezentují vektory pozorovaných dat,  $\mathbf{A}_{P \times N}$  je mixážní matice,  $\mathbf{S} = (s_1 \dots s_N)^T$  jsou vektory nezávislých komponent,  $P$  je dimenze prostoru pozorování (počet elektrod) a  $N$  vyjadřuje počet nezávislých zdrojů.

Pro rozklad EEG je pro účely tohoto příspěvku používán algoritmus EFICA [2], který je rozšířením dobře známého algoritmu FastICA [3]. EFICA je iterativní algoritmus pro nalezení maximálně nezávislých komponent, jehož výsledky jsou závislé na inicializaci. Spolehlivost jeho výsledků, může být tímto faktem zpochybněna. Dva parametry ovlivňují inicializaci, počáteční hodnota mixážní matice a množství vzorků pozorovaných signálů použité k výpočtu. Pro výpočet ICA jsou zde použity úseky dat dlouhé 10 s vystředěné okolo počátku pohybu. Spolehlivost komponent lze zkoumat pomocí jejich klastrování tak, že výpočet komponent je několikrát opakován pro různé počáteční podmínky. Nezávislé komponenty příslušné velikým klastřům pak mohou být považovány za spolehlivé.

V článku [4] je zkoumána spolehlivost výsledků FastICA algoritmu v závislosti jak na počáteční hodnotě mixážní matice, tak na množství použitých dat. Aglomerativní hierarchické klastrování s average-linkage kritériem je použito k seskupení komponent. Autoři v [5] využívají metodu nazvanou „split-half comparison“, která seskupuje komponenty spočítané ze všech pokusů, z jedné poloviny pokusů a z druhé poloviny pokusů. Komponenty jsou seskupovány jak na základě topologické vzdálenosti, tak na základě vzdálenosti v časové oblasti.

## 1.4. Předchozí výzkum

Tato práce rozšiřuje experimenty, které byly provedeny na našem pracovišti nad databází obsahující distální (pravý ukazováček) a proximální (pravé rameno) pohyby na jedné straně těla. Zároveň se snaží o jejich validaci na jiné databázi.

Práce [6][7] ukázaly, že je možné rozložit EEG na pohybové a nepohybové nezávislé komponenty. V průměru polovina komponent byla shledána pohybová na základě známých vlastností pohybového EEG. Výzkum v [8] ukazuje pozitivní dopady výběru pohybových

komponent a nulování nepohybových na výsledky klasifikace pohybů. Pohybové komponenty byly vybírány manuálně na základě jejich lokalizace na skalpu a krátkodobého spektra. Práce [9] používá automatickou klasifikaci nezávislých komponent a představuje míry použité k rozhodnutí, zda je komponenta pohybová, či nikoliv. V článku [10] je pak důkladně popsáno klastrování komponent na základě jejich topologické vzdálenosti s použitím hierarchického klastrování.

## 2. Metodika identifikace pohybů

### 2.1. Výběr parametrů pro ICA

Algoritmus EFICA má několik vstupních parametrů, mezi kterými jsou například: nelinearita pro počáteční odhad nezávislosti, nelinearita pro super-gaussovské zdroje, nelinearita pro test na sedlové body, velikost stop kritéria pro estimaci nezávislosti, či maximální počet iterací.

Aby šlo výsledky ICA nějakým způsobem porovnat, je třeba definovat objektivní kritéria měření kvality komponent [9]. Pro jejich výpočet jsou použity p hodnoty znaménkového testu počítaného vůči mediánu úrovní ve spektrogramu v referenčním intervalu 4,5 až 3,5 s před začátkem pohybu. Síla ERD jedné komponenty se definuje jako součet těchto hodnot v rozmezí -2 až +2 s okolo počátku pohybu pro frekvence od 5 do 30 Hz. Síla ERS pak jako součet v rozmezí 0 až 5 s pro frekvence 5 až 35 Hz. Hodnoty intervalů jsou zvoleny v souladu s výzkumem vlastností ERD a ERS [11].

Kvalitu dekompozice můžeme posuzovat podle toho, jak je síla ERD a ERS rozložena v jednotlivých komponentách. Snahou je, aby bylo ERD a ERS koncentrováno do několika komponent. Pohybové komponenty se vyznačují silným ERD a ERS. Pokud nepohybové komponenty se slabým ERD a ERS vynulujeme, docílíme tím vyfiltrování elektrických aktivit nesouvisejících s pohybem.

Ukazuje se, že dostupné nelinearity pro estimaci nezávislosti poskytují velice podobné výsledky. Jako nejlepší vyšla z testu nelinearita 'rat2'. Pokud ovšem vezme v potaz i rychlost konvergence algoritmu, ukazuje se jako nejlepší volba nelinearita 'tanh'. Vliv velikosti stop kritéria estimace byl také zkoumán. Rozdíly mezi hodnotami  $10^{-4}$ ,  $10^{-5}$  a  $10^{-6}$  jsou zanedbatelné, byla tedy zvolena hodnota  $10^{-4}$ . Hodnoty ostatních parametrů byly ponechány ve výchozí hodnotě zvolené autorem algoritmu, jelikož byly specificky zvoleny pro daný účel a v minulosti [9] se ukázaly jako vhodné pro dekompozici EEG.

### 2.2. Klastrování

Algoritmy pro výpočet ICA jsou iterativními algoritmy, které jsou závislé na inicializaci. U takovýchto algoritmů je třeba dávat pozor na spolehlivost jejich výsledků. Dva průběhy algoritmu s různou inicializací nedají shodné výsledky. Pokud výpočet neskonal nestandardně (například nebyl překročen definovaný maximální počet iterací), komponenty získané v rámci každého průběhu splňují požadovaná kritéria nezávislosti, napříč průběhy toto již ale neplatí. Z tohoto důvodu je rozumné nechat algoritmus proběhnout několikrát s různými počátečními podmínkami a sledovat vzájemnou podobnost výsledných komponent.

Každé komponentě připadá jeden řádek separační matice  $\mathbf{W}$ , která je inverzní k mixážní matici  $\mathbf{A}$ . Tento řádek určuje lineární kombinaci elektrod, která dá dohromady určitý nezávislý zdroj. Popisuje tedy topologii zdroje na skalpu. Rozložení nezávislé komponenty na skalpu nabízí velice přirozenou cestu, jak jednotlivé komponenty porovnávat.

Řekněme, že budeme opakovat výpočet  $M$  krát a že algoritmus vypočítává čtvercový model, tedy že počet elektrod je stejný jako počet určených nezávislých zdrojů. Můžeme

identifikovat  $M \times N$  vypočtených komponent, respektive  $M \times N$  řádků separačních matic. Přestože výsledky různých běhů nejsou zcela shodné, lze nalézt komponenty z každého běhu, které jsou si velice podobné. Podobné komponenty lze nalézt pomocí shlukování.

Komponentu můžeme nazvat spolehlivou, pokud je součástí klastru, který obsahuje komponenty i většiny ostatních běhů ICA. Takováto komponenta bude nejspíše obsahovat nějaký silný zdroj. Komponenty, které nejsou součástí žádného shluku, nebo jen malého počtu shluků, jsou nestabilní a mohou reprezentovat například šum, nebo jevy, které se dostatečně neprojevují ve zkoumaných datech.

Algoritmus EFICA byl spouštěn nad daty obsahujícími okolí pohybu a zkoumaných klidových intervalů. Délka použitých úseků byla vždy 10 s. Každý úsek byl před vstupem do algoritmu filtrován pomocí FFT pásmové propusti od 5 do 40 Hz a následně decimován faktorem 2. Všechny úseky byly pro každou osobu spojeny a použity jako vstupní data algoritmu. Algoritmus byl spuštěn 50x s různou inicializací separační matice. Při studiu výsledných komponent je třeba brát v potaz jednu ze základních nejednoznačností ICA, měřítko. Separační matice je hledána tak, aby nalezené komponenty měli jednotkový rozptyl. To má však za následek nejednoznačnost znaménka. Při hledání shluků musí být tedy počítáno jak s maticí  $\mathbf{W}$ , tak s maticí  $-\mathbf{W}$ .

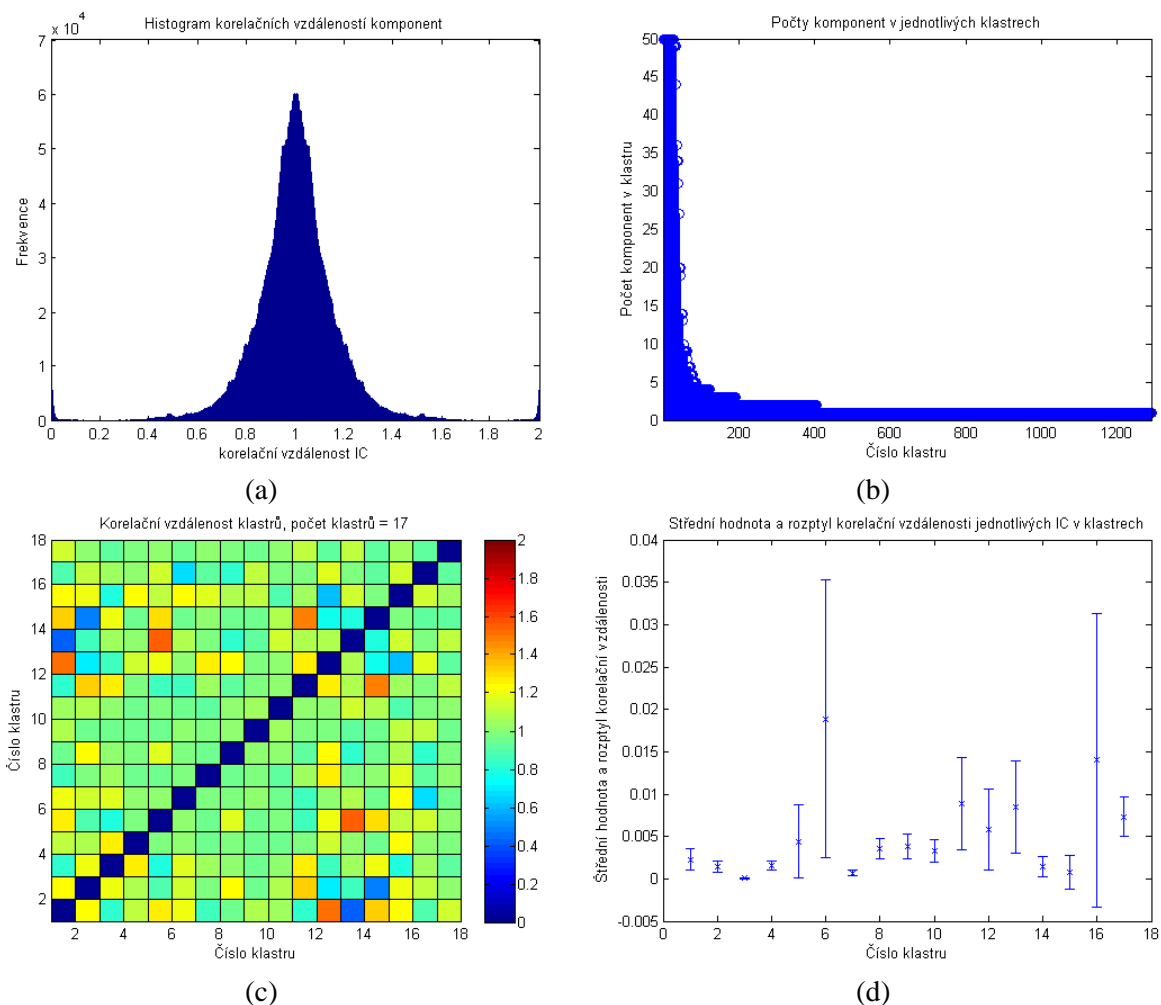
Existuje mnoho shlukovacích algoritmů. Mezi nejznámější patří samo-organizující se mapy (SOM), k-means, či hierarchické klastrování. K-means algoritmus není pro účel shlukování nezávislých komponent příliš vhodný, jelikož vyžaduje předem známý počet klastrů  $K$ . To je však jeden z parametrů, které se snažíme zjistit. Hierarchické klastrování má oproti SOM výhodu v interpretovatelnosti dílčích výsledků. Hierarchické shlukování tvoří binární strom, takzvaný dendrogram, kdy klastry jsou děleny na základě metriky tak, že klastry indikované nižší úrovní stromu jsou vždy součástí klastru indikovaného vyšší úrovní. Tento strom lze pak uříznout na nějaké úrovni podle toho, kolik chceme získat shluků, či jako v našem případě jaké vlastnosti klastru chceme.

Shluky tvoříme na základě korelační vzdálenosti jednotlivých komponent. Použitá aglomerační strategie se nazývá average linkage. Tato strategie se jeví jako nevhodnější a často se používá při shlukování nezávislých komponent [4, 10]. Klastry jsou tvořeny tak, aby průměrná korelační vzdálenost každých dvou komponent v jakýchkoliv dvou klastrech byla vždy větší než 0,1. Korelační vzdálenost je definována jako:

$$d_{corr} = 1 - \frac{(\mathbf{w}_i - \bar{\mathbf{w}}_l)(\mathbf{w}_k - \bar{\mathbf{w}}_k)^T}{\sqrt{(\mathbf{w}_i - \bar{\mathbf{w}}_l)(\mathbf{w}_i - \bar{\mathbf{w}}_l)^T} \cdot \sqrt{(\mathbf{w}_k - \bar{\mathbf{w}}_k)(\mathbf{w}_k - \bar{\mathbf{w}}_k)^T}},$$

kde  $\mathbf{w}_i$  a  $\mathbf{w}_k$  jsou řádky matice  $\mathbf{W}$  reprezentující jednotlivé komponenty a  $\bar{\mathbf{w}}_l$  a  $\bar{\mathbf{w}}_k$  jsou aritmetické průměry všech prvků vektorů  $\mathbf{w}_i$  a  $\mathbf{w}_k$ .

Na obrázku 1 jsou znázorněny výsledky klastrování pro osobu 3. V (a) je zobrazena četnost korelačních vzdáleností mezi spočtenými komponentami. V oblasti kolem 0 a 2 jsou vidět vzdálenosti mezi komponentami tvořícími shluky. V (b) je vidět počet komponent připadajících na klastr. Klastry s alespoň 48 komponentami odpovídají stabilní komponentě. Zde je třeba mít na paměti, že počet zobrazených stabilních shluků je dvojnásobný, jelikož každá komponenta je použita dvakrát s různým znaménkem. (c) a (d) ukazují vlastnosti stabilních shluků. Vidíme, že většina shluků je silně nekorelovaná a jen několik málo se vyznačuje slabou korelací. V (d) jsou vidět vlastnosti odhadu jednotlivých komponent, kdy například klastr 3 obsahuje téměř identické komponenty a například shluk 6 je více rozptýlený.



**Obr. 1 – Klastrování komponent pro osobu 3**

Zde je vidět velká síla shlukování výsledků ICA. Vidíme nejen, které komponenty jsou stabilní a spolehlivé ale můžeme získat také jejich lepší odhad. Ten získáme tak, že za správný odhad nezávislé komponenty považujeme centroid shluku, tedy komponentu, která má nejmenší korelační vzdálenost od všech ostatních v daném shluku.

Tabulka 1 ukazuje, kolik shluků komponent (po vyřazení duplicitních) bylo nalezeno pro jednotlivé osoby, kolik z nich mělo dostatečně mnoho prvků, aby mohli být považovány za spolehlivé a kolik z těch spolehlivých vykazovalo pohybovou aktivitu.

Pro zajímavost je zajímavé se podívat, jak se mění počet stabilních komponent v závislosti na výběru vstupních dat pro ICA. ICA spočtená pouze z částí nahrávky EEG obsahujících klasifikované úseky a má pro většinu osob menší počet stabilních komponent a zhruba stejné, nebo a trochu menší množství pohybových komponent, než ICA spočtená, z celé nahrávky, která obsahuje i úseky s artefakty a klidu mezi jednotlivými realizacemi pohybu.

**Tab. 1 – Počty stabilních a pohybových komponent**

Osoba:	1	2	3	4	5	6	7	8
Počet shluků:	151	206	645	735	133	167	575	82
Počet stabilních komponent:	24	18	17	12	23	22	10	31
Z nich pohybových:	10	11	8	6	5	11	5	13



### 2.3. Výběr pohybových komponent

Pohybové komponenty mohou být identifikovány na základě jejich topologie, nebo jejich časově-frekvenčním vývoji. Rozložení komponent na skalpu se v [9] ukázalo jako velice vhodné pro odstranění nehybových komponent, kdy například komponenty v blízkosti očí byly automaticky odstraněny. Tento přístup však není možný pro studovanou databázi, jelikož jsou všechny elektrody umístěny na centrální části skalpu, nad pohybovým centrem mozku.

Pro klasifikaci komponent se tedy pro databázi palec-malíček používá pouze spektrogramů komponent pro každou osobu rozdělených po realizacích pohybu. V první fázi se odstraní komponenty, které obsahují známé artefakty. V dalším kroku se zpracují spektrogramy pomocí znaménkového testu vztaženému k mediánu referenčního intervalu (4,5 – 3,5 s před pohybem) všech realizací, tak, že každé kombinaci časového okna a frekvenční čáry je přiřazena p hodnota znaménkového testu opatřená znaménkem podle toho, zda pro danou kombinaci došlo v průměru k poklesu (minus), či ke zvýšení (plus) aktivity (viz Obr. 1a). Znaménkový test se následně vyhodnotí na hladině významnosti 0,05 a na celý vývoj výsledků se aplikuje mediánový filtr pro odstranění malých seskupení a zvýraznění těch velikých.

Počet statisticky významných poklesů v a statisticky významných zvýšení v okolí pohybu se používá pro klasifikaci komponenty. Pouze komponenty vybrané jako pohybové jsou promítnuty zpět do elektrod, které jsou následovně použity pro klasifikaci pohybů. Počet komponent, které byly shledány pohybové je uveden v Tab. 1. Detailnější popis automatického klasifikátoru lze nalézt v [12].

### 2.4. Klasifikační systém

Klasifikační systém na bázi skrytých markovských modelů (Hidden Markov Models - HMM), který byl vyvinut v naší laboratoři je použit pro klasifikaci pohybů. HMM mají několik výhod [1]. Využívají informaci o tom, jak se pohybové EEG mění v čase. Jsou schopny přiřadit skryté psychologické procesy v mozku pohybu a mohou modelovat pohybové EEG. V neposlední řadě jdou jejich výsledky, na rozdíl od některých druhů např. neuronových sítí, poměrně dobře interpretovatelné.

Modely použité pro klasifikaci obsahují 4 emitující stavy a mají levo-pravou strukturu. Stavy vyjadřují klid – desynchronizaci – synchronizaci – klid. Lineární FFT koeficienty v rozmezí 5 – 40 Hz v intervalu  $\pm 5$  s okolo začátku pohybu jsou použity jako vstupní parametry pro klasifikaci. Jednotlivé realizace pohybu jsou rozděleny na trénovací (75%) a testovací (25%) podmnožiny. Trénování a testování je opakováno 20x za použití čtyřnásobné krosvalidace.

### 2.5. Výsledky klasifikace

V tabulce 2 jsou uvedeny výsledky klasifikace. Použití ICA a shlukování komponent přineslo zlepšení klasifikace oproti nefiltrovanému EEG, nepodařilo se avšak zlepšit výsledky oproti filtrování pomocí laplacovského prostorového filtru.

**Tab. 2 – Výsledky klasifikace**

Osoba	Typ prostorového filtru	Klasifikační skóre [%]				Elektróda
		Malíček	Palec	Klid	Průměr	
1	-	38.8±12.2	54.1±15.4	86.1±7.3	59.7	8
	8-neighbour Laplacian	44.0±21.9	56.1±24.6	86.4±8.6	62.2	18
	ICA + shlukování	56.0±10.1	50.9±11.2	79.4±7.8	62.1	17
2	-	65.6±10.1	48.8±13.3	96.2±5.5	70.2	17
	8-neighbour Laplacian	82.6±11.1	39.6±11.4	97.4±4.3	73.2	17
	ICA + shlukování	81.8±11.2	40.4±13.0	94.7±6.2	72.3	17
3	-	40.9±17.0	51.1±13.7	78.4±12.5	56.8	25
	8-neighbour Laplacian	36.1±20.3	64.7±27.0	86.5±6.5	62.4	25
	ICA + shlukování	41.9±17.6	67.4±12.8	86.0±7.5	65.1	25
4	-	37.5±9.2	44.5±12.5	73.0±10.0	51.7	7
	8-neighbour Laplacian	50.3±12.0	54.3±10.8	77.0±7.7	60.5	1
	ICA + shlukování	46.7±15.9	38.5±16.8	55.7±8.2	47.0	27
5	-	50.0±14.9	41.9±15.0	68.7±10.0	53.5	28
	8-neighbour Laplacian	48.7±19.3	38.3±16.8	85.6±7.5	57.5	19
	ICA + shlukování	34.2±16.2	57.4±16.5	80.5±10.2	57.4	16
6	-	47.6±14.8	47.1±12.8	64.6±12.0	53.1	14
	8-neighbour Laplacian	48.9±8.8	44.6±12.1	82.1±8.6	58.5	25
	ICA + shlukování	50.9±14.5	44.1±14.0	78.7±8.3	57.9	25
7	-	56.6±15.6	50.8±16.8	71.1±11.0	59.5	18
	8-neighbour Laplacian	56.6±16.8	51.4±17.4	86.2±6.6	64.7	18
	ICA + shlukování	42.6±13.5	57.3±15.4	83.6±5.4	61.2	28
8	-	43.1±15.3	41.6±12.1	61.7±11.3	48.8	14
	8-neighbour Laplacian	67.2±9.4	72.7±11.3	86.9±6.1	75.6	20
	ICA + shlukování	52.6±14.3	40.1±11.0	74.8±8.2	55.8	38

### 3. Závěr

Porovnáme-li dílčí výsledky jednotlivých kroků pro distální/proximální databázi a pro palec/malíček/ticho, lze pozorovat velikou podobnost výsledků až na finální klasifikaci pohybů. Jednotlivé komponenty jde podle dobře dokumentovaných znaků rozdělit na pohybové a nepohybové a po odstranění těch nepohybových zle pozorovat zlepšení kontrastu na průměrovaných spektrogramech pohybů.

Počet spolehlivých komponent se snižuje s tím, jak omezujeme data, a to jak ve frekvenční oblasti, tak v časové oblasti. Omezení se pouze na zkoumané úseky dat však výrazně nesnižuje počet nalezených pohybových komponent.

Při zkoumání nestabilních komponent bylo objeveno několik (řádově méně než nepohybových) pohybových komponent, avšak nebyla zjištěna žádná závislost typu komponenty na velikosti klastru. Vliv těchto komponent na klasifikaci pohybů není zatím znám a jistě si zaslouží bližší průzkum.

V budoucnu je naplánováno vyzkoušet jiné algoritmy pro slepou separaci zdrojů, které by mohli být schopny lépe rozložit EEG, například Denoising Source Separation.

### Poděkování

Výzkum popsany v tomto příspěvku byl podporován interním grantem ČVUT, SGS12/143/OHK3/2T/13 "Algoritmy a hardwarové realizace číslicového zpracování signálů".

## Reference

- [1] ŠŤASTNÝ J., SOVKA P. High-resolution movement EEG classification. *Intell. Neuroscience*, 2007:5:1–5:12, srpen 2007.
- [2] KOLDOVSKÝ Z., TICHAVSKÝ P., OJA E. Efficient variant of algorithm FastICA for independent component analysis attaining the Cramér-Rao lower bound. *IEEE Transactions on Neural Networks*, 17(5):1265–1277, 2006.
- [3] HYVÄRINEN A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, 1999.
- [4] HIMBERG J., HYVARINEN A., ESPOSITO F. Validating the independent components of neuroimaging time series via clustering and visualization. *NeuroImage*, 22(3):1214–1222, červenec 2004.
- [5] GROPE D. M., MAKEIG S., KUTAS M. Identifying reliable independent components via split-half comparisons. *NeuroImage*, 45(4):1199 – 1211, 2009.
- [6] KONOPKA O., ŠŤASTNÝ J., SOVKA P. Movement-related EEG separation using independent component analysis. In *3rd European Medical and Biological Conference on Biomedical Engineering*, s. 250–255, 2005.
- [7] RUČKAY L., ŠŤASTNÝ J., SOVKA P. Movement-related EEG decomposition using independent component analysis. V *Analysis of Biomedical Signals and Images, 18-th Biennial International EURASIP Conference BIOSIGNAL 2006*, s. 78–80, červen 2006.
- [8] RUČKAY L., ŠŤASTNÝ J., SOVKA P. Selection and classification of eeg movement-related independent components. In *Analysis of Biomedical Signals and Images, 19-th Biennial International EURASIP Conference BIOSIGNAL*, 2008.
- [9] RUČKAY L. EEG decomposition for simple movement classification. *dizertační práce, České vysoké učení technické v Praze*, říjen 2010.
- [10] RUČKAY L., ŠŤASTNÝ J., SOVKA P. ICA model order estimation using clustering method. *Radioengineering*, 16(4):51–57, 2007.
- [11] STANČÁK A., FEIGE B., LÜCKING C. H., KRISTEVA-FEIGE R. Oscillatory cortical activity and movement-related potentials in proximal and distal movements. *Clin Neurophysiol*, 111(4):636–650, duben 2000.
- [12] DOBIÁŠ M. Application of Independent Component Analysis in the BCI field. *diplomová práce, České vysoké učení technické v Praze*, květen 2011.

# Lokalizace zdrojů epileptiformních výbojů při peroperační kortikografii

Radek JANČA<sup>1</sup>, Petr MARUSIČ<sup>2</sup>, Roman ČMEJLA<sup>1</sup>, Petr JEŽDÍK<sup>1</sup>

<sup>1</sup> České vysoké učení v Praze, Fakulta elektrotechnická, Katedra teorie obvodů

<sup>2</sup> Univerzita Karlova v Praze, 2. lékařská fakulta, Neurologická klinika

[jancarad@fel.cvut.cz](mailto:jancarad@fel.cvut.cz)

**Abstrakt:** Článek se zabývá možnostmi detekce a analýzy epileptiformních výbojů v peroperační kortikografii. Zařazení jednotlivých výbojů do klastrů umožňuje definovat oblasti, ve kterých výboje vznikají, a které jsou zodpovědné za vznik epileptických záchvatů. Lokalizace epileptiformních ložisek během operace může sloužit jako diagnostický nástroj pro určení rozsahu resekcí oblasti při epileptochirurgické léčbě.

## 1. Úvod

Mozkové postižení s paroxysmálními symptomy se nazývá epilepsie, kterým trpí asi procento lidské populace. Mozky pacientů generují abnormální elektroencefalografickou (EEG) aktivitu v místech, která jsou zodpovědná za vznik záchvatových stavů. Neurologové vizuálně hledají patologické artefakty v EEG k přesné lokalizaci tzv. epileptogenní zóny. Epileptogenní zóna většinou reprezentuje část mozku obklopující primární patologii (nádor, nekróza, tuber, skleróza, dysplazie, trauma aj.). Záchvaty nezpůsobuje primární patologie, nýbrž utlačovaná nebo narušená tkáň v okolí. Farmakorezistentní pacienti proto podstupují epileptochirurgickou léčbu, při které je jim odstraněna část poškozeného mozku.

Jeden z důležitých ukazatelů patologické aktivity je výskyt epileptiformních výbojů (Spike-Wave Complex – SWC) v interiktálním EEG. Výboje mají charakter rychlých tranzientů trvajících mezi 20-70 ms jasně vystupujících nad aktivitu pozadí [1]. Ve skalpovém EEG jsou výboje zkresleny a silně tlumeny, protože lebku lze chápat jako dolní propust s vysokou impedancí. Frekvence nad 30 Hz jsou potlačeny a dochází k rozmazávání signálů, takže lokalizovat přesný zdroj vzniku výboje je velmi obtížné. Z tohoto důvodu se přechází k peroperační kortikografii (ECoG), kdy je snímána EEG aktivita přímo na povrchu mozku během operace. Signály jsou snímány standardní hlavici za použití elektrod umístěných na silikonových páscích (stripech).

Výskyt patologických výbojů může být současně i ve vzdálenějších oblastech mozku, protože přirozené nervové struktury jsou navzájem funkčně propojeny. Nicméně výboje jsou spouštěny lokálně a dále se šíří tkání nebo nervovými strukturami. Místo vzniku často koresponduje s epileptogenní zónou. Z tohoto důvodu nemusí pouhé kvantitativní hodnocení výskytu výbojů v jednotlivých kanálech ukazovat na konkrétní ložiska. Při chirurgické léčbě epilepsií temporálního laloku je peroperační monitorace důležitým ukazatelem rozsahu resekce. Cílem lékařů je odstranit celé epileptiformní ložisko a přitom pacienta co nejméně poškodit. Obr. 1 ukazuje tři nejběžnější rozsahy resekce. V současnosti neurochirurg po částech resekuje temporální lalok. Po každém kroku je prováděna monitorace za účelem zjištění, zdali ponechaná tkáň stále generuje výboje. Pokud tomu tak je, chirurg rozšíří resekovanou oblast. Nicméně tento postup vyžaduje přítomnost zkušeného neurologa pro hodnocení peroperačních ECoG signálů. Traumatizovaná tkáň při zákroku generuje taktéž

abnormální aktivitu, která svým charakterem připomíná epileptiformní výboje. Z tohoto důvodu mnoho pracovišť od ECoG monitorace upouští.

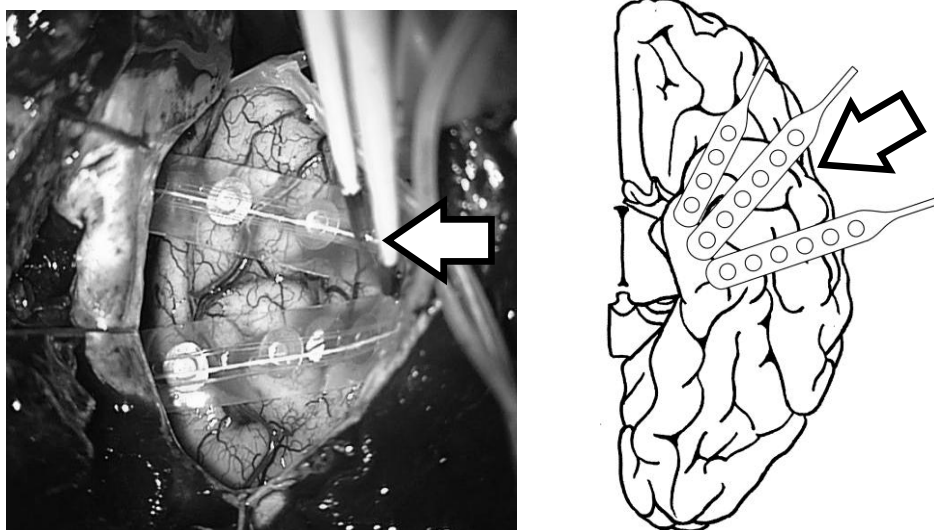
Naším cílem je analyzovat peroperační ECoG signál před samotným resekcčním zákrokem a lokalizovat zdroje výbojů při prvotní monitoraci. Výsledky by měly určit rozsah resekce bez nutnosti další monitorace, což by zkrátilo délku operace, tím se snížilo riziko pro pacienta a taktéž zjednodušilo práci celého operačního týmu.



**Obr. 1: Typické rozsahy resekcí temporálního laloku (kraniální pohled)**

## 2. Materiál a metody

Data byla nahrána na pacientech z oddělení neurologie ve Fakultní nemocnici Motol. Diagnóza analyzovaných pacientů byla epilepsie temporálního laloku. Data byla snímána peroperačně ECoG elektrodami. Operátor z malého operačního vstupu zasouvá elektrody naslepo pod mozkové pleny na kortex tak, aby rozložení elektrod odpovídalo vždy stejnému schématu, viz obr. 2. Následuje prvotní pětiminutový záznam 16 kanálů ECoG při vzorkovacím kmitočtu 250 Hz.

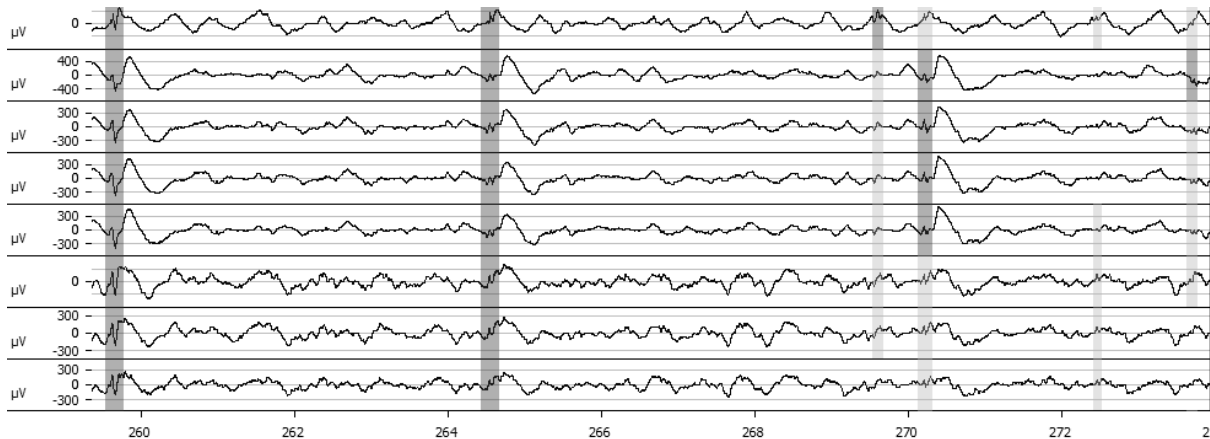


**Obr. 2: Rozložení ECoG elektrod při peroperační monitoraci**

## 3. Analýza epileptiformních výbojů

Pro detekci výbojů je využíván vlastní detektor oddělující základní (baseline) aktivitu a epileptiformní výboje na základě statistické distribuce energetické obálky filtrovaných

signálů [2]. Dříve publikovaný detektor byl upraven tak, aby rozděloval signál do tříd *výboj/nejasný výboj/baseline*. Původní rozhodovací práh byl nahrazen dvojicí prahů. Prahy jsou nově počítány adaptivně pro pětisekundové úseky signálu se sekundovým překryvem. *Nejasný výboj* je akceptován jen tehdy, pokud se ve stejný okamžik vyskytuje *výboj* v jiném kanále. Na obr. 3 je úsek multikanálového ECoG záznamu s vyznačenými výboji.



**Obr. 3: Příklad ECoG signálu s vyznačenými výboji**

Jak již bylo zmíněno v úvodu, celkové kvantitativní hodnocení výbojů v jednotlivých kanálech nemusí ukazovat na konkrétní ložiska. Snahou je tedy rozřadit výboje do podskupin (klastřů), které by prokázaly, zdali výboje vycházejí z jednoho místa (malá resekce) nebo vycházejí z různých částí (rozsáhlá resekce).

