

Kompresija zvuka i zvukovnih datoteka

Martinek, Karlo

Master's thesis / Diplomski rad

2020

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:217:291470>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-04-16**



Repository / Repozitorij:

[Repository of the Faculty of Science - University of Zagreb](#)



SVEUČILIŠTE U ZAGREBU
PRIRODOSLOVNO–MATEMATIČKI FAKULTET
MATEMATIČKI ODSJEK

Karlo Martinek

KOMPRESIJA ZVUKA I ZVUKOVNIH
DATOTEKA

Diplomski rad

Voditelj rada:
dr. sc. Goran Igaly

Zagreb, rujan, 2020

Ovaj diplomski rad obranjen je dana _____ pred ispitnim povjerenstvom u sastavu:

1. _____, predsjednik
2. _____, član
3. _____, član

Povjerenstvo je rad ocijenilo ocjenom _____.

Potpisi članova povjerenstva:

1. _____
2. _____
3. _____

Sadržaj

Sadržaj	iii
Uvod	1
1 Digitizacija zvuka	2
2 Kompresija podataka kvantizacijom	9
2.1 Neuniformna kvantizacija - μ -zakon i A-zakon	9
2.2 DPCM	12
2.3 ADPCM	13
2.4 DM	15
3 Vokoderi	17
3.1 Kanalni vokoderi	17
3.2 LPC	18
4 MPEG	23
4.1 Psihoakustika	23
4.2 MPEG-1	26
4.3 MPEG-2	37
5 Kompresija zvuka bez gubitaka	43
5.1 Metode kodiranja	43
5.2 FLAC	47
5.3 MPEG 4 - ALS	49
6 Zaključak	53
Bibliografija	54

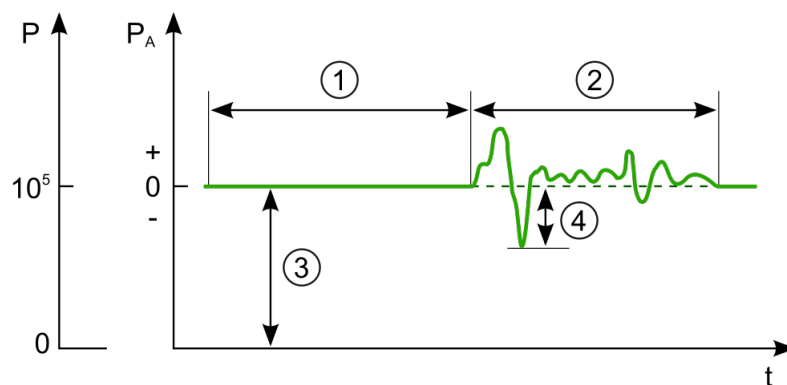
Uvod

Zvukovni podaci čine sastavni dio multimedije i telekomunikacije. Nameće se pitanje kako dobiti što kvalitetniji audio zapis, a da se pritom zauzima što manje prostora. Kompresija audio podataka je proces transformiranja digitalnih podataka kako bi se uštedio memorijski prostor, a da se pritom sačuvaju bitne zvukovne informacije. Postoje dvije vrste kompresije, kompresija s gubitcima i kompresija bez gubitaka. Kompresija bez gubitaka reducira veličinu datoteka na način da originalni signali ostanu u potpunosti očuvani. Kompresija s gubitcima uspijeva još više sažeti podatke, ali se pritom dio podataka bespovratno gubi. U ovom radu opisat će se osnovne karakteristike zvuka i procesa njegove pretvorbe u digitalni oblik. Prikazati će se koderi građeni specifično za kodiranje govornih podataka. Zatim će se opisati iskorištavanje psihoakustičnih svojstava zvuka pri kompresiji s gubitcima. Na kraju će se prezentirati principi kompresije bez gubitaka.

Poglavlje 1

Digitizacija zvuka

Zvuk je mehanički, longitudinalni val koji se giba u elastičnom mediju. Prilikom vibracije nekog predmeta pobuđuju se okolne čestice koje počinju titrati oko ravnotežnog položaja. Energija koju proizvode čestice se širi i time nastaju zvučni valovi. Prilikom širenja zvučnih valova, energija čestica remeti atmosferski tlak. Taj poremećaj naziva se zvučni tlak (eng. *sound pressure*). Kada valovi dođu do ljudskog uha, prolaze ušnim kanalom i udaraju u bubnjić. Mehanički zvučni val pretvara se u vibracije na opni bubnjića koje pokreću tri košćice: čekić, nakovanj i stremen. Košćice dalje prenose vibracije na tekućinu kojom je ispunjena pužnica. Unutar pužnice nalaze se osjetne stanice koje pretvaraju vibracije u neuronske impulse koji se šalju u mozak gdje se stvara osjet sluha.



Slika 1.1: Dijagram zvučnog tlaka: 1) Tišina; 2) Čujni zvuk; 3) Atmosferski tlak; 4) Zvučni tlak

Zvuk određuju dvije fizikalne veličine, frekvencija i valna duljina. Frekvencija zvuka f je definirana kao broj titraja čestice medija kojim se zvuk prenosi. Valna duljina je

udaljenost između dva identična susjedna zgušnjenja, odnosno razrjeđenja medija kojim se val širi. Odnos frekvencije i valne duljine obrnuto je proporcionalan, što znači da zvukovi viših frekvencija imaju veću valnu duljinu i obrnuto. Ljudsko uho registrira zvukove čija frekvencija iznosi između 20 Hz i 20 kHz, no te granice nisu fiksne i variraju od osobe do osobe. Ljudi starenjem postepeno gube osjetljivost na tonove iz višeg spektra frekvencija. Zvuk frekvencije više od 20 kHz naziva se ultrazvuk, a onaj frekvencije manje od 20 Hz naziva se infrazvuk.

Intenzitet ili jakost zvuka I je količina energije koja u jednoj sekundi prostruji kroz plohu površine 1 m^2 okomitu na smjer kretanja zvučnog vala. Mjerna jedinica SI sustava za intenzitet je W/m^2 . Minimalni intenzitet potreban da osoba primijeti zvuk iznosi 10^{-12} W/m^2 i prema dogovoru se ta vrijednost uzima kao referentni zvučni intenzitet I_0 . Intenzitet zvuka pri kojem čovjek osjeća bol iznosi između 1 i 100 W/m^2 . Zbog jako velikog raspona intenzitet se iskazuje na drugačiji način.

Razina intenziteta zvuka (eng. *sound intensity level*, SIL) logaritamski je omjer intenziteta i referentnog zvučnog intenziteta i izražava se u decibelima (dB).

$$\text{SIL} = 10 \log_{10} \left(\frac{I}{I_0} \right)$$

Decibel je relativna mjerna jedinica za izražavanje logaritamskih omjera imenovana po Alexandru Bellu.

Na isti način se izražava i razina zvučnog tlaka (eng. *sound pressure level*, SPL). Za referentni zvučni tlak p_0 uzima se vrijednost koja je potrebna da se dobije referentni zvučni intenzitet i ona iznosi $20 \mu\text{Pa}$

$$\text{SPL} = 20 \log_{10} \left(\frac{p}{p_0} \right)$$

U idućoj tablici dani su primjeri, intenziteti i odgovarajuće razine intenziteta zvukova.

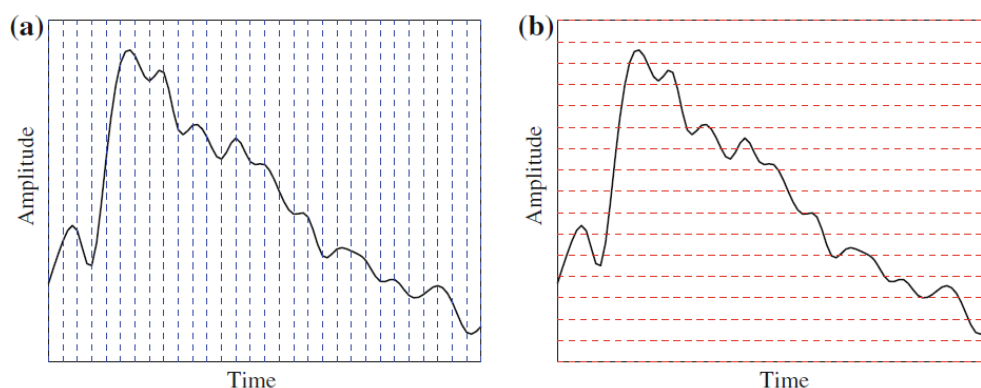
Visina zvuka (eng. *pitch*) je auditorno svojstvo zvuka prema kojemu se zvukovi mogu poredati na ljestvici od niskog prema visokom. Visina je povezana s frekvencijom zvuka. Viši tonovi imaju više frekvencije, a niži tonovi imaju niže frekvencije. Visina zvuka ograničena je granicama ljudske sposobnosti percepcije zvuka.

Glasnoća je subjektivna percepcija intenziteta zvuka ovisna o amplitudi, frekvencijskom sastavu i vremenskom obliku zvučnog signala. Glasnoća nije u linearnoj vezi sa zvučnim tlakom ili razinom zvučnog tlaka. Budući da je glasnoća subjektivna percepcija, nije ju moguće izravno izmjeriti već se mjerenje provodi ispitivanjem slušača. Ispitivanje se provodi na način da se pusti čisti ton određene frekvencije na 1 kHz. Nakon toga se pušta drugi ton kojem se mijenja intenzitet sve dok slušač ne kaže da percipira drugi ton jednako glasnim poput prvog. Mjerna jedinica za glasnoću je fon. Iznos glasnoće u fonima odgovara razini zvučnog tlaka samo na zvukovima frekvencije 1 kHz dok su vrijednosti za ostale frekvencije dobivene subjektivno na gore opisani način.

SIL (dB)	Intenzitet zvuka (W/m^2)	Primjer zvuka
0	10^{-12}	Čujni prag
10	10^{-11}	Šuškanje lišća
20	10^{-10}	Tiha soba
40	10^{-8}	Tihi radio
60	10^{-6}	Normalni razgovor
70	10^{-5}	Promet
80	10^{-4}	Lupanje vratima
90	10^{-3}	Automobilska truba
100	10^{-2}	Sirena na udaljenosti od 30 m
110	10^{-1}	Pneumatska bušilica
125	3	Prag boli
140	100	Mlažnjak na udaljenosti od 30 m

Nakon što smo opisali zvuk i njegove karakteristike, zanima nas kako pohraniti zvukove u digitalnom obliku. Proces pretvaranja informacija iz analognog u digitalni oblik naziva se digitizacija (eng. *digitization*). Korištenjem mikrofona, zvučni valovi udaraju membranu mikrofona koja počinje titrati. Membrana je spojena s pretvornikom koji titranje membrane pretvara u električni napon. Time smo pretvorili zvuk u kontinuirani analogni signal.

Kodek (eng. *codec*, skraćeno od *coder-decoder*) je hardverski sklop koji obavlja posao pretvorbe analognog u digitalni signal i digitalni u analogni signal. Nemoguće je odrediti i pohraniti signal u kontinuiranom obliku, već je signal potrebno diskretizirati.

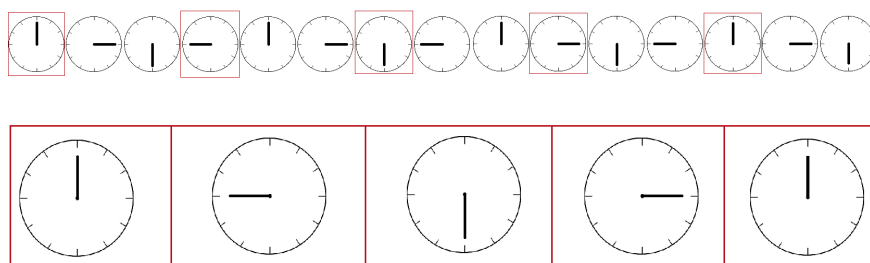


Slika 1.2: a) Uzorkovanje; b) Kvantizacija

Uzorkovanje (eng. *sampling*) je postupak diskretizacije kontinuiranog signala po vremenu. Vrijednost signala mjeri se u jednakim vremenskim periodima T . Broj skupljenih uzoraka po sekundi naziva se stopom ili frekvencijom uzorkovanja f_s (eng. *sample rate*).

$$f_s = 1/T$$

Odabirom veće frekvencije uzoraka, sakuplja se veći broj informacija i pravilnije se reprezentira kontinuirani signal. Budući da se bilježe podaci samo u trenutku uzorka, informacije između dva uzorka se gube. Odabere li se pogrešna frekvencija uzorkovanja, može doći do krivog tumačenja podataka i neželjenih posljedica. Proučimo primjer sa slike 1.3.



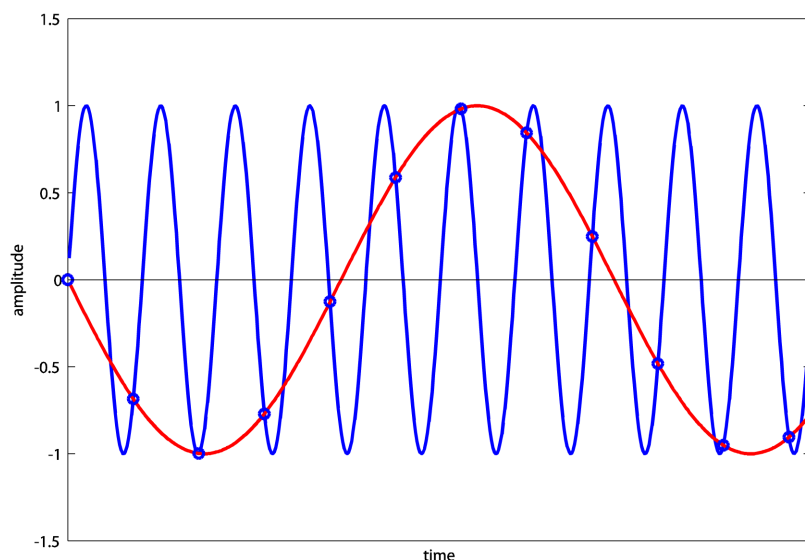
Slika 1.3: Uzorkovanje vrijednosti sata

Promatramo zidni sat koji na sebi ima samo sekundnu kazaljku. Neka je frekvencija uzorkovanja $f_s = 0.022$ Hz tj. period između uzoraka iznosi 45 sekundi. Stavimo li niz dobivenih uzoraka jedan do drugoga, zaključili bismo da se kazaljka kreće u smjeru suprotnom od kazaljke na satu.

Sličan efekt može se vidjeti u snimkama kotača, gdje se, zbog razlike u frekvencijama okretaja kotača i uzorkovanja tj. broju slika u sekundi prilikom snimanja, može činiti da se kotač okreće u suprotnom smjeru od stvarnog smjera vrtnje.

Taj efekt se naziva frekvencijsko poklapanje (eng. *aliasing*). Frekvencijsko poklapanje je pojava izobličenja (eng. *distortion*) koje se javlja kada je signal rekonstruiran iz uzoraka različit od originalnog. Do frekvencijskog poklapanja dolazi ako nije uzeta dovoljno velika frekvencija uzorkovanja. Frekvencijsko poklapanje kod audio signala rezultira "mutnijim" zvukom. Javlja se pozadinski šum (eng. *noise*), koji se pojačava sa sve većim smanjenjem stope uzorkovanja, te se gube visoke frekvencije.

Postavlja se pitanje kako odabrati pravu stopu uzorkovanja da bi se digitizirao upravo onaj signal koji se šalje kao ulaz našem računalu. Odgovor na to pitanje dali su Harry Nyquist i Claude Shannon.



Slika 1.4: Frekvencijsko poklapanje

Nyquist-Shannon teorem o uzorkovanju kaže da ako funkcija $x(t)$ ne sadrži frekvencije više od B Hz, onda je u potpunosti određena uzorcima uzetima u razmacima od $1/(2B)$ sekundi. [12]

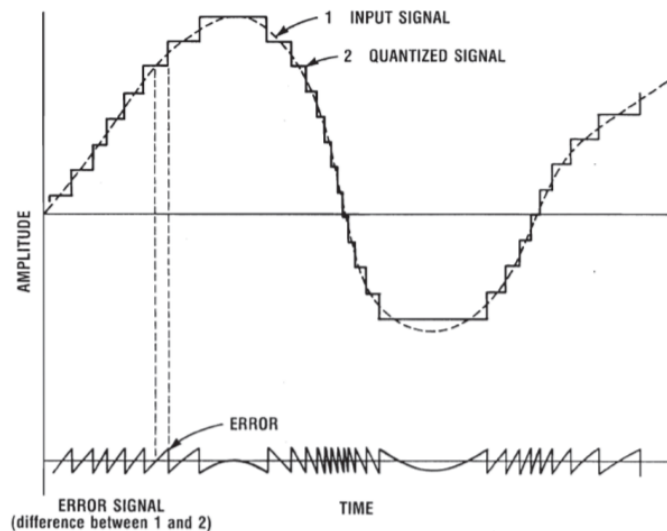
Dakle, za dani zvuk maksimalne frekvencije f , potrebna je stopa uzorkovanja $f_s \geq 2 \cdot f$ kako bi se izbjeglo frekvencijsko poklapanje. Najmanja takva stopa uzorkovanja se naziva Nyquistova stopa. Za danu stopu uzorkovanja f_s , najviša frekvencija koja može biti točno uzorkovana je $f_s/2$ i ona se naziva Nyquistovom frekvencijom.

Obzirom da ljudski raspon čujnih frekvencija leži u intervalu 20 Hz do 20 kHz, frekvencija uzorkovanja koja će biti dovoljno velika za savršenu reprodukciju originalnog signala je veličine 40 kHz ili više. Standardna frekvencija uzorkovanja za CD iznosi 44.1 kHz. Ona je uzeta kao standard kodiranja zbog kompatibilnosti sa tadašnjim VCR, te PAL i NTSC video sustavima.

Vidjeli smo kako uzorkovanje transformira kontinuirani vremenski signal iz $\mathbb{R}$ u diskretni signal u $\mathbb{N}$, no to je samo prvi korak pretvaranja analognog signala u digitalni. Drugi korak je pretvaranje kontinuiranog raspona mogućih amplituda na diskretni raspon $\Gamma \subset \mathbb{R}$. Taj proces naziva se kvantizacija (eng. *quantization*).

Uzorkovanje predstavlja vrijeme mjerenja, a kvantizacija iznos tog mjerenja. Kod zvučnih signala iznos mjerenja je amplituda signala. Amplituda se dijeli na dijelove koji se nazivaju reprezentacijske razine (eng. *representation levels*). Ako je udaljenost između

svih dijelova jednaka, kvantizacija je uniformna. Broj razina na koje se amplituda dijeli ovisi o odabiru fiksnog broja bitova koji će bilježiti podatke. Taj broj bitova naziva se dubina bitova (eng. *bit-depth*). Standardna dubina bitova za CD snimke je 16 bitova što daje 65536 mogućih reprezentacijskih razina. Vrijednost amplitude između dvije susjedne razine naziva se korakom kvantizacije (eng. *quantization step*).



Slika 1.5: Primjer kvantizacije i kvantizacijske pogreške

Kvantizacija se može prikazati funkcijom $Q : \mathbb{R} \rightarrow \Gamma$. Funkcija Q se naziva kvantizatorom i ona pridružuje svakoj očitanoj amplitudi a , vrijednost $Q(a)$ prema idućoj formuli:

$$Q(a) := \text{sign}(a) \cdot \Delta \cdot \left\lfloor \frac{|a|}{\Delta} + \frac{1}{2} \right\rfloor$$

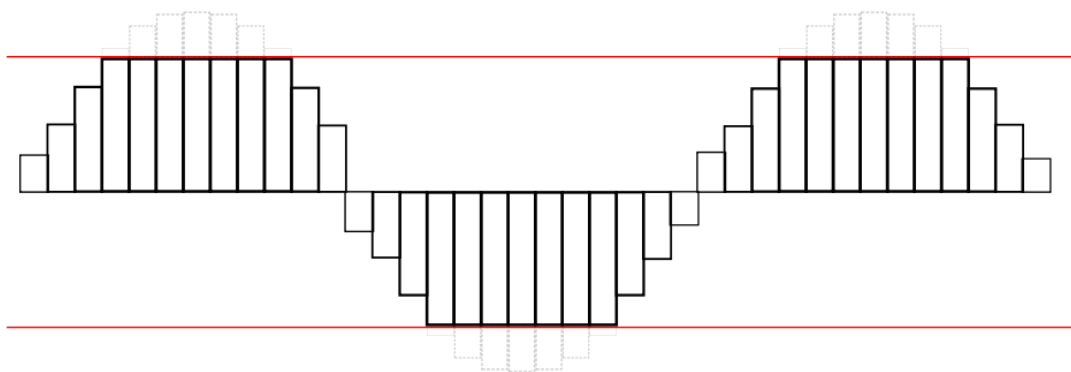
gdje je Δ veličina koraka kvantizacije.

