

# Vyhodnocování promluv dětí s poruchami řeči

*Petr Zlatník, Roman Čmejla*

Katedra teorie obvodů, Fakulta elektrotechnická, ČVUT, Praha

## Abstrakt

**Příspěvek popisuje metodu, která byla vyvinuta pro vyhodnocení vad řeči dětí. Princip metody je založen na algoritmu borcení časové osy, pro zvýšení robustnosti je využito více vstupních charakteristik řeči a výsledky jsou průměrovány. V článku jsou uvedeny dílčí výsledky rozboru vlivu jednotlivých řečových parametrizací s vyhodnocením modifikovaných přístupů, kdy je použito méně parametrizací. Dále je uvedeno vyhodnocení vlivu funkce průchodu maticí vzdáleností DTW na přesnost klasifikace a je zmíněno hodnocení vlivu změn fonetické struktury jako je vynechání hlásky nebo slabiky v testovaných slovech.**

## 1 Úvod

Uvedená metoda slouží ke klasifikaci vad řeči dětí postižených vývojovou dysfázií (vývojová nemluvnost dítěte, dítě má problém s řečí již od doby, kdy začíná mluvit) popřípadě afázií (porucha mozkových center, která odpovídá za tvorbu řeči za stavu, když již dítě umělo mluvit, pokud se přidají epileptické výboje v mozku, jedná se o tzv.: Landau-Kleffnerův syndrom). Děti mají problémy s vyslovováním slov (některé hlásky jsou prodlouženy, zaměněny nebo vynechány), nejsou schopny vyslovit souvislejší větu.

Projekt je řešen ve spolupráci s Fakultní nemocnicí v Motole, kde jsou nahrávány promluvy postižených dětí včetně léčby. Metoda byla navržena s cílem oddělit promluvy nemocných dětí od zdravých a hlavně by měla být schopna vyhodnotit, zda se stav dítěte v průběhu léčby lepší nebo ne.

## 2 Princip metody

### 2.1 Algoritmus DTW a realizace klasifikátoru

Algoritmus DTW [9] byl původně určen pro realizaci rozpoznávače izolovaných slov. Byla provedena jeho modifikace [1, 2, 10] tak, aby ho bylo možno využít pro klasifikaci nesrozumitelnosti dětských promluv. Jednotlivé promluvy jsou segmentovány s překryvem 50 %, segmenty jsou váhovány Hammingovým oknem a parametrizovány. Popis parametrizací je v následujícím odstavci. Následně je vypočítána z porovnávaných dvou promluv matice euklidovských vzdáleností [9]. Ta je procházena z jednoho rohu do protilehlého a je vypočítán celkový součet akumulovaných vzdáleností, který charakterizuje míru odlišnosti obou slov. Toto je základní princip popisované metody klasifikace slov, kdy se porovnávaly promluvy jednoho slova od zdravých a nemocných dětí a pro promluvu nemocného dítěte se předpokládala větší akumulovaná vzdálenost.

Protože při využití jedné parametrizace (první pokusy byly prováděny pro keprstrální koeficienty) docházelo občas ke vzniku chyb a špatné klasifikaci slov, byl navržen systém [1, 2, 10], kdy se využívá více různých vstupních charakteristik řeči. Tím jsou promluvy hodnoceny z více fonologických aspektů a dojde k eliminaci chyb.

