

Tehnici de Compresie a Semnalelor Multimedia
Lucrare de laborator

Compresia audio

I. Obiectivul lucrării

Lucrarea își propune familiarizarea studenților cu principiile compresiei semnalelor audio, punând accentul pe două aplicații de interes ridicat: codarea vorbirii prin metoda codării liniar predictive (LPC, Linear Predictive Coding) și codarea perceptuală în domeniul frecvență, specifică compresiei MPEG (MP3, MPEG-Layer III). Aplicațiile practice folosesc implementări în Matlab ale acestor două tehnici de compresie.

II. Compresia vorbirii prin codarea liniar predictivă (LPC)

Codarea vorbirii reprezintă un aspect important al telecomunicațiilor moderne. Codarea vorbirii înseamnă procesul de reprezentare numerică a semnalului vorbire. Obiectivul de baza al codării vorbirii este de a reprezenta semnalul vorbire cu un număr mic de biți, cu menținerea unui nivel de calitate suficient pentru refacerea vorbirii originale cu un grad de dificultate rezonabil.

Din punct de vedere constructiv, algoritmi de codare/decodare a vorbirii se pot clasifica în felul următor:

1. Tehnici de codare a formei de undă (*waveform coders*): realizează codarea formei de undă a semnalului vocal
 - a. În domeniul timp: PCM, DPCM, ADPCM
 - b. În domeniul frecvență: codare pe sub-benzi de frecvență
2. Tehnici de codare parametrică (*vocoders*): au la bază un model al generării semnalului vocal, problema fiind a se determina, prin analiză, parametrii acestui model pentru un semnal vocal de intrare
 - a. Codarea liniar predictivă (LPC): presupune un model liniar de generare a semnalului vocal

Codarea liniar predictivă (LPC) reprezintă o aplicație clasică a teoriei semnalelor în domeniul compresiei, găsindu-și o aplicativitate deosebită în domeniul compresiei semnalului vocal. Ca atare, LPC stă la baza unor standarde de compresie larg folosite în prezent, cum ar fi ITU G.729 și iLBC, folosite în rețelele de telefonie și aplicații VoIP (de ex. Skype). Codarea LPC oferă rezultate performante în special pentru rate de bit

reduse; când rata de bit nu reprezintă o problemă, codări de calitate mai ridicată se obțin prin tehnicile de codare a formei de undă.

Modelul generării vorbirii

La baza metodei LPC stă modelul de generare a vorbirii prezentat în Fig.1. Un tren de impulsuri este trecut prin filtrul glotei (coardele vocale), eventual amplificat/atenuat, rezultând un semnal cvasi periodic, apoi printr-o serie de cavități rezonante (gură, nas) ce reprezintă traiectul vocal, apoi printr-un al treilea filtru ce reprezintă buzele. Parametrii acestor filtre se modifică rapid în timp, pentru fiecare sunet pronunțat; filtrele sunt deci variabile în timp. Parametrii lor rămân însă relativ constanți pe durata unui singur sunet, ceea ce permite modelarea lor ca filtre invariante în timp pe durate scurte (zeci de ms). Între cuvinte se produc sunete non-vorbire, reprezentate de generatorul de zgomot.

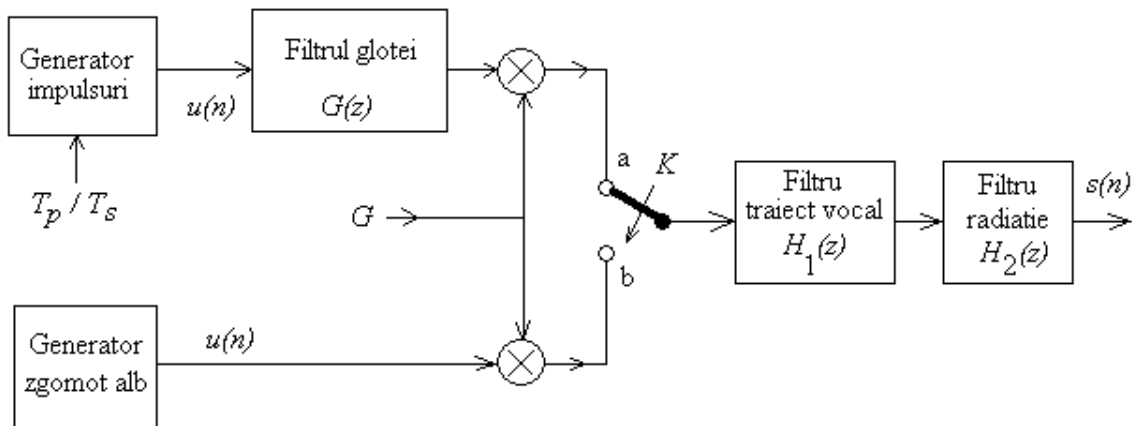


Fig.1. Modelul de generare a vorbirii

În practică se folosește un model simplificat, ca în Fig.2, în care toate filtrele sunt contopite într-unul singur, $H(z)$.

