COMPRESIA AUDIO.

- **Semnalul necomprimat:**
  - Frecvenţa maximă percepută de urechea umană este de aprox. 20kHz;
  - frecvenţa de eşantionare este de minim 40 kHz;
  - cuantizare cu 16 biti/eşantion;
  - pentru un semnal stereo calitate CD (eşantionat cu 44,1 kHz) rezultă o rată de transmisiune pentru semnalul necomprimat de 1.4 Mbps.
- **Metodele de compresie fără pierderi** (Huffman, LZW, etc.) în general nu funcţionează bine pentru compresia audio.
- **Metode de compresie cu pierderi**:
  - **Silence Compression**
    - detectează zonele de "linişte", asemănătoare cu codarea run-length;
  - **Adaptive Differential Pulse Code Modulation** (ADPCM)
    - în CCITT G.721 -- 16 sau 32 kbiţi/sec.
    - codează diferenţa între două eşantioane consecutive;
    - adaptează pasul de cuantizare aşa încât să se micşoreze varianţa (puterea) zgomotului de cuantizare.
    - se obţine o compresie de aproximativ 4:1.

- **Linear Predictive Coding (LPC)**
  - se transmit, conform modelului vorbirii, parametrii de model ai corzilor vocale, laringelui, cavităţii bucale.
  - sună ca şi vorbirea sintetizată pe calculator.
  - rată de 2.4kbiti/sec.
- **Code Excited Linear Predictor (CELP)**
  - efectuează LPC, dar transmite şi termenul de eroare
  - calitate de audio-conferinţă la o rată de 4,8 kbiţi/sec.

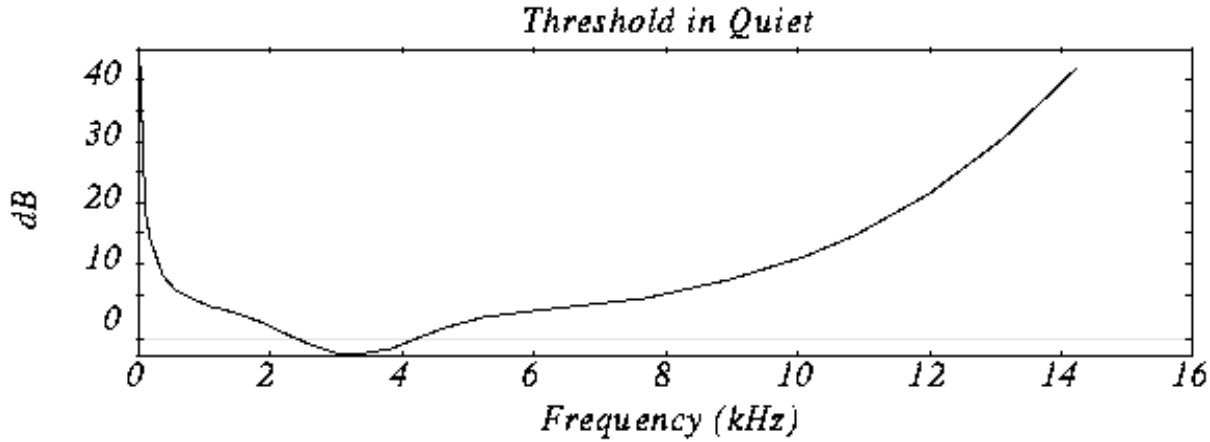
- Codarea audio poate fi făcută în:
  - TIMP
    - complexitate redusă;
    - necesită mai mult de 10 biţi/eşantion pentru păstrarea calităţii;
  - FRECVENŢĂ
    - se poate obţine o calitate înaltă cu numai 3 biţi/eşantion;
    - se utilizează codarea în subbenzi şi prin transformări;
- Pentru obţinerea unor rate de compresie mari toate metodele de codare se bazează pe percepţia audio umană (**PSIHOACUSTICĂ**).

Auzul si vocea umană

- Domeniul audibil este între 20 Hz şi 20 kHz, cel mai sensibil la frecvenţe de la 2 la 4 kHz.
- Dinamica auzului (încet la tare) e de aproximativ 96 dB.
- Vocea are domeniul normal de frecvenţă între 500 Hz şi 2 kHz
- Fonemele sonore (m, v, l) au frecvenţe joase.
- Fonemele insonore (f, s) au frecvenţe înalte.

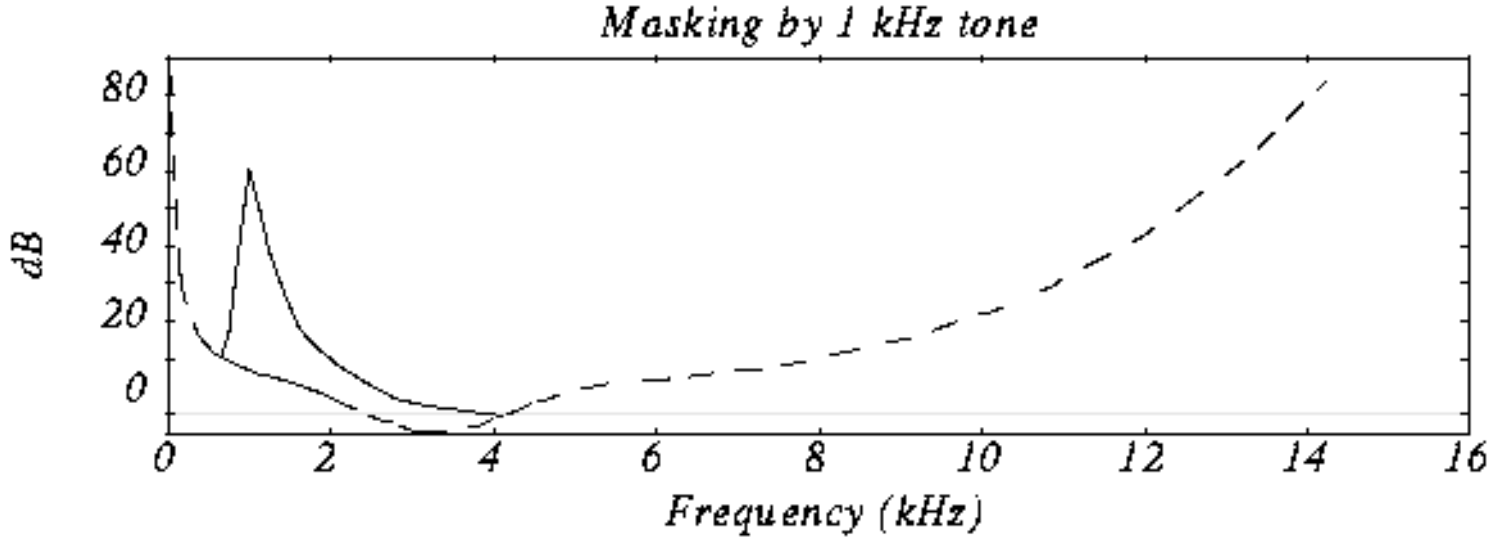
Cât de sensibila este urechea umana?

- Experiment: O persoană ascultă un semnal de 1 kHz într-o cameră liniştită. Se reduce nivelul semnalului până când acesta nu se mai aude. La fel se reprezintă pentru toată gama de frecvenţe audio şi rezultă **curba de mascare în linişte**:

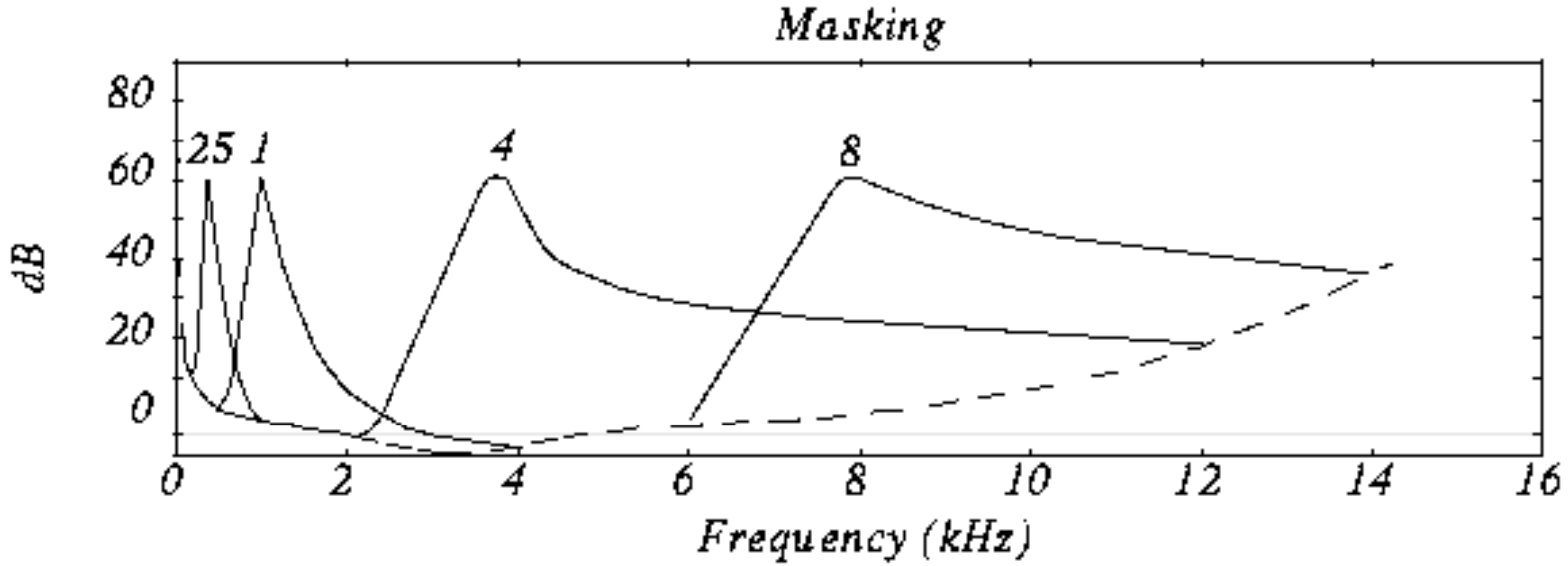


Mascarea în frecvenţă

- Experiment: Se asculta un ton de 1 kHz (**ton de mascare**) la un nivel fixat (60 dB).
- Se asculta un **ton de test** cu nivel variabil până când acesta începe să se audă.
- Se variază frecvenţa semnalului de test în jurul lui 1 kHz.



- Se repetă experimentul pentru mai multe frecvenţe ale **tonului de mascare** obţinându-se **curbele pragului de mascare în frecvenţă**.

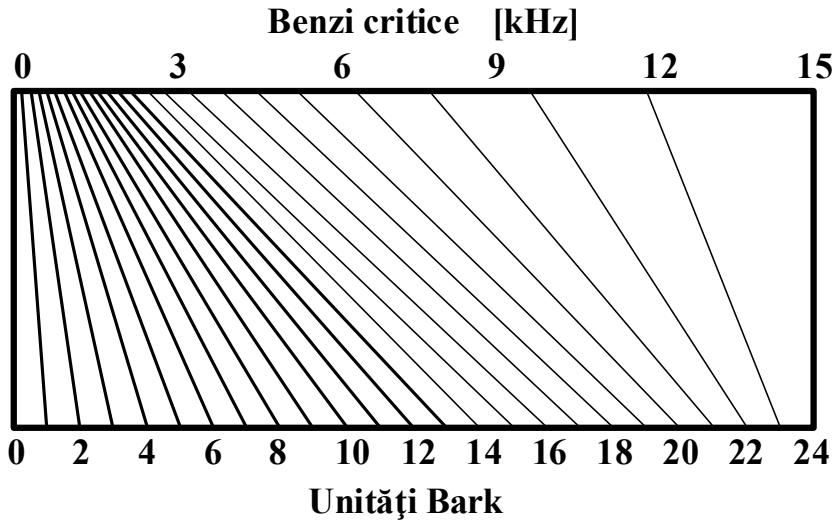


Benzi critice

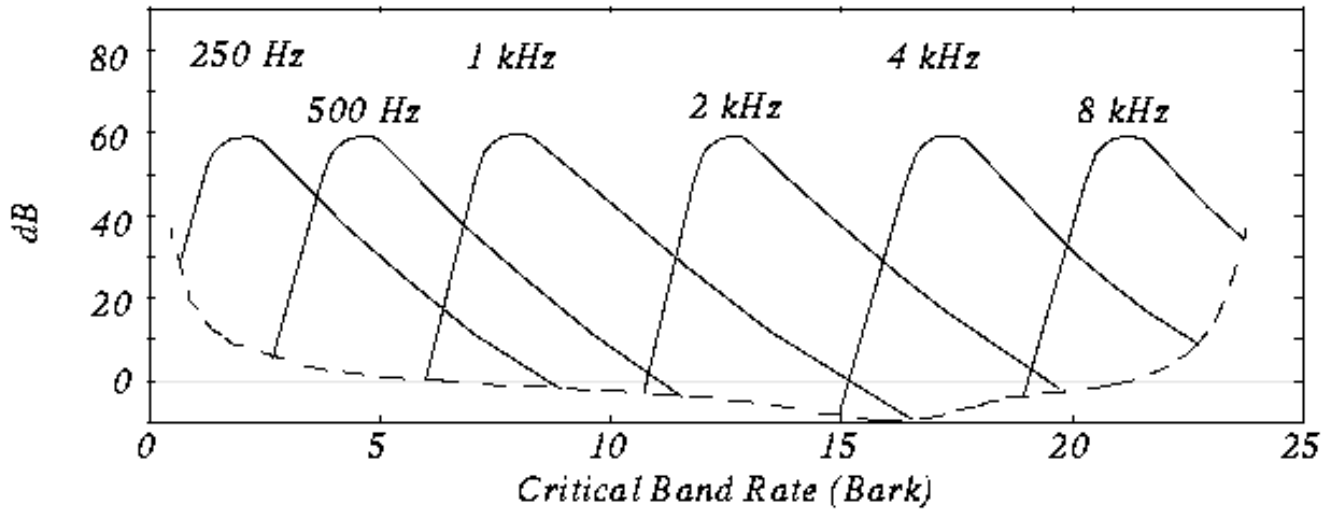
- Măsură uniformă de percepţie a frecvenţei neproporţională cu lăţimea curbei de mascare.
- Aproximativ 100 Hz pentru frecvenţe de mascare <500 Hz, creşte din ce în ce mai mult peste 500 Hz.
- Lăţimea benzii se numeşte mărimea benzii critice.