### 3.1. Klastrování

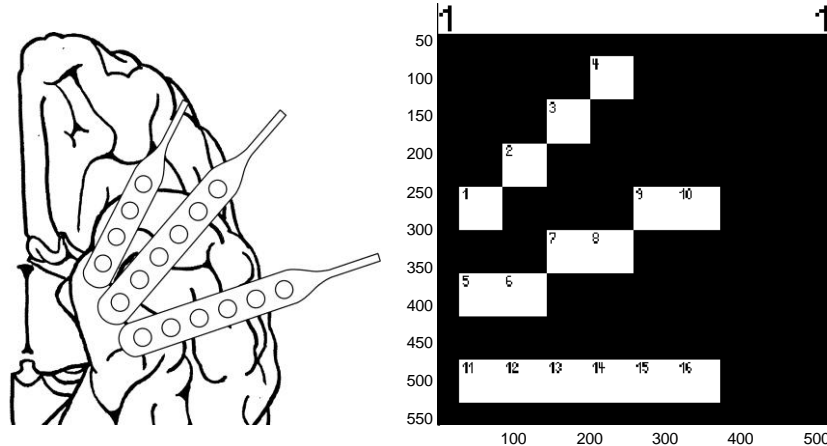
Událost, při které se s minimálním zpožděním objeví výboje v několika kanálech, lze parametrizovat pomocí vektoru. Prvky vektoru odpovídají kanálům; hodnoty prvků reprezentují typ signálu: 1 – *výboj*, 0.5 – *nejasný výboj*, 0 – *baseline* aktivita. Jednotlivé události tvoří prvky 16 dimenzionálního prostoru, ve kterém je snaha nalézat shluky (klastry). Byly testovány samoshlukové algoritmy typu k-mean, EM (Expectation–maximization), nicméně se ukázaly jako nepraktické. Důvodem byla vysoká dimenze prostoru s malým počtem zjištěných realizací. V pětiminutovém signálu byly detekovány desítky, maximálně stovky událostí, které nestačily k dostatečnému zahuštění prostoru. Redukce dimenze např. pomocí Kernelovského PCA (Principal component analysis) z principu problému nemůže fungovat. Pokud by byly výskyty výbojů pouze v jednotlivých kanálech, po redukci dimenze by samoshlukové algoritmy zařadily všechny události do společného klastru, protože společným parametrem by byla „absence výbojů v ostatních kanálech“. Z tohoto důvodu byl vytvořen vlastní klastrovací algoritmus založený na podobnosti událostí.

Shlukování je založeno na výpočtu korelačního koeficientu a testování hypotézy o nekorelaci. U dat se nedá předpokládat normální rozložení, proto je využívána Spearmanova výpočetní metoda [3]. V prvním kroku je spočtena vzájemná korelace všech událostí, čímž je získána symetrické korelační matice. Pro každý sloupec matice je spočten počet prvků, které mají kladný korelační koeficient a hypotézu o nekorelaci pod kvantilem 0.01. Sloupec matice, který obsahuje největší počet prvků splňující podmínku je označen jako rodič klastru. Všechny prvky splňující podmínku rodičovského sloupce jsou označeny za potomky a spadají do téhož klastru. Události vzniklého klastru jsou z celkové matice korelací vyřazeny. Algoritmus vytváří nové klastry stejným postupem do doby, než jsou všechny události rozříděny.

## 4. Výsledky

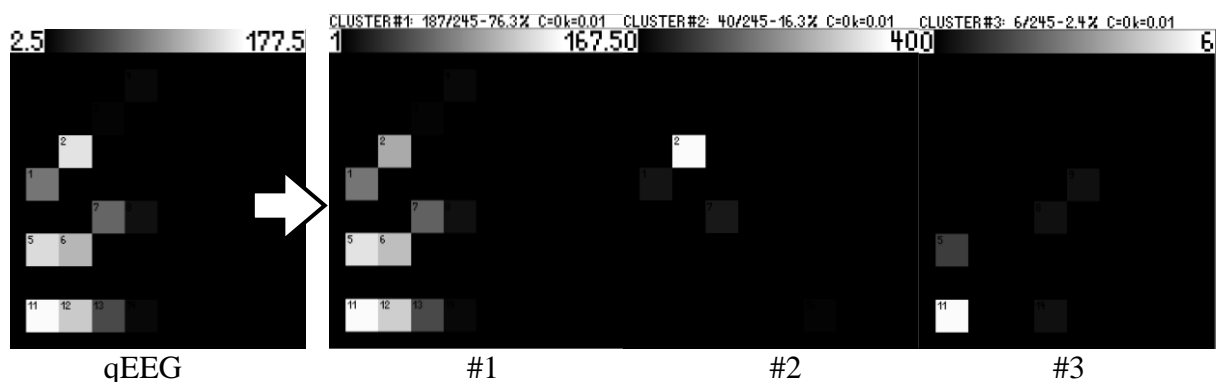
Pro předběžné výsledky byla testována omezená skupina šesti pacientů. Skupina obsahovala i pacienty s rušenými signály a výpadky celých kanálů apod. Vadné kanály byly z analýz vyjmuty. V signálech byly detekovány desítky, maximálně však stovky událostí.

Výsledky jsou zobrazeny barevnou škálou v kortikální mapě. Kortikální mapa reprezentuje fyzické rozložení elektrod, viz obr. 4. Kvantitativní hodnocení odpovídá součtu vah událostí jednotlivých klastrů v jednotlivých kanálech (1 – výboj, 0,5 – nejasný výboj).



Obr. 4: Vizualizace elektrod v kortikální mapě

Pro příklad jsou uvedeny výsledky dvou pacientů s odlišným rozsahem resekce. Pro názornost byly vybrány tři nejpočetnější klastry. U prvního pacienta největší tři klastry obsahují 95 % všech událostí, u druhého pacienta 57.1 %. Celkové kvantitativní hodnocení výbojů (qEEG) a rozdělení do klastrů prvního pacienta ukazuje obr. 6, druhého obr. 7.



Obr. 6: Rozklastrování výbojů – první pacient



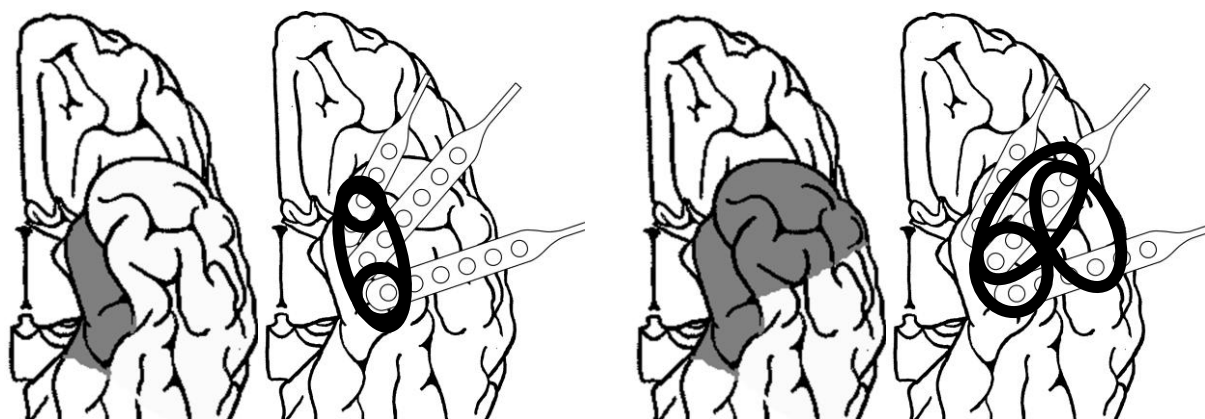
Obr. 7: Rozklastrování výbojů – druhý pacient

## 5. Diskuze

U prvního i druhého pacienta lze pozorovat největší výskyt výbojů v blízkosti parahipokampálního gyru (elektroda 1, 5, 6, 11 a 12). Nicméně až klastry ukazují na jasné rozdíly.

U prvního pacienta 80 % výbojů vychází z oblasti parahipokampálního gyru, ostatní výboje byly detekovány pod elektrodou 2. Vzhledem k slepému zavedení elektrod může i elektroda dva náležet na stejný gyrus. Dle vyšetření MRI se v této oblasti nacházel nádor. Neurologický nález se zcela shodoval s výsledkem analýzy. Při operaci byla zvolena nejmenší oblast resekce odpovídající zmíněnému gyru, obr. 8 vlevo.

Druhý pacient byl postižen dysplazií a hipokampální sklerózou. Analýzy ukázaly na více zdrojů výbojů, které zasahovaly do oblastí gyru occipito-temporalis lateralis (7, 8, 13, 14), temporalis medius a inferior (9, 10, 15, 16) a do pólu laloku (2, 3, 4). Výsledky tedy směřují ke kompletní resekci taktéž ve shodě s lékaři. Obr. 8 vpravo znázorňuje přibližný výskyt klastrů a resekovaných oblastí.



**Obr. 8: Orientační porovnání rozsahu resekce a lokalizace klastrů (první pacient vlevo, druhý vpravo).**

Výsledky ostatních pacientů vycházely v dobré shodě s neurology. Nicméně s přihlédnutím k velikosti testovaného souboru pacientů je předčasné hodnotit úspěšnost metody. Předběžné výsledky však vypadají dobře a poskytují základ pro další metody.

## Poděkování

Tento projekt je podporován granty IGA NT11460-4/2010 Intracranial EEG signal processing; epileptogenic zone identification in non-lesional refractory epilepsy patients, SGS 10/272/OHK4/3T/13 Analysis of intracranial EEG recording a výzkumným programem MSM6840770012 Transdisciplinary Research in Biomedical Engineering.

## Reference

- [1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529–551, April 1955. (references)
- [2] Janca R., Cmejla R. Jahodova A. Rules for Spike Detection in Multichannel Intracranial Electroencephalography - *19th Annual Conference Proceedings Technical Computing Prague 2011*, Prague, 2011.
- [3] Kendall, M.G. *Rank Correlation Methods*, Griffin, 1970



# Zpracování povrchových elektromyografických signálů

Daniel ŠPULÁK<sup>1</sup>

<sup>1</sup> České vysoké učení v Praze, Fakulta elektrotechnická, Katedra teorie obvodů

[spuladan@fel.cvut.cz](mailto:spuladan@fel.cvut.cz)

*Abstrakt:* Náš výzkum je zaměřen na zpracování povrchových elektromyografických signálů. V této oblasti spolupracujeme s Fakultou tělesné výchovy a sportu (FTVS) Univerzity Karlovy v Praze, konkrétně s Katedrou sportů v přírodě.

Předmětem výzkumu kolegů z FTVS jsou rozbor pohybu během různých sportovních aktivit. Jedná se například o karate (úder naprázdno, do pytle, s expandérem či bez něj), lyžování, jízdu na rotopedu (s různými druhy treter), chůze bez holí či s holemi (Nordic walking, v různě skloněném terénu), chůze a běh (s botami a bez bot).

Ve všech případech se nahrává vícekanálový elektromyografický signál (EMG, signál elektrické aktivity svalů), a to výhradně pomocí povrchových (nalepovacích) elektrod. Ty jsou umístěny na pokožce nad sledovanými svaly a propojeny s přenosným záznamovým zařízením, které podle potřeby nahrává až 16 kanálů EMG.

Cílem je zejména sestavení grafů, z nichž je patrné pořadí svalových aktivit a deaktivací. V nich se pak hledají shody a odlišnosti u různých subjektů nebo různých pohybů.

Naše aktivity se zaměřují na tři problémy: segmentaci signálu z periodického pohybu, detekci nástupu a konce svalové aktivity a způsob prezentace výsledků s přihlédnutím k odlišnostem v provedení jednotlivých opakování pohybu.

Je-li k dispozici záznam akcelerometrického nebo polohového snímače, odvozuje se od něj segmentace na jednotlivé periody pohybu. Obtížnější situace nastává, je-li k dispozici pouze záznam EMG. Naše výsledky ukazují, že při výběru vhodného kanálu EMG lze v nezbytných případech dosáhnout výsledků srovnatelných se segmentací založenou na akcelerometrické segmentaci.

K detekci nástupu a konce svalové aktivity používáme obálku EMG (tedy usměrněné EMG filtrované dolní propustí). Dobrých výsledků lze dosáhnout detekováním minima a maxima každé periody, přičemž za počátek aktivity se považuje takový bod na obálce EMG, který společně s minimem a maximem utvoří trojúhelník o maximálním obsahu. Předmětem výzkumu je srovnávání s jinak definovanými body trojúhelníka, případně zcela jinak řešené metody detekce nástupu.

Při zpracování periodických pohybů se přikláníme k detekci uvedených bodů v každé periodě zvlášť, přestože obvyklým postupem je průměrování obálek z jednotlivých period a detekce prováděná na průměrovaném průběhu. Jak jsme ukázali na simulovaných i měřených signálech, využití průměrované obálky EMG přináší určitá rizika, neboť realizace obálky se v jednotlivých periodách může nezanedbatelně lišit a průměrovaný průběh nemusí být dostatečně reprezentativní.

## Zpracování povrchových elektromyografických signálů

doktorand: Ing. Daniel Špulák  
vedoucí: doc. Ing. Roman Čmejla, CSc.  
K13131  
5/2012



## Elektromyografie

- snímání a vyhodnocení elektrické aktivity svalů
- elektromyografický/myoelektrický (EMG) signál snímáný pomocí povrchových nebo jehlových elektrod
- analýza v časové a frekvenční oblasti

2

## Využití elektromyografie

- v medicíně
  - diagnostika neuromuskulárních chorob a poruch motorického systému
  - řízení umělých končetin
- ve sportu
  - analýza svalové únavy
  - tréninkové programy založené na detekci prahu únavy
  - detailní analýza pohybů
- v komerční sféře
  - rozhraní člověk/stroj u počítačových her

3

## Experimenty

- spolupráce s katedrou sportů v přírodě (Univerzita Karlova v Praze, fakulta tělesné výchovy a sportu)
- výhradně povrchové EMG
- studium činnosti jednotlivých svalů během pohybu při sportu

4

## Experimenty – sportovní aktivity

- karate
  - údery naprázdno, do pytle, s expandérem či bez něj
- lyžování
- jízda na rotopedu
  - různé druhy treter
- chůze
  - bez holí či s holemi (Nordic walking)
  - různý sklon terénu
- běh
  - s botami či bez bot

5

## Experimenty – signály

- přenosný nahrávací aparát ME 6000
- až šestnáctikanálové EMG (sledování celých svalových skupin)
- možnost připojení akcelerometrického či polohového snímače

6

## Experimenty – cíle

- sestavení grafů znázorňujících pořadí svalových aktivací a deaktivací během pohybu
- porovnání pořadí a časových rozestupů svalových aktivací a deaktivací
  - u různých osob při provádění téže aktivity
  - u stejné osoby při různých aktivitách

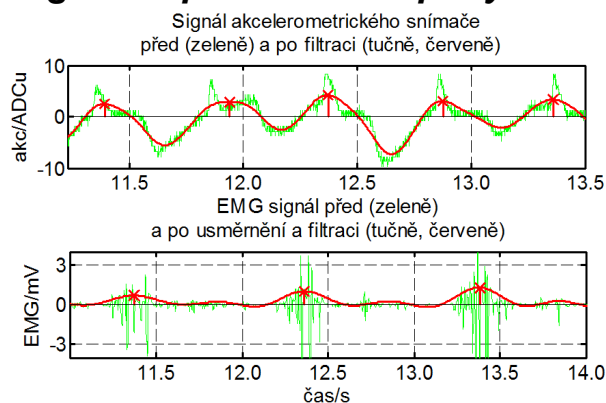
7

## Řešené problémy – segmentace signálu z periodického pohybu

- segmentace signálu na jednotlivé periody pohybu
- odvození segmentace
  - ze signálu akcelerometru nebo polohového snímače
  - ze signálu EMG, pokud výše uvedené signály nejsou k dispozici
- porovnání různých způsobů segmentace

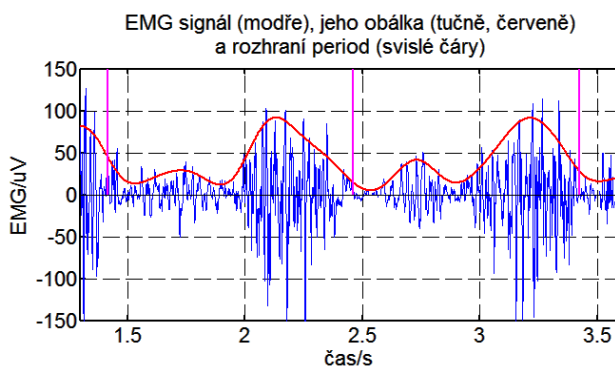
8

## Řešené problémy – segmentace signálu z periodického pohybu



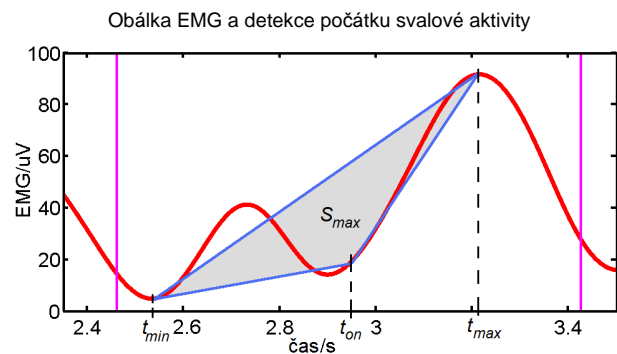
9

## Řešené problémy – detekce počátku a konce svalové aktivity



10

## Řešené problémy – detekce počátku a konce svalové aktivity



11

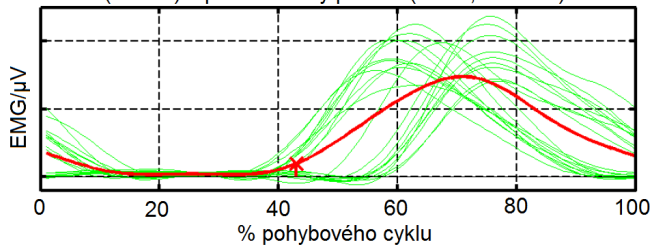
## Řešené problémy – vliv průměrování a prezentace výsledků

- Je vhodné provádět detekci počátku a konce svalové aktivity
  - na průběhu vzniklém zprůměrováním obálek ze všech period pohybu
  - nebo v každé periodě pohybu zvlášť a případně zprůměrovat získané hodnoty?

12

## Řešené problémy – vliv průměrování a prezentace výsledků

Interpolované obálky EMG z jednotlivých period pohybu (zelené) a průměrovaný průběh (tučně, červeně)

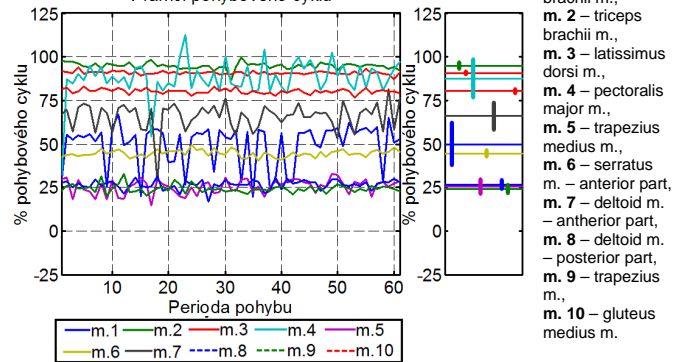


- užití průměrované křivky nemusí poskytnout reprezentativní výsledky

13

## Řešené problémy – vliv průměrování a prezentace výsledků

Poloha nástupu svalové aktivity v rámci pohybového cyklu



14

## Závěry

- segmentaci na periody pohybu lze v nutných případech odvodit z EMG
- detekce počátků a konců svalové aktivity je nejednoznačná a vhodné způsoby se u jednotlivých pohybů a svalových skupin liší
- používání průměrovaného průběhu obálky je nevhodné
- vyhodnocování jednotlivých pohybových cyklů zvláště je přesnější, ale náročnější na zobrazení výsledků

15

**Děkuji za pozornost!**

# Kompaktní senzorický systém pro sběr biologických signálů

Matouš POKORNÝ<sup>1</sup>, Jan HAVLÍK<sup>1</sup>

<sup>1</sup> České vysoké učení v Praze, Fakulta elektrotechnická, Katedra teorie obvodů

[pokormat@fel.cvut.cz](mailto:pokormat@fel.cvut.cz)

*Abstrakt:* Tento článek pojednává o kompaktním senzorickém systému pro sběr biologických signálů se zaměřením na asistivní technologie a telemedicínské aplikace. Senzorický systém je určen pro dlouhodobý sběr biologických signálů v běžném životě pacienta, hendikepované osoby či seniora. Základním požadavkem na senzorický systém je věrohodnost naměřených dat a komfort uživatele. Na základě získaných signálů by mělo být možné indikovat a predikovat krizové stavy monitorované osoby, jako např. srdeční arytmie, zástava dechu, pád osoby na zem a podobně. V článku je uveden návrh řešení takového to senzorického systému a diskuze možných problémů při realizaci a aplikaci.

## 1. Úvod

Rychlý vývoj elektroniky a techniky obecně umožňuje konstrukci stále dokonalejších diagnostických a monitorovacích lékařských přístrojů. Tyto přístroje pomáhají lékařům a zdravotnímu personálu při diagnóze, léčbě a zdravotnické péči převážně v lékařských zařízeních, nemocnicích a klinikách. Zde jsou rizikové osoby, u kterých je vyšší pravděpodobnost ohrožení zdraví či života, jako jsou pacienti s chronickým onemocněním, hendikepované osoby a senioři pod dohledem a v případná riziková situace je okamžitě indikována a řešena. Popsaný projekt má pomoci implementovat jednoduché dohledové a diagnostické funkce (kontrola dechu, srdečního rytmu, fyzické aktivity) do běžného života rizikových osob tak, aby se bez většího omezení mohli věnovat svým běžným činnostem. Senzorický systém tvoří část mobilní jednotky, kterou má monitorovaná osoba neustále u sebe. Pokud mobilní jednotka vyhodnotí daný okamžik jako krizovou situaci, upozorní své okolí (zvukově, opticky, vibrací) a neprodleně reportuje (bezdrátově pomocí mobilní sítě) stav např. ambulanci, nejbližší pohotovosti, blízkým postižené osoby.

Popisovaný senzorický systém se skládá z několika více či méně inteligentních snímačů, sbírá a případně předzpracovává měřené biosignály tak, aby krizové stavy byly vyhodnoceny spolehlivě a v co nejkratší možné době (desítky sekund). Konkrétní provedení senzorického systému musí být kompaktní a pohodlné na užívání, jako například naslouchátko pro nedoslýchavé, náramkové hodinky nebo pouzdro upevněné na paži pro běžce, a zároveň funkční (v závislosti na fyzikálním principu snímačů) tak, aby naměřená data byla věrohodná a aktuální. Takové to požadavky na senzorický systém přinášejí mnoho teoretických a praktických problémů, které není snadné odstranit. Je třeba vyřešit například principy měření požadovaných biosignálů, napájení celého systému, konstrukční provedení vzhledem k užívání 24 hodin denně / 7 dní v týdnu a uživatelskou přívětivost.

Senzorický systém ve formě prstenu je popsán v článku (4), článek (5) popisuje systém pro monitorování osob v domácím prostředí. V (6) je uveden komplexní systém detekce a reportu krizových situací pomocí mobilní sítě.

## 2. Návrh senzorického systému

Uvedený sensorický systém snímá biologické signály, které jsou vstupem algoritmů indikujících a predikujících krizové stavy monitorované osoby. Výběr snímaných biologických signálů má zásadní vliv na výslednou funkci celého systému. Samotný snímač musí být neinvazivního charakteru a musí spolehlivě měřit fyzikální veličinu v běžném životě monitorované osoby (aktivní pohyb, spánek, atd.). Biologické signály a způsoby snímání uvažované pro detekci krizových stavů jsou shrnuty v tabulce 1. Snímače byly vybírány tak, aby jejich princip, konstrukce a umístění co nejméně obtěžovaly monitorovanou osobu.

**Tab. 1: Biologické signály vhodné pro detekci krizových stavů**

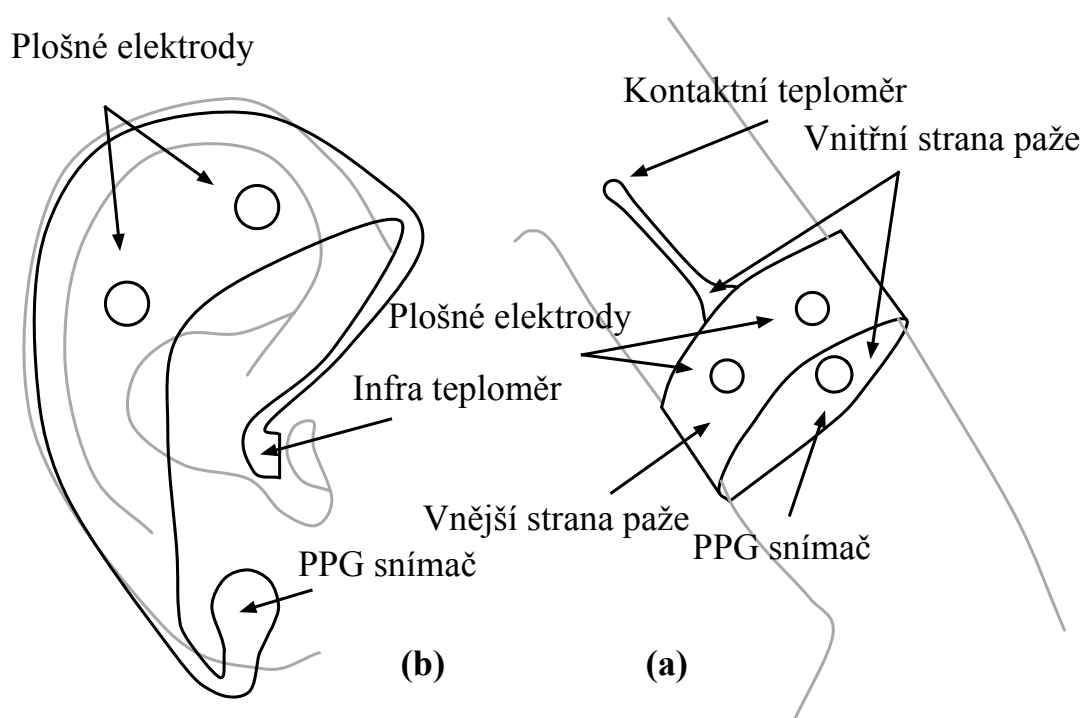
Měřený signál	Snímač	Aplikace
Tělesná teplota	kontaktní nebo infračervený teploměr	celkový stav organismu, detekce spánku
Okolní teplota	kontaktní teploměr	referenční hodnota
Povrchová vodivost kůže	plošné elektrody	psychické rozpoložení, fyzická aktivita, detekce spánku
Poloha, zrychlení a dráha	inerciální vztažná soustava	fyzická aktivita, výdej energie, poloha monitorované osoby
Photoplethysmogram	optický snímač (infračervená a viditelná červená vlnová délka)	poruchy srdečního rytmu, detekce dechu, okysličení krve

Z uvedených biologických signálů je velmi užitečný především photoplethysmogram (PPG) a povrchová vodivost kůže (Galvanic Skin Response – GSR), protože obsahují mnoho informací, které lze využít, a princip snímání je velmi jednoduchý. Z PPG lze určit většinu důležitých veličin souvisejících s fyzickou kondicí monitorované osoby jako je srdeční rytmus, dech, okysličení krve, krevní tlak, srdeční výdej a stav cév (1). PPG se měří pomocí optického snímače tak, že se dobře prokrvená tenká část těla (ušní lalůček, prst na ruce nebo noze) prosvěcuje infračervenou LED diodou a fotodioda měří intenzitu prošlého záření. Pro určení okysličení krve je třeba jako zdroj záření použít dvě LED diody, vyzařující infračervené záření (905, 910, nebo 940 nm) a viditelné světlo červené barvy (660 nm). Fotodioda měří intenzitu prošlého záření z každé LED diody samostatně, poměr těchto hodnot vyjadřuje okysličení krve. Povrchová vodivost kůže úzce souvisí s psychickým (stres, napětí, únava) i fyzickým stavem (teplo, fyzická aktivita) monitorované osoby. Nervový systém člověka automaticky řídí potní žlázy, které produkcí potu zvětšují vlhkost kůže a tím přímo ovlivňují vodivost kůže. Prudký pokles vodivosti kůže slouží jako indikátor, že nervový systém reaguje na nečekanou událost, například šok.

Pokud se naměřené signály, které se významově doplňují nebo překrývají, vhodně sloučí, zvýší se spolehlivost a věrohodnost detekce krizových stavů.

Navrhované provedení senzoričkého systému jsou na obrázku 1. Varianta na obrázku 1(a) je inspirovaná MP3 přehrávačem pro běžce. PPG senzor je reflexní, umístěný z vnitřní strany paže, a je zaměřen na pažní tepnu. Snímač teploty je v tomto případě kontaktní a pracuje obdobně jako běžný lékařský teploměr, je umístěn v blízkosti podpaží. Výhodou této varianty je též, že na pásku senzoričkého systému může být přímo umístěna jednotka detekce kritických stavů, tvoří tak společně kompaktní celek. Problém může nastat v případě, že paže monitorované osoby neumožňuje (tuková vrstva, vrásčitá kůže) dostatečný průchod záření z PPG senzoru, tento problém nehrozí u senzoričkého systému inspirovaného naslouchátkem, obrázek 1(b), kde PPG senzor snímá prokrvení ušního lalůčku (lze i reflexně uvnitř boltce). Tělesná teplota je snímána bezkontaktně uvnitř ucha pomocí inteligentního infračerveného teploměru. Umístění inerciální vztažné soustavy (Inertial Measurement Unit – IMU) a elektrod pro určení povrchové vodivosti kůže již není tak kritické.

**Obr. 1: Navrhované varianty provedení kompaktního senzoričkého systému**



### 3. Realizace

Navržené varianty kompaktního senzoričkého systému budou vyrobeny v několika prototypch. Mechanická, ergonomicky tvarovaná, konstrukce se navrhne pomocí 3D CAD softwaru a vytiskne se z plastické hmoty (ABS, PA) na 3D tiskárně. Tento postup je velmi vhodný pro takoveto výroby a vzhledem ke kusové výrobě je i finančně nenáročný.

Finální prototypy se otestují v laboratorních podmínkách a hlavně na koncových uživatelích. Na základě dlouhodobě zaznamenaných dat ze snímačů (spolehlivost snímání, reprodukovatelnost dat) a dojmů uživatelů se vybere nejvhodnější varianta.

Funkční prototypy budou též zařazeny do výuky lékařské techniky jako demonstrace systému pro vzdálené monitorování pacienta.

## 4. Závěr

Článek pojednává o kompaktním senzorním systému pro sběr biologických signálů, který je určen pro asistivní technologie a telemonitoring pacientů. Popsány jsou požadavky na takovýto systém a dva návrhy možného řešení, senzorní systém snímající na paži a na uchu. Dále jsou diskutovány problémy při nasazení těchto řešení. Předložený projekt má navázat na úspěšný projekt Intelligent Primer Nurse, se kterým členové týmu Biomedical Electronics Group (2) získali první cenu v soutěži Autonomie: Integrated Electronic Solutions for Human Assistance (3).

## Poděkování

Práci na tomto projektu vede školitel specialista Ing. Jan Havlík, Ph.D., projekt je podporován grantem České vysokého učení technického v Praze číslo SGS12/143/OHK3/2T/13 a také výzkumným projektem číslo MSM 6840770012 České vysokého učení technického v Praze (sponzoruje Ministerstvo školství, mládeže a tělovýchovy České republiky).

## Reference

- [1] Allen, J. Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement* (2007), vol. 28 no. 3.
- [2] Webové stránky Biomedical Electronics Group, <http://bmeg.fel.cvut.cz/>, duben 2012.
- [3] Webové stránky soutěže Autonomie: Integrated Electronic Solutions for Human Assistance, <http://www.stm32circle.com/projects/contest.php>, duben 2012.
- [4] Asada, H. H.; Shaltis, P.; Reisner, A.; Rhee, S.; Hutchinson, R. C. Mobile Monitoring with Wearable Photoplethysmographic Biosensors. *IEEE Engineering in Medicine and Biology Magazine* (2003).
- [5] Kaňtoch, E.; Jaworek J.; Augustiniak, P. Design of a wearable sensor network for home monitoring system. *Proceedings of the Federated Conference on Computer Science and Information Systems*, pp. 401–403
- [6] Rahman, A.; Alhamid, M. F.; Gueaieb W.; Saddik, A. An Ambient Intelligent Body Sensor Network For e-Health Applications. In International Workshop on Medical Measurements and Applications, Cetraro, Italy, May 29-30, 2009.



# Klasifikácia život ohrozujúcich stavov

Jakub PARÁK<sup>1</sup>, Jan Havlík<sup>1</sup>

<sup>1</sup> České vysoké učení v Praze, Fakulta elektrotechnická, Katedra teorie obvodů

[parakjak@fel.cvut.cz](mailto:parakjak@fel.cvut.cz)

**Abstrakt:** V nasledujúcom príspevku sú popísané návrh a implementácia jednoduchých metód a algoritmov na klasifikáciu život ohrozujúcich stavov. Metódy sú zamerané predovšetkým na monitorovanie a klasifikáciu základných srdečných arytmii a fyzickej aktivity monitorovanej osoby. Navrhnuté algoritmy sú prispôbené hlavne nízkym nárokom na systémové zdroje a výpočtový výkon, aby mohli byť použité v Real-Time embedded aplikáciách. Metódy boli implementované a testované v jednoduchom vývojovom modulárnom telemonitorovacom systéme, ktorého základ tvorí vývojový kit STM EvoPrimer s mikroprocesorom rady ARM Cortex. Na klasifikáciu život ohrozujúcich stavov obsahuje telemonitorovací systém moduly na snímanie biologických signálov a akcelerometer na snímanie fyzickej aktivity monitorovanej osoby. Príspevok taktiež popisuje návrh ďalších klasifikačných metód s využitím zložitejších algoritmov štatistického rozpoznávania, ktoré sa pripravujú na implementáciu.

