UNIVERSITATEA TEHNICĂ DIN CLUJ-NAPOCA

Facultatea de Electronică, Telecomunicații și Tehnologia Informației

Tehnici de compresie a semnalului vocal

Prof. Mircea GIURGiu
Mircea.Giurgiu@com.utcluj.ro

www.etti.utcluj.ro

Str. George Barițiu nr. 26-28, 400027, Cluj-Napoca
Telefon: 0264-401224, tel/fax: 0264-591689

Activitate

Cursuri:

Microprocesoare (III. Engl),
Prelucrarea semnalului vocal (IV TST),
Tehnici de compresie a semnalului vocal (I TM)

Cercetare (http://speech.utcluj.ro):

recunoasterea automata a vorbirii
sinteza din text a vorbirii
sisteme de invatare automata bazate pe AI (deep neural networks)

Proiecte de cercetare (relevante)

Sound2Sense (2007-2011), Key2nature (2009-2011)
http://simple4all.org (2011-2014)
http://speech.utcluj.ro/swara (2014-2018)
http://speech.utcluj.ro/sintero (2018 – 2021)

CONTINUT

I.1 Prezentare curs

- a. Obiectul
- b. Organizare
- c. Evaluare
- d. Continut

I.2 Standarde in compresia de semnal vocal

I.1.a Obiectul cursului

- Studiul tehnicilor de compresie a semnalelor audio, experimente de laborator si dezvoltarea de aplicatii

- Cunostinte preliminare
  - prelucrarea numerica a semnalelor (esantionare, cuantizare, filtrare numerica, analiza spectrala FFT, autocorelatie,
  - teoria informatiei (entropie, surse de informatie, codare CRC/Hamming/convolutionala)
  - Laborator / proiect: basic programare Matlab / algoritmi

I.1.b Organizare

- Curs – 2h/sapt, MS Teams
  - PPT + resurse bibliografice (sufficient pentru examen)
  - Cartea “Wai C. Chu - Speech_Coding_Algorithms” (vezi repository)
  - 1 prezentare PPT / student – la curs (dezbatere, urmata de evaluare)
  - participarea studentilor / dialog / problematizare
- Laborator
  - 2h/ la 2 saptamani – de eficientizat (e.g. cu grupa pentru scenariu online)
  - Raport de laborator cu rezultatele obtinute
- Activitate cercetare
  - 1h / 2 saptamani
  - modele de teme disponibile in resursele online
  - raportare progress (continuu, formal pe 1 slide/luna)
  - prezentare finala (ultima saptamana)

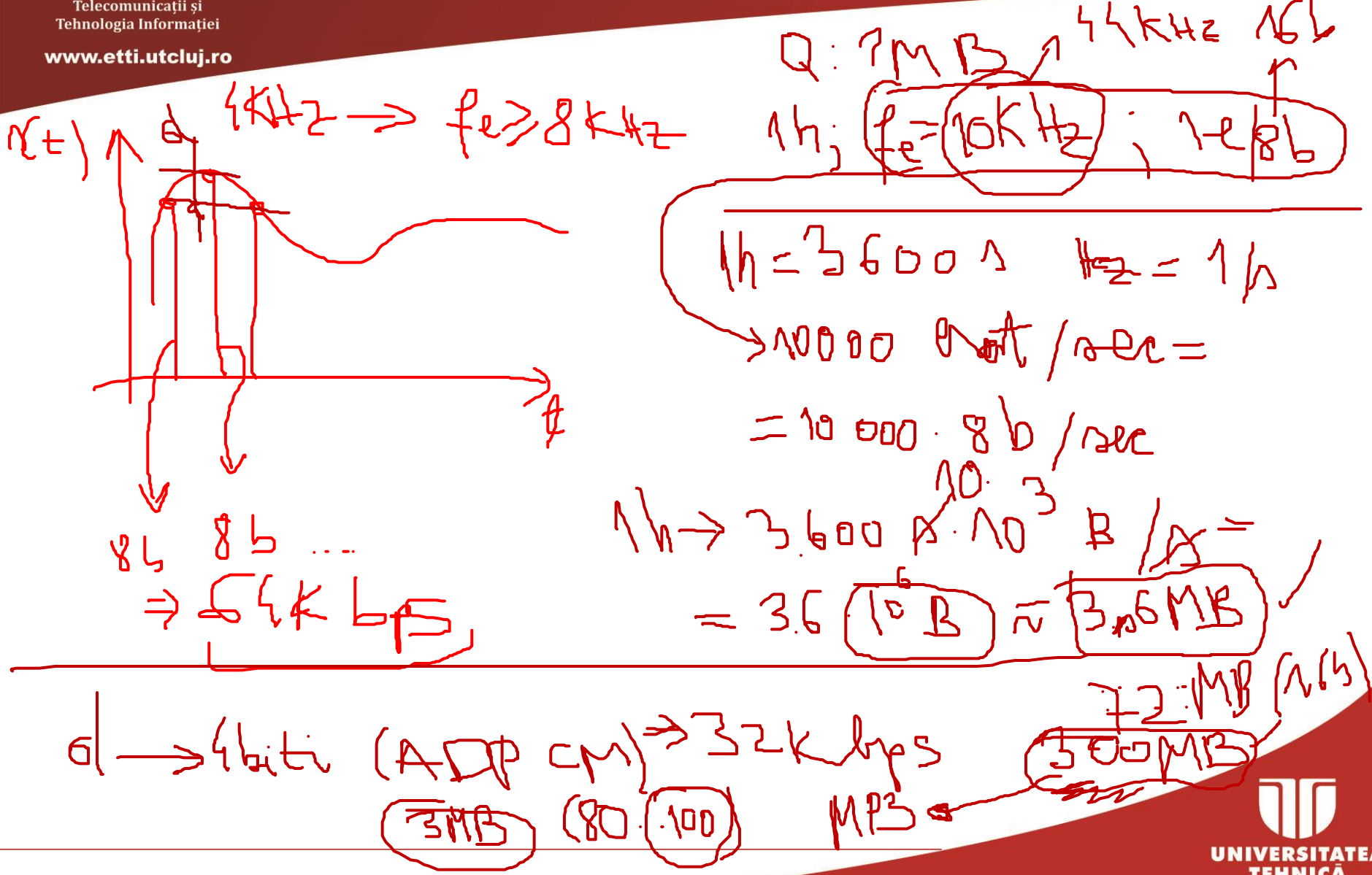
I.1.c Evaluare

- 1,0 pct = prezentare PPT la curs
- 2,0 pct = 0,4 x 5 evaluari dupa prezentare PPT
- 1,5 pct = 0,5 x 3 rapoarte lab (L1&2, L3&4, L5&6)
- 2,5 = activitate cercetare
- 3,0 = evaluare examen

