

Válcový model hlasového traktu

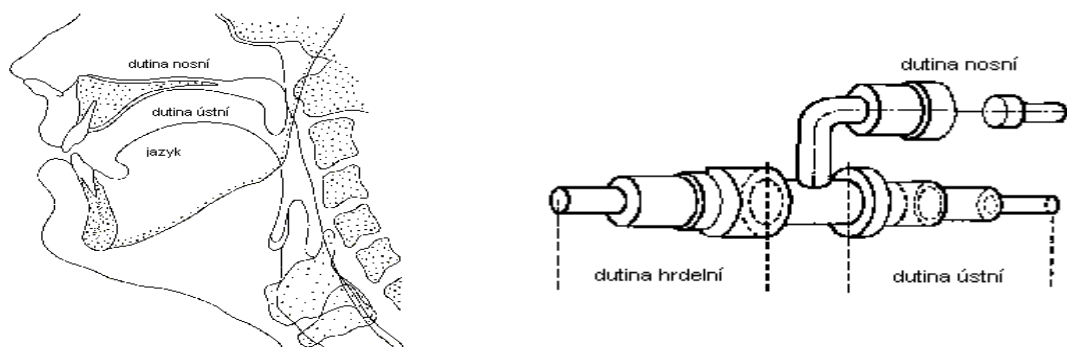
Filip Orság

Abstract

This article shows how to calculate a shape of a human vocal tract and a practical use of the vocal tract model. A simple way to find out the shape of the vocal tract is to use PARCOR (PARTIAL CORrelation) coefficients, which could be obtained during computation of the LPC (Linear Prediction Coefficients). The curve, which represents the vocal tract shape, could be used to recognize speech or speaker. It could be useful also for teachers and students of the foreign languages to improve their articulation or for the deaf people as a tool of the articulation training.

Generace řeči

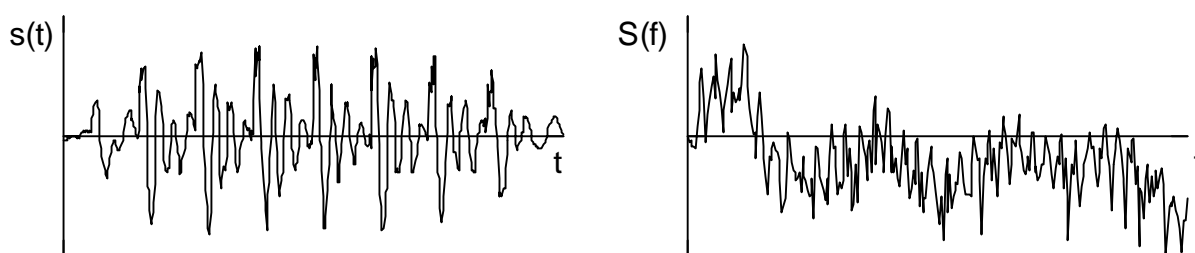
Pochopení vzniku řeči je důležitou součástí při rozpoznání obsahu řeči nebo mluvího. Řeč je generována následujícím způsobem: vibrací hlasivek, jež je způsobena proudem vzduchu vycházejícího z plic, je produkován signál se základní frekvencí f_0 a vyššími harmonickými frekvencemi, nebo některý z několika druhů šumu. Frekvence f_0 reprezentuje základní tón řeči (tj. výšku hlasu). Takto vzniklý signál prochází artikulačním traktem člověka a je modifikován v závislosti na tvaru traktu. Artikulační trakt lze modelovat jako dutou trubici [1], [4] s měnícím se průřezem, nebo také jako skupinu válců jdoucích za sebou s různým průměrem (obr. 1).



Obr. 1 Průřez artikulačním traktem člověka a úplný válcový model hlasového traktu.

Výpočet válcového modelu hlasového traktu

Délka artikulačního traktu bývá u dospělého člověka asi 15 – 19 cm. Zvuková vlna generovaná hlasivkami je ovlivňována tvarem (konfigurací) artikulačního traktu, v němž jednotlivá místa (především rozsáhlejší dutiny jako např. dutina ústní) působí jako dutinové rezonátory. Jejich rezonanční kmitočty se nazývají formanty a lze je vidět na spektru takto vygenerovaného signálu (obr. 2). Formanty jsou nejlépe patrné na spektrech samohlásek, dále se tedy budeme zabývat především samohláskami.



Obr. 2 Samohláska *a* v detailu, časový průběh a odpovídající kmitočtové spektrum.

