Matematički fakultet

Seminarski rad
iz tehničkog i naučnog pisanja

Algoritmi za kompresiju zvuka: kako MP3 „razume" ljudsko uvo

Student
Mihajlo Janićijević 2/2025

Profesor
dr Jelena Graovac

Beograd, 3. februar 2026.

Sadržaj

Sažetak

Ovaj rad istražuje principe rada MP3 algoritma za kompresiju zvuka, sa fokusom na primenu psihoakustičkih modela ljudskog sluha. MP3 format je revolucionirao distribuciju digitalne muzike postižući kompresiju od približno 10:1 uz zadržavanje prihvatljivog kvaliteta zvuka. Ključ ove efikasnosti leži u eksploataciji ograničenja ljudske percepcije zvuka, prvenstveno fenomena frekventnog i temporalnog maskiranja. Rad objašnjava kako MP3 enkoder koristi filter banke i Modified Discrete Cosine Transform (MDCT) za frekvencijsku analizu audio signala, zatim primenjuje psihoakustički model za identifikaciju zvučnih komponenti koje ljudsko uvo ne može da percipira, i konačno selektivno odbacuje te komponente kroz proces kvantizacije i Huffman kodiranja. Posebna pažnja posvećena je razumevanju kritičnih frekvencijskih opsega, praga maskiranja i njihovoj ulozi u određivanju koje informacije mogu biti uklonjene bez primetnog gubitka kvaliteta. Rad demonstrira kako duboko razumevanje ljudske fiziologije i percepcije omogućava razvoj efikasnih kompresionih algoritama koji su postali temelj moderne digitalne audio industrije.

1 Uvod

Kompresija zvuka predstavlja jedan od ključnih tehnoloških izazova digitalne ere. Njena važnost proizilazi iz nesklada između ogromne količine podataka potrebnih za kvalitetan audio signal i ograničenih resursa za skladištenje i prenos.

Nekompresovani digitalni zvuk CD kvaliteta zahteva frekvenciju uzorkovanja od 44,100 Hz sa rezolucijom od 16 bita po uzorku i dva kanala za stereo. Ovo daje protok od 1,411,200 bita po sekundi - približno 10.6 megabajta po minuti zvuka. Album od 60 minuta zauzima oko 635 megabajta. Tokom ranih 1990ih godina, kada je nastao MP3 format, ova veličina bila je ozbiljno ograničenje.

Tehnološka ograničenja tog perioda bila su izražena. Hard diskovi prosečnih računara imali su 100-500 megabajta kapaciteta, što je omogućavalo skladištenje svega nekoliko albuma. Prenos podataka putem dial-up internet konekcija, sa brzinama od 28.8-56 kilobita po sekundi, zahtevao bi više od 20 minuta za preuzimanje jedne pesme.

Razvoj efikasnih algoritama za kompresiju zvuka bio je neophodan kako bi digitalna muzika postala praktična za svakodnevnu upotrebu i omogućila revoluciju u distribuciji muzike preko interneta.

2 Osnove digitalnog zvuka

Pre nego što se razume kako MP3 postiže kompresiju, važno je shvatiti kako se zvuk predstavlja digitalno.

Proces digitalizacije zvuka odvija se kroz dva koraka. Prvi je uzorkovanje (eng. *sampling*), pri čemu se amplituda zvučnog talasa meri u regularnim vremenskim intervalima. Frekvencija uzorkovanja određuje koliko puta u sekundi se vrši merenje. Prema Nyquist-Shannon teoremi, frekvencija uzorkovanja mora biti najmanje dvostruko veća od najviše frekvencije koja se želi reprodukovati. Pošto je gornja granica ljudskog sluha oko 20 kHz, CD standard koristi 44.1 kHz, što omogućava preciznu reprodukciju svih čujnih frekvencija.

Drugi korak je *kvantizacija* (eng. *quantization*), gde se svaki izmereni uzorak reprezentuje konačnim brojem bita. CD kvalitet koristi 16-bitnu kvantizaciju, što daje 65,536 različitih nivoa amplitude. Veći broj bita znači precizniju reprezentaciju originalnog signala i manji šum kvantizacije, ali i veću veličinu podataka. Ova kombinacija frekvencije uzorkovanja i rezolucije kvantizacije naziva se Pulse Code Modulation (PCM) i predstavlja standardni format za nekompresovani digitalni audio.