Prilikom kvantizacije također dolazi do gubitaka podataka, jer se različite analogne vrijednosti pridružuju istoj digitalnoj vrijednosti. Razlika između stvarne analogne vrijednosti i kvantizirane vrijednosti naziva se kvantizacijska pogreška. Ako je iznos pogreške značajan tada se javlja kvantizacijski šum.

Omjer signala i šuma, SNR (od eng. *signal-to-noise ratio*) je karakterizacija šuma u sustavu na način da se u odnos stave amplitude najvećeg korisnog signala A_{max} i šuma signala A_{ϵ} . Omjer je definiran kao

$$SNR = 20 \log_{10} \frac{A_{max}}{A_{\epsilon}} \text{ dB.}$$

Osim šuma, još jedno izobličenje zvuka je moguće prilikom kvantizacije. Izobličenje radi preopterećenja (eng. *overload distortion*) javlja se kada ulazni signal premašuje raspon kvantizatora. Kada je raspon premašen, izlaz će ostati na maksimalnoj ili minimalnoj vrijednosti (ovisno s koje strane je premašen) sve dok se ulaz ne vrati u okvire raspona. Taj efekt naziva se rezanje ili ograničavanje amplituda (eng. *clipping*)



Slika 1.6: Primjer rezanja amplituda frekvencije

Frekvenciju bitova (eng. *bit-rate*) kvantiziranog audio signala dobivamo kao umnožak frekvencije uzorkovanja, dubine bitova i broju audio kanala koji su kodirani.

Upravo opisani postupci uzorkovanja i kvantizacije daju pulsno-kodnu modulaciju. Pulsno-kodna modulacija (eng. *pulse code modulation*, PCM) najraširenija je digitalna reprezentacija analognog zvuka. Svaki uzorak je reprezentiran fiksnim brojem bitova i kodiran je neovisno o ostalim uzorcima. Kvaliteta dobivenog zvuka ovisit će o frekvenciji uzorkovanja i kvaliteti kvantizacije. Velika frekvencija uzorkovanja daje bolje reprezentirani signal, a uz veliku dubinu bitova dobivamo preciznije izmjerenu amplitudu signala. Velika kvaliteta, međutim, sa sobom donosi i veliku količinu podataka koju je potrebno pohraniti. Prosječna veličina petominutne WAV datoteke dobivene pulsno kodnom modulacijom sa frekvencijom uzorkovanja od 44.1 kHz i dubinom od 16 bitova iznosi oko 50 MB.

Poglavlje 2

Kompresija podataka kvantizacijom

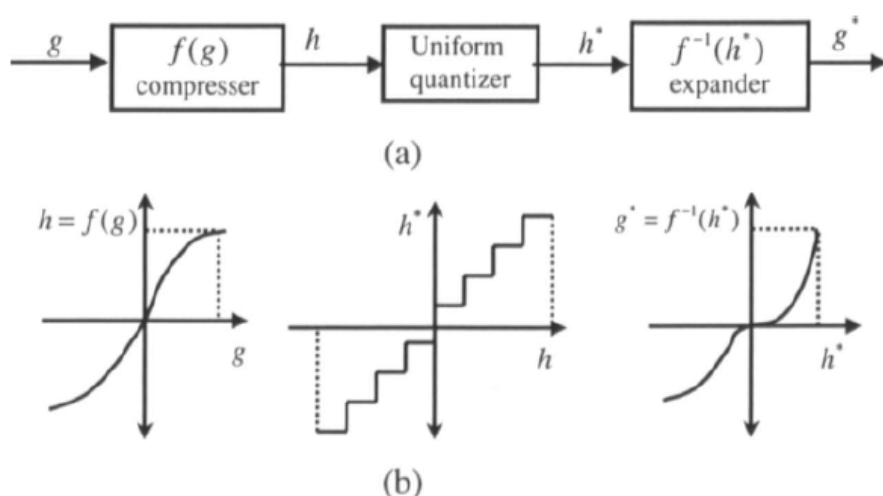
U prošlom poglavlju vidjeli smo da su zvukovne datoteke visoke razine kvalitete dosta velike. Nameće se pitanje može li se veličina datoteke nekako smanjiti, a da se zadrži kvaliteta zvuka te je li potrebno uhvatiti apsolutno sve frekvencije unutar ljudskog čujnog spektra? U ovome poglavlju prikazat ćemo nekoliko modela kvantizacije koji modificiraju osnovnu ideju PCM-a kako bi se uštedio prostor i zadržala primjerena kvaliteta zvuka za područje u kojem se ti zvukovni podaci koriste.

2.1 Neuniformna kvantizacija - μ -zakon i A-zakon

Frekvencije signala nisu uniformno distribuirane cijelim čujnim spektrom. Ako je većina signala u nižem spektru, korištenje uniformne kvantizacije nije učinkovita metoda jer se manje vrijednosti kvantiziraju istom veličinom koraka kao i uzorci velikih amplituda. Veća učinkovitost dobiva se neuniformnim kvantizatorom koji će kvantizirati češće uzorke preciznije od ostalih. Dizajniranje takvog kvantizatora je zahtjevan proces, no sličan efekt može se dobiti nelinearnom transformacijom koju prati uniformna kvantizacija.

Nad originalnim analognim signalom provodi se logaritamska transformacija, kompresijskom funkcijom. Kompresija je veća s povećanjem amplitude signala. Taj proces naziva se komandiranje (eng. *companding*). Nakon komandiranja slijedi standardna uniformna kvantizacija kao kod PCM-a. Prilikom reprodukcije signala, kvantizirani signal se dekvantizira i zatim se primjenjuje inverzna nelinearna transformacija koja se naziva ekspanziranje (eng. *expanding*).

Tehnika komandiranja i ekspanziranja koristi se u telekomunikaciji. Dinamički raspon glasa je dosta širok no rijetko se koristi u punom spektru. Prilikom običnog razgovora signal većinu vremena signal ostaje na relativno niskim amplitudama što pogoduje neuniformiranoj kvantizaciji. Također frekvencijski raspon glasa u prosjeku je oko 4 kHz, pa se u telekomunikaciji najčešće koristi frekvencija uzorkovanja od 8 kHz. Generalno, logari-



Slika 2.1: a) Shematski prikaz komandiranja i ekspandiranja; b) fukcije komandiranja i ekspandiranja

tamskim transformacijama potrebno je 8 bitova po uzorku da bi pokrili raspon od 12 do 14 bitova PCM-a. U širokoj upotrebi su kodeci koji koriste μ -zakon i A-zakon.

μ -zakon transformacija je kompresijska tehnika koju je specificirao CCITT (od fr. *Comité Consultatif International Téléphonique et Télégraphique*). Logaritamske je prirode i dozvoljava da uzorak od 8 bitova pokriva dinamički raspon od 14 bitova linearne kvantizacije. Frekvencija bitova koju postiže kodek s μ -zakonom iznosi 64 kbps.

μ -zakon je definiran jednađbom:

$$y = \frac{\log(1 + \mu x)}{\log(1 + \mu)} \text{ za } x \geq 0$$

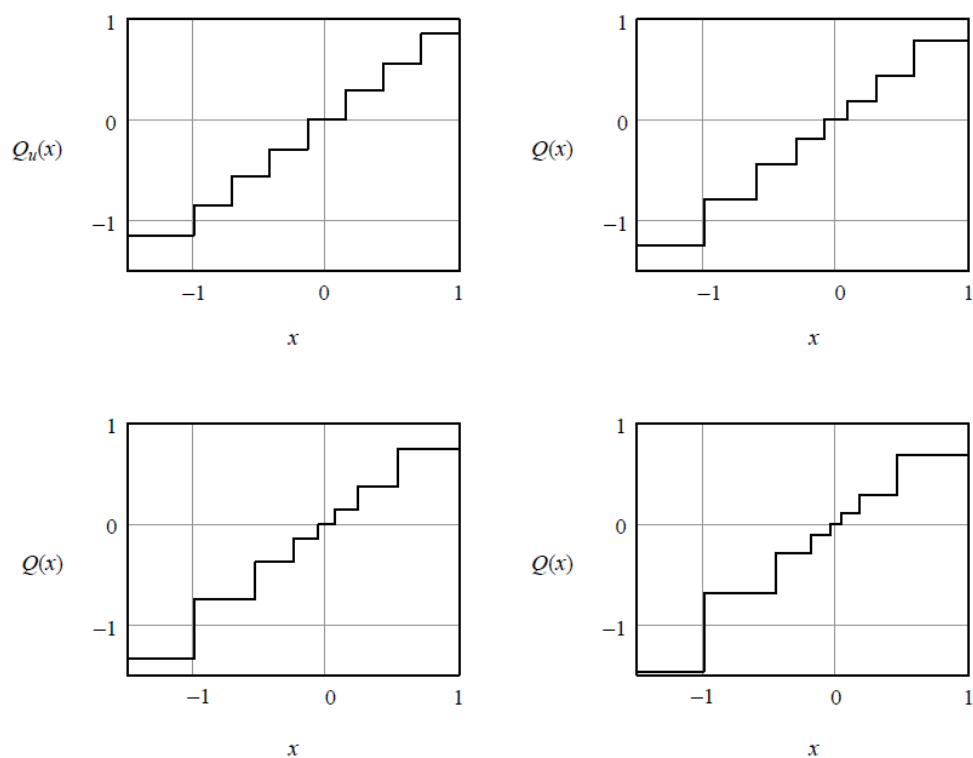
gdje je μ parametar koji određuje količinu komandiranja.

Na slici 2.2 vidimo utjecaj parametra μ na kvantizaciju kojoj dubina bitova iznosi 3, tj. kvantizaciju s 8 reprezentacijskih razina. U gornjem lijevom kutu prikazana je uniformna kvantizacija, dok ostala tri grafa prikazuju kvantizaciju kodeka s μ -zakonom. Gornji desni graf prikazuje rezultat dobiven korištenjem parametra $\mu = 2$, donji lijevi graf koristi $\mu = 4$, a donji desni $\mu = 10$. U telefoniji se koristi $\mu = 255$.

Nakon kvantizacije signal se može rekonstruirati inverznom funkcijom:

$$x = \frac{(1 + \mu)^{|y|} - 1}{\mu}.$$

μ -zakon je u uporabi u Sjevernoj Americi i Japanu, dok se u Europi koristi A-zakon. A-zakon također je logaritamska transformacija koja radi na istom principu kao i μ -zakon



Slika 2.2: Usporedba razina kvantizacije

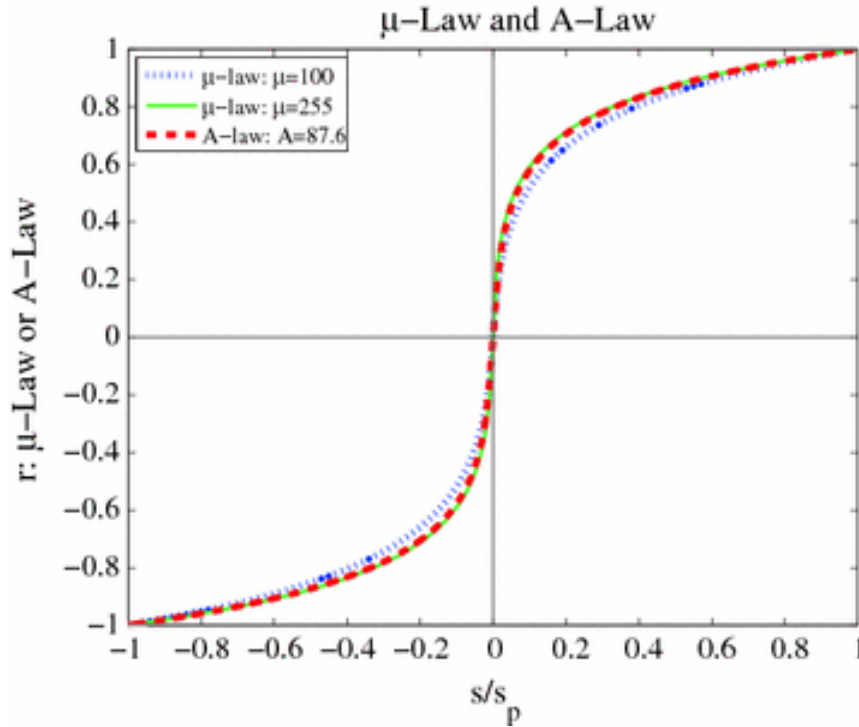
i zadana je formulom:

$$y = \begin{cases} \frac{Ax}{1+\ln A}, & 0 \leq |x| < 1/A \\ \frac{1+\ln Ax}{1+\ln A}, & 1/A \leq |x| \leq 1 \end{cases}$$

gdje je $A = 87.6$. Ekspandiranje A -zakona je zadano inverznom jednačbom:

$$x = \begin{cases} \frac{|y|(1+\ln(A))}{A}, & |y| < \frac{1}{1+\ln(A)} \\ \frac{e^{|y|(1+\ln(A))-1}}{A}, & \frac{1}{1+\ln(A)} \leq |y| < 1 \end{cases}$$

A -zakon postiže jednake rezultate kao i μ -zakon.

Slika 2.3: Usporedni prikaz μ -zakona i A-zakona

2.2 DPCM

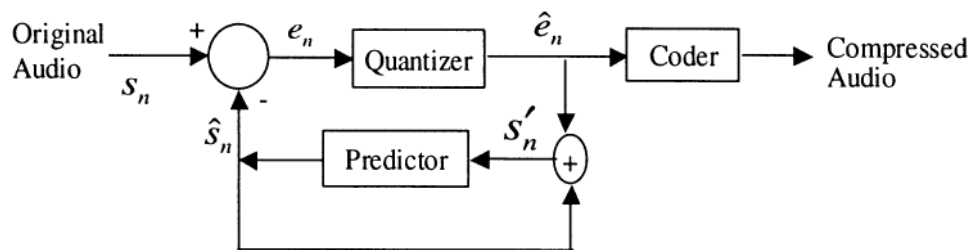
Drugi način modificiranja kvantizacije kako bi se uštedio prostor je iskorištavanje korelacije uzastopnih signala. Ako je frekvencija uzorkovanja veća od Nyquistove frekvencije, uzastopni signali imaju relativno bliske vrijednosti. Najjednostavniji oblik iskorištavanja takve korelacije, bilo bi da se umjesto signala s_n kvantizira razlika e_n između dva susjedna signala s_n i s_{n-1} , no postoji efikasniji način kojeg predstavlja DPCM.

Diferencijska pulsno kodna modulacija (eng. *differential pulse code modulation*, DPCM) je proces pretvaranja analognog signala u digitalni gdje se uzorkuje analogni signal i potom se kvantizira razlika e_n između stvarnog uzorka i njegove procijenjene vrijednosti $\hat{s}_n$.

$$e_n = s_n - \hat{s}_n$$

Razlika e_n naziva se predikcijskom greškom (eng. *prediction error*).

Vrijednost trenutnog uzorka procjenjuje se na temelju vrijednosti prethodnih uzoraka. U većini slučajeva predviđena vrijednost $\hat{s}_n$ bit će slična stvarnoj vrijednosti uzorka s_n u trenutku n . Procijenjena vrijednost dobiva se pomoću tehnike kodiranja koja se naziva linearno prediktivno kodiranje (eng. *linear predictive coding*, LPC).



Slika 2.4: Blok dijagram DPCM-a

Vrijednost $\hat{s}_n$ se dobiva iz prethodnih M vrijednosti signala s'_{n-i} gdje $i \in 1, \dots, M$ formulom:

$$\hat{s}_n = \sum_{i=1}^M \alpha_i s'_{n-i}$$

gdje su α_i , $i = 1, 2, \dots, M$ su koeficijenti prediktora. Vrijednost M naziva se red prediktora. Nakon što je izračunat $\hat{s}_n$, računa se predikcijska greška koja se potom kvantizira u vrijednost $\hat{e}_n$. Kvantizirana vrijednost se zatim šalje koderu za pohranu informacija, ali i prediktoru jer se upravo iz nje dobiva vrijednost s'_n koja će se koristiti za procjenu vrijednosti idućeg uzorka.

Kako bi se dobila bolja procjena, koeficijenti mogu biti odabrani tako da minimiziraju razliku greške e_n . Optimalni skup LPC koeficijenata $\alpha_{i,opt}$ dobiva se rješavanjem skupa uzastopnih jednadžbi:

$$\sum_{i=1}^M \alpha_{i,opt} R(j-i) = R(j)$$

gdje je $R(j)$ autokorelacijska funkcija za ulazne uzorke, definirana sa:

$$R(j) = \sum_{m=1}^N s_m \cdot s_{m+j}$$

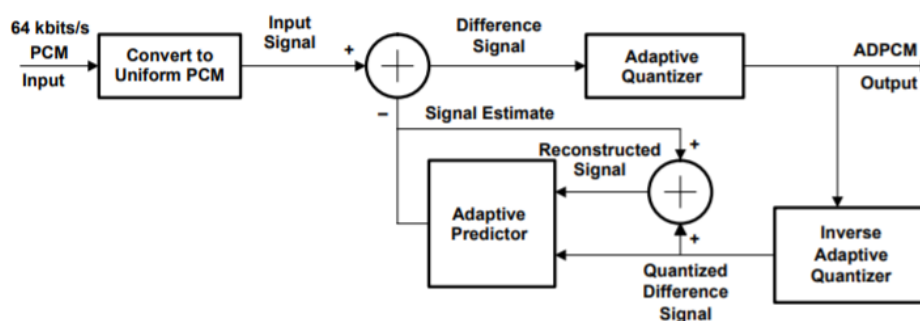
Gore navedene jednadžbe nazivaju se Yule-Walkerove jednadžbe. LPC će biti detaljnije opisan u idućem poglavlju.

Kao rezultat, podaci dobiveni DPCM-om su dva do četiri puta manji od onih dobivenih PCM-om, a pritom ne dolazi do značajne promjene kvalitete zvuka.

2.3 ADPCM

Adaptivna diferencijska pulsno kodna modulacija (eng. *adaptive differential pulse code modulation*) varijanta je DPCM-a koja omogućava promjenu raspona kvantizatora prilikom same kvantizacije. Kao i kod DPCM-a ne prenose se vrijednosti uzorka, već razlika

između stvarne i procijenjene vrijednosti uzorka, no ADPCM dinamički mijenja veličinu koraka kvantizacije. Velike razlike se kvantiziraju velikim koracima, a male razlike malim koracima. Time je količina očuvanih podataka skalirana veličinom razlika.



Slika 2.5: Blok dijagram ADPCM-a

ADPCM se koristi prvenstveno u telekomunikaciji, makar se tehnika može primijeniti i za druge tipove podataka poput slika i modemskih podataka. Prije provedbe ADPCM-a ulazni signali moraju prvo biti obrađeni PCM-om kako bi se smanjila količina bitova za obradu.

Adaptivna kvantizacija dobivena je variranjem veličine koraka prema vrijednosti kvadratne sredine ulaznog signala.

Adaptivna kvantizacija postiže se na dva načina:

- Adaptivna kvantizacija s procjenom unaprijed (eng. *adaptive quantization with forward estimation*, AQF)
- Adaptivna kvantizacija s unatražnom procjenom (eng. *adaptive quantization with backward estimation*, AQB)

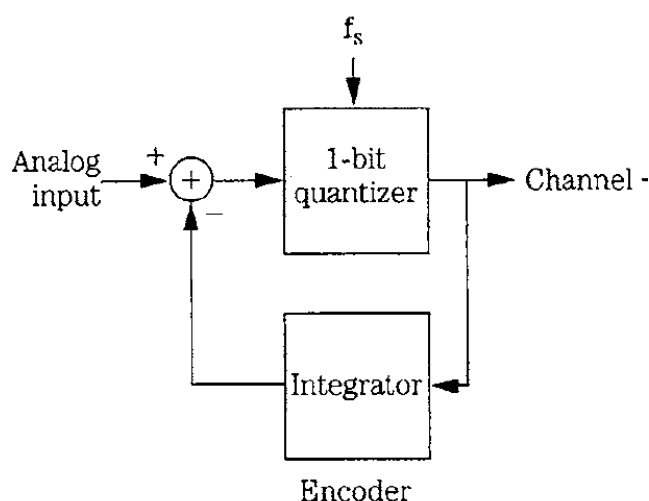
Kod AQF-a izlazni podaci podijeljeni su u blokove podataka koji se analiziraju prije kvantizacije i tada se postavljaju pripadajući parametri, a postavke kvantizatora se šalju u obliku dodatnih informacija. Nedostaci AQF-a su ti da se uvodi zastoje pri obradi bloka te je smanjena količina kompresije zbog dodatnih informacija o postavkama. Također, ako se uzimaju mali blokovi, povećava se preciznost promjena i veće su varijacije no time se povećava količina dodatnih informacija, dok se povećanjem blokova smanjuje količina informacija no ne uhvate se sve varijacije u statistici signala.