Bark

- O altă unitate de măsură pentru frecvenţă (după Barkhausen).
- 1 Bark = lăţimea unei benzi critice.
- Pentru frecvenţe < 500 Hz, f/100
- Pentru frecvenţe > 500 Hz, $9+4\cdot\log_2(f/1000)$

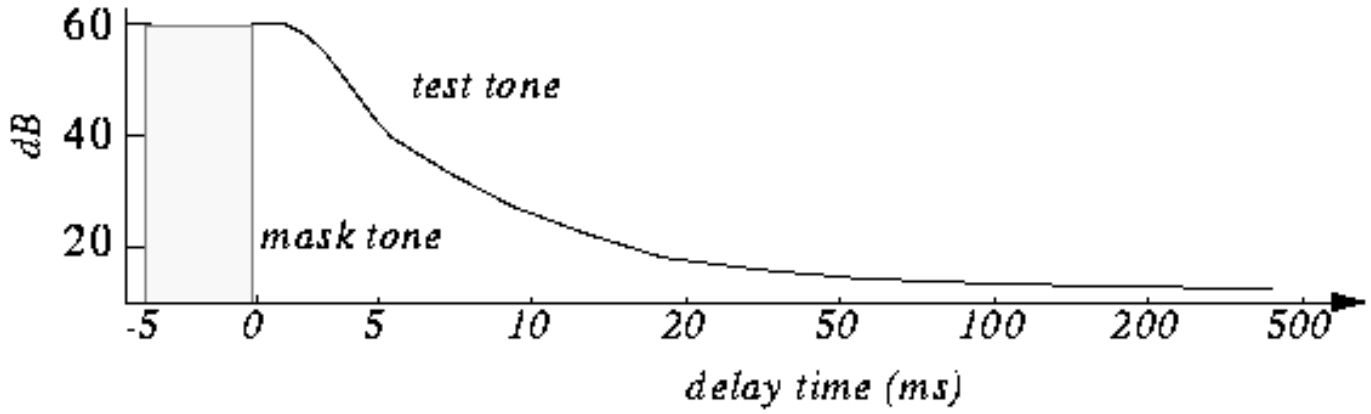


- Pragurile de mascare reprezentate în funcţie de banda critică:

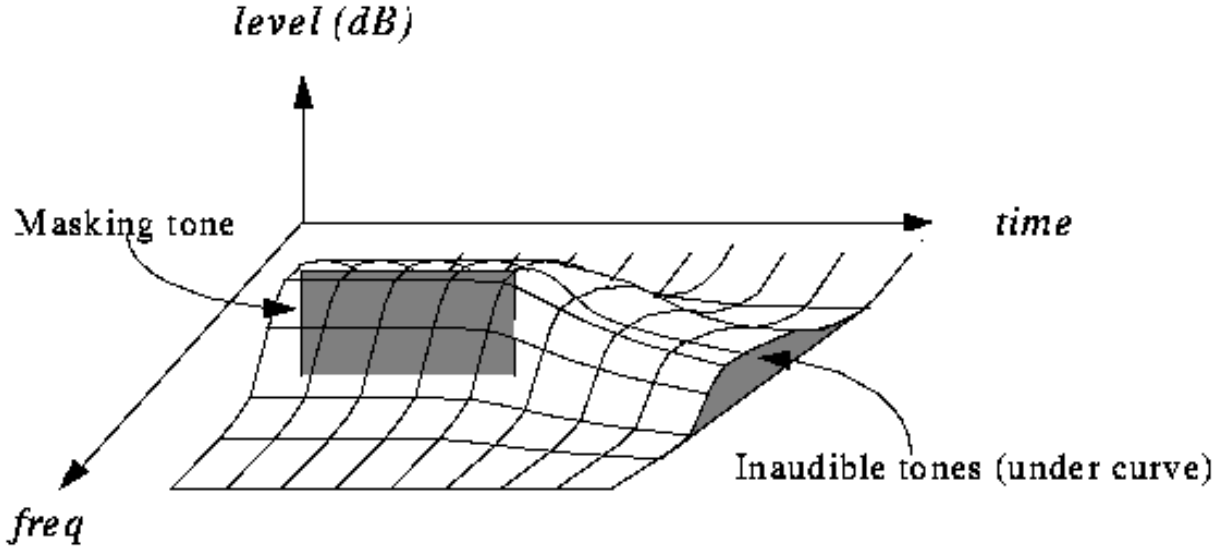


Mascare temporală

- Dacă ascultăm un sunet puternic, apoi acesta se opreşte, trebuie să treacă un timp scurt pentru ca să putem auzi un sunet slab în apropiere.
  - Experiment: Se ascultă un **ton de mascare** de 1 kHz, 60 dB şi un **ton de test** de 1.1kHz, 40 dB. Tonul de test nu se poate auzi (e mascat).
  - Se opreşte **tonul de mascare**, apoi, după o scurtă întârziere, se opreşte tonul de test.
  - Se ajustează întârzierea la durata minimă la care tonul de test mai poate fi auzit (aprox. 5 ms).
  - Se repetă cu niveluri diferite ale tonului de test.



- Se încearcă alte frecvenţe pentru tonul de test (durata tonului de mascare rămâne constantă).
- Efectul total al mascării:

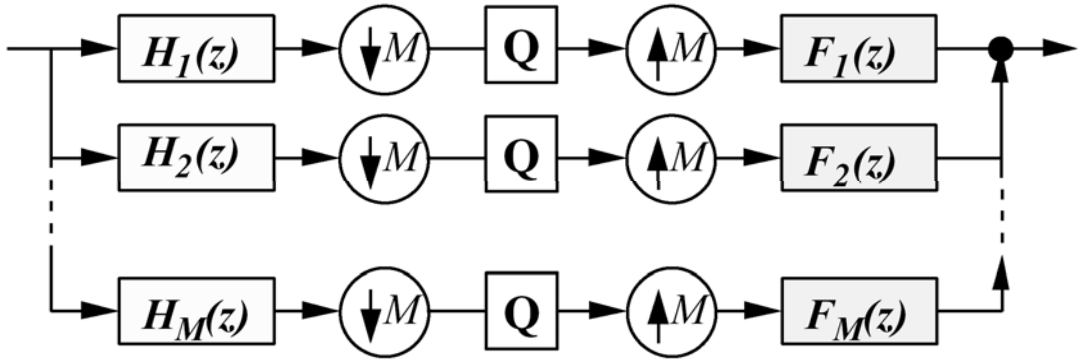


Concluzii:

- Dacă avem un ton puternic (de exemplu 1 kHz), atunci tonurile mai slabe, apropiate în frecvenţă şi timp, sunt mascate.
- Comparaţiile se fac pe scala benzilor critice (dimensiunea mascării e aproximativ o bandă critică).
- Există doi factori de mascare: mascare în frecvenţă şi mascare temporală.

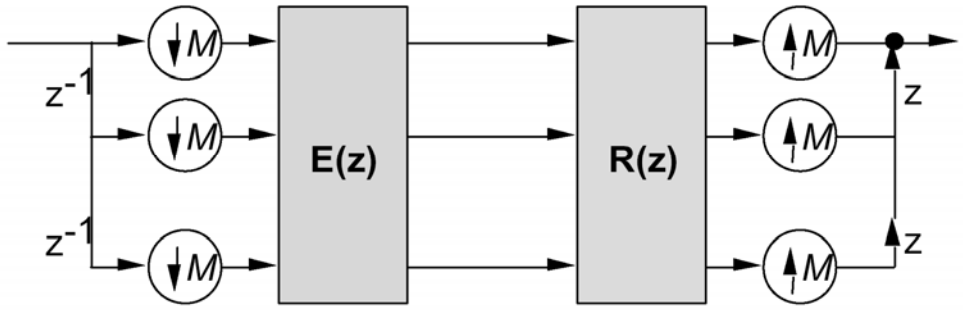
Cum se poate folosi mascarea în compresia audio?

- Un semnal mascat de altul mai puternic este comparabil cu zgomotul de cuantizare.
- Funcţia de mascare oferă distorsiunea maximă acceptabilă pentru fiecare bandă critică.
- Codorul determină mascarea din fiecare banda cauzată de semnalele din benzile apropiate.
- Dacă puterea în bandă este sub pragul de mascare aceasta nu se codează.
- Altfel, se determină numărul de biţi necesari pentru cuantizarea fiecărui coeficient astfel încât zgomotul introdus de cuantizare este sub pragul de mascare. (1 bit de cuantizare introduce 6 dB zgomot).



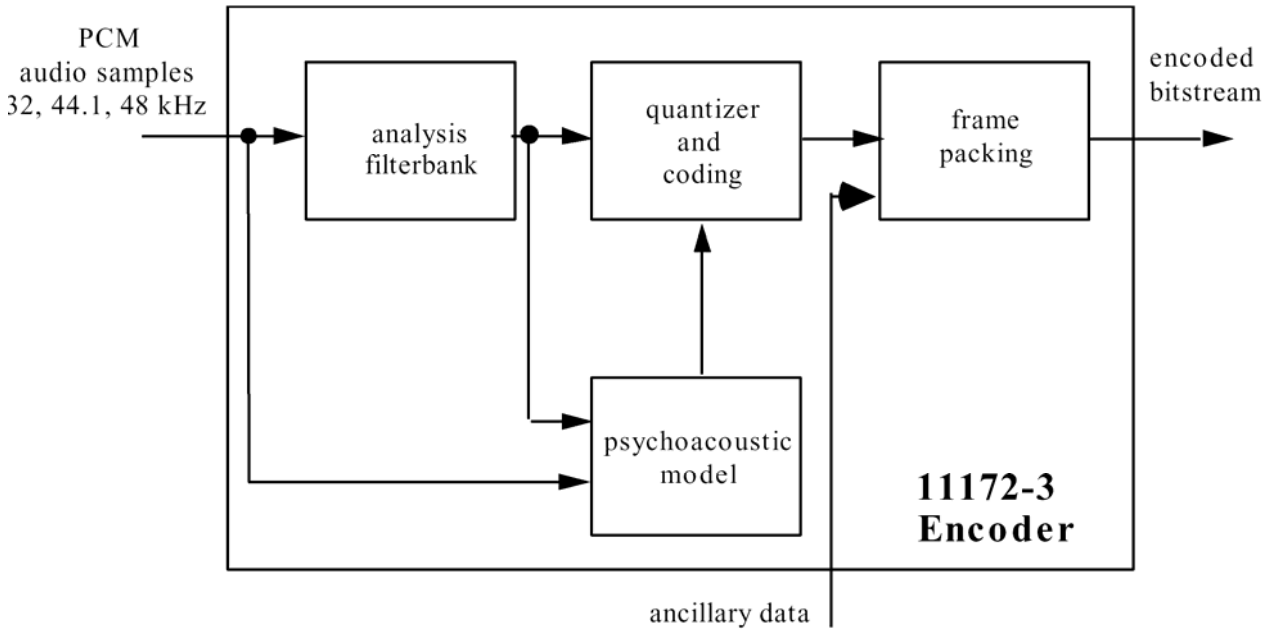
Filtre de analiză Filtre de sinteză

- Benzile de frecvenţă percepute de ureche nu sunt uniforme ci logaritmice.
- Bancul de filtre de analiză ar trebui să aproximeze benzile critice.
- Minimizarea ratei de biţi în limitele date de mascare conduce la o compresie audio optimă.
- Se poate folosi pentru analiza în subbenzi transformata cosinus dacă E(z) este matricea DCT şi R(z) este matricea IDCT.

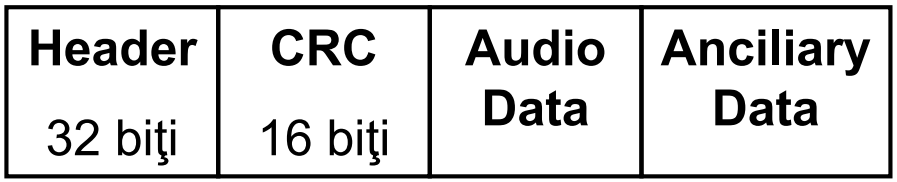


Codarea MPEG-1 audio

- Standardul ISO/IEC 11172-3 elaborat între 1988 şi 1991.
- Este primul standard de compresie audio de înaltă calitate.
- Codează semnale audio cu frecvenţele de eşantionare de 32, 44.1 şi 48kHz.
- Rata de bit comprimată pentru un semnal de calitate CD-audio stereo este între 64kbiţi/s şi 256kbiţi/s faţă de 1.4Mbiţi/s necomprimat.
- Schema bloc a codorului:

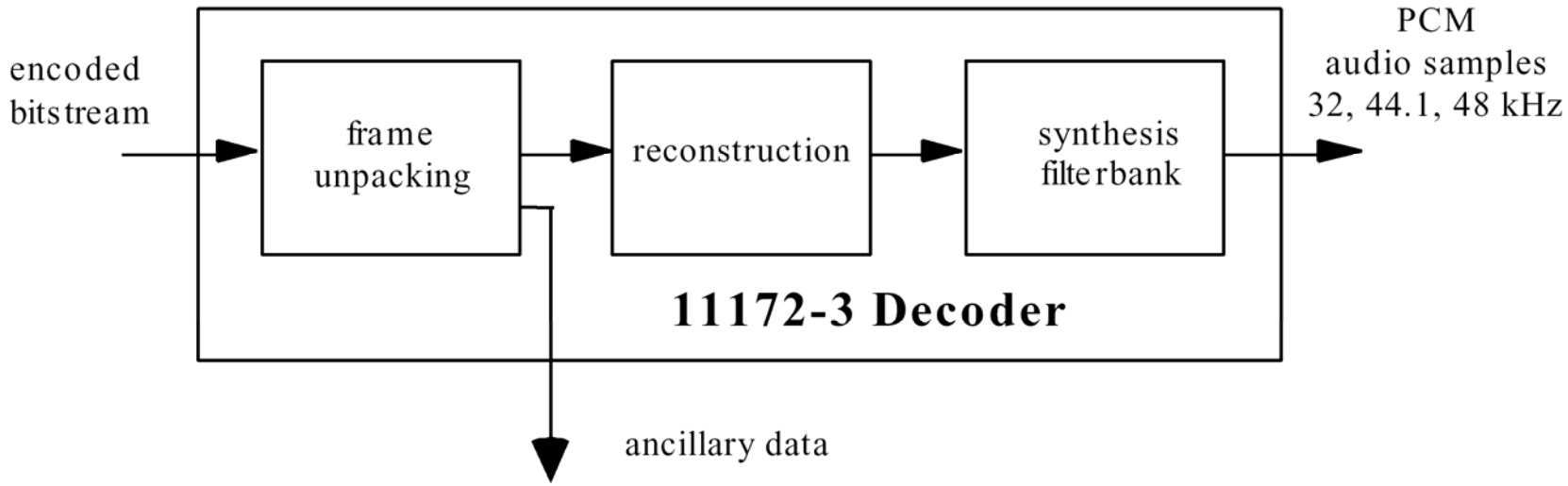


- Codorul analizează componentele spectrale ale semnalului audio cu ajutorul bancului de filtre sau a transformării MDCT (layer 3).
- Aplica un model psihoacustic pentru a estima nivelul minim de zgomot.
- Se furnizează SMR (Signal-to-Mask Ratio) pentru alocarea biţilor sau a zgomotului.
- Se formează fluxul de biţi după cum urmează:



- Header-ul
  - Syncword (12 biţi)
  - Layer code (2 biţi) reprezentând layerele I, II si III
  - Bit-rate index (4 biţi) indexul debitului utilizat (diferă pentru fiecare layer în parte)
  - Frecvenţa de eşantionare (2 biţi) poate fi 48, 44.1 şi 32kHz
  - Padding bit
  - Mod (2 biţi) stereo, joint stereo, unu sau două canale

- Schema bloc a decodorului



- Standardul MPEG audio include 3 layere diferite corespunzător diverselor aplicaţii, cu creşterea complexităţii codorului dar şi a performanţelor (calitatea sunetului raportată la rata de bit).
- Layerele sunt compatibile în sensul ierarhic (layerul N poate decoda fluxul de date codate în layerul N şi în toate layerele inferioare).
- Toate layerele au aceeaşi structură de bază.