I.1.d Continut

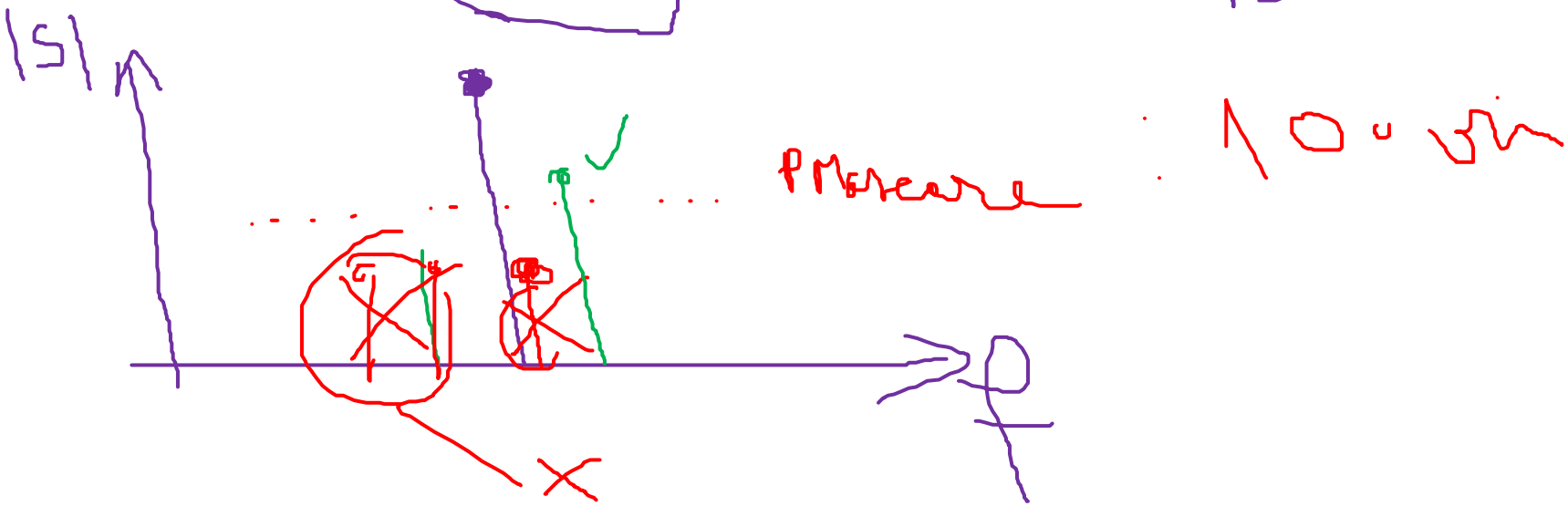
- Sisteme de codare in domeniul timp
  - PCM, APCM, DPCM, ADPCM, Delta

I.1.d Continut



I.1.d Continut

- Sisteme de codare in domeniul frecventa
  - Subbenzi, MPEG Layer I, II, III



I.2. Aplicatii care folosesc Prelucrarea Semnalului Vocal

(b) Aplicatii pentru industria multimedia (semnale de foarte inalta calitate, muzica)

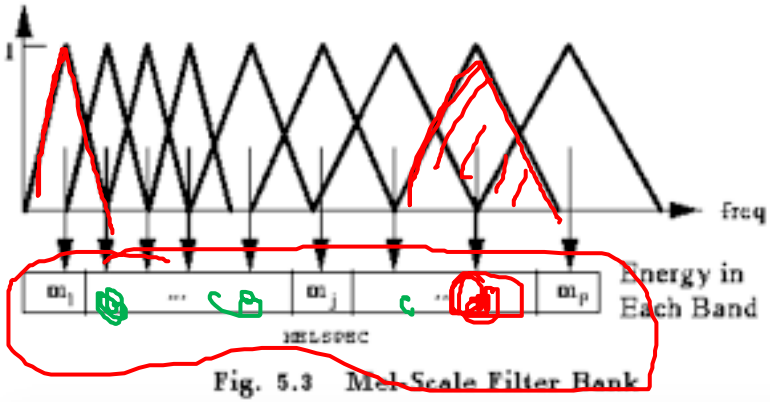
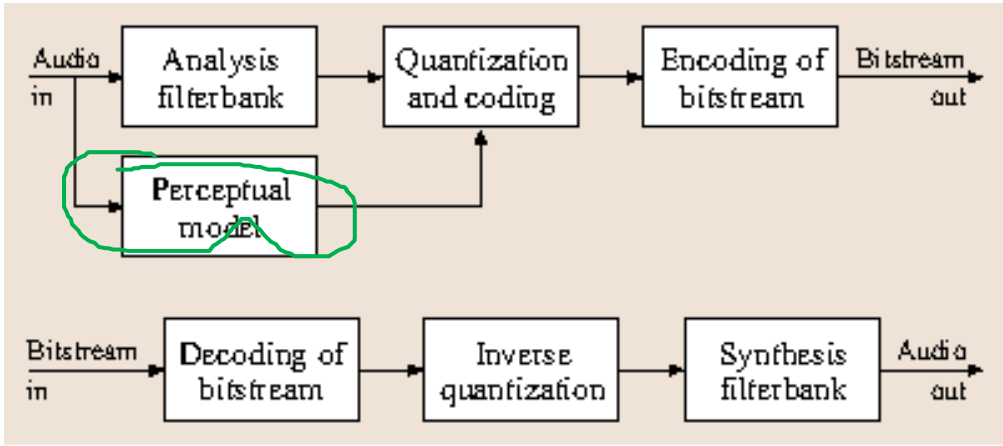
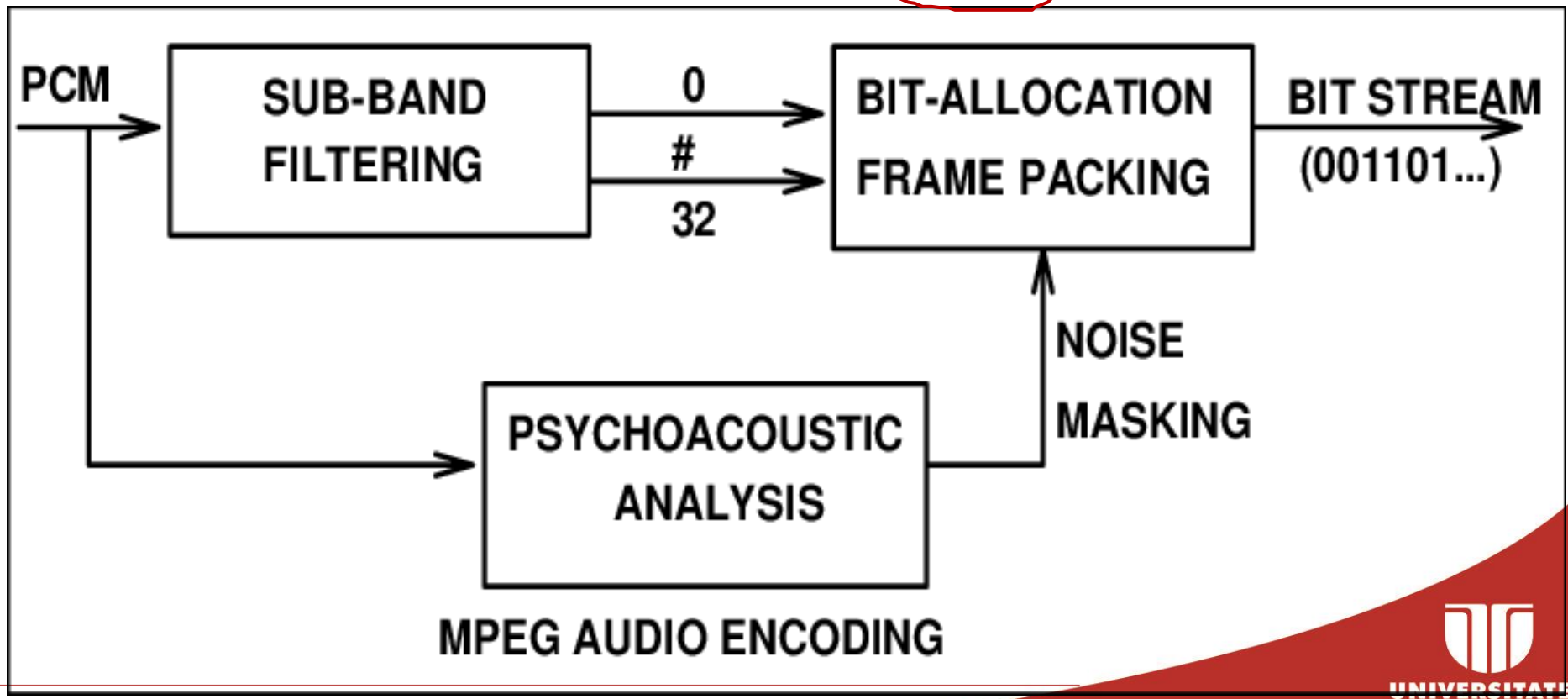