AQB koristi prethodne kvantizirane uzorke kako bi prilagodio parametre kvantizatora. Te uzorke potrebno je pamtit. Jedan od jednostavnijih modela je taj da se u obzir uzima samo zadnje kvantizirani izlazni podatak i taj kvantizator je poznat kao Jayantov kvantizator.

IMA (Interactive Multimedia Association) ADPCM je jednostavniji oblik ADPCM konverzije stvoren s ciljem pružanja dobre kvalitete zvuka uz dobru stopu kompresije, a da pritom bude dovoljno jednostavan da omogući softversku dekompresiju stereo zvuka u stvarnom vremenu, sa stopom uzorkovanja od 44.1 kHz. Jednostavnost je postignuta na način da je procijenjena vrijednost zapravo dekodirana vrijednost prethodnog uzorka. Time je procjenitelj pretvoren u sklop s vremenskim kašnjenjem (eng. *time-delay*) čiji je izlaz zapravo ulaz odgođen za jedan interval uzorkovanja. Time procjenitelj nije adaptivan čime se izbjegla potreba za slanjem sporednih informacija.

2.4 DM

Delta modulacija (eng. *delta modulation*, DM) je najjednostavniji oblik DPCM-a, i koristi samo jedan bit po uzorku. Signal se uzorkuje frekvencijom četiri ili pet puta većom od Nyquistove frekvencije čime se dobiva jaka korelacija između susjednih signala.



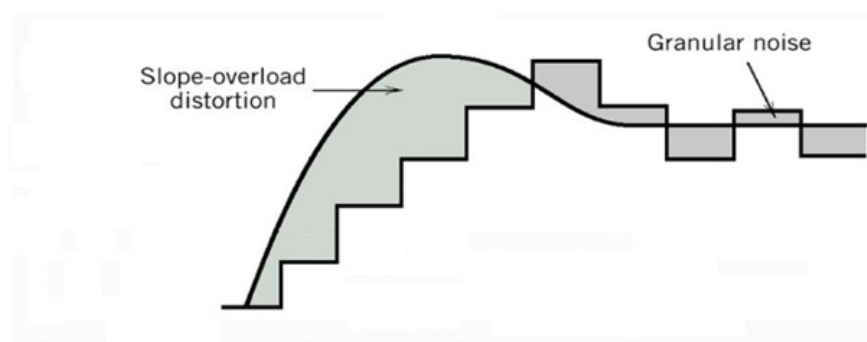
Slika 2.6: Blok dijagram DM-a

Digitizacija se dobiva tako da se umjesto procjenitelja koristi blok za vremensko kašnjenje zvan integrator, te se uspoređuju trenutni i prethodni uzorak. Ako je trenutni uzorak veći od prethodnog, kvantizira se jedinica, u suprotnom kvantizira se nula. Glavna prednost DM-a je njegova jednostavnost zbog koje je znatno smanjena hardverska kompleksnost u odnosu na PCM i DPCM kodeke.

No jednostavnost sa sobom donosi dva problema. Prvi problem je izobličenje uslijed prevelikog nagiba (eng. *slope overload*). Ono se javlja prilikom nagle promjene amplitude

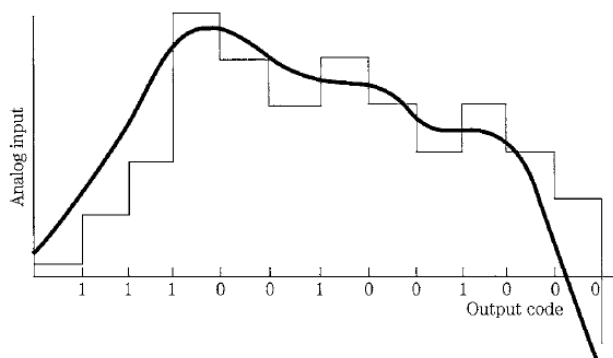
signala kada koraci kvantizacije ne mogu pratiti rast odnosno pad signala. Problem se može otkloniti još većim povećanjem frekvencije uzorkovanja.

Drugi problem javlja se kada su promjene u signalu izrazito male. Ako su promjene manje od veličine koraka kvantizacije, kvantizirat će se niz oblika 101010101. Niz nije dobra reprezentacija danog signala i prilikom reprodukcije zvuka se javlja zrnati šum (eng. *granular noise*).



Slika 2.7: Primjer izobličenja zvuka prilikom delta modulacije

Oba problema mogu se ublažiti adaptiranjem veličine koraka kvantizacije kao u ADPCM-u čime dobivamo adaptivnu delta modulaciju (eng. *adaptive delta modulation*, ADM).



Slika 2.8: Primjer adaptivne delta modulacije

Poglavlje 3

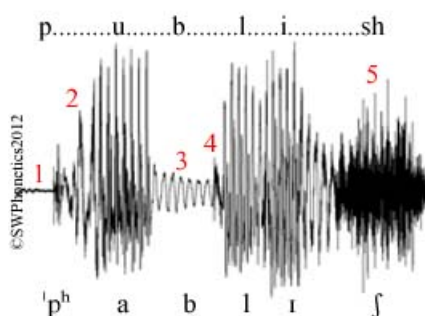
Vokoderi

Svi koderi koje smo dosada spomenuli, mogu se primijeniti na bilo koji tip signala. Vokoderi su kodeci specijalizirani za modeliranje glasa. Njihova glavna namjena je uhvatiti najbitnije govorne karakteristike glasa i pohraniti ih u što manje bitova, a da i dalje zvuk ostane razumljiv. Prilikom obrade signala, vokoderi koriste model valnog oblika u vremenu (LPC) ili rastavljaju signal na komponente frekvencija te ih potom modeliraju (kanalni vokoderi i formant vokoderi).

3.1 Kanalni vokoderi

Kanalni vokoderi prilikom obrade signala koriste potpojasno filtriranje. To je proces primjene skupa potpojasnih filtara kojima se provodi Fourierova analiza. Frekvencije se rastavljaju na svoje komponente i određuje se njihova jačina. Uz filtriranje, vokoderi provode analizu signala kako bi se odredila visina tona te je li glas zvučan ili bezvučan. Signal bezvučnih zvukova nalikuje šumu (npr. glasovi *s* i *f*) dok zvučni zvukovi imaju periodičan valni oblik. Prilikom izgovora samoglasnika zrak je izbačen kroz glasnice u obliku pravilnih kratkih dašaka koji se pojavljuju frekvencijom od 75-150 impulsa kod muškaraca i 150-250 impulsa kod žena. Osim samoglasnika, pod zvučne glasove spadaju i neki suglasnici poput *b*, *d*, *g*, *m* i *n*. Snaga zvučnih samoglasnika jača je od bezvučnih, ali slabija od samoglasnika. Na slici 3.1 se vide razlike u izgledu zvučnog vala prilikom izgovora riječi "publish". Prilikom izgovora "a" i "i" vidi se jaki periodičan signal, slabiji periodičan prilikom izgovora zvučnog samoglasnika "b", i šumoviti signal prilikom izgovora "š".

Za rekreiranje signala koriste se dva elementa. Zvučne glasove rekreira periodični pulsni generator jer se oni mogu aproksimirati sinusoidama. Budući da su bezvučni glasovi nalik šumovima, njih generira generator pseudoslučajnog šuma (eng. *pseudo-noise generator*). Sve vrijednosti se skaliraju vrijednostima dobivenim kroz skup potpojasnih filtara. Kanalni vokoder u konačnosti može proizvesti razgovijetan, ali sintetički glas koristeći 2.4



Slika 3.1: Valni oblik riječi "publish"

kbps.

3.2 LPC

LPC (eng. *linear predictive coding*) digitalna je metoda kodiranja analognog signala koja se temelji na pretpostavci da se trenutni uzorak bilo kojeg frekvencijskog signala može predvidjeti pomoću linearne kombinacije konačnog broja prošlih uzoraka. Svi iznosi prošlih uzoraka ne doprinose jednako iznosu trenutnog uzorka, pa se prije sumacije moraju pomnožiti odgovarajućim težinskim koeficijentima. Ti težinski koeficijenti su parametri ili koeficijenti linearnog prediktora. Broj tih težinskih koeficijenata ujedno određuje i red prediktora.

Cilj LPC analize je određivanje parametara linearnog prediktora tako da razlika između originalnih uzoraka i uzoraka dobivenih predikcijom bude minimalna budući da se na taj način dobiva najveća ušteda prostora. Parametri prediktora ne mogu biti konstantni u dužem vremenskom intervalu te se zato računaju nad okvirima analize signala konačnog trajanja (eng. *frame*) na kojima se mogu smatrati stalnima (tipično oko 20-30 ms).

Česta metoda koja se koristi za računanje parametara je metoda autokorelacije. Prvo se dio signala izdvaja pomoću vremenskog otvora, a zatim se izračunava vrijednost autokorelacijske funkcije. Nakon toga se iz dobivenih rezultata izračunaju vrijednosti koeficijenata prediktora rješavanjem linearnog sustava jednadžbi.

Za autokorelacijski postupak koeficijenti u jednadžbi imaju poseban oblik. Zbrajanje pogreške predikcije se provodi za indekse uzoraka pogreške predikcije od $m = 0$ do $N + p - 1$. Dobivamo idući izraz za koeficijente:

$$\Phi(i, j) = \Phi(I) = \sum_{m=0}^{N-1-I} s_n(m) \cdot s_n(m + I)$$

gdje je I razlika indeksa retka i stupca $I = i - j$. I predstavlja pomak autokorelacije segmenta govornog signala koji je izdvojen vremenskim otvorom.

Autokorelacija signala je simetrična oko nule, pa vrijedi:

$$\Phi(I) = \Phi(-I)$$

odnosno:

$$\Phi(i, j) = \Phi(j, i)$$

Koristeći nove oznake i svojstvo simetričnosti dobivamo:

$$R_n(i - j) = R_n(j - i) = R_n(|i - j|) = \Phi(i, j)$$

Sada koeficijente prediktora možemo dobiti iz jednadžbe:

$$\sum_{j=1}^p \alpha_j \cdot R_n(|i - j|) = R_n(i), \text{ za } i = 1, 2, \dots, p$$

koja se može prikazati u matričnom obliku:

$$\begin{bmatrix} R_n(0) & R_n(1) & \dots & R_n(p-1) \\ R_n(1) & R_n(0) & \dots & R_n(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ R_n(p-1) & R_n(p-2) & \dots & R_n(0) \end{bmatrix} \cdot \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} R_n(1) \\ R_n(2) \\ \vdots \\ R_n(p) \end{bmatrix}$$

Rješenja jednadžbi dobivaju se provođenjem Levinson-Durbinovog algoritma opisanim idućim pseudo-kodom:

```

i = 0;
E(0) = R(0);
while i ≠ p do
    i = i + 1;
    ki =  $\frac{R(i) - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} R(i-j)}{E^{(i-1)}}$ ; // pomoćni koeficijent
    αi(i) = ki; // najviši koeficijent prediktora reda i
    j = 1;
    while 1 ≤ j ≤ i - 1 do
        αj(i) = αj(i-1) - kiαi-j(i-1); // preostali koeficijenti
        j = j + 1;
    end
    E(i) = (1 - ki2)E(i-1); // pogreška predikcije za prediktor reda i
end
αj = αj(p); // konačno rješenje

```

Za kodiranje govora pokazalo se da je dovoljno koristiti prediktore reda 10, te se taj tip kodeka naziva LPC-10.

LPC vokoderi kodiraju određeni niz parametara koji oblikuju zvuk i glasovne karakteristike. Iz razloga što je kôd analiza zvuka, a ne zvuk sam, frekvencija bitova koju LPC koristi može biti dosta mala. Analiziraju se vremenski okviri u trajanju 22.5 ms, što odgovara broju od 180 uzoraka pri frekvenciji uzorkovanja od 8 kHz. Prvo se određuje je li trenutni segment zvučan ili bezvučan. Za bezvučne glasove, generator šuma kreira uzorke $f(n)$ koji će služiti kao ulaz za simulator vokalnog trakta (eng. *vocal-tract simulator*). Za zvučne glasove tu ulogu preuzima generator impulsnog niza (eng. *pulse-train generator*).

Označimo li izlazne vrijednosti sa $s(n)$, tada za ulaz $f(n)$ izlaz ovisi o 10 prethodnih izlaznih vrijednosti:

$$s(n) = \sum_{i=1}^{10} \alpha_i s(n-i) + G \cdot f(n)$$

gdje su α_i koeficijenti prediktora dobiveni autokorelacijskim postupkom, a G je gain faktor dobiven formulom:

$$G = R(0) - \sum_{j=1}^{10} \alpha_j R(j)$$

Visina tona P trenutnog govornog okvira dobiva se traženjem indeksa vrha (eng. *peak*). Za govorno uzorkovanje od 8 kHz, raspon pretrage najčešće je [12,140]. Ako je $v(P)$ ispod zadanog praga, tada se trenutni okvir bilježi kao bezvučan okvir i bit će rekonstruiran s primateljeve strane kao bijeli šum.

U dosta slučajeva, govorni okvir ne može u potpunosti biti označen kao zvučan ili bezvučan. Ako okvir uhvati uzorke na prijelazu zvučnoga u bezvučni, ili obrnuto, LPC model neuspješno označuje okvir. Ta netočnost dovodi do izobličenja u reprodukciji zvuka. Također, uporaba šumova i strogo periodičnih impulsa kao rezultat daje rezultat drugačiji od prirodnog glasa. Konačni izlaz je "robotizirani", sintetički glas.

LPC koderi u konačnici pružaju prijenos podataka sa malom frekvencijom bitova i razgovijetnim izlazom loše kvalitete. Kvaliteta se dodatno smanjuje pri korištenju u glasnom okruženju. Pozadinska buka povećava mogućnost krivog prepoznavanja zvučnog glasa kao bezvučnog.

Zbog niske kvalitete, LPC koderi više se ne koriste aktivno u komunikaciji, ali određene aplikacije ih i dalje koriste, prvenstveno za glasovnu sintezu kao glasovni efekt. Osnovne značajke LPC-a mogu se dodatno preraditi kako bi se postigli bolji rezultati, što će biti prikazano u idućem poglavlju.

CELP

Kodno pobuđeno linearno predviđanje (eng. *code excited linear prediction*, CELP) je složeniji vokoder koji pokušava unaprijediti lošiju kvalitetu zvuka LPC modela korištenjem kodne knjige vektora pobuda (eng. *codebook of excitation vectors*). Uzorci cijele knjige uspoređuju se sa stvarnim govorom te se šalje indeks najboljeg podudaranja. Uz korištenje kodne knjige, CELP uvodi dvije vrste predviđanja koje zamjenjuju proces analize zvučnih i bezzvučnih zvukova.

Redundancija govornog signala postiže se kroz kratkotrajno predviđanje STP (eng. *short-time prediction*) i dugotrajno predviđanje LTP (eng. *long-time prediction*). STP provodi LPC analizu signala nad nekoliko uzastopnih uzoraka poput LPC-10 kodera. Nakon STP-a na red dolazi LTP koji utvrđuje korelaciju između uzoraka udaljenih za jedan period glasa.

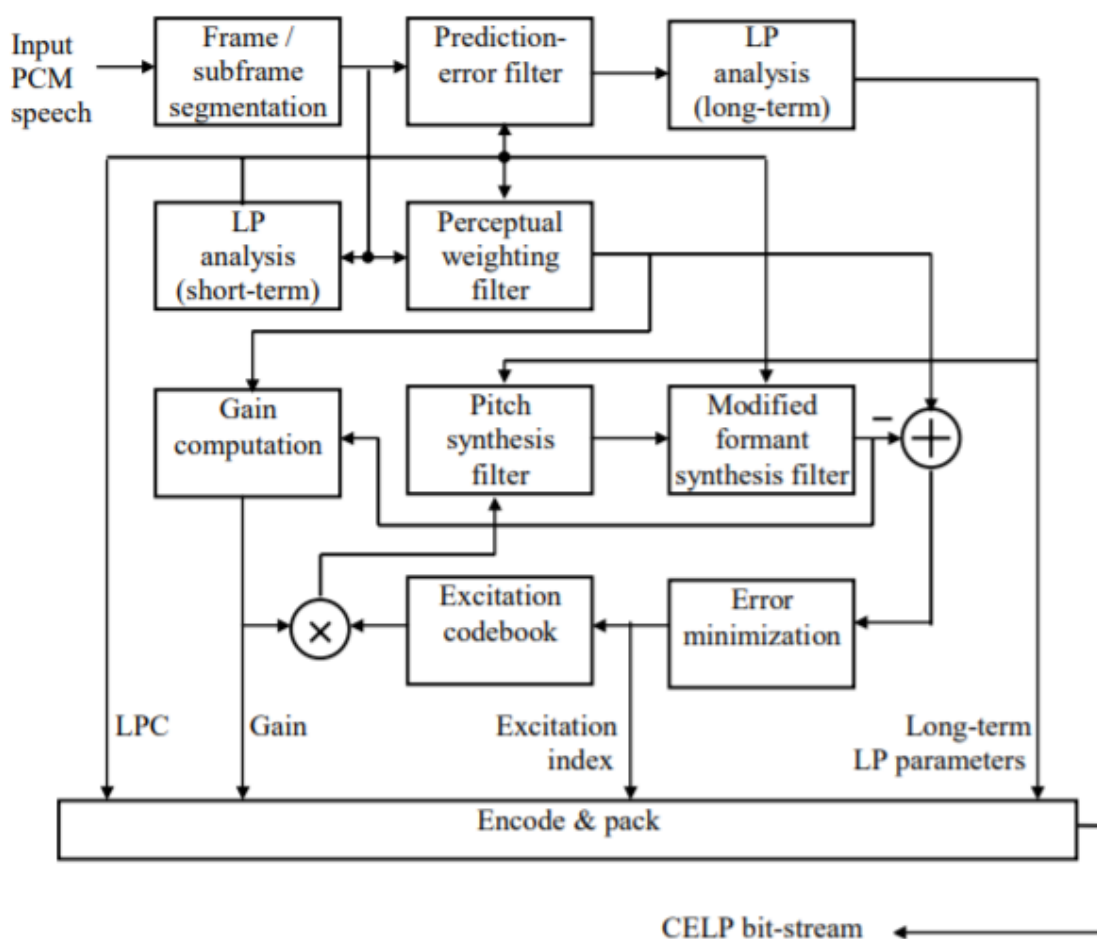
U prethodnom poglavlju vidjeli smo da se parametri koji reprezentiraju signal (pojačanje (eng. *gain*), zvučnost/bezzvučnost, visina tona) dobivaju izravno iz ulaznog signala, te se potom kvantiziraju i kasnije koriste za sintezu zvuka. Takav pristup se naziva sustav otvorene petlje (eng. *open-loop system*). Efektivniji pristup je korištenje parametara da se stvori signal prilikom kodiranja, a zatim se preciznije podešavaju kako bi generirali što vjerniju rekonstrukciju. Ta metoda naziva se optimizacija zatvorenom petljom (eng. *closed-loop system*). Budući da je signal sintetiziran prilikom kodiranja sa svrhom analiziranja, princip se još naziva "analiza-sintezom" (eng. *analysis-by-synthesis*).

U idealnoj situaciji, svi parametri bi se zajedno optimizirali za najbolji rezultat, no to nije izvedivo zbog prevelike složenosti izračunavanja. Zato se u praksi samo određen dio parametara koristi za pristup zatvorenom petljom, dok se ostatak određuje otvorenom petljom. CELP koristi analizu-sintezom gdje se najbolje podudaranje u kodnoj knjižici određuje metodom zatvorene petlje, a ostali parametri otvorenom petljom.