U každé samohlásky lze nalézt určité uspořádání formantů. Pro rozpoznání obsahu řeči jsou významné první tři formanty, které jsou energeticky nejbohatší a přibližně odpovídají těmto dutinám: první formant – dutina hrdelní, druhý formant – dutina ústní a třetí formant – dutina nosní. Jelikož vznik formantů je vázán na konfiguraci artikulačního traktu, lze z jeho tvaru zjistit, jaká samohláska byla vyřčena. Tvar artikulačního traktu lze zjistit pomocí koeficientů PARCOR (PARTIAL CORrelation), jež jsou meziproduktem při výpočtu LPC (Linear Prediction Coefficients). Při určování tvaru artikulačního traktu se zanedbává vliv dutiny nosní. K výpočtu LPC lze využít následující soustavu lineárních rovnic [1]

$$\begin{aligned} a_1 R_0 + a_2 R_1 + \dots + a_M R_{M-1} &= R_1 \\ a_1 R_1 + a_2 R_0 + \dots + a_M R_{M-2} &= R_2 \\ \vdots & \\ a_1 R_{M-1} + a_2 R_{M-2} + \dots + a_M R_0 &= R_M \end{aligned} \quad (1)$$

kde M je řád predikce, koeficienty a_i jsou LP koeficienty a R_k jsou autokorelační koeficienty, jejichž hodnotu lze vypočítat z následujícího vztahu

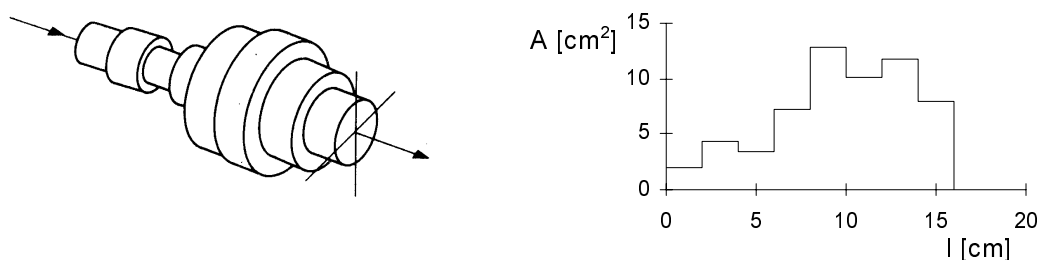
$$R_k = \sum_{n=0}^{N-k-1} s(n)s(n+k), \quad (2)$$

kde N je délka analyzovaného signálu, k je řád autokorelace a $s(n)$ je analyzovaný signál. K řešení této soustavy lze využít rekurzivní algoritmus (např. Levinson-Durbinův). Během

tohoto algoritmu je možné získat koeficienty PARCOR, pomocí nichž lze namodelovat tvar hlasového traktu, tak že pomocí následujícího vztahu vypočteme relativní obsah plochy průřezu traktu [2]

$$\frac{A_{m+1}}{A_m} = \frac{1 - k_m}{1 + k_m}, \quad (3)$$

k_m jsou koeficienty PARCOR, A_m je m -tý obsah průřezu a $m = 1, 2, \dots, M$, kde M je řád predikce – tedy počet získaných koeficientů. Jelikož je pro výpočet nutné znát i plochu první a poslední (tj. A_0 a A_{M+1}), platí: $A_0 = \infty$ a $A_{M+1} = 1$. Tyto hodnoty jsou však pouze relativní. Absolutní hodnotu (tj. skutečný obsah průřezu) lze vypočítat pomocí základní frekvence hlasu a znalosti závislosti velikosti průřezu hlasivkové štěrbiny a pohlaví a věku mluvčího. Příklad vypočtených hodnot průřezu je na obr. 3 [3].

