

FORANA

Pavel GRILL¹, Jana TUČKOVÁ²

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

Abstrakt

Jedním z příznaků vývojové dysfázie je částečná porucha tvorby a porozumění řeči nebo její úplná ztráta. Aby bylo možné provádět výzkum řeči nemocných dětí, je nejprve nutné provádět správnou parametrizaci řečového signálu u zdravých dětí. Tato práce se věnuje programu FORANA, který byl vytvořen z důvodu správné extrakce formantových kmitočtů z řečového signálu.

1. Úvod

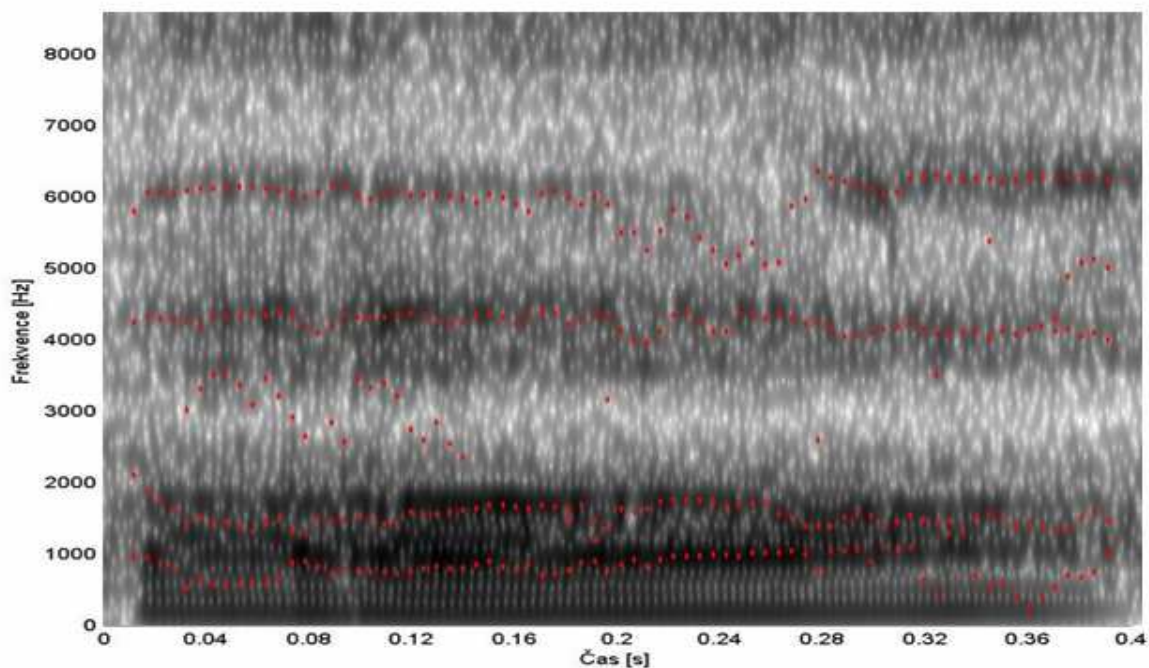
Vývojová dysfázie je onemocnění, které se projevuje ztíženou schopností nebo neschopností naučit se verbálně komunikovat, i když podmínky pro rozvoj řeči jsou přiměřené. Toto onemocnění je způsobené poruchou centrálního zpracování řečového signálu. Je zde předpoklad, že k příčině tohoto stavu dochází díky difuznímu postižení centrálního nervového systému, který zasahuje v podstatě celou centrální korovou oblast. Příčiny vzniku mohou být jednak v poškození mozku v období prenatálním, perinatálním a postnatálním, ale dále se uvažuje, že na onemocnění má vliv i dědičnost. Poruchy v řeči a opoždění vývoj řeči se spíše objevují u chlapců než u dívek, a to asi v poměru 4:1. Z EEG vyšetření byly též u některých pacientů zjištěny epileptické výboje, kde se nalézají řečová a sluchová centra. Nejvýraznější znak vývojové dysfázie je nápadně opožděný vývoj řeči. Další příznak je nerovnoměrný vývoj osobnosti. Vývojovou dysfázi lze nejen detekovat a analyzovat ze signálu mozkové aktivity, tj. z EEG vyšetření, ale i z analýz pacientových promluv. Předpokládá se, že je u této poruchy ovlivněna tvorba řeči, dochází k posunu formantových kmitočtů, které lze z řečového signálu extrahovat. Cílem celého projektu, který je společný pro katedru teorie obvodů FEL ČVUT v Praze a Dětské neurologické kliniky 2. Lékařské fakulty UK v Praze - Motole, je nalezení vztahů mezi klinickými a elektrofyziologickými projevy u dětí, které trpí touto poruchou. Dílčím úkolem tohoto projektu je volba optimální parametrizace s využitím umělých neuronových sítí (UNS). S tímto úkolem souvisí vznik programu FORANA (Formant Analysis).