Fig. 5.3 Mel-Scale Filter Bank



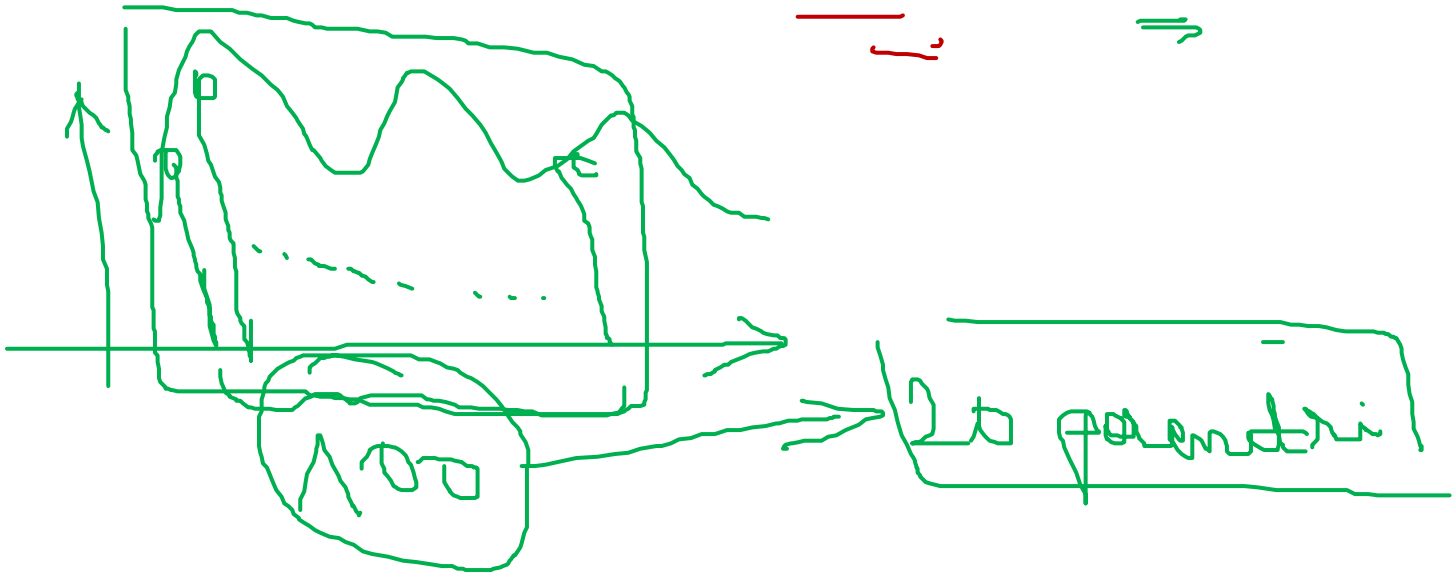
I.2. Aplicatii care folosesc Prelucrarea Semnalului Vocal

(a)Aplicatii pentru industria multimedia (semnale de foarte inalta calitate, muzica) – ex. Standardul MPEG / MP3