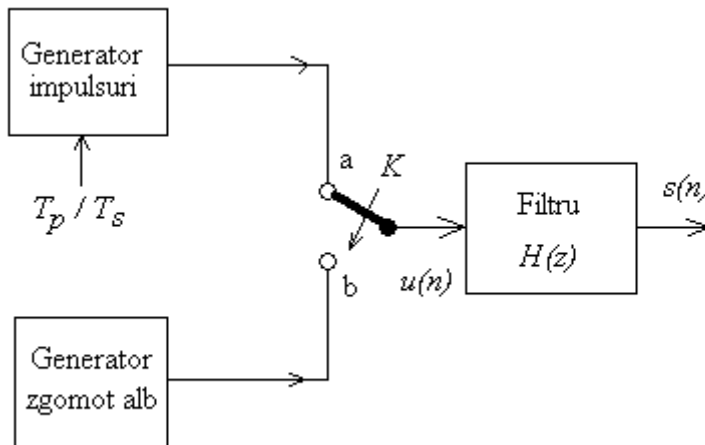


Fig.2. Reprezentarea simplificata a procesului vorbirii

Ideea LPC este de a analiza semnalul vocal, pe segmente scurte de timp, a stabili dacă semnalul este vocal sau zgomot, și în cazul în care este semnal vocal a determina parametrii filtrului $H(z)$ precum și semnalul excitație de la intrarea sa. Astfel, în loc de a se transmite forma de undă a semnalului, se vor transmite acești parametri pentru fiecare segment de timp, urmând ca la recepție semnalul să fie re-sintetizat prin aceeași schemă.

În general, pentru filtru se folosește structura ARMA (*Autoregressive Moving Average*). Eșantionul vorbirii $s(n)$ este modelat ca o combinație liniară a ieșirilor vechi și prezente și a intrărilor vechi din semnalul excitație $u(n)$, după relația:

$$s(n) = \sum_{k=1}^p a_k \cdot s(n-k) + G \sum_{k=0}^q b_k \cdot u(n-k), \quad b_0 = 1 \quad (1)$$

unde G este câștigul filtrului și $\{a_k, b_k\}$ sunt parametrii modelului. Numărul p implică folosirea a p eșantioane trecute, fiind **ordinul** predicției liniare. Funcția de transfer $H(z)$ a modelului este, așadar:

$$H(z) = \frac{S(z)}{U(z)} = G \cdot \frac{1 + \sum_{k=1}^q b_k z^{-k}}{1 - \sum_{k=1}^p a_k z^{-k}} \quad (2)$$

cea ce arată un model de tip poli-zero. În spectrul vorbirii, sunetele nazale sunt determinate de zero-uri și cele vocale de poli.

Există două cazuri speciale ale acestui model:

1. Cazul modelului de tip auto-regresiv, când $H(z)$ are numai poli, deci coeficienții b_k sunt nuli. Acest model se folosește din motive de simplitate și eficiență a implementării., motivul fiind că în rezolvarea unui model poli-zero este nevoie de rezolvarea unui set de ecuații neliniare în timp ce modelul numai cu poli necesită numai rezolvarea unui set liniar de ecuații.
2. Cazul modelului de tip medie alunecătoare, când $H(z)$ are numai zero-uri, deci coeficienții a_k sunt zero.

În cazul modelului AR (pe care îl studiem în această lucrare), coeficienții $\{a_k\}$ sunt numiți **coeficienții LP (liniari predictivi)** ai filtrului liniar.

Analiza și sinteza semnalului vocal

Cele trei etape ale codării LPC a unui segment de date vocal sunt următoarele:

1. Estimarea parametrilor filtrului $H(z)$, pentru un ordin dat, prin metoda autocorelației și algoritmul Levinson-Durbin
2. Analiza semnalului prin filtrarea cu inversul lui $H(z)$ (pentru un model AR, filtrul de analiză $A(z) = 1 / H(z)$ este un filtru MA, numai cu zerouri). Se obține un **semnal rezidual**, care nu este altceva decât semnalul excitație $u(n)$ de la intrarea filtrului.
3. La recepție, cunoscând semnalul rezidual și parametrii filtrului, se obține semnalul inițial prin filtrarea semnalului rezidual cu filtrul de sinteză, identic cu $H(z)$.

Ca urmare, modelul global poate fi descompus în două părți, partea de analiză și partea de sinteză, așa cum se prezintă în figura 3.