2 Vznik akustického signálu řeči v mluvidlech

Všechny zvukové složky řeči [PAL94] jsou složeny zvuky mající tónovou a šumovou povahu. Zdrojem zvuku tónového charakteru je hlasové ústrojí. Základním prvkem řeči je základní tón, který se označuje F0. Charakterizuje jej kmitání hlasivek a lidského zabarvení nabývá po průchodu dutinami nad hrtanem, kde dochází k rezonancím, které jej zesilují a přidávají svrchní harmonické tóny - formanty. Takto obohacený zvuk vnímáme jako lidský hlas.

2.1 Formanty

Formant [PAL94], [GRI08] je obecně charakterizován jako koncentrace akustické energie v okolí určité frekvence. V řečovém pásmu se jich nalézají více. Jednotlivé formanty se vyskytují na různých frekvencích odpovídajících zhruba intervalům 1000 Hz pro dospělého mluvčího. Velice zřetelně si je můžeme prohlédnout na spektrogramu, kde jsou zobrazeny tmavými pásy. Čím jsou tyto pásy tmavší, tím více akustické energie charakterizují. Na spektrogramu (Obr. 1) je namapováno pět formantových pásem, které jsou zobrazeny červenými body.



Obr. 1 Zobrazení prvních pěti formantových pásem červenými body ve spektrogramu.

Změny základního tónu a formantů ovlivňují vnímání hlasu. Změny v základním tónu interpretujeme jako změny v melodii řeči. Změnám prvního formantu (F1) odpovídají změny ve vertikálním pohybu jazyka, změnám druhého formantu (F2) odpovídají změny v horizontálním pohybu jazyka a třetí formant (F3) je ovlivňován změnami v dutině nosní. Pro samohlásky jsou tedy nejvýznamnější první dva až tři formanty, a to v pořadí F2, F1, F3. Vyšší formanty jsou pro všechny samohlásky prakticky stejné, obsahují značnou část informace o barvě hlasu mluvčího. Pokud si dáme první dva formanty (F1 a F2) do vzájemného vztahu, dostaneme tzv. vokální trojúhelník. Ten nám jednotlivé samohlásky dělí do tří tříd v závislosti na poloze těchto formantů. První třídu tvoří samohláska **a**, druhou třídu tvoří samohlásky **e** a **i** a třetí třída je tvořena samohláskami **o** a **u**.

Vokální trojúhelník						
Vertikální poloha jazyka	Horizontální poloha jazyka					
		Přední	střední	Zadní		
	vysoké	i,í				u,ú
	středové		e,é		o,ó	
nízké			a,á			

Tab. 1 Vokální trojúhelník [PAL94].

2.2 Výpočet formantových frekvencí

Většina postupů [GRI08], [PSU06], [TUC09], [UHL07] pro identifikaci formantových kmitočtů pracuje ve frekvenční oblasti a vychází z analýzy spektrální obálky stanovené metodou LPC. Existují v podstatě dva postupy jak určit ze spektra LPC hodnoty formantových frekvencí. První z nich zjišťuje kořeny polynomu $A(z)$ (tedy póly přenosové funkce $H(z)$) a druhý, který hledá na spektrální obálce lokální maxima odvozená z lineárního prediktoru. Výpočet pólů přenosové funkce zjistíme z kořenů polynomu $A(z)$, které se zjistí vyřešením rovnice (1)