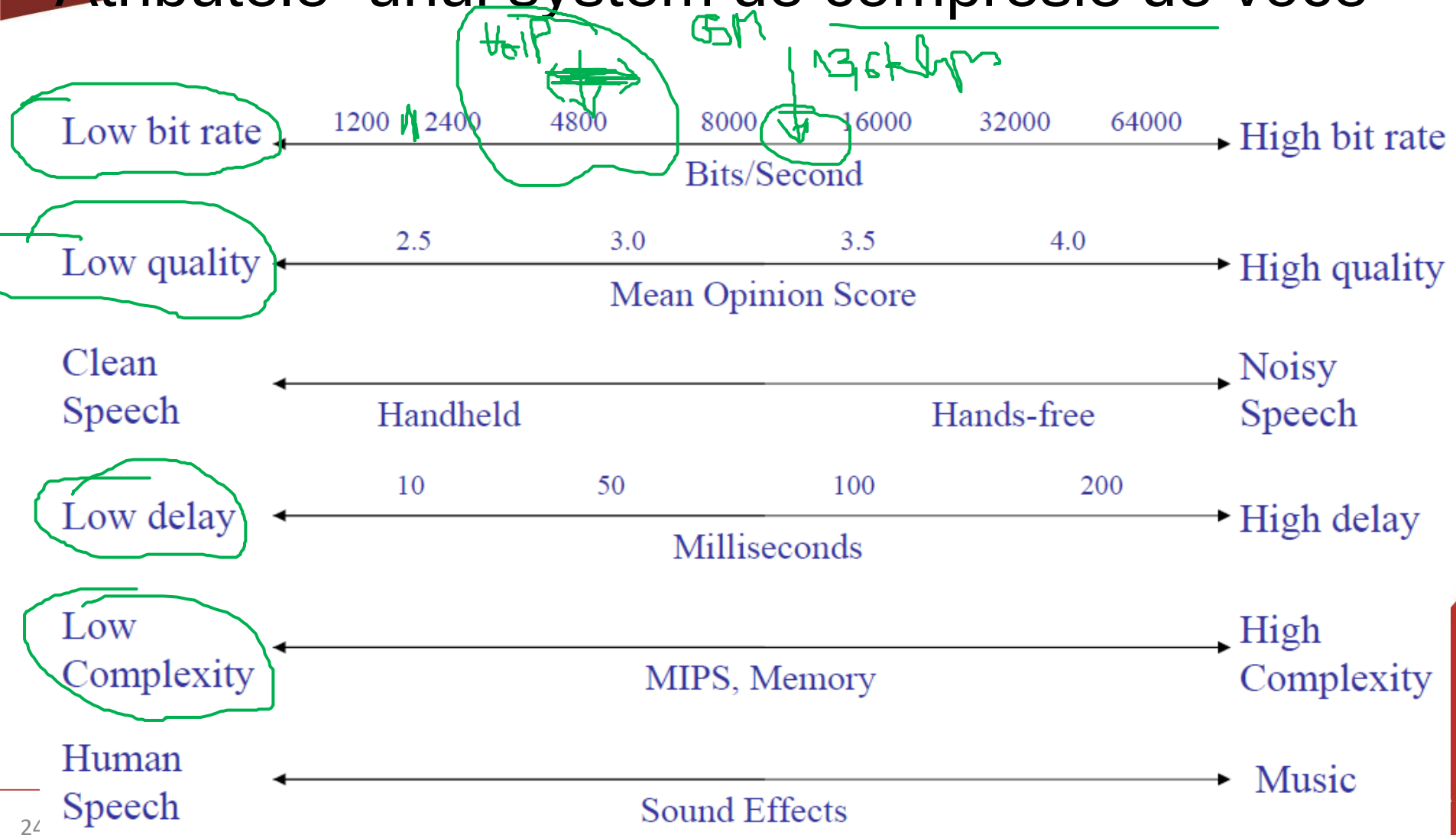
I.1.d Continut

- Sisteme de codare parametrica & codarea folosind analiza prin sinteza
  - LPC, MPE, RPE-LTP (GSM), CELP, ACELP



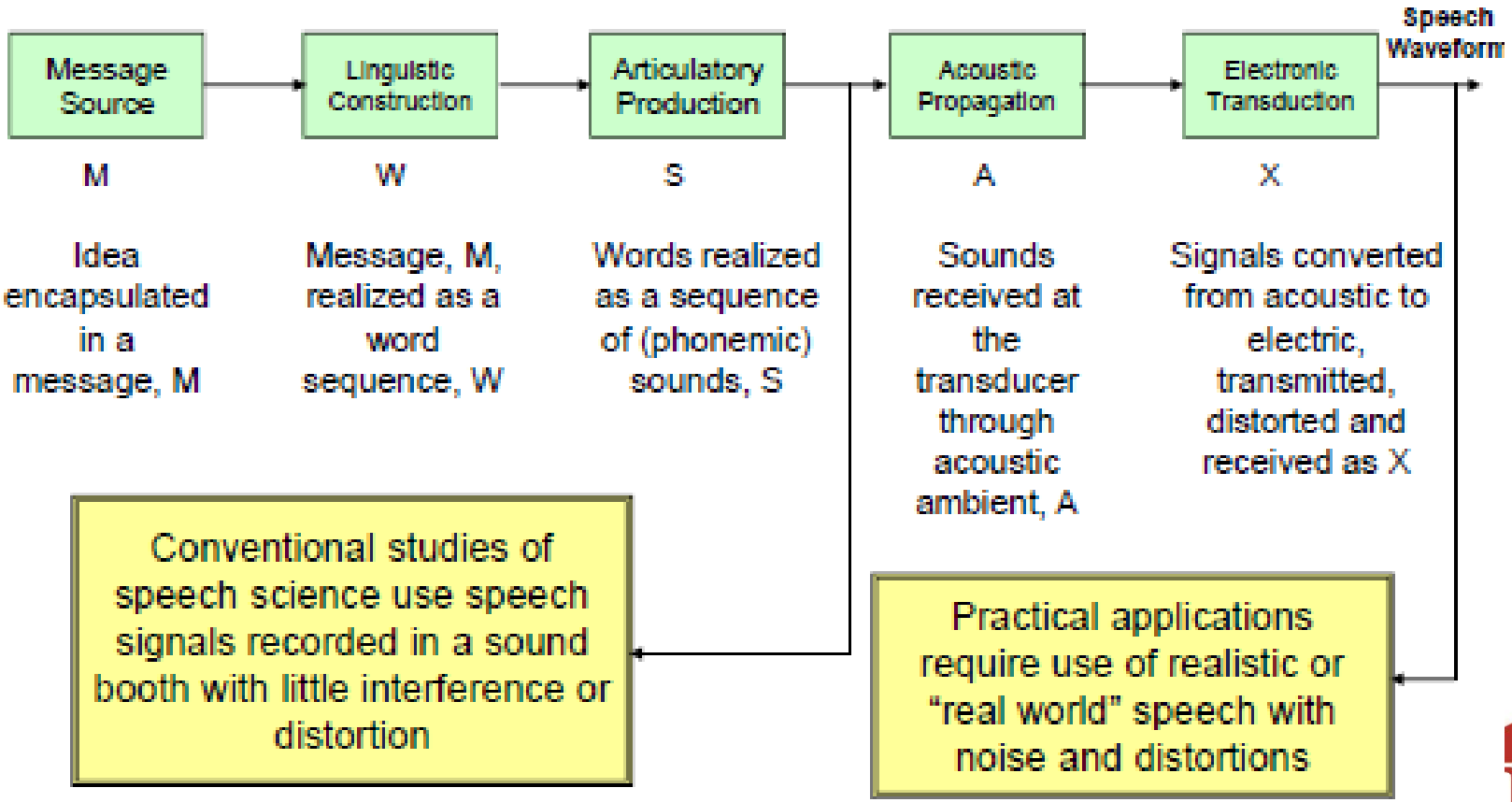
I.2. Aplicatii care folosesc Prelucrarea Semnalului Vocal