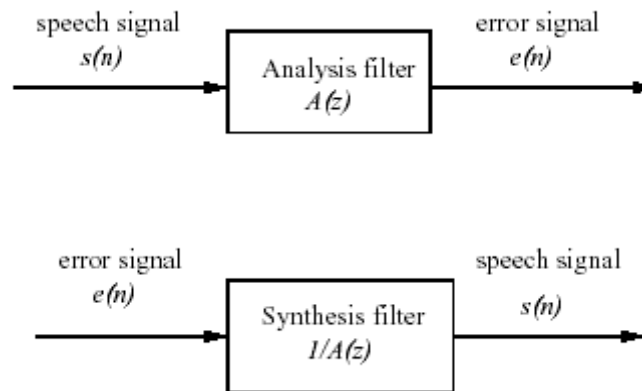


Fig.3. Modelele de analiză și de sinteză

Partea de *analiză* analizează semnalul vorbire și produce semnalul rezidual. Partea de *sinteză* preia semnalul rezidual ca semnal de intrare, îl filtrează cu $1/A(z) = H(z)$ și reconstruiește semnalul vorbire.

Există două tipuri de codare bazate pe predicție liniară:

- Codor adaptiv înainte sau direct (*Forward adaptive coder*). Predicția liniară se bazează pe eșantioanele de intrare vechi. Analiza LP este efectuată în codor și se transmit coeficienții LP.
- Codor adaptiv înapoi sau invers (*Backward adaptive coder*). Analiza LP este făcută din nou la recepție, deci în cadrul decodurului, coeficienții LP sunt calculați din eșantioanele vechi reconstruite. În acest fel, nu este necesar să se transmită coeficienții LP de către codor.

Estimarea coeficienților predicției liniare

Există două metode larg folosite în estimarea coeficienților predicției liniare: *autocorelația* și *covarianța*. Ambele metode folosesc coeficienții filtrului LP $\{a_k\}$ astfel încât energia reziduală (energia semnalului rezidual) să fie minimizată.

Ponderarea pe ferestre (*windowing*)

Semnalul vorbire este un semnal variabil în timp și unele variații sunt aleatoare. Uzual, în timpul vorbirii cu intensitate mică, forma tractului vocal și a excitației nu se modifică în 200 ms. Dar *fonemele* au o durată medie de 80 ms. Cele mai multe schimbări apar mai frecvent decât intervalul de 200 ms.

Analiza semnalului presupune ca proprietățile semnalului se schimbă lent în timp, ceea ce permite o analiză în timp scurt a semnalului. Semnalul este divizat în segmente succesive, analiza se face pe aceste segmente și se extrag anumite proprietăți dinamice.

Semnalul $s(n)$ este înmulțit cu o fereastră de analiză $w(n)$ pentru a extrage un anumit segment de analiză. Tehnica se numește *ponderare (windowing)*.

Alegerea formei ferestrei este importantă întrucât eșantioanele vor fi ponderate în mod diferit. Folosirea unei ferestre dreptunghiulare determină apariția unor lobi secundari mari și anumite efecte nedorite în domeniul frecvență. Pentru a înlătura aceste oscilații mari, se utilizează ferestre de filtrare fără schimbări abrupte în domeniul timp, așa cum este fereastra Hamming, Blackman, Kaiser sau Bartlett. Se pot folosi și combinații de ferestre, așa cum se folosește în GSM, unde fereastra de analiză are două jumătăți de ferestre Hamming de dimensiuni diferite.

Metoda autocorelației

După segmentare și ponderare, al doilea pas este minimizarea energiei semnalului rezidual. Energia reziduală se obține cu expresia

$$E = \sum_{n=-\infty}^{\infty} e^2(n) = \sum_{n=-\infty}^{\infty} \left[s_w(n) - \sum_{k=1}^p a_k s_w(n-k) \right]^2 \quad (2)$$

Valorile parametrilor $\{a_k\}$ care minimizează energia E se găsesc prin calcularea derivatelor parțiale ale lui E în raport cu $\{a_k\}$ și egalarea lor cu zero, $\frac{\partial E}{\partial a_k} = 0$, pentru $k = 1, \dots, p$, de unde rezultă un set de p ecuații cu p necunoscute:

$$\sum_{k=1}^p a_k \sum_{n=-\infty}^{\infty} s_w(n-i) s_w(n-k) = \sum_{n=-\infty}^{\infty} s_w(n-i) s_w(n), \quad 1 \leq i \leq p \quad (3)$$

În ultima ecuație semnalul ponderat $s_w(n)=0$ în afara ferestrei de filtrare.

Ecuațiile liniare pot fi exprimate în funcție de funcția de autocorelație. Funcția de autocorelație a unui segment ponderat de vorbire este definită de relația

$$R(i) = \sum_{n=-i}^{N_w-1} s_w(n)s_w(n-i), \quad 1 \leq i \leq p \quad (4)$$

unde N_w este lungimea ferestrei. Prin înlocuirea valorilor autocorelației în penultima ecuație se obține

$$\sum_{k=1}^p a_k R(|i-k|) = R(i), \quad 1 \leq i \leq p \quad (5)$$

Setul de ecuații liniare poate fi scris matricial sub forma

$$\begin{bmatrix} R(0) & R(1) & \cdots & R(p-1) \\ R(1) & R(0) & \cdots & R(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ R(p-1) & R(p-2) & \cdots & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ \vdots \\ R(p) \end{bmatrix} \quad (6)$$

adică

$$\mathbf{R} \cdot \mathbf{a} = \mathbf{r} \quad (7)$$

Matricea \mathbf{R} este o matrice Toeplitz (fiecare diagonală conține același element). Acest lucru permite rezolvarea ecuației matriciale prin algoritmul Levinson-Durbin sau prin algoritmul lui Schur.

