USE ERROR LPC FOR HUMANITY TTS SYSTEM

Milan VOJTĚCH, Master Degree Programme (5)
Dept. of Radio Electronics, FEEC, BUT
E-mail: milvojt@email.cz

Supervised by: Ing. Pavel Matějka

ABSTRACT

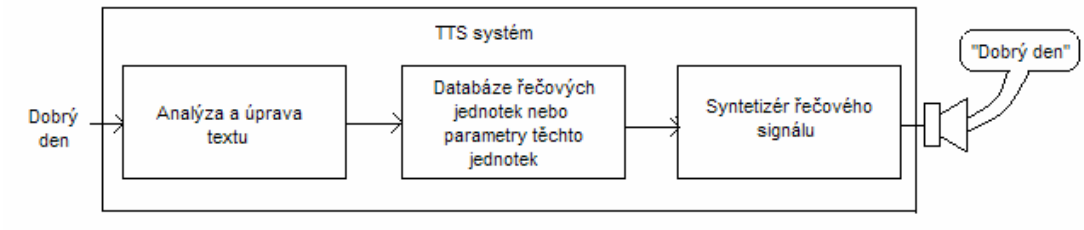
This article discuss how to get Text-To-Speech system with better humanity output voice, if we get parameters for different models of excitation pulses for produced human speech.

1 ÚVOD

TTS tato zkratka pocházející z anglického termínu Text-To-Speech, označuje systémy umožňující konverzi textu na řeč. Tyto systémy umožňují nevidomým spoluobčanům zpřístupňovat elektronická skripta či knihy. Díky TTS systémům se mohou nevidomý zapojit do normálního života. Ovšem ani tyto důmyslné systémy nepřekonají lhostejnost, některých vidících jedinců.

2 SEZNÁMENÍ S TTS

Vstupem do TTS systému je textová informace a výstupem je řečový signál. Abychom získali ze vstupního textu výsledný řečový signál, musí TTS systém obsahovat funkční bloky zakreslené na Obr.1, v prvním bloku je provedena fonetická transkripce (každému grafému je přiřazen odpovídající foném) a například se zde může provádět substituce typu: Dr. -> doktor,



Obr. 1: *Blokové schéma TTS systému*

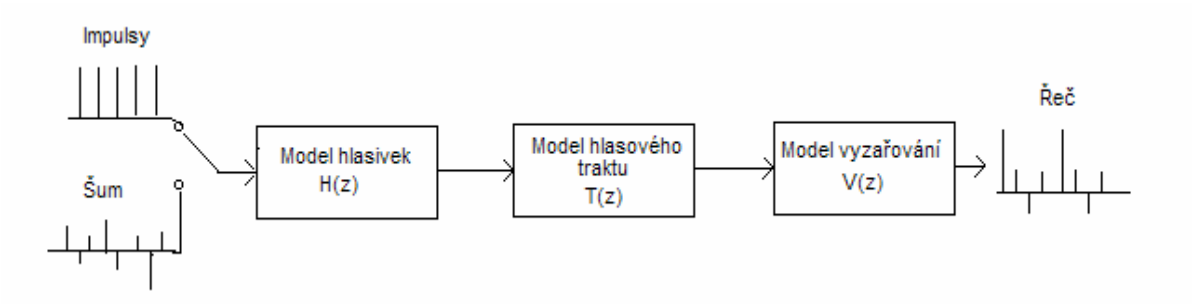
v dalším bloku pak rozhodneme, jak budeme provádět samotnou syntézu řeči a zvolíme řečové jednotky pro syntézu, poslední blok syntetizuje řečový signál.

3 SYNTÉZA ŘEČI

K syntéze řečového signálu můžeme přistupovat v časové oblasti nebo ve frekvenční oblasti. Já se zaměřuji na syntézu ve frekvenční oblasti, neboť ta umožňuje snadněji pracovat s prozodii, neboli větnou melodií, jejíž hlavní parametry jsou základní tón řeči, hlasitost, tempo a ráz.

3.1 SYNTÉZA ŘEČI VE FREKVENČNÍ OBLASTI

Syntetizér obsažený v TTS systému využívá diskrétní model produkce řeči člověka. Tento model využívá pro syntézu znělých úseků řeči generátor impulsů a pro syntézu neznělých úseků řeči generátor bílého šumu. Impulsy se mohou opakovat s libovolnou frekvencí, tím máme možnost modelovat prozodii věty a tak uživateli TTS systému zpříjemnit poslech. Elektronický model produkce řeči je na Obr.2.