Atributele unui system de compresie de voce



I.4. Mecanismul producerii vorbirii

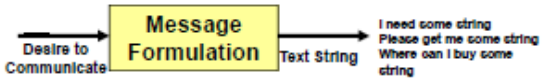
Speech Signal Production



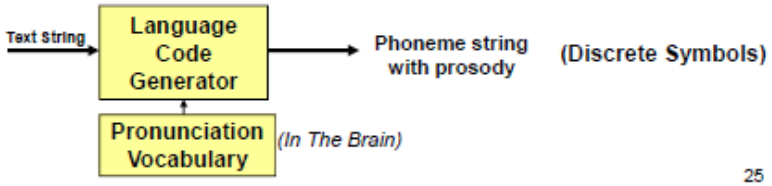
I.4. Mecanismul producerii vorbirii

Speech Production/Generation Model

- ***Message Formulation*** → desire to communicate an idea, a wish, a request, … => express the message as a sequence of words



- ***Language Code*** → need to convert chosen text string to a sequence of sounds in the language that can be understood by others; need to give some form of emphasis, prosody (tune, melody) to the spoken sounds so as to impart non-speech information such as sense of urgency, importance, psychological state of talker, environmental factors (noise, echo)

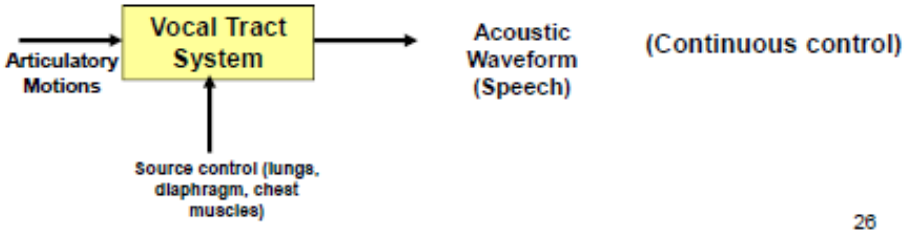


25

Speech Production/Generation Model

- ***Neuro-Muscular Controls*** → need to direct the neuro-muscular system to move the articulators (tongue, lips, teeth, jaws, velum) so as to produce the desired spoken message in the desired manner

- ***Vocal Tract System*** → need to shape the human vocal tract system and provide the appropriate sound sources to create an acoustic waveform (speech) that is understandable in the environment in which it is spoken

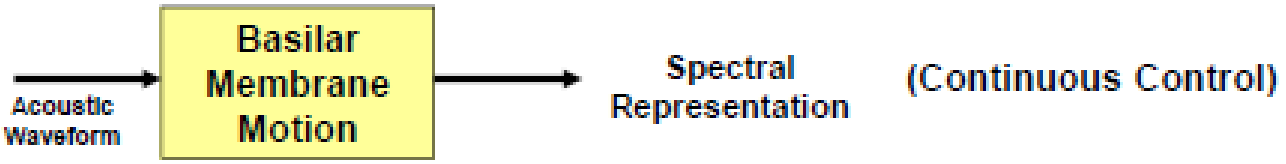


26

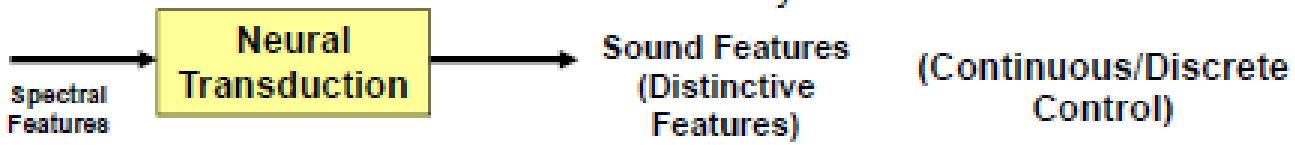
I.4. Mecanismul producerii vorbirii

Speech Perception Model

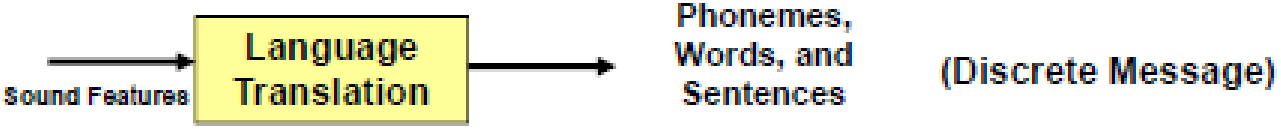
- The acoustic waveform impinges on the ear (the basilar membrane) and is spectrally analyzed by an equivalent filter bank of the ear



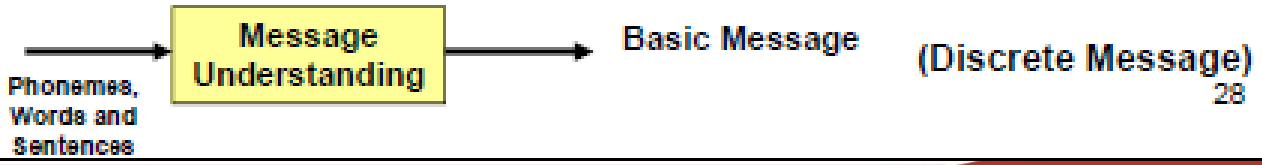
- The signal from the basilar membrane is neurally transduced and coded into features that can be decoded by the brain