Datorită structurii matricii \mathbf{R} , $A(z)$ este de tip fază minimă. Filtrul folosit la sinteză are funcția de transfer $H(z) = 1 / A(z)$, deci zerourile lui $A(z)$ devin poliile lui $H(z)$. Astfel, faza minimă a lui $A(z)$ garantează stabilitatea filtrului $H(z)$.

Algoritmul Levinson-Durbin

Algoritmul rezolvă ecuația matricială $\mathbf{Ax}=\mathbf{b}$, în care \mathbf{A} are o structură Toeplitz, matrice simetrică și pozitiv definită, iar \mathbf{b} este un vector arbitrar. Aceste cerințe sunt satisfăcute de matricea de autocorelație \mathbf{R} , definită anterior.

Algoritmul LD necesită o formă specială a lui \mathbf{b} , unde \mathbf{b} constă din anumite elemente ale lui \mathbf{A} . Si aceasta condiție este satisfăcută de ecuațiile autocorelației.

Fie $a_k(m)$ coeficientul k pentru un cadru oarecare m al iterației. Algoritmul LD rezolvă iterativ setul de ecuații pentru $m=1,2,\dots,p$ după relațiile:

$$k(m) = R(m) - \sum_{k=1}^{m-1} a_k(m-1)R(m-k) \quad (8)$$

$$a_m(m) = k(m) \quad (9)$$

$$a_k(m) = a_k(m-1) - k(m)a_{m-k}(m-1), \quad 1 \leq k \leq m \quad (10)$$

$$E(m) = (1 - k(m))^2 \cdot E(m-1) \quad (11)$$

Se consideră inițial $E(0) = R(0)$ și $a(0) = 0$. La fiecare iterație, coeficientul $a_k(m)$ pentru $k=1,2,\dots,m$ descrie predictorul liniar de ordin m ; eroarea de energie $E(m)$ este redusă cu un factor de $(1-k(m)^2)$.

Întrucât $E(m)$ nu este negativă, rezultă că $|k(m)| \leq 1$. Această condiție pentru coeficientul de reflexie $k(m)$ garantează că rădăcinile lui $A(z)$ vor fi în interiorul cercului unitate. Rezultă că filtrului $H(z)$ va fi stabil.

Transmisia semnalului rezidual

În vederea reducerii ratei de bit totale, codoarele vorbirii de tipul CELP (*code excited linear prediction*), cum este de ex. G.729, nu transmit întreg semnalul rezidual, întrucât se folosește un tabel de codare vectorial pentru a coda semnalul de excitație. Tehnica este numită cuantizare vectorială (VQ): codorul selectează unul dintre semnalele de excitație dintr-un tabel predeterminat, și va transmite indexul semnalului de excitație care se află cel mai aproape de cel care trebuie transmis. Tabelul cu semnalele de excitație este cunoscut atât de codor cât și de decodor. Semnalul excitație este selectat astfel încât distorsiunea dintre cadrul original și cel reconstruit să fie minimă.

III. Codarea perceptuală a semnalelor audio

Codarea pe sub-benzi

În codarea pe sub-benzi, semnalul audio este supus unei analize cu un banc de filtre constând din M filtre trece bandă care sunt continue în frecvențe, astfel încât semnalele rezultate pot fi recombinate aditiv pentru a produce semnalul original. Fiecare ieșire a filtrului este decimată la limită (se mai spune *critic decimată*) (adică eșantionată la dublul benzii nominale) cu un factor egal cu M , numărul de filtre trece bandă. Această decimare rezultă într-un număr de eșantioane subbandă ce egalează acelea din semnalul original. În receptor, rata de eșantionare a fiecărei subbenzi este crescută la cea a sursei de semnal prin completarea cu eșantioane zero. Procesul de eșantionare poate introduce distorsiuni de alias din cauza suprapunerii subbenzilor. Există bancuri de filtre ce permit o reconstrucție perfectă, termenii de aliasing fiind eliminați și suma ieșirilor filtrelor trece bandă fiind egală cu semnalul de intrare, în absența cuantizării.