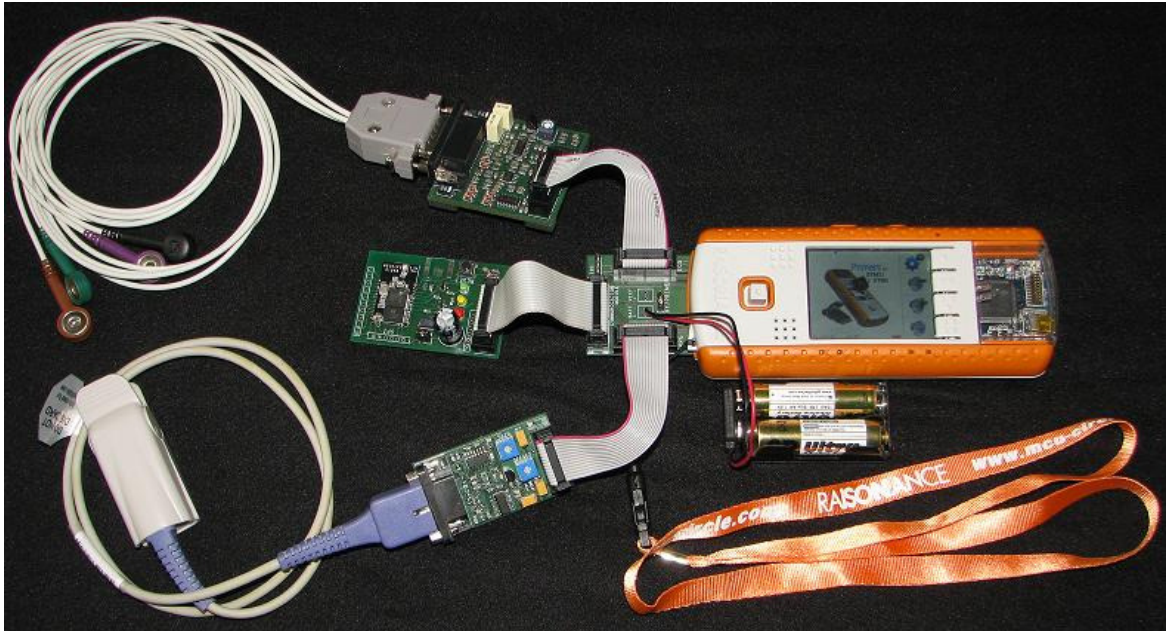
## 1. Úvod

Klasifikácia život ohrozujúcich stavov začína byť stále častejšie implementovaná ako súčasť asistívnych technológií. V súčasnosti vznikajú návrhy rôznych systémov, ktoré umožňujú monitorovať životné funkcie. Vzhľadom na starnúcu populáciu sú tieto systémy určené najmä pre osamelých seniorov. Hlavnou funkciou týchto systémov je možnosť aktivácie alarmu pri život ohrozujúcej situácii a následné upovedomenie záchranných zložiek, ošetrojúcich lekárov alebo príbuzných. Takýto alarm je aktivovaný na základe klasifikačných algoritmov a metód, ktoré spracúvajú dáta z rôznych senzorov, ktoré sú súčasťou zariadenia alebo sú pripevné k monitorovanej osobe [1].

Klasifikácia život ohrozujúcich stavov je vykonávaná predovšetkým na základe zmien srdečného tepu, ktorý je vypočítaný z meraného signálu elektrokardiogramu. Na správnu klasifikáciu sa okrem monitorovania zmien srdečnej tepovej frekvencie využíva aj meranie fyzickej aktivity monitorovanej osoby predovšetkým pomocou akcelerometrov [2].

Mnohé z navrhnutých systémov sú prepojené so Smart-Phone zariadením, do ktorého sú pomocou bezdrôtových protokolov prenášané dáta s rôznych senzorov umiestnených na monitorovanej osobe. Tieto systémy na spracovanie signálov a klasifikáciu život ohrozujúcich stavov využívajú pomerne vysokého výpočtového výkonu mobilného telefónu. Pre zariadenie je vytvorená mobilná aplikácia, ktorá spracováva snímané signály, vyhodnocuje život ohrozujúce situácie a spúšťa signalizačné alarmy. [3].

Navrhnuté zariadenie na spracovanie snímaných signálov a klasifikáciu život ohrozujúcich stavov v reálnom čase využíva iba výpočtový výkon jedného mikroprocesoru, v ktorom sú implementované algoritmy na spracovanie signálov, klasifikáciu život ohrozujúcich stavov signalizáciu alarmov a ovládanie komponentov riadiacej jednotky. Bezdrôtové pripojenie je využité iba na vizualizáciu alarmov a signálov na PC pomocou špeciálnej aplikácie. Aplikácia v PC nevykonáva už žiadne spracovanie prenesených dát.



Obr. 1: Modulárny telemonitorovací systém

## 2. Popis testovacieho telemonitorovacieho systému

Testovací telemonitorovací systém je zložený z hlavnej riadiacej jednotky a modulov na snímanie biologických signálov a bezdrôtový prenos dát. Riadiacu jednotku tvorí vývojový kit STM EvoPrimer osadený mikroprocesorom rady ARM Cortex F103. Kit má v sebe implementovaný trojosí MEMS akcelerometrický senzor, LCD display, bzučiak, slot na SD kartu, ovládací joystick a rozširujúci konektor na pripájanie vlastných externých modulov. Vývojový kit obsahuje vlastnú nabíjateľnú batériu, z ktorej môžu byť napájané aj externé moduly.

Na snímanie biologických signálov boli vyvinuté externé moduly pre elektrokardiogram, fotopletyzomogram a fonokardiogram. Keďže sa jedná o telemonitorovací systém, je k nemu možné pripojiť aj moduly na prenos dát pomocou Wi-Fi, Bluetooth alebo GSM.

Ovládanie jednotlivých komponentov vývojového kitu je vykonávané pomocou operačného systému CircleOS, ktorý je napísaný v jazyku C. Do tohto operačného systému boli vytvorené aj knižnice na snímanie signálov a prenos dát pomocou externých modulov. Bližšie informácie o telemonitorovacom systéme sú uvedené v publikácii [4].

Na implementáciu navrhnutých metód na klasifikáciu život ohrozujúcich stavov bola využitá kombinácia modulov na snímanie elektrokardiogramu, fotopletyzomogramu a prenos dát cez Bluetooth. Celkové riešenie je zobrazené na obr. 1.

## 3. Metódy a algoritmy na klasifikáciu život ohrozujúcich stavov

### 3.1. Navrhnuté metódy a algoritmy

Navrhnuté metódy sú zamerané predovšetkým na monitorovanie a klasifikáciu základných srdečných arytmií a fyzickej aktivity monitorovanej osoby, pri ktorých môžu nastať život ohrozujúce stavy. Tieto metódy boli navrhnuté tak, aby ich bolo možné jednoducho zrealizovať a implementovať do vývojového telemonitorovacieho systému. Vzhľadom na obmedzené systémové prostriedky mikroprocesoru v riadiacej jednotke, navrhnuté algoritmy nie sú výpočtovo ani pamäťovo náročné a teda dokážu bežať v reálnom čase a vyhodnocovať život ohrozujúce situácie.

Základné srdečné arytmie, ako sú zvýšená a znížená srdečná tepová frekvencia (bradykardia a tachykardia), sú detekované pomocou jednoduchého pevne stanoveného prahu. Hodnoty tepovej srdečnej frekvencie zvýšené nad 90 bpm (úderov za minútu) sú klasifikované ako tachykardia a hodnoty znížené pod 40 bpm sú klasifikované ako bradykardia.

Fyzická aktivita je monitorovaná pomocou zrýchlenia, ktoré je merané trojosím akcelerometrom. Na klasifikáciu sa využíva výsledný vektor, ktorý je zložený z vektorového súčtu zrýchlení vo všetkých troch osiach. Fyzická aktivita je klasifikovaná pomocou dát z akcelerometru, ktoré prekročujú stanovený prah alebo na základe reakcie monitorovanej osoby, od ktorej sa vyžaduje aby počas zadaného času niekoľkonásobne stlačila tlačidlo joysticku na riadiacej jednotke. Jedná sa o podobný systém ako je tlačidlo bdelosti v lokomotívach, ktoré musí rušňovodič stlačiť v presne stanovenom maximálnom čase, inak sa spustí zvukový alarm. V našom prípade je fyzická aktivita klasifikovaná iba do dvoch tried a to, či sa monitorovaná osoba pohybuje, alebo sa dlhší čas nepohybuje, čo môže signalizovať život ohrozujúci stav.

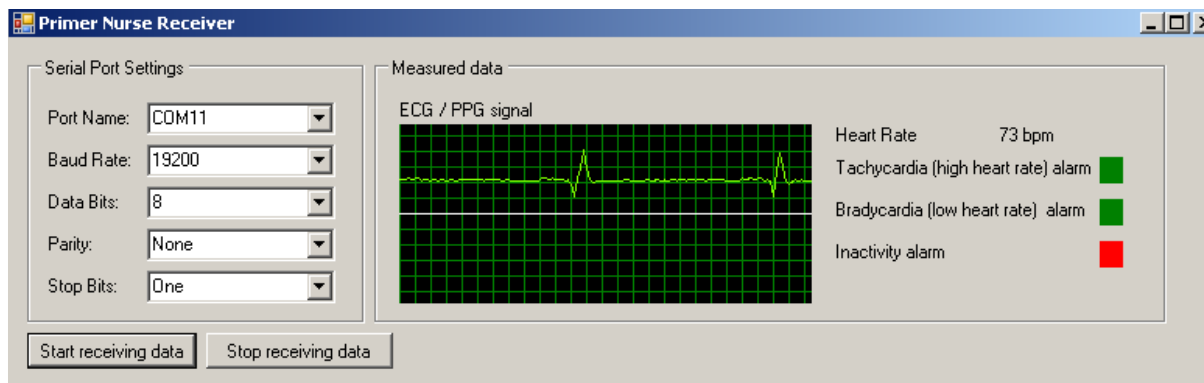
### **3.2. Implementácia navrhnutých algoritmov**

Klasifikácia srdečných arytmií je robená na základe srdečnej tepovej frekvencie, ktorá je počítaná z II. zvodu elektrokardiogramu. Algoritmus na výpočet srdečnej frekvencie z elektrokardiogramu a jeho implementácia v jazyku C je bližšie popísaná v článku [5]. Hodnota srdečného tepu je vyhodnocovaná každých 5 sekúnd. Samotná klasifikácia je implementovaná ako premenná typu „semafor“ z dôvodu eliminácie chybné vypočítanej srdečnej frekvencie napríklad kvôli značným pohybovým artefaktom v biologickom signáli elektrokardiogramu. Znamená to, že napríklad pri vyhodnocovaní tachykardie sa inkrementuje čítač alarmu vždy pri prekročení prahu a dekrementuje sa pri hodnote nižšej ako prah. Čítač beží v rozmedzí od 0 do 5, to znamená, že až po príchode 5 nadprahových hodnôt je daný stav klasifikovaný ako rizikový a je spustený alarm. Alarm je signalizovaný blikajúcim nápisom na displeji a zvukovým znamením zo zvučiaku na riadiacej jednotke. Pokiaľ sa čítač alarmu pri ďalšom vyhodnovení zníži na hodnotu menšiu ako 5, alarm sa automaticky vypne.

Klasifikácia fyzickej aktivity, ktorá je popísaná vyššie, je implementovaná ako voľne bežiaci čítač. Tento čítač sa inkrementuje každú sekundu alebo reštartuje na nulovú hodnotu pokiaľ bolo stlačené tlačidlo joysticku alebo výsledné zrýchlenie zaznamenané akcelerometrom dosiahlo nadprahovú hodnotu, ktorá bola nastavená na  $1,4 \times 9,8 \text{ ms}^{-2}$ . Pokiaľ čítač alarmu presiahne hodnotu 30, je automaticky spustený alarm, ktorý je signalizovaný blikajúcim nápisom na displayi a zvukovým znamením zo zvučiaku na riadiacej jednotke. Alarm je vypnutý pri následnom stlačení tlačidla alebo zachytený pohybu pomocou akcelerometrického senzoru.

### **3.3. Vizualizácia a prenos alarmov rizikových stavov**

Alarmy, ktorými sú signalizované rizikové stavy na základe výslednej klasifikácie, sú prenášané pomocou bezdrôtového Bluetooth prenosu do PC. V PC sú alarmy rizikových stavov pomocou jednoduchej aplikácie vizualizované spolu s meraným signálom. Vizualizácia signálu a alarmu je vhodná napríklad pre operátora záchranného strediska alebo lekára, ktorý môže vidieť či sa jedná o chybnú klasifikáciu a následný planý poplach. Ukážka vizualizácie sa nachádza na screenshots aplikácie na obr. 2.



Obr. 2: PC aplikácia na vizualizáciu alarmov a signálov

#### 4. Návrh pripravovaných pokročilých metód klasifikácie

V predchádzajúcej časti bol popísaný návrh a implementácia jednoduchých metód klasifikácie život ohrozujúcich stavov. Na základe predchádzajúcich skúseností a testovaní sa pripravuje implementácia klasifikácie pomocou metód štatistického rozpoznávania. Merané dáta budú rozšírené o signály z gyroskopického senzoru a na výpočet srdečnej frekvencie sa použije ako druhý referenčný signál fotopletysmografická krivka. Namiesto prahovacích metód bude použitá implementácia lineárneho klasifikátora typu Perceptron, ktorý nemá veľké nároky na systémové zdroje pri klasifikácii daného stavu, vzhľadom k tomu, že na určenie správnej triedy je nutné vypočítať jednu lineárnu funkciu. Tento klasifikátor bude natrénovaný na dátach, ktoré budú získané na základe rozdelenia do jednotlivých tried pomocou zhukovanej analýzy K-Means [6]. Trénovanie klasifikátora bude prebiehať na PC v prostredí Matlab.

Pri implementácii algoritmov sa bude klásť dôraz na rýchlosť výpočtov, ktorá je pri vyhodnocovaní život ohrozujúcich situácií a malý výpočtový výkon mikropočítačov dôležitým faktorom správneho návrhu algoritmov.

Klasifikácia rizikových stavov bude rozšírená o priamu detekciu pádov a vylepšenú detekciu fyzickej aktivity a srdečných arytmií, pretože bude využitá fúzia dát z pohybových senzorov a srdečná tepová frekvencia získaná z elektrokardiogramu a fotopletysmogramu [7].

Tieto metódy budú implementované a testované na vývojom modulárnom telemonitorovacom systéme, ktorý bol navrhnutý na tieto účely. Konečným výsledkom má byť implementácia algoritmov do jedného kompaktného systému, ktorý bude zložený z riadiacej jednotky a senzorického systému [8].

#### 5. Záver

Klasifikácia život ohrozujúcich stavov je jedna z významných funkcií systémov určených pre asistívne technológie a monitorovanie životných funkcií.

Pri návrhu algoritmov, ktoré majú bežať na zariadeniach s malým výpočtovým výkonom a majú byť pritom dostatočne rýchle a spoľahlivé, treba vyberať z jednoduchých metód a ich správnej implementácie.

Implementáciou algoritmov na klasifikáciu základných srdečných arytmií a fyzickej aktivity monitorovanej osoby do modulárneho telemonitorovacieho systému bola zlepšená jeho funkcionálnosť, ktorá môže byť využitá na ďalšie výskumné a výukové účely.

Systém s touto rozšírenou funkcionálnosťou komplexne demonštruje získavanie dát z monitorovaného pacienta, ich následné spracovanie, vyhodnotenie život ohrozujúceho stavu, prenos dát do PC a ich následnú vizualizáciu.

Funkčné prototypové riešenie celého telemonitorovacieho systému, bolo pod názvom projektu „Intelligent Primer Nurse“, ohodnotené prvou cenou v súťaži 2011 OpenWorld

Design Contest Promoting Human Assistance, ktorá bola organizovaná pod záštitou firiem Raisonance a STMicroelectronics [9].

Na základe úspešných výsledkov pri testovaní jednoduchého riešenia sú pripravované pokročilé klasifikačné metódy, ktoré boli popísané v poslednej časti príspevku. Pri výbere a použití týchto metód je nutné dbať na ich správnu implementáciu v Real-Time systémoch.

## Podakovanie

Tato práca bola podporená výskumným zámerom Ministerstva školstva, mládeže a telovýchovy Českej republiky č. MSM 6840770012.

## References

- [1] Pantelopoulos, A.; Bourbakis, N. G. A Survey on Wearable Sensor-Based Systems for Health Monitoring and Prognosis. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*. Jan. 2010, vol. 40, 1–12.
- [2] Wan-Young Chung; Bhardwaj, S.; Purwar, A.; Dae-Seok Lee; Myllylae, R. W. A Fusion Health Monitoring Using ECG and Accelerometer sensors for Elderly Persons at Home. In *Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Lyon, August 2007.
- [3] Leijdekkers, P.; Gay, V. Personal Heart Monitoring System Using Smart Phones To Detect Life Threatening Arrhythmias. In *Proceedings of the 19th IEEE International Symposium Computer-Based Medical Systems*. Salt Lake City, June 2006.
- [4] Parak, J.; Dvořák, J.; Pokorný, M.; Havlík, J. Modular development telemonitoring system. In *Proceedings of the 35th IEEE International Conference on Telecommunications and Signal Processing*, Prague, July 2012. (in press)
- [5] Parak, J.; Dvořák, J.; Havlík, J. Device for Long Term Measurement of Heart Rate. In *Proceedings on the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies*, Barcelona, Spain, October 2011.
- [6] Bidargaddi, N.; Sarela, A.; Korhonen, I. Physiological state characterization by clustering heart rate, heart rate variability and movement activity information. In *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vancouver, August 2008.
- [7] Anh Dinh; Teng, D.; Li Chen; Yang Shi; McCrosky, C.; Basran, J.; Del Bello-Hass, V. Implementation of a Physical Activity Monitoring System for The Elderly People With Built-in Vital Sign and Fall Detection. In *Proceedings of the IEEE International Solid-State Circuits Conference, 2008. Digest of Technical Papers*, San Francisco, Feb. 2008.
- [8] Wong, A.C.-W.; McDonagh, D.; Kathiresan, G.; Omeni, O. C.; El-Jamaly, O.; Chan, T. C.-K.; Paddan, P.; Burdett, A. J. A 1V, Micropower System-on-Chip for Vital-Sign Monitoring in Wireless Body Sensor Networks. In *Proceedings of the 6th International Conference on Information Technology: New Generations*, Las Vegas, April 2009.
- [9] STMicroelectronics. STMicroelectronics Announces Results of the Autonomie 'OpenWorld Design Contest' Initiative to Promote Human-Assistance Technologies. [http://www.st.com/internet/com/press\\_release/t3291.jsp](http://www.st.com/internet/com/press_release/t3291.jsp), 4/2012.

# Návrh mechanického modelu kardiovaskulárního systému

Miroslav LOŽEK<sup>1</sup>, Jan HAVLÍK<sup>1</sup>, Zdeněk HORČÍK<sup>1</sup>

<sup>1</sup> České vysoké učení v Praze, Fakulta elektrotechnická, Katedra teorie obvodů

[lozekmir@fel.cvut.cz](mailto:lozekmir@fel.cvut.cz)

**Abstrakt:** Model kardiovaskulárního systému je mechanický simulátor hemodynamických parametrů lidského krevního řečiště. Jedná se o soustavu trubic, ventilů, čerpadla a dalších pomocných prvků, které jsou uspořádány tak, aby věrohodně napodobily tok krve zdravého lidského jedince. Na modelu jsou sledovány základní hemodynamické parametry, mezi které patří krevní tlak, srdeční výdej, rychlost šíření pulsní vlny a další. Nepostradatelnou součástí celého systému je mechanické čerpadlo, které svou činností napodobuje pulzatilní tok krve vydávaného zdravým lidským srdcem. Tato publikace popisuje návrh a konstrukci modelu, způsob řízení mechanického čerpadla a metody měření hemodynamických parametrů. Cílem celého projektu je vytvoření věrného modelu kardiovaskulárního systému s možností měření hemodynamických parametrů za pomoci konvenčních monitorovacích zařízení.

## 1. Úvod

Kardiovaskulární systém (KVS) lidského organismu je složitá soustava sestávající se z různých druhů cév s různým průměrem, který se pohybuje v řádech  $\mu\text{m}$  až  $\text{cm}$ . Nepostradatelnou součástí KVS je srdeční pumpa, která slouží jako adaptivní iniciátor proudění krve. Složení krve také velkou měrou ovlivňuje hemodynamické parametry krevního řečiště. Řízení oběhové soustavy je zabezpečeno chemicky a elektricky pomocí hormonálního a inervačního systému. [1].

Vzhledem k velké složitosti a rozsahu kardiovaskulárního systému není možné vytvořit přesný model zahrnující všechny elementy cévního řečiště. Z toho důvodu je nutné zaměřit se pouze na určité celky systému a jejich vlastnosti.

Mechanický model KVS slouží ke sledování hemodynamických parametrů měřených ve velkých cévách. Na základě tohoto předpokladu je možné značnou část modelu zjednodušit a nahradit ji jednoduchými prvky, které simulují chování celého systému.

## 2. Kardiovaskulární systém

Kardiovaskulární systém je soustava orgánů zajišťující cirkulaci krve, která okysličuje organismus, transportuje životodárné látky a udržuje homeostázu celého těla.

### 2.1. Anatomie kardiovaskulárního systému

Oběhový systém lze rozdělit na část, která vede okysličenou krev. Toto vedení je v systémovém oběhu zajištěno arteriálním systémem, který vykazuje vysokou pevnost a elasticitu. Krev v arteriích proudí pulzatilním charakterem směrem k orgánům. Odvod deoxygenované krve z orgánů zajišťuje venózní systém, který je charakteristický svou

rigiditou. Proudění krve v žilách má kontinuální charakter. Propojení výše popsaných systémů je zabezpečeno kapilární sítí, ve které dochází k přenosu krevních plynů mezi krví a buňkami. Kapilární řečiště představuje z hemodynamického hlediska silně odporový člen, který, mimo jiné, má za následek transformaci pulzativního proudění na proudění kontinuální.

Středem kardiovaskulárního systému je srdce, které představuje hlavní akční člen proudění krve. Srdce je rozděleno na 4 části (2 síně a 2 komory). V době systolické kontrakce komor dochází k vypuzení okysličené krve z levé komory do těla, odtud se následně odkysličená krev vrací žilním systémem do pravé síně. Správnou sekvenci srdečních kontrakcí zabezpečuje převodní systém srdeční, který se sestává ze specifických, tzv. pacemakerových buněk. Pacemakerová buňka nemá klidový membránový potenciál a slouží jako autonomní generátor elektrických impulzů v srdci. [1, 2].

## 2.2. Hemodynamické parametry

Hemodynamika kardiovaskulárního systému sleduje fyzikální chování toku kapaliny v krevním řečišti. Hemodynamiku ovlivňují mechanické parametry cév, výstupní tok srdeční pumpy a také vlastnosti krve.

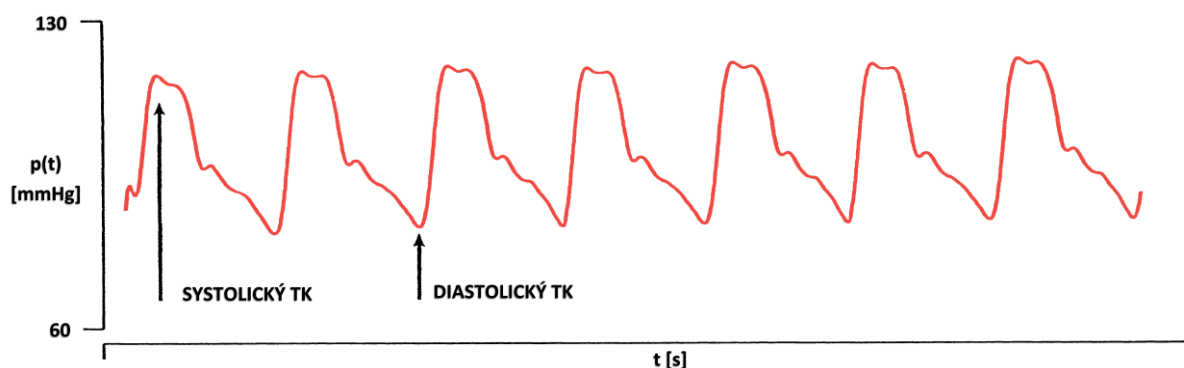
Jedním z parametrů popisujících tok krve cévou je tzv. rychlost pulzní vlny (**Pulse Wave Velocity**) [m/s]. Je to parametr, který je závislý pouze na vlastnostech cévy a hustotě kapaliny  $\rho$  [kg/m<sup>3</sup>]. Mezi vlastnosti cév patří vnitřní poloměr cévy  $r$  [m], tloušťka stěny cévy  $h$  [m] a její modul elasticity  $E$  [Pa]. Rychlost pulzní vlny může sloužit k diagnostice stavu cévního řečiště. Výpočet popisuje Moens-Kortewegova rovnice (1). [1,3]

$$PWV = \sqrt{\frac{E \cdot h}{2r \cdot \rho}} \quad (1)$$

Dalším důležitým parametrem KVS je srdeční výdej. Srdeční výdej popisuje, jaké množství krve je vypuzeno srdcem do oběhového systému. Srdeční výdej je obvykle vztažen k jednotce času (**Minute Cardiac Output** [l/min]) nebo k jednomu srdečnímu cyklu (**Stroke Volume** [l]). Vztah těchto dvou veličin popisuje rovnice (2), kde  $HR$  [1/min.] je srdeční frekvence. [1,3,4]

$$MCO = SV \cdot HR \quad (2)$$

Krevní tlak je takový tlak, kterým krev působí na stěnu cévy, kterou prochází. Tlak je vyvolán silovým působením srdečního svalu v době své kontrakce. Konvenčně se krevní tlak měří na velkých arteriích. Arteriální tlaková křivka má pulzativní charakter (viz Obr. 1), jejíž hodnota lokálních extrémů má největší diagnostický význam. V souvislosti se srdeční činností jsou tyto dva tlakové parametry nazývány systolickým a diastolickým krevním tlakem. [1, 3].



Obr. 1: Typický průběh fyziologické tlakové křivky

### 2.3. Parametry cévního řečiště

Proudění krve řečištěm velmi zásadně ovlivňuje materiál, ze kterého je krevní řečiště sestaveno. Chování cévního řečiště lze popsat pomocí parametrů: odpor, compliance a inertiance.

Krevní řečiště představuje toku krve určitý odpor, ten zásadně ovlivňuje velikost tlaku v řečišti. Organismus je schopen za pomoci konstričních a dilatačních mechanismů měnit odpor krevního řečiště, čímž zajišťuje regulaci krevního tlaku.

Stěny cév jsou více či méně poddajné. Zvýšení tlaku způsobí roztažení cévy a tím i změnu jejího objemu. Elastické vlastnosti cévy popisuje parametr compliance  $[m^5/N]$ . Vztah mezi *compliance*, změnou tlaku  $\Delta p$  [Pa] a objemem  $V$  [ml] vyjadřuje rovnice (3), kde  $V_0$  [ml] je objem cévy bez přiloženého tlaku. [3 - 5]

$$V = V_0 + C \cdot \Delta p \quad (3)$$

Setrvačnost toku krve cévní soustavou popisuje parametr zvaný inertiance  $[Ns^2/m^5]$ . Inertiance závisí na délce cévy  $l$  [m], ploše průřezu cévou  $S$   $[m^2]$  a hustotě krve  $\rho$   $[kg/m^3]$  (4).

$$I = \frac{\rho \cdot l}{S} \quad (4)$$

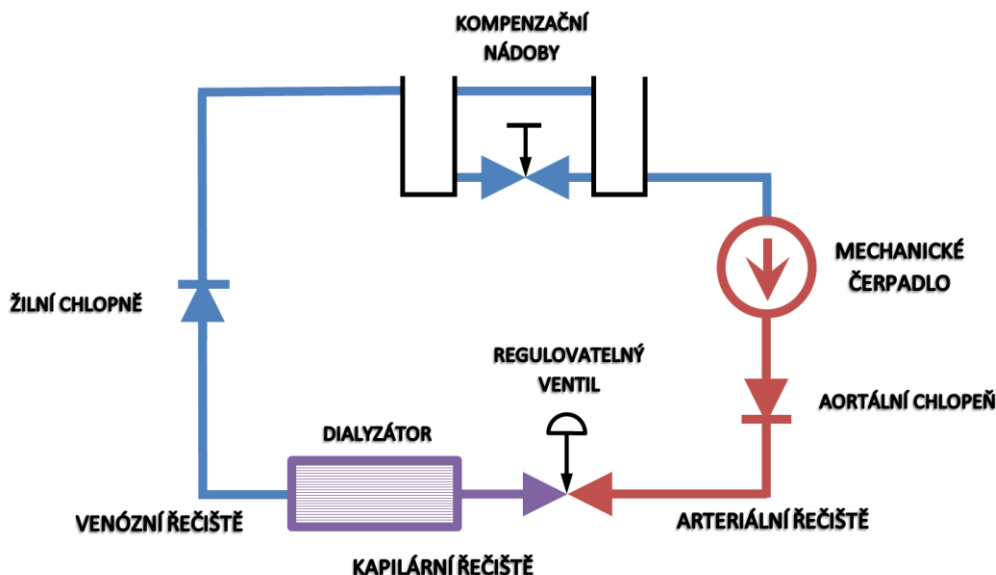
## 3. Mechanický model kardiovaskulárního systému

Základní mechanický model KVS vychází z předpokladu, že lze sdružit parametry jednotlivých elementů oběhového systému do větších funkčních celků. Výsledný model se následně sestává z několika základních částí. Hlavním elementem modelu je mechanická pumpa, která napodobuje činnost lidského srdce. Cévní systém je modelován pomocí sériově zapojených vedení s různými parametry odporu, compliance a inertiance, která reprezentují arteriální, venózní a kapilární řečiště. [4 - 6].

### 3.1. Stavba modelu

V základním modelu pro první experimenty postačí soustava dvou hadic a regulovatelného ventilu. Hadice napodobující arteriální systém se vyznačuje velkou elasticitou, což zajistí nízký parametr compliance. Naopak hadice simulující žilní systém je velmi rigidní, její compliance je podstatně vyšší. Délkou hadic je možné ovlivnit parametr inertiance. Obě hadice jsou propojeny regulovatelným ventilem, kterým je možné měnit odpor celé soustavy. Simulace kapilárního řečiště je dále podpořena zahrnutím kapilárního dialyzátoru. Dalšího zvýšení věrohodnosti výsledků bylo dosaženo zařazením dvou zpětných ventilů do obvodu řečiště. Ventil zařazený za srdeční pumpou simuluje aortální chlopeň. Druhý zpětný ventil reprezentuje žilní chlopeň a zabraňuje zpětnému žilnímu toku. Stabilitu soustavy udržují kompenzační nádoby, které jsou umístěny tak, aby umožňovaly zběžné měření srdečního výdeje. (viz Obr. 2)





Obr. 2: Schématické zobrazení modelu kardiovaskulárního systému

### 3.2. Mechanická pumpa

Tento model je zaměřen na modelování hemodynamických parametrů krevního řečiště. Na základě tohoto faktu není nutné modelovat složitou čtyř-komorovou pumpu. Důležitý je pouze charakter srdečního výstupu krve do velkého krevního oběhu, ten musí věrně napodobovat výstup reálného srdce.

Mechanickou pumpou může být vodní čerpadlo s dostatečným průtokem a dynamikou chodu. Lze použít běžné, vhodně řízené zubové čerpadlo.

#### 3.2.1. Řízení čerpadla

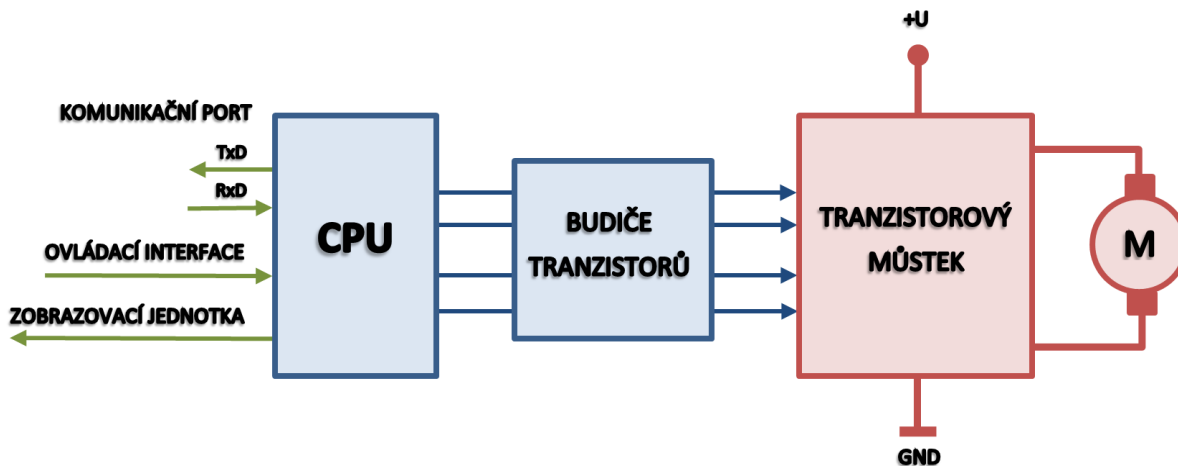
Cílem správného řízení čerpadla je namodelovat výstupní tlakovou vlnu tak, aby se věrohodně podobala fyziologickému tlakovému gradientu. Toho lze dosáhnout nelineárním řízením otáček motoru.

Vzhledem k faktu, že vinutí stejnosměrného motoru čerpadla má pro střídavé buzení induktivní charakter je možné řídit motor čerpadla v pulzním režimu. K řízení motoru je použita pulzně-šířková modulace (**P**ulse-**W**ide **M**odulation). Střední hodnota výstupního napětí PWM částečně koreluje s vytvořenou tlakovou vlnou. Díky vhodně navrženému algoritmu řízení je možné velmi dobře napodobit tvar fyziologické tlakové vlny.

#### 3.2.2. Řídicí jednotka mechanického čerpadla

Řídicí jednotka se principiálně dělí na dvě části. Část signálová obsahuje mikroprocesor generující řídicí signály přivedené do části výkonové, která obsahuje spínací prvky.

Signálový procesor generuje řídicí signály podle předepsaného algoritmu pro jednotlivé spínací prvky. Mikročip dále řídí celý proces srdečního cyklu a zaštiťuje ovládací interface. Výkonový blok je tvořen čtveřicí výkonových spínacích součástí, které jsou uspořádány do dvou spínacích můstků. Toto zapojení umožňuje chod v proti-fázi, tento proces dále zvětší dynamický rozsah čerpadla.