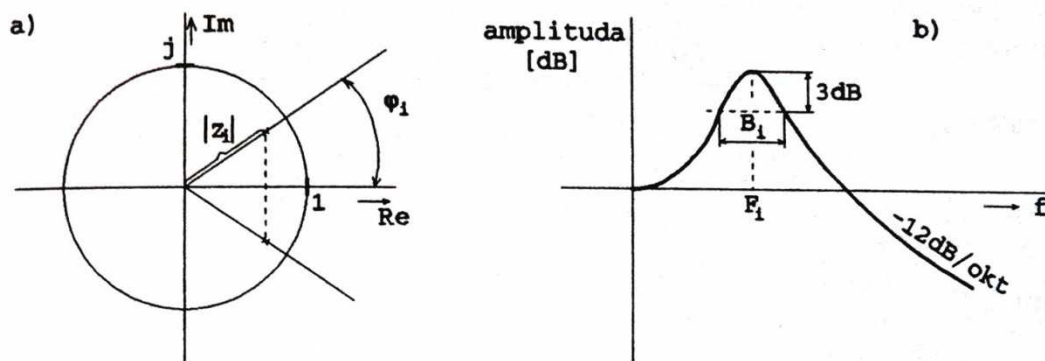
$$z^Q + a_1 z^{Q-1} + a_2 z^{Q-2} + \dots + a_{Q-1} z + a_Q = 0 \quad (1)$$

Tato rovnice je Q -tého řádu s reálnými koeficienty. V jejím řešení převažují páry komplexně sdružených kořenů. Uvažme jednu dvojici komplexně sdružených kořenů $z_i = |z_i| e^{j\varphi}$ a $\bar{z}_i = |z_i| e^{-j\varphi}$ rozloženou v z -rovině podle Obr. 2a (φ_i je argument z_i [rad]). Odpovídající formantovou frekvenci F_i a šířku pásma formantu B_i pro pokles charakteristiky o 3dB lze vyjádřit pomocí vztahů (2 a 3):

$$F_i = \frac{f_s \cdot \arg z_i}{2\pi} [Hz] \quad (2)$$

$$B_i = -\frac{f_s \cdot \arg z_i}{2\pi} [Hz] \quad (3)$$

kde T je perioda vzorkování původního akustického signálu. Souvislost veličin F_i a B_i lze znázornit ve frekvenční oblasti obrázkem Obr. 2b.



Obr. 2 Ilustrace rozložení: a) dvojice komplexně sdružených kořenů polynomu $A(z)$, b) odpovídající frekvenční charakteristika [PSU95].