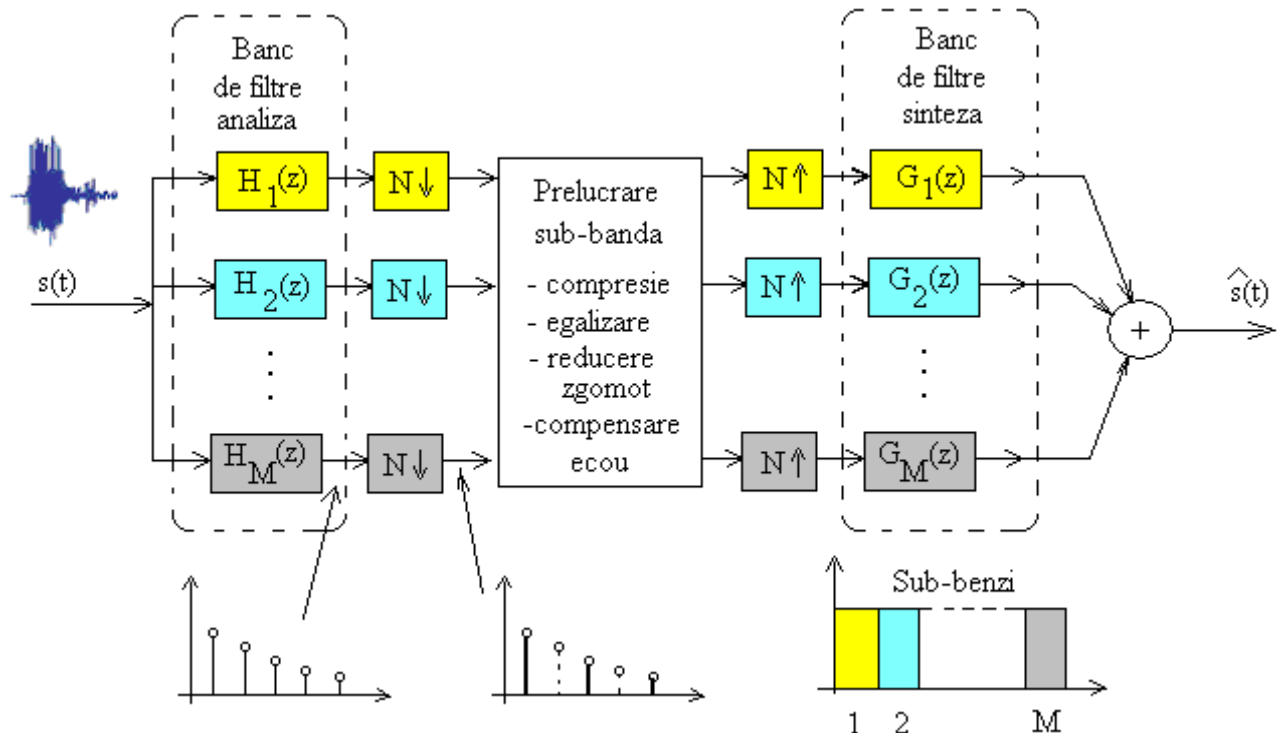


Fig.4. Descompunerea semnalului în sub-benzi

Ideea codării pe sub-benzi este de a cuantiza fiecare sub-bandă în mod diferit, în funcție de sensibilitatea urechii umane, frecvențele mai puțin perceptibile fiind cuantizate mai puternic. Vorbim așadar de un model *psihacoustic* al urechii umane care estimează ce benzi de frecvență sunt mai puțin perceptibile la un moment dat, alocând mai puțini biți (cuantizare mai puternică) celor mai puțin audibile (*codare perceptuală*), fig.5.