Ulazni signal dijeli se na okvire duljine 20 do 30 ms i podokvire dužine 5 do 7.5 ms. Zatim se provodi STP za svaki okvir kako bi se dobili koeficijenti linearnog predviđanja. Nakon toga se provodi LTP nad podokvirima.

Kada se izračunaju svi koeficijenti za filtere kreće određivanje pobudnog niza. Za svaki podokvir traži se najbolji uzorak iz kodne knjižice. Traženje počinje stvaranjem generiranjem skupa filtriranih uzbudnih nizova s odgovarajućim pojačanjima. Za svaki niz se izračuna pogreška u odnosu na uzorke te se odabire onaj s najmanjom pogreškom. U konačnici se indeks kodne knjižice, pojačanje, LTP parametri i LPC kodiraju, pakiraju i šalju kao CELP bit-stream.

Obzirom da svi govorni segmenti koriste isti skup predložaka iz kodne knjižice, preneseni govor puno je prirodniji u odnosu na rezultat LPC-10 vokodera. Pобољшanje zvuka i složenosti sa sobom donosi i povećanje frekvencije bitova na 4.8 do 9.6 kbps-a. CELP se pokazao veoma uspješnim u praksi i iz njega su proizašli mnogi standardizirani kodeci temeljeni na istim temeljnim principima.



Slika 3.2: Blok dijagram CELP koda

Glavna mana CELP-a je da je specijaliziran isključivo za kodiranje govora. Iznimno je efikasan u reprezentaciji govora, no pokazuje jako loše performanse kada je ulazni signal glazba.

Poglavlje 4

MPEG

Moving Picture Experts Group (MPEG) je radna skupina koju je oformila Međunarodna organizacija za standardizaciju (eng. International Organization for Standardization - ISO) i Međunarodna komisija za elektrotehniku (eng. International Electrotechnical Commission - IEC) kako bi se postavili standardi za audio i video kompresiju i za njihov prijenos.

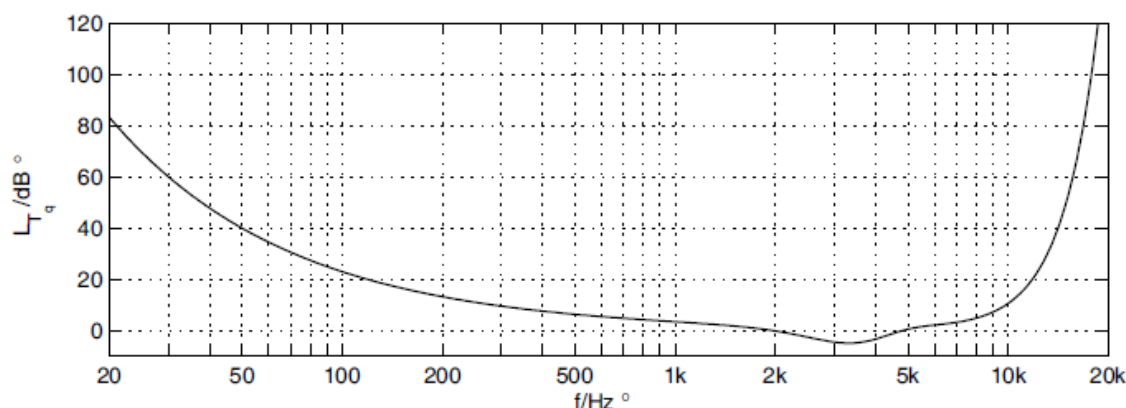
Zvukovna kompresija u MPEG standardima temelji se na iskorištavanju načina na koji ljudi percipiraju zvuk. U ovom poglavlju obradit će se MPEG-standardi te psihoakustična svojstva na kojima se oni temelje.

4.1 Psihoakustika

Psihoakustika je grana znanosti koja proučava statističke odnose između akustičnih podražaja i osjeta sluha.

Pustimo li istovremeno dva čista tona iste amplitude, ali različitih frekvencija najčešće ćemo jedan ton registrirati glasnijim od drugoga. Razlog tome je da ljudsko uho nije jednako osjetljivo na području cijelog čujnog spektra (20 Hz - 20 kHz). Izofonske krivulje (eng. *equal-loudness contour*) prikazuju odnos između opažene glasnoće (izražene u fonovima) u odnosu na razinu zvučnog tlaka za tonove različitih frekvencija. Sve krivulje su postavljene na način da je opažena glasnoća jednaka glasnoći tona frekvencije 1 kHz. Time je dobiveno podudaranje da će glasnoća od 80 fona iznositi 80 dB na frekvenciji od 1 kHz. Iz grafa je vidljivo da je ljudsko uho najosjetljivije na frekvencije u rasponu od 1 do 5 kHz.

Prag čujnosti (eng. *hearing threshold*) je najmanja razina zvuka koju je moguće čuti na danoj frekvenciji. Ljudsko uho ne može registrirati zvukove čija je razina zvučnog tlaka ispod praga ni u izuzetno tihim uvjetima. Prag čujnosti je bitan za kodiranje zvuka jer svi signali koji su manji od praga nisu bitni te ih se ne treba bilježiti. Prag je relativno visok za niske frekvencije gdje potrebna razina zvučnog tlaka od otprilike 40 dB kako bi zvuk



Slika 4.1: Prag čujnosti

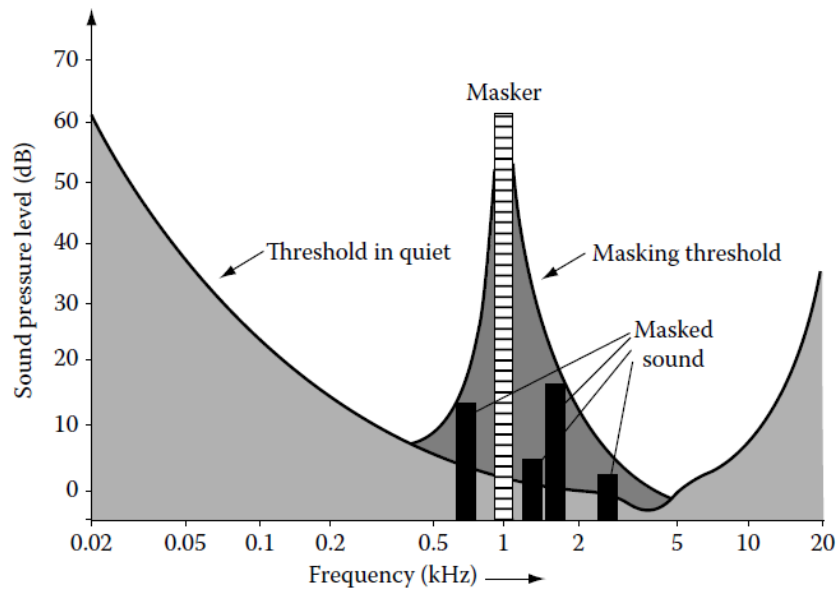
na 50 Hz postao čujan. U rasponu od 500 Hz do 5 kHz je prag oko 0 dB, nakon čega prag ponovo počinje rasti. Za frekvencije ispod 2 kHz, prag je neovisan o starosti osobe, dok je za frekvencije iznad 2 kHz vidljiv pomak praga. Na primjer, razlika u pragu za frekvenciju od 10 kHz je 30 dB između dvadestogodišnjaka i šezdesetogodišnjaka.

Osim praga čujnosti, još jedan psihoakustični efekt vezan uz glasnoću nalazi primjenu pri kodiranju audio zapisa. Maskiranje je efekt pri kojem snažniji zvuk posve prekriva neki drugi zvuk i čini ga nečujnim za slušatelja. Sa prisutnošću maskirajućeg tona mijenja se prag čujnosti, takav prag se naziva maskirajući prag. Dio praga oko maskirajućeg tona je asimetričan. Prag brže raste nego što pada, što znači da maskirajući ton prekriva više viših frekvencija nego nižih. Iz tog razloga, niži tonovi efikasnije maskiraju više tonove. Također, što je maskirajući ton jači time je širi spektar frekvencija koje maskira. Ako se maskiranje događa kada se tonovi javljaju istovremeno takvo maskiranje se naziva simultano maskiranje.

Vremensko maskiranje je naziv za maskiranje koje nastupa kada se tonovi ne javljaju u istom trenutku. Postoje dva tipa vremenskog maskiranja:

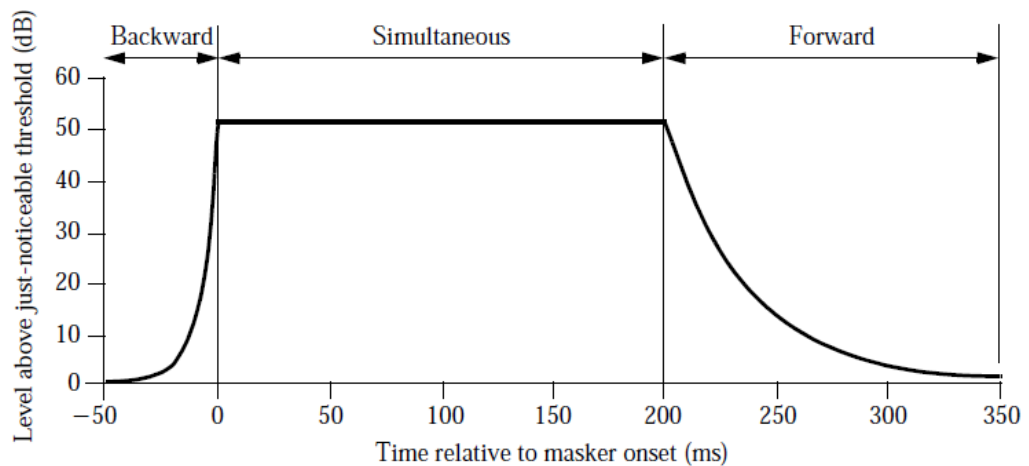
- Maskiranje unaprijed (eng. *forward masking*) javlja se nakon što zvuk nestane. Učinak na prag čujnosti je vidljiv između 100 i 200 ms u ovisnosti o jačini maskirajućeg tona.
- Maskiranje unatrag (eng. *backward masking*) javlja se kada slabiji zvuk prethodi maskirajućem tonu. Efekt se javlja zbog vremena koje je potrebno mozgu da obradi zvukove.

Ljudski zvučni spektar može se prirodno podijeliti u kritične pojaseve (eng. *critical bands*). Uho je frekvencijski selektivno i dijeli područje čujnosti na 24 pojasa (25 se ko-



Slika 4.2: Maskiranje

riste u kodiranju). Selektivnost je vidljiva prilikom razlikovanja ulaznih signala. Ako je frekvencijski razmak između ulaznih signala manji od širine kritičnog pojasa, uho ih neće razlikovati. Širina kritičnih pojasa manja je u niskom registru (100 Hz), dok za visoke frekvencije može biti šira od 4 kHz.



Slika 4.3: Vremensko maskiranje

Zbog širine raspona uvodi se nova jedinica za frekvenciju - Bark. Bark odgovara širini jednog kritičnog pojasa. Pretvorba između frekvencije i rednog broja njenog odgovarajućeg kritičnog pojasa izraženog u Barkovima zadana je na idući način:

$$b = \begin{cases} f/100 & \text{za } f < 500 \\ 9 + 4 \log_2(f/1000) & \text{za } f \geq 500. \end{cases}$$

Band #	Lower bound (Hz)	Center (Hz)	Upper bound (Hz)	Bandwidth (Hz)
1	–	50	100	–
2	100	150	200	100
3	200	250	300	100
4	300	350	400	100
5	400	450	510	110
6	510	570	630	120
7	630	700	770	140
8	770	840	920	150
9	920	1000	1080	160
10	1080	1170	1270	190
11	1270	1370	1480	210
12	1480	1600	1720	240
13	1720	1850	2000	280
14	2000	2150	2320	320
15	2320	2500	2700	380
16	2700	2900	3150	450
17	3150	3400	3700	550
18	3700	4000	4400	700
19	4400	4800	5300	900
20	5300	5800	6400	1100
21	6400	7000	7700	1300
22	7700	8500	9500	1800
23	9500	10500	12000	2500
24	12000	13500	15500	3500
25	15500	18775	22050	6550

Slika 4.4: Tablica kritičnih pojasa

4.2 MPEG-1

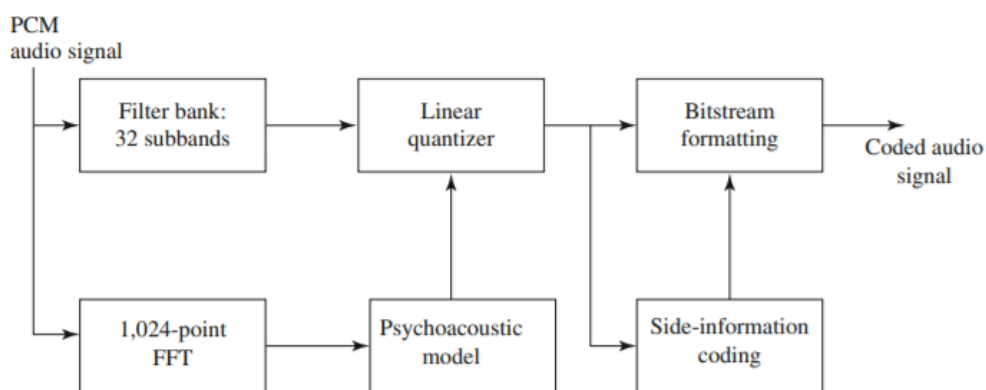
MPEG-1 standard je prvi standard za kompresiju s gubitcima za video i audio kojeg je donijela MPEG grupa 1993. godine. Audio standard sastoji se od tri sloja kompresija. Svaki idući sloj je složeniji, ali daje bolju kvalitetu zvuka. Viši slojevi grade se na temelju prethodnih i kompatibilni su s prethodnima jer imaju identično zaglavlje informacija.

Standard podržava stopu uzorkovanja od 32, 44.1 ili 48 kHz, a bit-stream podržava jedan ili dva audio kanala na jedan od četiri moguća načina:

- Monofoni način sa jednim audio kanalom.
- Dvostruki monofoni način za dva neovisna audio kanala koji je funkcionalno identičan stereo načinu.

- Stereo način za stereo kanale koji dijele bitove, ali ne koriste zajedničko stereo kodiranje.
- Zajednički stereo način koji iskorištava korelaciju između stereo kanala ili nevažnost faznih razlika između kanala, ili oboje.

MPEG-1 Layer 1



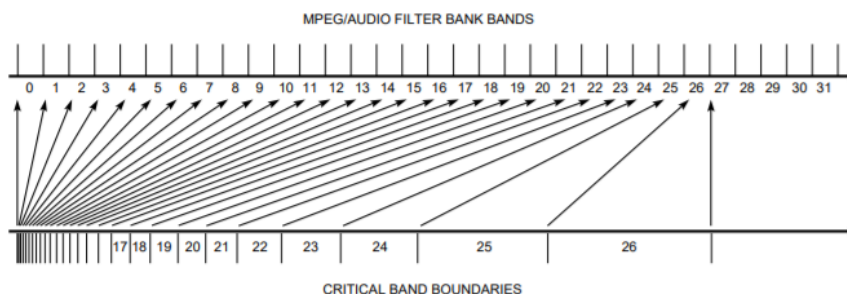
Slika 4.5: Blok dijagram MPEG-1 Layer 1 i 2

Kodiranje započinje puštanjem ulaznog signala kroz polifazni filter koji dijeli signal na 32 frekvencijska potpojasu jednake širine. Budući da su potpojasi jednake širine, podjela ne odražava kritične pojase ljudskog uha. Kod nižih frekvencija jedan frekvencijski potpojas obuhvaćat će nekoliko kritičnih pojasa, pa se broj bitova za kvantizaciju neće moći prilagoditi svakom kritičnom pojasu. Umjesto toga, kritični pojas koji ima najveću osjetljivost na šum određuje broj bitova za cijeli frekvencijski potpojas.

Svaki potpojas prima 12 uzoraka što čini jedan segment. Prilikom stvaranja segmenata javlja se kašnjenje dok se ne prikupi svih 384 uzoraka. Za svaki od 32 segmenata traži se maksimalna amplituda od 12 uzoraka i ona postaje koeficijent skaliranja za taj potpojas. Maksimum se nakon toga šalje bloku za alokaciju bitova zajedno sa uzorcima potpojasa.

Paralelno s time odvija se spektralna analiza zvuka. Psihoakustični model za Layer 1 se sastoji od idućih devet koraka.

- **FFT analiza.** Signal prolazi kroz FFT sa 512 točaka kako bi se signal prebacio iz vremenske u frekvencijsku domenu.
- **Određivanje razine zvučnog tlaka.** Za svaki potpojas izračunava se maksimalan SPL odabirom najveće amplitude spektralne linije u potpojasu.

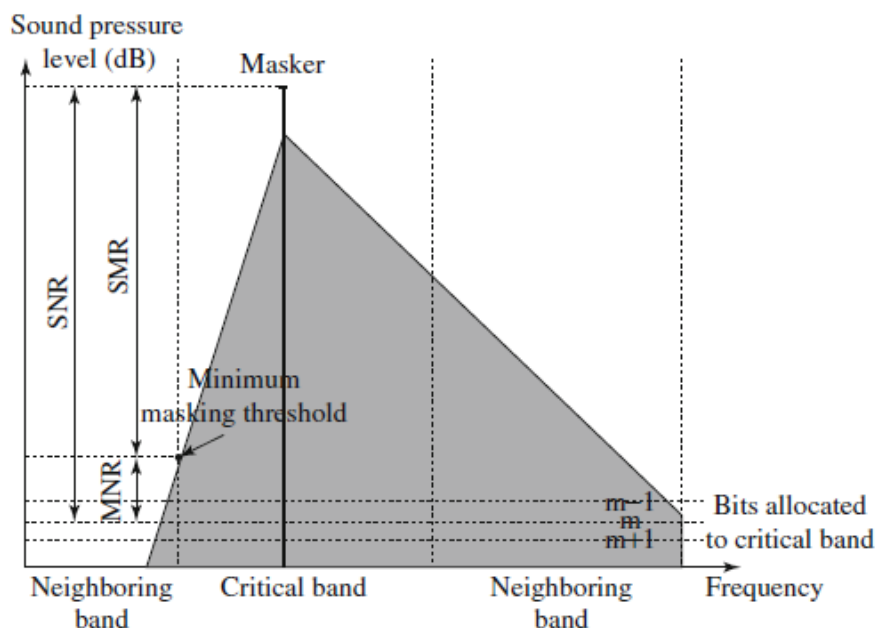


Slika 4.6: Usporedba frekvencijskih potpojasa i kritičnih pojasa

- **Minimalni prag čujnosti.** Minimalni prag čujnosti se postavlja kao donja granica za maskiranje.
- **Pronalazak zvučnih i bezzvučnih komponenata.** Određuju se sinusoidni (zvučni) i šumoviti (bezzvučni) dijelovi. Pritom se bilježi SPL zvučnih elemenata. Intenzitet preostalih komponenata se smatra bezzvučnima i zbraja se, te se potom računa njihov SPL.
- **Rezanje zvučnih i bezzvučnih maskirajućih komponenata.** Smanjuje se broj maskirajućih zvukova kako bi ostali samo oni relevantni. Relevantni maskirajući tonovi su oni čiji intenzitet prelazi minimalni prag čujnosti i zvučne komponente koje su najjače unutar 1/2 Barka.
- **Izračunavanje maskirnog praga potpojasa.** Maskirni prag dobije se na temelju zvučnih i bezzvučnih komponenata nakon što im se dodijeli maskirajući indeks i nakon toga se izračuna maskirajuća funkcija.
- **Izračunavanje globalnog maskirnog praga.** Kombinacijom podataka o pragovima potpojasa i minimalnog praga čujnosti dobiva se globalni maskirni prag koji točno određuje koje će signale ljudsko uho prepoznati, a koji će biti maskirani.
- **Određivanje minimalnog maskirnog praga za svaki potpojas.**
- **Izračunavanje SMR omjera.** SMR je omjer vrijednosti signala i maskirnog praga (eng. *signal-to-mask ratio*).

$$\text{SMR} = \text{SPL}_{\max} - \text{Minimalni maskirni prag potpojasa}$$

Izračunati SMR se potom šalje bloku za alokaciju bitova. Potpojasu može biti alocirano 0 do 15 bitova.