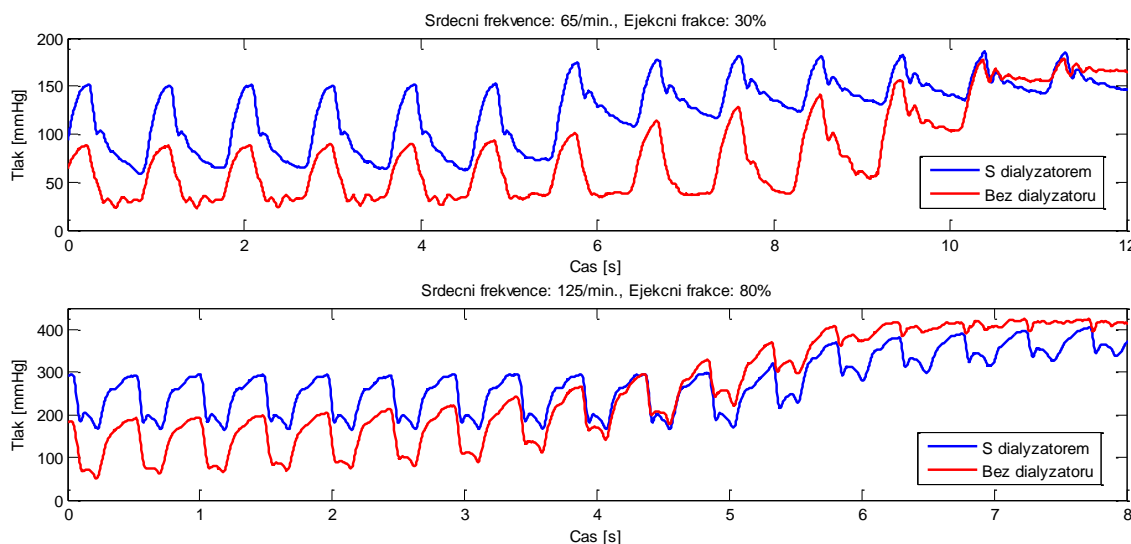


Obr. 3: Blokové schéma řídicí jednotky čerpadla

#### 4. Měření hemodynamických parametrů modelu

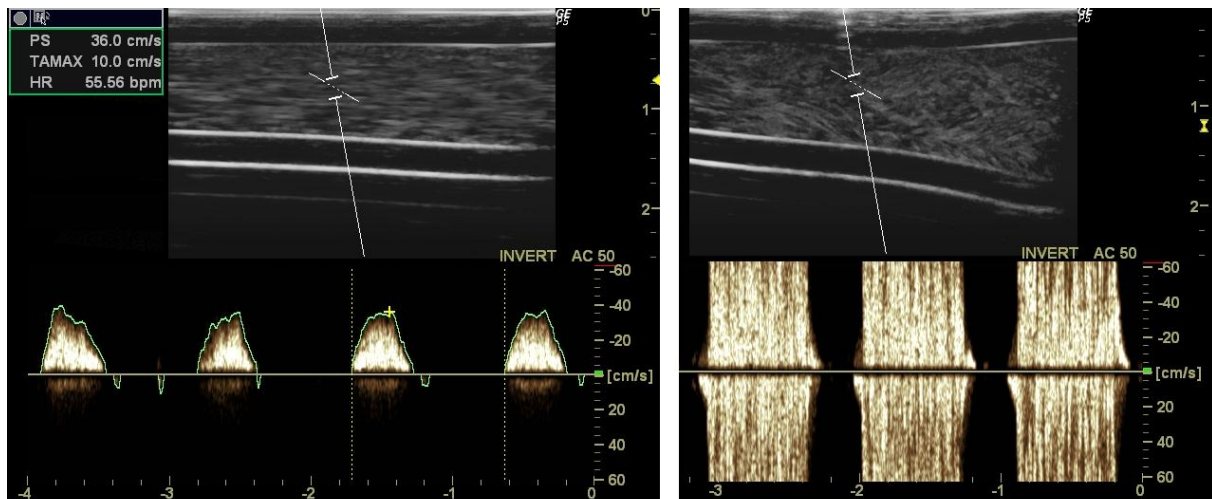
Ověření funkčnosti bylo provedeno pomocí externího měření hemodynamických parametrů. K měření bylo použito tlakových senzorů a dopplerovské sonografie.

Sledování krevního tlaku bylo zajištěno pomocí analogového tlakového čidla MPX53GP, výstupní signál byl zesílen a vzorkován měřicí kartou. Měření proběhlo na arteriální části modelu při změnách srdeční frekvence, ejekční frakce a odporu krevního řečiště (Obr. 4). Na obrázku je také patrný vliv dialyzátoru zahrnutého do oběhového systému.



Obr. 4: Naměřená tlaková křivka

Měření srdečního výdeje bylo provedeno na hadici představující arterii pomocí dopplerovského sonografu GE Logiq 5. Sonografií byl změřen časový profil rychlosti toku (viz Obr. 5 nalevo), jehož jednoduchou integrací lze při znalosti světlosti hadice dopočítat srdeční výdej. Mimo jiné lze touto metodou prokázat turbulentnost proudění (viz Obr. 5 napravo).



Obr. 5: Měření průtoku a turbulentnosti toku pomocí dopplerovské sonografie

## 5. Závěr

Projekt zaměřený na modelování kardiovaskulárního systému umožnil vznik mechanického modelu zabývajícího se tokem tekutin v cévním řečišti. Model byl sestaven pomocí jednoduchých elementů simulujících jednotlivé funkční bloky reálného oběhového systému.

Měření prokázalo dílčí úspěšnost navrženého systému. Tlaková křivka naměřená v modelovaném oběhu věrohodně napodobuje tlakovou křivku fyziologického krevního řečiště. Měření srdečního výdeje pomocí sonografie odpovídá skutečnému přenesenému objemu s maximální odchylkou 10%. Naměřené profily rychlosti toku arterií také odpovídají fyziologické skutečnosti. Pomocí této metody lze velmi přesvědčivě prokázat turbulentnost a laminárnost proudění, charakteristické rysy těchto typů proudění lze pozorovat vizuálně i pomocí časových profilů dopplerovského měření rychlosti toku.

## Poděkování

Tato práce byla podpořena grantem č. SGS12/143/OHK3/2T/13 a výzkumným programem č. MSM 6840770012.

## Reference

- [1] Dylevský, I. *Základy funkční anatomie*. 2011. ISBN: 978-80-87419-06-9
- [2] Čihák, R. *Anatomie 3*. Grada Publishing, 2004. ISBN: 80-247-1132-X
- [3] Batzel, J. J., Kappel, F., Schneditz, D., Tran, H. T. *Cardiovascular and Respiratory Systems: Modeling, Analysis, and Control*. 2006. ISBN 978-0898716177.
- [4] Westerhof, N., Stergiopoulos, N., Noble, M. I. M. *Snapshots of Hemodynamics*. Basic Science for the Cardiologist, 2005.
- [5] Cobelli, C., Carsin, E. *Introduction to modeling in physiology and medicine*. Academic Press. 2008. ISBN: 978-0-12-160240-6
- [6] Le Rolle, V., Hernandez, A. I., Richard, P.Y., Buisson, J., Carrault, G. *A Bond Graph Model of the Cardiovascular System*. Acta Biotheorica, 2005.

# Použitie MLLR adaptácie akustických modelov trifónov v rozpoznávači spojitej reči na bázi HTK

Michal Borský

České vysoké učení v Praze, Fakulta elektrotechnická, Katedra teorie obvodů  
borskmic@fel.cvut.cz

*Abstrakt:* Adaptácia akustického modelu na rečníka, prostredie alebo kanál je všeobecne používaná metóda zvýšenia úspešnosti v systémoch automatického rozpoznávania reči. Jednou z často používaných je technika maximálne vierohodnej lineárnej regresie. Tento prístup obmedzuje počet voľných parametrov modelu združovaním komponent do tried a dokáže tak priniesť zvýšenie úspešnosti aj v situácii s malým množstvom adaptačných dát. V príspevku je analyzovaný prínos tejto techniky v reálnej aplikácii systému rozpoznávania spojitej reči na množine 11 rečníkov. Porovnané sú dva prístupy k tvorbe regresných tried, automatické delenie algoritmom a ručné delenie založené na fonetike. Uskutočné experimenty ukazujú, že oba prístupy sú si rovnocenné z hľadiska zlepšenia chybovosti.

## 1. Úvod

Automatické rozpoznávanie reči (ASR - automatic speech recognition) na bázi skrytých Markovových modelov (HMM - Hidden Markov Model) je v dnešnej dobe najpoužívajúci zo štatistických prístupov k riešeniu tejto problematiky. Dve hlavné súčasti každého automatického roznávača spojitej reči s veľkým slovníkom (LVCSR - Large Vocabulary Continuous Speech Recognition) je akustický model (AM), ktorý modeluje akustickú stránku reči a jazykový model (LM - language model). Správne natrénovanie AM si vyžaduje obrovské množstvo dát, rádovo niekoľko desiatok hodín, keď takto natrénovaný model je vo výsledku prispôbený na tieto dáta. Jedným z problémov ale je, že vo väčšine prípadov dáta pochádzajú od viacerých rečníkov, preto takto natrénované modely nazývame nezávislé na rečníkovi (SI - speaker independent). Veľká variabilita akustickej reprezentácie tej istej výpovede medzi rečníkmi ale obmedzuje ich použitie v reálnych aplikáciach, keďže výsledky ktoré s nimi je možné dosiahnuť sú len priemerné. Z tohto dôvodu všetky súčasné kvalitné LVCSR systémy riešia tento problém adaptáciou AM [3], [4], keď sa snažia prispôbiť SI model na jedinečnosť s akou každý človek rozpráva. Takto zadaptované modely potom nazývame závislé na rečníkovi (SA - speaker adapted).

Všeobecne je ale možné povedať, že úloha adaptácie akustického modelu je prispôbenie natrénovaného AM na špecifické vlastnosti prostredia, kanálu alebo rečníka. Spomeňme preto aj situácie, keď prispôbujeme AM na prostredie alebo kanál. V prvom prípade sa jedná o situáciu, keď použitý AM bol vytvorený v iných podmienkach, než v akých sa ho snažíme aplikovať. Ide predovšetkým o rozdielnú úroveň šumu na pozadí alebo odlišné akustické vlastnosti prostredia ako prítomnosť odrazov atď. V prípade adaptácie na vlastnosti kanálu je cieľom potlačiť vplyv rozdielných frekvenčných charakteristik

nahrávacích zariadení, prípadne následnej kompresie. Mnoho nahrávacích zariadení, za účelom uchovania väčšieho množstva dát, automaticky komprimuje vstupný signál, čo sa vo výsledku prejaví ako skreslenie spektra signálu. Použitie takto skresleného spektra môže spôsobiť v krajnom prípade až zlyhanie rozpoznávia. Medzi najpoužívanejšie kompresie uvedme MP3 alebo kompresiu do telefónneho kanálu.

Všetky tieto vlastnosti sú prítomné v každom signáli súčasne a pokiaľ sa nepoužijú osobitné techniky na ich potlačenie (spektrálne odčítanie šumu, normalizácia vokálneho traktu, atď) prebieha adaptácia na prostredie, kanál aj rečníka zároveň. Cieľ adaptácie je preto určený správnym výberom dát. Tieto musia obsahovať čo najviac informácie ohľadom parametru ktorý sa snažíme zadaptovať a zvyšné vlastnosti, vzhľadom k neadaptovanému modelu, musia zostať zachované.

## 2. Adaptácia akustického modelu

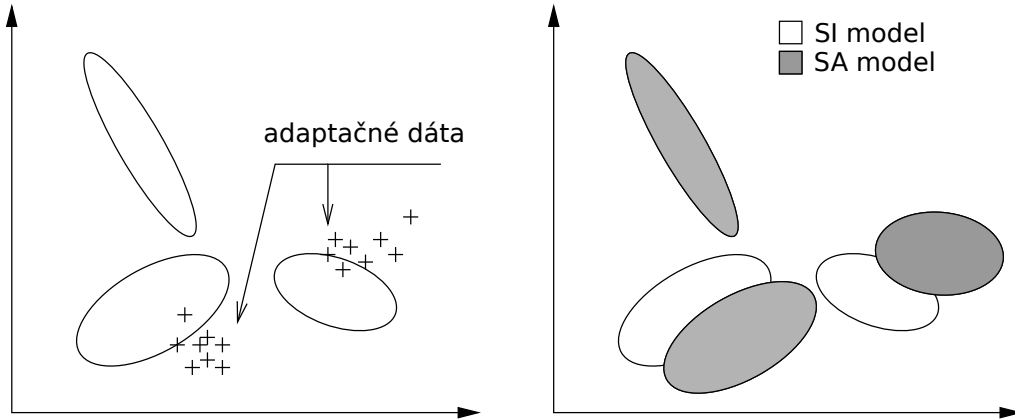
Dostupné metódy adaptácie sa líšia vzhľadom na situáciu a prístup k riešeniu problému v ktorom ich aplikujeme. V zásade rozoznávame niekoľko samostatných postupov adaptácie

- Riadená (supervised) a neriadená (unsupervised) - líšia sa podľa dostupnosti prepisu adaptačných dát.
- Dávková (batch) a inkrementálna (incremental) – rozlíšenie podľa dostupnosti adaptačných dát. Pokiaľ sú všetky dáta dostupné v momente adaptácie hovoríme o statickej, alebo tiež dávkovej, pokiaľ sa adaptácia vykonáva zároveň s tým ako prichádzajú nové dáta hovoríme o inkrementálnej.
- On-line a off-line – rozlíšenie podľa toho, či adaptácia prebieha zároveň s rozpoznávaním alebo nie.

Ako už bolo zmienené, výber konkrétnej adaptačnej metódy je závislý na situácii a možnostiach, ktoré su k dispozícii. V prípadoch ako sú napr. systémy na prepis zvukových záznamov do textovej podoby, kde je čas aj priestor a identita rečníka je dopredu známa, sa využíva zväčša riadená, dávková adaptácia. Za asistencie systému užívateľ nahovorí dopredu určený text, ktorý bol navrhnutý tak, aby obsahoval čo najviac foneticky bohatého obsahu. Tento sa spolu s jeho prepisom použije ako adaptačná množina na úpravu SI modelu.

Pokiaľ identita rečníka nieje dopredu známa, musí sa systém vyrovnáť so situáciou, keď neexistuje ani množina na ktorej by bolo možné adaptáciu vykonať. Jedným z riešení v tomto prípade je postup, keď sa výpovede určené na prepis použijú zároveň aj na adaptáciu modelu. Prepis je vytvorí pomocou SI modelu a jeho správny obsah nieje preto zaručený, hovoríme preto o neriadenej inkrementálnej adaptácii. Množstvo adaptačných dát je preto priamo úmerné množstvu dát, ktoré systém rozpoznáva.

Vo väčšine dnešných ASR systémoch sa na úrovni akustického modelovania používajú skryté Markovove modely so spojitými rozloženími hustôt (CDHMM - Continuous Density HMM). Výstupné pravdepodobnosti jednotlivých stavov sú určené multivariačnou zmesou Gaussovských hustôt pravdepodobností (GMM - Gaussian Mixture Model). Tieto sú určené svojimi stredmi, zahrnutými do vektoru stredných hodnôt a maticou kovariancií. Pre systémy určené na prepis spojitaj reči (LVCSR) sa počet týchto komponent pohybuje v desiatkách tisíc, niekedy viac než stotisíc.



Obr. 1: Ukážka adaptácie akustického modelu lineárnou transformáciou

Pri riešení úlohy adaptácie akustických modelov na rečníka sa najčastejšie stretáme s dvoma prístupmi, prvý je adaptácia pomocou maximálnej aposteriornej pravdepodobnosti (MAP) a druhý je adaptácia použitím maximálne vierohodnej lineárnej regresie (MLLR - Maximum Likelihood Linear Regression).

Jedným z hlavných problémov procesu adaptácie modelu je malé množstvo adaptačných dát na úpravu všetkých parametrov modelu. Toto je možné obísť združením viacerých komponent AM do jednej triedy, zn. ( $\Omega$ ), a určením príslušnej adaptácie pre celú túto triedu. MLLR, ktorá spadá do kategórie lineárnych transformácií, je typickým príkladom techniky, ktorá využíva tento postup a obmedzuje tak počet volných parametrov modelu. Efektívne tak dokáže nájsť transformačné vektory aj pre tie parametre, pre ktoré nejestvujú dáta v adaptačnej množine.

## 2.1. MLLR

MLLR je metóda výpočtu transformačných vektorov tak, aby akustické modely reprezentujúce výpoveď, presnejšie zodpovedali rozloženiam adaptačných dát, podľa obr.1. Z pohľadu adaptácie pre GMM sa ako najdôležitejšie javia hustoty výstupných pravdepodobností stavov modelu, vyjadrené funkciou  $\mathbf{b}_{jm}(\mathbf{o}_t)$ , transformačné vektory teda určujú posun ich stredov a kovariančných matic. MLLR takýmto spôsobom maximalizuje pravdepodobnosť generovania adaptačných dát novo zadaptovaným (SA) modelom. Transformačné matice stredov, resp. rozptylov, sú výsledkom hľadania maxima pomocou EM algoritmom z pomocnej funkcie, ktorú možno prepísať do tvaru

$$\frac{\partial Q(\boldsymbol{\lambda}, \bar{\boldsymbol{\lambda}})}{\partial \mathbf{W}_\Omega} = \sum_{e=1}^E \sum_{b_{jm} \in \Omega} \sum_{t=1}^{T_e} \gamma^e_{jm}(t) \log(e) \mathbf{C}_{jm}^{-1}(\mathbf{o}_t - \mathbf{W}_\Omega \tilde{\boldsymbol{\mu}}_{jm}) \tilde{\boldsymbol{\mu}}_{jm}^T \quad (1)$$

kde  $\mathbf{W}_\Omega$  je transformačná matica pre triedu  $\Omega$ ,  $\mathbf{C}_{jm}$  je kovariančná matica,  $\mathbf{o}_t$  je vektor pozorovania a  $\boldsymbol{\mu}_{jm}$  je vektor stredných hodnôt.

V prípade adaptácie stredných hodnôt rozložení je možné vektor nových stredov  $\bar{\boldsymbol{\mu}}_{jm}$  zapísať v tvare

$$\bar{\boldsymbol{\mu}}_{jm} = \mathbf{W}_\Omega \tilde{\boldsymbol{\mu}}_{jm} \quad (2)$$

keď transformačná matica  $\mathbf{W}_\Omega$  sa dá rozložiť na regresnú maticu

$\mathbf{A}_\Omega$  a vektor posuvu  $\mathbf{b}_\Omega$ .

$$\mathbf{W}_\Omega = [\mathbf{A}_\Omega, \mathbf{b}_\Omega] \quad (3)$$

Výpočet transformácie pre vektory stredov je najčastejší spôsob použitia MLLR adaptácie, preto je výklad transformácii obmedzený iba na tento prípad.

MLLR adaptácia si vyžaduje združovanie komponent do regresných tried. Pre naše experimenty sa testoval prínos dvoch rozdielných združovacích postupov. V prvom prípade sa nechali všetky komponenty modelu združiť do tried automaticky. Bol použitý prístup pomocou binárneho regresného stromu metódou od zhora-nadol. Pre rozhodovací algoritmus bola použitá Euklidovská miera v akustickom priestore. V druhom prípade boli komponenty združené na základe ich príslušnosti k fonetickým triedam, tak ako ich rozoznáva fonetika. Keďže sa na úrovni akustického modelu pracovalo s trifónmi, a fonetické rozdelenie je dobre zmapované len pre monofóny, muselo sa spraviť niekoľko apriorných predpokladov.

V práci [9] sa uvádza, že model každého trifónu je vo svojom strednom stave identický s modelom stredného monofónu a že iba okrajové stavy sú ovplyvnené kontextom. Prvý predpoklad preto bol, že všetky trifóny s rovnakým stredným monofónom majú patriť do jednej triedy. Druhý predpoklad bol, že pre jeden trifón by sa mali všetky jeho komponenty zadaptovať rovnako, preto boli do tejto triedy zahrnuté všetky komponenty daných trifónov.

Na základe týchto predpokladov boli testované tri varianty rozdelení (ďalej značené ako *Setup\_1*, *Setup\_2* a *Setup\_3*). *Setup\_1* mal rozdelenie podľa klasického delenia fonémov na : samohlásky krátke aj dlhé, dvojhásky, frikatívy, afrikáty, nazály, likvidy, explozívny a samostatná skupina pre ticho. *Setup\_2* mal rozdelenú skupinu samohlások na dve podskupiny podľa postavenia pier (zaokrúhlené a nezaokrúhlené) a *Setup\_3* podľa postavenie jazyka. Zvyšné triedy zostali pre tieto rozdelenia zachované.

### 3. Experimenty

Vyššie uvedené metódy boli testované pri konštrukcii rozpoznávača s pomocou nástrojovej sady HTK.

#### 3.1. Nastavenie rozpoznávača

Procesu adaptácie predchádzalo natréňovanie SI modelu, keď pre extrakciu príznakov bolo použité nasledovné nastavenie : 13 MFCC doplnených ich prvými a druhými deriváciami, dĺžka okna 25ms s 10ms prekryvom a Hammingovo váhové okno.

Pri akustickom modelovaní sa vychádzalo zo štandardnej sady 43 českých monofónov, doplnených o model ticha a krátkej pauzy. Táto bola rozšírená na sadu trifónov, u ktorých bolo vykonané štatistické zviazanie stavov. Ich celkový počet sa tak znížil z cca. 83k na približne 15k. V poslednom kroku sa pre každý stav modelu pridalo 5 Gaussiánov. Výsledný model teda obsahoval sadu 15k trifónov, kde každý model tvorili tri emitujúce stavy a 6 Gaussiánov na stav.

Pre jazykové modelovanie bol použitý trigramový jazykový model vytvorený na katedre /citeLM, ktorý obsahoval 340k slov.

Postup adaptácie bol nasledovný : SI model bol zadaptovaný v doch ktoroch, keď najprv bola určená globálna transformácia pre všetky komponenty. Táto sa v druhom kroku použila ako vstupná transformácia pre presnejšie určenie pre jednotlivé triedy.

### 3.2. Databáza signálov

Pre experimenty s rozpoznávaním a adaptáciami boli použité audio nahrávky z databáze SPEECON [12]. Táto databáza obsahuje audio nahrávky od 580 rečníkov, keď pre každého rečníka bolo nahraných cca. 300 signálov v prostrediach s rôznou úrovňou šumu na pozadí. Signály obsahujú jednoduché jednoslovné príkazy, slovné spojenia ale aj celé vety. Pre účely experimentov bola táto databáza rozdelená na tri disjunktné časti pre akustické tréningovanie, adaptáciu a testovanie. Všetky nahrávky v týchto množinách pochádzajú výhradne z prostredia OFFICE, ktoré sa vyznačuje minimálnou úrovňou šumu.

Trénovacia množina obsahovala cca. 60k signálov z množiny 190-ich rečníkov v celkovej dĺžke asi 51 hodín, nahraných kvalitných headset mikrofonom.

Testovacia množina obsahovala 275 signálov v celkovej dĺžke 27.5 minúty. Pre účely rozpoznávania bolo vybraných 11 dospelých rečníkov, keď sa nebral ohľad na ich pohlavie alebo rečový dialekt. Zastúpené boli Moravské, České aj Slezské nárečia, takisto muži ako aj ženy. Opäť boli použité len signály z prostredia OFFICE, nahrané pomocou headset mikrofónu. Obsahovo sa jednalo o signály so štruktúrou celej vety, ktoré sú obzvlášť vhodné na testovanie systému rozpoznávania plynulej reči.

Adaptačná množina obsahovala 170 signálov pre každého rečníka v celkovej dĺžke 4 minút, väčšinou išlo o nahrávky jednoslovných príkazov alebo výpovedí mien a adries.

### 3.3. Hodnotie výsledkov

Štandardným parametrom hodnotenia úspešnosti rozpoznávača je presnosť (ACC - Accuracy) definovaná ako

$$ACC = \frac{N - S - D - I}{N} \cdot 100\% \quad (4)$$

kde N je počet všetkých slov, S je počet nahradených slov, D je počet slov ktoré chýbajú a I počet slov vložených navyše. Pre účely hodnotenia prínosu adaptácie sa však častejšie používa chyba na úrovni slov (WER - Word Error Rate)

$$WER = 100 - ACC \quad (5)$$

### 3.4. Výsledky

V tab. 1 sú zhrnuté výsledky automatickej adaptácie pre rozličný počet regresných tried. Uvádzané sú len stredné hodnoty WER pre daný počet tried cez všetkých rečníkov. Z týchto hodnôt je zrejmé, že najlepších výsledkov bolo dosiahnutých pre 8 tried, a s  $mean(WER) = 20.08 \%$ . Oproti neadaptovanému systému (*Baseline*) to predstavuje zlepšenie o 20.25%.

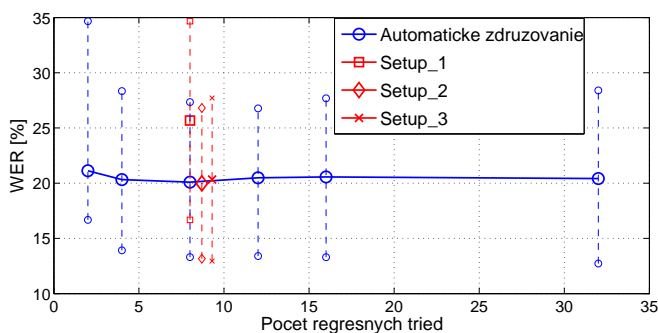
V tab. 2 sú zhrnuté výsledky ručného rozdelenia. Ako najlepšia sa ukázala varianta základného delenia (*Setup-1*) s  $mean(WER) = 19.98 \%$ , čo predstavuje zlepšenie o 20.91%.



	<i>WER</i> [%]						
	<i>Baseline</i>	<i>Classes_2</i>	<i>Classes_4</i>	<i>Classes_8</i>	<i>Classes_12</i>	<i>Classes_16</i>	<i>Classes_32</i>
<i>mean(WER)</i>	25.67	22.12	20.32	20.08	20.49	20.56	20.41

Tabuľka 1: Výsledky adaptácii pre automatické rozdelenie, 11 rečníkov

Počet regresných tried bol v tomto prípade identický ako u najlepšieho výsledku pre automatické delenie. Zvyšné dve nastavenia dosiahli o niečo horšie výsledky, s priemernými hodnotami *WER* 20.32% a 20.41%. Počet tried pre tieto nastavenia bol o jednu vyšší (9), je preto možné, že množina adaptačných dát neobsahovala dostatočné množstvo výpovedí so zodpovedajúcim fonetickým obsahom. Takto určené transformačné vektory preto nepresnejšie určovali posuv v akustickom priestore v porovnaní s prvým prípadom. Tento predpoklad potvrdzuje aj vývoj *WER* na obr. 2. pre automatické nastavenie, v ktorom je zrejmný nárast chybovosti pre vyšší počet tried ako 8. Podobných výsledkov redukcie chybovosti kombináciou vicerých prístupov bolo dosiahnutých aj v práci [4], samotnou MLLR adaptáciou o niečo horších.



	<i>mean(WER)</i> [%]
<i>Baseline</i>	25.67
<i>Setup_1</i>	19.98
<i>Setup_2</i>	20.32
<i>Setup_3</i>	20.41

Obr. 2: Stredné hodnoty a rozptyly *WER* pre automatické aj ručné rozdelenie v závislosti na počte regresných tried

Tabuľka 2: Výsledky adaptácii pre ručné rozdelenie, 11 rečníkov

## 4. Záver

Cieľom tejto práce bolo zmapovať prínos dvoch rozdielných klastrovacích prístupov na výsledok adaptácie akustického modelu na jedného rečníka v systéme rozpoznávania spojitých reči na bázi HTK.

Výsledky prezentované v predchádzajúcej časti ukazujú, že oba tieto prístupy dávajú veľmi podobné výsledky. Pre oba prístupy boli dosiahnuté najlepšie výsledky pre 8 regresných tried. Pri automatickom rozdelení je počet regresných tried určených množstvom adaptačných dát, pri ručnom rozdelení fonémov nebola táto informácia braná do úvahy. Pokiaľ by sa teda ukázalo, že pre zvolené rozdelenie nieje dostatok dát, muselo by ručné zloženie tried upraviť.

V ďalšej práci sa plánujeme zamerať na rozšírenie výzkumu adaptačných techník a ich použitia nielen na rečníka, ale aj prostredie. Jedná sa predovšetkým o možnosti aplikácie adaptácie na signály z prostredia s vysokou úrovňou šumu alebo z prostredia vyznačujúceho sa ďalekým dozvukom. Druhým plánovaným smerom budúceho výzkumu je aplikácia

dosiaľ nepoužívaných adaptačných techník ako MLLR vektoru príznakov (fMLLR - feature MLLR) alebo diskriminatívne adaptácie.

## Literatúra

- [1] Nouza, J., Zdansky, J., David, P. Fully automated approach to broadcast news transcription in Czech Language. *Text, Speech and Dialogue: Lecture Notes in Computer Science*, 2004, vol. 3206/2004, p.401-408.5
- [2] Psutka, J., Psutka, J., Ircing, P., Hoidekr, J. Recognition of spontaneously pronounced TV ice-hokey commentary. In *Proceedings of ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*. Tokyo (Japan), 2003 p. 83-86.
- [3] Cerva, P., Nouza, J., Kolerenc, J., David, P. Improved Transcription of Czech Parliamentary Speeches by Acoustic and Language Model Adaptation. In *SPEECOM'2006* St. Petersburg, 2006 p.25-29.
- [4] Cerva, P., Zdansky, J., Silovsky, J., Nouza, J. Study on Speaker Adaptation Methods in the Broadcast News Transcription Task. In *Lecture Notes in Artificial Intelligence, Text, Speech and Dialogue, LNAI 5246*, Springer-Verlag, 2008, p. 277-284, ISSN 0302-9743.
- [5] CERVA, P., NOUZA, J. MAP Based Speaker Adaptation in Very Large Vocabulary Speech Recognition of Czech. *Radioengineering*, September 2004, Vol. 13, No 3, p. 42-46, ISSN 1210-2512.
- [6] Rajnoha, J., Pollak, P. ASR Systems in Noisy Environment : Analysis and Solutions for Increasing Noise Robustness. *Radioengineering*, 2011, vol. 20, no. 1, p. 74 - 84.
- [7] Prochazka, V., Pollak, P. Performance of Czech Speech Recognition with Language Models Created from Public Resources. *Radioengineering*, 2011, vol. 20, no. 4, p. 1002 - 1008.
- [8] Leggetter, C.J., Woodland, P.C. Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models, 1995.
- [9] R. M. Schwartz, Y. Chow, S. Roucos, M. Krasner, and J. Makhoul. Improved hidden Markov modeling of phonemes for continuous speech recognition. In Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing 1984, pages 35.6.1–35.6.4, 1984.
- [10] Young, S. *The HTK Book for HTK Version 3.4*. Cambridge University Engineering Department, 2006.
- [11] Psutka, J., Müller, L., Matoušek, J., Radová, V., *Mluvíme s počítačem česky*, Academia, Praha 2006, p.47 - 56, ISBN:80-200-1309-1.
- [12] SPEECON database distributed through the European Language Resources Association. Available at : [http://catalog.elra.info/product\\_info.php?products\\_id=1095](http://catalog.elra.info/product_info.php?products_id=1095)

# Rozpoznávání řeči pomocí signálových procesorů

Robert KREJČÍ<sup>1</sup>, Václav HANŽL<sup>1</sup>

<sup>1</sup> České vysoké učení v Praze, Fakulta elektrotechnická, Katedra teorie obvodů

robert.krejci@centrum.cz, hanzl@noel.feld.cvut.cz

*Abstrakt:* Automatické rozpoznávání řeči je obecně výpočetně velmi náročná úloha. Jedním z cílů našeho výzkumu je hledat metody snížení výpočetní náročnosti a vyvinout takové algoritmy rozpoznávačů řeči, aby bylo možné je využívat i na relativně méně výkonných HW systémech a umožnit tak jejich fungování v přenosných a miniaturních aplikacích. Typickým příkladem využití může být hlasové ovládání některých funkcí strojů a přístrojů (výtah, kuchyňské spotřebiče, pomůcky pro nevidomé, chytré domy). Proto se v současné době zaměřujeme na algoritmy rozpoznávání samostatných slov – povelů.

Na *Letních doktorandských dnech 2011* byl prezentován softwarový nástroj *phpHMM*, který jsme vyvinuli a používáme ho na generování zdrojových kódů a datových struktur rozpoznávače řeči. Vstupem pro tento nástroj je model řeči ve standardním formátu HTK a seznam slov (povelů), které má rozpoznávač umět rozeznat. Výstupem jsou dva soubory v programovacím jazyce C – hlavičkový soubor a soubor se všemi funkcemi a datovými strukturami sestavený přesně na míru předloženému modelu řeči.

V současné době máme vyvinut rozpoznávač řeči, který je výsledkem přepisu matematických rovnic a pracuje s operandy v plovoucí řádové čárce (`float`). Výsledky této základní verze jsou brány jako referenční. Na základě tohoto referenčního rozpoznávače provádíme výzkum vlivu optimalizací na rychlost výpočtů a úspěšnost rozpoznávání.

Při vývoji rozpoznávače řeči tohoto typu je nutné vzít v úvahu specifika hardwarových architektur, které jsou v současné době na trhu. Konkrétně to znamená např. skutečnost, že moderní mikrokontroléry (tj. mikropočítače se zabudovanou datovou a programovou pamětí) mohou mít k dispozici poměrně velkou programovou paměť typu Flash (řádově až megabajty), zatímco datová paměť bývá značně menší (řádově stovky kilobajtů). Oproti klasickému přístupu je proto nutné jednoznačně rozdělit datové struktury rozpoznávače řeči na konstanty a proměnné a zvolit jejich umístění do vhodného typu paměti.

Jedna z optimalizačních metod spočívá ve spočítání výsledků předem u těch částí algoritmů, které nejsou závislé na vstupním signálu. V našem případě jsou výsledky vygenerovány jako datová pole umístěná v programové paměti. Ukazuje se, že tuto metodu lze aplikovat na mnoho částí rozpoznávače řeči, a to v podstatě bez snížení přesnosti rozpoznávání. Např. v algoritmu diskretní kosinové transformace (DCT) se vyskytuje výpočet kosinu, jehož argumenty však nejsou závislé na vstupním signálu. Pokud se vygenerované konstanty vhodně seřadí, je možné je během výpočtu číst z paměti postupně za sebou, tedy s nulovou spotřebou výpočetního výkonu, protože moderní HW architektury běžně mohou při čtení z paměti zároveň posunout ukazatel na další záznam. Pomocí tohoto typu optimalizací je možné nejen ušetřit značné množství času při výpočtu goniometrických funkcí, ale také zcela vynechat knihovní funkce `sin()` a `cos()` z programové paměti.

Další typ optimalizací spočívá ve vynechání konstant a příslušných operací, které se sice v modelu řeči vyskytují, ale ve skutečnosti se s nimi nepočítá.

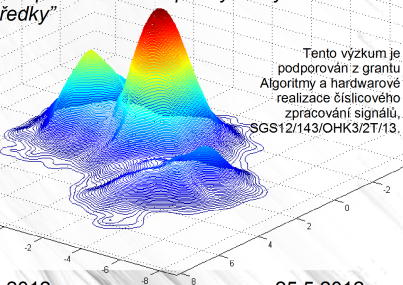
Vyvinuli jsme také optimalizační metody, které pomocí transformace parametrů vedou na výpočetní úkony typické pro číslicovou filtraci, a jsou tedy na architekturách signálových procesorů proveditelné rychleji oproti původní formě.

# Rozpoznávání řeči pomocí signálových procesorů

Robert KREJČÍ, Václav HANŽL

„Optimalizace algoritmů rozpoznávačů řeči pro systémy s omezenými HW prostředky“

ČVUT v Praze  
Elektrotechnická fakulta  
Katedra teorie obvodů  
Laboratoř zpracování řeči



Tento výzkum je podporován z grantu Algoritmy a hardwarové realizace číslicového zpracování signálů. SGS12/143/OHK3/2T/13.