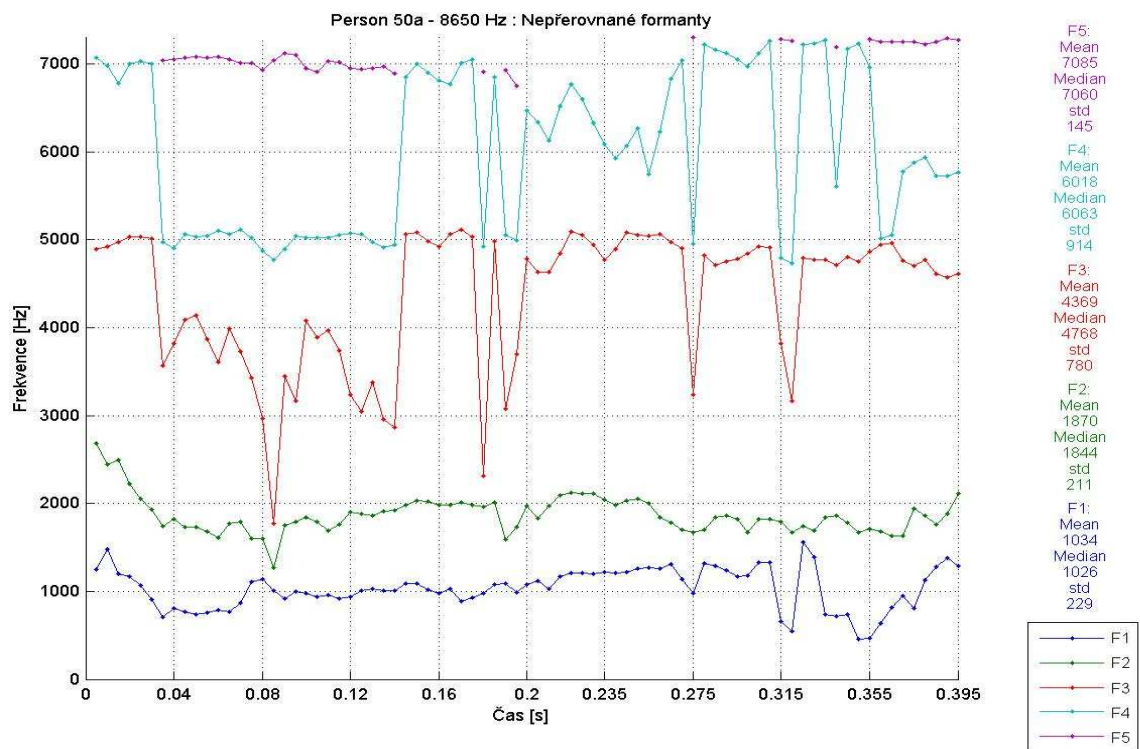
3. FORANA

Pro náš výzkum v oblasti zpracování řečového signálu dětí s neurologickou poruchou se snažíme preferovat popis řečového signálu pomocí formantové analýzy. Softwarový program FORANA [GRI08] byl vytvořen v programovém prostředí MATLAB. K vývoji programu nás vedl především požadavek korektní formantové analýzy. Původně se extrakce formantových kmitočtů z řečového signálu zajišťovala pomocí programu PRAAT [1]. Tento program byl vytvořen Paulem Boersmanem a Davidem Weeninkem z katedry fonetiky na amsterdamské univerzitě a je uznáván a používán fonetiky na celém světě. Při analýze řečového signálu se objevují chyby v klasifikaci formantů. Nelze proto považovat výsledky získané programem PRAAT za bezchybné. Bylo přistoupeno k vývoji programu, který by formanty klasifikoval s minimálními chybami. Dalším požadavkem, který ovlivňoval vznik FORANY, byla nutnost zautomatizovat celý proces extrakce formantů z řečového signálu. Bez toho nelze provádět efektivně výzkum dětské řeči, jelikož množství dat, které je zapotřebí zpracovat, je velké. Na následujícím obrázku (Obr. 3) je vidět grafická ukázka programu FORANA.

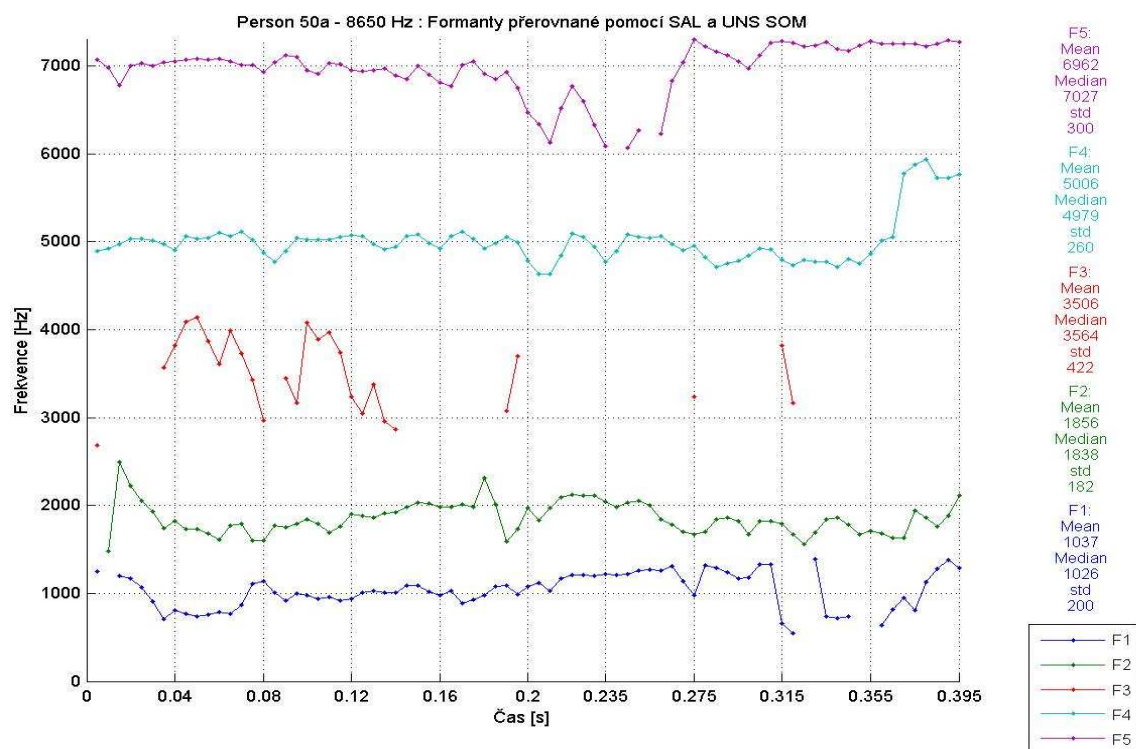
O správnost klasifikace formantových kmitočtů se stará implementovaný přerovnávací algoritmus SAL a umělá neuronová síť, a to samoorganizující se mapa (SOM). Nejprve jsou vypočtené formanty zkontrolovány a případné chyby v jejich určení opraveny, aby došlo k jejich správné klasifikaci. Na následujících grafech jsou ukázány vypočtené formanty pomocí tohoto programu. V prvním grafu (Obr. 4) vidíme formanty, na které nejsou použity korekční metody opravující jejich chybnou klasifikaci. V druhém grafu (Obr. 5) jsou zobrazeny formanty získané po použití korekčních metod.