Slika 4.7: Prikaz MNR-a i SMR-a u potpojasu

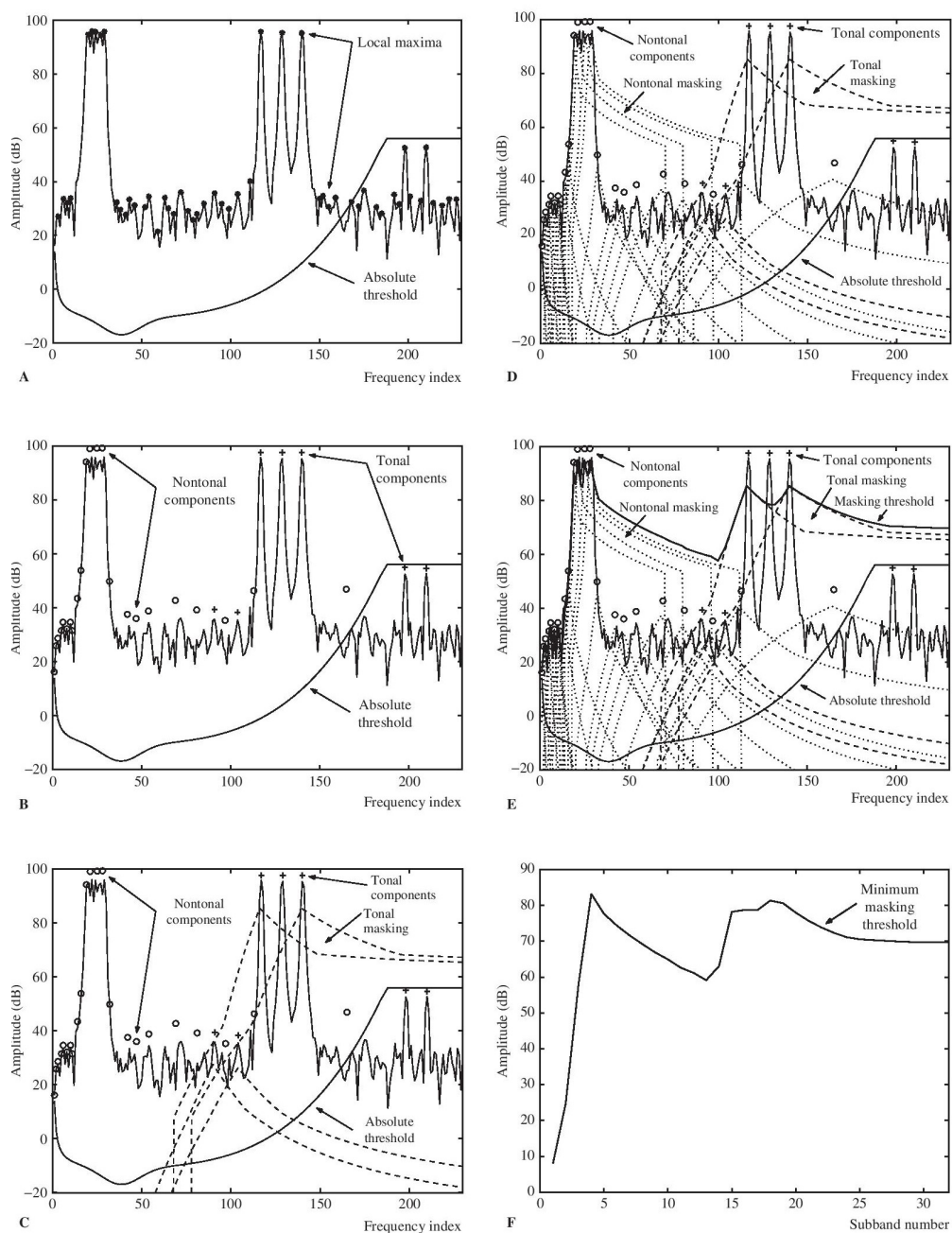
Na slici 4.8 možemo vidjeti primjer rada psihoakustičnog modela za slojeve 1 i 2.

Cilj alokacije bitova je osigurati da svi šumovi nastali prilikom kvantizacije ostanu ispod maskirnog praga.

- SMR dobiven iz psihoakustičnog modela određuje kvantizaciju, odnosno minimalni broj bitova koji su potrebni. Signal koji je iznad praga se treba kodirati, a onaj ispod se izostavlja.
- Izračuna se SNR za sve signale iz lookup tablice.
- Omjer maskirnog praga i razine šuma (eng. *mask-to-noise ratio*, MNR) definira se kao iduća razlika izražena u decibelima:

$$MNR = SNR - SMR$$

- Idući postupak se iterira dok nema više bitova za alokaciju:
 - Alocira se bit potpojasa koji ima najniži MNR.
 - Ponovo se pogleda SNR potpojasa koji ima alociran više bitova te se ponovno izračuna MNR.

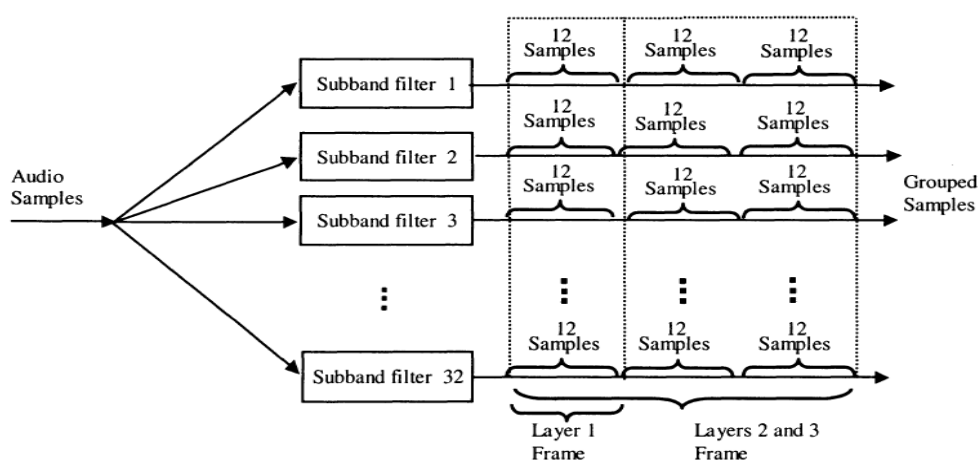


Slika 4.8: A) Lokalni maksimumi i prag čujnosti; B) Zvučne i bezzvučne komponente; C) Zvučno maskiranje; D) Bezvučno maskiranje; E) Maskirni prag; D) Minimalni maskirni prag

MPEG-1 Layer 2

Drugi sloj MPEG-1 standarda direktna je nadogradnja prvog sloja. Koder prati iste korake kao u prvom sloju uz male razlike, kako bi se smanjio frekvenciju bitova i popravio kvalitetu zvuka.

Za razliku od prvog sloja, koder drugog sloja uzima tri grupe od 12 uzoraka za jedan potpojas. Time se povećava broj uzoraka u okviru sa 348 na 1152. Koeficijent skaliranja se određuje za svaku grupu unutar segmenta, a ovisno o veličini razlike između koeficijenata, poslat će se jedan, dva ili sva tri koeficijenta.



Slika 4.9: Grupiranje uzoraka u MPEG-1 slojevima 1, 2 i 3

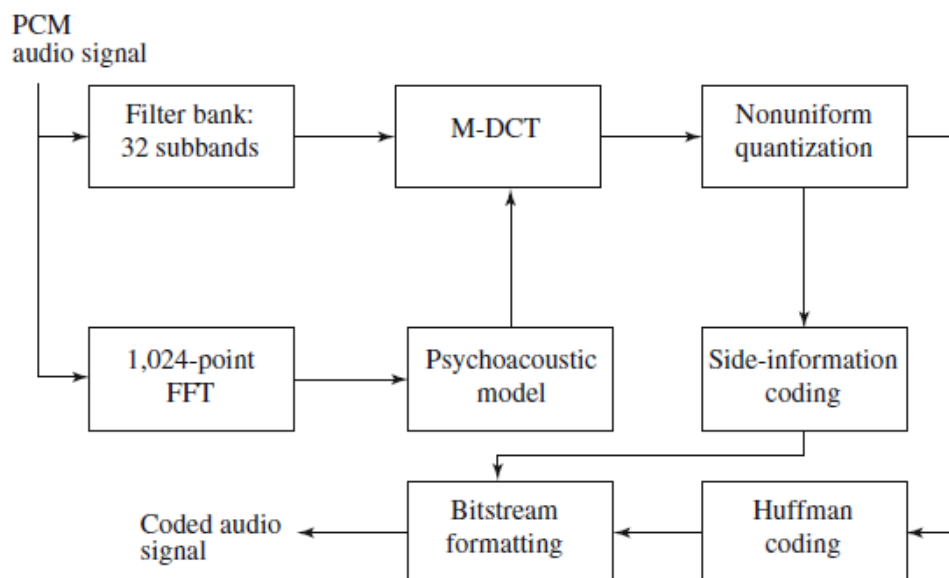
Psihoakustični model drugog sloja identičan je prvom sloju, jedina je razlika što se sada FFT provodi na 1024 točke.

Zadnja razlika javlja se prilikom kvantizacije. Uvedene su restrikcije na moguće alokacije bitova za srednje i visoke potpojase. Niskim potpojasima može biti alocirano do 15, srednjim potpojasima 7, a visokima do 3 bita. Frekvencije u potpojasima 27-31 se ne šalju. Ako slojevi koriste 3, 5 ili 9 bitova za kvantizaciju tada koder reprezentira tri uzastopne kvantizirane vrijednosti sa jednom, kompaktnijom riječju.

MPEG-1 Layer 3

Poznatiji kao MP3, treći sloj koristi sličnu frekvenciju bitova kao i slojevi 1 i 2, no proizvodi mnogo bolju kvalitetu zvuka, opet po cijenu složenosti. Shematski prikaz trećeg sloja prikazan je na slici 4.10. U trećem sloju uvedeni su novi elementi poput hibridnog polifaznog filtera, analize sintezom i neuniformnom kvantizacijom sa entropijskim kodira-

njem. Novitet je i spremnik bitova koji omogućava uštedu frekvencije bitova. Layer 3 je jedini sloj koji ima mogućnost za promjenjivo kodiranje frekvencije bitova.



Slika 4.10: Blok dijagram MPEG-1 Layer 3

Kako bi se ostvarila frekventijska podjela koja je sličnija kritičnim pojasima, 32 potpojasu se dodatno dijele na frekventijske elemente na način da se nad svakim potpojasom provede modificirana kosinusna transformacija (MDCT) s dugim blokom od 18 uzoraka ili kratkim blokom od 6 uzoraka. Između uzastopnih transformiranih prozora je preklapanje od 50 % pa su veličine prozora 36 i 12.

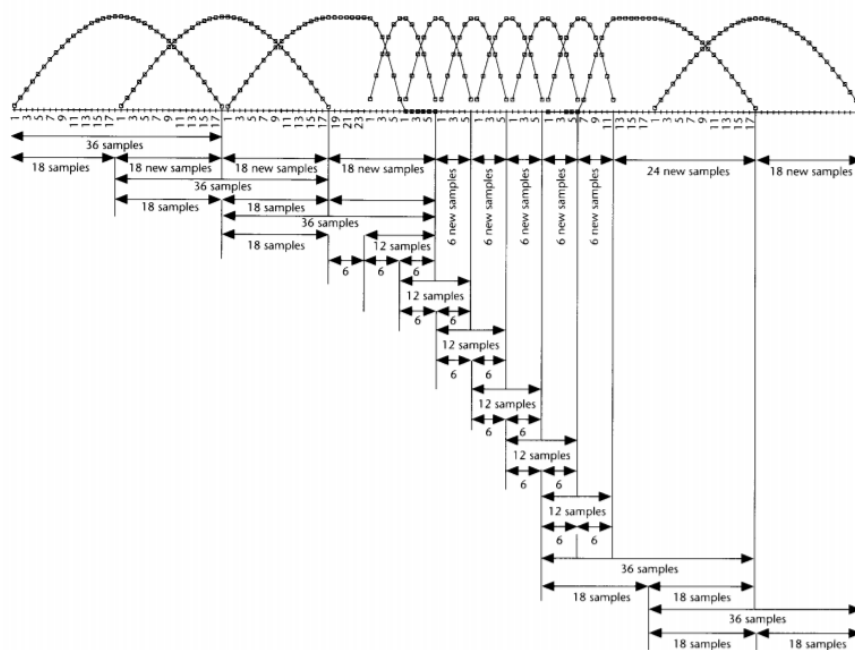
$$F(u) = \sum_{i=0}^{N-1} f(i) \cos \left[\frac{2\pi}{N} \left(i + \frac{N/2 + 1}{2} \right) (u + 1/2) \right], \text{ za } u = 0, \dots, N/2 - 1$$

gdje je broj uzoraka $M = N/2$, a N veličina prozora.

Duljina velikog bloka omogućuje bolju rezoluciju frekvenciju za signale sa stacionarnim karakteristikama dok su kratki bolji za tranzijente. Duljina kratkih blokova je trećina dugih blokova kako bi MDCT uzorci okvira ostali nepromijenjeni neovisno o duljini bloka.

Za dani okvir audio uzoraka, MDCT može imati sve blokove iste duljine, ili mogu biti pomiješani. Prilikom miješanih blokova, dugi blokovi su za MDCT donja dva frekventijska potpojasu, dok ostalih 30 viših potpojasu koriste kratke blokove. Prijelaz između dugih i kratkih blokova ne događa se odmah, već postoje posebni dugi-u-kratki i kratki-u-dugi podatkovni prozori prikazani na slici 4.11 koji omogućuju prijelaze. Vrijednosti dobivene

MDCT-om se nakon toga obrađuju kako bi se spriječilo moguće frekvencijsko preklapanje zbog preklapajućih pojava prilikom polifaznog filtriranja.



Slika 4.11: Prikaz preklapajućih MDCT prozora

Psihoakustični model trećeg sloja provodi detaljniju analizu od ranijeg modela po cijenu složenosti. Model se provodi u idućih 14 koraka i prikazan je na slici 4.12:

- **Rekonstrukcija ulaznih uzoraka.** Sastavlja se skup od 1024 ulaza.
- **Računanje kompleksnog spektra.** Vremenski signali se stavljaju u prozore putem Hannovog prozora od 1024 točaka. Računa se FFT i izlaz je reprezentiran magnitudom i fazom.
- **Računanje predviđene magnitude i faze.** Oni se određuju ekstrapolacijom iz dva prethodna bloka.
- **Računanje mjere nepredvidivosti.** Mjera nepredvidivosti mjeri koliko dobro mogu biti predviđeni složeni DFT koeficijenti unutar frekvencijskog pojasa. Računa se putem euklidske udaljenosti između predviđenih i stvarnih vrijednosti u magnitudnoj/faznoj domeni. Kako bi se smanjila kompleksnost računanje može biti provedeno samo na nižim frekvencijama.

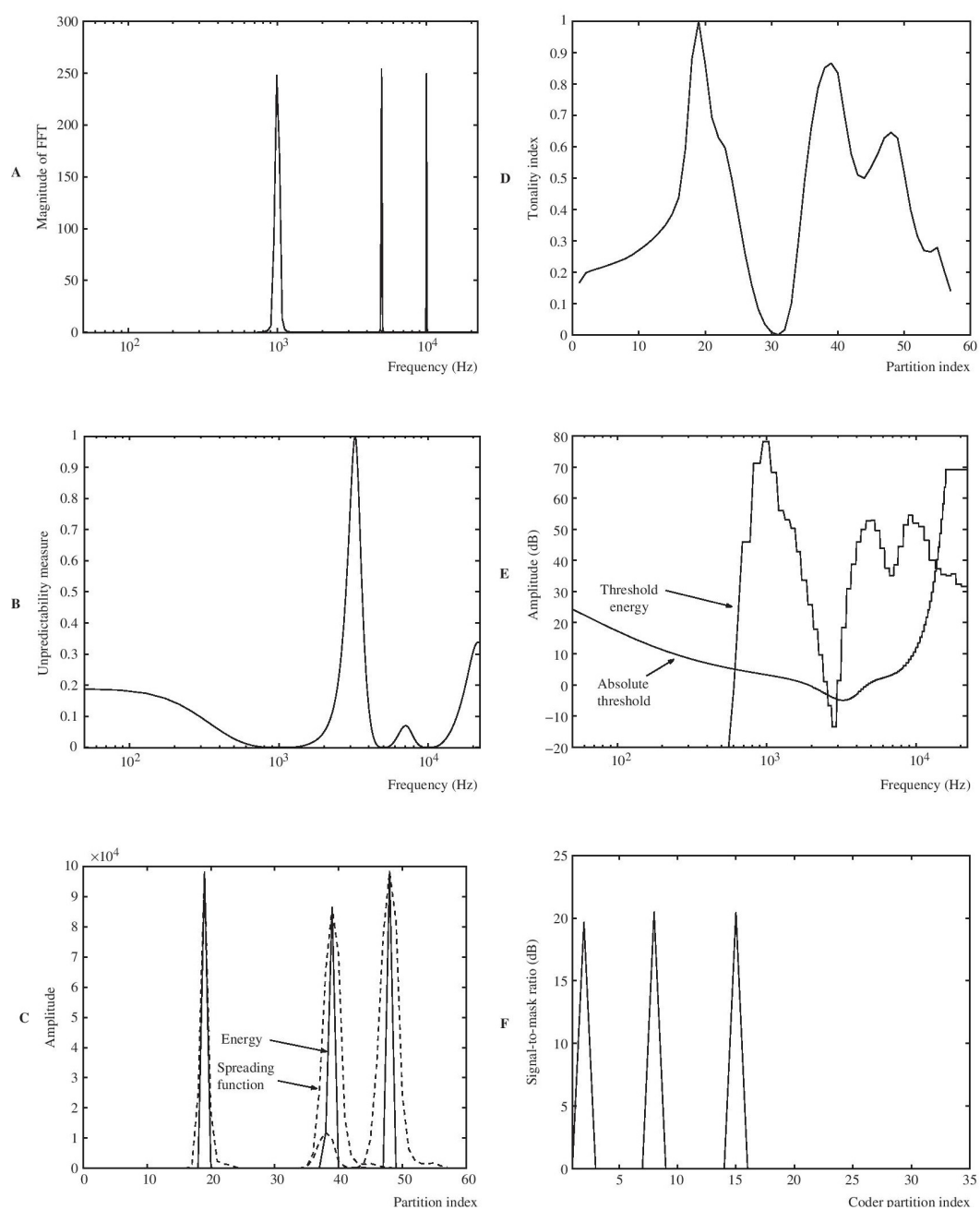
- **Računanje energije i nepredvidivosti u particijama.** Particije su širine jedne spektralne linije ili $1/3$ kritičnog pojasa, ovisno koji iznos je širi.
- **Konvolucija energije i nepredvidivosti s funkcijom širenja** (eng. *spreading function*).
- **Određivanje zvučnih indeksa.** Zvučni indeksi su u rasponu od 0 (jaka nepredvidivost) do 1 (slaba nepredvidivost). Oni određuju relativnu zvučnost maskirnih tonova prilikom računanja maskirnog praga za svaku particiju.
- **Izračunavanje potrebnog SNR-a.**
- **Izračunavanje omjera snage.**
- **Računanje praga energije za svaku particiju.**
- **Širenje energetskog praga.**
- **Računanje završne energije praga čujnosti.**
- **Računanje kontrole pred-jeke.** Tim postupkom pokušava se spriječiti čujnost greške signala prilikom filtriranja.
- **Računanje SMR-a.** SMR se računa za svaku particiju i izražen je u decibelima. Vrijednosti SMR-a se zatim šalju algoritmu za alokaciju.

Za razliku od prethodna dva sloja, Layer 3 koristi alokaciju šuma. Iterativnom petljom računa se optimalna kvantizacija za svaki potpojas. Umjesto alokacije bitova izravno iz vrijednosti SNR-a, prilikom alokacije šuma koristi se analiza sintezom kako bi se izračunao kvantizacijski spektar koji drži šumove ispod maskirnoga praga. Koriste se dvije ugniježdene petlje kako bi se izračunale dvije vrijednosti potrebne za alokaciju:

- Globalno pojačanje koje određuje veličinu koraka kvantizacije
- Koeficijenti za skaliranje koji oblikuju šum za svaki potpojas.

Treći sloj provodi neuniformu kvantizaciju. Podaci s ulaza se potenciraju sa potencijom $3/4$ prije kvantizacije kako bi se dobio konzistentniji SNR.

