

## Využití grafických karet pro akceleraci parametrizace řeči

Josef Michálek<sup>1</sup>

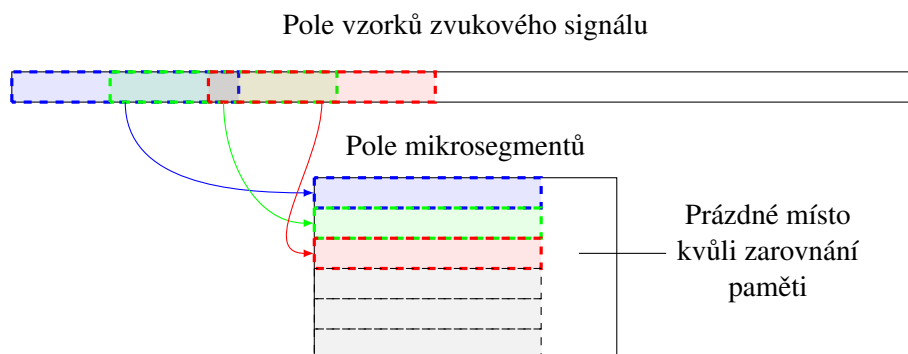
### 1 Úvod

Postupem času se vyvíjí stále nové metody zpracování řeči. Jejich zdokonalování probíhá současně se zdokonalováním technologie a je vhodné pro zpracování řeči používat nejmodernější prostředky. Jedním z nejrychleji se vyvíjejících dílů domácích počítačů jsou grafické karty, které jsou schopny paralelně vykonávat velké množství operací. Narozdíl od procesoru mohou na grafické kartě běžet najednou stovky vláken a přepínání mezi nimi je téměř okamžité. Výpočetní výkon na jedno vlákno je velmi malý, ale celkový výkon všech vláken je několikanásobně vyšší. Vhodně navržený algoritmus je schopný tento výkon využít a zpracovat velké množství dat rychleji než procesor. V této práci se zabývám úpravou algoritmů parametrizace řeči pro použití na grafických kartách.

### 2 Parametrizace řeči

Parametrizace řeči je proces, kterým se pro každý úsek řečového signálu získá vektor příznaků popisující tento úsek řeči. V práci se zabývám metodami MFCC (Mel-frequency cepstral coefficients), PLP (Perceptual linear predictive analysis) a TRAPS (Temporal patterns). Tyto metody mají několik společných kroků:

- Rozdělení vstupního signálu na segmenty pevné délky
- Fourierova transformace
- Filtrace bankou filtrů



**Obrázek 1:** Znárodnění segmentace

<sup>1</sup> student doktorského studijního programu Aplikované vědy a informatika, obor Kybernetika, specializace Umělá inteligence, e-mail: orcus@kky.zcu.cz

Poté každá metoda se segmenty pracuje rozdílně, ale důležitá je jedna společná vlastnost: segmenty se zpracovávají nezávisle na sobě. To umožňuje algoritmus výpočtu paralelizovat a proto jsou grafické karty pro tuto úlohu velmi vhodné.

Nejdříve bylo třeba přesunout vstupní data do vhodného uspořádání, viz obr. 1. Poté byly jednotlivé kroky algoritmů analyzovány a vyladěny pro optimální dobu výpočtu. Výsledná implementace byla nejdříve vytvořena pro platformu CUDA a poté přepsána pro OpenCL.

### 3 Výsledky

Výsledkem práce je program pro operační systémy Windows i GNU/Linux dostupný na adrese <http://sourceforge.net/projects/acclfeatextr/>. Program umožňuje parametrizovat zvukové soubory pomocí výše uvedených metod na platformách CUDA, OpenCL a také na procesoru. V tabulce 1 jsou časy zpracování 4 různých korpusů pro metodu MFCC. Každý obsahuje soubory určité délky a vzorkovací frekvence, celková délka každého korpusu je 10 hodin. Porovnání bylo provedeno s programy HTK a OpenSMILE, výsledky této práce jsou ve sloupci nadepsaném Afet.

Z tabulky je vidět, že pro všechny případy bylo dosaženo velkého urychlení. V nejlepším případě bylo celých 10 hodin dat zpracováno dokonce za 6,1 s, to odpovídá 0.00017 RTF (Real time factor). Také bylo ověřeno, že pro malé soubory může SSD disk zpracování několikanásobně urychlit. Zpracování dat je oproti HTK v nejlepším případě až 20krát rychlejší, porovnání s OpenSMILE vychází ještě lépe. Z důsledné analýzy programu bylo zjištěno že pro metodu MFCC připadá z celkové doby zpracování pouze 6,8 % času na samotný algoritmus parametrizace, zbytek času jsou převážně vstupněvýstupní operace. Další zrychlení algoritmů by se proto na výsledku už příliš neprojevalo. Pro další 2 implementované metody vychází výsledky srovnatelně.

Vstup	HDD → HDD			SSD → SSD			SSD → HDD			HDD → SSD		
	HTK	oS	Afet	HTK	oS	Afet	HTK	oS	Afet	HTK	oS	Afet
Krátký 8 kHz	143.2	347.3	<b>37.5</b>	109.2	346.4	<b>28.6</b>	61.1	346.5	<b>33.9</b>	137.1	344.6	<b>29.7</b>
Dlouhý 8 kHz	109.3	282.9	<b>6.1</b>	88.6	281.9	<b>6.0</b>	88.6	283.4	<b>6.5</b>	97.5	283.2	<b>6.1</b>
Krátký 44 kHz	364.6	998.2	<b>172.3</b>	273.8	1002.5	<b>57.1</b>	275.3	997.1	<b>63.1</b>	336.7	1001.0	<b>109.1</b>
Dlouhý 44 kHz	272.2	875.7	<b>46.8</b>	252.6	875.1	<b>28.9</b>	254.2	874.2	<b>28.4</b>	266.6	873.9	<b>37.9</b>

**Tabulka 1:** Doba zpracování souborů metodou MFCC v sekundách

### Poděkování

Tento příspěvek byl podpořen grantovým projektem SGS-2013-032 (Inteligentní metody strojového vnímání a porozumění 2).

### Literatura

- Prof. Ing. Psutka J., CSc., Doc. Ing. Matoušek J., Ph.D., Doc. Ing. Müller L., Ph.D., Doc. Dr. Ing. Radová V., 2006. *Mluvíme s počítačem česky*. Praha, Academia
- Schwarz P., 2008. *Phoneme recognition based on long temporal context*, disertační práce. Brno, Vysoké Učení Technické v Brně
- Steve J. Young, D. Kershaw, J. Odell, D. Ollason, V. Valtchev, P. Woodland, 2006. *The HTK Book Version 3.4*. Cambridge University Press