- The brain decodes the feature stream into sounds, words and sentences

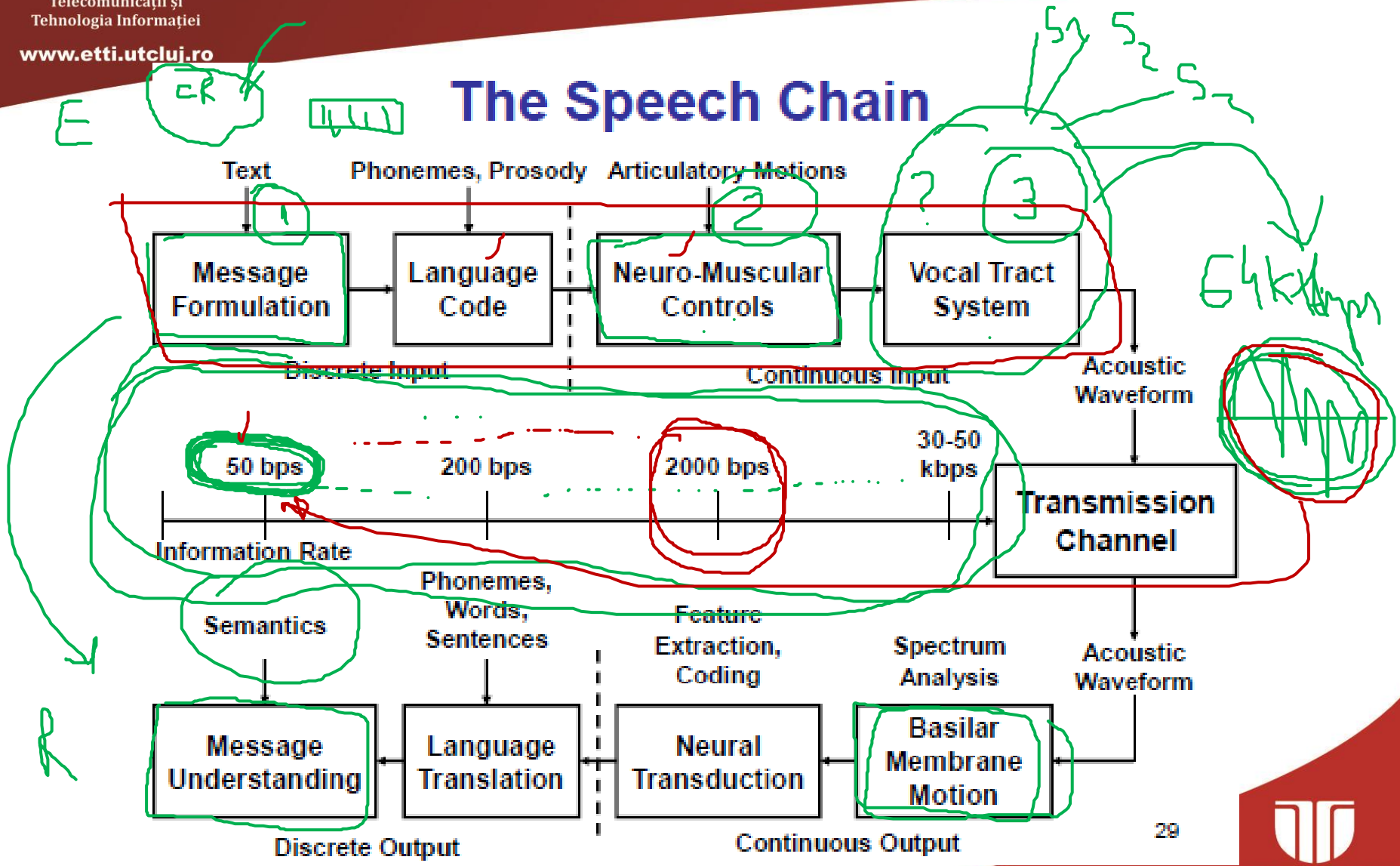


- The brain determines the meaning of the words via a message understanding mechanism



28

I.4. Mecanismul producerii vorbirii



I.1.d Continut

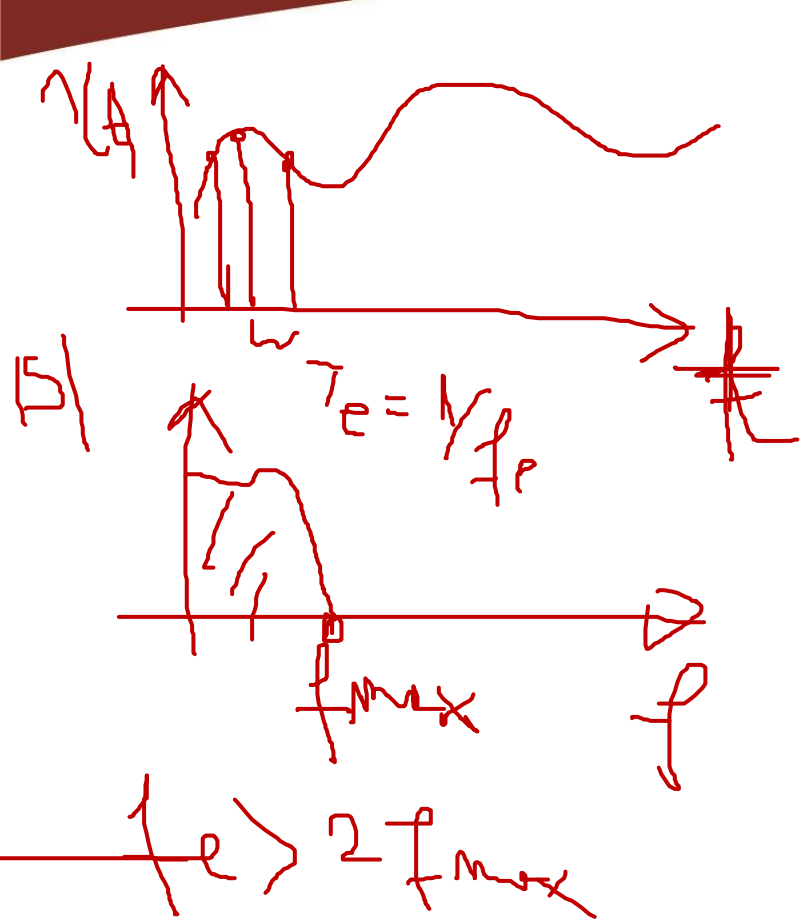
- Compresia prin cuantizare vectoriala
- Compresia prin Transformata Wavelet

- Sisteme comerciale (in documentare si prezentari student)
- Vocodere neuronale

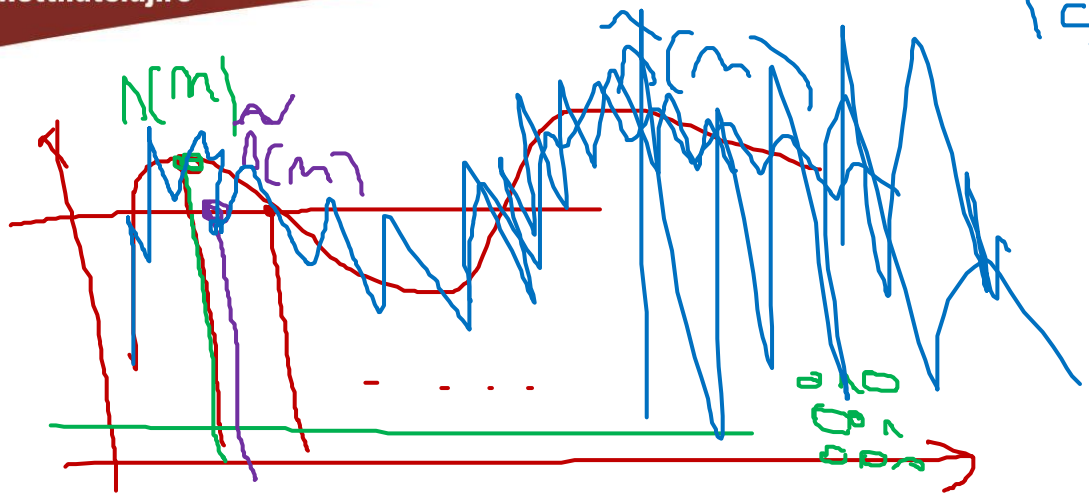
I.1.d Cunostinte preliminare (dezbatere, revizuire)