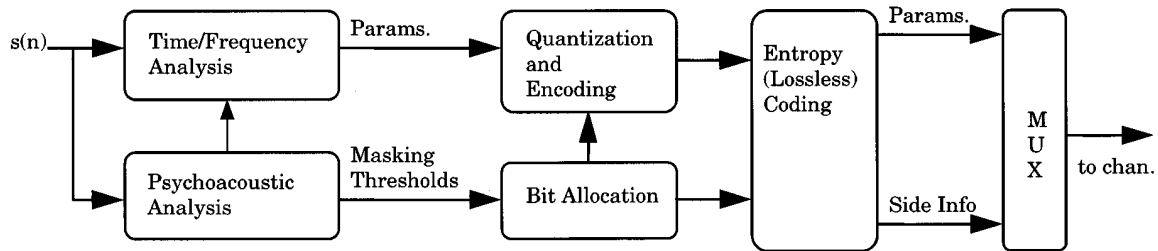


Fig.5. Schema de principiu a codării pe sub-benzi

Modele psihoacustice

Cel mai simplu model psihoacustic folosește graficul general al sensibilității urechii la diverse frecvențe, precum cel din fig.6. Astfel, pasul de cuantizare pentru o sub-bandă rezultă din sensibilitatea medie a urechii în acea sub-bandă, astfel încât zgomotul de cuantizare să fie invers proporțional cu sensibilitatea (la frecvențele mai puțin sensibile se poate tolera un zgomot de cuantizare mai puternic, la frecvențele mai perceptibile este necesar un zgomot de cuantizare mai redus). În plus, urechea are o capacitate redusă de a discrimina frecvențele, în special înspre capetele benzii – o abatere a frecvenței de la valoarea corectă poate trece neobservată. Ca atare, se pot defini *benzile critice* ca fiind benzile de frecvență în interiorul cărora urechea nu distinge diferența între două tonuri; acestea sunt mai înguste în jurul valorii de 4kHz. Dintr-o bandă critică este suficient a se transmite un ton dominant. Ca atare, în sistemele ce folosesc codarea perceptuală, semnalul este descompus în sub-benzi care corespund exact benzilor critice ale sistemului auditiv uman.

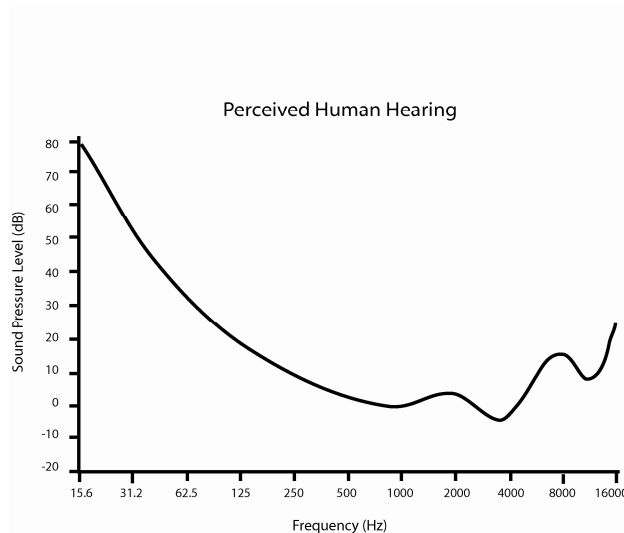


Fig.6. Pragul de sensibilitate în frecvență, în condiții de liniște

Efecte de mascare auditivă

Un fenomen interesant, exploatat în standardele de compresie MP3 și AAC (și nu numai), este cel de mascare auditivă, prezentat în fig.7

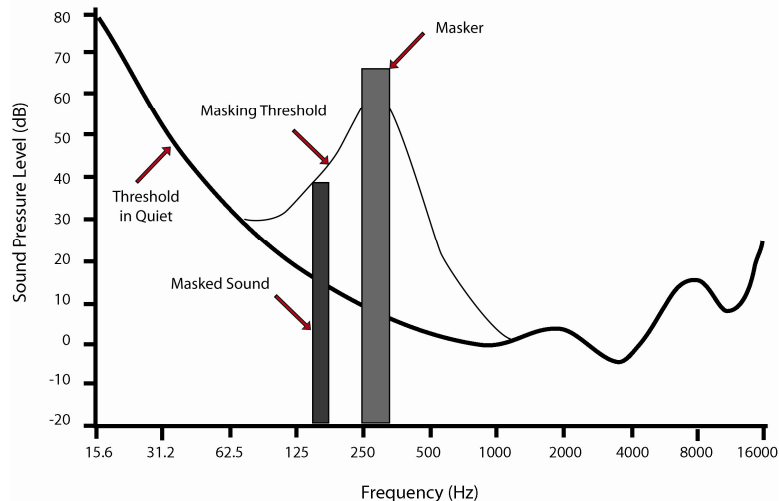


Fig.7. Efectul de mascare auditivă

Fenomenul este următorul: prezența unui ton de intensitate ridicată, pe o anumită frecvență, maschează prezența simultană a unui ton pe o frecvență apropiată, de intensitate mai redusă, ton care însă ar fi fost audibil în absența primului. Cu alte cuvinte, în prezența unui ton dominant, tonurile de intensitate scăzută de pe frecvențele apropiate sunt inaudibile. Acest efect se micșorează pe măsură de diferența de frecvență este mai mare. Ca atare, tonurile ce intră sub această *mască auditivă* pot fi cuantizate foarte puternic sau chiar eliminate din semnal, fără ca urechea umană să sesizeze o diferență. Această analiză trebuie repetată însă pe segmente scurte de timp, întrucât compoziția spectrală a semnalului variază în timp.

Ca atare, etapele de funcționare ale unui codor perceptual (cum sunt MP3 / AAC) sunt următoarele:

1. Semnalul este segmentat în segmente de durată redusă, care sunt apoi descompuse în frecvență pe sub-benzi
2. Se determină frecvențele tonurilor dominante și se construiește masca auditivă pentru acest segment
3. Sub-benzile se cuantizează în acord cu masca auditivă calculată, cuantizând mai puternic benzile de frecvență mascate, astfel încât zgomoul de cuantizare să rămână ușor sub pragul audibil.