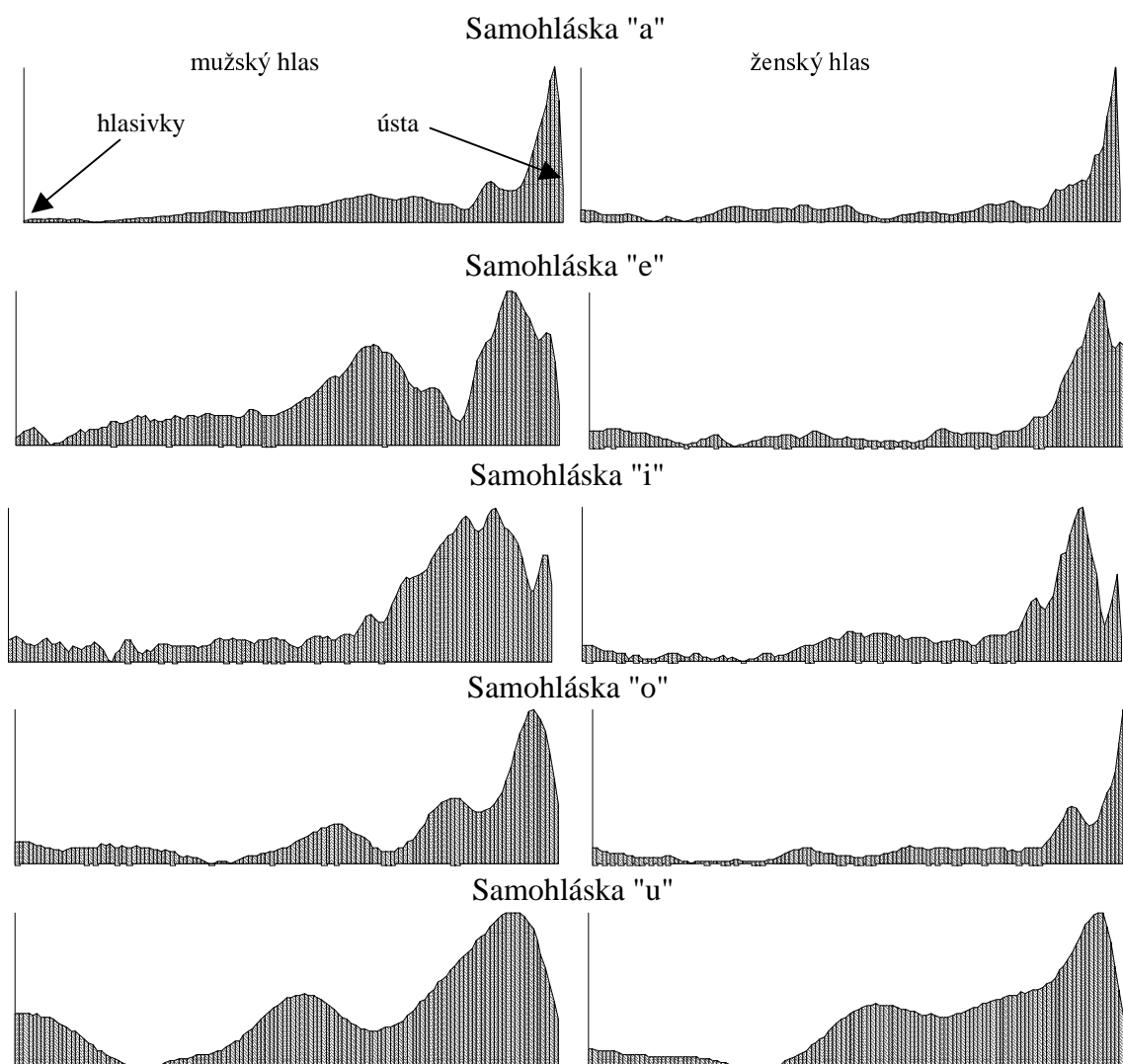


Obr. 3 Zjednodušený válcový model hlasového traktu a vypočítané průřezy traktem pro samohlásku *a*.

Získané výsledky

Na obr. 4 jsou vypočítané průřezy artikulačním traktem muže a ženy pro české samohlásky. Počátek souřadné soustavy odpovídá hlasivkové štěrbině a konec odpovídá ústům. Levý průběh je pro mužský hlas, pravý pro ženský. Při srovnání průřezů je velmi dobře patrný rozdíl mezi mužským traktem a ženským traktem (na něm jsou patrné menší průřezy, tj. menší velikost dutin, což mj. způsobuje vyšší tón ženského hlasu), ale pro rozpoznání pohlaví existují jiné metody. Mnohem užitečnější je tvar traktu, který jednoznačně rozlišuje samohlásky (např. specifické zakulacené průběhy samohlásek „u“ a „o“, nebo ojedinělý průběh pro samohlásku „a“). Při použití válcového modelu hlasového traktu pro účely rozpoznání mluvčího, hraje roli statistika a dlouhodobé průřezy hlasovým traktem, které jsou pro každého mluvčího charakteristické. Je také možné vysledovat odchylky a zvláštnosti při vyslovení některé z hlásek a pomocí těchto unikátních vlastností rozpoznat mluvčího. Samostatné použití válcového modelu pro identifikaci mluvčího však