- Layer 1
  - de la 32 kbps pâna la 448 kbps
  - rata de compresie 1:4
- Layer 2
  - de la 32 kbps pâna la 384 kbps
  - rata de compresie 1:6..8
- Layer 3
  - de la 32 kbps pâna la 320 kbps
  - rata de compresie 1:10..12

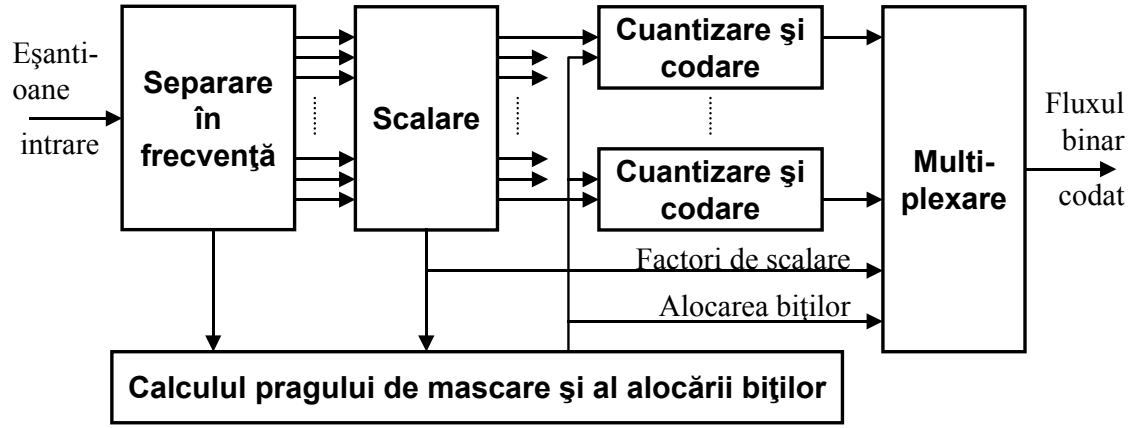
- Layer I - algoritmul de bază pentru codarea audio
  - Bancul de filtre împarte semnalul audio în 32 de subbenzi de frecvenţă cu lăţime egală. Acestea nu corespund cu benzile critice.
  - Codorul calculează pragul de mascare folosind benzile critice.
  - Eroarea care apare la reconstrucţie nu este mare.
  - Cadrul este format din 12 eşantioane/subbandă.
  - Conţine un model psihoacustic pentru determinarea adaptivă a alocării biţilor şi pentru cuantizare.
  - Domeniile de aplicaţie includ înregistrarea digitală pe bandă sau disc.

- Layer II - algoritm îmbunătăţit faţă de layer I
  - Îmbunătăţirea constă într-o codare suplimentară a alocării biţilor, a factorilor de scalare şi o structură diferită a cadrului.
  - Codorul formează 3 blocuri cu 12 eşantioane/bloc şi 32 de subbenzi (1152 eşantioane).
  - Se transmite un tip de alocare a biţilor şi maxim 3 factori de scalare pentru 3 blocuri (câte un factor de scalare pentru fiecare bloc).
  - Aplicaţii în studiourile profesionale (radiodifuziune, înregistrări), telecomunicaţii, multimedia etc.
- Layer III - cea mai bună compresie
  - Creşte complexitatea codorului/decodorului.
  - Conţine un banc de filtre hibrid (filtre plus MDCT- modified discrete cosine transform).
  - Se obţine o rezoluţie mai bună în frecvenţă prin utilizarea MDCT.
  - Două lungimi ale blocului MDCT: 36 eşantioane şi 12 eşantioane.
  - Aplicaţii în telecomunicaţii pe canale de banda îngustă ISDN, mp3 şi alte aplicaţii cu debit foarte redus.

Algoritmi propuşi:

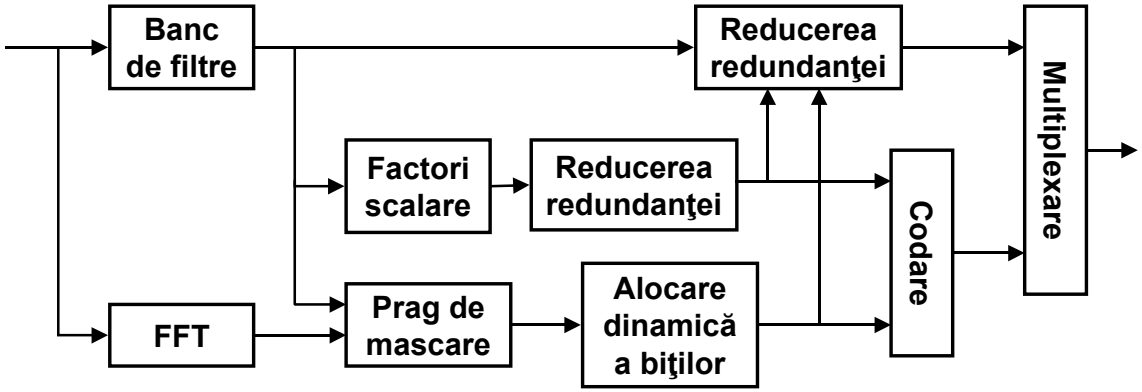
- ASPEC (Audio Spectral Perceptual Entropy Coding): codare cu transformate cu suprapunerea blocurilor;
- ATAC (Adaptive Transform Aliasing Cancellation): codare cu transformate fara suprapunerea blocurilor;

- SB/DPCM (Subband Coding and DPCM): codare pe subbenzi cu mai putin de 8 subbenzi;
- MUSICAM (Masking-pattern Universal Subband Integrated Coding and Multiplexing): codare pe subbenzi cu mai mult de 8 subbenzi;

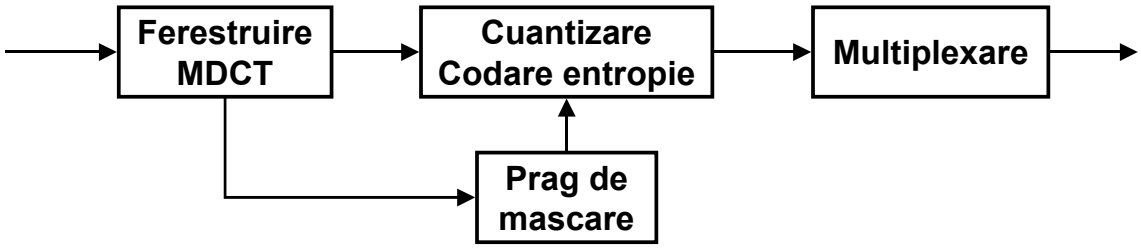


- Eşantioanele audio sunt mapate în frecvenţă printr-o transformare sau cu un banc de filtre.
- Coeficienţii audio din domeniul de frecvenţă sunt normaţi cu un factor de scalare detreminat din pragul de mascare al răspunsului psihoacustic.

- Codorul MUSICAM



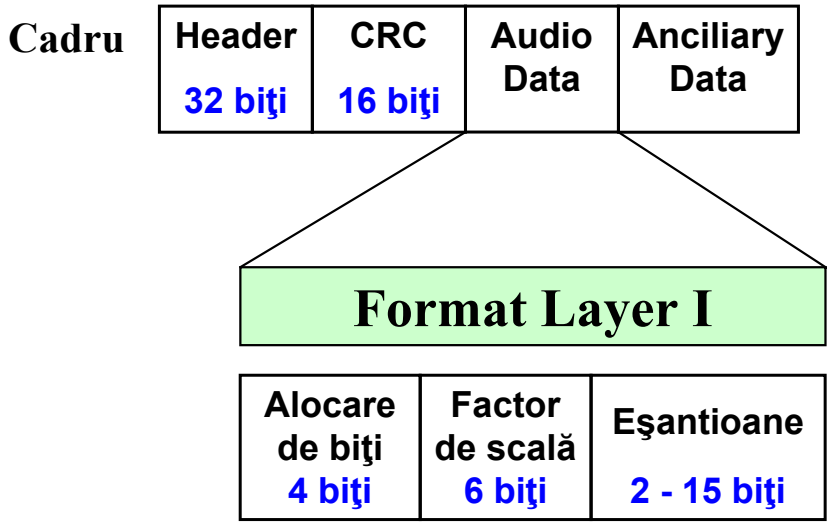
- Filtrele polifazice au complexitate de calcul redusă iar faza liniară permite reconstrucţia perfectă.
- Un semnal eşantionat cu 48kHz este împărţit în 32 de subbenzi, fiecare bandă având lăţimea de 0.75kHz.
- Semnalele de subbandă sunt împărţite în cadre digitale de 12 eşantioane succesive (8 ms).
- Intervalul de eşantionare în fiecare subbandă este de 2/3 ms.
- Pragul de mascare se calculează dintr-un estimat pe termen scurt al densităţii spectrale de putere prin medierea transformatei FFT.
- Calculul se repetă la fiecare 24 ms.
- Lăţimea constantă a subbenzilor nu coincide cu benzile critice.
- După calculul puterii zgomotului de mascare, biţii se aloca cuantizoarelor minimizându-se NMR.
- Factorii de scalare pot fi calculaţi folosind cuantizarea adaptivă aşa încât eşantioanele să fie între [-1,1].
- Factorii de scalare au redundanţă mare şi pot fi codaţi, urmând a fi transmişi împreună cu informaţia de alocare a biţilor în fluxul de date.
- Codorul ASPEC



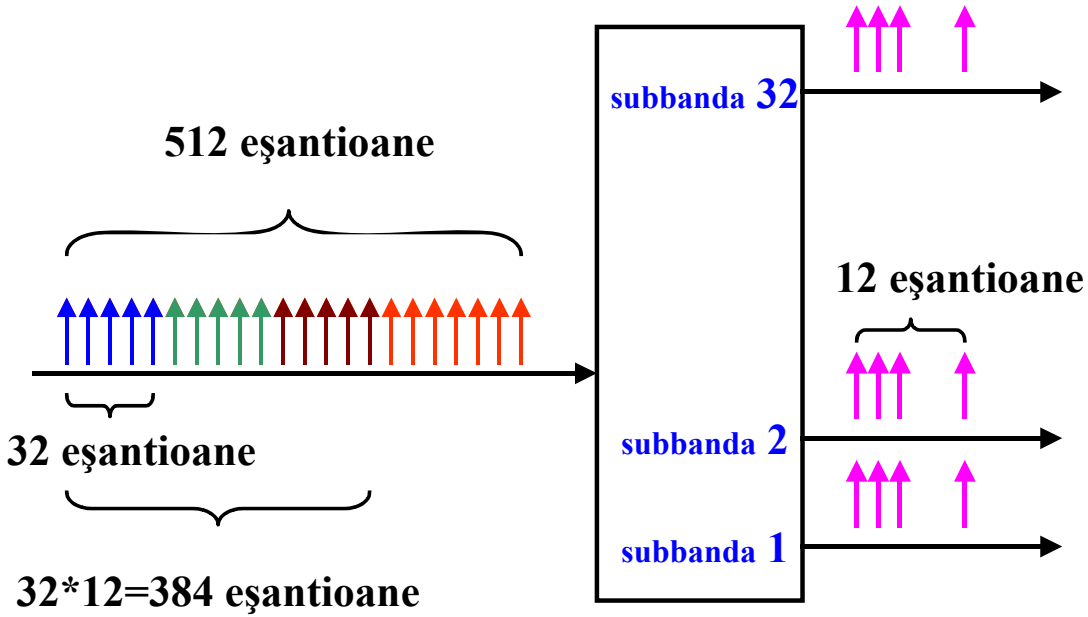
- Pentru separarea în frecvenţă se utilizează MDCT.
- Eşantioanele sunt ferestruite pentru limitarea alierii în domeniul timp.
- MDCT împreună cu subeşantionarea creează dintr-un bloc de 2N eşantioane, N coeficienţi în frecvenţă.
- Calculul pragului de cuantizare:
  - Este calculată energia semnalului în domeniul frecvenţă (amplitudine şi fază);
  - Se calculează energia în fiecare bandă critică. Această energie dă pragul neîmprăştiat.
  - Împraştierea e calculată cu o funcţie de împrăştiere.
  - În final se calculează entropia perceptuală utilizată pentru estimarea numărului de biţi necesari pentru blocul curent.
- Datele cuantizate sunt codate cu cod Huffman.
- Factorii de scalare şi alocarea biţilor sunt multiplexati în fluxul de date.

- **MPEG Layer I**

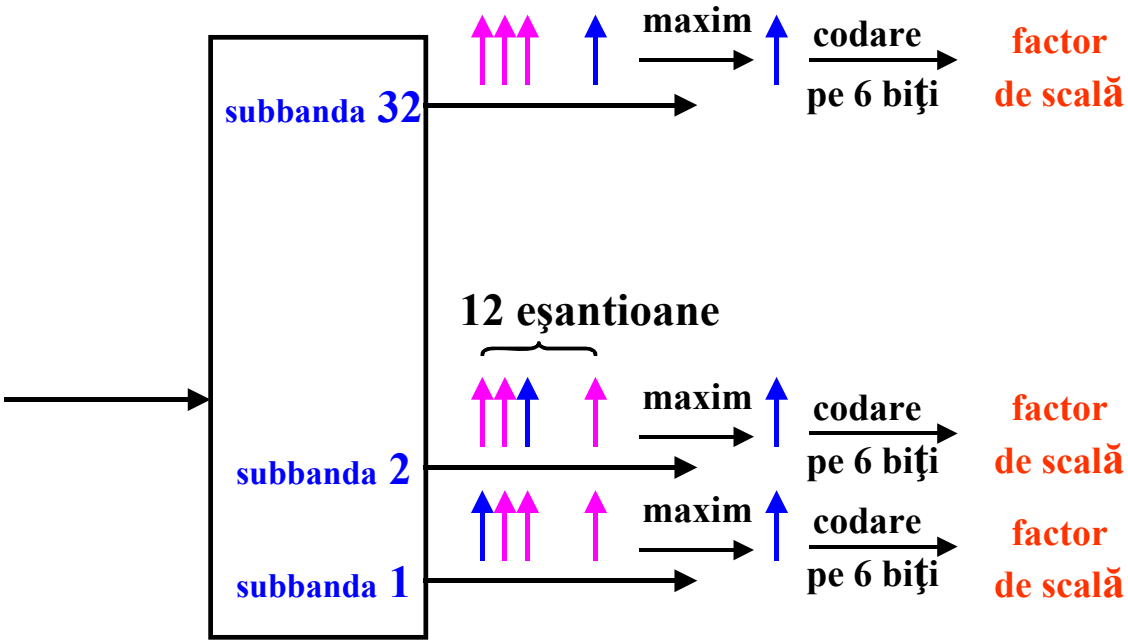
- Filtrarea în subbenzi;
- Modelare psihoacustică;
- Scalare şi alocarea biţilor;
- Cuantizare şi codare
- Formarea fluxului de date



- **Filtrarea în subbenzi**
- Se foloseşte un banc de filtre pentru a transforma semnalul audio din domeniul timp în frecvenţă.
- Filtrele împart semnalul iniţial în 32 de benzi de frecvenţă echidistante cu frecvenţă de eşantionare $F_S/32$.