Letní doktorandské dny 2012

25.5.2012

## Obsah

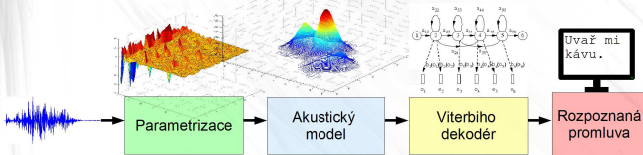
- Rozpoznávání řeči
- Nástroj *phpHMM* pro generování zdrojových kódů rozpoznávače řeči
- Laboratorní rozpoznávač řeči
- Metody optimalizací pro signálové procesory a mikrokontroléry
- Další vývoj

Robert Krejčí: Rozpoznávání řeči pomocí signálových procesorů

2/12

## Úvod

- 3 základní fáze rozpoznávače samostatných slov – povelů:
  - parametrizace
  - výpočet akustické podobnosti
  - průchod stavů modelu

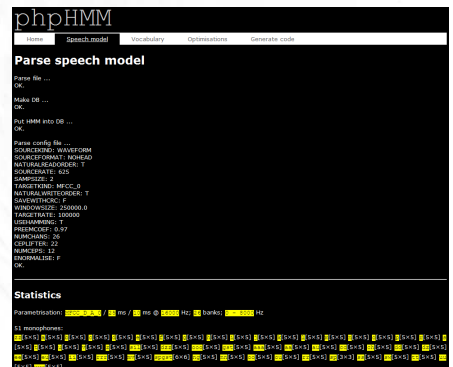


Robert Krejčí: Rozpoznávání řeči pomocí signálových procesorů

3/12

## Nástroj phpHMM

- Model řeči + seznam slov → sr.h, sr.c



Robert Krejčí: Rozpoznávání řeči pomocí signálových procesorů

4/12

## Laboratorní rozpoznávač řeči

- Čistý přepis matematických rovnic
- Bez optimalizací
- Plná přesnost v plovoucí řádové čárce (float)
- Reference pro další optimalizace
  - doba výpočtů
  - přesnost rozpoznávání

Robert Krejčí: Rozpoznávání řeči pomocí signálových procesorů

5/12

## Specifické vlastnosti mikro počítačů

- Velká paměť Flash [~1 MB] – vhodná pro uložení konstant
- Malá paměť RAM [~256 kB]
- Architektura přizpůsobená pro zpracování signálů:

$$c_j = \sum_{i=0}^{N-1} a_i b_i$$

Robert Krejčí: Rozpoznávání řeči pomocí signálových procesorů

6/12

## Optimalizace tabulkou konstant

- Princip: spočítat předem výsledky, které nejsou závislé na signálu
- Nemá vliv na přesnost rozpoznávání
- Hamming, FFT, banka mel-filtrů, DCT, lífrace kepstra
- Vynechání goniometrických funkcí z programové paměti

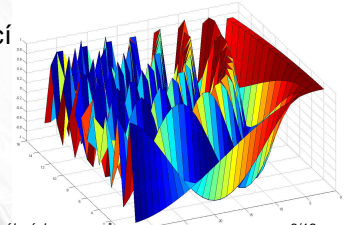
## Příklad: výpočet diskrétní kosinové transformace bez kosinu

- Matematické vyjádření:

$$c_i = \sqrt{\frac{2}{P}} \sum_{j=1}^P g_j \cos\left(\frac{\pi i}{P}(j-0,5)\right)$$

- Optimalizace pomocí tabulky konstant:

$$c_i = \sum_{j=1}^P g_j \text{tab}_{ij}$$



## Vynechání nevyužitých dat

- Původní tvar: 8/25 float

$$a_{gst} = \begin{bmatrix} 0 & 0,81 & 0 & 0 & 0,19 \\ 0 & 0,73 & 0,27 & 0 & 0 \\ 0 & 0 & 0,80 & 0,20 & 0 \\ 0 & 0 & 0 & 0,84 & 0,16 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

- Optimalizovaný tvar: 8 char +8 float

$$a_{gst} = \begin{bmatrix} 1 & 4 & 6 & 7 & 12 & 13 & 18 & 19 \\ 0,81 & 0,19 & 0,73 & 0,27 & 0,80 & 0,20 & 0,84 & 0,16 \end{bmatrix}$$

## Transformace dat

- Transformace výstupní pravděpodobnostní funkce na rychleji proveditelnou formu na architekturách signálových procesorů

- Původní tvar:  $\ln b_j(o) = \ln g - \frac{1}{2} \sum_{k=1}^N \frac{(o_k - \mu_k)^2}{r_k}$

- Transformovaná funkce s vektorovým součinem:

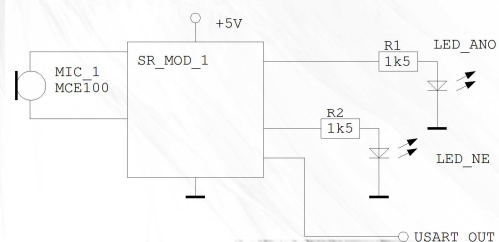
$$\ln b_j(o) = K - \sum_{k=1}^{2N} p_k x_k$$

## Další vývoj

- Převod částí rozpoznávače do pevné řádové čárky
- 32, 16, (8) bitů
- Úspora paměti (32 → 16 bitů)
- Možnost použití instrukcí SIMD
- Testování na architekturách TMS320C674x, ARM, MIPS

## Rozpoznávač povelů jako modul

- Komunikace s nadřazeným modulem pomocí definovaného protokolu
- Hlasové ovládání funkcí strojů a přístrojů



# Prozodická analýza a modelování

Jan BARTOŠEK<sup>1</sup>, Václav HANŽL<sup>1</sup>

<sup>1</sup> České vysoké učení v Praze, Fakulta elektrotechnická, Katedra teorie obvodů

[bartoj11@fel.cvut.cz](mailto:bartoj11@fel.cvut.cz)

*Abstrakt:* Při přepisu mluvené řeči na textové řetězce slov klasickými systémy pro rozpoznávání řeči (ASR) obvykle ztrácíme jistou část informace v podobě chybějících interpunkčních značek. To jednak snižuje čitelnost výsledného textu a v druhé řadě může být otázka lehce zaměněna za větu oznamovací (při zcela shodném slovosledu, který čeština v tomto případě připouští). Navíc v určitých situacích, kdy je například možné rozdělit danou promluvu na slova více způsoby (typickým příkladem je fráze „světlovnímají“), klasické rozpoznávače klíčovou prozodickou informaci pro rozhodování nevyužívají. Příspěvek se zabývá problematikou obou zmíněných případů. V popisovaném výzkumu se zaměřujeme hlavně na češtinu, avšak užití principy jsou zobecnitelné i na ostatní jazyky.

Termínem prozodie se označuje souhrn jevů vznikajících na suprasegmentální úrovni řeči, tedy na úrovni vyšší než jednotlivé segmenty (hlásky). Množina zvukových kvalit, které se pod tímto termínem skrývá je ve většině literatury následující: intonace, přízvuk, rytmus a mluvní tempo. Intonací myslíme průběh výšky hlasivkového tónu (F0) v čase, nejčastěji vnímáme větnou intonaci (melodii), která určuje modalitu vnímaného sdělení (větu oznamovací, otázku, ...). Intonační analýzu řečového signálu provádíme obvykle pomocí algoritmu pro detekci základní frekvence (PDA), kterých existuje celá řada a liší se mezi sebou jak přesností, tak odolností vůči rušení v signálu. PDA musí dále umět správně detekovat znělé a neznělé úseky řeči (V/UV), což v tranzientních úsecích řečového signálu není vždy jednoznačné. Výsledná podoba získané křivky je často dále upravována tak, aby zachycovala jen pro posluchače podstatnou informaci.

Přízvukem obvykle rozumíme takové slabiky, které mají oproti ostatním prominenci. Lexikální přízvuk je dán jazykem, v češtině se vyskytuje vždy na první slabice přízvukového taktu. Ten je obvykle tvořen jedním slovem, avšak jednoslabičná slova se často pojí k přecházejícím přízvukovým taktům (příklonky), nebo tvoří začátek následujícího přízvukového taktu (předklonky) a vznikají tak přízvukové takty nové. V češtině spolu zvukové kvality realizující přízvuk ne zcela korespondují (například oproti maďarštině), a proto je strojová detekce přízvuku složitější úlohou.

V prvním představeném experimentu jsme se snažili detekovat modality vět pouze z intonačního průběhu promluvy. Klasifikace byla realizovaná vícevrstvou dopřednou neuronovou sítí a byla snaha o nezávislosti na systému pro rozpoznávání řeči. Pro tento experiment byl vytvořen minikorpus skládající se z audioknih čtených předními českými herci, zaručující prozodickou bohatost dat. Lepší stylizace zjištěných intonačních kontur by jistě mohla vést k vyšší úspěšnosti klasifikace, avšak k tomuto účelu by bylo vhodné znát segmentaci promluvy na slabiky.

Druhým představeným experimentem je úloha rozdělení promluvy na přízvukové takty, tentokrát pomocí techniky skrytých Markovových modelů (HMM). Příznaky jsou extrahovány vždy ve střezech slabik, získaných force-alignmentem. Převod textových předloh na výstupní třídy (N-slabičné přízvukové takty) je provedeno vytvořeným lexikálním modulem. Podle fonologických studií by klíčovou roli pro určení hranic přízvukových taktů v češtině měla nést intonace, avšak výsledky tomu zatím nenasvědčují.

## Prozodická analýza a modelování

- Doktorand: Ing. Jan Bartošek
  - Školitel: Ing. Václav Hanžl, CSc.
- **Rámcové téma dizertace:**
- Využití prozodie v rozpoznávání souvislé řeči

LDD 2012

1

## Osnova přednášky

- Úvod - Co je prozodie
- Intonace, větná melodie
- Přízvuk, přízvukový takt
- Experiment 1 - klasifikace modalit v češtině
- Experiment 2 - přízvukové takty v češtině
- Závěr
- Použitá literatura

LDD 2012

2

## Úvod

- v našem výzkumu se zaměřujeme hlavně na češtinu, avšak užití principy jsou zobecnitelné i na ostatní jazyky

### Prozodie

- označuje souhrn jevů na suprasegmentální úrovni řeči
- intonace, přízvuk, rytmus a mluvní tempo
- patologická vsuvka: při vývojové dysartrii je narušena i prozodie řeči

LDD 2012

3

## Intonace

- průběh výšky hlasivkového tónu v čase (řeč = znělé + neznělé hlásky + pauzy)
- intonace na úrovni věty = větná melodie
- melodém – intonační předpis určité větné modality (ukončující klesavý pro oznam.větu)
- Intonace existuje i na nižších úrovních: v rámci slova, ale i slabiky (mikrointonace)
- F0 - základní hlasivkový tón
- Pitch - vnímaná výška tónu

LDD 2012

4

## Intonační analýza

- řečový signál je komplexní (zdrojem hlasivkový pulz, dále filtrován), zejména v transiencích často dvě sousední periody nemají shodný průběh
- klíčový je algoritmus pro detekci výšky základního tónu ze signálu (PDA)
- mnoho používaných, rozdíly v přesnosti odhadu F0, robustnosti proti šumu, algoritmus musí být schopen rozlišit znělost úseku (VUV)
- nejpoužívanější: ACF, AMDF, CEPS, CC
- postprocessing: median filtering, Viterbiho algoritmus

LDD 2012

5

## "Stylizace" průběhu F0

- posluchač vnímá F0 konturu spojitě
- snaha o jednodušší popis F0 kontury bez ztráty potřebné komunikační informace
- hlavním nositelem intonační informace je znělý střed slabiky (foneticky nukleus)
- ke změnám výšky ale může docházet i uvnitř těchto středů → pohyb intonace
- prozodické anotační standardy (ToBI, ...)

LDD 2012

6

## Přízvuková analýza

- dva typy přízvuku (prominence):
  - explicitní (větný) - určujeme ho sami → "důraz" na určitém slově ve větě
  - implicitní (lexikální) "přízvuk" → segmentace řeči na "přízvukové takty", klíčová role pro rytmus řeči
- umístění přízvuku v přízvukovém taktu závislí na konkrétním jazyce, v češtině je přízvuk na první slabice (fixed-stressed) vs. angličtina
- běžný mluvčí/posluchač dokáže určit hranice

LDD 2012

7

## Obecné zvukové prostředky pro realizaci prominence

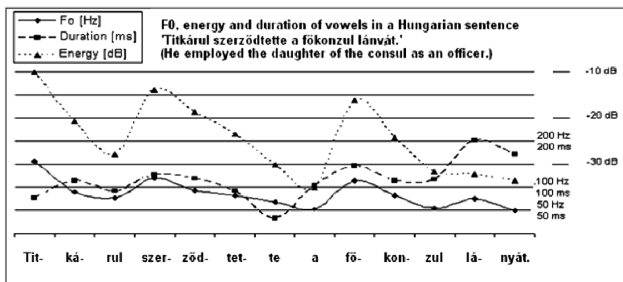
- 1) akustická intenzita slabik  
~ energie signálu
- 2) rozdílná délka slabik
- 3) intonace
- pro češtinu není strojové určení hranic přízvukových taktů pomocí těchto prostředků přímočaré (viz dále)

LDD 2012

8

## Ukázka realizace lexikální prominence

- maďarština - zvukové kvality spolu korespondují na přízvukných slabikách:



LDD 2012

9

## Realizace lexikální prominence v češtině

v češtině spolu zvukové kvality takto nekorespondují, přízvuk je realizován:

- jak zvýšením, tak snížením intonace
- jak zvýšením, tak snížením intenzity
- délka slabik nehraje významnou roli

LDD 2012

10

## Experiment 1

- Cíl experimentu: klasifikovat modalitu promluvy jen na základě intonačního průběhu bez znalosti hranic slabik promluvy (nezávislost na ASR)
- použitá trénovací a testovací data: české audioknihy namluvené předními českými herci → prozodicky bohatý materiál
- extrakce příznaků:
  - F0 – Praat Cross-Corellation
- Klasifikátor: ANN-klasická dopředná MLP síť

LDD 2012

11

## Experiment 1 - postprocessing

- Převod F0 na logaritmickou škálu
- Ořez okrajů intonační kontury
- Interpolace chybějících neznělých úseků
- Odstranění mikrointonace
- Rekonstrukce úrovní extrémů
- Podvzorkování výsledného signálu na požadovaný počet vstupů testované konfigurace ANN

LDD 2012

12



## Experiment 1 - výsledky

- matice záměn v %, MLP 64 vstupů, 10 neuronů ve skryté vrstvě, redukovaná množina dat, klasifikace na základě posledních 1200ms intonační kontury

Actual class → Predicted class ↓	?	!	,	.
?	46	30	26	28
!	1	7	1	1
,	28	40	57	22
.	25	23	16	49

LDD 2012

13

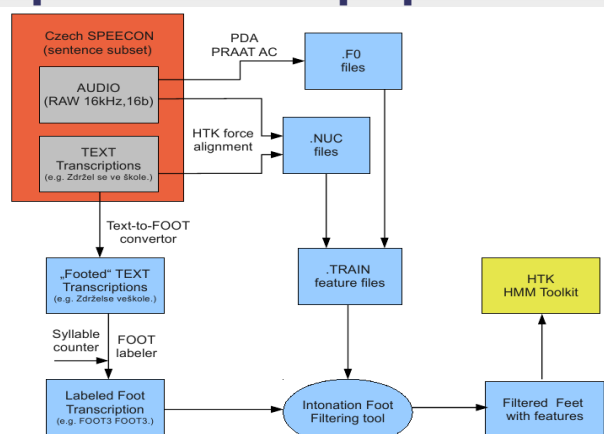
## Experiment 2

- cíl: pokusit se strojově rozdělit promluvu na přízvukové taktly pomocí prozodické informace, což může klasickému rozpoznávači pomoci při nerozhodnosti mezi hypotézami
- zdrojová řečová databáze: CZ SPEECON
- příprava trénovacích dat nebyla zcela jednoduchá, vytvořen lexikální modul pro převod vět na sekvenci přízvukových taktů podle mluvnických pravidel
- Realizace: HTK HMM Toolkit

LDD 2012

14

## Experiment 2 - příprava dat



LDD 2012

15

## Experiment 2 - výsledky

- Experiment je stále aktuální, není ukončen
- první výsledky při použití pouze intonačních příznaků dávají úspěšnost umístění hranice přízvukových taktů 49% na trénovací množině a 46% na testovací
- Poměrně značný prostor pro další vylepšování (kombinace ostatních zvukových kvalit ve vektoru příznaků)

LDD 2012

16

## Závěr

- Byly představeny dva nejdůležitější prozodické aspekty češtiny, jejichž využití může zlepšit úspěšnost a použitelnost současných ASR
- Experiment 1 – vypadá to, že je třeba znát umístění jednotlivých slabik v promluvě
- Experiment 2 – pro dotažení experimentu bude zřejmě nutná kolaborace s fonetiky
- práci od počátku ztěžuje nedostatek prozodicky označovaných dat

LDD 2012

17

## Použité zdroje

- Z. Palková, Fonetika a fonologie češtiny [Phonetics and phonology of Czech]. Karolinum, Praha, 1994.
- K. Vicsi and G. Szaszak, "Automatic segmentation of continuous speech on word level based on supra-segmental features", 2005, International Journal of Speech Technology, vol. 8, pp. 363–370,
- Hermes, D.J.: Stylization of pitch contours. In: Methods in Empirical Prosody Research, 2006, pp. 29-62.
- K. Silverman, M. Beckman, J. Pierrehumbert, M. Ostendorf, C. Wightman, P. Price, and J. Hirschberg, "Tobi: A standard scheme for labeling prosody," in International Conference on Spoken Language Processing, 1992.
- Bartošek, J. - Hanžl, V.: Intonation Based Sentence Modality Classifier for Czech Using Artificial Neural Network. NOLISP 2011, In Lecture Notes in Artificial Intelligence (LNAI) series , vol. 7015 - Advances in Nonlinear Speech Processing. Heidelberg: Springer, 2011, p. 162-169. ISBN 978-3-642-25019-4.

LDD 2012

18

# Automatický výpočet formantů a testování správnosti promluv

Pavel GRILL<sup>1</sup>

<sup>1</sup> České vysoké učení v Praze, Fakulta elektrotechnická, Katedra teorie obvodů

[grillpav@fel.cvut.cz](mailto:grillpav@fel.cvut.cz)

*Abstrakt:* Tento rozšířený abstrakt je věnován zčásti problematice zautomatizování formantové analýzy. Dále je zde popsána metoda pro testování správnosti řečených promluv a nakonec se letmo dotknu nově vytvořených webových stránek pro správu a tvorbu dětské řečové databáze vznikající v LANNĚ na ČVUT FEL v Praze.

Pro výzkum řečového signálu dětí trpících neurologickým onemocněním označovaným jako vývojová dysfázie preferuji metodu formantové analýzy a to především z toho důvodu, že formanty mají fyziologickou podstatu svého vzniku, kde charakterizují výskyt akustické energie v řečovém spektru a lze je dát do souvislosti s činností mozku a pohybem artikulačního aparátu. V minulých doktorandských dnech jsem se již věnoval problematice správnosti výpočtu a klasifikaci formantových kmitočtů, ale je nasnadě zopakovat, že pro správnou klasifikaci formantových kmitočtů je zapotřebí určit optimální šířku pásma vztaženou k určité hlásce a k určitému počtu formantů. Není-li tato podmínka splněna, tak často dochází k chybné klasifikaci formantů a k nezobrazení tzv. vokalického trojúhelníka. Tento fakt zabraňuje v použití formantové analýzy na analýzu kontinuálního signálu, ve kterém jsou obsaženy různé hlásky, promluvy slov a i celé věty. Metoda, která umožňuje použít formantovou analýzu na zmíněný signál je metoda založená na skládání formantů. Velice zjednodušeně ji lze popsat v několika krocích. Nejprve je nutné signál rozdělit na kvazistacionární úseky s překryvem, neboli provedeme segmentaci signálu. Pak se na těchto úsecích provede formantová analýza. Zde je několik kontrolních mechanismů, které se starají o její správnost. V posledním kroku se provede vlastní skládání formantů. Po jejich složení obdržíme správně klasifikované formanty pro celý signál.

Prozatím jsem se snažil o popis neurologického onemocnění u dětí používat pouze technické prostředky, čili byl prováděn rozklad řečového signálu pomocí formantové analýzy. Pro komplexní pohled na tuto neurologickou poruchu je ale zapotřebí získat i informace z tzv. druhé strany a zahrnout i obsahovou stránku zvukového projevu dítěte, která nebyla zatím řešena. Z tohoto důvodu byla vytvořena metoda testování správnosti řečených promluv. Tato metoda neřeší jenom správnost vyřčené promluvy, ale přidává další informace jako je počet hlásek ve slově a v promluvě, které hlásky jsou správně vyřčené, které hlásky jsou špatně, případně které hlásky chybí nebo přebývají a která hláska je nahrazena jinou hláskou.

Webové stránky, které se postupně upravují a přidávají další funkce vhodné k našemu výzkumu dětí s vývojovou dysfázií, byly původně vytvořeny za účel možnosti tvorby textu, který se používá pro nahrávání. Tento text byl vytvořen podle zkušeností předních logopedů z uznávaných testů. Jelikož v novém motolském grantu se děti řadí do tří skupin podle závažnosti onemocnění (lehká, střední a těžká) není úplně nejvhodnější nahrávat text s konstantním zněním tak, jak se to dělalo doted', protože děti jsou vícekrát nahrávány. Nyní je možné si pro každé dítě vytvořit tři různá řešení čteného textu a to tak, že buď vygeneruje text v originálním znění (originální pořadí slov), nebo se náhodně generují slova v rámci svých skupin. Poslední možností je náhodného generování pořadí všech slov.



## Automatický výpočet formantů a testování správnosti promluv

Pavel GRILL



- > Úvod Obecný úvod  
Formanty
- > Řešení problémů Automatický výpočet formantů  
Metoda testování promluv  
Generování promluv
- > Závěr Zhodnocení práce

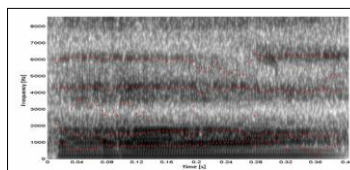
24. - 25. květen 2012

LDD 2012

2



## Formanty



Formanty

- F5
- F4
- F3
- F2
- F1

Vokalický  
trojúhelník

Vertikální pozice jazyka	Vokalický trojúhelník			
	Horizontální pozice jazyka			
	vysoké	přední	střední	zadní
středově	i, j	e, é	o, ó	u, ú
nízké			ɔ, ɔ̃	
hlída	řída	drůha	prvci	Třetí

24. - 25. květen 2012

LDD 2012

3



## Jak na to?

- > Problémy formantové analýzy
  - správná klasifikace formantů
  - správná šířka pásma
- > Problémy kontinuálního signálu
  - „nekonečnost“ signálu
  - měnnost typu promluv

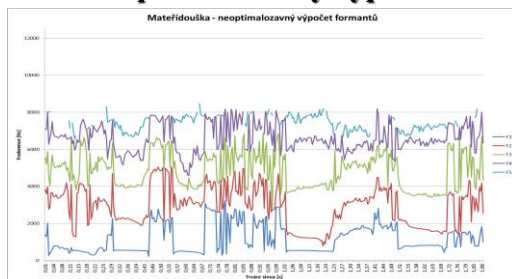
24. - 25. květen 2012

LDD 2012

4



## Neoptimalizovaný výpočet



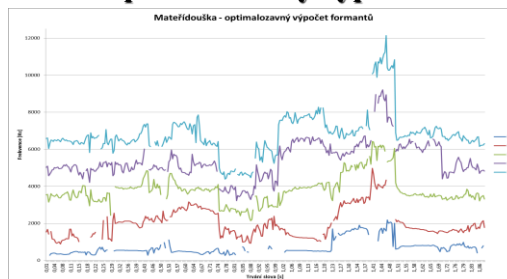
24. - 25. květen 2012

LDD 2012

5



## Optimalizovaný výpočet



24. - 25. květen 2012

LDD 2012

6





## Ukázka

Forma: Ukázka

24. - 25. květen 2012 LDD 2012 13



## Ukázka

### Test Lanina (TL)

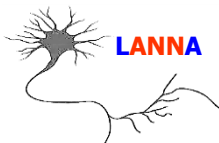
Datum nacizování: 2009-09-09  
Diagnóza: Vývojová dysfázie - Síťová  
Poznámky:

#### Nabrávky

1. 2012-05-01. Originální pořadí slov  
Poznámky: Nabrávka je diktována v databázi. --- Je zpracováno (rozlišitelné do adresářů 1 až 13). --- Je odlišováno (až po "Nacizovatelná slova" - adresáře 1 až 7). --- Chybí odlišování adresářů 8 - 11.  
a - o - u - e - i - ... - m - b - t - d - l - k - g - h - ch - ... - pr - la - va - pro - bo - mo - ber - tek - prít - ... - kolo - pávo - sokol - ...  
- papír - mělo - ... - děláček - pohádka - pokřmen - květen - ... - motýlek - šelma - popelice - ... - římskokatol - ...  
- pohádková vlna - kouzelný mláček - čarovný hamec - ... - voda - živa voda - živa a mrtvá voda - pramen a živa a mrtvá voda  
- Křídla lá Čarovná Křídla k hradě, pohádková vlna - ... - per - nes - ten den - kůl vln - hrad brat - jist jist - hle pte - kříd - kříd - květ květ - ... - nej - na - vln vln - ... - Podívej se na smetákové křídla.  
2. 2012-05-01. Nahodná pořadí ve skupinách  
Poznámky: Nabrávka je diktována v databázi. --- Je zpracováno (rozlišitelné do adresářů 1 až 13). --- Je odlišováno (až po "Nacizovatelná slova" - adresáře 1 až 7). --- Chybí odlišování adresářů 8 - 11.  
a - o - u - e - i - ... - l - ch - m - k - d - g - h - t - r - ... - va - pro - prít - pr - tek - ber - la - nes - ... - papír - pávo - sokol - ...  
- kolo - mělo - ... - pohádková - pohádková - křídla - pohádková - ... - motýlek - popelice - ... - římskokatol - ...  
- pohádková vlna - čarovný hamec - kouzelný mláček - ... - živa voda - pramen a živa a mrtvá voda - živa a mrtvá voda - voda  
- Křídla lá Čarovná Křídla k hradě, pohádková vlna - ... - kříd - kříd - vln vln - ten den - nej - na - per - nes - květ květ - ...  
- jist jist - hle pte - kůl vln - hrad brat - ... - Podívej se na smetákové křídla.  
3. 2012-05-01. Nahodná pořadí ve skupinách  
Poznámky:  
a - o - u - e - i - ... - l - d - m - b - ch - g - k - h - t - r - ... - pro - mo - pro - ber - prít - la - va - h - kř - ... - kolo - pávo - papír - ...  
- mělo - sokol - ... - pohádková - pohádková - děláček - květen - ... - motýlek - popelice - ... - římskokatol - ...  
- kouzelný mláček - pohádková vlna - čarovný hamec - ... - voda - pramen a živa a mrtvá voda - živa a mrtvá voda  
- Křídla lá Čarovná Křídla k hradě, pohádková vlna - ... - nej - na - ten den - kříd kříd - vln vln - per - nes - květ květ - ...  
- hle pte - hrad brat - květ květ - per - nes - ... - Podívej se na smetákové křídla.



## Konec



➤ Děkuji za vaši pozornost

Tato práce byla podpořována částečně grantem „Korelace MR traktografie, EEG analýzy a počítačového zpracování řečového signálu u dětí s vývojovou dysfázií“ 10-14, NT11643 a částečně grantem „Algoritmy a hardwarové realizace číselového zpracování signálu“ 1p-1, SGS SGS121420H3Z7F13.

# Odhad základní hlasivkové frekvence v patologické řeči

Lukáš BAUER

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

[bauerlu3@fel.cvut.cz](mailto:bauerlu3@fel.cvut.cz)

**Abstrakt:** Detekce základní hlasivkové frekvence ( $f_0$ ) je užívaná v řadě aplikací mezi které patří: rozpoznání mluvího, komunikace, lingvistika nebo detekce řečových patologií. Ve světě existuje již řada algoritmů detekujících  $f_0$ , ale žádný z těchto algoritmů nefunguje dostatečně přesně a robustně vzhledem k problematice detekce řečových patologií. To je dáno tím, že v patologické řeči se vyskytuje velké množství šumu, který vzniká právě poruchami hlasového traktu nebo nervové autonomní oblasti ovládající systém vytvářející hlas. Samotná detekce  $f_0$  však nevede k detekci řečových patologií, ale pouze pomáhá k detekci dalších řečových parametrů, které na patologii hlasu poukazují. Mezi tyto parametry patří amplitudová nestabilita (shimmer), frekvenční nestabilita (jitter) a odstup harmonické od šumu (HNR). Tyto parametry se dají detekovat pouze na prodloužených fonacích samohlásek, což zjednodušuje detekci. Tento článek shrnuje současnou situaci algoritmů detekce  $f_0$  ve světě sloužících k detekci řečových patologií.

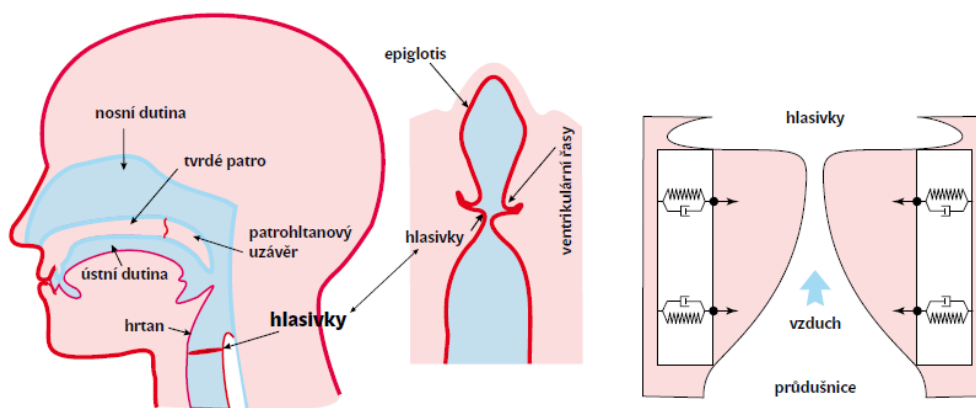
## 1. Úvod

Záznam prodloužené fonace je tvořen vibracemi hlasivek a dalším nastavením částí hlasového traktu, viz obrázek 1. Signál se pak skládá z periodické části tvořené hlasivkami (námi hledané základní hlasivkové frekvence  $f_0$ ) a přidaného šumu, který je tvořen převážně poruchami řečového ústrojí. Základní hlasivková frekvence je pro každého člověka charakteristická. Záleží převážně na pohlaví a věku daného jedince viz tabulka 1.

Výpočet  $f_0$  je důležitým krokem v řadě řečových aplikací jako je syntéza řeči, detekce pohlaví, rozpoznání mluvího [1]. Řečové patologie, poruchy hlasového ústrojí, se vyznačují velkou mírou šumu v signálu. Řada metod vedoucích k odhalení řečových patologií je založena na správné detekci základní hlasivkové frekvence signálu.

Běžně používané metody detekce základní hlasivkové frekvence však nejsou využitelné v patologických promluvách, protože nevedou k označení okamžiků potřebných k detekci ostatních vokálních parametrů. Mezi tyto parametry patří jitter (frekvenční nestabilita), shimmer (amplitudová nestabilita) a HNR (odstup harmonické od šumu) [2]. Pro detekci těchto parametrů je ale nutné mít dostatečně přesně detekované jednotlivé periody signálu.

Z tohoto důvodu je nutné navrhnout nové automatizované a dostatečně robustní metody detekce základní hlasivkové frekvence, sloužící k detekci ostatních řečových parametrů, pomocí nichž jsme schopni detekovat řečové patologie.



Obrázek 1: Vznik řečového signálu (Vesmír 87, prosinec 2008)

## 2. Existující algoritmy detekce základní hlasivkové frekvence

Řada algoritmů detekci  $f_0$  v řeči s výskytem patologií řeší danou problematiku v časové a některé ve frekvenční oblasti. Každá tato oblast má své přednosti a svá omezení obzvláště z hlediska přesnosti a časové výpočetní náročnosti. Hlavním představitelem časové oblasti je korelační funkce, zatímco hlavním představitelem frekvenční oblasti keprální analýza [3]. Když odhlédneme od řešení v dané oblasti, existují ve světě dva směry detekce: 1) pomocí porovnání podobnosti period (waveform matching – korelace), 2) pomocí detekce špiček (peak picking).

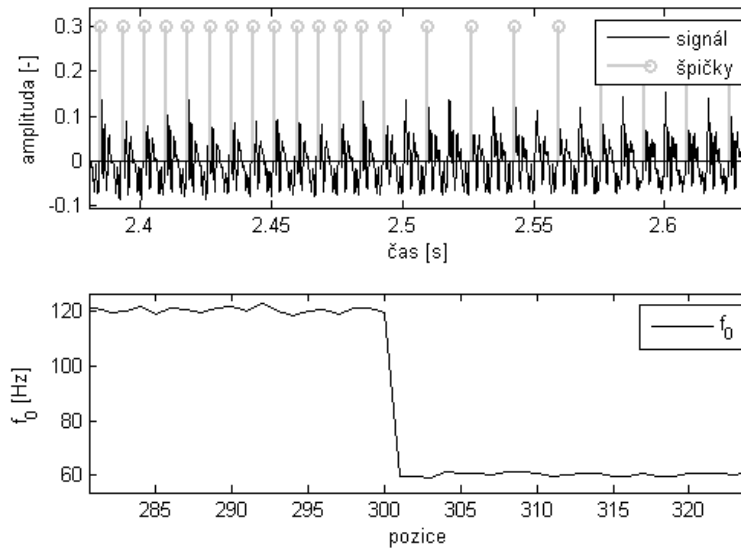
V dnešní době patří mezi nejpoužívanější světově uznávané algoritmy algoritmus Praatu [3], který je on-line dostupný a komerční algoritmus MDVP [4]. Řada dalších algoritmů, pracujících se statistickými metodami rozpoznávání [5] nebo například vlnkové transformace [6] se občas objeví, ale většinou demonstruje jenom sílu algoritmu v dané specifické oblasti, nemá však širší využití.

Algoritmus Praatu je založen na autokorelaci nebo vzájemné korelaci signálu, záleží na výběru uživatele. Základní hlasivková frekvence je hledána v rozmezí  $f_{0min} : f_{0max}$ , který je daný tabulkou 1. Je vybráno okno délky  $3 \times f_{0max}$  ve kterém je prováděna korelace. V signálu jsou nalezena korelační maxima a poté je nalezena optimální cesta. Největším problémem korelace je zdvojení (pitch doubling) a půlení (pitch halving) vyskytující se  $f_0$ . Tento problém je zobrazen na obrázku 2, je však problematicky potlačitelný.

Tabulka 1: Frekvenční rozsah hlasivkového tónu

	$f_{0typ}$ [Hz]	$f_{0min}$ [Hz]	$f_{0max}$ [Hz]
<b>Muži</b>	125	60	200
<b>Ženy</b>	225	150	350
<b>Děti</b>	300	200	500

Algoritmus MDVP využívá metody detekce špiček v signálu. Tato metoda dosahuje poměrně velké přesnosti. V případech, kdy dochází k velkému zastoupení šumu v signálu, silnější patologii, dochází u tohoto algoritmu k nejasně definovaným místům detekce špiček, viz obrázek 3.