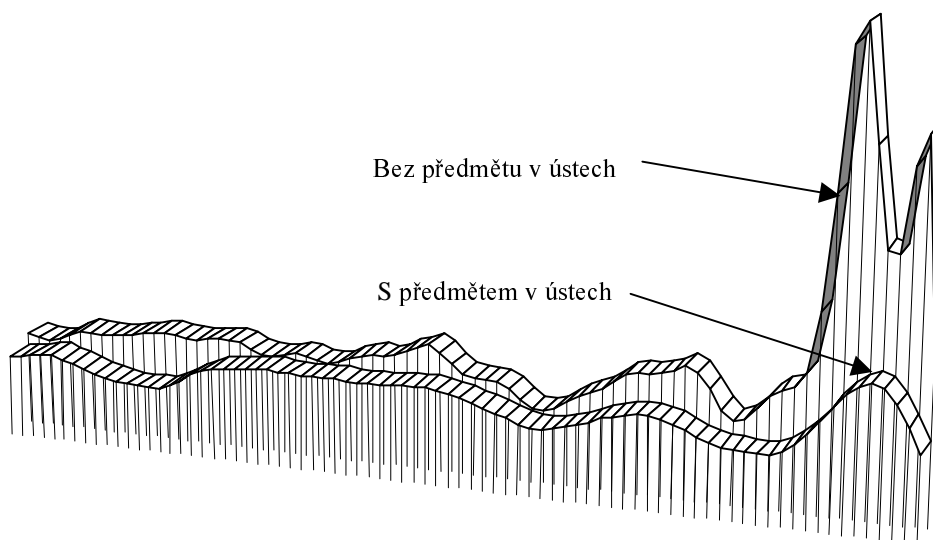
není možné, neboť model neposkytuje dostatečné informace pro kvalitní separaci mluvčích.



Obr. 4 Vypočítané průřezy hlasovým traktem při vyslovení českých samohlásek.

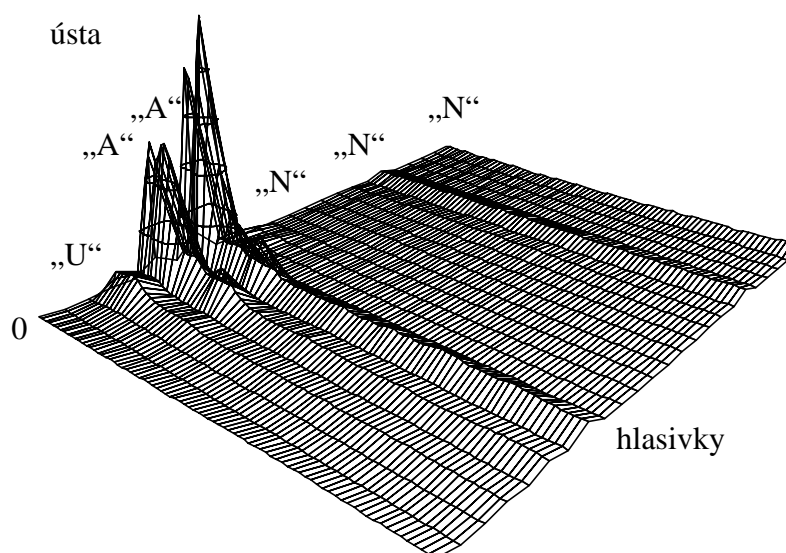
Praktické využití této metody spočívá například v lékařské analýze hlasového traktu. Na obr. 5 jsou znázorněny 2 průběhy – průřez hlasovým traktem při vyslovení hlásky „a“ mužským mluvčím a průřez při vyslovení téže hlásky tímž mluvčím, ale s předmětem umístěným na jazyku. Z průběhů je patrné, že přítomnost předmětu ovlivnila signál, a toto ovlivnění se projevilo i na modelu hlasového traktu.

Další možností využití je při výuce jazyků. Rodilý mluvčí vysloví slovo a parametry promluvy jsou uloženy ve tvaru časové změny průřezu hlasovým traktem. Příklad takového průběhu je na obr. 6. Student poté může vyřknout stejné slovo, nechat je



Obr. 5 Průřezy hlasovým traktem při vyslovení hlásky „a“ s předmětem v ústech a bez předmětu v ústech.

analyzovat, poté porovnat s průběhem traktu rodilého mluvčího a snažit se, co nejdříveji, tento průběh napodobit.



Obr. 6 Časový průběh tvaru hlasového traktu při vyslovení anglického slova „one“.

Závěr

V příspěvku je uveden způsob výpočtu geometrických rozměrů zjednodušeného modelu hlasového traktu a uvedeny výsledky pro české samohlásky. Získané hodnoty lze použít jako biometrické parametry mluvčího. Vizualizaci této metody lze uplatnit také při výuce cizích jazyků a jako pomůcku pro výuku sluchově postižených osob, které takto mohou vizuálně kontrolovat, zda správně artikulují.

This work was supported by the Ministry of Education of Czech Republic - Project No.VS97060.

Literatura

- [1] Psutka, J.: Komunikace s počítačem mluvenou řečí. Academia, Praha 1995.
- [2] Juang,B.H.-Rabiner,L.R.: Fundamentals of Speech Recognition. Englewood Cliffs, NJ, Prentice Hall, 1993.
- [3] Sigmund,M.: Speaker Recognition - Identifying People by their Voices. Habilitační práce FEI VUT Brno, 1999.
- [4] <http://www.ims.uni-stuttgart.de/phonetik/EGG>