Još jedna novost u trećem sloju je entropijsko kodiranje koje iskorištava statistička svojstva podataka kako bi se postigla bolja kompresija. Oblik entropijskog kodiranja koje se primjenjuje u trećem sloju je Huffmanovo kodiranje. Uzorci se sortiraju prema visini frekvencije i dijele se u tri regije od kojih se svaka kodira pomoću različitog skupa tablica Huffmanovog kodiranja.

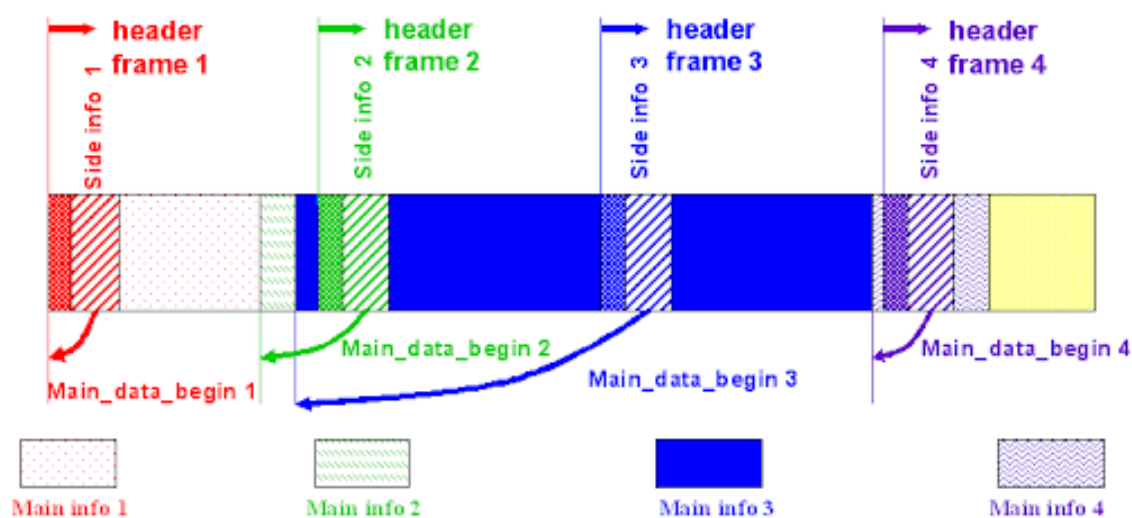


Slika 4.12: A) Vrijednosti nakon FFT; B) Mjera nepredvidljivosti; C) Energija i funkcija širenja; D) Indeks zvučnosti; E) Energija praga i prag čujnosti; F) SMR vrijednosti

Regija najviših frekvencija naziva se "*count0*" i prepoznaje se po uzastopnom nizu nula. Ta regija se ne treba kodirati jer se njena veličina može zaključiti iz veličine drugih dviju regija. Jedini uvjet koji se nameće toj regiji je da mora sadržavati paran broj nula jer preostale dvije regije grupiraju svoje vrijednosti u parnim nizovima.

Iduća regija se naziva "*count1*". Ona sadržava niz vrijednosti koje su -1, 0 ili 1. Četiri uzastopne vrijednosti gledaju se zajedno te se kodiraju pomoću Huffmanove tablice, stoga veličina regije mora biti višekratnik broja 4.

Posljednja regija je "*regija velikih vrijednosti*". Ona se ponovo dijeli na tri dijela i svaka je pokrivena jednom Huffmanovom tablicom. Odabir tablice za grupu radi se iz skupa od 16 tablica. Ona tablica koja najbolje odgovara statistici signala bit će dodijeljena toj grupi. Vrijednosti se ovdje uparuju i zatim kodiraju u ovisnosti o tablici.



Slika 4.13: Primjer spremnika bitova

Treći sloj MPEG-a koristi mehanizam koji se naziva spremnik bitova. Kodirani podaci ne moraju nužno stati u okvir fiksne duljine, već koder može donirati bitove spremniku kada je nekom okviru potrebno manje bitova od prosjeka. Ako je kasnije nekom okviru potrebno više bitova za kodiranje podataka tada on može posuditi bitove iz spremnika. Bitovi se mogu posuditi samo iz prošlih okvira, ne mogu se posuđivati unaprijed.

U konačnici MP3 ostvaruje jako dobru kompresiju u odnosu na traženu kvalitetu zvuka.

Kvaliteta zvuka	Bit-rate	Način	Omjer kompresije
Telefonija	8 kbps	Mono	96:1
Bolje od kratkih valova (eng. <i>shortwave</i>)	16kbps	Mono	48:1
Bolje od AM radija	32 kbps	Mono	24:1
Slično FM radiju	56-64 kbps	Stereo	26:1 - 24:1
Približno CD-u	96 kbps	Stereo	16:1
CD	112-128 kbps	Stereo	14:1-12:1

4.3 MPEG-2

1994. godine donesen je još jedan standard za kodiranje audio i video zapisa - MPEG-2. Novi standard uvodi definiciju zvuka sa više kanala kako bi se dobio još kvalitetniji doživljaj zvuka, pogotovo u audiovizualnim sustavima poput kino dvorana. Višekanalni sustavi pružaju mogućnost višejezičnih kanala ili dodatnih kanala za ljude sa slušnim i vidnim poteškoćama. MPEG-2.

Pri opisu višekanalnih zvučnih sustava koristi se p/q prikaz gdje p označava prednje, a q stražnje kanale. Kao najuspješniji višekanalni sustav nametnuo se 5.1 sustav koji se sastoji od tri prednja kanala (lijevi-L, desni-R i središnji-C), dva stražnja surround kanala (LS i RS) te kanala za efekte niskih frekvencija (eng. *Low frequency effects*, LFE).

MPEG-2 standard ostvaruje kompatibilnost sa prethodnim MPEG standardom pri čemu i dalje omogućava višekanalno kodiranje sa formatom MPEG-2 BC, ali donosi i nekompatibilni format kao unapređenje prethodnog standarda u obliku MPEG-2 AAC.

MPEG-2 LSF

Prvi oblik MPEG-2 standarda je MPEG-2 LSF (od. eng *low sampling frequencies*). Cilj formata je ostvariti kvalitetu jednaku ili bolju od MPEG-1 uz smanjenje količine podataka. Prvi način ostvarenja uštede prostora dobiven je smanjivanjem frekvencije uzorkovanja sa 48, 44.1 i 32 kHz kod MPEG-1 na 24, 22.05 i 16 kHz. Prepolovljavanje uzorkovanja sa sobom donosi i smanjenje uhvaćenih frekvencija u skladu sa Nyquist-Shannonovim teoremom, no gubitak viših frekvencija je prihvaćen kao razuman kompromis za određene aplikacije u zamjenu za dvostruku uštedu prostora.

Struktura LSF koda je veoma slična strukturi MPEG-1. Zadržana je troslojna struktura koja svakim slojem postaje kompleksnija. Također, omogućeno je kodiranje mono i dva stereo kanala. Zbog smanjene frekvencije uzorkovanja, potrebno je prilagoditi parametre u psihoakustičnom modelu. Veličina okvira smanjena je sa 1152 uzorka na 576 uzoraka u trećem sloju kako bi se ostvarilo lakše pakiranje za prijenos podataka internetom.

U konačnici, MPEG-2 LSF uspijeva smanjiti frekvenciju bitova koja je kod MPEG-1 standarda varirala od 32 do 224 kbps ovisno o frekvenciji uzorkovanja, na 32 do 128 kbps za prvi sloj te 8 do 80 kbps po kanalu za drugi i treći sloj.

MPEG-2 BC

MPEG-2 BC (od eng. *backward compatible*) je dio MPEG-2 standarda kompatibilan s prethodnim standardom. Format koristi 5 kanalno kodiranje, $3/2 + 1$ stereo kanal, tj. 3 prednja, dva surround kanala i niskofrekvencijski kanal koji je opcionalan. Osim takvog oblika, MPEG-2 BC može biti kodiran u $3/1$, $3/0$, $2/2$, $2/1$, $2/0$ i $1/0$ načinima.

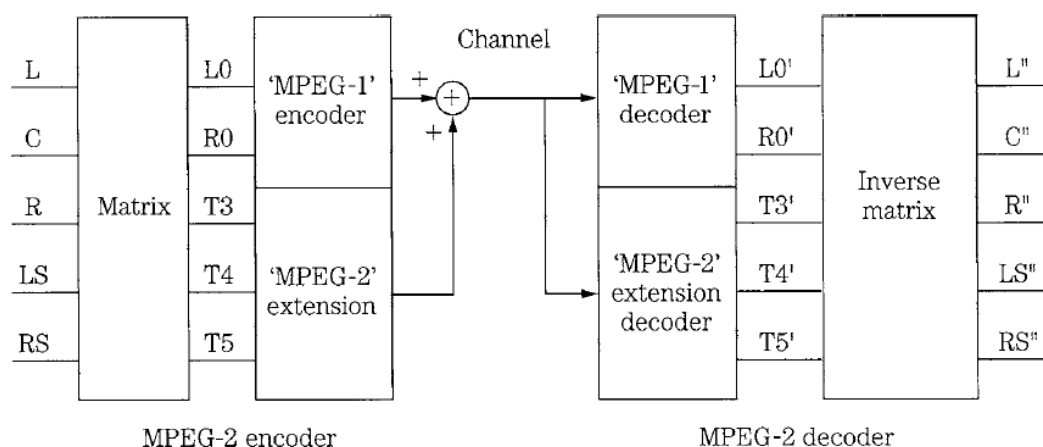
Kompatibilnost se postiže korištenjem kodirajuće matrice koja omogućuje dvokanalnim dekoderima da dekodiraju kompatibilan signal koji je podskup multikanalnog bitstreama. Tipičan oblik dobivanja kanala slijedi iz idućih jednažbi:

$$LO = \alpha(L + \beta C + \delta LS)$$

$$\alpha = \frac{1}{1 + \sqrt{2}}; \beta = \delta = \sqrt{2}$$

$$RO = \alpha(R + \beta C + \delta RS).$$

Koeficijenti α, β i δ služe za ublažavanje signala kako bi se izbjeglo prepunjenje okvira. Na slici 4.14 vidimo princip rada MPEG-2 BC-a. Signali LO i RO su poslani u MPEG-1 formatu kanalima T1 i T2, a kanali T3, T4 i T5 zajedno čine višekanalni ekstenzijski signal. Dekoder treba biti u stanju proizvesti ponovno i originalni $3/2$ stereo signal, pa se T3, T4 i T5 najčešće kodiraju kao T3 = C, T4 = LS i T5 = RS.

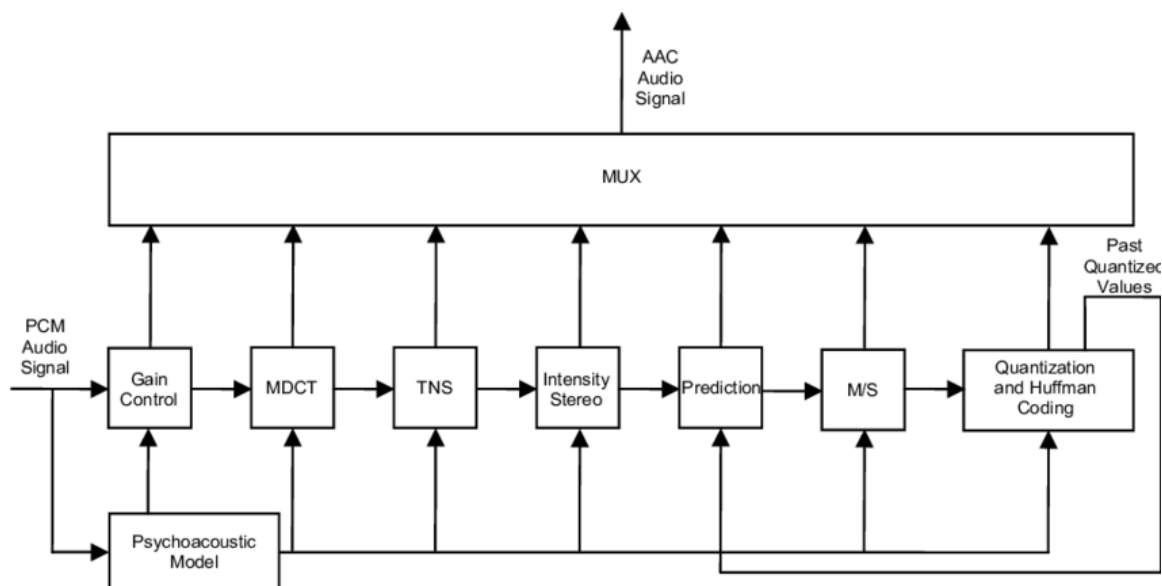


Slika 4.14: Primjer kompatibilnosti 5 kanalnog MPEG-2 BC-a sa MPEG-1

MPEG-2 BC ostvaruje frekvenciju bitova od 384 kbps.

MPEG-2 AAC

1994. godine započet je proces stvaranja standarda koji bi trebao stvoriti visoko kvalitetni višekanalni standard. Nakon tri godine dovršen je AAC (od eng. *Advanced audio coding*). AAC odbacuje kompatibilnost sa prethodnim standardom. Glavni cilj je bio dobivanje kvalitete s neprimjetnim gubitcima za ljudsko uho na frekvenciji bitova od 384 kbps ili manje. AAC uspješno zadovoljava dane zahtjeve jer uspijeva prenijeti informacije sa 320 kbps za svaki od pet kanala. Kako bi omogućio balans između kvalitete, memorijskih potreba i vremena izračunavanja, AAC ima tri različita profila: glavni profil, profil male složenosti (eng. *Low complexity*, LC) i profil s podesivom frekvencijom uzorkovanja (eng. *Scalable sampling rate*, SSR).



Slika 4.15: Blok dijagram AAC enkodera

Kodiranje započinje primjenom MDCT-a nad transformacijskim blokom od 2048 ili 256 vremenskih uzoraka. Unapređenje donosi mogućnost odabira prozorskog oblika ovisno o ulaznom signalu. Za signale sa gušćim harmonijskim sadržajem koristi se sinusni prozor, a za ostale signale se koristi Kaisler-Bessel prozor [14].

Novi koncept unutar AAC-a je blok za vremensko oblikovanje šuma (eng. *temporal noise shaping*, TNS). TNS modificira koeficijente dobivene MDCT-om kako bi se daljnje kodiranje bolje prilagodilo vremensko/frekvencijskim karakteristikama signala. Koder ima kontrolu nad kvantizacijom šuma na način da bira hoće li šum biti kodiran direktno svojim spektralnim vrijednostima ili će se nad signalom vršiti prediktivno kodiranje. Kvantizirajući šum se oblikuje u vremenskoj domeni putem linearnih prediktora u frekvencijskoj

domeni. Time se dopušta koderu kontrola nad vremenskom strukturom šuma unutar filtriranog prozora. Tehnika može biti provedena nad cijelim spektrom ili samo na dijelovima. TNS-om se uvelike povećava kvaliteta signala pri kodiranju na niskim frekvencijama bitova.

Blok za zajedničko kodiranje kanala (eng. *joint-channel coding*) omogućuje kodiranje stereo kanala i koristi dvije tehnike: kodiranje jačine i M/S (eng. *middle/side*) stereo kodiranje. Obje tehnike mogu se selektivno koristiti na različitim frekvencijama.

M/S stereo kodiranje pretvara lijeve i desne kanale u srednji i sporedni kanal na idući način:

$$M = L + R ; S = L - R.$$

Originalni signal se kasnije dobiva kao:

$$L = \frac{M + S}{2} ; R = \frac{M - S}{2}$$

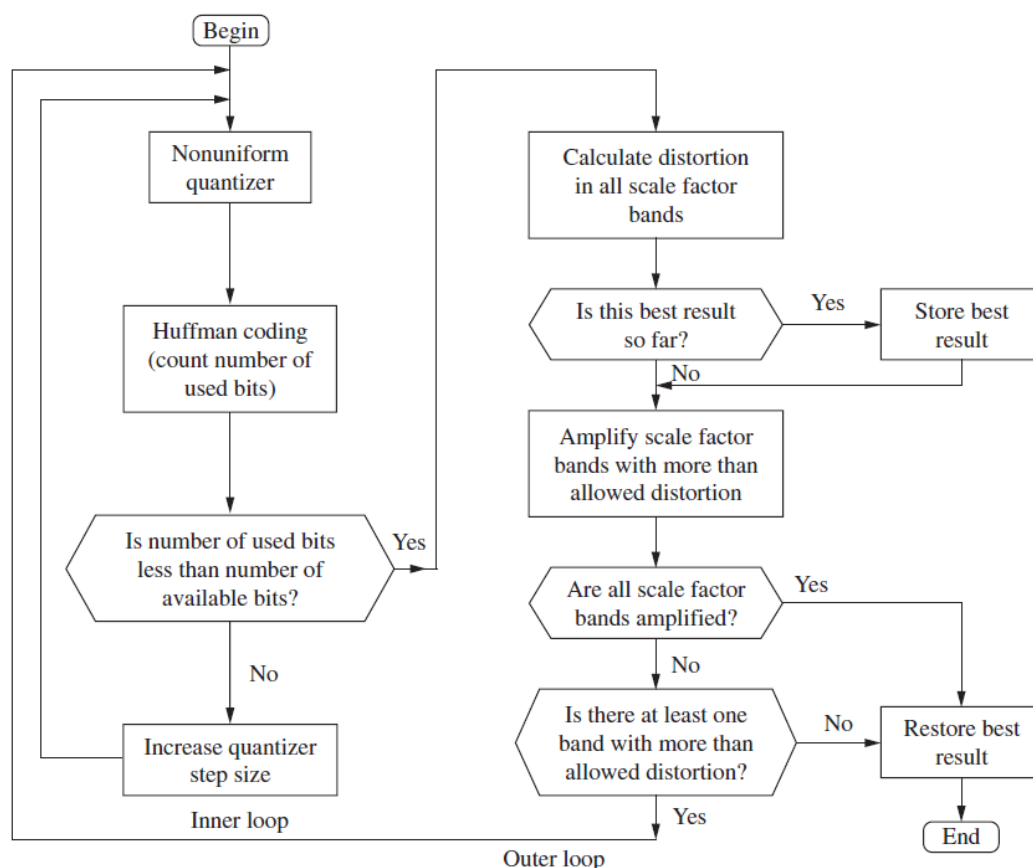
Kodiranje jačine je oblik kodiranja na principu lokalizacije zvuka koji iskorištava interauralnu vremensku razliku (eng. *interaural time difference*, ITD) i interauralanu razinsku razliku (eng. *interaural level difference*, ILD).

ITD je razlika u vremenu dolaska zvuka između dva uha i važna je za lokalizaciju zvuka jer daje znak za smjer ili kut izvora zvuka u odnosu na glavu. Ako signal dolazi s jedne strane, signal treba više proputovati da dođe do daljeg uha. ITD je izraženiji kod zvukova niske frekvencije.

ILD je razlika u glasnoći i raspodjeli frekvencija između dva uha. Kako zvuk putuje njegova snaga slabi. Ako se nalazimo kraj izvora glasnog zvuka, zvuk će biti intenzivan, no ako smo dosta udaljeni od istog tog izvora, zvuk će biti tih. Uz to glava stvara akustičnu sjenu koja mijenja glasnoću i raspodjelu zvuka koja dolazi svakom uhu. Akustična sjena izraženija je kod visokofrekventnih zvukova nego kod niskofrekventnih.

Idea kodiranja jačine je spojiti gornji spektar ulaza u jedan kanal i kao kratku dodatnu informaciju poslati kako konfigurirati određene regije frekvencija da se rekonstruiraju glasnoće kako bi se dobio stereo efekt. Rekonstrukcija originalnog signala neće biti savršena zbog gubitka podataka prilikom pojednostavljivanja stereo signala, no i dalje daje rezultate iznimno velike kvalitete.

Prediktor se primjenjuje za svaki kanal nad spektralnim komponentama dobivenim dekompozicijom kroz filter. Za svaku spektralnu komponentu do 16 kHz postoji odgovarajući prediktor. Predikcija se može i ne treba koristiti prilikom kodiranja. Ako je aktivna tada se kvantizatoru šalje greška predviđanja umjesto originalnih spektralnih informacija, čime se dobiva veća efikasnost koderu. Prediktori iskorištavaju korelaciju između spektralnih vrijednosti uzastopnih okvira. Koristi se predikcija unatrag drugog reda, tj. nova vrijednost se predviđa na temelju prethodne dvije komponente. Pritom nije potrebno slati koeficijente predikcije jer dekoder sve informacije može dobiti iz samog kvantiziranog signala.