- Semnale, Teorema esantionarii
- Eroare de cuantizare
- Zgomot
- SNR
- Spectru, frecventa fundamentala
- Functie de transfer, caracteristica de transfer
- Filtru (FTJ, FTS, FTB), tipuri
- Informatie
- Entropie
- Modulatie (MA, MF)

I.1.d Cunostinte preliminare (dezbatere, revizuire)



I.1.d Cunostinte preliminare (dezbatere, revizuire)



$P_s = 10 P_z$ 10 dB

$= 1000 P_z$ 30 dB

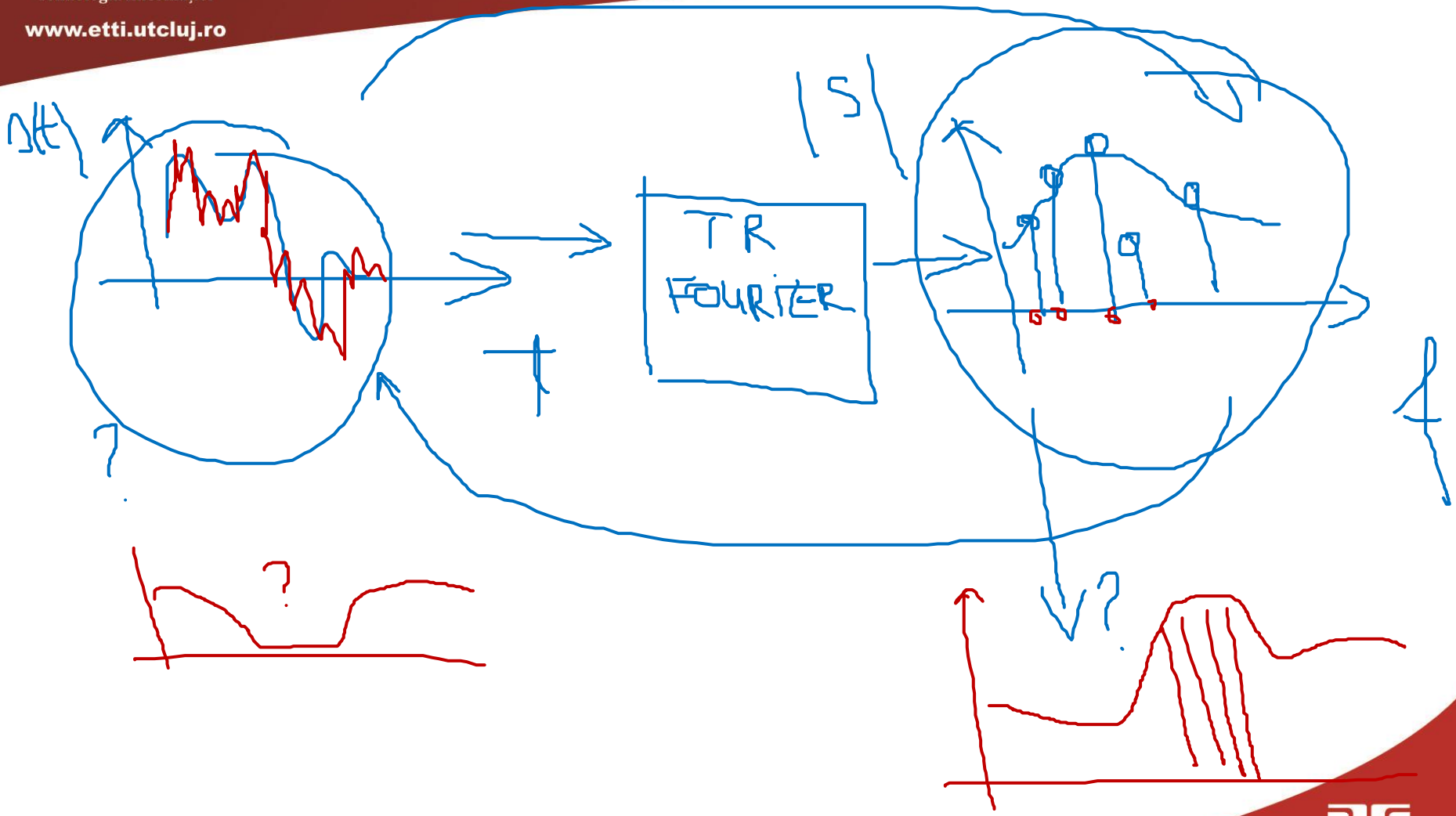
$e(n) = s(n) - \hat{s}(n)$

$z(n) = s(n) - \hat{s}(n)$

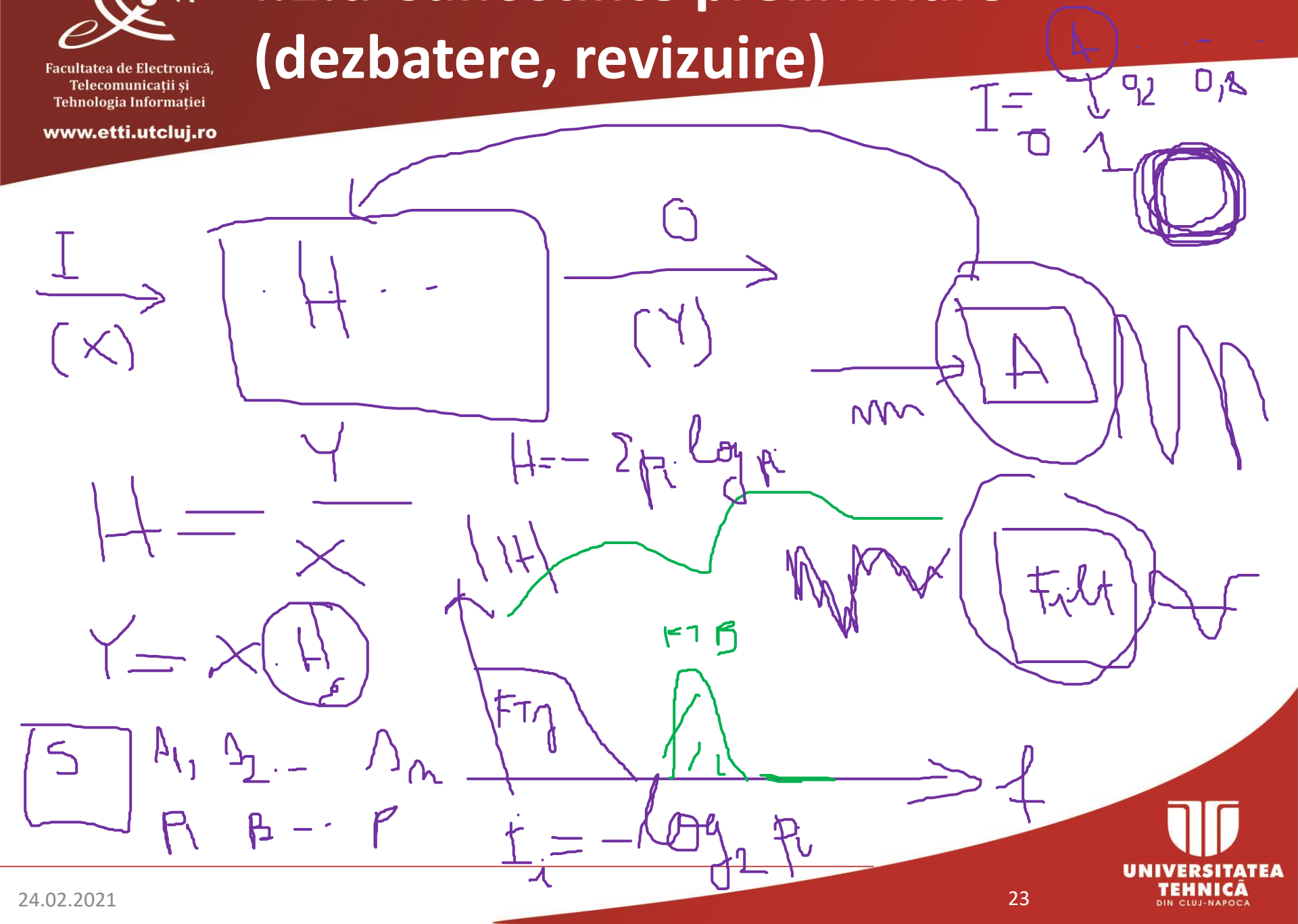
$SNR = 10 \log_{10} \frac{P_s}{P_z}$ [dB] $M \gg$ 30 dB

Signal to Noise Ration

I.1.d Cunostinte preliminare (dezbatere, revizuire)



I.1.d Cunostinte preliminare (dezbatere, revizuire)



SISTEME DE CODARE SI COMPRESIE IN DOMENIUL TIMP