Ključna razlika za razumevanje kompresije je između bezgubitne i gubitne kompresije. Bezgubitna kompresija (eng. *lossless compression*), kao FLAC ili ALAC, smanjuje veličinu fajla bez odbacivanja informacija, omogućavajući identičnu rekonstrukciju originalnog signala. Međutim, tipična kompresija je samo 2:1 ili 3:1. Gubitna kompresija (eng. *lossy compression*), koju koristi MP3, postiže mnogo veću kompresiju odbacivanjem informacija koje ljudsko uvo verovatno neće primetiti. Ovaj pristup zasniva se na perceptualnom modelovanju - zadržava samo podatke relevantne za ljudsku percepciju zvuka.

Bitrate (eng. *bitrate*) definiše koliko bita podataka se koristi za reprezentaciju jedne sekunde zvuka, izražen u kilobitima po sekundi (kbps). Nekompresovani CD kvalitet ima bitrate od 1,411 kbps, dok tipični MP3 fajlovi koriste 128-320 kbps. Niži bitrate znači manju veličinu ali potencijalno niži kvalitet, dok viši bitrate obezbeđuje bolji kvalitet uz veću veličinu. U tabeli 1 je prikazana razmera kompresije za različite vrednosti bitrate-a. [3, 5]

Bitrate (kbps)	**Veličina (MB)**	**Kompresija**
1411 (CD)	31.7	1:1
320	7.2	4.4:1
192	4.3	7.4:1
128	2.9	11:1
96	2.2	14.6:1

Tabela 1: Poređenje MP3 bitrate-a za audio zapis od 3 minuta

3 Psihoakustika

Efikasnost MP3 algoritma proizilazi iz razumevanja načina na koji ljudsko uvo percipira zvuk. Psihoakustika je disciplina koja proučava psihološku i fiziološku percepciju zvuka, i upravo ova saznanja omogućavaju da se identifikuju koje komponente audio signala mogu biti uklonjene bez da slušalac primeti razliku. MP3 format ne pokušava da sačuva sve podatke već samo one koje ljudsko uvo može da registruje.

3.1 Struktura i ograničenja ljudskog sluha

Ljudsko uvo je kompleksan organ sa specifičnim mogućnostima i ograničenjima. Spoljašnje uvo prikuplja zvučne talase i usmerava ih ka bubnoj opni, koja vibrira prenoseći mehaničku energiju na tri male kosti u srednjem uvu. Ove kosti pojačavaju vibracije i prenose ih ka pužnici u unutrašnjem uvu, spiralnom organu ispunjenom tečnošću koji sadrži hiljade dlakastih ćelija raspoređenih duž bazilarne membrane.

Različite frekvencije zvuka izazivaju maksimalne vibracije na različitim lokacijama duž bazilarne membrane - visoke frekvencije stimulišu ćelije blizu ulaza u pužnicu, dok niske frekvencije stimulišu ćelije na suprotnom kraju. Ljudsko uvo može da percipira frekvencije od približno 20 Hz do 20 kHz, pri čemu je osetljivost najizraženija u oblasti od 2-5 kHz, što odgovara frekvencijskom opsegu ljudskog govora.

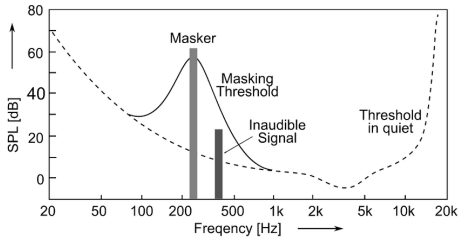
Međutim, uvo ne analizira svaku pojedinačnu frekvenciju nezavisno. Ono grupuje frekvencije u takozvane kritične opsege (eng. *critical bands*). U ljudskom sluhu postoji oko 24 kritična opsega, pri čemu je širina svakog opsega približno 100 Hz ispod 500 Hz, a iznad te granice raste eksponencijalno i može dostići nekoliko hiljada Hz. Ova organizacija je fundamentalna za MP3 kompresiju jer znači da uvo ne može da razlikuje između dve vrlo bliske frekvencije unutar istog kritičnog opsega.[1],

3.2 Maskiranje

Najvažniji psihoakustički fenomen koji MP3 koristi je maskiranje (eng. *masking*) - pojava gde prisustvo jednog zvuka sprečava percepciju drugog zvuka. Postoje dva osnovna tipa maskiranja: frekventno i temporalno.

