

VEŽBA 4 – Osnove kompresije audio signala

Potrebno predznanje

- Poznavanje programskog jezika C
- Diskretna Furijeova transformacija
- Overlap-add metoda
- Kvantizacija

Šta će biti naučeno tokom izrade vežbe

Tokom izrade ove vežbe biće uvedena nova transformacija koja se koristi za prevođenje signala u frekvencijski domen – modifikovana diskretna kosinusna transformacija (MDCT) i biće objašnjene njene prednosti u odnosu na diskretnu Furijeovu transformaciju. Takođe će biti naučene osnovne metode kompresije audio signala bazirane na MDCT transformaciji, kvantizaciji i entropijskom kodovanju.

Motivacija

Kompresija signala je bila ključni faktor koji je omogućio razvoj digitalne televizije, jer je omogućila prenos digitalnih signala kroz kanale čiji je propusni opseg mnogo manji od propusnog opsega originalnog signala koji se prenosi, zadržavajući pri tom visok nivo kvaliteta kodovanog signala. U ovoj vežbi biće ilustrovane osnovne metode kompresije audio signala na kojima se zasniva i većina današnjih audio kodeka (MP3, AAC, WMA, AC-3, OGG Vorbis idr.)

1 Teorijske osnove

Kompresija signala omogućava značajno smanjenje propusnog opsega potrebnog za prenos digitalnog signala, zadržavajući pri tome njegov percipirani kvalitet. Tehnike kompresije signala generalno mogu da se podele u 2 velike kategorije:

- **tehnike kompresije bez gubitaka** (eng. lossless compression) i
- **tehnike kompresije sa gubicima** (eng. lossy compression).

Kod kompresije bez gubitaka posle procesa kodovanja i dekodovanja dobija se signal koji je identičan originalnom ulaznom signalu. Ove tehnike uglavnom se zasnivaju na nekoj vrsti entropijskog kodovanja kod kog se umesto kodova fiksne dužine koriste kodovi kod kojih dužina svakog koda odgovara količini informacije koju odgovarajući element signala nosi, čime se iz signala eliminiše redundansa, odnosno elementi signala koji ne nose nikakvu korisnu informaciju. Dve metode

entropijskog kodovanja koje se najčešće koriste u praksi su Hafmanov (Huffman) kod i aritmetičko kodovanje.

S druge strane, kod kompresije sa gubicima posle kodovanja i dekodovanja dobija se signal koji nije istovetan sa originalnim signalom, već ima određeno izobličenje. Osnovna ideja kod ovih tehnika kompresije je da se izvrši kvantizacija signala, odnosno da se kompresovani signal koduje sa manjim brojem bita nego originalni signal, ali tako da izobličenje signala bude što je moguće manje primetno. U kompresiji audio signala ovo se postiže korišćenjem različitih psihoakustičkih efekata, pre svega efekta maskiranja. Kada u zvučnom signalu postoji komponenta relativno velike snage na nekoj frekvenciji, prag čujnosti u okolini ove frekvencije se podiže, čime se povećava i snaga kvantizacionog šuma koju je moguće uneti u maskirani opseg frekvencija bez velikog uticaja na percipirani kvalitet. Pošto raspodela šuma kvantizacije direktno zavisi od amplitudskog spektra kompresovanog signala, jasno je da se kvantizacija mora vršiti u frekvencijskom domenu. Iako se za obradu signala u frekvencijskom domenu najčešće koristi diskretna Furijeova transformacija, ona zbog nekih svojih osobina ne predstavlja dobar izbor kada je u pitanju kompresija signala. Pre svega, diskretna Furijeova transformacija je definisana tako da transformiše **kompleksni** signal $x(n)$ u **kompleksni** spektar $X(k)$:

$$X(k) = DFT\{x(n)\}, \quad x, X \in \mathbb{C}^N$$

U slučaju kada je signal $x(n)$ realan, za N realnih odbiraka na ulazu dobija se N kompleksnih koeficijenata, odnosno $2N$ realnih brojeva, što je dvostruko više podataka nego na ulazu. Ovo je u direktnoj suprotnosti sa osnovnom idejom kompresije, a to je da kompresovani signal bude predstavljen sa **manje** bita nego ulazni signal. Pošto kompleksni koeficijenti $X(k)$ u ovom slučaju zadovoljavaju relaciju

$$X(N - k) = X^*(k)$$

broj koeficijenata se efektivno ipak može svesti na $N/2$, odnosno na N realnih brojeva. Međutim, ako predstavimo koeficijente $X(k)$ u polarnom obliku

$$X(k) = A(k)e^{i\varphi(k)}$$

vidimo da samo polovina od ovih N realnih brojeva nosi informaciju o amplitudskom spektru, dok druga polovina nosi informaciju o faznom spektru koji je nebitan za proces kompresije. Transformacija koja rešava ovaj problem zove se diskretna kosinusna transformacija (eng. discrete cosine transform, DCT). U kompresiji audio signala najčešće se koristi tzv. DCT-IV varijanta ove transformacije definisana izazom

$$X(k) = \sum_{n=0}^{N-1} x(n) \cos \left[\frac{\pi}{N} \left(n + \frac{1}{2} \right) \left(k + \frac{1}{2} \right) \right], \quad x, X \in \mathbb{R}^N$$

koja transformiše blok od N **realnih** odbiraka u blok od N **realnih** koeficijenata, što je čini boljim izborom od diskretne Furijeove transformacije.

Drugi problem ogleda se u činjenici da se za rekonstrukciju signala koristi Overlap-add metoda kod koje je veličina bloka obrade dvostruko veća od veličine okvira. To znači da se čak i u slučaju DCT transformacije za svaki okvir od N ulaznih odbiraka dobija $2N$ koeficijenata, što je opet dvostruko više podataka nego na ulazu. Ovaj problem rešava se korišćenjem **modifikovane** diskretne kosinusne transformacije (MDCT) koja je definisana izrazom

$$X(k) = \sum_{n=0}^{2N-1} x(n) \cos \left[\frac{\pi}{N} \left(n + \frac{N}{2} + \frac{1}{2} \right) \left(k + \frac{1}{2} \right) \right], \quad x, X \in \mathbb{R}^N$$

MDCT transformiše blok od $2N$ realnih odbiraka u blok od N realnih koeficijenata, čime se za svaki okvir od N odbiraka dobija open N realnih koeficijenata. Inverzna modifikovana kosinusna transformacija (IMDCT) transformiše blok od N koeficijenata u blok od $2N$ realnih odbiraka i ima oblik

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) \cos \left[\frac{\pi}{N} \left(n + \frac{N}{2} + \frac{1}{2} \right) \left(k + \frac{1}{2} \right) \right], \quad x, X \in \mathbb{R}^N$$

MDCT transformacija definisana je tako da u potpunosti eliminiše izobličenje signala nastalo podelom signala na blokove koji se preklapaju. Ako posmatramo tri uzastopna okvira A , B i C dužine N odbiraka, oni čine dva uzastopna bloka obrade $X=[A, B]$ i $Y=[B, C]$ dužine $2N$ odbiraka. Posle direkne i inverzne MDCT transformacije blokova X i Y dobijaju se blokovi

$$\hat{X} = \frac{1}{2} [A - A', B + B']$$

$$\hat{Y} = \frac{1}{2} [B - B', C + C']$$

gde A' , B' i C' predstavljaju blokove A , B i C obrnute u vremenu. Očigledno, sabiranjem druge polovine prvog bloka i prve polovine drugog bloka, kao u overlap-add metodi, dobija se ponovo originalni okvir B .