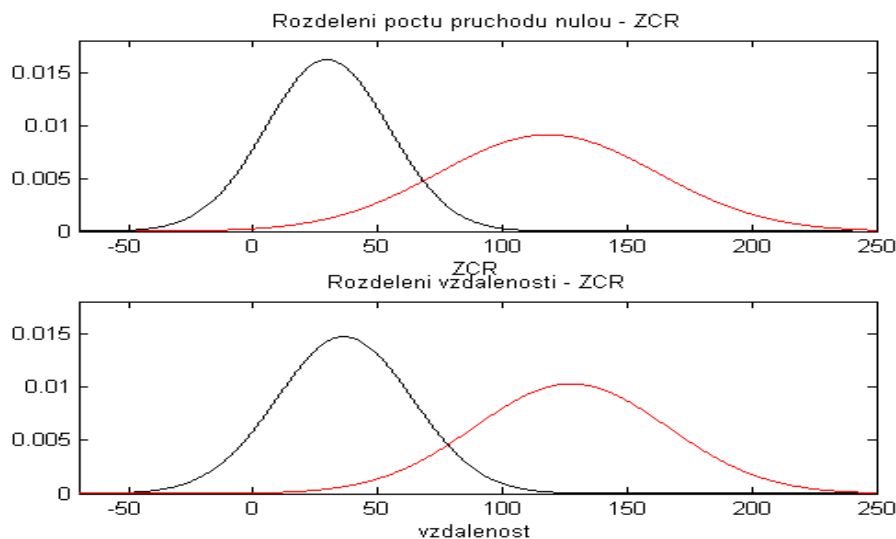
Nejprve jsou všechny promluvy (pro konkrétní testované slovo, např.: mateřídouška) od zdravých dětí porovnány pomocí algoritmu DTW navzájem a je vypočítána průměrná akumulovaná vzdálenost každého slova zdravého dítěte ode všech ostatních. Následně jsou porovnány promluvy nemocných dětí (nebo obecněji nové testované promluvy) s promluvami zdravých (již dříve zpracovaných a uložených v referenční databázi) a opět je vypočítána průměrná vzdálenost promluv nemocných dětí ode všech zdravých. Tyto výpočty jsou provedeny pro všechny využitě

parametrizace. Aby bylo možné výsledky spojit dohromady, je nutné provést pro každou parametrizaci normování vzdáleností (maximální hodnota pro zdravé děti je rovna jedné) a výsledky je možné sečíst přes všechny parametrizace, tím dojde při správné funkci algoritmu ke spolehlivějšímu a výraznějšímu oddělení promluv nemocných dětí.

## 2.2 Vybrané řečové charakteristiky a jejich vliv

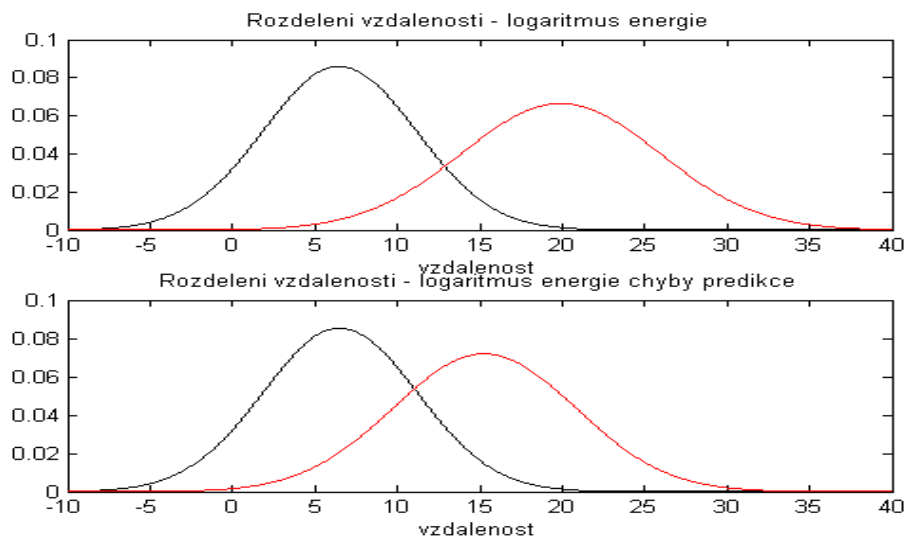
Původně bylo vybráno šestnáct různých parametrizací: koeficienty odrazu - *RC* [4]; první koeficient lineární predikce - *LPC1* [6]; dvanáct koeficientů lineární predikce - *LPC* [6]; logaritmus energie chyby predikce - *Ep* [6]; logaritmus energie signálu - *Ep* [7]; první kepstrální koeficient - *CC1* [9]; druhý kepstrální koeficient - *CC2* [9]; jedenáct kepstrálních koeficientů - *CC* [9]; počet průchodů nulou - *ZCR* [7]; melovské spektrální koeficienty - *MELSPEC* [4], [5]; logaritmované melovské spektrální koeficienty - *FBANK* [4], [5]; melovské; kepstrální koeficienty - *MFCC* [4], [5]; spektrální PLP koeficienty - *SPEC PLP* [8]; (kepstrální) PLP koeficienty - *PLP* [8]; spektrální PLP-RASTA koeficienty - *SPEC PLP-RASTA* [8]; (kepstrální) PLP-RASTA koeficienty - *PLP-RASTA* [8].