Frekventno maskiranje (eng. *frequency masking*) nastaje kada glasniji zvuk na određenoj frekvenciji "maskira"ili čini nečujnim tiše zvukove na sličnim frekvencijama. Efekat je najizraženiji unutar istog kritičnog opsega ali se može proširiti i na susedne opsege. Praktičan primer je kada bas bubanj svira glasno - tihi zvukovi drugih instrumenata na sličnim niskim frekvencijama postaju potpuno nečujni. Kada MP3 enkoder detektuje jak signal na nekoj frekvenciji, on može značajno redukovati ili potpuno eliminisati podatke o tišim zvucima u okolnom opsegu jer slušalac ih neće moći da čuje.

Prag maskiranja (eng. *masking threshold*) predstavlja nivo ispod kojeg zvukovi postaju nečujni zbog prisustva drugih, glasnijih zvukova. MP3 algoritam izračunava ovaj prag za svaki frekventni opseg i koristi ga kao kriterijum za određivanje koliko bitova treba dodeliti. Signali čija je amplituda ispod praga maskiranja mogu biti potpuno odbačeni, dok oni blizu praga mogu biti kodirani sa mnogo manje bitova.



Slika 1: Ilustracija praga maskiranja.

Temporalno maskiranje (eng. *temporal masking*) opisuje fenomen gde glasan zvuk maskira tiše zvukove koji se dešavaju neposredno pre njega ili nakon njega. Pre-maskiranje (eng. *pre-masking*) traje kratko, obično 5-20 milisekundi pre glasnog zvuka, i nastaje zbog kašnjenja u neuralnom procesiranju. Post-maskiranje (eng. *post-masking*) traje duže, do 50-200 milisekundi nakon glasnog zvuka, jer auditorni sistem treba vreme da

se vrati u normalno stanje osetljivosti. Ovaj fenomen je posebno izražen kod naglih, perkusivnih zvukova. MP3 enkoder koristi ovu informaciju da smanji broj bitova potrebnih za kodiranje audio segmenata koji se dešavaju neposredno pre i posle glasnih delova. Slika 1 je ilustracija na kojoj se vidi da glasniji ton podiže prag čujnosti za okolne frekvencije. Signali ispod praga mogu biti odbačeni jer neće biti čujni [1, 2, 7]

3.3 Primer maskiranja

Da bi se ilustrovalo kako maskiranje funkcioniše, razmotrimo pojednostavljen scenario. Pretpostavimo da imamo dva tona: Ton A na frekvenciji od 1000 Hz sa nivoom od 60 dB, i Ton B na frekvenciji od 1100 Hz sa nivoom od 30 dB. Ove dve frekvencije spadaju u isti kritični opseg (širok oko 160 Hz na 1000 Hz).

Prisustvo glasnog Tona A podiže prag čujnosti u okolnom frekventnom opsegu. Za frekvenciju od 1100 Hz, prag maskiranja može biti podignut na, recimo, 35 dB. Pošto je Ton B (30 dB) ispod ovog praga maskiranja (35 dB), on neće biti čujan kada se oba tona reprodukuju istovremeno. MP3 enkoder može potpuno zanemariti podatke o Tonu B jer ga slušalac neće moći da detektuje. Međutim, kada bi Ton B imao nivo od 40 dB, on bi bio iznad praga maskiranja i morao bi biti kodiran, mada možda sa manjom preciznošću nego kada bi se reprodukovao samostalno.

