

VÝPOČET KOEFICIENTŮ MODIFIKACE PRO NELINEÁRNÍ METODU WSOLA

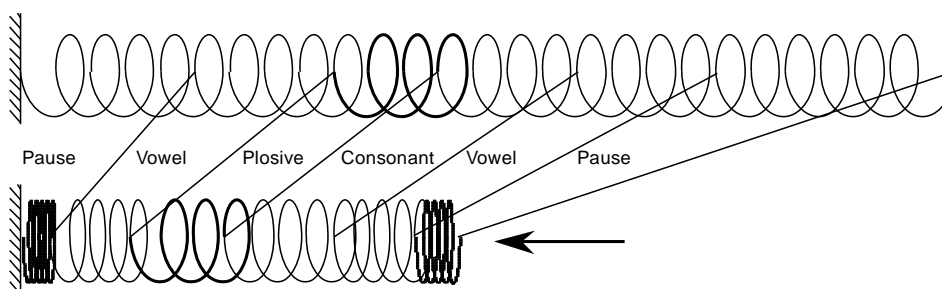
Martin MÉNER¹

1 ÚVOD

Tato práce se zabývá návrhem metody, která by byla schopna řešit problém určování koeficientů modifikace pro metodu WSOLA (Waveform Similarity Over-Lap Add), resp. pro její nelineární verzi. Využití této metody je úzce spjato s TTS (Text-To-Speech) systémem ARTIC (Artificial Talker In Czech). Tento systém generuje na základě textu syntetizovanou řeč. Jeho principem je vhodné skládání skupin fonémů z předem vytvořeného korpusu tak, aby byla co nejpřirozeněji vytvořena požadovaná zvuková stopa. Proto je tempo syntetizované řeči dáno tempem zmíněného korpusu, což může v praxi limitovat použitelnost TTS systému. Např. při vytváření audio titulků k cizojazyčným filmům je potřeba tempo řeči měnit, zrakově postižení lidé zase vyžadují velmi rychlé tempo řeči, neboť jsou na takový příjem informací zvyklí. Na druhou stranu je vždy žádoucí zachovat charakter řeči přirozený. Jedním z mnoha způsobů, jak toho docílit, tkví právě v algoritmu WSOLA. Technika WSOLA je často implementována lineárně. To znamená, že každý foném je modifikován stejným faktorem. Výhodou takového přístupu je jednoduchost. Naproti tomu může ovšem docházet ke ztrátě kvality zvukového signálu i pro relativně malé změny tempa. Proto je upřednostňována nelineární metoda WSOLA, která umožňuje modifikovat teoreticky každý foném jiným koeficientem. Způsob, jak tyto koeficienty obecně určit tak, aby byla dosažena požadovaná modifikace, je konkrétním cílem této práce.

2 ZOObECNĚNÁ NELINEÁRNÍ MODIFIKACE

Aby byla myšlenka nelineární modifikace formulována naprosto obecně, bude zaveden pojem segment. Ten bude označovat skupinu fonémů, která se bude modifikovat lineárně. Určitou analogii k těmto segmentům tvoří pružiny. Stejně jako má každá pružina svoji tuhost, bude mít každý tento segment definovanou konstantu „modifikovatelnosti/tuhosti“. Rečový signál, složený ze segmentů, si pak lze představit jako soustavu pružin s různě definovanými tuhostmi. Pokud na jeden konec pružiny bude působit síla, dojde k nelineárnímu zkrácení, jak je ilustrováno na obr. 1.



Obr. 1: Představa nelineární modifikace

¹Martin Méner, student navazujícího studijního programu Aplikované vědy a informatiky, obor Kybernetika a řídicí technika, e-mail: mmener@students.zcu.cz

Matematicky lze problém formulovat následujícím způsobem. Necht' k_i je tuhost i -tého segmentu, x_i je jeho přirozená délka, f_i je neznámý koeficient a nová délka segmentu je $\hat{x}_i = f_i x_i$. Pak lze podle teorie pružnosti nadefinovat tuhost N pružin/segmentů jako

$$\frac{1}{k} = \sum_i^N \frac{1}{k_i} \quad (1)$$

Dále lze podle Hookova zákona vyčíslit celkovou sílu F nutnou pro vychýlení celé soustavy pružin s tuhostí k o $\Delta x = x - \hat{x}$. Tuto sílu lze pak vydělit jednotlivými koeficienty k_i a získat tak vychýlení jednotlivých pružin/segmentů (2), ze kterého se zpětně určí \hat{x}_i .

$$\Delta x_i = \frac{F}{k_i} \quad (2)$$

Poté se již použije algoritmus WSOLA a modifikuje každý segment faktorem $f_i = \frac{\hat{x}_i}{x_i}$.

Protože ale deformace pružiny nezávisí na její délce, může vyjít faktor f_i záporný. Proto byl vztah (1) upraven na

$$\frac{1}{k} = \sum_i^N \frac{1}{\frac{x k_i}{x_i}} \quad (3)$$

Přístup podle (3) přináší dobré výsledky pouze při malých modifikacích, a tak byl naprogramován iterativní algoritmus, který dodrží stanovený limit zkrácení pro každý segment. Algoritmus v prvním kroku vypočte faktory modifikace všech segmentů podle (2) a (3). V dalším kroku se zkontroluje, zda žádný faktor nepřekročil stanovenou mez. Pokud ne, jsou určeny všechny faktory správně a lze přistoupit k WSOLA. Pokud je nalezen záporný faktor, je předefinován na svou mezní hodnotu a dojde k přepočtu nové původní a celkové délky segmentů a opakuje se první krok.

3 ZÁVĚR

Výsledkem této práce je implementace algoritmu, který je schopen obecně určovat koeficienty modifikace pro nelineární metodu WSOLA. Na základě původní délky řečového signálu, velikosti všech přítomných segmentů, jejich „tuhosti“ a limitů modifikace, je tak exaktně vypočten faktor modifikace každého z nich.

Poděkování: Příspěvek byl podpořen grantovým projektem SGS-2010-054.

LITERATURA

Demol M., Verhelst W., Struyve K., Verhoeve P., “Efficient Non-Uniform Time-Scaling of Speech with WSOLA”. *Proceedings of the Speech and Computers 2005 (SPECOM-2005)*, pp. 163–166, Patras, Greece, 2005.

Hanzlíček Z., Matoušek J., Tihelka D., “Towards Automatic Track Generation for Czech TV Broadcasting: Initial Experiments with Subtitles-to-Speech Synthesis”. *Proceedings of the 9th International Conference on Signal Processing, ICSP'08*, vol. 3, pp. 2721-2724, IEEE Press, Beijing, China, 2008.

Méner, M., Tihelka, D., 2011. Generalized Non-Uniform time Scaling distribution Method for Natural-Sounding Speech Rate Change