- Pentru fiecare subbandă se calculează maximul (în modul) pentru fiecare set de 12 eşantioane.
- Factorul de scalare se alege dintr-un tabel şi este valoarea imediat superioară maximului găsit.
- Se codează indexul factorului de scalare din tabel, pe 6 biţi pentru fiecare subbandă.
- Acesta se transmite doar dacă a fost alocat benzii un număr nenul de biţi.



- **Modelare psihoacustică**
- Layer I suporta atât modelul psihoacustic I cât şi modelul psihoacustic II.
- Totusi, modelul psihoacustic I este suficient pentru Layer I, care implică un FFT de 512 elemente.
- SMR (signal-to-mask ratio) se determina din modelul psihoacustic folosit.

- **Modelul psihoacustic I**

- Calculul FFT în paralel cu filtrarea în subbenzi compensează lipsa de selectivitate a filtrelor în zona de joasă frecvenţă. FFT este de 512 eşantioane pentru layer I şi de 1024 eşantioane pentru layer II.
- Se cunoaşte pragul de mascare în linişte.
- Se extrag din spectrul de putere FFT componentele tonale şi netonale deoarece ele influenţează pragul de mascare în benzile critice.
- Componentele tonale sunt cele care respecta relaţiile:

$$power_x(i-j) < power_x(i) - 7 \le power_x(i+j) \qquad j \in \{2,3,6\}$$

$$power_x(i-1) < power_x(i) \le power_x(i+1)$$

- Se elimina componentele vecine componentelor tonale.
- Se elimina componentele tonale şi netonale care sunt sub pragul de mascare în linişte.
- Daca mai multe componente tonale sunt la distanţă mai mică de 0.5 Bark se păstrează maximul lor.
- Calculul pragului global de mascare (în dB):

$$LT_G(i) = 10 \ \log_{10}\left[10^{LT_q(i)/10} + \sum_{j=1}^{m} 10^{LT_{tm}(j,i)/10} + \sum_{j=1}^{n} 10^{LT_{nm}(j,i)/10}\right]$$

  unde LT_q este pragul în linişte, iar LT_{tm} şi LT_{nm} sunt pragurile de mascare datorate componentelor tonale şi netonale.
- Pragul global de mascare minim din subbanda *n* se utilizează pentru determinarea raportului semnal-mascare (SMR):

$$SMR_{sb}(n) = L_{sb}(n) - LT_{\min}(n) \ \text{dB}$$

  unde Lsb(n) este nivelul semnalului în subbanda *n*.

- Se calculează SMR pentru fiecare subbandă.

- **Modelul psihoacustic II**
- Dimensiunea FFT şi a ferestrei Hann poate fi variată. Layer III calculează modelul de două ori în paralel cu FFT de 192 şi de 576 esantioane (bloc scurt / lung).
- Se consideră o funcţie de împrăştiere între benzile critice vecine bazată pe mascarea temporală (sunetele se "sting" în timp iar curba de mascare este influenţată de sunetele precedente).
- Pragul audibil final se calculează prin convoluţia energiei împrăştiate şi a energiei parţiale iniţiale.

- SMR e calculat ca raport între energia parţială e_{part} şi nivelul zgomotului n_{part}:

$$SMR_n = 10\log_{10}\left(epart_n / npart_n\right)$$

- **Alocarea biţilor**
- Conceptul de bază în alocarea biţilor este minimizarea **MNR** din cadru cu constrângerea ca numărul total de biţi utilizaţi să nu depăşească numărul de biţi disponibili în cadru B_f. B_f se calculează cu formula:

$$B_f = \frac{Bit\ \ rate}{f_s} \cdot 384\ \ \ biti / cadru$$

- Procedura de alocare de biţi e iterativă şi porneşte din starea "zero bit allocation".
- Intâi se calculează "mask-to-noise ratio" **MNR** care se obţine cu formula:

MNR = SNR - SMR (dB)

unde :

SNR se găseşte în tabelul următor
SMR este furnizat de modelul psihoacustic.

Biţi	Codul	Număr de niveluri	SNR (dB)
0	0000	0	0.00
2	0001	3	7.00
3	0010	7	16.00
4	0011	15	25.28
5	0100	31	31.59
6	0101	63	37.75
7	0110	127	43.84
8	0111	255	49.89
9	1000	511	55.93
10	1001	1023	61.96
11	1010	2047	67.98

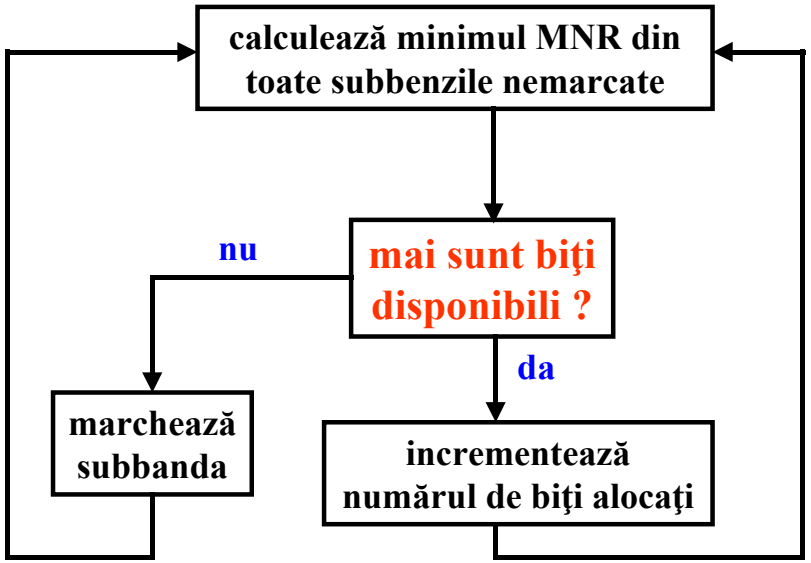
12	1011	4095	74.01
13	1100	8191	80.03
14	1101	16383	86.05
15	1110	32767	92.01
invalid	1111	-	-

- **MNR** arată diferenţa dintre eroarea de cuantizare şi măsurarea perceptuală.
- Eşantioanele audio pot fi comprimate de MNR ori.
- De aceea minimul MNR din fiecare subbandă e determinat la fiecare iteraţie.

- Procedura iterativă se repetă până când MNR e minimizat şi numărul de biţi folosiţi pentru cele 4 componente se apropie de numărul de biţi disponibili.
- Biţii marginali calculaţi la fiecare iteraţie B_{mg}, pot fi calculaţi ca:

$$B_{mg}=B_{tav}-(b_{bal}+b_{scf}+b_{spl}+b_{anc})$$

unde:
b_{bal} - numărul de biţi de alocare
b_{scf} - numărul de biţi pentru factorul de scală
b_{spl} - numărul de biţi pentru eşantioane
b_{anc} - numărul de biţi pentru “ancillary data”



- **Cuantizarea şi codarea**
- Eşantioanele de subbandă sunt codate şi cuantizate de un cuantizor uniform cu o reprezentare simetrică faţă de 0.
- Fiecare eşantion de subbandă S_i este normat la factorul de scală şi cuantizat utilizînd formula :

$$S_{qi} = \left(A\left(\frac{S_i}{scf} \right) + B \right)\Big|_N$$

- Coeficienţii A şi B sunt tabelaţi.

Număr de	A	B

niveluri		
3	0.750000000	-0.250000000
7	0.875000000	-0.125000000
15	0.937500000	-0.062500000
31	0.968750000	-0.031250000
63	0.984375000	-0.015625000
127	0.992187500	-0.007812500
255	0.996093750	-0.003906250

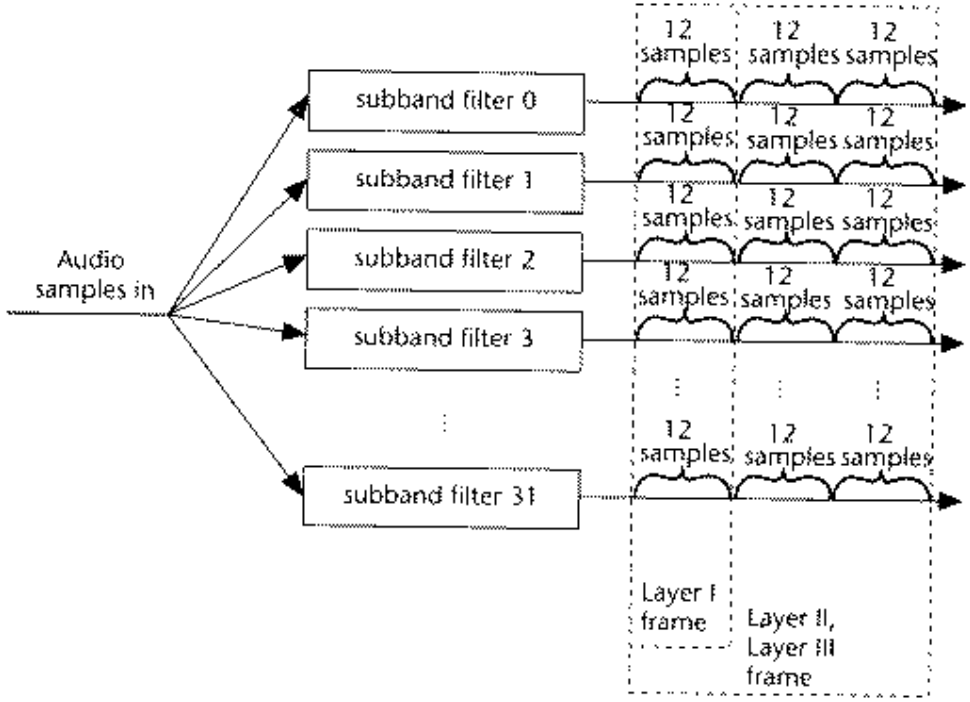
- **Fluxul de biţi**
- Informaţia codată din subbenzi e multiplexată în cadre. Această operaţie nu presupune o codare suplimentară.
- Un cadru este compus dintr-un număr întreg de sloturi pentru a ajusta fluxul mediu de biţi.
- În Layer I un slot are 32 de biţi în timp ce în Layer II şi III un slot are 8 biţi.
- Numărul de sloturi dintr-un cadru se obţine împărţind B_f la numărul de biţi dintr-un slot.
- Dacă frecvenţa de eşantionare este 44.1 kHz numărul de sloturi nu este întreg. În asemenea cazuri cadrul trebuie ajustat prin adăugarea de biţi (padding). Astfel numărul de sloturi dintr-un cadru poate fi N sau N+1.

  Ex:

  F_S=44.1 kHz, 114.84 cadre/sec, 1 cadru=8.70ms

  rezultă 17.41 sloturi => 18 sloturi

- **MPEG Layer II**
- Layer II urmăreşte în principiu aceleaşi reguli de codare şi decodare ca şi Layer I.
- Principala diferenţă este ca Layer II introduce corelatie între subbenzi. Layer II contine informatii pentru 1152 de esantioane
  (3 x 12 x 32 esantioane = 1152 de esantioane).
- In fluxul de date apare şi un selector al factorului de scală.

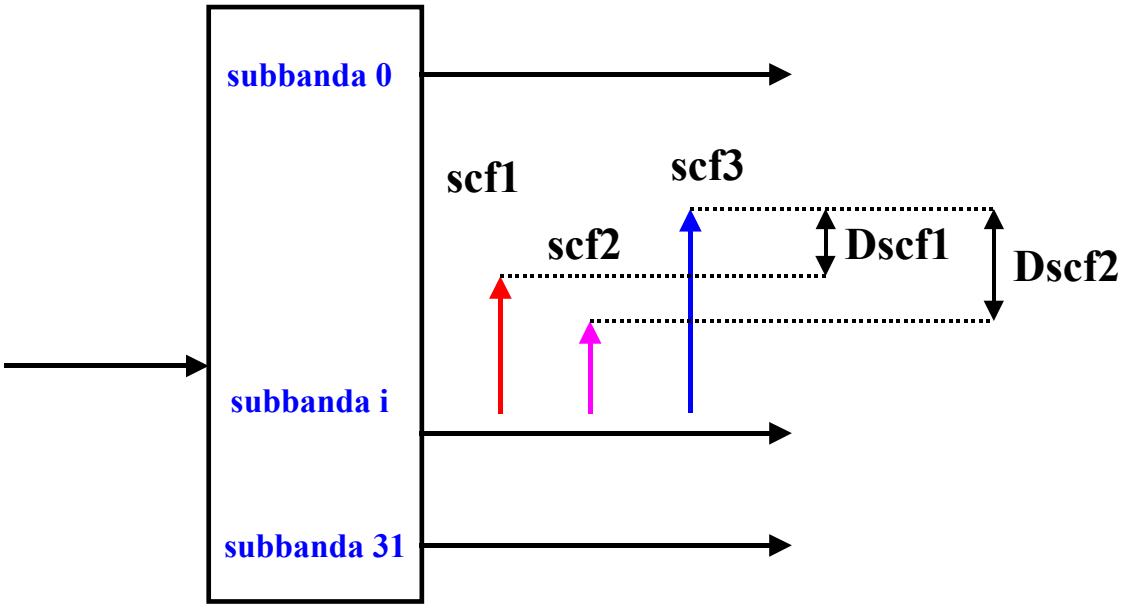


- Layer II suportă atât modelul psihoacustic I cât şi modelul psihoacustic II.

- Modelul psihoacustic I implică un FFT de 1024 esantioane iar modelul II 512 eşantioane.
- SMR din fiecare subbandă se determină din modelul psihoacustic folosit.
- **Codarea factorilor de scalare**
- Se poate folosi aceeaşi analiză şi sinteză a filtrelor ca în cazul Layer I.
- În Layer II un cadru conţine 36 (3 x 12) eşantioane de subbandă (12 granule) şi 3 factori de scală pe subbandă.