Obr. 3 Programové prostředí FORANA.



Obr. 4 Nepřerovnané formanty.



Obr. 5 Přerovnané formanty pomocí SAL a následně UNS SSOM.

Pro ověření správnosti našich postupů zavedeme statistické jednotky [ZVA04]. Těmi jsou průměr a směrodatná odchylka. Průměr (4), neboli střední hodnotu, vypočítáme, jestliže součet všech hodnot dělíme jejich počtem:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (4)$$

Pokud jsou pozorování soustředěna kolem svého průměru, je jejich variabilita malá. Jsou-li data naopak roztroušena ve značné vzdálenosti od průměru, pak je jejich variabilita velká. Toho si můžeme všimnout i u formantových kmitočtů. Proto je zapotřebí zavést pojem směrodatná odchylka, která se vypočte z následujícího vztahu (5).

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^k n_i (x_i - \bar{x})^2} \quad (5)$$

V Tab. 2 je statisticky zpracovaná směrodatná odchylka. V procentech je vyjádřena úspěšnost klasifikace přerovnaní. Čím je hodnota vyjádřená v procentech menší, tím je věci,

vysvětleny zkratky k metodám, které jsou použity v předchozí tabulce. V Tab. 3 jsou popisky k Tab. 2.

Směrodatná odchylka		
	N	SAL → SOM
$\sum_{i=1}^5 F_i$	2283	1497
%	100	66

Tab. 2 Směrodatná odchylka.

Pojmy	
N	nepřerovnané formanty
SAL → SSOM	Formanty přerovnané korekčními metodami

Tab. 3 Popis zkratk použitých v Tab.1.

4. Závěr

Z prezentovaných výsledků vyplývá, že použití korekčních metod, které přerovňávají formanty má své opodstatnění, jelikož nám zaručuje jejich správnou klasifikaci. Předností ve využití kombinací klasické metody a umělé neuronové sítě spočívá v tom, že se odstraní nedostatky obou zmíněných metod a dojde se k relevantním a správným výsledkům analýzy řečového signálu.

Poděkování

Tato práce je podporována z výzkumného záměru č.MSM6840770012 „Transdisciplinární výzkum v biomedicínském inženýrství 2“ (experimentální část) a grantů GAČR č. 102/09/0989 „Nové perspektivní metody vysoce kvalitní syntézy mluvené češtiny“ (teoretická část), „Analýza a modelování biologických a řečových signálů“, GAČR č. 102/08/H008, a "Rozpoznávání mluvené řeči v reálných podmínkách" ("Speech Recognition under Real-World Conditions"), GACR 102/08/0707.

Reference

- [GRI08] GRILL, P.: *Analýza promluv dětí v závislosti na věku*. Diplomová práce ČVUT, 2008
- [PAL94] PALKOVÁ, Z.: *Fonetika a fonologie češtiny*. Univerzita Karlova, Praha 1994, ISBN 80-7066-843-1.
- [PSU06] Psutka, J.-Muller, L.-Matousek, J.-Radová, V.: *Mluvíme s počítačem česky*, Academia Praha, 2006, ISBN 80-200-0203-0.
- [TUC09] TUCKOVÁ, J.: *Vybrané aplikace umělých neuronových sítí při zpracování signálu*. Nakladatelství CVUT Praha, leden 2009. ISBN 978-80-01-04229-8.
- [UHL07] UHLÍŘ, J., SOVKA, P., POLLÁK P., HANŽL V., ČMEJLA R.: *Technologie hlasových komunikací*. Vydavatelství ČVUT, 2007, 276 s., ISBN 978-80-01-03888-8.

[ZVA04] ZVÁROVÁ J.: *Základy statistiky pro biomedicínské obory*. Karolinum Praha 2004, třetí dotisk 1. vydání, ISBN 80-7184-786-0.

[1] PRAAT

<http://www.fon.hum.uva.nl/praat/>

Kontaktní informace:

1. autor:

Ing. Pavel Grill,

grillpavel@seznam.cz, tel: +420 737955555

2. autor:

Doc. Ing. Jana Tučková

tuckova@fel.cvut.cz, tel: +420 721502556