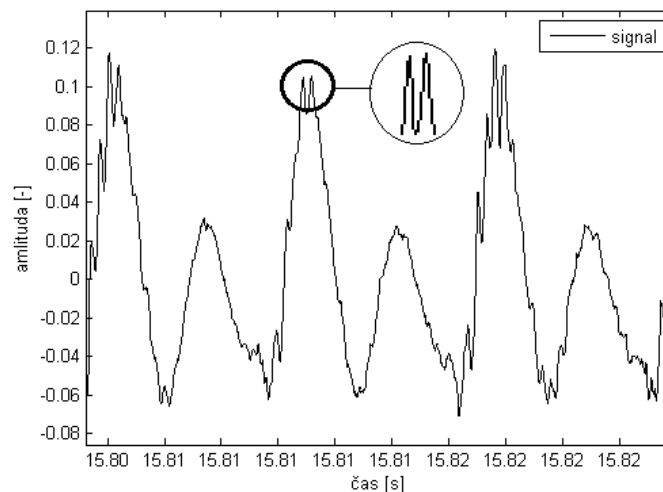


Obrázek 2: Přeskok na poloviční  $f_0$

### 3. Nově navržené algoritmy

Během návrhu nových algoritmů byla odhalena řada nectností některých přístupů. Nejdříve bylo snahou vytvořit algoritmus fungující na základě korelační funkce, ale v průběhu bylo zjištěno, že chybné přeskoky na poloviční nebo dvojnásobnou frekvenci jsou složitě odstranitelné a vyvolávají řadu dalších chyb. Proto nebyla zatím tato metoda úspěšně implementována.

Následně byl navržen algoritmus detekce špiček pomocí maxim v signálu, viz popis algoritmu 1, obrázek 4. Při návrhu tohoto algoritmu byla zjištěna problematická detekce správné pozice špičky, kde není zjevné, které maximum je správné. Algoritmus se snaží udržovat správnou pozici špičky vzhledem k předchozím výskytům špiček v signálu.

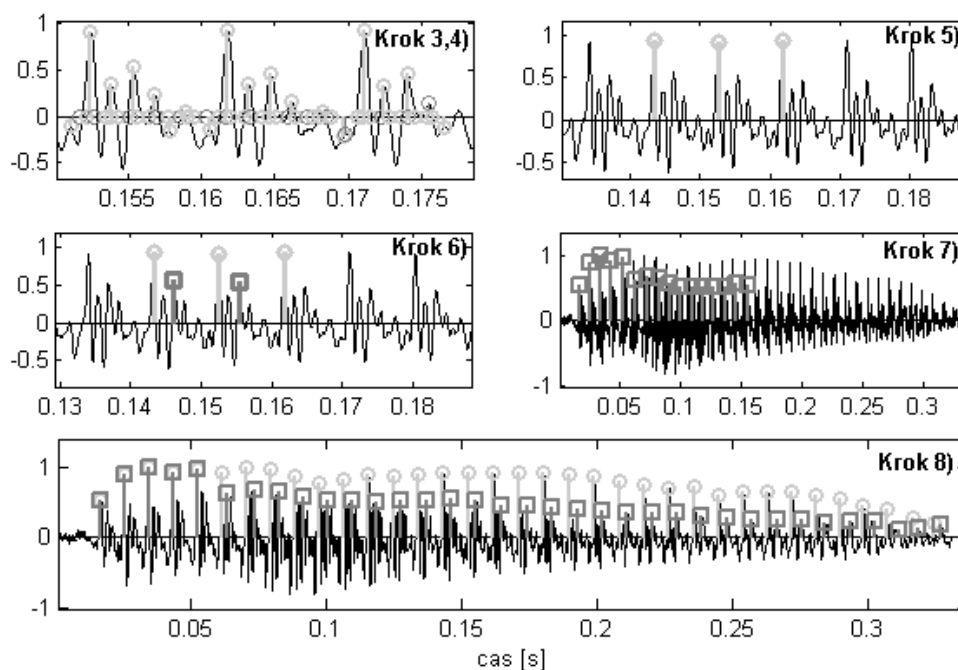


Obrázek 3: Problematická detekce periody pomocí detekce špiček



#### Popis algoritmu 1

- 1) Určení průměrné  $f_0$  pomocí Welchovy metody
- 2) Určení dominantní energie
- 3) Detekce všech maxim (minim) v segmentech signálu o délce  $T_0$
- 4) Maxima (minima), která nejsou lokálními extrémy, jsou vynulována
- 5) Detekce maxim (minim) probíhá od středu signálu k jeho krajům
- 6) Detekce maxima mezi dvěma prvními detekovanými maximy v kroku 5, tedy posun o přibližně půl periody vede k odstranění špatné počáteční detekce, viz obrázek 4 krok 6, špičky označené čtverečky
- 7) Detekce maxim (minim) podle nově detekovaných pozic z bodu 6) od středu signálu k jeho krajům
- 8) Výběr mezi výsledky kroku 5 nebo kroku 7 (špičky s vyšší energií)



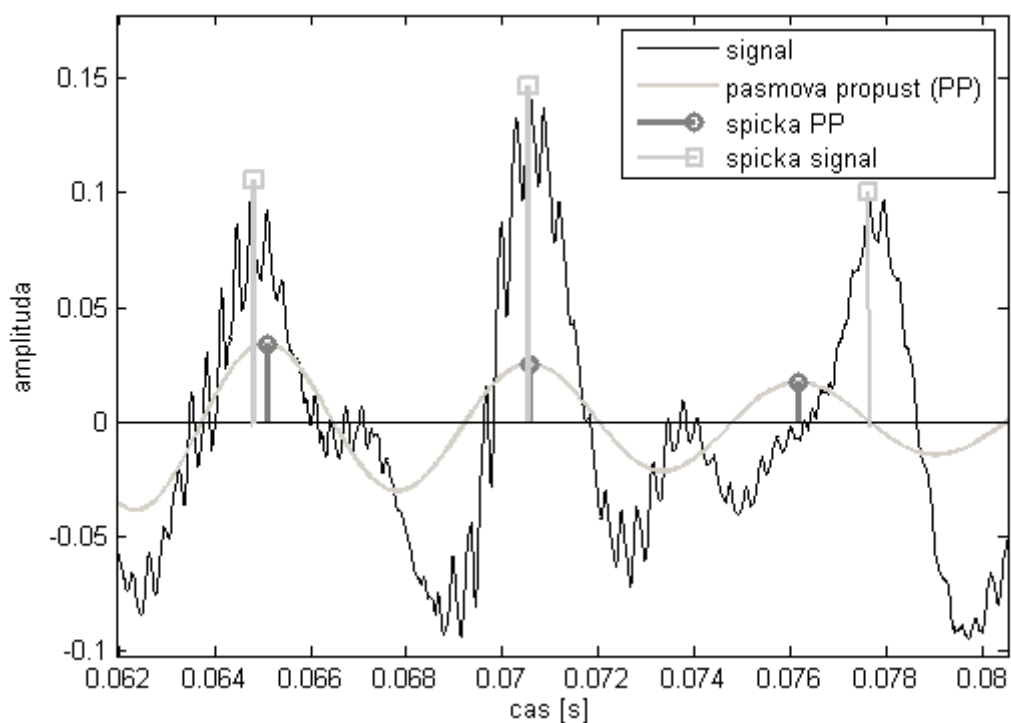
**Obrázek 4: Metoda detekce  $f_0$  pomocí maxim**

Pro potlačení výše popsaného nedostatku jsou hledána maxima v signálu filtrovaném pásmovou propustí. Tím by mělo dojít k potlačení šumu v signálu a vyplynutí správné pozice maxima. Nedostatkem tohoto algoritmu je jeho časová náročnost. Rozbor algoritmu viz popis algoritmu 2, obrázek 5.

V průběhu návrhu algoritmů byla navržena metoda pro detekci průměrné  $f_0$  v signálu pomocí Welchovy metody s nastaveným prahem. Hlavní nevýhoda Welchovy metody je v průměrování okna. U pacientů, kterým v průběhu fonace plynule klesá  $f_0$ , dochází ke zkreslení průměrné  $f_0$  a k ovlivnění výsledků algoritmu. To je částečně potlačené detekcí špiček od středu signálu směrem do krajů.

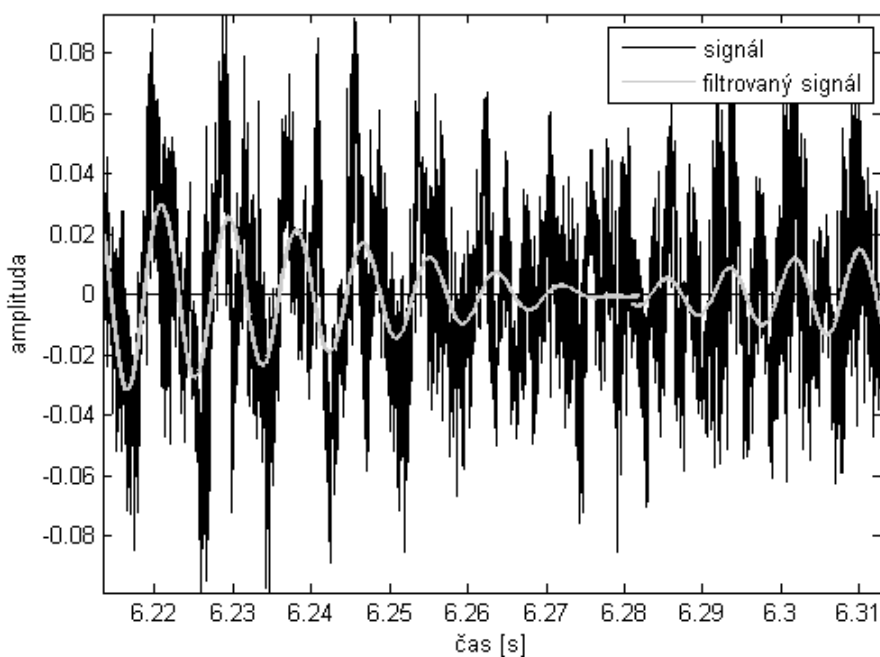
#### Popis algoritmu 2

- 1) Určení průměrné  $f_0$  pomocí welchovy metody
- 2) Výběr úseku a opětovná detekce  $f_0$  úseku, pokud se liší o více jak 50 Hz od průměrné  $f_0$  celého záznamu, je stanovena na její hodnotu
- 3) Aplikace pásmové propusti s šířkou pásma 10 Hz  $f_0$  úseku
- 4) Nalezení maxim pásmovou propustí
- 5) Nalezení maxim v okolí maxim detekovaných pásmovou propustí, ( $2/5 T_0$  před,  $1/5 T_0$  za pozicí maxim, kde  $T_0=1/f_0$ )



**Obrázek 5: Detekce  $f_0$  pomocí pásmové propusti**

U některých pacientů dochází k takové míře řečové patologie, že poté není možné detekovat základní hlasivkovou frekvenci. Tyto části signálu jsou stanoveny na základě poklesu amplitudy pásmové propusti pod stanovený práh, viz obrázek 6.



**Obrázek 6: Detekce chybné části signálu pomocí pásmové propusti**

## 4. Závěr

Návrh a vývoj metod detekce základní hlasivkové frekvence začal již v 70. letech minulého století. Dosud však zatím nebyla vyvinuta metoda, která by detekovala přesné okamžiky pro určení základní hlasivkové frekvence. Námi navržené algoritmy byly porovnány na databázi krátkých záznamů (do 0,5 sec) fonace vokálů /a/, /i/ a /u/ označované pomocí programu Praat. Databáze obsahuje 63 záznamů zdravých pacientů a 70 záznamů pacientů patologických. Výsledky dosahují 88% přesnosti pro metodu maxim a téměř 84% úspěšnosti pro metodu pásmové propusti.

Důležité je si uvědomit problematiku porovnatelnost algoritmů. Ve světě neexistuje žádný návod, jak správně detekovat základní hlasivkovou frekvenci. Krátká databáze záznamů není v tomto případě ideální, protože algoritmus se nestačí plně stabilizovat. Porovnáme-li však naše výsledky ostatních řečových parametrů, mezi které patří jitter, shimmer a HNR s výsledky programu Praat [7], který bereme jako referenční, začínáme dosahovat zajímavých hodnot korelace. Výsledky jsou porovnány pomocí Pearsonova korelačního koeficientu na databázi obsahující nahrávky vokálů /a/, /e/, /i/ a /o/, jejichž délka se pohybuje od 5 do 40 sekund. Databáze obsahuje 107 promluv pacientů s ranou Parkinsonovou nemocí a 57 promluv zdravých mluvčích kontrolní skupiny. Výsledky pro parametr HNR dosahují 97% míry korelace, pro shimmer 87% míry korelace a pro jitter 84% míry korelace pro metodu maxim a 66% míry korelace pro metodu pásmové propusti.

Z výsledků vyplývá, že algoritmy fungují poměrně dobře pro parametr HNR a shimmer, pro parametr jitter je vhodné přejít na detekci  $f_0$  pomocí metody waveform-matching. Tu se zatím snažíme implementovat metodou, která převádí signál do obrazové oblasti, tedy informace uložená v signálu narůstá o druhý rozměr. Poté pomocí jednotlivých pixelů je EM algoritmem určována správná pozice periody. Metoda však zatím není plně implementována a otestována.

## Reference

- [1] Camacho, A.; A sawtooth waveform inspired pitch estimator for speech and music, *University of Florida, Dissertation thesis*, 2007
- [2] Baken, R. J.; Orlikoff, R. F.: *Clinical Measurement of speech and voice 2nd edition – Singular in San Diego*, 2000
- [3] Boersma P.: Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound, In *Proceedings. Of the Institute of Phonetics Sciences*, 17: 97-112, 1993
- [4] Kay Elemetrics Corp.: *Multi-Dimensional Voice Program (MDVP) Model 5106: Software Instruction Manual*. Lincoln Park, Kay Elemetrics, 2003.
- [5] Mitev, P.; Hadjitodov, S.: Fundamental frequency estimation of voice of patients with laryngeal disorders, *Information Sciences Journal*, Vol.156, Issues 1-2, 1 November 2003 , pp. 3-19
- [6] Ercelebi, E.: Second generation wavelet transform-based pitch period estimation and voiced/unvoiced decision for speech signals, *Applied Acoustics*, Vol. 64, Issue 1, 2003, Pages 25–41
- [7] Boersma, P.; Praat, a system for doing phonetics by computer. *Glott Int* 2001;5:341-345

# Automatická segmentace hlásek při rychlém opakování slabik (/pa/ – /ta/ – /ka/) u hypokinetické dysartrie

Michal Novotný

České vysoké učení v Praze, Fakulta elektrotechnická, Katedra teorie obvodů  
novotm26@fel.cvut.cz

*Abstrakt:*

Analýza artikulace může být užitečným nástrojem pro rozlišování zdravých a dysartrických mluvčích s Parkinsonovou nemocí (PN). Cílem této práce je návrh algoritmu pro automatickou segmentaci konsonant a vokálů v promluvě, založené na rychlém opakování slabik /pa/-/ta/-/ka/. Data určená pro tvorbu algoritmu byla složena z PN i zdravých promluv v nichž byly manuálně označeny pozice explozí (E), počátků vokálů (V) a okluzí (O). Navíc byl proveden výpočet délky doby mezi E a V tzv. voice onset time (VOT). Při testování algoritmu byly nastaveny tři meze 5 ms, 10 ms a 20 ms pro rozdíl mezi délkou VOT a pro rozdíly mezi ručním a automatickým označení jednotlivých pozic. Míra úspěšnosti se při srovnání délky VOT pohybovala mezi 68.2 – 90.5 %, 44.1 – 75.2 % a 57.2 – 83.5 % pro zdravé, dysartrické a pro všechny dohromady. Tato práce ukazuje algoritmus založený na spektrálních, Bayesovských a polynomiálních přístupech, který je schopen detekovat pozice konsonant a vokálů ve zdravých a dysartrických promluvách.

## 1. Úvod

PN postihuje mediální oblast mozku, jmenovitě oblast substantia nigra (SN), ze které vedou dopaminoergní neuronové cesty, do putamen. Poškození SN vede k poklesu koncentrace dopaminu a způsobuje motorické poruchy [1]. Mezi čtyři základní motorické poruchy vyskytující se u parkinsoniků, patří tremor (neovladatelný třes), akinesie nebo bradykinesie (nepohzblivost nebo pomalost pohybů), ztuhlost (zvýšení svalového tonu) a nestabilní držení těla [1]. Dalšími běžnými příznaky PN jsou poruchy nálad, chování, vnímání a různé změny řeči označované jako hypokinetická dysartrie [1]. Podle uveřejněné studie [2] se u lidí s PN vyskytují vady řeči s pravděpodobností 70 – 90 %.

Jsou definována tři stadia dysartrie, mírná (mild), střední (moderate) a těžká (severe)[3]. Existuje mnoho hlasových charakteristik spojených s hodnocením PN, tři základní oblasti tvoří *fonace, artikulace a prozódie* [4].

Mezi jednu z úloh pro hodnocení přesnosti artikulace u dysartrie patří diadochokinetická (DDK) úloha, která měří schopnost jedince provádět rychlé opakované pohyby. Pro účely analýzy se obvykle používá metoda opakování konsonanty a vokálu, kdy je pacient požádán, aby co nejrychleji a co nejdéle opakoval sekvenci slabik /pa/ – /ta/ – /ka/ [5].

Tato práce se zabývá segmentací promluv získaných v DDK úloze a jejím cílem je navržení algoritmu pro automatické označení pozic vokálů a exploziv v signálu.

## 2. Metodika

### 2.1. Data

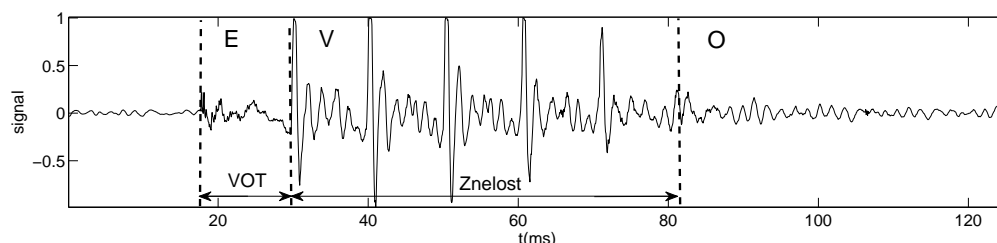
Data použitá pro vývoj algoritmu jsou součástí předchozí studie [6], v rámci které byly pořízeny zvukové nahrávky 46 rodilých mluvčích, z nichž 24 (20 mužů a 4 ženy) bylo diagnostikováno s časným stádiem PN a jejich nahrávky byly pořízeny před zahájením symptomatické léčby. Data zdravé kontrolní skupiny byla pořízena od 22 účastníků (15 mužů a 7 žen) bez historie neurologických onemocnění.

V rámci studie [6] byly nahrány promluvy při úloze pro charakterizaci DDK, při které byli účastníci vyzváni, aby co možná nejdéle a nejrychleji opakovali /pa/ - /ta/ - /ka/, nejméně pětkrát za sebou.

Výsledná trénovací množina algoritmu byla složena z databáze o velikosti 80 záznamů (1644 slabik /pa/, /ta/ nebo /ka/). Z tohoto počtu bylo 40 záznamů lidí trpících počátečním stádiem PN (753 slabik) a 40 záznamů zdravých lidí (891).

### 2.2. Segmentace signálu

Pro účely automatické analýzy promluvy je nutné segmentovat řečový záznam a detekovat tři základní pozice. První pozicí je exploze (E), počátek konsonant (/p/, /t/, /k/), která je charakterizována uvolněním orálního závěru a nárůstem energie explozivního šumu. Druhou pozicí je počátek vokálu (V), který představuje počátek kmitání hlasivek. Poslední pozicí je okluze (O) nebo-li závěr, během které je v řečovém signálu minimální energie. Jedna slabika /pa/ s vyznačenými polohami E, V a O je uvedena na obrázku 1.



Obrázek 1: Slabika /PA/ s vyznačenými pozicemi E, V a O a vyznačením VOT a znělosti

### 2.3. Segmentace na jednotlivé slabiky

Vzhledem k neznalosti počtu slabik v jednotlivých signálech je efektivnější nejprve signál rozdělit na menší segmenty obsahující vždy jednu pozici E, V a O. Výsledkem rozdělení jsou přibližné hranice slabik v signálu.

Signál je nejprve převzorkován na vzorkovací frekvenci  $f_s = 16\text{ kHz}$  a filtrován dolnoproputným (DP) filtrem s šířkou propustného pásma  $300\text{ Hz}$  a druhým (DP) filtrem s šířkou propustného pásma  $1100\text{ Hz}$ . Takto jsou získány dva průběhy. Na oba průběhy je použit špičkový detektor

$$s(n) = (1 - k(n)) |y(n)| + k(n)s(n - 1), \quad (1)$$

kde  $|y(n)|$  značí absolutní hodnotu  $n$ -tého vzorku signálu a  $k(n)$  je definováno jako

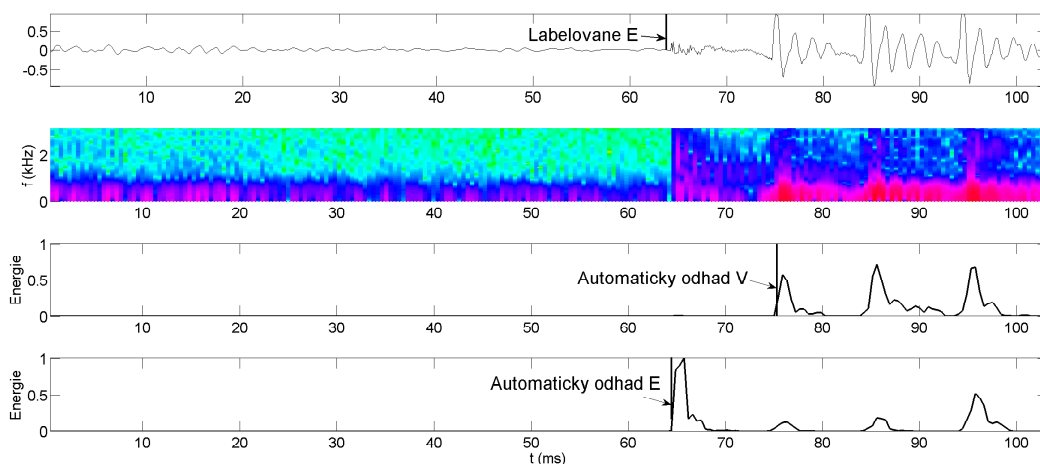
$$k(n) = \begin{cases} 0.9 & |y(n)| > s(n-1) \\ 0.997 & |y(n)| \geq s(n-1). \end{cases} \quad (2)$$

Výstupy špičkového detektoru jsou vyhlazeny metodou klouzavého průměrování a normovány. Ve vyhlazených průbězích jsou hledána lokální maxima s minimální vzdáleností 800 vzorků. Tímto postupem je předcházeno falešným detekcím.

Vektor nalezených slabik je vytvořen tak, že nalezená maxima ze signálu filtrovaného 1100 Hz DP filtrem jsou porovnávána s polohami maxim z druhého signálu a v případě chybějící polohy je vektor doplněn.

Z tohoto vektoru je poté vypočítána největší vzdálenost dvou sousedních poloh tato vzdálenost je rozšířena o určitý počet vzorků, čímž je získána délka jednotlivých segmentů. Z této délky je poté poměrově určena délka segmentu před a za dříve nalezenou polohou. Tímto postupem získáme hranice jednotlivých segmentů obsahujících pouze jednu slabiku, viz obrázek 1.

## 2.4. Detekce exploze



Obrázek 2: Princip detekce exploze, odshora labelovaná část segmentu, spektrogram segmentu, celková energie spektrogramu, energie spektrogramu nad 1500 Hz

Detekce exploze je realizována pouze na části signálu předcházející polohu udanou segmentací na slabiky, pomocí filtrace spektrogramu, k němuž je přistupováno jako k matici  $\mathbf{P}$  s  $m$  řádky a  $n$  sloupci. Pro zvýraznění jednotlivých špiček je stanovena filtrační mez. Všechny hodnoty v  $\mathbf{P}$ , které jsou menší než stanovená mez jsou položeny rovny nule.

Z filtrovaného spektrogramu jsou vysčítáním slouců získány dvě energetické obálky, první pro celý signál a druhá pouze pro frekvence nad 1500 Hz. Z obou obálek jsou vypočítána těžiště signálu, která značí medián signálu na časové ose. Poté jsou určeny relativní vzájemné polohy těžišť a polohy absolutní a v případě velkého absolutního přesunutí a nevhodné vzájemné polohy, jsou signály vyřazeny jako falešné detekce. Toto je založeno

na předpokladu, že většina energie je uložena ve vokálu, který se nachází na konci pro-  
 věřovaného segmentu.

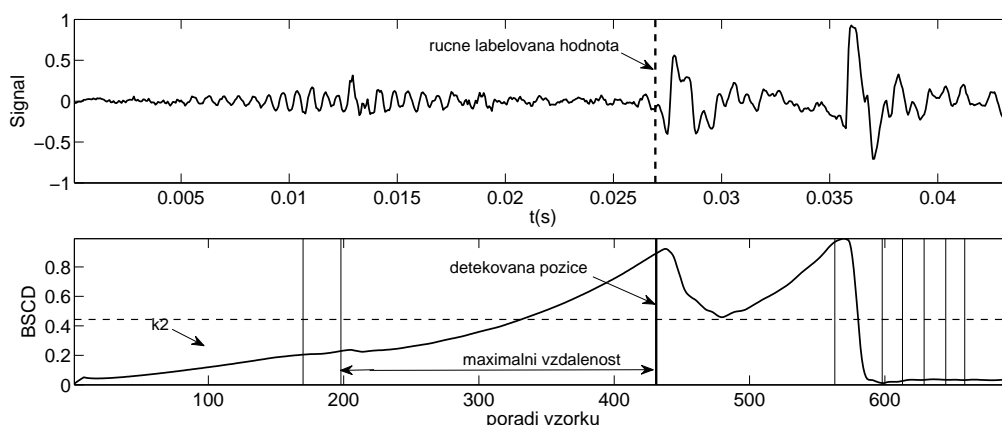
První energetická obálka je posléze filtrována klouzavým průměrováním a je v ní  
 detekována přibližná pozice  $V$  a v případě, že je tato pozice opět příliš v předu a hrozí  
 že pozice  $E$  je za hranicemi segmentu, je segment posunut podle odhadnuté pozice  $V$ .

Detekce  $E$  probíhá v energetické obálce signálu nad 1500 Hz, jelikož energie vokálu z  
 nižších frekvencí zakrývá peak exploze. Detekce je znázorněna na obrázku 2.

## 2.5. Detekce počátku vokálu

Detekce počátku vokálu je založena na prudkém nárůstu energie signálu, detekované po-  
 pomocí Bayesovského skokového detektoru (BSCD, Bayesian step changepoint detector)[7].  
 Ve výstupu BSCD jsou nalezena lokální maxima, mezi nimiž je i lokální maximum ozna-  
 čující  $V$ .

Při učení správného maxima byl využit tvar výstupu, který má před hledaným  $V$   
 dlouhý pozvolný nástup, proto bylo jako  $V$  označeno maximum následující po největší  
 mezeře mezi dvěma sousedními maximy. Tento postup je znázorněn na obrázku 3.

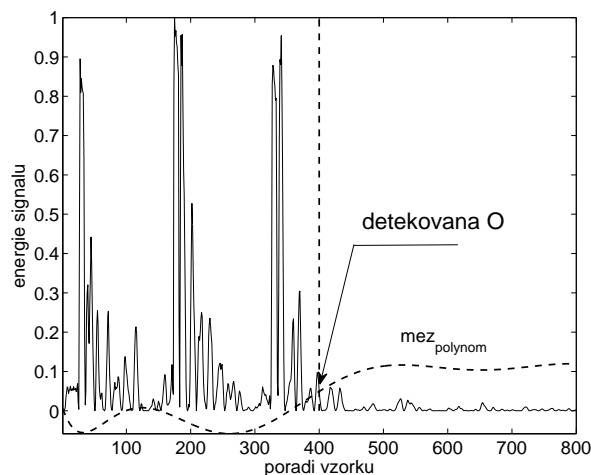


Obrázek 3: Postup detekce počátku vokálu

## 2.6. Detekce okluze

Principem detekce okluze je nalezení  
 proměnlivé meze, která se optimali-  
 zuje vzhledem ke změnám signálu.  
 Signál je nejprve filtrován dolno-  
 propustním filtrem se šířkou pásma  
 500 Hz. Následně je z filtrovaného  
 signálu vypočítán kvadrát.

Mez je poté tvořena invertova-  
 nou polynomiální aproximací devá-



Obrázek 4: Postup detekce O

tého řádu, která je zároveň posunuta o offset tvořený dvojnásobkem střední hodnoty. Tato mez může být vyjádřena jako

$$mez_{polynom} = \prod_{i=1}^9 (a_i \mathbf{x} + b(i)) + 2\bar{x}, \quad (3)$$

kde  $\mathbf{x}$  je vektor hodnot osy  $x$  s první hodnotou rovnou jedné a s délkou rovnou délce zkoumaného segmentu  $n$ . Koeficienty  $a_i$  a  $b_i$  jsou koeficienty  $i$  – tého řádu polynomu a  $\bar{x}$  je střední hodnota energie signálu.

## 2.7. Vyhodnocení úspěšnosti

Vyhodnocování funkčnosti detekcí jednotlivých pozic bylo prováděno vzhledem k celkovým počtům jednotlivých slabik, nikoliv k úspěšnosti detekcí v jednotlivých signálech.

Pro posouzení úspěšnosti byly využity ručně označené pozice jednotlivých E, V a O. Pro porovnání byl vypočítán rozdíl automaticky detekovaných pozic a těchto referenčních hodnot. Absolutní hodnota rozdílu pak byla porovnána se třemi mezemi a to 5 ms, 10 ms a 20 ms, jako úspěšná byla hodnocena každá detekce, která byla menší, nebo se rovnala stanovené mezi. Procentuální úspěšnost byla stanovena jako podíl součtu všech úspěšných detekcí slabik ku celkovému počtu.

Pro hodnocení úspěšnosti detekce VOT byly stanoveny stejné meze jako pro hodnocení detekce E, V a O. Porovnáván byl rozdíl mezi detekovanou a labelovanou délkou VOT, vypočítanou jako rozdíl  $VOT = V - E$ .

## 3. Závěr

Úspěšnost detekce VOT je možné rámcově porovnat s pracemi [8] a [9]. Studie [8] se zabývá hodnocením délky VOT pro účely rozlišování akcentů. Pro porovnání byly vybrány výsledky hodnocení subjektů majících jako mateřský jazyk americkou angličtinu, jelikož se nejvíce podobaly signálům v naší databázi. Úspěšnost nalezení VOT pod mez rovnou 10 % délky VOT je 74,9 % a průměrná chyba ve správných detekcích je 0,735 ms [8]. Náš algoritmus dosahuje v celkových výsledcích horšího skóre 57,2 % a 1,273 ms. Výsledky uvedené v práci [9], která se zabývá návrhem algoritmu pro měření VOT znělých i neznelých konzontan(/b/, /p/, /d/, /t/, /g/, a /k/) jsou pro mez 10 ms 72,6 % a pro mez 20 ms 87,8 %. Naše výsledky jsou na téměř srovnatelné úrovni, 68,1 % pro 10 ms mez a pro 20 ms mez jsou 83,5 %.

Při porovnávání je však nutné uvážit i rozdílné zaměření a hodnocení úspěšnosti jednotlivých prací. Data použitá v článku [8], byla poskytnuta lidmi s různým akcentem (americká angličtina, hindština, čínština), ve článku [9] byla využita databáze založena na americké angličtině, dalším limitujícím faktorem je zastoupení PN lidí, které detekci VOT výrazně ztěžuje.

Algoritmus prezentovaný v článku pracoval, při uplatnění nejsilnější meze  $\leq 5$  ms, s úspěšnostmi pro všechny participanty pro E, V, a VOT rovnými  $\leq \acute{u}sp_E = 64,0\%$ ,  $\acute{u}sp_V = 71,2\%$ ,  $\acute{u}sp_{VOT} = 57,2\%$ . Úspěšnost detekce O pro všechny účastníky na relevantnější  $\leq 10$  ms mezi je rovna  $\leq \acute{u}sp_O = 64,6\%$  Tyto úspěšnosti jsou uspokojivé a srovnatelné s ostatními pracemi, ovšem další zlepšení robustnosti jsou nezbytná pro zvýšení



použitelnosti v případech PN. V těchto případech je úspěšnost detekce stále příliš nízká.

V současné době pracují všechny tři detekce nezávisle na sobě, proto je zde prostor pro zlepšení výsledků zpětnou kontrolou. Jedním z přístupů může být výpočet VOT nebo délky znělosti a v případě zjištění hodnot, které nejsou fyziologicky pravděpodobné (předpokladem je, že minimální délku VOT je možné experimentálně odhadnout), opětovné přeměření a detekce v mezích založených na předchozích měřeních.

Zlepšení detekce O by mohlo pomoci plánované zavedení algoritmu pro výpočet ideálního stupně polynomu pro určení tvaru meze. V současné době je stupeň polynomu nastaven na pevnou hodnotu a ne ve všech případech je tato hodnota tou nejlepší.

## Poděkování

Tato práce je podporována z grantů SGS12/185/OHK4/3T/13, GACR 102/12/2230 a NT 12288-5/2011 a z výzkumných záměrů MSM 0021620849 a MSM 6840770012.

## Reference

- [1] Rodríguez-Oroz, M., C., Jahanshahi, M., Krack, P., Litvan, I., Macias, R., Bezard, E., Obeso, J., A.: Initial clinical manifestations of Parkinson's disease: features and pathophysiological mechanisms, *The Lancet Neurology*, 8 (12), 1128 – 1139, 2009.
- [2] Duffy, J., R.: Motor Speech Disorders: Substrates, Differential Diagnosis and Management, 2nd ed. *Mosby, New York, NY, 2005*, pp. 1 – 592.
- [3] Darley, F., L., Aronson, A., E., Brown, J., R.: Differential diagnostics patterns of dysarthria, *J. Speech. Hear. Res.*, 12, 426 – 496, 1969.
- [4] Rusz, J., Čmejla, R., Ruzickova, H., Ruzicka, E.: Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease, *J. Acoust. Soc. Am.*, 129 (1), 350 – 367, 2011.
- [5] Fletcher, S., Time – by – count measurement of dyadochokinetic syllable rate, *J. Speech. Hear. Disord.*, 15, 757 – 762, 1972.
- [6] Rusz, J., Čmejla, R., Růžičková, H., Klempíř, 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.
- [7] Čmejla, R., Sovka, P.: Recursive Bayesian Autoregressive Changepoint Detector for Sequential Signal Segmentation, *EUSipco Proceedings, Wien(2004)*, 245 – 248.
- [8] Hansen, J., H., L., Gray, S., S., Kim, W.: Automatic voice onset time detection for unvoiced stops (/p/,/t/,/k/) with application to accent classification, *Speech Communication*, 52, 777–789, 2010. Prague, 2001.
- [9] Stouten, V., Van Hame, H.: Automatic voice onset time estimation from reassignment spectra, *Speech Communication*, 51, 1194 – 1205, 2009.