- Cele două diferenţe se obţin din cei trei factori de scală după cum urmează:

Dscf1=scf3-scf1
Dscf2=scf3-scf2



- Fiecare diferenţă este clasificată în una din cele 5 clase după cum urmează:

Clasa	Condiţia
1	$Dscf_i \leq -3$
2	$-3 < Dscf_i < 0$
3	$Dscf_i = 0$
4	$0 < Dscf_i < 3$
5	$Dscf_i \geq 3$

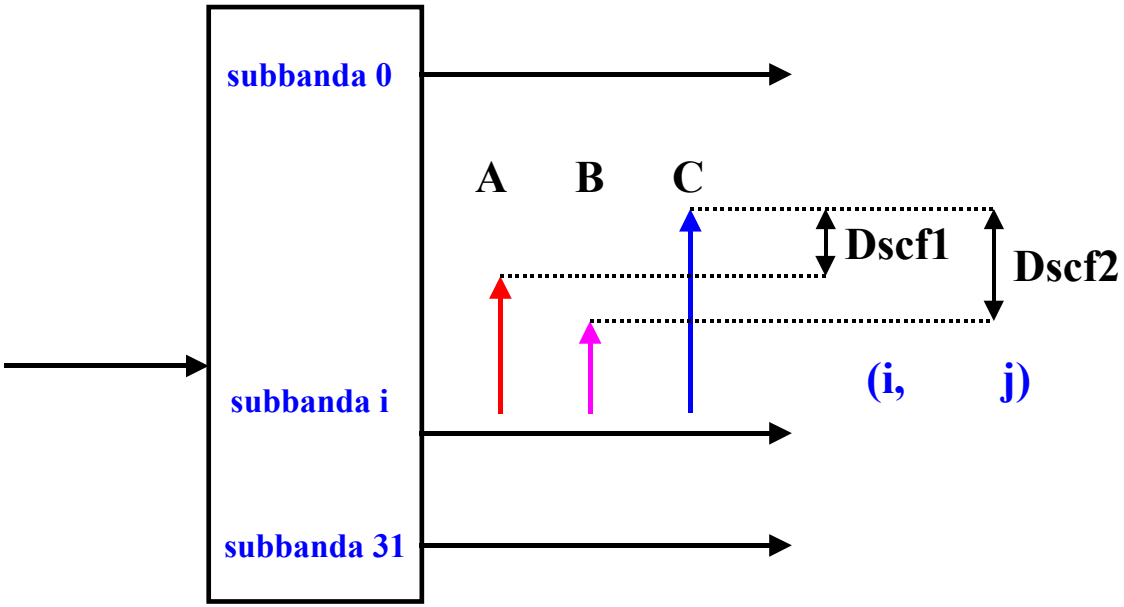
- Cele doua clase ce rezultă corespund transmisiei unui pattern (cei trei factori de scală care trebuie transmişi).
- Redundanţa e redusă cu preţul codării informaţiei de selectare a factorului de scală (2 biţi).

(Clasa1,Clasa2)	Pattern transmis	Factor de scală selectat
(1,1), (1,5), (4,5), (5,1),(5,5)	123	0
(1,2), (1,3), (5,2),(5,3)	122	3
(1,4), (5,4)	133	3
(2,1), (2,5), (3,5)	113	1
(2,2), (2,3),(3,1),(3,2),(3,3)	111	2
(2,4)	444	2
(3,4), (4,4)	333	2
(4,1), (4,2), (4,3)	222	2

- Biţii de selecţie a factorului de scală reprezintă numărul şi poziţia factorilor de scală din fiecare subbandă.

scfsi	Factori de scală	Factor de scală

	codaţi	decodat
0 (00)	3	scf1, scf2, scf3
1 (01)	2	primul □□scf1 şi scf2 al doilea □ scf3
2 (10)	1	scf=scf1=scf2=scf3
3 (11)	2	primul □□scf1 al doilea □ scf2 siscf3



Ex: Presupunem că 3 factori de scală A, B, C sunt obţinuţi într-o subbandă.

Clasa	Factori de scală transmişi	scfs i	Factori de scală decodaţi
(1,1)	ABC	00	ABC
(1,3)	AB	11	ABB
(3,2)	A	10	AAA

- **Alocarea biţilor**

- **SMR** din modelul psihoacustic este folosit pentru a obţine **MNR** şi operaţia iterativă este similară cu cea din Layer I, incluzând şi câmpul de selecţie a factorului de scală.

- **Cuantizarea şi codarea:**
- Acelaşi algoritm folosit la cuantizare în Layer I se aplică şi aici.
- Trei eşantioane succesive (1 granulă) sunt codate ca un singur cuvânt de cod.

- La decodare se va folosi următorul algoritm (s(0), s(1) şi s(2) sunt cele 3 eşantioane codate):

  for i=0 to 2
  s(i)=(code) MOD (număr de nivele)
  code=(code) DIV (număr de nivele)

- Cele trei eşantioane sunt decuantizate după formula:

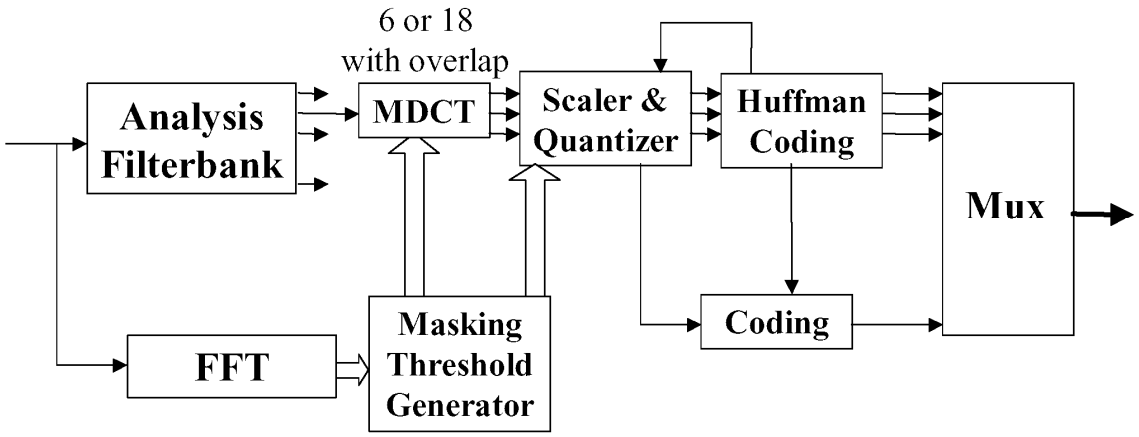
  $$S(i)=C(S''(i)+D)$$

  Unde C şi D sunt constante tabelate.

- **Formarea fluxului de biţi:**

- Aceleaşi operaţii ca şi în Layer I sunt efectuate. Această operaţie nu presupune o codare suplimentară.
- În Layer II un slot are 8 biţi.
- Aceiaşi algoritmi de padding se aplică şi aici.

- **MPEG Layer III**
- Codarea în MPEG Layer III e mult mai sofisticată decât cea din Layer I/II.
- Cu ajutorul unui banc de filtre hibride se obţine o mai bună rezoluţie în frecvenţă.
- Filtrele hibride sunt obţinute prin cascadarea filtrelor polifazice de analiză (folosite şi în Layer I şi II) cu operaţia MDCT (Modifed DCT).
- Modelul perceptual combină calculul energiei cu FFT şi cu bancul de filtre.
- Ieşirile modelului perceptual sunt valorile pragului de mascare echivalent cu valoarea acceptată a zgomotului în fiecare bandă.
- Benzile de frecvenţă sunt egale cu benzile critice.
- Cuantizarea nu mai este uniformă, se introduce codarea entropică, se introduc mai multe bucle pentru modelul psihoacustic şi pentru alocarea de biţi.
- Codarea Huffman se face în funcţie de statistica semnalului muzical alegându-se tabelul de codare optim.



- Ferestrele definite pentru MDCT sunt pentru blocuri lungi şi scurte.
- Pentru blocuri lungi (N=36) formula este:

$$h(k) = x(k)\sin\left(\frac{\pi}{N}\left(k+\frac{1}{2}\right)\right) \quad \text{k=0, 1, ..., 35, N=36}$$

- Pentru blocuri scurte se aplică aceeaşi formulă doar că N=12.
- Comutarea între blocuri nu e instantanee. Pentru aceasta se definesc ferestre de tranziţie (lung => scurt şi scurt => lung).
- Decizia de comutare se ia din curba de mascare obţinută din estimatul entropiei psihoacustice. Dacă valoarea entropiei psihoacustice (PE) depăşeşte un anumit nivel (PE>1800) atunci se va trece la blocul scurt.

- **Transformarea Cosinus Modificată (MDCT)**
- Următoarea ecuaţie se foloseşte pentru a obţine N/2 coeficienţi S_i din N eşantioane de intrare x_k:

$$S_i = \sum_{k=0}^{N-1} x_k \cos\left(\frac{\pi}{2N}\left(2k+1+\frac{N}{2}\right)(2i+1)\right)$$

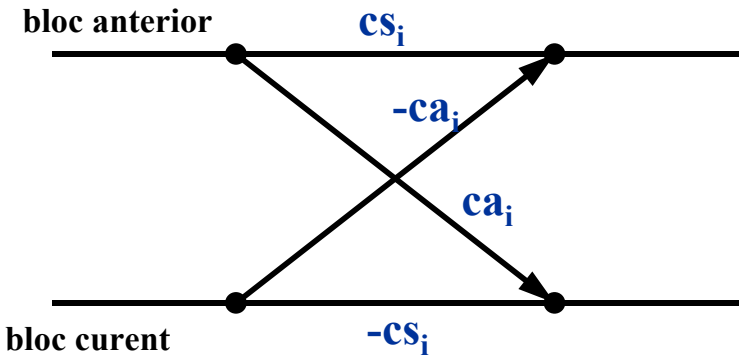
unde : $i = 0,\ 1,\ ...,\ \frac{N}{2}-1$

- N poate fi 12 pentru blocuri scurte şi 36 pentru blocuri lungi.
- Transformarea MDCT inversă are expresia:

$$x_k = \sum_{i=0}^{\frac{N}{2}-1} S_i \cos\left(\frac{\pi}{2N}\left(2k+1+\frac{N}{2}\right)(2i+1)\right)$$

unde k=0, 1, ..., N-1

- **Reducerea efectului de aliere**
- Calculul de reducere a alierii se face atât în codor cât şi în decodor.
- Numai blocurilor lungi li se aplică această procedură.
- Transformarea MDCT dă 18 coeficienţi din 36 de eşantioane de intrare. Între 2 seturi de 18 coeficienţi se aplică un operator fluture ca în figura următoare.



unde i=0, 1, ..., 7 iar c_{si} şi c_{ai} se calculează cu formulele:

$$cs_i = \frac{1}{\sqrt{1+c_i^2}} \qquad ca_i = \frac{c_i}{\sqrt{1+c_i^2}}$$

- Cei 8 coeficienţi c_i sunt tabelaţi:

i	c_i
0	-0.6
1	-0.535
2	-0.33
3	-0.185
4	-0.095
5	-0.0041
6	-0.0142
7	-0.0037

- **Cuantizarea şi codarea**
- Cuantizorul MPEG Layer III este neliniar. Legea de cuantizare este de forma:

$$Q\left(\alpha \cdot x^{\frac{3}{4}}\right)$$

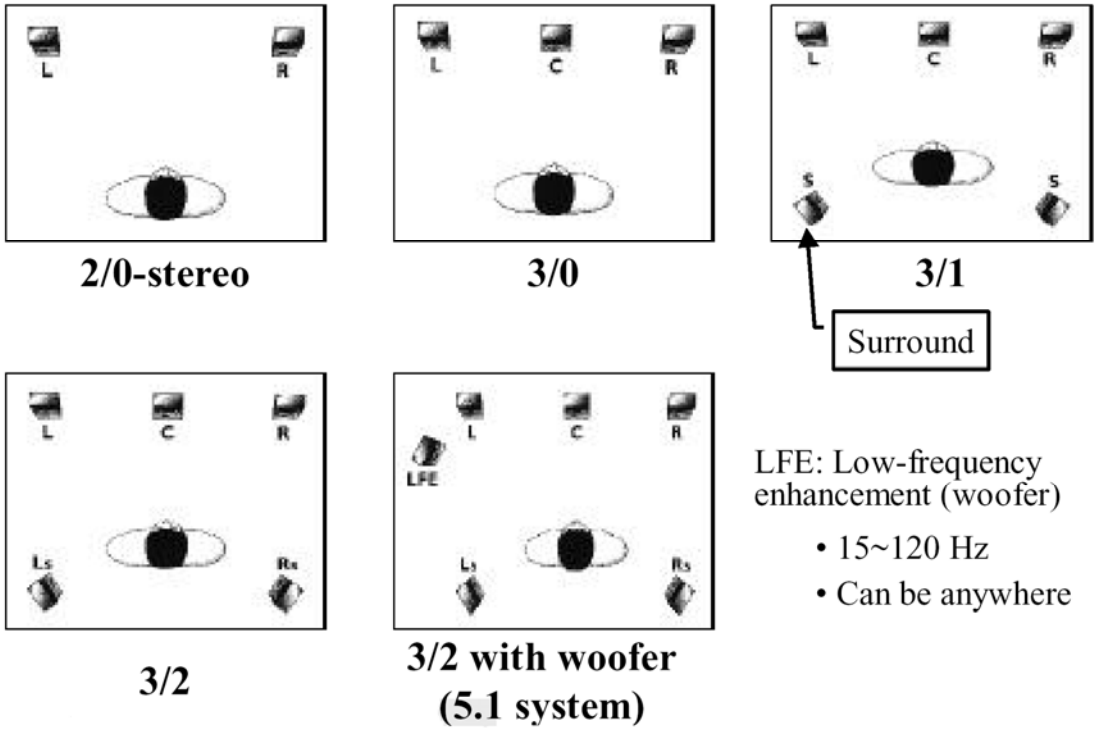
- La decodare va trebui efectuată operaţia inversă adică ridicarea la puterea 4/3.
- Codorul Huffman este utilizat pentru codare entropică.
- Procesul de găsire a câştigului şi factorilor de scalare optimi pentru un bloc, rata de bit şi ieşirea modelului perceptual este realizat în două cicluri iterative prin analiză-sinteză.
- Ciclul interior (ciclul de rată):
  - Codul Huffman alocă valorilor cuantizate mici (cele mai frecvente) cuvinte de cod de lungime minimă.
  - Dacă numărul de biţi rezultat depăşeşte numărul de biţi disponibili pentru codarea unui bloc de date, aceasta se poate ajusta prin modificarea câştigului global care rezultă într-un pas de cuantizare mai mare, ceea ce conduce la valori cuantizate mai mici.
  - Operaţia este repetată cu diferiţi paşi de cuantizare până când cererea de biţi pentru codarea Huffman este suficient de mică.

- Ciclul exterior (ciclul de control al zgomotului):
  - Pentru a dimensiona zgomotul de cuantizare în funcţie de pragul de mascare, se aplică un factor de scalare fiecărei benzi.
  - Sistemul porneşte cu un factor de scalare 1.

