

ZRE Projekt 2022 — Spektra a RELP

Honza Černocký, FIT VUT Brno

March 28, 2022

Úkolem projektu je pocvičit se ve výpočtu spekter pomocí FFT, Mel-banky filtrů a LPC, a zkusit napsat jednoduchý RELP (Residual-Excited Linear Prediction) kodek. Můžete pracovat v čemkoliv včetně Matlabu a jeho klonů (Octave, atd.), ale asi budete preferovat Python. Můžete použít jakékoliv funkce ze cvičení či odjinud. V případě použití cizích funkcí bych byl ale rád, kdybyste věděli, **co dělají**. Kdekoliv je tedy “Napište” můžete nahradit za “Napište nebo sežeňte s porozuměním” ;)

Organisace projektů

Projekty se řeší individuálně.

Odevzdáváte

1. Protokol, kde ke každé úloze krátce popíšete, co jste dělali — neopisujte prosím teorii, vložte obrázky a odpovídejte na otázky, kód vkládejte jen pokud je nezbytně nutný (budeme ho mít v archívu).
2. TGZ archiv s Vaším kódem a signály po kódování a dekodování.

Zadání

Úloha 1 - základy - 2 body

Načtěte řečový signál z <https://www.fit.vutbr.cz/study/courses/ZRE/public/2021-22/a.wav> (použijte vhodnou knihovnu ke čtení WAVek). Odečtěte střední hodnotu (tím se zbavíte stejnosměrné složky). Rozdělte jej na rámce o délce 20 ms s nulovým překrytím (budou tedy těsně navazovat jeden na druhý). Vyberte pěkný znělý rámec, zobrazte ho a ručně určete periodu základního tónu ve vzorcích a v sekundách.

Úloha 2 - spektrální hustota výkonu pomocí FFT - 2 body

Napište funkci pro odhad spektrální hustoty výkonu (power spectral density, PSD) pomocí FFT¹. Výstup nechť je 129 čísel od 0 do poloviny vzorkovací frekvence (FFT tedy musíte počítat na 256 vzorcích - pokud je jich v rámci méně, použijte doplnění nulami). Zobrazte logaritmus PSD vybraného rámce se slušnou frekvenční osou v Hz.

Úloha 3 - Odhad sady LPC filtrů - 2 body

Napište funkci pro odhad LPC filtru reprezentujícího daný rámec pomocí algoritmu Levinsona Durbin. Zajímají nás **všechny** filtry od řádu $P = 1$ až po $P = 20$. Nezapomeňte pro každý filtr uložit i gain.

Úloha 4 - analýza sady LPC filtrů - 2 body

Proveďte LPC analýzu vybraného znělého rámce pomocí všech LPC filtrů. Pro každý vygenerujte chybový signál lineární predikce $e[n]$. Spočítejte jeho energii. Ukažte závislost energie chybového signálu na řádu prediktoru. Ověřte že $P = 10$ je “rozumný” řád prediktoru.

Úloha 5 - spektrální hustota výkonu pomocí LPC - 2 body

Napište funkci pro odhad spektrální hustoty výkonu (power spectral density, PSD) pomocí LPC filtru a jeho gainu. Výstup nechť je opět 129 čísel od 0 do poloviny vzorkovací frekvence. Zobrazte logaritmus PSD vybraného rámce pro vybraný řád filtru se slušnou frekvenční osou.

Úloha 6 - spektrální hustota výkonu pomocí Mel banky filtrů - 2 body

Napište funkci pro odhad spektrální hustoty výkonu (power spectral density, PSD) pomocí Mel-banky filtrů s 23 filtry. Výstup nechť je opět 129 čísel od 0 do poloviny vzorkovací frekvence. Zobrazte logaritmus PSD vybraného rámce se slušnou frekvenční osou. Využijte funkce z knihovny `pytel`².

¹FFT použijte knihovnu, neimplementujte ji, toto není ISS ;)

²<https://www.fit.vutbr.cz/study/courses/ZRE/public/2021-22/features.py>

Úloha 7 - Srovnání PSD - 2 body

Ukažte (stále na vybraném rámci) všechny vypočítané PSD (pomocí FFT, Mel-banky filtrů a LPC pro alespoň 5 různých řádů filtrů) v jednom obrázku. Pokud “nesedí” jejich úrovně, srovnajte je buď teoreticky (úprava gainů, atd.) nebo experimentálně. Komentujte frekvenční rozlišení jednotlivých metod. Uveďte, kde vidíme vliv hlasivek a kde artikulačního ústrojí.