# Objektivizace logopedického věku dítěte

Jan JANDA<sup>1</sup>, Roman ČMEJLA<sup>1</sup>

<sup>1</sup> České vysoké učení v Praze, Fakulta elektrotechnická, Katedra teorie obvodů

[jandaja2@fel.cvut.cz](mailto:jandaja2@fel.cvut.cz)

*Abstrakt:* Tato studie je zaměřena na hledání věkově závislých parametrů v dětské řeči. Věková závislost těchto parametrů často pochází ze skutečnosti, že děti mají kratší vokální trakt a drobnější hlasivky. To se projevuje zejména vyšší základní frekvencí hlasivkového tónu a vyššími formanty. Dále pak větší spektrální a suprasegmentální změny jsou přičítány nepřesné dětské výslovnosti.

Práce přináší přehled věkově závislých charakteristik fonace, artikulace, srozumitelnosti a prozodie.

Pro posouzení srozumitelnosti dětské řeči je představen parametr založený na kumulovaném součtu funkce DTW. Jednotlivé parametry jsou z hlediska věkové závislosti porovnány a statisticky zhodnoceny.

V rámci této studie byla na základě dosavadních poznatků a zkušeností nově pořízena databáze dětských promluv. Obsah je cíleně navržen tak, aby se v databázi vyskytovaly především typy promluv, na nichž se dařilo věkově závislé parametry naměřit. Databáze je jednotná pro děti předškolního a školního věku.

Každému dítěti bylo při nahrávacím sezení zadáno několik úloh: prodloužená fonace samohlásek, složitější izolovaná slova, říkánka „En ten týky...“, prodloužené sykavky a další. Databáze obsahuje promluvy od 250 dětí ve věku od 4 do 15 let s přibližným zastoupením 20 dětí na ročník.

Silná věková závislost byla naměřena u základního hlasivkového tónu F0, menší u parametrů frekvenčního a amplitudového kolísání (jiter, schimmer, HNR) a zanedbatelná u formantových frekvencí F1, F2. Neuspokojivá byla rovněž věková závislost spektrálních momentů sykavek.

Při posuzování srozumitelnosti (resp. patlavosti) dětské řeči byla použita vedle analyzované promluvy i promluva referenční stejného obsahu, precizně vyřčená. U analyzované i referenční promluvy byla provedena parametrizace 10 RASTA-PLP koeficienty s délkou segmentu 25ms. V matici vzdáleností jednotlivých segmentů v prostoru dané řečové parametrizace byla nalezena křivka DTW. Kumulativní vzdálenost podél křivky DTW značně korelovala s nesrozumitelností zkoumané promluvy. Tato vzdálenost měla značnou věkovou závislost (u slova „fotbalista“ činila korelace s věkem -0,72). U prozodických parametrů byla největší věková závislost nalezena u rychlosti řeči a podílu pauz v promluvě.

Na základě výběru významných věkově závislých akusticko-fonetických parametrů byl natrénován modelový strom. Pomocí tohoto stromu natrénovaného z 22 řečových parametrů se podařilo strojově odhadnout věk dítěte s průměrnou absolutní chybou 0,92 roku. Míra korelace takto určeného logopedického věku s věkem skutečným dosáhla hodnoty 0,92.

# Objektivizace logopedického věku dítěte

Jan Janda

školitel: Doc. Ing. Roman Čmejla, CSc

1

## Obsah

- Úvod
- Vývoj vokálního traktu dítěte
- Databáze dětských promluv
- Akusticko-fonetické analýzy
  - Samohlásky
  - Sibilanty
  - Srozumitelnost slov
  - Prozodické parametry
- Strojové určení věku dítěte
- Další cíle

2

## Definice

- Chronologický věk  
*je určen jednoznačně datem narození.*
- Logopedický věk  
*je věk, na který usuzujeme z akusticko-fonetické stránky promluvy člověka.*
- Percepční věk  
*je věk, který odpovídá výsledku poslechového testu*

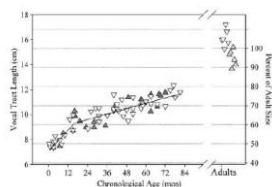
3

## Vývoj řeči

- Od narození do dospělosti se délka vokálního traktu prodlouží přibližně na dvojnásobek
- Výrazně se mění geometrické proporce jednotlivých tkání relativně k délce vok. traktu (zakřivení v oblasti nosohltanu, sestoupení hrtanu a příklopky hrtanové, pokles zadní části jazyka)
- Různé struktury dorůstají dospělých rozměrů od 7 do 18 let věku

4

## Vývoj řeči



- Očekávaný pokles F0 a formantových frekvencí
- Absence pohlavního dimorfismu zkoumaných charakteristik do 6 let
- Zvětšující se prostor pro pohyb jazyka

5

## Řečová databáze

- Spolupráce s Foniatrikou klinikou 1.LF UK
- Soustředí se na nadějně řečové charakteristiky
- Použití i pro zkoumání vývojových vad řeči
- Spontánní realizace slov

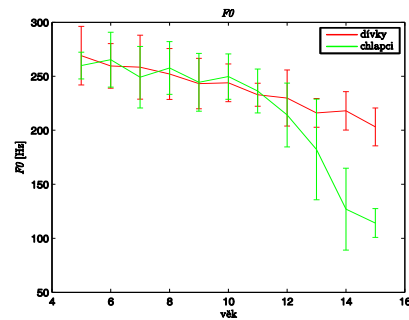
6

## Řečová databáze

- Prodloužené fonace samohlásek
- Syčení s,š
- Sekvence pa, ta, ka; ba, da, ga
- Složitější slova (babička, čokoláda, popelnice, sluníčko, košile, silnice...)  
- spontánní popis obrázků
- Popis činností na obrázcích
- Říkáňka
- **Aktuálně 250 dětí; 4-15 let**

7

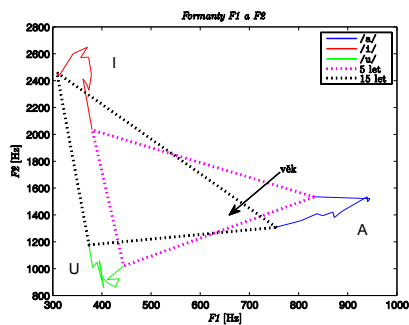
## Fonetické analýzy – věková závislost F0



- Vokál /i/ - prodloužená fonace cca 5s

8

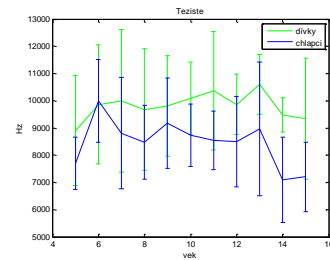
## Fonetické analýzy – věková závislost F1 a F2



- Vokály /a,i,u/ - prodloužená fonace cca 5s

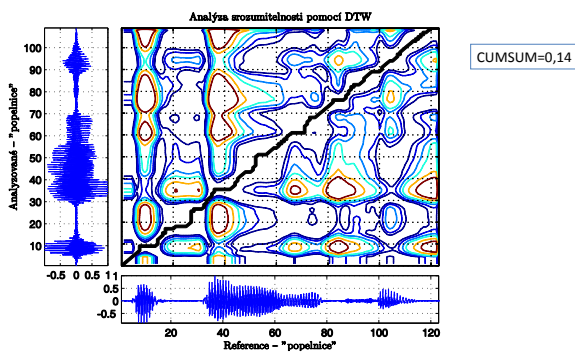
9

## Fonetická analýza – sibilanty spektrální těžiště /s/



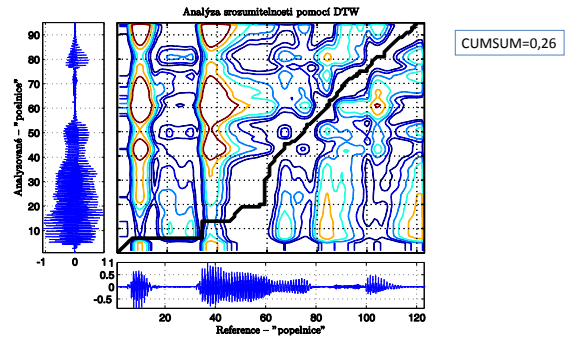
10

## Srozumitelnost – zřetelná promluva



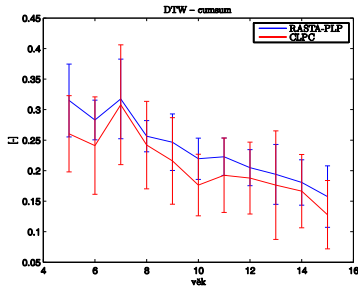
11

## Srozumitelnost – patlaná promluva



12

## Srozumitelnost řeči DTW – „fotbalista“

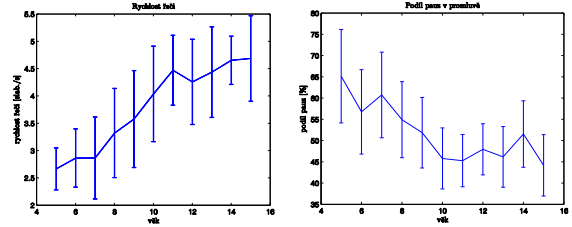


10 koeficientů, délka segmentu 25 ms

13

## Prozodie: Rychlost řeči, Podíl pauz

- Říkanka „ententýky...“ (z paměti)



14

## Věkově závislé charakteristiky - přehled

charakteristika	r		r
DTW (fotbalista)	-0,72	schimmer(apq5)	0,40
DTW (popelnice)	-0,71	jitter(rap)	0,38
DTW (velryba)	-0,70	schimmer(apq11)	0,38
DTW (čokoláda)	-0,70	F1	-0,30
F0	-0,66	sp. zešikmení /ss/	0,21
DTW (hamburger)	-0,66	F2	0,20
rychlost řeči	0,66	std /ss/	0,19
DTW (Rákosníček)	-0,53	sm. odchylka F0	0,17
HNR	-0,53	mezikvart. rozp. F0	0,08
podíl pauz	-0,51	sp. špičatost /s/	-0,08
schimmer(local,abs)	0,45	std /s/	-0,03
schimmer(local)	0,45	sp. těžiště /ss/	-0,03
schimmer(apq3)	0,45	sp. těžiště /s/	0,03
jitter(ppq5)	0,44	sp. zešikmení /s/	0,03
jitter(local,abs)	0,42	sp. špičatost /ss/	0,03
jitter(local)	0,40		

15

## Strojové určení věku dítěte

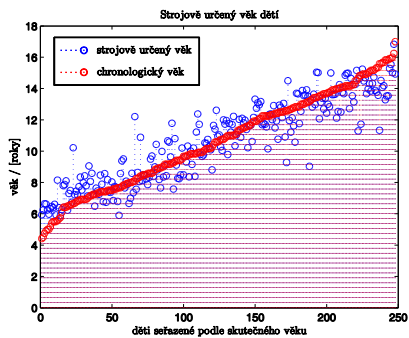
- Věk je považován za spojitou veličinu
- Natrénován regresní strom M5
- Trénování metodou 10 foldové krosvalidace

## Výsledky

- Korelace se skutečným věkem 0,92
- Průměrná absolutní chyba 0,92 roku
- Směrodatná odchylka 1,19 roku

16

## Regresní strom M5 - výsledky



17

## Další cíle

- Automatická detekce řečových charakteristik
- Realizace systému v reálném čase
- Porovnání s výsledkem percepčního testu
- Využití při analýzách řeči dětí s vývojovou disfázií

## Reference

- Janda, J.: Age Dependence of Children's Speech Parameters. Acta Polytechnica. 2009, vol. 49, no. 2-3, p. 40-43. ISSN 1210-2709
- Zlatník, P. - Čmejla, R.: Disordered Speech Assessment Using Different Speech Parameterizations. In 19th International Congress on Acoustics Madrid, Spain
- Y. Wang, I. H. Witten: Induction of model trees for predicting continuous classes. In: Poster papers of the 9th European Conference on Machine Learning, 1997
- Voperian K. H., Kent R. D. et al.: Development of vocal tract length during early childhood: A magnetic resonance imaging study. Journal of the Acoustical Society of America 117, January 2005

18

# Objektivní hodnocení závažnosti vývojové dysfázie dle mluveného projevu dětí

Martina NEJEPSOVÁ, Roman ČMEJLA<sup>1</sup>, Jan JANDA<sup>1</sup>, Jan VOKŘÁL<sup>2</sup>

<sup>1</sup> České vysoké učení v Praze, Fakulta elektrotechnická, Katedra teorie obvodů

<sup>2</sup> Foniatrická klinika 1. LF UK a VFN

[nejepmar@fel.cvut.cz](mailto:nejepmar@fel.cvut.cz)

**Abstrakt:** Cílem výzkumu, jenž je popsán v tomto článku je nalezení akustických parametrů vhodných ke stanovení závažnosti poruchy u dětských pacientů s vývojovou dysfázií. Na základě subjektivních poslechových testů provedených odborníky i laiky byly pacienti rozděleni do tří skupin dle závažnosti vývojové dysfázie (mírná, střední a závažná forma poruchy). Po provedených analýzách hlasu a řeči totožných promluv zdravých dětí i dětí s vývojovou dysfázií byly vybrány takové parametry vhodné pro kategorizaci mluvčích do jednotlivých skupin. Objektivní hodnocení promluv pacientů s vývojovou dysfázií by mělo usnadnit diagnostiku poruchy, dále zhodnotit stávající úroveň rozvoje řeči, artikulaci a klasifikaci závažnosti poruchy těchto pacientů prováděnou na klinických pracovištích.

## 1. Úvod

Řeč - jeden z hlavních dorozumivacích prostředků člověka se rozvíjí již od raného věku a mění se v průběhu dospívání. Často však bývá ovlivněna vrozenou nebo získanou vadou. V praxi se nejčastěji setkáváme s opožděným či nerovnoměrným vývojem řeči, vývojovou dysfázií, vývojovou dysartrií, dyslálií (patlavostí), koktavostí (zadržáváním) či různými poruchami hlasu. V důsledku výskytu lehčích forem těchto vad dochází k diagnostice až v pozdějším věku a tudíž i ke zpoždění se zahájením příslušné terapie. Včasným rozpoznáním vady hlasu či řeči je možné pomocí terapeutické péče zmírnit či úplně odstranit její projevy. Rozpoznání symptomů a určení původu poruchy vede ke stanovení konkrétní diagnózy a závažnosti dané poruchy. To je v některých případech obtížné i pro zkušené specialisty. Nejednotnost diagnostiky a převážně subjektivní hodnocení může zapříčinit odlišné určení nejen závažnosti, ale i typu poruchy.

Nutnost objektivních klasifikací, které by napomohli při subjektivních hodnoceních prováděných specialisty, je dána současným trendem sjednocování metodiky hodnocení a klasifikace dle předem definovaných pravidel. Na základě doposud získaných zkušeností probíhají nezávislé výzkumy v oblasti objektivních měření specifických promluv od pacientů s vadami hlasu a řeči.

## 2. Vývojová dysfázie

Vývojová dysfázie, neboli specificky narušený vývoj řeči je porucha zpracování řečového signálu [1]. Projevuje se úplnou či částečnou ztrátou schopností naučit se verbálně komunikovat, ačkoliv podmínky pro rozvoj řeči jsou přiměřené. Původ a období vzniku

nejdou zatím přesně známé (předpokládá se však, že vzniká v prenatálním období, při porodu či následně po něm).

Charakteristickými rysy vývojové dysfázie patrných z mluveného projevu jsou obtížně srozumitelný projev, opožděný vývoj řeči, slovní zásoba neodpovídající věku mluvčího, komolení slov či přehazování slovosledu a vynechávání některých slovních druhů. Projevuje se také výraznými rozdíly mezi verbálními a neverbálními schopnostmi, narušením krátkodobé verbální paměti, sluchovým vnímáním či gramatickými chybami. Dalšími symptomy jsou kresby postav, motorika a koordinace pohybů, špatná orientace v čase a prostoru a nesprávné stanovení vztahů mezi členy rodiny.

Diagnóza vývojové dysfázie je ve většině případů stanovena na základě hodnocení několika specialistů. Během foniatrického vyšetření jsou sledovány promluvy a porozumění pacienta, provádí se vyšetření všech složek řeči, včetně testování motorické aktivity řečových orgánů. Tónová audiometrie, tympanometrie a elektrofyziologické testování pomocí sluchově evokovaných potenciálů se používají ke kontrole sluchu. Mezi důležité ukazatele patří i výsledky dichotických testů, testu fonemického sluchu, Heidelberský test či vyšetření indexu vnitřní informace řeči stanoveného dětským percepčním testem. Klinický logoped stanovuje přehled schopností dítěte a míru opoždění vývoje řeči. Neurologická vyšetření poskytují pro diagnózu EEG záznam nebo CT. Nedílnou součástí je také psychologické vyšetření.

### 3. Databáze promluv

Pro účely hodnocení akustických parametrů, jejich závislostí a sledování trendu závažnosti bylo třeba pořídit totožné promluvy od pacientů s vývojovou dysfázií a od zdravých mluvčích. Ve spolupráci s Foniatrickou klinikou 1. LF UK a VFN byl sestaven seznam logopedických jevů vhodných pro následné analýzy (obsah databáze je v Tab. 1).

Promluvy zdravých dětí pořizoval Ing. J. Janda v mateřských a základních školách v Praze pro účely výzkumu věkově závislých akustických parametrů v dětské řeči [2]. Promluvy dětí s vývojovou dysfázií byly pořizovány autorem na Foniatrické klinice 1. LF UK a VFN v rámci výzkumu akustických parametrů pro klasifikaci vývojové dysfázie [3].

**Tab. 1: Obsah databáze**

izolované vokály	/A/, /E/, /I/, /O/, /U/	výdrž reprodukce – spontánní promluva, opakovaná promluva dle vzoru (u předškolních dětí)
izolovaná slova	/máma/, /babička/, /čokoláda/, /sluníčko/, /popelnice/, /košile/, /silnice/, /Rákosníček/, /hamburger/, /velryba/, /ucho/, /ježek/, /ředkvička/, /fotbalista/	označování obrázků – spontánní promluva, opakovaná promluva dle vzoru (dle znalosti slov a závažnosti poruchy)
říkanka	/En ten týky, dva špalíky, čert vyletěl z elektriky. Bez klobouku bos, natloukl si nos./	spontánní promluva, promluva s náповědou (u dětí se závažnější formou poruchy)
rychlé opakování slabik	/PA-TA-KA/ /BA-DA-GA/	spontánní promluva (pouze u školních dětí)
izolované sykavky	/S/, /Š/	výdrž reprodukce – spontánní promluva, opakovaná promluva dle vzoru (u předškolních dětí)
popis, spontánní promluva	ranní vstávání a následující činnosti před cestou do školky/školy	popis rutinní činnosti pomocí sekvence obrázků

Všechny promluvy byly pořizovány pomocí přenosného digitálního rekordéru EDIROL R-09HR (ROLAND) se směrovým kondenzátorovým mikrofonem OPUS 55.09 Mk II SC (BEYERDYNAMIC) v běžných uzavřených prostorách (nejen z důvodu nedostupnosti speciálních odhlučněných prostor při nahrávání, ale i z důvodu následného užití klasifikátoru v běžných ambulancích). Mluvčím byl umístěn hlavový mikrofon ve vzdálenosti 2cm od úst. U zdravých mluvčích byla nastavena konstantní úroveň vstupu. U pacientů s vývojovou dysfázií tento postup nebyl možný, poněvadž rozsah intenzity projevu pacientů byl velmi rozdílný a nebylo by následně možné pořízené zvukové záznamy analyzovat. Před samotným nahráváním pacientů s vývojovou dysfázií byl veden krátký rozhovor na běžná témata, ze kterého byl patrný aktuální stav dítěte.

V současné době obsahuje databáze 178 nahrávek od zdravých dětí a 195 nahrávek od dětí s vývojovou dysfázií ve věku od 4 do 12 let.

#### **4. Subjektivní poslechové testy**

Pro získání referenčních hodnocení závažnosti poruchy vývojové dysfázie u dětí byly připraveny subjektivní poslechové testy.

První subjektivní poslechový test byl zaměřen na testování hodnotitelů – schopnost určit skutečný věk mluvčího na základě vyslechnuté promluvy. Test byl předložen k hodnocení laikům i specialistům a porovnáván s automatickým klasifikátorem Ing. J. Jandy (viz. [4]). Sledována byla i opakovatelnost testu. Výsledky tohoto subjektivního testu jsou popsány v [5].

Pro hodnocení promluv pacientů s vývojovou dysfázií byly připraveny další dva subjektivní poslechové testy: 1. hodnocení logopedického věku mluvčího na základě vyslechnuté promluvy, 2. hodnocení závažnosti poruchy na základě vyslechnuté promluvy a znalosti biologického věku mluvčího. Podrobný popis testů a jejich realizace je uvedena v [6]. Stupeň závažnosti vývojové dysfázie hodnocených odborníky z foniatrické kliniky v subjektivních poslechových testech byl užit jako referenční hodnota pro kategorizaci promluv do tří skupin (1 – mírná, 2 – střední a 3 – závažná porucha). Určený stupeň závažnosti poruchy významně koreluje s rozdílem určeného logopedického věku a biologického věku mluvčího.

#### **5. Objektivní hodnocení promluv**

##### **5.1. Výběr promluv pro hodnocení závažnosti vývojové dysfázie**

Pro objektivní hodnocení promluv s vývojovou dysfázií bylo vybráno 10 promluv zdravých dětí a 30 promluv dětí s vývojovou dysfázií (10 promluv v každém stupni závažnosti) ve věku 5-7 let. Tyto promluvy byly z databáze vybírány na základě subjektivních poslechových testů. U zdravých dětí byly voleny promluvy s největší shodou určení jednotlivými hodnotiteli a nejmenší odchylkou od skutečného věku a určeného věku v opakovaném hodnocení. U pacientů s vývojovou dysfázií byly vybírány promluvy s největší shodou určení závažnosti poruchy jednotlivými hodnotiteli. Dělení mluvčích dle pohlaví není nutné neboť odlišnost akustických charakteristik mezi mužským a ženským hlasem se projevuje až v průběhu dospívání.

##### **5.2. Realizované analýzy**

###### *5.2.1. Analýza vokálů*

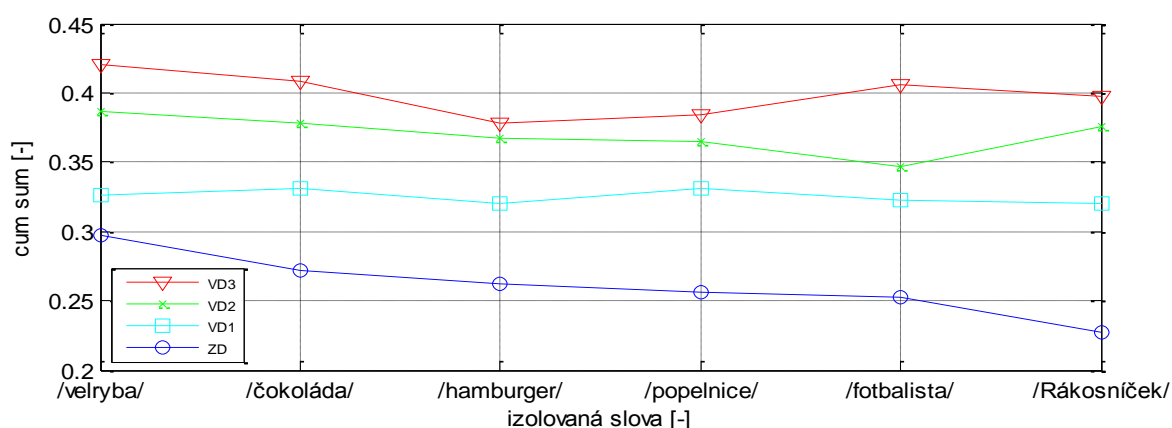
Vokální analýza byla prováděna na izolovaných vokálech a na vokálech extrahovaných z izolovaných slov v programu PRAAT [10] a ověřována v programovém WAVESURFER



[11]. Pro každý z vokálů byly sledovány trendy základního hlasivkového tónu (F0), prvního (F1) a druhého (F2) formantu, délka izolovaného vokálu a poměr délek izolovaného a extrahovaného vokálu.

### 5.2.2. Analýza izolovaných slov

Při analýze slov byla sledována srozumitelnost promluv pomocí algoritmu DTW (Dynamic Time Warping) v MATLABu (podrobně popsána v [9]). Metoda je založena na porovnání identických promluv. Nejprve byla testována referenční promluva zdravého mluvčího. Jako nejvhodnější byla zvolena promluva šestnáctiletého chlapce. Dalším bodem byla volba užití fázového vokodéru k normalizaci délek promluvy. U slov s menším počtem slabik není rozdíl mezi užitím a neužitím vokodéru výrazný. Avšak u víceslabičných slov (3 a více slabik) je užití vokodéru nutné, neboť pacienti s vývojovou dysfázií některé slabiky vynechávají či do slov přidávají. Poslední volbou bylo i správné nastavení parametrů pro algoritmus DTW (délka a typ okna, koeficienty, ...). Jedna z realizací je na Obr. 1..



**Obr. 1: Kumulovaná vzdálenost pro izolovaná slova (spektrální PLP koeficienty)**

### 5.2.3. Analýza říkanky

Analýza pomocí dětem známé čtyřveršové říkanky slouží ke sledování tempa řeči během promluvy, stupni souladu s veršem a rýmem. Jednotlivé kategorie byly porovnávány dle délky říkanky – doby trvání reprodukce říkanky, obsahu pauz v promluvě a poměru řeč/pauza. Kompletní reprodukce říkanky v podání dětí s vývojovou dysfázií trvá déle než u zdravých mluvčích. V praxi se však setkáváme s neúplnou reprodukcí (vynechané rýmy či slova). Deficit krátkodobé paměti je zřejmě příčinou, proč děti s vývojovou dysfázií si hůře zapamatují a hůře recitují říkanky.

### 5.2.4. Analýza rychlého opakování slabik

Tato analýza nebyla doposud prováděna, neboť počet pořízených promluv ve věkové kategorii 5-7 let je nedostatečný. Pro pacienty s vývojovou dysfázií je obtížné tento logopedický jev realizovat.

### 5.2.5. Analýza sykavek

Analýza sykavek byla prováděna na izolovaných sykavkách i na sykavkách extrahovaných z izolovaných slov v MATLABu. Spektrální parametry – spektrální těžiště, směrodatná odchylka, špičatost a zešíkmení byly vypočteny podle matematických vzorců pro výpočet spektrálních momentů. Dále je sledována délka obou izolovaných sykavek. Pacienti

s vývojovou dysfázií mají problém s vyslovením izolovaných sykavek, ačkoliv v běžném řečovém projevu je vyslovují správně. I z tohoto důvodu jsou sledovány poměry parametrů izolovaných sykavek a extrahovaných sykavek ze slov.

### 5.2.6. Analýza popisu, spontánní promluvy

Spontánní popis konkrétní situace je prováděn pomocí sekvence 9 obrázků znázorňující danou činnost. V rámci této analýzy byla posuzována míra rozmanitosti projevu. Pacienti se závažnou formou vývojové dysfázie jsou často schopni provést popis jen s použitím sloves znázorňující činnosti. Jednoduché věty, které obsahují podmět a přísudek je možné slyšet od pacientů se středně závažnou vývojovou dysfázií. Pouze zdravé děti a velmi málo pacientů s vývojovou dysfázií jsou schopni popsat sekvenci obrázků celou větou. Z tohoto důvodu jsou sledovány charakteristiky počtu použitých slov a slovních druhů (v současné době hodnoceno manuálně), počtu pauz v projevu, poměr řeč/pauza a v neposlední řadě délka samotného spontánního popisu.

## 6. Závěr

Na základě provedených analýz byly jednotlivé parametry rozděleny do několika skupin dle vhodnosti jejich užití pro kategorizaci promluv dle závažnosti.

Mezi nevhodné parametry patří F0, F1 a F2, jejichž charakteristiky jsou totožné pro zdravé děti i děti s vývojovou dysfázií. Dle [2] jsou věkově závislé.

Vhodné parametry pro určování závažnosti poruchy jsou uvedeny v Tab. 2.

**Tab. 2: Přehled parametrů vhodných pro hodnocení závažnosti vývojové dysfázie**

Délka izolovaných vokálů a sykavek
Poměr délek izolovaného a extrahovaného vokálu a sykavek
Kumulativní součet DTW (spektrální PLP koeficienty, RASTA koeficienty) u izolovaných slov
Spektrální těžiště, směrodatná odchylka, špičatost a zešikmení u sykavek
Délka dětské říkanky a spontánního projevu
Míra artikulace v dětské říkance a spontánním projevu
Poměr řeč/pauza, řeč/celá promluva, pauza/celá promluva v dětské říkance a spontánním projevu
Počet slov v říkance a spontánním projevu
Počet pauz a jejich celková délka v říkance a spontánním projevu

Významné jsou i parametry – počet slov použitých ve správném pořadí v říkance, dodržování rýmů a veršů, počet užitých slovních druhů v spontánním popisu, poměr spontánních/opakovaných/nevyslovených izolovaných slov, subjektivní hodnocení celého záznamu promluvy nahravatelem, které jsou v současné době hodnoceny pouze manuálně.

Cílem výzkumu by měl být softwarový klasifikátor, který by měl být schopen na základě analýzy vhodných parametrů rozpoznat zdravé děti, děti s vývojovou dysfázií a jejich závažnost tohoto onemocnění. Dále bude rozšiřována databáze pacientů ve věkových skupin 3-4 a 8-12 let. Samozřejmostí je i hledání nových parametrů pro hodnocení vývojové dysfázie a realizace subjektivních poslechových testů.

## Poděkování

Tento výzkum je prováděn pod odborným vedením Doc. Ing. Romana Čmejly, CSc., ve spolupráci s Foniatričskou klinikou 1. LF UK a VFN a je podporován z grantu GAČR 102/12/2230 “ Analýza hlasu a řeči pacientů s onemocněními centrální nervové soustavy”.

Poděkování patří Ing. Janu Jandovi za poskytnutí databáze promluv zdravých dětí a dále všem hodnotitelům, kteří se účastnili subjektivních poslechových testů pro hodnocení vývojové dysfázie.

## Reference

- [1] Škodová, E., Jedlička, I. a kolektiv: *Klinická logopedie – 2. aktualizované vydání*, PORTAL, Praha, 2007.
- [2] Janda J.: *Studie věkově závislých akustických parametrů v dětské řeči*. Studie k odborné rozpravě, ČVUT FEL, Praha, 2010.
- [3] Nejeřmová, M.: *Analýza a možnosti objektivního hodnocení promluv dětských pacientů s vývojovou dysfázií*. Studie k odborné rozpravě, ČVUT FEL, Praha, 2011.
- [4] Janda, J.: Odhad logopedického věku z řeči dítěte. *Analýza a zpracování řečových a biologických signálů - sborník prací 2010*, ČVUT FEL, Praha, 2010.
- [5] Nejeřmová, M.: Analýza subjektivního hodnocení dětského věku dle promluv. *Analýza a zpracování řečových a biologických signálů - sborník prací 2010*, ČVUT FEL, Praha, 2010.
- [6] Nejeřmová, M.: Subjective listening tests for the evaluation of developmental dysphasia. *16th International Student Conference on Electrical Engineering POSTER 2012 – Conference proceedings (on CD-ROM)*, ČVUT FEL, Praha, 2012.
- [7] Melka, A.: *Základy experimentální akustiky*, AMU, Praha 2005.
- [8] Nejeřmová, M., Janda, J., Čmejla, R., Vokřál, J.: Acoustic Analysis of Utterances from Children with Developmental Dysphasia. *Akustické listy 2010*, č. 4, ČVUT FEL, Praha, 2010.
- [9] Zlatník, P., Čmejla, R.: Disordered Speech Assessment Using Different Speech Parameterizations, *19th International Congress on Acoustics*, Sociedad Española de Acústica, Madrid, 2007.
- [10] Boersma, P., Weenink, D.: *Praat: doing phonetics by computer* [Computer program], version 5.1.05, <http://www.praat.org/>
- [11] Sjölander, K., Beskow, J.: *WaveSurfer* [Computer program], version 1.6.3, <http://www.speech.kth.se/wavesurfer/>
- [12] Schötz, S: *Perception, Analysis and Synthesis of Speaker Age*. Doctoral Thesis, Lund University, 2006.
- [13] Vick, J., Moore, C.A., Campbell, T., Shriberg, L., Green, J., Truemper, K.: Multivariate Classification of Children with Speech Delay of Unknown Origin. *The Annual Convention of the American Speech-Language-Hearing Association*, New Orleans, 2009.
- [14] Bunnell, H.T., Schanen, N.C., Vallino, L.D., Morlet, T.G., Polikoff, J.B., Driscoll, J.D., Mantell, J.T.: Speech Perception in Children with Speech Sound Disorder. *8th Annual Conference of the International Speech Communication Association*, Belgium – Antwerp, 2007.

# Hodnocení koktavosti

Tomáš LUSTYK<sup>1</sup>, Petr BERGL<sup>1</sup>, Roman ČMEJLA<sup>1</sup>

<sup>1</sup> České vysoké učení v Praze, Fakulta elektrotechnická, Katedra teorie obvodů

[lustytom@fel.cvut.cz](mailto:lustytom@fel.cvut.cz)

**Abstrakt:** Článek se zabývá hodnocením koktavosti pomocí automatických metod. Experimenty jsou provedeny na čtených promluvách. Parametry zpracovávají vždy celou řečovou promluvu. Výstupy parametrů jsou srovnány s kontrolními daty vytvořenými pomocí hodnotícího systému LBDL (the Lidcombe Behavioral Data Language of Stuttering).

## 1. Úvod

Koktavost je jednou z poruch plynulosti řeči, má mnoho podob a její příčiny nejsou dosud uspokojivě vysvětleny. Příznaky koktavosti mohou být rozlišovány na *vnitřní* a *vnější* [1]. Mezi vnější patří hlavně příznaky pozorované v řeči mluvčí, ale také poruchy dýchání, zášklby končetin, těla, a další. Příznaky pozorované na řeči jsou repetice (opakování hlásek, slabik částí nebo celých slov), prolongace (nepřirozené prodloužení hlásky), tonoklony (kombinace dvou předchozích), časté pauzy a neřečové události.

Mluvčí si svoji poruchu uvědomují a tato skutečnost má vliv na psychiku mluvčího (vnitřní příznaky). Vnitřní příznaky se vyvíjí v průběhu nemoci a mohou vést až ke strachu komunikovat řečí (logofobii). Z uvedených faktů vyplývá, že koktavost je vážný lékařský a sociální problém. Správné posouzení tíže poruchy a volba následujícího léčebného postupu jsou obtížné úkoly.

Stanovení diagnózy a léčbu provádí specialisté na logopedii. Postup při stanovení tíže poruchy plynulosti řeči je zdlouhavý a namáhavý proces. Metoda, která by umožnila automaticky a objektivně určit tíži poruchy by byla přínosná. Umožnila by zejména [1, 2]: 1) stanovení tíže poruchy; 2) hodnocení výsledků léčby; 3) porovnání léčebných postupů.

Pro stanovení tíže poruchy mohou automatické algoritmy hledat v nahrávkách jednotlivé symptomy zmíněné v úvodním odstavci. Článek [3] se zaměřuje na hledání repeticí a prolongací v řečovém signálu. Jednoduchý VAD (Voice Activity Detector) a časové prahy jsou použity pro detekci repetice. Formantové frekvence jsou využívány pro detekci prolongací. Více komplexní metody zahrnující HMM (Hidden Markov Models) jsou použity pro rozpoznávání bloků s repeticemi a prolongací u frikatvních hlásek v článku [4] nebo pro hledání repetice a prolongací v článku [5].