- Dacă zgomotul de cuantizare într-o bandă depăşeşte pragul de mascare (zgomotul permis), factorul de scalare pentru această bandă este ajustat pentru a reduce zgomotul de cuantizare.
- Deoarece pentru a reduce zgomotul de cuantizare sunt necesari mai mulţi paşi de cuantizare deci o rată de bit mai mare, ciclul interior de rată este repetat de fiecare dată când se modifică factorii de scalare.
- Ciclul exterior este repetat până când zgomotul (calculat ca diferenţa între valorile spectrale originale şi cuantizate) este sub pragul de mascare.

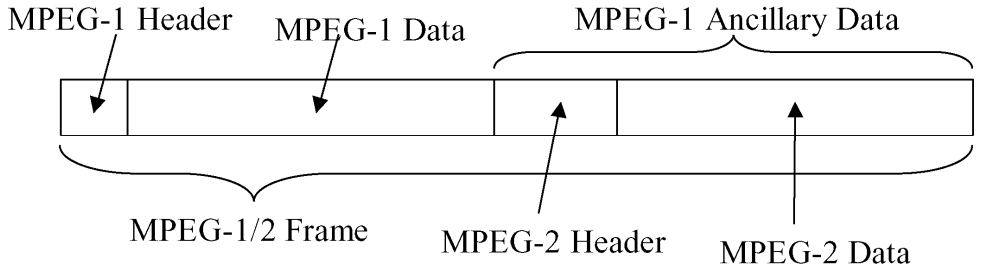
- **Codarea semnalului stereo.**
- MPEG-1 audio codează atât cu semnal mono cât şi stereo.
- Sunt patru moduri de codare: mono, stereo, două canale separate şi joint stereo.
- O tehnică de codare eficientă a semnalului stereo se numeşte ***joint stereo coding***:
  - **Codarea stereo a intensităţii** exploatează redundanţa din semnalele stereofonice bazată pe perceperea la frecvenţe mai mari de 2kHz numai a anvelopei energiei canalelor drept si stâng.
  - **Codarea MS(middle/side) stereo** exploatează redundanţa din semnalele stereofonice bazată pe codarea sumei şi diferenţei dintre canalele drept şi stâng.

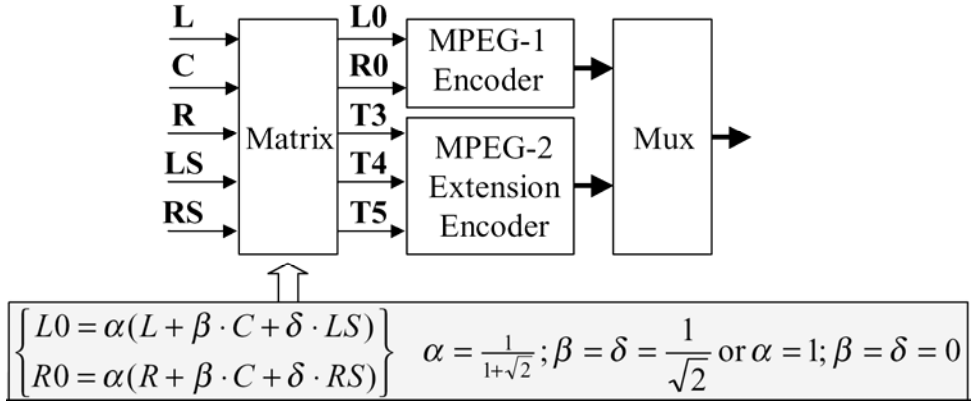
- **MPEG-2 AUDIO**
- Permite şi codarea semnalelor cu frecvenţe mai mici de eşantionare: 16, 22 şi 24kHz.
- Realizează o analiză în frecvenţă cu rezoluţie mărită.
- Include codorul MPEG-1 (Layer I, II şi III)
- Codare multicanal:
  - Permite codarea a 2 până la 5 canale: sunet surround sau coloana sonoră pentru mai multe limbi

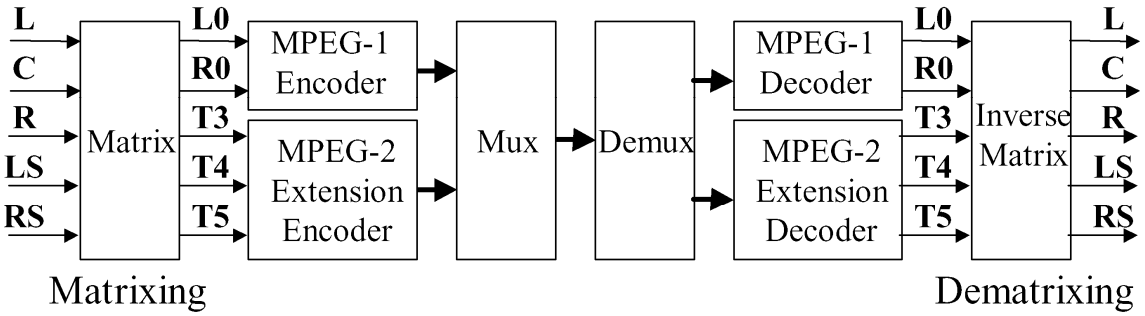


- **Compatibilitatea MPEG audio.**
- Compatibilitate directă (forward):
  - Un decodor nou poate decoda un flux de biţi creat de un codor mai vechi.
  - Se poate obţine relativ uşor.
- Compatibilitate inversă (backward):
  - Un decodor mai vechi poate decoda un flux de biţi creat de un codor nou, cel puţin parţial.
  - Limitează eficienţa codării.

- Codorul audio MPEG-2 **compatibil în sens invers** (ISO/IEC 13818-3):

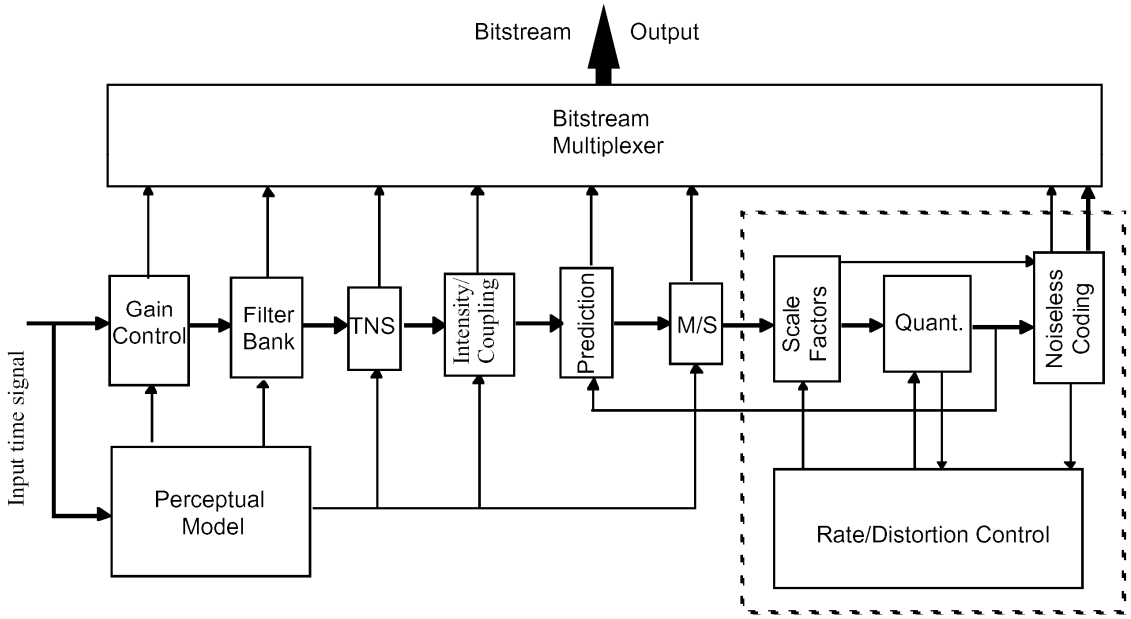






- Codarea **Non Backward Compatible** (NBC)
- MPEG-2 **Advanced Audio Coding (AAC)** ISO/IEC 13818-7 (Aprilie 1997).
- Rata de codare: 320-384 kbiţi/s pentru 5 canale, 64 kbiţi/canal.
- Semnal codat NBC la 320kbiţi/s are aceeaşi calitate ca semnalul codat BC la 640kbiţi/s.

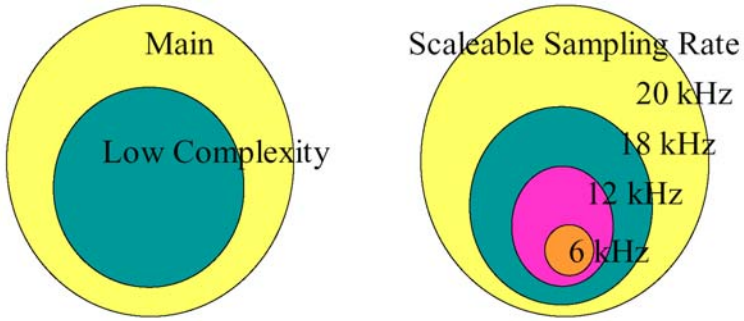
- Permite codarea multicanal: 1-48 canale audio, 0-16 canale LFE (low frequency enhancement), 0-16 canale de date.
- Aceeaşi structură (codare perceptuală pe subbenzi) ca la MPEG-1 cu unele îmbunătăţiri.



- Îmbunătăţiri
  - Banc de filtre cu rezoluţie mărită (MDCT în 1024 sau 128 puncte) cu răspuns la impuls micşorat la 5.3 ms (faţă de 18.6 ms la Layer III) reduce distorsiunile de tip pre-echo (zgomotul de cuantizare se aude înaintea muzicii care îl produce).
  - Cuantizarea dependentă de evoluţia în timp a semnalului (Temporal noise shaping TNS).
  - Predicţie inversă în subbenzi oferă o codare eficientă a semnalelor tonale.

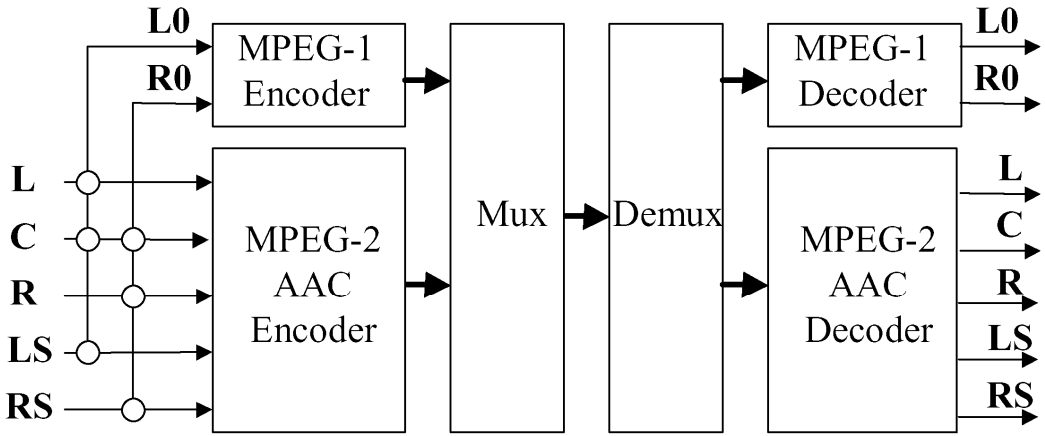
- Codare stereo Middle/Side şi de intensitate mai flexibilă reduce rata de bit.
- Codare Huffman cu tabele de codare pe fiecare bloc al codorului.

- **Profiluri MPEG-2 AAC**



- Profilul principal
  - Cea mai bună calitate, complexitate maximă
  - MDCT în 1024 sau 128 puncte
- Profilul de complexitate redusă
  - Fără predicţie şi TNS
- Profil cu frecvenţa de eşantionare scalabilă
  - Complexitatea şi frecvenţa de eşantionare sunt scalabile
  - Foloseşte filtre hibride ca la MPEG-1 Layer III
  - Fără predicţie şi intercorelare canal

- Pentru a obţine compatibilitate în sens invers dar cu o rată de bit mai mare se poate folosi schema (Simulcast):



- **MPEG-4 AUDIO**
- MPEG-4 Audio integrează codarea audio sintetizată şi naturală.
- Partea de codare sintetizată cuprinde realizarea muzicii şi vorbirii definite simbolic. Include sisteme MIDI şi Text-to-Speech. În plus, sunt incluse tehnici de localizare 3-D a sunetului, permiţând crearea unor medii de sunet artificiale folosindu-se surse artificiale şi naturale.

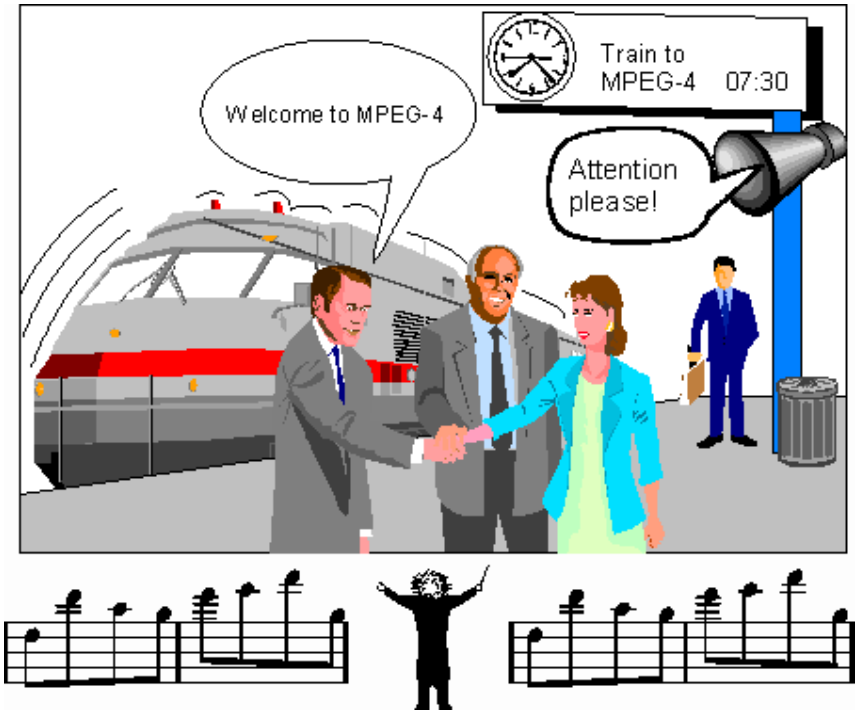
- **Codarea audio naturală**
- pentru debite între 2 kbiti/s şi 64 kbiti/s.
- trei tipuri de codecuri:
  - un **codec parametric** pentru cele mai mici debite

- un **codec CELP** (Code Excited Linear Predictive) pentru debite medii:
- **codecuri timp-frecvenţă** (TF) incluzând MPEG-2 AAC şi Cuantizare Vectorială.

- Sunt oferite facilităţi pentru o gamă largă de aplicaţii de la vorbirea inteligibilă la audio-multicanal de înaltă calitate.
- În MPEG-4 sunt incluse funcţii adiţionale
  - controlul vitezei la redare.
  - modificarea înălţimii sunetului.
  - înlăturarea erorilor.
  - scalabilitatea.