Standardul MPEG – Layer III (MP3)

Structura de bază este prezentată în figura 8, și are următoarele caracteristici:

Mod de operare: MPEG-1 lucrează pentru semnale mono dar și stereo. O tehnică numită codare stereo compusă (joint stereo coding) poate fi utilizată pentru a coda mai

eficient canalele stânga și dreapta ale semnalului audio stereofonic. Nivelul 3 permite codarea duala mijloc/stereo și a intensității stereo. Modurile de operare sunt:

- un canal;
- canal dublu (doua canale independente, de exemplu conținând informație audio în 2 limbi diferite);
- stereo;
- stereo compus;

Debitul informației (Bit-rate): Alegerea ratei de bit este lăsată, între anumite limite, la opțiunea celui care implementează codorul MPEG audio. Pentru nivelul 3, standardul definește un domeniu al debitelor de la 8Kbit/s la 320 Kbit/s. Mai mult, decodorul pentru nivelul 3 trebuie să suporte schimbarea ratei de bit de la un cadru la altul.

Codorul: Codarea MPEG audio este lăsată complet la latitudinea celui care implementează standardul.

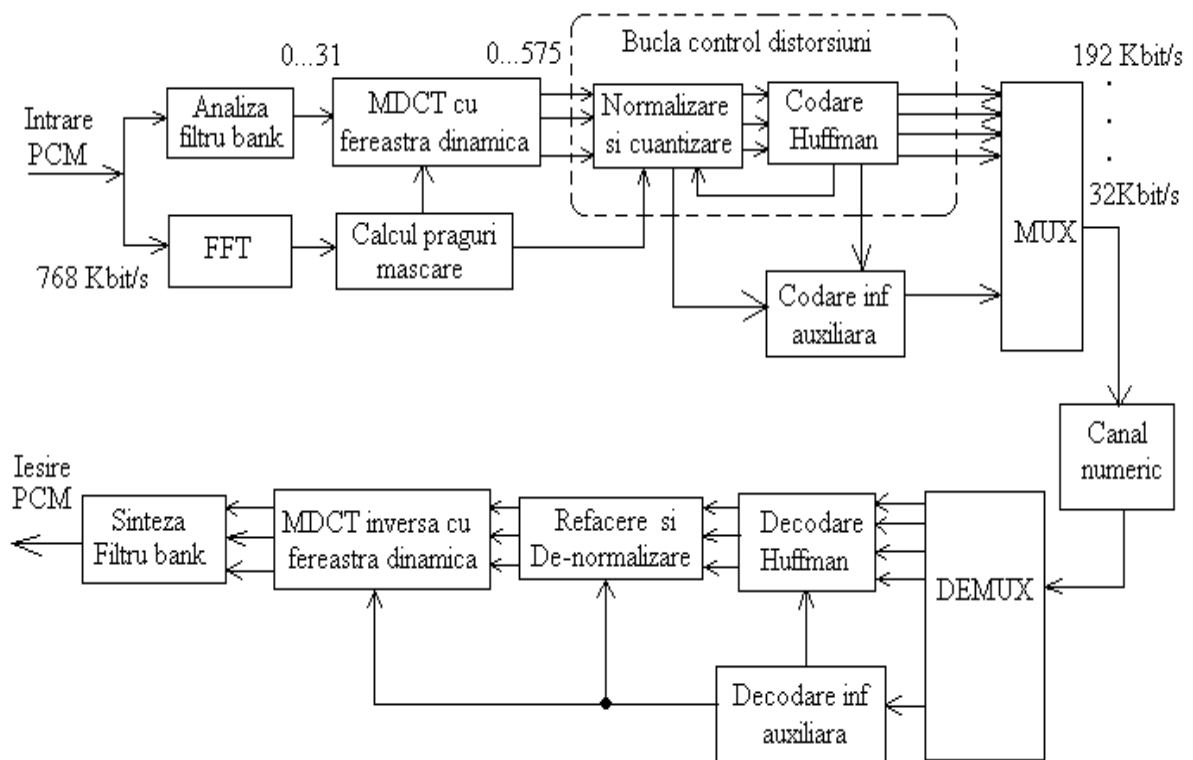


Fig.8. Structura standardului audio MPEG-1, stratul III

Banc de filtre hibride comutat: Pentru a obține o rezoluție în frecvență îmbunătățită, aproape de partiția în benzi critice, cele 32 de semnale sub-bandă sunt divizate în continuare prin aplicarea pe fiecare sub-bandă a unei transformate DCT modificate (MDCT) în 6 sau 18 puncte, cu 50% suprapunere. Numărul maxim de componente în frecvență este $32 \cdot 18 = 576$, fiecare reprezentând o bandă de $24000/576 =$

41.67 Hz. Transformarea în 18 puncte se aplică pentru a furniza o rezoluție îmbunătățită în frecvență, în timp ce transformarea în 6 puncte oferă o rezoluție mai bună în domeniul timp.