Parametry nemusí hledat pouze symptomy řečové poruchy, ale mohou zpracovávat řečový signál jako celek. Skupina více než 20 parametrů v časové a frekvenční oblasti byla představena v práci [6]. V časové oblasti jsou to například: průměrná délka ticha, poměr celkové délky ticha a řeči a parametry zkoumající energii signálu. Ve frekvenční oblasti: směrodatná odchylka vzdáleností spektrálních změn BACD (Bayesian Change-point Detector) nebo počet spektrálních změn.

Článek popisuje dosavadní výsledky experimentů prováděných na čtených promluvách. V krátkosti popisuje jednotlivé parametry používané pro stanovení tíže poruch. Výsledky a vhodnost parametrů pro automatické hodnocení jsou uvedeny v závěru článku.

## 2. Databáze signálů a hodnocení

Databáze signálů byla vytvořena v posledních několika letech na 1. Foniatrické klinice Univerzity Karlovy a Všeobecné fakultní nemocnice v Praze. Databáze obsahuje nahrávky přibližně 160 mluvčích rozdílného věku a stupně koktavosti. Promluvy jsou *čtené* a *spontánní*, nahrávané se zpožděnou akustickou vazbou (DAF – Delayed Auditory Feedback). Rozsah DAF se pohybuje od 10 do 110 ms. Třicet promluv jsou nahrávky zdravých mluvčích bez poruchy plynulosti řeči (kontrolní skupina). Vzorkovací frekvence při nahrávání byla 44 kHz, signály byly pro následující zpracování převzorkovány na 16 kHz.

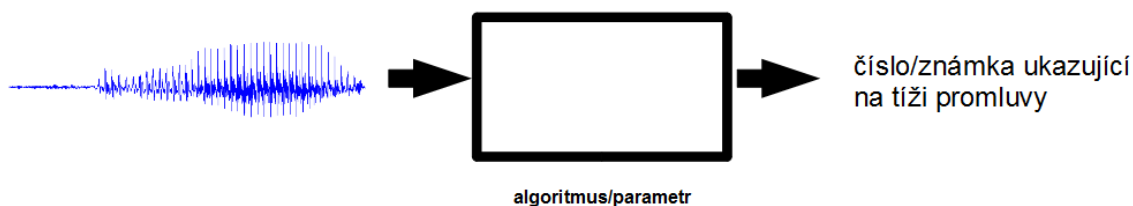
Čtená část databáze, použitá v popsaných experimentech, obsahuje 121 nahrávek. Čtený text se skládá ze 70 slov. Délka promluv kolísá mezi 40 až 60 s. Spontánní promluvy jsou vytvářeny jako popis obrázků.

Pro ověření vhodnosti parametrů pro stanovení tíže poruchy jsou nutná kontrolní data. Kontrolní data byla vytvořena pomocí LBDL [7]. Tento systém je velmi rozšířený, je snadné ho používat. Článek [8] např. popisuje hodnocení symptomů koktavosti v řeči pacientů s Parkinsonovou chorobou, kde LBDL byl použit pro tvorbu referenčních dat. LBDL uvažuje sedm charakteristik symptomů koktavosti: *repetice slabik (1)*; *repetice hlásek (2)*; *opakování víceslabičných útvarů (3)*; *fixovaný stav se slyšitelným zvukem - prolongace (4)*; *fixovaný stav bez slyšitelného zvuku – pauza, přerušeni (5)*; *nadbytečné řečové projevy (6)*; *nadbytečné neřečové projevy (7)*. Všechny charakteristiky kromě 7 jsou detekovatelné v řečovém signálu, proto charakteristika 7 není použita v tomto experimentu. Navíc od základních charakteristik jsou použity ještě *celková známka, opakované (1 + 2 + 3)* a *fixované (4 + 5)*. Hodnocení vytvořené pomocí LBDL bylo použito jako kontrolní data pro všechny popsané experimenty.

Všechny nahrávky byly ohodnoceny jedním hodnotitelem. 30 promluv bylo ohodnoceno dvakrát pro získání shoda hodnocení (konzistentnosti) pro hodnotitele. Stejných třicet promluv bylo ohodnoceno druhým hodnotitelem pro získání shody u dvou hodnotitelů. Výsledky srovnání souhlasu jednoho hodnotitele a souhlasu dvou různých hodnotitelů jsou uvedeny v tabulce 1.

**Tab. 1: Výsledky souladu hodnocení pro jednoho hodnotitele a dvou různých hodnotitelů, pro všechny charakteristiky hodnotícího systému LBDL korelační koeficient  $r$  a hodnota  $p$  ( $p < 0,05$  v případě, že jsou korelace  $r$  významné)**

charakteristika	shoda pro jednoho hodnotitel		shoda pro dva hodnotitele	
	$r$	$p$	$r$	$p$
<i>1</i>	0,94	0	0,91	0
<i>2</i>	0,97	0	0,94	0
<i>3</i>	0,89	0	0,76	0
<i>4</i>	0,96	$0,02 \cdot 10^{-9}$	0,89	0
<i>5</i>	0,96	0	0,83	0
<i>6</i>	0,87	0	0,32	0,08
<i>opakované</i>	0,99	$0,3 \cdot 10^{-9}$	0,97	0
<i>fixované</i>	0,95	0	0,84	0
<i>celková známka</i>	0,98	0	0,94	0



**Obr. 1: Schématické znázornění zpracování řečového signálu.**

### 3. Popis parametrů pro měření tíže poruchy

Tíže poruchy se určuje na základě analýzy řečového signálu. Zpracovávána je celá promluva, do algoritmů není vložen žádný ruční zásah. Schématicky může být celé zpracování zobrazeno jako na obrázku 1, na vstup algoritmu (černé skříňky) přivedeme řečový signál. Ten je zpracován a na výstupu obdržíme číslo odpovídající tíži poruchy.

V nyní budou ve stručnosti popsány vybrané parametry navržené pro odhad tíže poruchy, podrobněji je více parametrů popsáno v práci [6]. Těmito parametry jsou:

- 1) pravidelnost energie (snaží se postihnout nepravidelnost výdeje energie při promluvách koktavých);
- 2) průměrná délka ticha (popisující velké množství ticha v promluvách koktavých);
- 3) počet repetice (inspirován článkem [3], hledá opakující se úseky v řečovém signálu);
- 4) počet maxim Bayesova detektoru (PMBD, využívá detektoru spektrálních změn BACD, popsaném v [9] vychází z předpokladu, že mluvčí s poruchou plynulosti řeči v promluvě za určitou dobu vysloví méně slov/hlásek než mluvčí z kontrolní skupiny);
- 5) rozestupy při překročení prahu (RPP, využívá detektor spektrálních změn, snaží se postihovat prolongace a pauzy v řeči koktavých);
- 6) směrodatná odchylka z 11 po sobě jdoucích intervalů (SOI11, opět využívá BACD, snaží se postihovat celkovou plynulost/dojem z promluvy);
- 7) počet spektrálních změn v krátkém intervalu (využívá Bayesova detektoru, cílem návrhu parametru bylo postihnout častá přerušení a pauzy v řeči)
- 8) počet rozestupů delších než práh (využívá VAD i BACD, snaží se zaznamenat prolongace v řečovém signálu)
- 9) rychlost řeči (inspirován článkem [10], využívá VAD i BACD, snaží se popisovat pauzy a přerušení v promluvě).

### 4. Výsledky

V tabulce 2. jsou shrnuty výsledky pro nejlepší nastavení parametrů. Sloupce určují pořadí parametru, řádky jsou jednotlivé charakteristiky LBDL. V tabulce je zaznamenán název parametru a velikost korelačního koeficientu. Pokud je v tabulce vynecháno místo, výsledky dalších parametrů nebyly tak výrazné a nebyly proto zaznamenány.

Celkově nejvyšší shoda 0,94 je pro charakteristiku 2 (*opakování hlásek*) pro parametr počet repetice. Velmi dobré výsledky také můžeme pozorovat pro charakteristiku 5 (*fixovaný stav bez slyšitelného zvuku – pauza, přerušení*), *opakované, fixované a celková známka*. Horší

výsledky pozorujeme pro charakteristiky 1 (*opakování slabik*) a 3 (*opakování víceslabičných výrazů*). Tyto projevy koktavosti jsou automatickými metodami velmi těžko zachytitelné.

Velmi dobrým výsledkem je, že parametr detekující repetice má nejvyšší shodu právě s charakteristikou 2 (*opakování hlásek*).

Pro charakteristiku 5 (*fixovaný stav bez slyšitelného zvuku – pauza, přerušeni*) dosáhly nejlepších výsledků právě parametry, které byly navrženy za účelem popsat co nejlépe pauzy a přerušeni v neplynulé řeči, rychlost (-0,86), průměrné ticho (0,85) a počet změn v krátkém intervalu (-0,75).

Úspěchem je také vysoký korelační koeficient (0,72) pro parametr popisující prolongace s charakteristikou 5.

Velmi dobré výsledky pro charakteristiku 6 mohou být způsobeny podobností některých neřečových událostí a repetice nebo také velmi podobným počtem těchto projevů s počtem repetice v promluvách.

Parametry, které mají dobré výsledky pro jednu charakteristiku, která je součástí dalšího celku *opakované* nebo *fixované*, mají právě pro tyto kombinované charakteristiky také uspokojivé výsledky. Těmito příklady jsou parametry počet repetice a rychlost.

Velkým plusem je také velmi dobrý výsledek pro *celkovou známku* velkého množství parametrů. Nejvyšší korelační koeficient 0,85 dosáhl parametr pravidelnost energie, následují průměrné ticho, RPP a SOI11.

**Tab. 2: Pořadí nejlepších parametrů pro jednotlivé charakteristiky LBDL.**

charakteristika	pořadí			
	1	2	3	4
<b>1</b>	PMBD (-0,5257)	počet rozestupů (0,5127)	-	-
<b>2</b>	počet repetice (0,9491)	prům. Ticho (0,8348)	SOI11 (0,7917)	-
<b>3</b>	RPP (0,5933)	PMBD (-0,5804)	-	-
<b>4</b>	počet rozestupů (0,7244)	-	-	-
<b>5</b>	rychlost (-0,8611)	prům. ticho (0,8569)	počet změn za dobu (-0,7525)	-
<b>6</b>	počet repetice (0,8517)	pravidelnost energie (0,6926)	-	-
<b>opakované</b>	počet repetice (0,8895)	prům. ticho (0,8671)	pravidelnost energie (0,7912)	-
<b>fixované</b>	rychlost (-0,8568)	RPP (0,8163)	PMBD (-0,8121)	počet změn za dobu (-0,8106)
<b>celková známka</b>	pravidelnost energie (0,8589)	prům. Ticho (0,8583)	RPP (-0,8364)	SOI11 (0,8293)

## 5. Závěr

V článku jsou uvedeny výsledky automatických metod pro hodnocení koktavosti a neplynulosti. Experimenty byly provedeny na čtených promluvách. Parametry zpracovávají vždy celou řečovou promluvu. Výstupy parametrů jsou srovnány s kontrolními daty vytvořenými pomocí hodnotícího systému LBDL.

Výsledky ukazují, že uvedené parametry jsou vhodné pro automatické a objektivní hodnocení koktavosti. Například shoda parametru pravidelnost energie s celkovou známkou je 0,8589. Další práce by mohly směřovat ke kombinaci parametrů a rozšířením použití na spontánní promluvy.

## Poděkování

Tento výzkum je podporován grantem GACR 102/12/2230 a SGS12/185/OHK4/3T/13. Rádi bychom také poděkovali Tereze Tykalové za ohodnocení signálů.

## Reference

- [1] Škodová, E., Jedlička, I, a kolektiv. *Klinická logopedie*. Praha, Portál, 2003.
- [2] Lechta, V. *Poruchy plynulosti řeči*. Praha: Scriptorium, 1999.
- [3] Howell, P., Hamilton, A., Kyriacopoulos, A. Automatic detection of repetitions and prolongations in stuttered speech, *Speech Input/Output: Techniques and Applications*, IEE Publications, 252–256, 1986.
- [4] Wiśniewski, M., Kuniszyk-Józkowiak, W., Smołka, E., and Suszyński, W. Automatic detection of disorders in a continuous speech with the hidden Markov models approach. In *Computer Recognition Systems 2*, vol. 45 of *Advances in Soft Computing*, pp. 445–453, Springer, Berlin, Germany, 2007.
- [5] Nöth, E., Niemann, H., Haderlein, T., Decher, M., Eysholdt, U., Rosanowski, F., Wittenberg, T. (2000): *Automatic stuttering recognition using hidden Markov models*, In *ICSLP-2000*, vol.4, 65-68.
- [6] Bergl, P., *Objektivizace poruch plynulosti řeči*. Disertační práce, Fakulta elektrotechnická, ČVUT v Praze, 2010.
- [7] Teesson, K., Packman, A., and Onslow, M. (2003). The Lidcombe behavioral data language of stuttering. *Journal of Speech, Language, and Hearing Research*, 46, 1009–1015.
- [8] Goberman, A., M., Blomgren, M., Metzger, E. Characteristics of speech disfluency in Parkinson disease. *Journal of Neurolinguistics*, 23, 470–478, 2010.
- [9] Cmejla, 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.
- [10] C. Cucchiaroni, H. Strik, and Boves L. Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *J. Acoustic Soc. Am*, 107:989-999, 2000.



# Objektivizace chraptivosti - porovnání výpočtu HNR ve frekvenční oblasti s Praatem

Adam Stráník, Roman Čmejla

České vysoké učení v Praze, Fakulta elektrotechnická, Katedra teorie obvodů  
stranada@fel.cvut.cz

*Abstrakt:* Akustická analýza hlasu je neinvazivní metoda, která je obecně uznávaná jako efektivní a objektivní vyšetřovací nástroj. Při analýze patologických hlasů (například chraptivosti) je poměr harmonické a šumové složky (HNR) hlasu jedním z velmi často používaných parametrů – je totiž schopen odhalit míru patologické šumové příměsi ve znělých úsecích řeči. Zvýšená šumová příměs může být způsobena nedokonalým uzavřením hlasivek během fonace a to např. z důvodů otoku hlasivek, polypů atd. V příspěvku je popsán algoritmus odhadu HNR v řeči, který pracuje iterativně ve frekvenční oblasti. Jsou popsány silné a slabé stránky algoritmu a je ukázáno, jak změna vstupních parametrů algoritmu ovlivní výslednou vypočtenou hodnotu HNR. Dále je provedeno porovnání hodnot vypočtených výše zmíněným algoritmem s hodnotami vypočtenými programem Praat - celosvětově používaným a uznávaným nástrojem pro analýzu hlasu.

## 1. Úvod

Patologický hlas je často ovlivněn vysokou šumovou příměsí, přičemž míra šumové příměsi může být úspěšně použita pro stanovení míry patologie hlasu YUMOTO et al. [1984], SHAMA et al. [2007].

Akustický signál hlasu může být popsán následující rovnicí

$$x(k) = s(k) + w(k), \quad (1)$$

kde  $x(k)$  je výsledný řečový signál,  $s(k)$  je periodická část hlasu, která je generována hlasivkami a  $w(k)$  je šumová příměs generovaná proudem vzduchu z plic. V normálním (zdravém) hlasu je míra šumové příměsi  $w(k)$  nízká a téměř zanedbatelná v porovnání s harmonickou složkou  $s(k)$ . Vlivem špatné funkce hlasivek, jejich nedomykavosti, vzrůstá podíl šumové složky na úkor harmonické složky. Nedomykavost hlasivek může být způsobena otokem hlasivek, polypem na hlasivkách atd.

Dobře známý a často používaný parametr HNR (harmonics-to-noise ratio – poměr harmonické a šumové složky) je definován následujícím vztahem

$$\text{HNR} = 10 \log \left( \frac{En_{s(k)}}{En_{w(k)}} \right) \quad [\text{dB}], \quad (2)$$

kde  $En_{s(k)}$  je energie periodické složky hlasu a  $En_{w(k)}$  je energie šumové složky hlasu.

Hodnota HNR 0 dB znamená, že obě složky hlasu, harmonická a šumová, mají stejnou energii – jsou v hlasu zastoupeny stejně intenzivně. Takový hlas není vnímán jako zdravý, ale jako dyšný. Zdravé hlasy mají hodnoty HNR vyšší než 15 dB.

Není zaveden jednotný postup jak nejlépe oddělit šumovou a harmonickou složku. Byla vypracována řada postupů: v časové oblasti YUMOTO et al. [1984], ve frekvenční oblasti SHAMA et al. [2007], YEGNANARAYANA et al. [1998], použitím vlnkové transformace MANFREDI [2000] nebo kepstrální analýzy MURPHY and AKANDE [2005].

V tomto článku je popsána metoda oddělení harmonické a šumové složky iterativním algoritmem pracujícím ve frekvenční oblasti, která byla poprvé představena YEGNANARAYANA et al. [1998]. V článku je provedena původní implementace v prostředí MATLAB a je provedeno porovnání hodnot a hodnotami spočtenými programem Praat BOERSMA and WEENINK [2011], který je uznávaným, volně dostupným nástrojem pro analýzu hlasu a řeči.

## 2. Databáze

Společnost Kay Elemetrics distribuuje CD-ROM obsahující databázi nahrávek cca 700 pacientů. Tyto nahrávky jsou pořízeny Messachtsetskou oční a ušní klinikou a laboratořemi pro vyšetření hlasu a řeči (Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Labs) PENTAX et al.. Nahrávky obsahují prodlouženou fonaci hlásky /a/ vždy 1 až 3 sekundy. Soubor obsahuje 53 nahrávek zdravých pacientů a 657 nahrávek od pacientů se širokou škálou organických, neurologických, traumatických a psychogenních poruch v různých stádiích progresu. Nahrávky byly pořízeny v kontrolovaném prostředí s jasně definovanými akustickými parametry. Při nahrávání byla u části nahrávek použita vzorkovací frekvence 25 kHz a u části 50 kHz, rozlišení 16 bitů na vzorek je použito ve všech případech. K experimentům byly všechny nahrávky převzorkovány na 25 kHz.

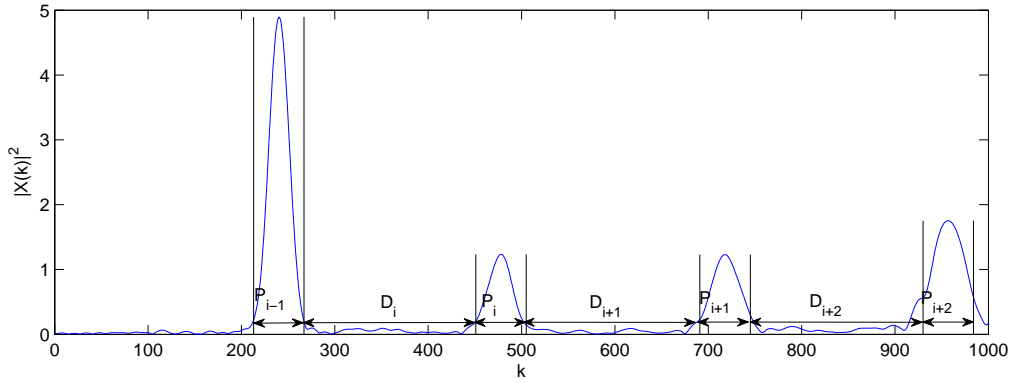
## 3. Algoritmus

Algoritmus použitý v této práci byl poprvé představen YEGNANARAYANA et al. [1998], dodatečné úpravy zvyšující stabilitu byly navrženy v STRÁNÍK and ČMEJLA [2011].

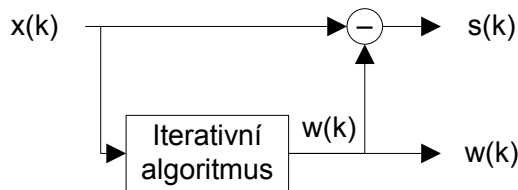
### 3.1. Popis algoritmu

Algoritmus je založen na následující myšlence: frekvenční spektrum znělých mikrosegmentů řeči obsahuje harmonické úseky  $P_i$  a mezery mezi harmonickými úseky  $D_i$ , viz Obr. 1. Zatímco šumová komponenta je v určité míře rozprostřena přes celé frekvenční spektrum, harmonická složka je soustředěna pouze do harmonických oblastí  $P_i$  ve frekvenčním spektru.

Vynulování harmonických komponent  $P_i$  ve frekvenčním spektru způsobí, že signál po převodu zpět do časové oblasti obsahuje méně harmonické složky  $s(k)$ . Několikerým vynulováním harmonických oblastí  $P_i$  ve spektru lze dosáhnout stavu, kde výsledný signál v časové oblasti obsahuje pouze šumovou složku  $w(k)$ . Pokud tuto šumovou složku v časové oblasti odečteme od původního analyzovaného signálu  $x(k)$ , získáme harmonickou složku v časové oblasti  $s(k)$ . Pak lze k výpočtu výsledné hodnoty HNR aplikovat vztah (2). Pro ilustraci procesu separace harmonické a šumové složky viz Obr. 2.



Obrázek 1: Ukázka harmonických oblastí  $P_i$  a mezer mezi harmonickými úseky  $D_i$  ve výkonovém frekvenčním spektru.



Obrázek 2: Rozklad původního signálu  $x(k)$  na šumovou a harmonickou komponentu  $w(k)$  a  $s(k)$  pomocí iterativního algoritmu.

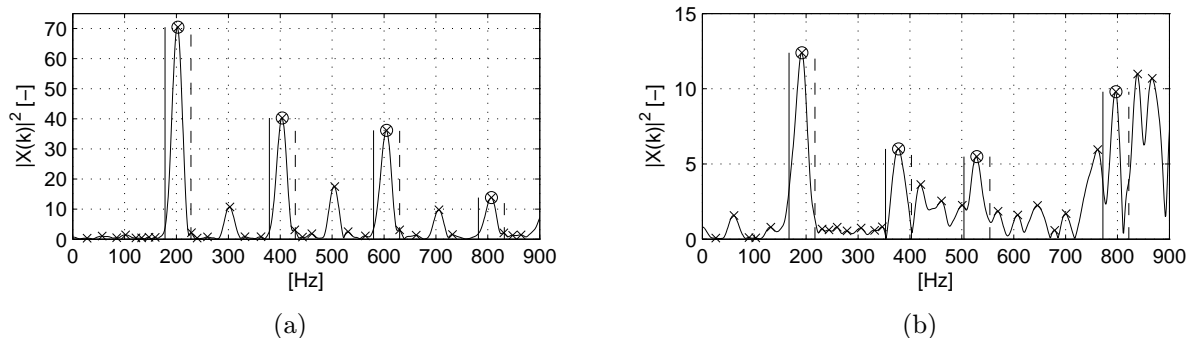
### 3.2. Omezení algoritmu

Hlavní omezení algoritmu jsou následující: 1) analyzované mikrosegmenty musí být relativně dlouhé ( $\sim 80$  ms), aby bylo zajištěno, že se harmonické oblasti ve frekvenčním spektru nepřekrývají (STRÁNÍK and ČMEJLA [2011]); 2) některé patologické hlasy mohou obsahovat tzv. subharmonické složky (OMORI et al. [1997]) – subharmonická složka spektra se projevuje jako výraznější lokální maximum mezi dvěma harmonickými; vzhledem k principu detekce harmonických oblastí nejsou tyto úseky spektra nulovány a falešně zvyšují energii šumové části signálu (ukázka signálu se subharmonickými je na Obr. 3(a)); 3) u velmi dyšných hlasů je obtížné detekovat správné harmonické oblasti a to z důvodu velkého množství „falešných“ harmonických oblastí, viz Obr. 3(b).

## 4. Experimenty

### 4.1. Počet iterací algoritmu

Jak je zmíněno výše, pro získání šumové komponenty jsou opakovaně nulovány harmonické regiony ve frekvenčním spektru signálu. V rámci analýzy algoritmu byl proveden následující experiment: jak ovlivňuje počet iterací algoritmu výslednou hodnotu HNR? K experimentu byly použity dvě nahrávky, jedna nahrávka zdravého pacienta s vysokým HNR a dobře definovanými harmonickými regiony a jedna nahrávka pacienta s nízkou hodnotou HNR a špatně definovanými harmonickými regiony.



Obrázek 3: Ukázky spekter problematických signálů: (a) subharmonické, (b) vysoká šumová příměs. Harmonické oblasti jsou části spektra ohraničené zleva plnými a zprava přerušovanými horizontálními čarami. Tyto oblasti jsou algoritmem nulovány.

## 4.2. Porovnání s Praatem

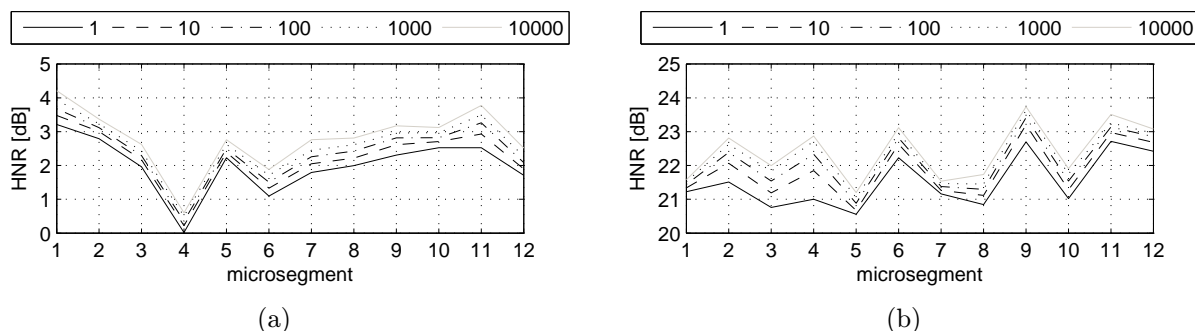
Hlavním experimentem je porovnání středních hodnot HNR jednotlivých nahrávek s hodnotami určenými programem pro analýzu hlasu a řeči Praat BOERSMA and WEENINK [2011].

Oba algoritmy byly nastavené tak, aby pro každou nahrávku vracely stejný počet hodnot (stejná délka mikrosegmentu a jejich překrytí). Střední hodnoty z programu Praat pro každou nahrávku poté byly seřazeny od nejmenší do největší, tzn. od nejvíce patologických (dyšných) k nejzdravějším hlasům.

## 5. Výsledky

### 5.1. Počet iterací algoritmu

Experiment lze shrnout následovně: časové průběhy HNR jsou tvarově velmi podobné pro všechny testované počty iterací, viz Obr. 4 a téměř všechny rozdíly byly v rozmezí 1 dB. Největší rozdíl je ovšem v čase nutném pro běh algoritmu, viz Tab. 1. Pro následující experimenty byl použit algoritmus s deseti iteracemi.



Obrázek 4: Ukázka vlivu počtu iterací na výslednou hodnotu HNR na signálu s nízkou HNR (a) a na signálu s vysokou HNR (b).

iterací	1	10	100	1 000	10 000
doba odhadu [s]	0,36	0,65	5,32	51,86	518,73

Tabulka 1: Čas potřebný k výpočtu HNR pro různé počty iterací algoritmu. Délka testovacího signálu je 1 s.

## 5.2. Porovnání s Praatem

Porovnání bylo provedeno pro obě skupiny – pro patologické hlasy (P) a pro zdravé hlasy (Z). V Tab. 2 jsou zobrazeny výsledky porovnání středních hodnot HNR: Spearmanův korelační koeficient a střední hodnota a směrodatná odchylka rozdílů středních hodnot HNR pro každou nahrávku.

	$\rho$ [-]	$\varnothing \pm \sigma$ [dB]
Patologičtí	0,95	-1,18 $\pm$ 3,91
Zdraví	0,87	1,10 $\pm$ 1,54

Tabulka 2: Spearmanův korelační koeficient mezi středními hodnotami HNR iterativního algoritmu a Praatu pro každou nahrávku a střední hodnota a směrodatná odchylka rozdílů středních hodnot pro každou nahrávku.

Pro patologické hlasy vychází Spearmanův korelační koeficient ( $\rho_P = 0,95$ ) překvapivě vyšší než pro zdravé hlasy ( $\rho_Z = 0,87$ ). To může být způsobeno výrazně rozdílnou velikostí souborů patologických (657) a zdravých (53) hlasů.

Za povšimnutí stojí také rozdílné znaménko střední hodnoty rozdílů hodnot z Praatu a z iterativního algoritmu. Zatímco patologické hlasy mají zápornou hodnotu  $\varnothing_P = -1,18$  dB, zdravé hlasy mají kladnou hodnotu  $\varnothing_H = 1,10$  dB. Je nutné si uvědomit, že záporná hodnota znamená, že iterativní algoritmus přiřadil nahrávce v průměru vyšší hodnotu HNR než Praat a vyšší hodnota znamená méně dyšný a tím pádem zdravější hlas.

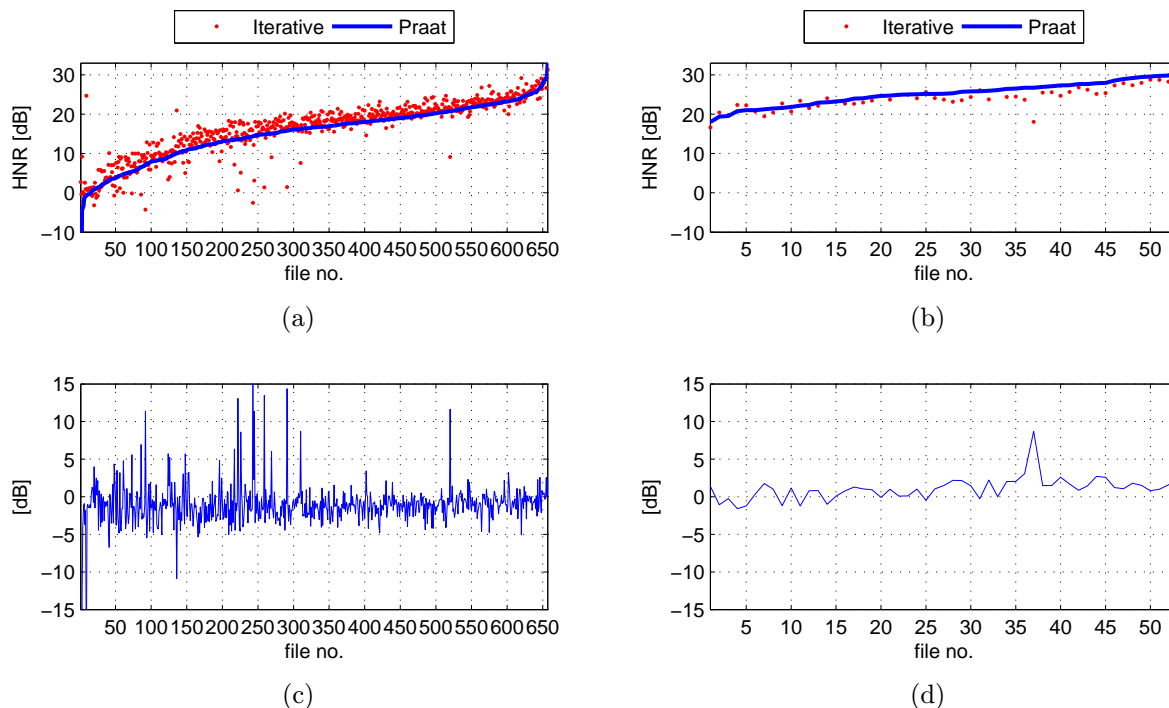
Obr. 5 ukazuje porovnání středních hodnot HNR pro patologické a zdravé hlasy. Na Obr. 5(c) stojí za povšimnutí větší rozdíly středních hodnot HNR nejvíce patologických hlasů patologických pacientů (hodnoty nejvíce vlevo). Jak je zmíněno výše, střední hodnoty HNR vypočtené Praatem byly seřazeny od nejmenších po největší (v Obr. 5(c) zleva doprava). Některé z nejvíce patologických hlasů jsou až afonické, tzn. že během fonace nedochází k pohybu hlasivek.

## 6. Závěr

V příspěvku je analyzován iterativní algoritmus pro odhad šumové složky signálu, který je dále použit k výpočtu HNR. Je provedeno porovnání těchto hodnot s hodnotami vypočtenými programem Praat. K testovacím účelům je použita databáze Kay Elmetrics.

Analýza iterativního algoritmu odhalila, že počet iterací algoritmu výrazně neovlivňuje výslednou hodnotu HNR: rozdíly mezi jednou a deseti tisíci iterací jsou převážně menší než 1 dB.

Porovnání středních hodnot HNR z testovací databáze spočtených jednak iterativním algoritmem a jednak programem Praat ukázalo velmi dobrou korelaci těchto hodnot:



Obrázek 5: Porovnání středních hodnot HNR získaných iterativním algoritmem a Praatem pro jednotlivé nahrávky. Přímé porovnání patologických (a) a zdravých (b) a porovnání rozdílů hodnot patologických (c) a zdravých (d).

$\rho_P = 0,95$  pro patologické a  $\rho_H = 0,87$  pro zdravé hlasy. Největší rozdíly jsou podle předpokladů u nejvíce patologických hlasů, které mohou být až afonické.

## Poděkování

Tato práce je podporována z grantů **GACR 102/12/2230** a **SGS12/185/OHK4/3T/13**.

## Reference

- P. BOERSMA and D. WEENINK. Praat: doing phonetics by computer [computer program], 2011. URL <http://www.praat.org/>. Version 5.2.35.
- C. MANFREDI. Adaptive noise energy estimation in pathological speech signals. *Biomedical Engineering, IEEE Transactions on*, 47(11):1538–1543, 2000. doi: 10.1109/10.880107.
- P. J. MURPHY and O. O. AKANDE. Quantification of glottal and voiced speech harmonics-to-noise ratios using cepstral-based estimation. In *NOLISP*, volume 3817, pages 150–160, 2005. doi: [http://dx.doi.org/10.1007/11613107\\_13](http://dx.doi.org/10.1007/11613107_13).
- K. OMORI, H. KOJIMA, R. KAKANI, D. H. SLAVIT, and S. M. BLAUGRUND. Acoustic characteristics of rough voice: Subharmonics. *Journal of Voice*, 11(1):40–47, 1997.

- K. PENTAX, M. Eye, and E. Infirmery. Disordered voice database and program, model 4337. online.
- K. SHAMA, A. KRISHNA, and M. U. CHOLAYYA. Study of harmonics-to-noise ratio and critical-band energy spectrum of speech as acoustic indicators of laryngeal and voice pathology. *EURASIP J. Appl. Signal Process.*, pages 50–50, 2007.
- A. STRÁNÍK and R. ČMEJLA. An analysis of iterative algorithm for estimation of harmonics-to-noise ratio in speech. In *Technical Computing Prague 2011*, pages 1–7, 2011.
- B. YEGNANARAYANA, C. d’ALESSANDRO, and V. DARSINOS. An iterative algorithm for decomposition of speech signals into periodic and aperiodic components. *IEEE Transactions on Speech and Audio Processing*, 6(1):1–11, 1998.
- E. YUMOTO, Y. SASAKI, and H. OKAMURA. Harmonics-to-noise ratio and physiological measurement of the degree of hoarseness. *JSHLR*, 27:2–6, 1984.