Pro vyhodnocování chování a vlivu jednotlivých parametrizací byly měřeny průměrné euklidovské vzdálenosti (jako při výpočtu matice vzdáleností DTW algoritmu) mezi jednotlivými hláskami v dětské řeči. Tyto vzdálenosti byly seříděny od nejkratších po nejdelší a pro jednotlivé parametrizace byly sestaveny skupiny "podobných" hlásek. Tímto způsobem byla např.: pro počet průchodů nulou (*ZCR*) nalezena skupina hlásek c, č, f, ř, s, š, která vykazuje vyšší průměrnou vzdálenost od ostatních. Je to dáno tím, že tyto hlásky vykazují vyšší počet průchodů nulou, než ostatní. Statistické gaussovské rozdělení vzdáleností této skupiny pro *ZCR* je na obrázku 1, kde v horní části je rozdělení počtu průchodů nulou (černý průběh znamená rozdělení nízkého *ZCR* zbytku abecedy a červený je pro uvedenou skupinu, vysoké *ZCR*) a v dolní části je rozdělení vzdáleností (černý průběh je pro vzdálenosti uvnitř obou skupin, pro obě jsou malé a červený je pro vzdálenosti mezi skupinami, velká).



Obrázek 1: Rozdělení počtů průchodů nulou (horní část, černý průběh je pro skupinu hlásek "a, b, d, d', e, g, h, ch, i, j, k, l, m, n, ň, o, p, r, t, t', u, v, z, ž" a červený pro "c, č, f, ř, s, š") a rozdělení průměrných euklidovských vzdáleností (dolní část, černý průběh je pro hodnoty vzdáleností uvnitř obou skupin a červený pro vzdálenosti mezi skupinami)

Na obrázku 2 jsou v horní části rozdělení vzdáleností pro logaritmus energie signálu (tato parametrizace odděluje skupinu c, č, f, k, š, t, t', která má nízkou energii od zbytku abecedy) a v dolní pro logaritmus energie chyby predikce, která odděluje skupinu a, e, i, f, h, l, n, o, ř, s, š, z, ž.



Obrázek 2: Rozdělení průměrných euklidovských vzdáleností pro logaritmus energie (horní část, černý průběh je pro hodnoty vzdáleností uvnitř skupin "c, č, f, k, š, t, ť" a "a, b, d, d', e, g, h, ch, i, j, l, m, n, ň, o, p, r, ř, s, u, v, z, ž" a červený pro vzdálenosti mezi skupinami) a rozdělení vzdáleností pro logaritmus energie chyby predikce (dolní část, černý průběh je pro hodnoty vzdáleností uvnitř skupin "a, l, n, o, f, e, h, i, ř, s, š, z, ž" a "c, č, ch, ť, b, d, d', g, j, k, m, ň, p, r, t, u, v" a červený pro vzdálenosti mezi skupinami)

### 2.3 Optimalizace DTW algoritmu

Aby DTW algoritmus co nejlépe odděloval promluvy nemocných dětí od zdravých, byly provedeny pokusy o jeho optimalizaci.

Jedny z prvních pokusů, které byly prováděny, bylo normování všech promluv jak od zdravých, tak i od nemocných dětí na stejný počet segmentů pomocí vztahů v [9], (7.26-7.29). Dále bylo zkoušeno vyhlazování a určité typy normalizace parametrizovaných hodnot signálu [9] (7.62-7.67). Ukázalo se, že tyto kroky jsou pro naše účely nevhodné, neboť byly určeny pro zlepšení vlastností algoritmu pro účely rozpoznávání slov, ale v naší aplikaci smazávají rozdíly mezi promluvy od zdravých a nemocných dětí, čímž se výsledky zhoršují.

