



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ

FACULTY OF INFORMATION TECHNOLOGY

ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ

DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

KODEK DO 16 KBIT/S

CODEC 16 KBIT/S

ZRE: PROJEKTOVÁ DOKUMENTACE

ZRE: PROJECT DOCUMENTATION

AUTOR PRÁCE

AUTHOR

Bc. DAVID KOLEČKÁŘ

BRNO 2018

Úvod

Hlavním cílem projektu je implementovat řečový kodek (tedy sekvenci kodér – dekodér) dosahující srozumitelnou řeč na bitové rychlosti nižší nebo rovno 16 kbit/s.

Popis řešení

Kodek využívá lineární prediktivní kódování řeči. Tedy ze signálu jsou získávány jeho charakteristické parametry, které nesou informaci potřebnou pro dekódování. Vychází z toho, že následující vzorky signálu se dají určit ze vzorků předchozích dle lineární závislosti. [2] Vstupní signál se vzorkovací frekvencí 8000 Hz je rozdělen na rámce (délka 25 ms), pro které je provedena LPC analýza i odhad základního tónu. Jelikož je výpočet LPC časově náročný, trvá kódování delšího signálu delší dobu. Do výstupního souboru ve formátu COD je na začátek souboru uloženo počet rámců poté vektor se zpožděním, gain vektor a nakonec matice s koeficienty prediktorů. [1] Kodér i dekodér vypisují na standartní chybový výstup počet rámců a vzorků signálu.

Překlad

Kodek je implementován v jazyce Python. V kořenovém adresáři lze najít dva skripty *zre_code* a *zre_decode*, které obsahují zdrojový kód. Pro spuštění skriptu je nutné mít nainstalovaný v systému Python. Kodek byl testován na školním serveru merlin, který obsahuje Python ve verzi 3. Testování probíhalo na testovacích datech, které lze nalézt ve složce *test_data*. Pro delší zvukovou stopu, kodér zpracovává vstup delší dobu. Což je zapříčiněno výpočtem LPC koeficientů. Příklad spuštění z příkazové řádky:

1. `python3 zre_code infile.raw outfile.cod`

- vstupem je soubor s řečí RAW 16 bitový znaménkový typ short s vzorkovací frekvencí 8000 Hz
- výstupem je binární soubor .cod s kódovanou řečí

2. `python3 zre_decode outfile.cod outfile.raw`

- vstupem je binární zakódovaný soubor .cod
- výstupem je soubor s řečí RAW 16 bitový znaménkový typ short s vzorkovací frekvencí 8000 Hz

Název souboru	Velikost bez kódování	Velikost po kódování	Délka stopy
test.raw	21 KB	3 KB	2 sekundy
long.raw	1099 KB	132 KB	120 sekund
train.raw	321 KB	39 KB	20 sekund

Tabulka 1: Informace o testovacích datech

Závěr

Projekt se podařilo naimplementovat dle zadání. Kodér produkuje binární soubory, které mají minimálně 8-krát menší velikost než zdrojový soubor ve formátu RAW. Po dekodování lze znovu přehrát srozumitelnou řeč. Testování probíhalo poslechem dekodované nahrávky a taky doporučenou funkcí pro výpočet logaritmického spektrálního zkreslení pro 2 signály (originálního a dekodovaného). Skripty obsahují jen základní chybové hlášky a nejsou tak ošetřeny všechny možné chybové stavy.

Literatura

- [1] Černocký, J.: *Základní tón, kódování a dekódování řeči*. 2018, [Online; navštíveno 16.04.2018].
URL https://www.fit.vutbr.cz/study/courses/ZRE/public/labs/03_pitch_codec/03_pitch_codec.pdf
- [2] Ozun, O.; Steurer, P.; Thell, D.: *Wideband Speech Coding with Linear Predictive Coding (LPC)*. 2002, [Online; navštíveno 16.04.2018].
URL <http://www.seas.ucla.edu/spapl/projects/ee214aW2002/1/report.html>