Slika 1: Blok dijagram tipičnog audio kodera

Pojednostavljeni blok dijagram tipičnog audio kodera prikazan je na slici 1. Posle transformacije MDCT koeficijenti se kvantizuju, pri čemu se broj bita svakog koeficijenta određuje na osnovu psihoakustičkog modela. Da bi se smanjio dinamički opseg između najvećih i najmanjih vrednosti, na koeficijente se pre kvantizacije primenjuje nelinearna transformacija kao što je

$$Y(k) = \text{sgn}(X(k))|X(k)|^{\frac{1}{2}} \quad (1)$$

Kvantizovani koeficijenti se zatim koduju entropijskim kodom koji vrednosti koeficijenata koduje kodovima različitih dužina, u zavisnosti od količine informacije koju data vrednost nosi. Količina informacije izražena u bitima računa se po formuli

$$B(x) = \log_2 \frac{1}{p(x)}$$

gde je $p(x)$ verovatnoća da kvantizovani MDCT koeficijent ima vrednost x . Verovatnoća $p(x)$ obično se računa na osnovu histograma $H(x)$, gde svaka vrednost $H(x)$ predstavlja broj pojavljivanja vrednosti x nad celim ulaznim signalom:

$$p(x) = \frac{H(x)}{\sum_i H(i)}$$

Zadaci

Zadatak 1

Modifikovati funkciju `void obrada(double *in, double *out, int N)` implementiranu u prethodnoj vežbi tako da koristi MDCT transformaciju u mesto FFT transformacije. Funkcija ima sledeće argumente:

- `in` – ulazni okvir
- `out` – izlazni okvir
- `N` – dužina okvira

U implementaciji koristiti sledeće pomoćne bafere koji su definisani u izvornoj datoteci:

- `time_buffer` – blok odbiraka u vremenskom domenu dužine $2N$
- `mdct_buffer` – blok koeficijenata u MDCT domenu dužine N
- `in_delay` – prethodni ulazni okvir dužine N
- `out_delay` – drugih N odbiraka prethodnog izlaznog bloka
- `window` – prozorska funkcija dužine $2N$

Direktna i inverzna modifikovana diskretna kosinusna transformacija realizovane su u okviru funkcija `void mdct(double *in, double *out)` i `void imdct(double *in, double *out)`. Rešenje testirati datim ulaznim datotekama. U izlaznoj datoteci ne sme biti приметnih izobličenja.

Zadatak 2

Proširiti funkciju `obrada` implementiranu u prethodnom zadatku tako da vrši nelinearnu transformaciju prema formuli (1) i kvantizaciju MDCT koeficijenata tako da budu predstavljenim sa zadatim brojem bita B . Rešenje testirati za $B = 12, 10, 8$ i 6 bita. Posle kvantizacije vratiti koeficijente u originalni opseg i rekonstruisati signal. Analizirati izobličenja koja se pojavljuju u izlaznoj datoteci.

Napomena: Vrednosti na izlazu iz MDCT transformacije dužine 2^M imaće $M-1$ bit više u odnosu na ulazne odbirke. Pošto su ulazni odbirci 16-bitni i pošto je veličina transformacije korišćena u zadatku 1024, nekvantizovani MDCT koeficijenti imaće $16+9 = 25$ bita. Posle nelinearne transformacije broj bita se dodatno smanjuje za pola, tako da ostaje približno 13 bita.

Zadatak 3

Proširiti funkciju `obrada` implementiranu u prethodnom zadatku tako da računa histogram vrednosti kvantizovanih MDCT koeficijenata. Implementirati funkciju `void statistika()` koja na osnovu histograma računa verovatnoće $p(x)$ za sve moguće vrednosti kvantizovanih koeficijenata za dati broj bita, kao i odgovarajuću količinu informacije u bitima $B(x)$. Na osnovu $B(x)$ izračunati ukupan broj bita potrebnih za kodovanje celog ulaznog signala i na osnovu toga izračunati faktor kompresije. Pretpostaviti da su u originalnom signalu odbirci kodovani sa 16 bita.