Při realizaci DTW algoritmu je je možno volit různé funkce průchodu maticí vzdáleností [9]. Experimentálně bylo ověřeno, že nejlepší výsledků bylo dosaženo pro směr průchodu definovaný následující rovnicí.

$$g(i, j) = \begin{cases} g(i-1, j) + d(i, j) \\ g(i, j-1) + d(i, j) \\ g(i-1, j-1) + d(i, j) \end{cases} \quad (1)$$

Kde  $i = 1, 2, \dots, I, j = 1, 2, \dots, J$ ,  $I$  a  $J$  jsou počty segmentů první a druhé promluvy, které se porovnávají,  $d(i, j)$  je lokální hodnota euklidovské vzdálenosti pro aktuální indexy  $i$  a  $j$ . Celkovou vzdálenost obou slov udává pak hodnota  $g(I, J)$ , která bývá dělena hodnotou  $(I + J)$  [9], při využití DTW jako rozpoznávače slov.

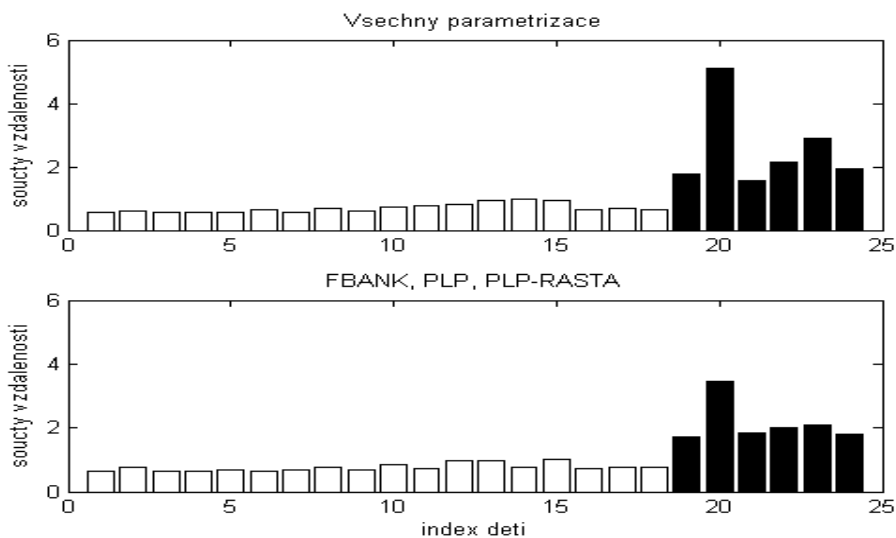
Protože v naprosté většině promluv nemocných dětí dochází jednak k záměně, ale hlavně k prodlužování hlásek i promluv, je vhodné vynechat dělení  $g(I, J)$  hodnotou  $(I + J)$ , čímž selepší odstup promluv nemocných dětí od zdravých, což je další krok při optimalizaci DTW. Problém může nastat tehdy, když nemocné dítě promluvu výrazně zkrátí oproti zdravým (např.: místo slova "materídouška" je řečeno slovo "matouška" a podobně). Zde se potom akumulovaná vzdálenost sníží až k nejvyšším hodnotám zdravých dětí a slovo je obtížně klasifikováno. Daná situace bude dále testována, budou se hledat parametrizace, které jsou na zkrácení slova méně citlivé. Je vhodné počítat celkové akumulované vzdálenosti součtem hodnot z několika různých

vyslovených slov dětí. čímž se zvýrazní odstup nemocných od zdravých a tento problém se potlačí. Dále je vhodné testy provádět pro více slabičná slova, protože některé nemocné děti vyslovují jedno a dvouslabičná slova téměř jako zdravé.

### 3 Experimentální výsledky

DTW algoritmus byl nastaven podle popisu v předchozím odstavci. Jednotlivé parametrizace byly vyhodnocovány pro různá slova a byly prováděny testy, jak dobře oddělují promluvy nemocných dětí od zdravých. Lze říci, že jako nejpřesnější se jeví keprální PLP a PLP-RASTA koeficienty, popřípadě velmi dobře pracují koeficienty FBANK. Tento závěr vyplývá i z pokusů, které naznačují, že ostatní parametrizace řadí promluvy od zdravých dětí k hranici nemocných, pokud některé zdravé dítě má chraptivý a zastřený hlas. Tento závěr odpovídá teoretickým předpokladům, protože PLP a PLP-RASTA koeficienty byly navrženy s ohledem na potlačení rozdílů mezi mluvčími pro zvýšení úspěšnosti rozpoznání řeči.