Úloha 8 - RELP - jen krátkodobá predikce - 5 bodů

Vytvořte funkci pro kódování – tvorbu signálu chyby lineární predikce (“residual”) pro celou nahrávku, po 20 ms budete měnit LPC koeficienty. Dejte pozor, abyste ve filtru $A(z)$ neresetovali počáteční podmínky, ale použili je z minulého rámce (prostudujte help k příslušné filtrovací funkci nebo si ji napište sami). Zobrazte reziduál pro vybraný rámec, měli byste jasně vidět, že krátkodobé korelace (“vlnky”) vzniklé díky rezonancím hlasového traktu v něm zmizely.

Vytvořte funkci pro dekódování – generování řeči z reziduálu pro celou nahrávku, po 20 ms budete měnit LPC koeficienty. Dejte pozor, abyste ve filtru $\frac{1}{A(z)}$ neresetovali počáteční podmínky, ale použili je z minulého rámce (prostudujte help k příslušné filtrovací funkci nebo si ji napište sami).

Vygenerujte signál po průchodu takovým kodekem a ověřte, že je shodný s originálem.

Úloha 9 - RELP - včetně základního tónu - 5 bodů

Pokud se podíváte na reziduál z úlohy 8, stále v něm uvidíte “rány” způsobené hlasivkami, které se periodicky opakují. Výsledný reziduál tak není úplně vhodný na kódování. V této úloze se jej pokuste dále “vybělit” (tak, aby se co nejvíce podobal bílému šumu) pomocí znalosti o periodě základního tónu³. Můžete postupovat různě:

1. odhadněte periodu základního tónu (viz cvičení) a analogicky ke krátkodobému prediktoru $A(z) = 1 + a_1 z^{-1} \dots a_{10} z^{-10}$ implementujte dlouhodobý prediktor $B(z) = 1 + b_1 z^{-L}$, kde L je perioda základního tónu ve vzorcích (lag). Koeficient b_1 můžete natvrdo nastavit na 1, odvodit/najít vzoreček pro jeho výpočet, nebo najít dobrou hodnotu hrubou silou pomocí minimalizace energie “druhého” reziduálu po průchodu tímto filtrem.
2. Dlouhodobý prediktor $B(z)$ můžete udělat i složitější s více koeficienty.
3. V historii krátkodobého reziduálu (na výstupu $A(z)$) pomocí korelace najít nejpodobnější úsek, zapamatovat si, jak daleko v historii je (pro znělé hlásky bude “jak daleko v historii je” odpovídat $L!$) a odečíst ho.

Získáte “druhý” reziduál. Zobrazte jej pro vybraný rámec a zkontrolujte, že je “bělejší” než ten krátkodobý.

Připravte si také funkci pro dekódování řeči (pro možnost 1 nebo 2 to bude filtrování filtrem $\frac{1}{B(z)}$, pro 3 přičtení “druhého” reziduálu ke správnému úseku v historii).

Vygenerujte signál po průchodu takovým kodekem a ověřte, že je shodný s originálem.

Úloha 10 - RELP s kvantováním - 5 bodů

Úkol je jednoduchý: stlačit počet bitů potřebných k zakódování celé nahrávky na minimum. Můžete použít cokoliv - jiné LPC parametry (PARCOR, LAR, LSF, cokoliv...), vektorové kvantování, kvantování reziduálu nízkým počtem bitů, atd.

Popište, co jste dělali a jak nízko jste se dostali, aby byla nahrávka stále srozumitelná. Do odevzdaného tgz dejte výslednou nahrávku a do protokolu počet bitů.

Nemusíte implementovat bitové operace, pokud např. víte, že LPC koeficienty kvantujete VQ s kódovou knihou o velikosti 1024, gain skalárně na 4 bitech, lag na 8 bitech a reziduál na 1 bitu na vzorek, stačí počet bitů spočítat jako

$$N_r \times (10 + 4 + 8 + 160 \times 1),$$

kde N_r je počet rámců.

Hodnocení

Viz zadání. Jak je u mě zvykem, student/ka s nejinovativnějším / nejzajímavějším řešením dostane láhev dobrého červeného.

³Kodeky typu CELP dělají toto zpracování většinou v pod-rámcích (40 vzorků), ale toto je projekt do předmětu, klidně použijte celý rámec.

Speciální projekty

Pokud máte zájem o zpracování jiného zadání (např. rozpoznávání řeči, jazyka, mluvčího), nebo na něčem děláte v rámci BP / DP / projektové praxe / ve firmě, atd., prosím kontaktujte mě co nejdříve. Co nejdříve se osobně domluvíme, zadání musím předem odsouhlasit.

Další informace

Cvičící a přednášející kursu ZRE. Přeji všem úspěšný projekt !