Ovaj princip se primenjuje na hiljade frekventnih komponenti istovremeno kroz ceo spektar, omogućavajući značajnu redukciju količine podataka koje treba sačuvati dok kvalitet zvuka ostaje visok.

4 Princip funkcionisanja MP3 algoritma

Osnovni princip rada MP3 kompresije zasniva se na sekvencijalnoj obradi audio signala kroz nekoliko povezanih koraka. Signal se prvo deli na male vremenske segmente (okvire), zatim analizira u frekventnom domenu, primenjuje se psihoakustički model za identifikaciju maskiranih komponenti, izvršava se kvantizacija sa različitom preciznošću za različite frekventne opsege, i konačno se dobijeni podaci dodatno kompresuju bezgubitnim kodiranjem. Svaki korak doprinosi ukupnoj kompresiji dok minimizuje perceptibilni gubitak kvaliteta.

4.1 Filter banka i MDCT transformacija

Prvi korak u MP3 enkodiranju je prolazak audio signala kroz polifaznu filter banku (eng. *polyphase filterbank*) koja deli signal na 32 jednaka frekventna podopsega. Svaki podopseg pokriva približno 689 Hz spektra (22,050 Hz / 32), od najnižih do najviših frekvencija. Ova inicijalna podela omogućava grubu frekventnu separaciju i olakšava dalje procesiranje.

Međutim, rezolucija od 32 podopsega nije dovoljna za preciznu primenu psihoakustičkog modela jer kritični opsezi ljudskog sluha nisu uniformne širine - oni su uži na nižim i širi na višim frekvencijama. Zato MP3 primenjuje dodatnu transformaciju na izlaz filter banke. Modified Discrete Cosine Transform (MDCT) je matematička transformacija koja dalje deli svaki od 32 podopsega na 18 spektralnih linija (za kratke blokove) ili 6 spektralnih linija (za duge blokove), rezultujući ukupno sa 576 frekventnih koeficijenata po kanalu. MDCT je izabran jer ima posebnu osobinu da

se blokovi preklapaju (50% overlap) što eliminiše artefakte na granicama između blokova. Takođe, MDCT koncentriše energiju signala u mali broj koeficijenata što olakšava kompresiju. [6, 5]

4.2 Primena psihoakustičnog modela

Dok se signal transformiše kroz filter banku i MDCT, paralelno se izvršava psihoakustička analiza. Psihoakustički model je srce MP3 algoritma i odgovoran je za određivanje koje komponente signala mogu biti odbačene ili redukovane. Model analizira spektar audio signala i izračunava prag maskiranja za svaki frekventni opseg.

Proces počinje identifikacijom tonalnih (harmoničnih) i ne-tonalnih (šumnih) komponenti u signalu. Tonalne komponente, kao čisti tonovi muzičkih instrumenata, obično imaju jači efekat maskiranja od šumnih komponenti. Model zatim mapira ove komponente na kritične opsege i za svaki opseg izračunava lokalnu funkciju maskiranja.

Signal-to-Mask Ratio (SMR) predstavlja odnos između stvarne jačine signala i praga maskiranja u svakom frekventnom opsegu. Visok SMR znači da je signal značajno iznad praga i mora biti pažljivo kodiran, dok nizak SMR ukazuje da je signal blizu praga i može biti kodiran sa manjom preciznošću ili čak potpuno odbačen.[5]

4.3 Kvantizacija i alociranje bitova

Na osnovu informacija iz psihoakustičkog modela, MP3 enkoder vrši neuniformnu kvantizaciju MDCT koeficijenata. Proces kvantizacije podrazumeva zaokruživanje realnih vrednosti koeficijenata na ograničen skup diskretnih nivoa, što neizbežno uvodi greške ali omogućava kompresiju.

Alokacija bitova je dinamički proces gde enkoder raspoređuje raspoložive bite između različitih frekventnih opsega. Frekventni opsezi sa visokim SMR vrednostima dobijaju više bitova i finiju kvantizaciju, dok opsezi sa niskim SMR vrednostima dobijaju malo ili nimalo bitova. Cilj je da kvantizacioni šum ostane ispod praga maskiranja u svim frekventnim opsegama, čineći ga nečujnim.