Ukázka výsledků pro slovo "mateřídouška" je na obrázku 3, kde jsou vyneseny vzdálenosti pro 18 nemocných dětí (bílý průběh) a pro 6 nemocných (černý průběh), horní část obrázku je pro normovaný součet přes všech 16 parametrizací, dolní část je jen pro 3 parametrizace FBANK, PLP a PLP-RASTA koeficienty. Z obrázku je vidět, že dochází skutečně ke smazávání rozdílů mezi zdravými dětmi, algoritmus je méně citlivý na mluvčího.



Obrázek 3: Výsledek klasifikace slova "mateřídouška" pro všech 16 parametrizací (horní část) a pro 3 parametrizace FBANK, PLP a PLP-RASTA (dolní část)

### 4 Závěr

V práci je popsán algoritmus pro vyhodnocování promluv dětí s vývojovou dysfázií, popřípadě afázií s popisem vhodných parametrizací a modifikace DTW algoritmu pro tyto účely. Dosavadní experimenty naznačují možnost využití metody pro objektivní posouzení nemoci a hodnocení jejího vývoje.

Další práce budou zaměřeny na testování algoritmu na rozšířenějších databázích promluv, bude vyhodnocen vliv šumu na úspěšnost klasifikace.

## 5 Poděkování

Práce je podporována granty GA ČR - 102/03/H085 "Modelování biologických a řečových signálů", IGA MZ ČR - NR 8287-3/2005 "Počítačová analýza řečového projevu a celonočních EEG záznamu u dětí" a MŠM6840770012 "Transdisciplinární výzkum v biomedicínském inženýrství 2".

## Reference

- [1] Čmejla R.: *Vyhodnocování výslovnosti s využitím DTW algoritmu*. Výzkumná zpráva LK01/2005, K13131, FEL ČVUT, Praha, srpen, 2005.
- [2] Zlatník P., Čmejla R.: *Application of the DTW algorithm for disordered speech evaluation*. Digital Technologies '05. University of Žilina, 2005, Slovak Republic.
- [3] Rabiner L., Juang P.: *Fundamental of speech recognition*. Prentice Hall, U.S.A., 1984.
- [4] Young S., et al.: *The HTK Book*. Version 3.2.1, Cambridge 2002, England.
- [5] ETSI.: *European Telecommunications Standards Institute* Nov. 2003, ETSI Standard, ETSI ES 202212, Version 1.1.1 France.
- [6] Harrington J., Cassidy S.: *Techniques in speech acoustics*. Kluwer Academic Publishers 1999, Netherlands.
- [7] Deller J. R., Hansen J. H. L., Proakis J. G.: *Discrete-time processing of speech signals*. IEEE Press 2000, U.S.A..
- [8] Hermansky H., Morgan N.: *Rasta processing of speech*. IEEE Transaction on Speech and Audio Processing, Vol. 2, No. 4, pp. 587-589, October 1994, U.S.A.
- [9] Psutka J.: *Komunikace s počítačem mluvenou řečí*. Academia Praha, 1995, printed by CENTA, spol. s. r. o., Veveří 39, Brno.
- [10] Zlatník P., Čmejla R.: *Disordered Speech Evaluation Using the DTW Algorithm*. Analysis of biomedical signals and images, ISSN 1211-412X, Biosignal 2006, Brno.

---

Ing. Petr Zlatník, České vysoké učení technické v Praze, Fakulta elektrotechnická, K13131, Technická 2, 166 27 Praha 6.

Tel.: (+420) 22435 2820, E-mail: zlatnip@fel.cvut.cz

Doc. Ing. Roman Čmejla, CSc, České vysoké učení technické v Praze, Fakulta elektrotechnická, K13131, Technická 2, 166 27 Praha 6.

Tel.: (+420) 22435 2236, E-mail: cmejla@fel.cvut.cz