- **Obiecte audio MPEG-4**
- MPEG-4 defineşte obiectele audio ca obiecte "realistice".
- Un obiect audio "real-world" poate fi definit ca o entitate semantică audibilă (vocea unor vorbitori, instrumente muzicale etc.).
- Acesta poate fi înregistrat cu un microfon (înregistrare mono) sau cu mai multe microfoane în direcţii diferite (înregistrare multicanal).
- Obiectele audio pot fi grupate sau mixate împreună dar nu pot fi (uşor) descompuse în sub-obiecte.
- Un singur obiect audio poate fi reprezentat pe unu sau mai multe canale audio, dacă definim canalele audio ca informaţia pentru poziţia unei boxe. De exemplu un flux audio MPEG-1 poate fi un obiect audio în MPEG-4. Acest obiect poate conţine un canal (mono) sau 2 canale (stereo etc.)

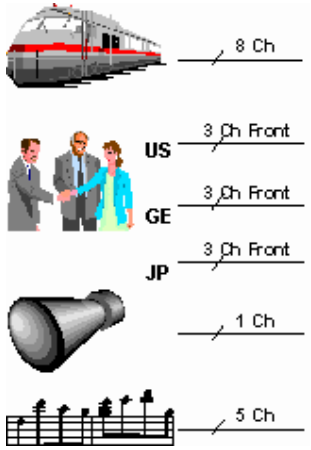
- **Exemple de aplicaţii tipice pentru MPEG-4 Audio**
- **Cântă N-1 Obiecte Audio**
  - Transmiterea a cinci semnale multicanal care reprezintă cinci instrumente ale unui cvintet. Ascultătorul poate asculta numai patru instrumente deoarece vrea sa cânte el la al cincilea instrument.
- **Servicii de difuzare în mai multe limbi**
  - Cei ce urmăresc programele sportive sunt frecvent distraşi de vocea comentatorului. MPEG-4 permite un "mix-minus" stil de prezentare unde să fie incluse toate sunetele, mai puţin vocea comentatorului.
  - Alternativ, într-un serviciu multi-limbi, poate fi inclus unul din comentariile în limbi străine.
- **Filme**
  - O scenă la gară dintr-un film poate conţine de exemplu patru tipuri de obiecte audio:



- **Obiectul conversaţie**:
  - Vocea ‘welcome’ este cu siguranţă cea mai importanta informaţie.
  - Vorbirea este întotdeauna localizată în faţa ascultătorului.
  - Această conversaţie poate fi de asemenea disponibilă în mai multe limbi.
- **Obiectul fundal**:
  - Trenul va veni din depărtare spre centrul scenei, va trece de ascultător şi va dispare în spatele lui.

- În plus canalul pentru efecte de joasă frecvenţă va produce un zgomot de huruit.
- Deşi includerea acestui obiect este dorită, el poate fi exclus în cazul unei conexiuni cu debit foarte redus.

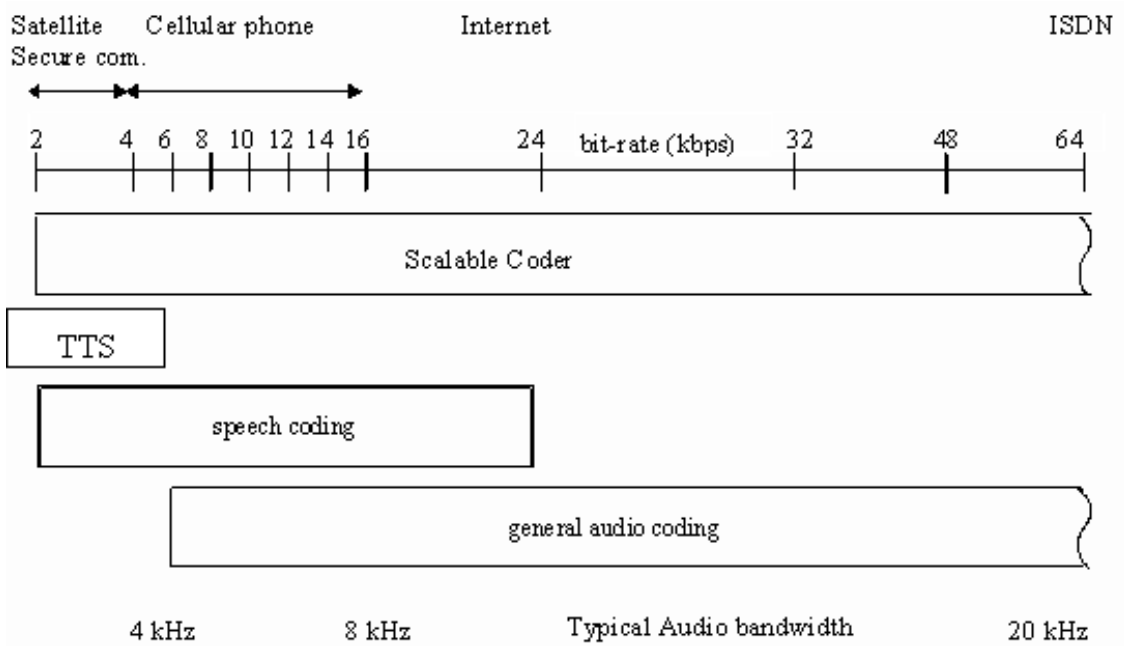
- **Obiectul anunţ**:
  - Pentru anunţ este suficient de transmis vorbire cu calitate redusă.
  - Pot fi generate uşor unele efecte pseudo 3D şi de ecou la prezentarea scenei.
- **Muzica de fundal**:
  - Orchestra poate fi codată cu MPEG-2 mutlicanal şi fluxul de biţi poate fi folosit fără necesitatea recodării.



- **Obiecte audio multi-limbă**
- Pentru o producţie internaţională mai mult de un obiect conversaţie este necesar.

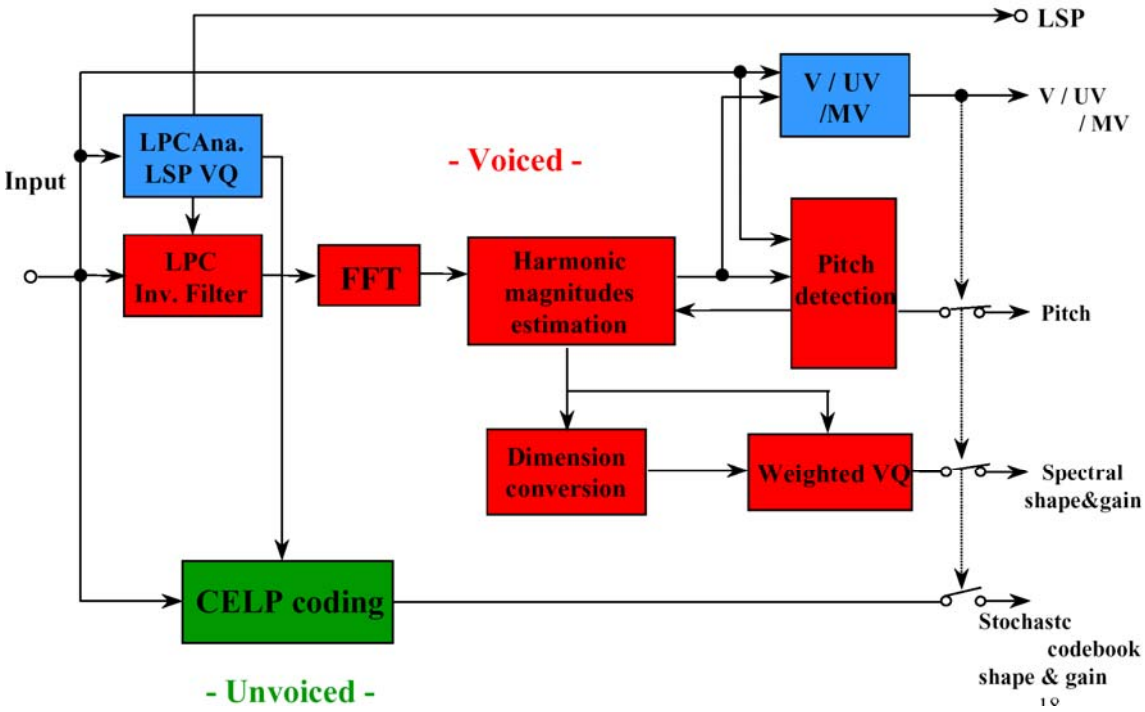
- Acelaşi obiect audio din scenă poate exista în mai multe limbi.
- Fiecare limbă este un obiect audio separat, va fi codată cu un codor independent, şi va fi selectată la cerere în decodor.

- **Codarea obiectelor audio**
- Codarea MPEG-4 a obiectelor audio oferă tehnici pentru reprezentarea sunetelor naturale şi pentru sunetele sintetizate pe baza descrierii structurii.
- Reprezentarea pentru sunetele sintetizate poate deriva dintr-un şir de date sau aşa numita descriere de instrument şi prin codarea parametrică pentru a furniza efecte ca reverberaţia şi spaţializarea.
- Această reprezentare avantajează compresia şi alte funcţii cum ar fi scalabilitatea şi redarea la diferite viteze.
- MPEG-4 standardizează codarea audio naturală pentru debite între 2 kbiţi/s şi 64 kbiţi/s.



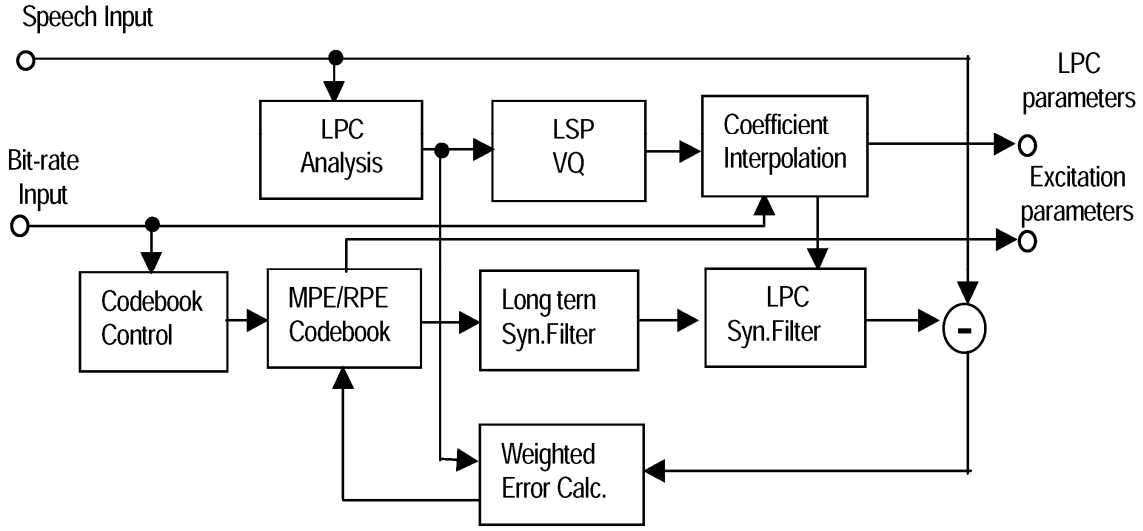
- Pentru obţinerea celei mai bune calităţi posibile pentru toate debitele şi să ofere şi funcţii suplimentare, în standard au fost incluse trei tipuri de structuri de codare:
- **Tehnici de codare parametrică (HVXC)**,
  - Codare voce cu 8 kHz frecvenţă de eşantionare la rate de bit foarte mici (între 2 – 4 kbiţi/s).
  - **Scalabilitatea ratei de bit**: Este posibilă decodarea la 2kbiţi/s dintr-un flux de bit codat cu 4kbiţi/s.
  - **Variaţia vitezei de redare şi a pitch-ului**: Utilă pentru căutarea în baze de date de vorbitori.
  - Sunt combinate două tipuri de scheme de codare: una pentru segmente vocale şi alta pentru segmente nevocale.

- Voce: Informaţia de fază este eliminată la reprezentarea spectrului de putere a erorii de predicţie a filtrului LPC.
- Nevocal: Parametrii consoanelor sunt obţinuţi cu codorul CELP.



- **Tehnici de codare Code Excited Linear Predictive (CELP)**.
  - Codarea vorbirii la debite medii între 6 –24 kbiţi/s.

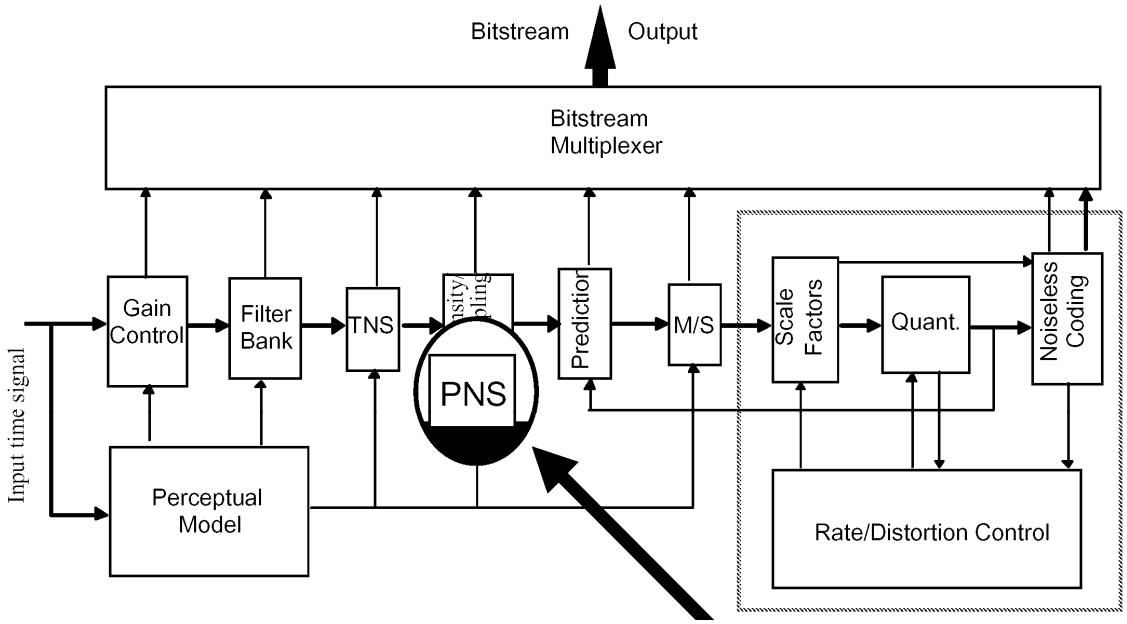
- În această zonă, două frecvenţe de eşantionare, 8 şi 16 kHz, sunt folosite pentru vorbirea de bandă îngustă şi bandă largă.
- Banda îngustă: 3,85-12,2 kbps, pentru cadre de 10-40 ms.
- Bandă largă: 10,9-23,8 kbps, pentru cadre de 10-20 ms.