Enkoder koristi iterativni proces da pronađe optimalnu alokaciju bitova. Počinje sa inicijalnom podjelom, zatim analizira nastali kvantizacioni šum i poredi ga sa pragom maskiranja. Ako je šum iznad praga u nekom opsegu, više bitova se dodeljuje tom opsegu. Proces se ponavlja dok se ne postigne zadovoljavajući balans između veličine kompresovanog podatka i perceptualnog kvaliteta.[5, 7]

4.4 Huffman kodiranje i formiranje bitstream-a

Nakon kvantizacije, kvantizovani koeficijenti se dalje kompresuju primenom Huffman kodiranja, tehnike bezgubitne kompresije koja koristi kraće kodove za češće vrednosti i duže kodove za ređe vrednosti. Pošto kvantizacija proizvodi mnogo nula (odbačenih koeficijenata ispod praga maskiranja) i malih vrednosti, Huffman kodiranje značajno redukuje veličinu podataka.

MP3 format organizuje kodirane podatke u okvire (eng. *frames*), gde svaki okvir reprezentuje 1152 uzorka originalnog audio signala što odgovara približno 26 ms zvuka na CD frekvenciji uzorkovanja. Svaki okvir sadrži

zaglavlje sa metapodacima (bitrate, frekvencija uzorkovanja, mod), glavne podatke (kvantizovani koeficijenti), i dodatne informacije (scale faktori, Huffman tablice).

Konačan MP3 fajl je sekvenca ovih nezavisnih okvira, što omogućava streaming reprodukciju i olakšava manipulaciju fajlom jer svaki okvir može biti dekodiran nezavisno od ostalih. [4, 2]

5 Zaključak

MP3 format predstavlja elegantnu demonstraciju kako duboko razumevanje ljudske biologije i percepcije može voditi razvoju revolucionarnih tehnoloških rešenja. MP3 algoritam postiže kompresiju zvuka od približno 10:1 zadržavajući visok perceptualni kvalitet.

Širi značaj MP3 formata nadilazi tehničke aspekte. MP3 je omogućio revoluciju u distribuciji muzike, razvoj prenosivih audio plejera i pojavu digitalnih muzičkih platformi, fundamentalno menjajući muzičku industriju. Iako su noviji formati kao AAC i Opus postigli dodatna poboljšanja, principi perceptualnog kodiranja koje je utvrdio MP3 ostaju temelj moderne audio kompresije i demonstriraju kako razumevanje ljudske biologije vodi ka inovativnim tehnološkim rešenjima.

Literatura

[1] TrueGeometry Blog. Psychoacoustic modeling in mp3 compression. https://blog.truegeometry.com/api/exploreHTML/166bbe34b1629d2c84feb7f48cd97092.exploreHTML, 2025. Blog post.

[2] Gabriel Bouvigne. Overview of the mp3 techniques. http://www.mp3-tech.org/tech.html, 2007. MP3'Tech website.

[3] John (Jake) Guckert. The use of fft and mdct in mp3 audio compression. https://www.math.utah.edu/~gustafso/s2012/2270/web-projects/Guckert-audio-compression-svd-mdct-MP3.pdf, 2012. Math 2270, University of Utah.

[4] Programiz. Huffman coding. https://www.programiz.com/dsa/huffman-coding, 2024.

[5] Open University. Exploring communications technology:. https://www.open.edu/openlearn/digital-computing/exploring-communications-technology/content-section-3.1, 2017. OpenLearn publication. Sekcije 3.1 do 3.6.

[6] Stanford University. Mp3: Hybrid filter bank. https://cs.stanford.edu/people/eroberts/courses/soco/projects/data-compression/lossy/mp3/hybrid_filter.htm, 2012. Project page, Stanford CS.

[7] Stanford University. Mp3: Psychoacoustic model. https://cs.stanford.edu/people/eroberts/courses/soco/projects/data-compression/lossy/mp3/psychoacoustics.htm, 2012. Project page, Stanford CS. Ukljucujuci sledecu lekciju pod nazivom "Bit Allocation".