Slika 4.16: Algoritam kvantizacije MPEG-2 AAC-a

Sama kvantizacija za cilj ima kvantizirati podatke na način da bude zadovoljen psihoakustični model. Vrijednosti koje će biti kvantizirane ovise o stopi uzorkovanja i o željenoj frekvenciji bitova. AAC kao i MP3 koristi spremnik bitova koji omogućuje raspodjelu bitova između uzastopnih blokova. Dvije umrežene petlje iterativno izvedu neuniformnu kvantizaciju i analizu sintezom. Unutarnja petlja koristi početne razine kvantizacije korištene za prikupljanje podataka te se provodi Huffmanovo kodiranje, nalik onome pri MP3 kodiranju, kako bi se odredio broj bitova potrebnih za kvantizaciju. Vanjska petlja skalira koeficijente kako bi se smanjila čujnost kvantizacijskog šuma. Najbolji dobiveni rezultat se sprema. Završetkom postupka dobije se optimalno raspoređen kvantizacijski šum preko cijelog spektra.

Iako glavni profil AAC-a daje najbolje rezultate, profil male složenosti je najrašireniji oblik AAC-a zbog svoje efikasnosti. Manja složenost je dobivena izbacivanjem predobrade

podataka i prediktora, a uz to je još i uvedeno ograničenje na red TNS-a. U odnosu na MP3, AAC bilježi rast od 30% u omjeru kvalitete naspram frekvencije bitova. Zadržana je kvaliteta zvuka, a smanjeno je opterećenje memorije i snage potrebne za izvršavanje izračunavanja.

U konačnici, AAC daje izvrsne rezultate, no njegova upotreba u praksi je ograničena. Zbog problema s licencom i nedostupnosti izvornog koda prednost ima MP3 format.

Poglavlje 5

Kompresija zvuka bez gubitaka

U prošlom poglavlju prikazani su audio formati koji pri kompresiji gube dio početnih podataka i oni su prigodni tamo gdje je potrebno dobiti najbolju kvalitetu za zadanu frekvenciju bitova. Iako u većini slučajeva MP3 i AAC audio formati odlično čuvaju kvalitetu zvuka za subjektivnog slušača, ponekad to nije dovoljno. Prilikom daljnje studijske obrade zvuka, kao što je slučaj kod miksanja glazbenih albuma, u procesu uređivanja i daljnjeg kodiranja mogu se otkriti originalno maskirana izobličenja zvuka. Kompresija zvuka bez gubitaka (eng. *lossless compression*) kodira audio signal na način da je omogućena savršena rekonstrukcija originalnog signala.

5.1 Metode kodiranja

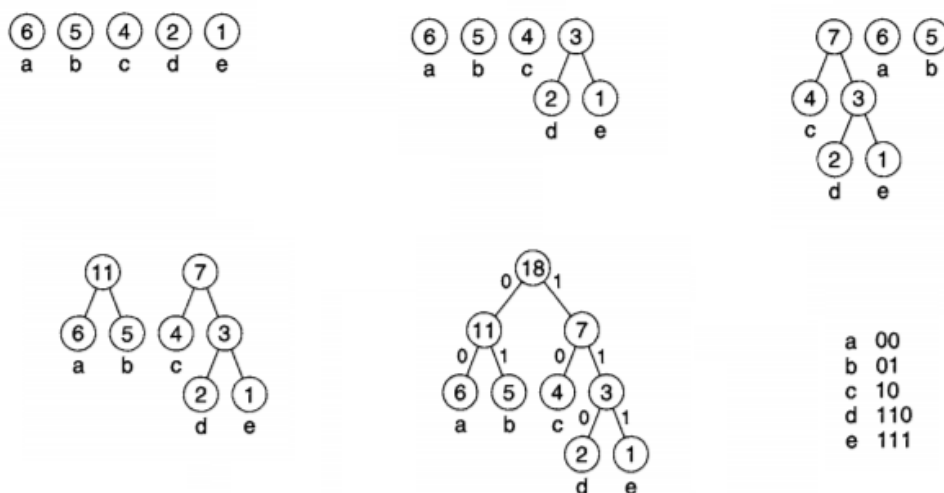
Prilikom kompresije zvuka potrebno je koristiti posebne metode kodiranja. Standardna zip kompresija postiže velike stope kompresije, no prilikom sažimanja audio datoteka ne pokazuje dobre rezultate. Kompresijom testne WAV datoteke, dobivena je ZIP datoteka veličine 9.4 MB, što je jako mala ušteda prostora u odnosu na polaznu veličinu od 10.1 MB. Razlog tako slabog rezultata leži u činjenici da algoritmi za ZIP kompresiju koriste uzorke koji se učestalo pojavljuju u podacima, no prilikom obrade audio zapisa takvi uzorci se vrlo rijetko ponavljaju. Čak i prilikom obrade zvukova čiji je valni oblik periodičan, uzorci se ne ponavljaju zbog razlike u frekvenciji perioda i frekvencije uzorkovanja signala.

Huffmanovo kodiranje

Huffmanovo kodiranje smatra se jednom od najuspješnijih tehnika za kompresiju podataka. Algoritam je razvio David Huffman 1952. godine za tekstualne datoteke, a danas čestu uporabu pronalazi pri kompresiji zvuka i slika. Kodiranje se temelji na učestalosti pojavljivanja određenih signala.

Općeniti algoritam sastoji se od idućih koraka (na primjeru tekstualnih simbola):

- Simboli se sortiraju u listi prema njihovoj učestalosti.
- Idući postupak se ponavlja sve dok ne ostane samo jedan simbol:
 - Odaberu se dva simbola sa najmanjim brojem pojavljivanja, stvori se podstablo s roditeljem kojem su odabrani simboli djeca.
 - Zbroje se vrijednosti pojavljivanja djece i pridodaju se njihovom roditelju. Nova vrijednost se zatim dodaje u listu na pripadajuće mjesto.
 - Brišu se vrijednosti djece iz liste
- Svakom čvoru se dodjeljuje kod ovisno o putu od korijena. Odlazak u lijevo dijete dodaje 0, a odlazak u desno dodaje 1 u kod.



Slika 5.1: Jednostavni primjer Huffmanovog kodiranja

Huffmanovo kodiranje zadovoljava dva bitna svojstva. Prvo svojstvo je jedinstvenost prefiksa. Niti jedan kod neće biti prefiks nekog drugog koda. Takvo svojstvo je iznimno bitno da bi se dobio efikasni dekodirer jer isključuje višeznačnost prilikom dekodiranja podataka. Drugo svojstvo je optimalnost. Huffmanovo kodiranje optimalno je za dani model jer će dva najrjeđa simbola imati kodove jednakih duljina koji se razlikuju samo u posljednjem. Simboli koji se pojavljuju češće i više puta se kodiraju, imaju kraće kodove.

Preciznije, za simbole s_i i s_j te njihove vjerojatnosti pojavljivanja p_i i p_j , ako je $p_i \geq p_j$ tada će vrijediti $l_i \leq l_j$, gdje l_k označuje duljinu koda simbola s_k u bitovima.

Tri glavna oblika Huffmanovog kodiranja su neadaptivno, poluadaptivno i adaptivno kodiranje.

Neadaptivni model najčešće se koristi pri kodiranju zvuka i spomenut je u prethodnom poglavlju. Unaprijed se konstruiraju tablice Huffmanovih kodova za različite frekvencijske pojaseve koje služe kao ogledne tablice prilikom kodiranja podataka.

Poluadaptivni model koristi dva prolaza kroz podatke. Prvi prolaz koristi se za prikupljanje informacija o učestalosti simbola, a zatim se pri drugom prolazu kodiraju podaci na temelju Huffmanove tablice dobivene u prethodnom prolazu. U ovom modelu tablica mora biti poslana zajedno s kodom kako bi dekodirer mogao uspješno dekodirati kod.

Adaptivni model koristi predikciju na temelju pročitanih simbola kako bi se izračunala učestalost daljnjih pojavljivanja. Prednost takvog modela je u omogućenom kodiranju u stvarnom vremenu no bilo kakva pogreška u predikciji dovodi do gubitka svih podataka.

Kroz godine su se razvile brojne metode i algoritmi koji nadograđuju originalnu ideju Huffmanovog kodiranja kako bi se uštedjelo na vremenu i prostoru prilikom izračunavanja. Više detalja može se pronaći u [15].

Riceovo kodiranje

Riceovo kodiranje, nazvano prema Robertu Riceu, koristi se kako bi se kodirao niz brojeva kojima duljina u bitovima varira. Cilj kodiranja nije da kodirani zapis broja bude kraći od pojedinog broja, već smanjivanje prosječne duljine svih vrijednosti u odnosu na fiksnu duljinu zapisa brojeva. Duljina koda za broj n i Riceov parametar S iznosi $S + \lfloor n/2^S \rfloor + 1$.

Kodiranje dijeli broj na dva ili tri dijela, ovisno o potrebi prvog bita koji određuje predznak broja. Algoritam se sastoji od dva dijela. Prvo se odvoji S najmanje značajnih bitova broja n koji postaju najmanje značajni bitovi samog koda, a zatim se broj $k = \lfloor n/2^S \rfloor$, dobiven micanjem zadnjih S bitova od n , kodira kao niz od k nula praćenih jedinicom ili obrnuto.

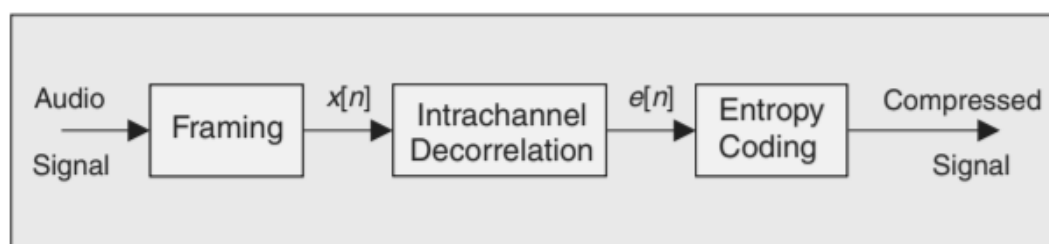
Kao rezultat, za manji S manji brojevi će biti efikasnije kodirani od velikih, dok za veći S veliki brojevi će biti relativno efikasno kodirani, a manje vrijednosti će se kodirati s puno većom količinom bitova od potrebnog.

Riceovo kodiranje optimalno je za Laplaceovu distribuciju, što je dovelo i do korištenja metode pri kompresiji audio podataka jer reziduali audio signala često prate Laplaceovu distribuciju. U idućoj tablici prikazan je primjer Riceovog kodiranja za $n \geq 0$ i n' koji može biti i negativan.

n	n'	$S = 0$	$S = 1$	$S = 2$	$S = 3$	$S = 1$ sa bitom predznaka za n'
0	0	1	1 0	1 00	1 000	0 1 0
1	1	01	1 1	1 01	1 001	0 1 1
2	-1	001	01 0	1 10	1 010	1 1 1
3	2	0001	01 1	1 11	1 011	0 01 0
4	-2	00001	001 0	01 00	1 100	1 01 0
5	3	000001	001 1	01 01	1 101	0 01 1
6	-3	0000001	0001 0	01 10	1 110	1 01 1
7	4	00000001	0001 1	01 11	1 111	0 001 0
8	-4	000000001	00001 0	001 00	01 000	1 001 1

Općeniti principi kompresije bez gubitaka

Kompresija se provodi u tri glavne faze. U prvoj fazi, signal se dijeli na okvire, zatim se uklanja unutarkanalna korelacija i na kraju se izvršava entropijsko kodiranje. Svaki audio kanal prolazi iste tri faze i zasebno se kodira.



Slika 5.2: Tri faze kodiranja za kompresiju bez gubitaka

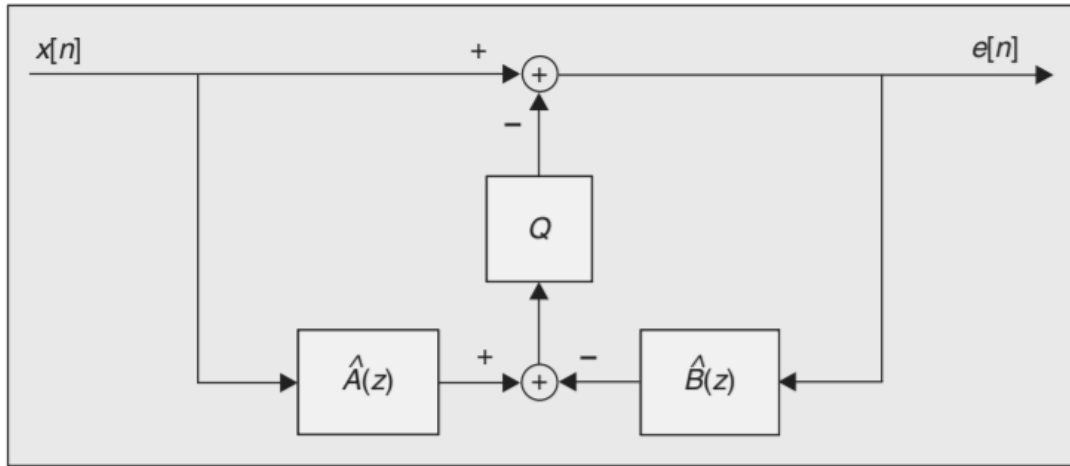
Na početku, ulazni signal se razbija na manje dijelove i slaže u okvire, čime se omogućuje brže i lakše uređivanje podataka. Također, time je omogućeno mijenjanje parametara kompresije iz okvira u okvir, što dovodi do boljih rezultata. Svaki okvir u svom zaglavlju prenosi informacije o parametrima kompresije te može sadržavati dodatne informacije o sinkronizaciji s ostalim podacima. Okviri najčešće sadržavaju 576 ili 1152 uzorka pri frekvenciji uzorkovanja od 44.1 kHz, što u vremenskom trajanju iznosi 13 do 26 ms.

Druga faza je uklanjanje korelacija uzoraka unutar okvira. Većina algoritama koristi neki oblik linearne predikcije kako bi modificirala signal. Princip je dosta sličan prediktorima koje smo susreli u poglavlju o LPC koderima. Predikcija se provodi nad uzorcima $x[n]$ prilikom čega se dobivaju predviđene greške $e[n]$ koje su prosjeku manje od originalnih uzoraka i time će biti potrebna manja količina bitova za pohranu.

Model korišten pri kompresiji prikazan je na slici 5.3. Greška se dobiva na idući način:

$$e[n] = x[n] - Q\left\{\sum_{k=1}^M \hat{a}_k x[n-k] - \sum_{k=1}^N \hat{b}_k e[n-k]\right\}$$

gdje Q označava kvantizaciju novodobivenih signala, a M i N redove prediktora.



Slika 5.3: Blok dijagram modela prediktora

Operacija kvantizacije čini prediktor nelinearnim i potrebna je za kompresiju bez gubitaka jer je zbog nje moguće rekonstruirati $x[n]$ iz danog $e[n]$ i na računalima koji imaju drugačiju arhitekturu. Za rekonstrukciju potrebno je poslati i koeficijente prediktora. Rekonstruirani signal dobije se iz iduće formule:

$$x[n] = e[n] + Q\left\{\sum_{k=1}^M \hat{a}_k x[n-k] - \sum_{k=1}^N \hat{b}_k e[n-k]\right\}$$

Zadnja faza kompresije sastoji se od entropijskog kodiranja koje uklanja redundantne podatke iz signala $e[n]$. Entropijsko kodiranje provodi se Huffmanovim i Riceovim kodiranjima koji su opisani na početku ovog poglavlja.

U nastavku će se opisati dva audio formata koji koriste kompresiju bez gubitaka.

5.2 FLAC

FLAC (od eng. *free lossless audio codec*) je audio format kojeg je stvorio Xiph.Org Foundation. Kodek je otvorenog formata i ne naplaćuje se licenca za njegovo korištenje što

je uvelike doprinijelo njegovoj popularnosti. Većina standardnih audio programa poput Windows Media Player-a, VLC player-a i GOM player-a podupiru FLAC kao tip formata.

Kodiranje podataka u FLAC-u prolazi standardne korake audio kompresije bez gubitaka. Ulaz se dijeli na blokove čija duljina može varirati, što ima izravan utjecaj na stopu kompresije. Ako je veličina bloka premala, okviri će većinom sadržavati podatke zaglavlja. Ako su blokovi preveliki neće biti moguće pronaći adekvatni prediktor zbog prevelikog variranja svojstava signala. FLAC propisuje minimalnu veličinu bloka koja iznosi 16 uzoraka i maksimalnu veličinu od 65535 uzoraka.

Uklanjanje unutarkanalnih korelacija provodi se za stereo kanale. FLAC dozvoljava da enkoder odabere najbolju reprezentaciju, okvir po okvir. Bira se jedna od četiri opcije:

- Nezavisni kanali - lijevi i desni kanal se kodiraju zasebno.
- M/S - lijevi i desni kanal se transformiraju u srednji i sporedni kanal. Srednji kanal dobije se kao prosjek lijevog i desnog kanala, a sporedni kao razlika lijevog i desnog.
- L/S - kodiraju se lijevi i sporedni kanal.
- R/S - kodiraju se desni i sporedni kanal.

Za predikciju FLAC koristi jednu od četiri metode:

- **Doslovno.** Prediktor nultog reda koji jednostavno propušta sve podatke. Predikcija je u svakom trenu 0 te se kodira sam signal.
- **Konstanta.** Predviđena vrijednost je konstantna i koristi se kada blok na svim uzorcima ima istu vrijednost.
- **Fiksni linearni prediktor** Prediktori reda 0, 1, 2, 3 ili 4. Budući da su prediktori fiksni, dovoljno je poslati broj reda kao dodatni parametar kako bi dekodirao signal.
- **FIR linearni prediktor.** Računaju se LPC koeficijenti iz autokorelacijskih koeficijenata. Pomoću njih se kodiraju audio podaci, a zatim se koeficijenti šalju zajedno s kodom kao dodatne informacije. Maksimalan red prediktora kojeg FLAC podržava je 32. Ovaj pristup daje najbolju predikciju, no povećava vrijeme izvršavanja.

Novodobiveni signal se zatim kodira. FLAC iskorištava činjenicu da signal greške ima Laplaceovu distribuciju. Takvi signali efikasno se kodiraju Riceovim kodovima i koristeći jedan od dva načina. Prvi način je da koder odredi jedan Riceov parametar i nakon čega se cijela greška kodira tim parametrom. Drugi način je da se greška podijeli na manje uzorke i da se potom svaki od njih kodira zasebnim parametrom.

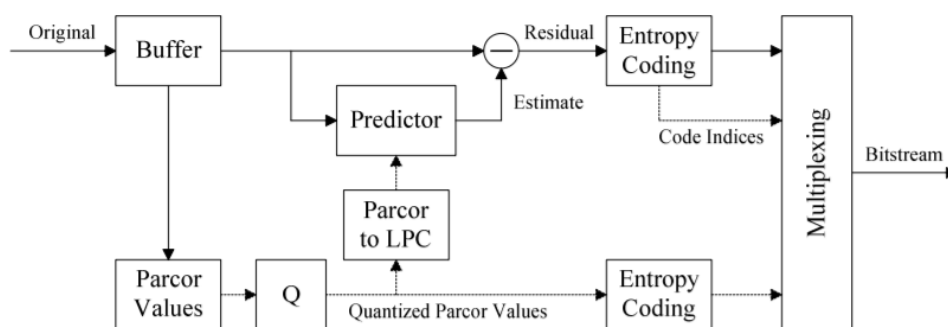
Krajnji rezultat je FLAC datoteka koja u prosjeku zauzima 55-75% originalne WAV datoteke. U idućoj tablici prikazani su rezultati FLAC kompresije nad WAV datotekom trajanja 3 min i 43 sekunde, veličine 37.65 MB sa različitim postavkama enkodera.