- **Tehnici de codare timp-frecvenţă (T/F)**,
  - Pentru debite peste 16 kbiţi/s semnale audio.
  - Se folosesc în principal codoarele TwinVQ şi AAC.
  - Frecvenţele de eşantionare sunt peste 8 kHz.

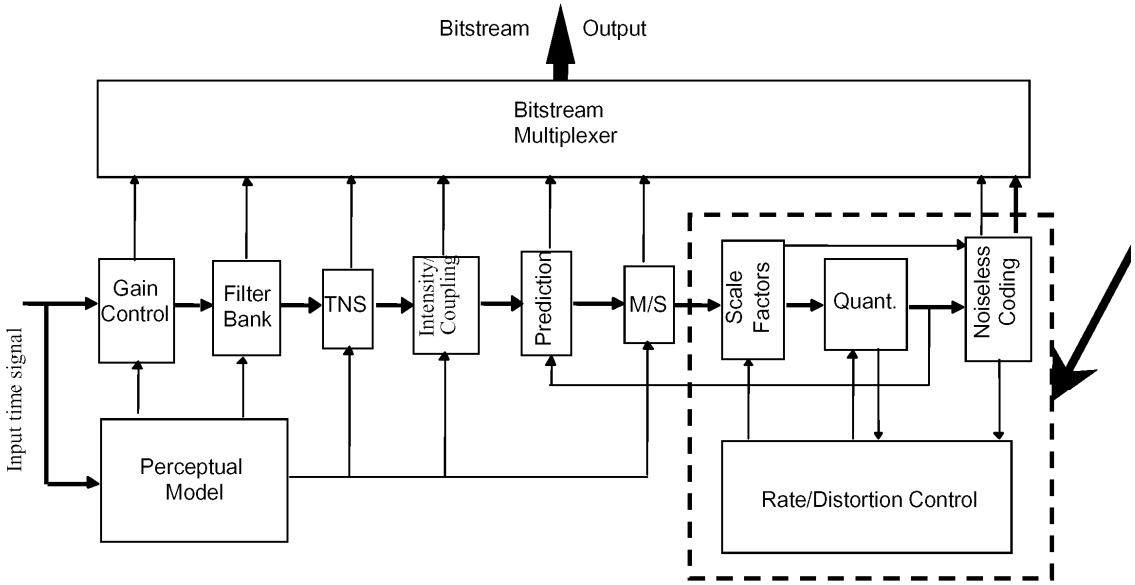
  - Extensii la AAC:

- **Substituţia zgomotului perceptual (PNS)**



- Codarea parametrică a semnalelor asemănătoare zgomotului se foloseşte în codarea vorbirii (consoane).
- **Perceptual Noise Substitution** (PNS) permite o codare selectivă a frecvenţelor pentru semnale similare zgomotului.
- Componentele ca de zgomot se detectează în funcţie de factorul de scalare al benzii.
- Coeficienţii spectrali corespunzători nu sunt cuantizaţi şi codaţi. In loc de aceştia se transmite un flag de înlocuire cu zgomot şi puterea totală a benzii substituite.
- Decodorul generează semnal pseudo aleator cu puterea echivalentă a coeficienţilor spectrali.

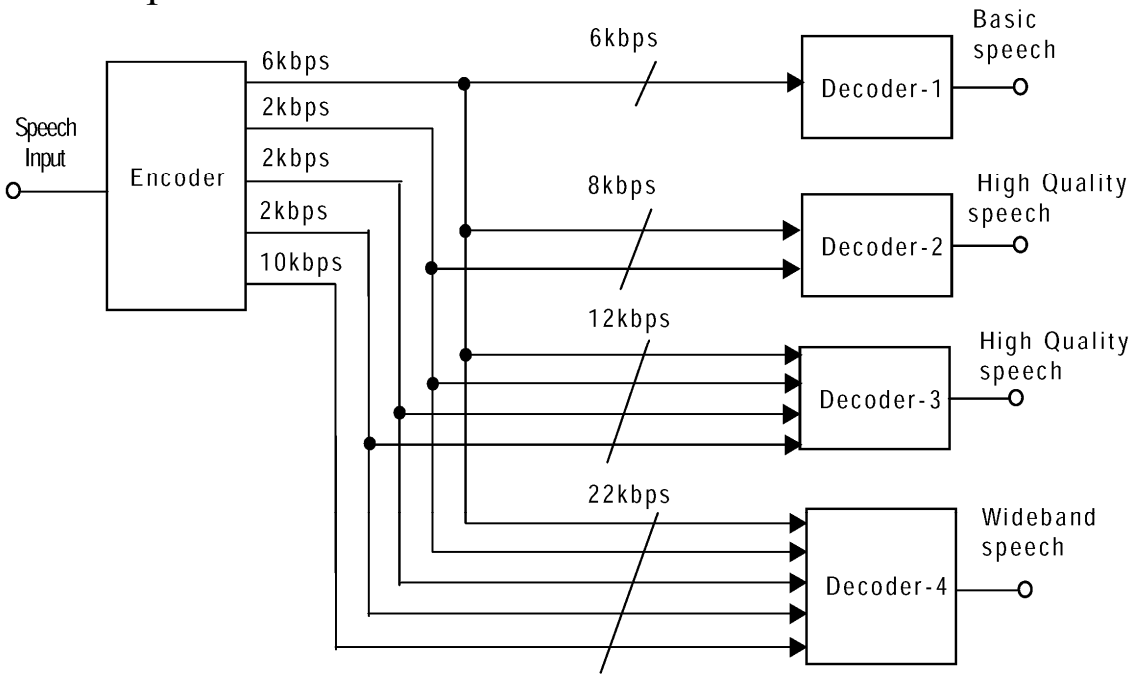
- **Predicţie pe termen lung**
- Semnalele tonale necesită precizie la codare mai mare decât semnalele similare zgomotului (netonale).
- Componentele tonale sunt predictibile
- Predicţia fiecărui coeficient spectral se face în MPEG-2 AAC cu un predictor invers adaptiv. Acesta are complexitate mare (50% din complexitatea decodării).
- În MPEG-4 se foloseşte Long Time Predictor (LTP) cunoscut în codarea vorbirii.
- Acesta are complexitate redusă (cu 50% mai mică faţă de MPEG-2 la aceleaşi performanţe)
- **Codecul TwinVQ** (Transform-Domain Weighted Interleave Vector Quantization)
- Codare audio la rate de bit extrem de mici (6-8 kbiţi/s)
- Codoarele CELP nu se comportă bine la codarea muzicii.
- La rata dorită se obţin 0,5 biţi pe componenta de frecvenţă!
- Selectează vectorul codat controlat de modelul perceptual.
- Este complet integrat în MPEG-4 AAC.
- Foloseşte aceeaşi reprezentare spectrală ca şi codorul AAC.
- Foloseşte facilităţile MPEG-4 (LTP, TNS, joint stereo)



- **Structura TwinVQ:**
- Normalizarea coeficienţilor spectrali:
  - Anvelopa LPC (curba globală a spectrului)
  - Codarea componentelor periodice (componente armonice)
  - Codarea curbei după scara bark.
- Cuantizarea Vectorială (VQ)
  - Întreţeserea coeficienţilor spectrali în sub-vectori
  - Cuantizarea vectorială se face cu două seturi de cuvinte de cod.

- **Scalabilitatea codorului audio MPEG-4**

- Există mai multe tipuri de scalabilitate:
  - **Scalabilitatea debitului** permite unui flux de biţi să fie partiţionat într-un flux cu debit mai mic care să poată fi încă decodat într-un semnal inteligibil. Partiţionarea poate fi efectuată fie în timpul transmisiei sau la decodor.

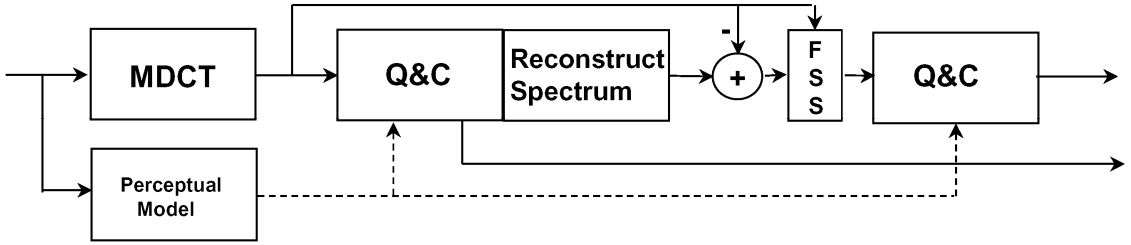


  - **Scalabilitatea benzii de frecvenţă** este un caz particular al scalabilităţii debitului, unde o parte din fluxul de biţi reprezintă o parte din spectrul de frecvenţă care poate fi ignorat în timpul transmisiunii sau la decodare.

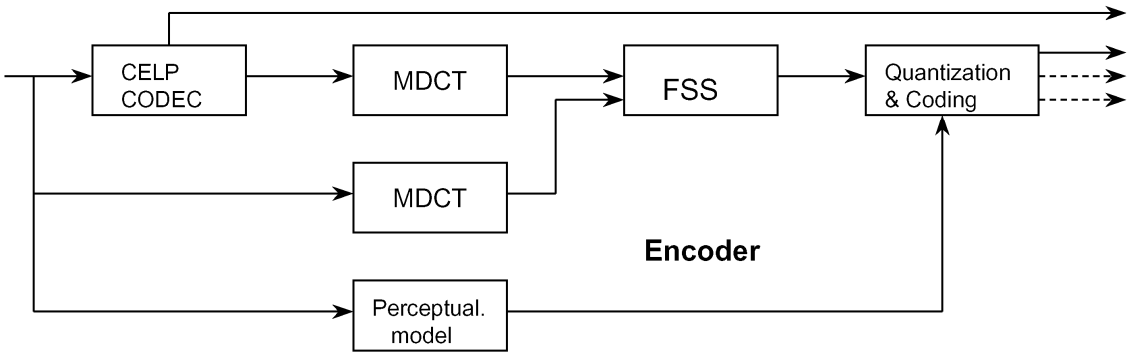
- **Scalabilitatea complexităţii codorului** permite ca codoare de complexitate diferită să genereze fluxuri de biţi valide şi inteligibile.
- **Scalabilitatea complexităţii decodorului** permite ca un flux de biţi să fie decodat de decodoare cu diferite niveluri de complexitate.

- Scalabilitatea funcţionează cu unele din tehnicile MPEG-4, dar poate fi aplicată şi unei combinaţii de tehnici (de exemplu cu Twin VQ ca layer de bază şi AAC pentru layere extinse).

- Exemplu: Codarea semnalului eroare de cuantizare al unui modul AAC sau TwinVQ ca intrare într-un al doilea modul cuantizare/codare în frecvenţă.



- Exemplu: Combinarea cu codor CELP:



- **Codarea audio sintetizată**

- **Codarea Text To Speech (TTS)**
- Codoarele TTS asigură un debit între 200 biţi/s şi 1.2 kbiţi/s şi permit ca să se genereze o vorbire sintetizată inteligibilă, primind la intrare text sau text şi parametrii prozodici (conturul înălţimii, durata fonemelor etc.)
- MPEG-4 oferă o interfaţă standard pentru operarea unui codor TTS şi nu standardizează un anume sintetizor TTS.
- Sunt incluse următoarele funcţionalităţi:
  - Sinteza vorbirii folosind prozodia vorbirii originale.
  - Controlul sincronizării buzelor cu informaţia despre foneme.
  - Pauză, reluare, derulare înainte/înapoi.
  - Suport pentru limbi străine şi dialecte pentru text.

- Suport pentru simboluri de foneme internaţionale, şi suport pentru specificarea vârstei, sexului, debitului verbal al vorbitorului.

- **Sinteza după partitură**
- Tehnicile de Structurare Audio decodează datele de intrare şi produc sunete.
- Această decodare este condusă de un limbaj special de sinteza numit SAOL (Structured Audio Orchestra Language), standardizat ca parte a MPEG-4.
- Acest limbaj e utilizat pentru a defini o “orchestră" alcătuită din “instrumente” (provenite din fluxul de biţi şi nu fixate în terminal) care creează şi procesează data de control.
- Un instrument este o mică reţea de primitive de procesare de semnal care poate emula sunete specifice ca ale instrumentelor acustice naturale.
- Reţeaua de procesare a semnalului poate fi implementată hardware sau software şi include generarea şi procesarea sunetelor şi manipularea sunetelor pre-stocate.
- MPEG-4 nu standardizează o metoda de sinteză ci mai degrabă o metodă de descriere a sintezei.
- Orice metodă curentă sau viitoare poate fi descrisă în SAOL, inclusiv sinteza wavetable, FM, aditivă, modelare psihică şi granulară, precum şi metode hibride non-parametrice.
- Controlul sintezei este desăvârşit prin extragerea “partiturii” sau “scenariului” din fluxul de biţi.
- O partitură este un set de comenzi în timp care invocă diferite instrumente la momente de timp specifice, fiecare contribuind la interpretarea globală a muzicii sau la generarea efectelor sonore.

- Descrierea partiturii, integrată într-un limbaj numit SASL (Structured Audio Score Language), poate fi folosită pentru a crea sunete noi şi de a include informaţii adiţionale de control pentru modificarea sunetului existent.
- Aceasta permite compozitorului un control mai fin asupra sunetului final sintetizat.
- Pentru sinteza care nu necesită un control aşa de fin, se poate utiliza protocolul MIDI pentru controlul orchestrei.
- Controlul fin împreună cu definirea de instrumente proprii, permite generarea unor sunete pornind de la simple efecte audio cum ar fi zgomot de paşi sau de uşi închise, până la simularea sunetelor naturale cum ar fi ploaia sau de la muzica cântată pe instrumente convenţionale pâna la sunete integral sintetizate pentru efecte audio complexe sau muzica futuristă.
- Pentru terminale cu mai puţine facilităţi şi pentru aplicaţii care nu necesita o sinteza atât de sofisticată, un "wavetable bank format" (SASBF) este standardizat.
- Cu acest format pot fi extrase eşantioane de sunet care vor fi folosite în sinteza wavetable, de asemenea şi procesări simple cum ar fi: filtre, reverberaţii şi efecte de cor.
- În acest caz, complexitatea de calcul pentru procesul de decodare poate fi determinată exact, examinându-se fluxul de biţi.

- **Efecte audio speciale**
- Decodorul bazat pe Structurarea audio/Efecte permite la decodare un flux de date care să includă atât canalele audio decodate cât şi parametrii necesari pentru controlul efectelor (desfăşurarea lor în timp etc.)

- Efectele sunt în esenţă descrieri de instrumente “speciale” servind procesoarelor de efecte aplicate asupra fluxului de intrare.
- Procesarea de efecte include reverberatoare, spaţializatoare, mixere, limitatoare, controlul dinamicii, filtre, flangere, coruri şi efecte hibride.
- Avându-se în vedere aceste facilităţi, se poate realiza pe lângă compoziţia muzicală, organizarea altor tipuri de audiţii cum ar fi voce, efecte sonore şi ambianţă generală.