Obr. 2: *Elektronický model produkce řeči*

Model hlasového traktu T(z) popisuje rezonance v dutině hrdelní, ústní a nosní. Tyto rezonance se projeví v kmitočtovém spektru hlásek jako výrazné laloky – *formanty*. Model hlasového traktu je realizován jako IIR filtr libovolného řádu, koeficienty tohoto filtru mohou být například tvořeny predikčnímy koeficienty a_i získanými lineární predikční analýzou (LPC) úseků řeči nebo mohou být tvořeny kepstrálnímy koeficienty.

Pokud použijeme LPC model hlasového traktu [1] s přenosovou funkci (1), do pouštíme se jisté chyby, protože neznáme u(k) v rovnici (2). Tato chyba je způsobena neznalostí vlastností budícího signálu u(k) při odvození T(z), buzení pouze impulsy nebo šumovým signálem je nedostatečné. Výsledná řeč je sice srozumitelná, ale je nepřirozená. Proto jsem se začal zabývat možností, jak tuto chybu odstranit. Samozřejmě je možné použít jiný přístup k syntéze řeči, ale LPC je výpočetně nenáročná metoda a tudíž ji lze využít pro aplikace s mikroprocesory nižší cenové kategorie.

$$T(z) = \frac{1}{1 + \sum_{i=1}^{M} a_i z^{-i}} \tag{1}$$

$$s(k) = -\sum_{i=1}^{M} a_i s(k-i) + u(k) \tag{2}$$

4 ZÍSKÁNÍ BUDÍCÍCH IMPULSŮ

Předpokládáme - li, že řečový signál je zpracováván lineárním prediktorem řádu M

$$\hat{s}(k) = -\sum_{i=1}^{M} \alpha_i s(k-i) \tag{3}$$

Vznikne chyba lineární predikce *e(k)* definována rovnicí (4) jako rozdíl originálního a predikovaného signálu [1]

$$e(k) = s(k) - \hat{s}\ (k) = s(k) + \sum_{i=1}^{M} \alpha_i s(k-i) \tag{4}$$

Z porovnání (2) a (4) vyplývá, jestliže $\alpha_i = a_i$ TTS, pak chyba predikce **e(k)=u(k)** obsahuje důležité informace o budícím signálu. Lze tedy získat časový průběh budících impulsů pomocí filtrace originálního signálu filtrem s inverzní přenosovou funkcí ve tvaru

$$T^{-1}(z) = 1 + \sum_{i=1}^{M} a_i z^{-i} \tag{5}$$

Nyní stačí parametrizovat časový průběh budících impulsů a využít tyto parametry v syntetizéru. K tomu můžeme použít Liljencrants-Fantova (6) modelu [2], který popisuje derivaci časového průběhu budícího signálu $v_g(t)$

$$v_g(t) = \begin{cases} \dfrac{-E_e}{\sin\left[\varpi_g\left(T_e - T_{op}\right)\right]} e^{\alpha(n-T_e)} \sin\left[\varpi_g\left(n - T_{op}\right)\right] & T_{op} \le n \le T_e \\ \dfrac{-E_e}{\varepsilon T_a}\left[e^{\varepsilon(T_e-n)} - e^{\varepsilon(T_e-T_c)}\right] & T_e < n < T_c \\ 0 & jinak \end{cases}$$

$$T_a = \frac{\left[1 - e^{\varepsilon(T_e - T_c)}\right]}{\varepsilon} \tag{6}$$

5 ZÁVĚR

Doposud byl implementován v prostředí Matlab systém využívající LPC model hlasového traktu pro syntézu řeči. Popisovaná možnost parametrizace budícího signálu, by měla zlepšit kvalitu syntetizované řeči a tím usnadnit poslech syntetizované řeči nevidomým spoluobčanům. Důležitým poznatek je to, že získaný budící signál obsahuje informace o psychickém stavu mluvčího nebo zda je ve stresu. Tím by se syntetizovaná řeč stala „lidštější“.

LITERATURA

[1] Psutka, J.: Komunikace s počítačem mluvenou řečí, Academia, 1995

[2] Boštík, M.: Analýza hlasu pro diagnostické účely, diplomová práce, 2002