Postavke	Veličina	Veličina u odnosu na početnu datoteku	Vrijeme kodiranja (s)	Vrijeme dekodiranja (s)
FLAC -0	25.54 MB	67.826 %	1.692	1.422
FLAC -5	23.12 MB	61.402 %	4.035	1.732
FLAC -6	23.12 MB	61.398 %	4.286	1.632
FLAC -7	23.11 MB	61.377 %	8.642	1.792
FLAC -8	23.06 MB	61.247 %	11.276	1.762
FLAC -8 -Ax2	23.05 MB	61.214 %	18.446	1.572

Slika 5.4: Primjer FLAC kompresije

5.3 MPEG 4 - ALS

ALS (od eng. *audio lossless coding*) je dodatak MPEG 4 standardu zamišljen da postane međunarodni standard za kodiranje zvuka bez gubitaka. Prva verzija standarda objavljena je 2006. godine, a zadnja dorađena verzija iz 2009. godine je u potpunosti integrirana u MPEG-4 standard.



Slika 5.5: Blok dijagram MPEG-4 ALS kodera

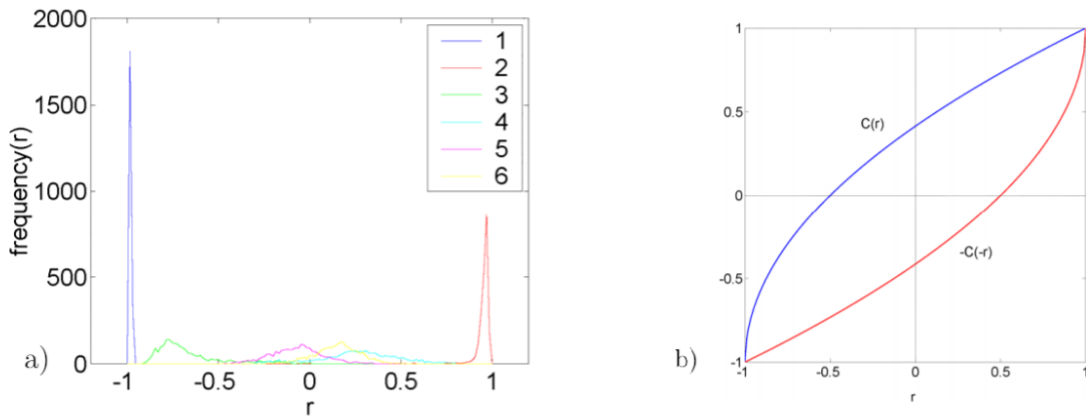
Ulazni podaci dijele se u okvire unutar kojih kanal može biti ponovno podijeljen u manje blokove za daljnju obradu. Za svaki blok se izračunaju koeficijenti refleksije, parcor koeficijenti, koji omogućuju alternativnu formu LPC-a. Nad svakim blokom provodi se linearna predikcija kojom se dobivaju reziduali za daljnje kodiranje. ALS koristi adaptivnu predikciju unaprijed što znači da se za svaki blok procjenjuju optimalni koeficijenti predviđanja.

Prilikom predviđanja, red prediktora može varirati od bloka do bloka. Promjena reda se provodi kako bi se smanjila količina dodatnih informacija prilikom slanja koeficijenata. Veći red daje bolju predikciju pa su kodirani podaci manji, ali sa sobom donosi veću količinu dodatnih podataka zbog povećanja broja koeficijenata. Optimalan red efikasno se pronalazi Levinson-Durbinovim algoritmom. ALS podupire predikcije do reda 1023.

Koeficijenti predikcije se ne kvantiziraju izravno jer njihov prijenos nije dovoljno efikasan. Kvantizacija se izvršava preko ranije spomenutih koeficijenata refleksije r_k . Koeficijenti se restringiraju na interval $[-1,1]$. Frekvencije r_1 i r_2 puno su veće od ostalih koeficijenata r_k , $k > 2$ pa se oni zato dodatno komandiraju idućom funkcijom:

$$C(r) = -1 + \sqrt{2} \sqrt{r+1}$$

$C(r_1)$ daje bolje rezultate za prvi koeficijent, a $-C(-r_2)$ za drugi.



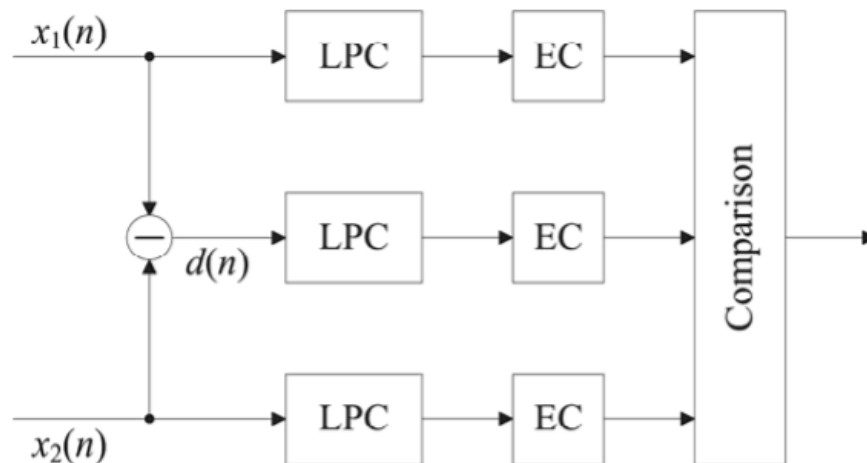
Slika 5.6: a) Distribucija koeficijenata refleksije $r_1, \dots, r_6$ za 48kHz, 16-bitni audio. b) Funkcija komandiranja $C(r)$

Nakon komandiranja prva dva koeficijenta slijedi 7-bitna uniforma kvantizacija:

$$a_k = \begin{cases} \lfloor 64 \cdot C(r_1) \rfloor & \text{za } k = 1 \\ \lfloor 64 \cdot C(-r_2) \rfloor & \text{za } k = 2 \\ \lfloor 64 \cdot r_k \rfloor & \text{za } k > 2 \end{cases}$$

Radi pojednostavljenja izračunavanja, prilikom kvantizacije drugog koeficijenta koristi se $+C(-r_2)$, pa je komandirana vrijednost suprotnog predznaka. Prilikom dekodiranja koeficijenti predikcije se dobivaju iz kvantiziranih vrijednosti a_k transformacijskom funkcijom.

Ako se kodira stereo signal, ALS koristi zajedničko kodiranje kanala kako bi postigao dodatnu kompresiju. Prva metoda je kodiranje razlike $d(n)$ između kanala x_1 i x_2 . Za svaki od kanala i njihovu razliku $d(n) = x_2(n) - x_1(n)$ provodi se LPC faza i entropijsko kodiranje, te se nakon toga uspoređuju dobiveni rezultati. Ako je postignuta određena razina kompresije, kodira se razlika, u suprotnom se kanali kodiraju zasebno. U slučaju da ulaz ima više od dva audio kanala, koder sparuje one kanale koji daju najbolje rezultate.



Slika 5.7: Blok dijagram zajedničkog kodiranja signala njihovom razlikom

Druga metoda je MCC (od eng. *multi-channel coding*). MCC može se koristiti kao dodatak kodiranju razlika kanala ili kao zasebna metoda. Kodiranje se izvršava nad rezidualima nakon što podaci prođu postupak predikcije. Metoda se temelji na ideji da svaki kanal $e_c(n)$ može biti predviđen koristeći bilo koji drugi kanal različit od njega. Kao i kod kodiranja razlika, za svaki blok se gleda je li postignuta bolja kompresija te se odlučuje hoće li se kodirati podaci dobiveni MCC-om ili originalni reziduali. Ako je odabrano korištenje MCC-a tada se kao dodatna informacija treba poslati informacija koji se kanal koristio za predikciju.

Zadnja faza ALS-a je entropijsko kodiranje reziduala $e(n)$. Kodiranje se provodi korištenjem Riceovog kodiranja ili blokovnog Gilbert-Moore koda (BGMC) [13]. Riceovo kodiranje se može provesti s jednim zajedničkim Riceovim parametrom S ili se blok može rastaviti na podblokove od kojih svaki ima različiti parametar S_i .

Za zadani parametar S i vrijednosti $e(n)$, izlazni podaci kodiranja se dobivaju na idući način:

Ako je $S = 0$

$$k(n) = \begin{cases} 2 \cdot e(n) & \text{za } e(n) \geq 0 \\ -2 \cdot e(n) - 1 & \text{za } e(n) < 0 \end{cases}$$

Kodirana vrijednost će se sastojati od $k(n)$ nula iza kojih dolazi jedinica.

Ako je $S > 0$

$$k(n) = \begin{cases} \lfloor e(n)/2^{S-1} \rfloor & \text{za } e(n) \geq 0 \\ \lfloor (-e(n) - 1)/2^{S-1} \rfloor & \text{za } e(n) < 0 \end{cases}$$

$$j(n) = \begin{cases} ((2^{S-1} - 1) \wedge e(n)) + 2^{S-1} & \text{za } e(n) \geq 0 \\ (2^{S-1} - 1) \wedge (-e(n) - 1) & \text{za } e(n) < 0 \end{cases}$$

Kodirana vrijednost će se sastojati od $k(n)$ nula iza kojih dolazi jedinica. Na taj broj nastavlja se broj $j(n)$ širine S bitova.

Krajnji rezultat je datoteka s prosječnom kompresijom od 60 do 40 %. U idućoj tablici su prikazani omjeri originalnih podataka i kompresiranih podataka za FLAC i MPEG-4 ALS algoritme.

Audio format	FLAC	MPEG-4 ALS
48 kHz / 16-bit	2.06	2.24
48 kHz / 24-bit	1.46	1.59
96 kHz / 24-bit	1.76	2.16
192 kHz / 24-bit	-	2.66

Vidimo da MPEG-4 ALS postiže bolje rezultate u odnosu na FLAC, no unatoč tome FLAC je popularniji i rašireniji radi svoje dostupnosti.

Poglavlje 6

Zaključak

Kompresija podataka je proces smanjivanja podataka kako bi se ostvarila ušteda prostora. Manja veličina datoteke omogućuje brže slanje podataka, što je ključno za pristupačnost usluga telekomunikacije i multimedije. Uz uštetu prostora, cilj je i zadržati kvalitetu podataka u mjeri potrebnoj za adekvatnu upotrebu. Ovisno o potrebama, koriste se kodeci s različitim pristupima i rezultatima kompresije. U ovom radu opisani su postupci digitizacije zvuka, odnosno pretvaranje zvuka u njegovu digitalnu reprezentaciju. Glavni postupci digitizacije su uzorkovanje i kvantizacija. Način kvantiziranja može se modificirati kako bi se postigla ušteda podatkovnog prostora. Postoje dvije glavne vrste kompresije, kompresija s gubitcima i kompresija bez gubitaka.

Kompresija s gubitcima u radu prikazana je na primjeru dva MPEG standarda, MPEG-1 i MPEG-2. Oni iskorištavaju psihoakustična svojstva zvuka, poput praga čujnosti ljudskog uha i maskirajućih efekata glasnih tonova, kako bi se postigla visoka kvaliteta zvuka s velikom uštedom prostora. Ovakav tip kompresije najčešće se koristi kod glazbenih datoteka, jer pruža najbolji omjer uštede podataka i zadržane kvalitete zvuka. U telekomunikaciji, kompresija se postiže posebno dizajniranim kodecima koji su specijalizirani za govorne podatke.

Kompresija bez gubitaka uspijeva smanjiti veličinu zvučnih datoteka, a da se pri tom ne gube informacije. Koriste se posebne metode kodiranja, poput Huffmanovog i Riceovog kodiranja. Kompresija bez gubitaka provodi se u tri glavne faze. U prvoj fazi, signal se dijeli na okvire, zatim se uklanja unutarkanalna korelacija i na kraju se izvršava entropijsko kodiranje. U radu su opisana dva primjera kompresije bez gubitaka FLAC i MPEG-4 ALS. Kompresija bez gubitaka koristi se u studijskim okruženjima gdje je prioritet očuvanje svih podataka.

Bibliografija

- [1] M. Bosi, R. E. Goldberg, *Introduction to Digital Audio Coding and Standards*. Springer, 2003.
- [2] K. Pohlmann, *Principles of Digital Audio, Sixth Edition*. McGraw-Hill, 2011.
- [3] Z.N. Li, M.S. Drew, J.Liu, *Fundamentals of Multimedia, Second Edition*. Springer, 2014.
- [4] U. Zölzer, *Digital Audio Signal Processing*. Wiley, 2008.
- [5] V.K. Madiseti, *The Digital Signal Processing Handbook, Second Edition*. CRC Press, 2010.
- [6] N. Chapman, J. Chapman, *Digital Multimedia, Third Edition*. Wiley, 2009.
- [7] G. Loy, *Musimathics: The Mathematical Foundations of Music, Volume 1*. The MIT Press, 2007.
- [8] G. Loy, *Musimathics: The Mathematical Foundations of Music, Volume 2*. The MIT Press, 2007.
- [9] D. Pan, *A Tutorial on MPEG/Audio Compression*. IEEE MultiMedia (Volume: 2, Issue: 2), 1995. 60-74.
- [10] J. Herre, S. Dick, *Psychoacoustic Models for Perceptual Audio Coding - A Tutorial Review*. Applied Sciences 9, MDPI, 2019.
- [11] D. Pan, *Digital Audio Compression*. Digital Technical Journal, Vol.5, No.2, 1993.
- [12] C.E. Shannon, *Communication in the presence of noise*. Proceedings of the Institute of Radio Engineers, 1949; Reprint: Proc. IEEE, Vol. 86, No. 2, 1998.
- [13] Y. A. Reznik, *Coding of Prediction Residual in MPEG-4 Standard for Lossless Audio Coding (MPEG-4 ALS)*. Proc. IEEE International Conference on Acoustics, Speech and Signal Processing vol. III, 2004, pp. III-1024-1027.

- [14] L. D. Fielder, M. Bosi, G. A. Davidson, M. Davis, C. Todd, S. Vernon, *AC-2 and AC-3: Low Complexity Transform-Based Audio Coding*. Collected Papers on Digital Audio Bit-Rate Reduction, pp. 54-72, AES 1996
- [15] K. Sayood, *Lossless Compression Handbook*. Academic press, 2003.
- [16] K. Sayood, *Introduction to Data Compression, Third edition* Elsevier, 2006.
- [17] M.K. Mandal, *Multimedia Signals and Systems*. Kluwer Academic Publishers, 2003.
- [18] W.C. Chu, *Speech Coding Algorithms* Wiley, 2003.
- [19] D. Petrinović, *Uvod u digitalnu obradu govora korištenjem matlaba*. Dostupno na: fer.unizg.hr
- [20] M. Hans, R.W. Schafer, *Lossless Compression of Digital Audio*. IEEE Signal Processing Magazine 18, 21-32, August 2001.
- [21] N. Harada, *Lossless Compression of Speech and Audio Signals and Its Applications*. Dostupno na: <http://www.kecl.ntt.co.jp/people/harada.noboru>
- [22] T. Liebchen, *MPEG-4 ALS - The Standard for Lossless Audio Coding*. The Journal of the Acoustical Society of Korea 28, January 2009.

Pojmovnik

Hrvatski	Engleski	Kratica
adaptivna delta modulacija	adaptive delta modulation	ADM
adaptivna diferencijska pulsno kodna modulacija	adaptive differential pulse code modulation	ADPCM
adaptivna kvantizacija s procjenom unaprijed	adaptive quantization with forward estimation	AQF
adaptivna kvantizacija s unatražnom procjenom	adaptive quantization with backward estimation	AQB
analiza sintezom	analysis-by-synthesis	
delta modulacija	delta modulation	DM
diferencijska pulsno kodna modulacija	differential pulse code modulation	DPCM
digitizacija	digitization	
dubina bitova	bit depth	
dugotrajno predviđanje	long-time prediction	LTP
efekt niskih frekvencija	low frequency effect	LFE
ekspandiranje	expanding	
frekvencija bitova	bit-rate	
frekvencija uzorkovanja	sample rate	
frekvencijsko poklapanje	aliasing	
funkcija širenja	spreading function	
generator impulsnog niza	pulse train generator	
generator pseudoslučajnog šuma	pseudo-noise generator	
interauralna razinska razlika	interaural level difference	ILD
interauralna vremenska razlika	interaural time difference	ITD
izobličenje	distortion	
izobličenje radio preopterećenja	overload distortion	
izobličenje uslijed prevelikog nagiba	slope overload	
izofonska krivulja	equal-loudness contour	
kodek	codec	
kodna knjiga vektora pobuda	codebook of excitation vectors	
kodno pobuđeno linearno predviđanje	code excited linear prediction	CELP
kompaniranje	companding	
kompresija bez gubitaka	lossless compression	
korak kvantizacije	quantization step	
kratki valovi	shortwave	
kratkotrajno predviđanje	short-time prediction	STP
kritični pojasi	critical bands	
kvantizacija	quantization	
linearno prediktivno kodiranje	linear predictive coding	LPC
maskiranje unaprijed	forward masking	
maskiranje unatrag	backward masking	
omjer maskirnog praga i razine šuma	mask-to-noise ratio	MNR

Hrvatski	Engleski	Kratica
omjer signala i maskirnog praga	signal-to-mask ratio	SMR
omjer signala i šuma	signal-to-noise ratio	SNR
podesiva frekvencija uzorkovanja	scalable sampling rate	SSR
pojačanje	gain	
prag čujnosti	hearing threshold	
predikcijska greška	prediction error	
pulsno kodna modulacija	pulse code modulation	PCM
razina intenziteta zvuk	sound intensity level	SIL
razina zvučnog tlaka	sound pressure level	SPL
reprezentacijske razine	representation levels	
rezanje amplituda	clipping	
simulator vokalnog trakta	vocal tract simulator	
sustav otvorene petlje	open-loop system	
sustav zatvorene petlje	closed-loop system	
šum	noise	
uzorkovanje	sampling	
visina zvuka	pitch	
vremensko kašnjenje	time delay	
vremensko oblikovanje šuma	temporal noise shaping	TNS
zajdničko kodiranje kanala	joint channel coding	
zrnati šum	granular noise	
zvučni tlak	sound pressure	

Sažetak

Kompresija podataka je proces smanjivanja podataka kako bi se ostvarila ušteda prostora. U ovom radu opisani su postupci digitizacije zvuka, odnosno pretvaranje zvuka u njegovu digitalnu reprezentaciju. Glavni postupci digitizacije su uzorkovanje i kvantizacija. Način kvantiziranja može se modificirati kako bi se postigla ušteda podatkovnog prostora. Postoje dvije glavne vrste kompresije, kompresija s gubitcima i kompresija bez gubitaka. Kompresija s gubitcima u radu je prikazana na primjeru dva MPEG standarda, MPEG-1 i MPEG-2. Oni iskorištavaju psihoakustična svojstva zvuka, poput praga čujnosti ljudskog uha i maskirajućih efekata glasnih tonova, kako bi se postigla visoka kvaliteta zvuka s velikom uštedom prostora. Kompresija bez gubitaka uspijeva smanjiti veličinu zvučnih datoteka, a da se pri tom ne gube informacije. Koriste se posebne metode kodiranja, poput Huffmanovog i Riceovog kodiranja. Kompresija bez gubitaka provodi se u tri glavne faze. U prvoj fazi, signal se dijeli na okvire, zatim se uklanja unutarkanalna korelacija i na kraju se izvršava entropijsko kodiranje. U radu su opisana dva primjera kompresije bez gubitaka FLAC i MPEG-4 ALS.

Summary

Data compression is the process of encoding information using fewer bits than the original representation. This paper describes the procedures for digitizing sound, ie converting sound into its digital representation. The main digitization procedures are sampling and quantization. The quantization method can be modified to save data space. There are two main types of compression, lossy and lossless compression. Two examples of lossy compression are described in this paper, MPEG-1 and MPEG-2. They use the psychoacoustic properties, such as the hearing threshold and the masking effects of loud tones, to achieve high sound quality with efficient space saving. Lossless compression manages to reduce the size of audio files without losing information. Special coding methods are used, such as Huffman and Rice coding. Lossless compression is performed in three main phases. In the first phase, the signal is divided into frames, then the interchannel correlation is removed and finally the entropy coding is performed. The paper describes two examples of lossless compression - FLAC and MPEG-4 ALS.

Životopis

Rođen sam 4.10.1993. godine u Zagrebu gdje sam pohađao OŠ Frana Krste Frankopana te opći smjer III. gimnazije. 2012. godine upisujem preddiplomski sveučilišni studij Matematike na PMF-u u Zagrebu i završavam preddiplomski nastavnički studij 2016. godine. Tada upisujem diplomski studij Računarstva i Matematike koji završavam ovim radom. Tijekom studiranja više puta sam volontirao na otvorenim danima matematike na PMF-u.