Cuantizare și codare: Componentele spectrale sunt cuantizate și codate cu scopul menținerii zgomotului de cuantizare sub pragul de mascare.

Ieșirea modelului perceptual constă în valorile pragurilor de mascare pentru fiecare bandă de frecvență a codorului. Benzile de frecvență ale codorului sunt echivalente – în mare – cu benzile critice ale auzului uman. Dacă zgomotul de cuantizare poate fi păstrat sub pragul de mascare pentru fiecare bandă, atunci rezultatul compresiei nu diferă de semnalul original din punct de vedere al audiției.

Cuantizarea este neliniară, astfel încât valorile mari sunt codate automat mai puțin precis. Valorile cuantizate sunt codate Huffman. Pentru adaptarea procesului de codare la diferite statistici ale semnalului audio, tabelul Huffman este selectat dintr-un număr de opțiuni posibile, astfel încât se pot folosi diferite tabele Huffman pentru părți diferite ale spectrului. Procesul de căutare al câștigului optim și a factorilor de scalare pentru un bloc dat se face uzual prin două bucle iterative, ce lucrează în modul analiză-sinteză:

- o buclă *internă* (bucla pentru rata de bit): Tabela Huffman alocă cuvinte de cod scurte valorilor cuantizate ce apar cel mai des. Dacă numărul de biți rezultat din operația de codare depășește numărul de biți disponibili pentru a coda un bloc de date, se poate corecta prin ajustarea câștigului global pentru a determina un pas de cuantizare mai mare, obținându-se un număr mai mic de valori cuantizate. Bucla se numește *rate loop*, pentru că modifică rata globală a codorului până când este suficient de mică;
- o buclă *externă* (bucla de control a zgomotului): pentru obținerea unui zgomot de cuantizare la un prag de mascare, factorii de scală sunt aplicați la fiecare factor de scală al benzii. Sistemul pleacă cu un factor de scală inițial de 1.0 pentru fiecare bandă. Dacă zgomotul de cuantizare într-o bandă dată depășește pragul audibil furnizat de modelul perceptual, factorul de scală pentru această bandă este ajustat pentru reducerea zgomotului de cuantizare. Întrucât obținerea unui zgomot de cuantizare mic necesită un număr mare de pași de cuantizare și deci o rată de bit mare, rata de ajustare a buclei trebuie să se repete de fiecare dată când se aplică noi factori de scală. Bucla externă este executată până când zgomotul actual (calculat ca diferența între valorile spectrului original și valorile spectrului cuantizat) este mai mic decât pragul de mascare pentru fiecare factor de scală al unei benzi.

IV. Desfășurarea lucrării

Codarea LPC

1. În Matlab, rulați aplicația LPC Vocoder. Ascultați semnalul original și cel resintetizat. Identificați semnalele vizualizate pe schema bloc din fig.3. Vizualizați mai multe segmente din semnal prin butoanele < și >.
 - a. Modificați ordinul filtrului LPC la 50 (și rulați din nou codarea cu „Run Program”), apoi la 200, și apoi la 4. Ce se întâmplă cu forma de undă a filtrului $H(z)$ și a semnalului de la ieșirea sa ? Dar cu forma de undă a semnalului excitație? Cum variază calitatea sunetului? Cum credeți că variază rata de compresie?
 - b. Pentru ordinul filtrului egal cu 12, setați lungimea segmentelor la 5ms cu 0% suprapunere, și alegeți fereastra rectangulară / Hamming / triangulară. În care caz calitatea este cea mai slabă? De ce ?
 - c. Setați fereastra Hamming, 0% suprapunere, și variați lungimea segmentelor la 5, 20, 50, 200 ms. Ce se întâmplă la creșterea segmentelor?
 - d. La lungimea segmentelor de 200ms, alegeți suprapunerea segmentelor la 0% / 50% / 75%. Cum se modifică calitatea semnalului? Care este rolul suprapunerii segmentelor?
 - e. Pentru un ordin de predicție egal cu 10, cu lungimea segmentului de 40ms, suprapunere de 50% și coeficienți cuantizați pe 4 biți, calculați rata de compresie față de un semnal PCM cu eșantionare de 8kHz cu 8biți/eșantion (pentru LPC luați în considerare doar coeficienții de predicție, nu și semnalul rezidual).

Codarea perceptuală

2. Rulați programul Matlab-MPEG. Cu Enter, derulați figurile pentru câteva segmente succesive. Identificați zonele cu sensibilitate ridicată și cele în care efectul de mascare permite un zgomot de cuantizare mare.

Întrebări:

1. Care este specificul codării parametrice a vorbirii?
2. Pentru ce rate de bit este preferată codarea LPC?
3. Ce este fenomenul de mascare auditivă?
4. În figurile de la punctul 2, măștile auditive de la frecvențele înalte sunt mai „largi” decât cele de la frecvențe joase. De ce?