- SCOP: **reducerea debitului binar** rezultat din sistemul de conversie analog/numerica, asociat cu **protectia la erorile** din canalul de comunicatii ➔ o codare combinata a sursei de semnal si a canalului de propagare.

- IMPORTANT: toate procesarile se efectueaza doar pe forma de unda a semnalului (in domeniul timp) ➔ tehnicile exploreaza doar caracteristici in domeniul timp (eg. corelatia intre esantioane, gama dinamica a semnalului, modul de distributie a valorii esantioanelor).

- EFECT: generarea unor debite standard (64Kbps, 32Kbps) la calitate buna a semnalului reconstituit;

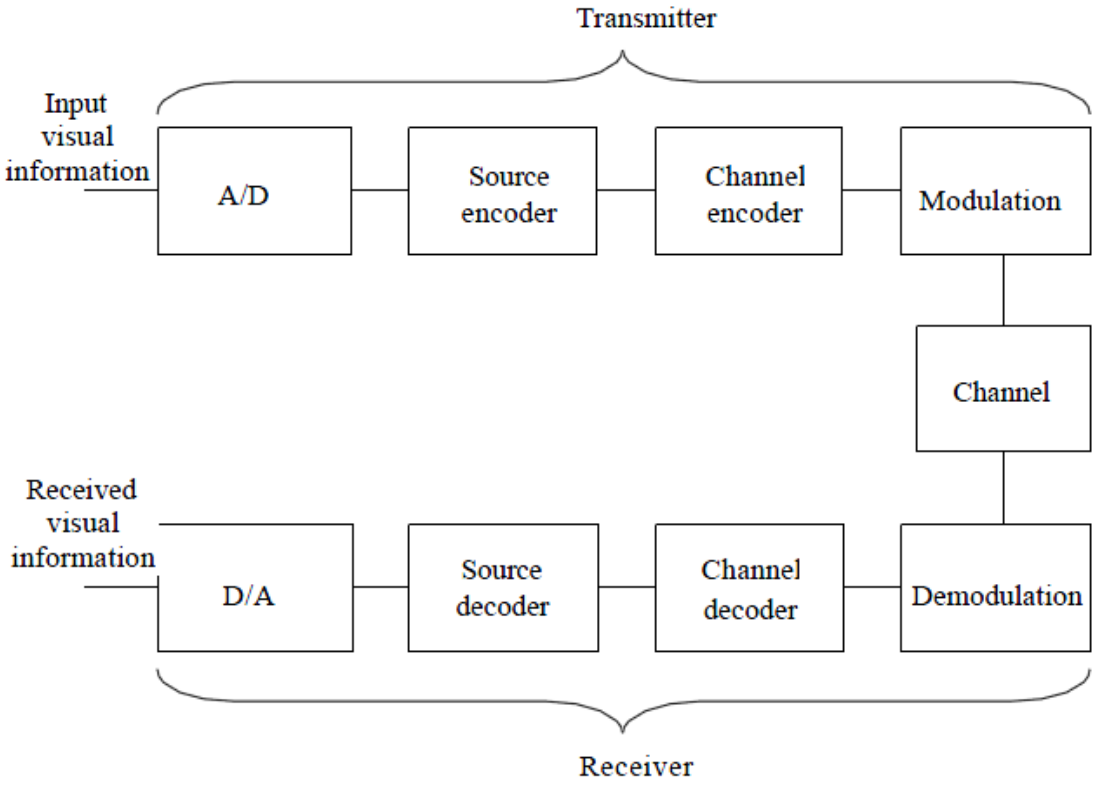
- APLICATII: telefonie digitala, VoIP, VideoConferinta.

Sisteme de codare in dom. timp - exemple

- PCM, WB-PCM, A-PCM
- DPCM
- ADPCM
- WideBand – ADPCM
- IMA-ADPCM
- Codec Delta

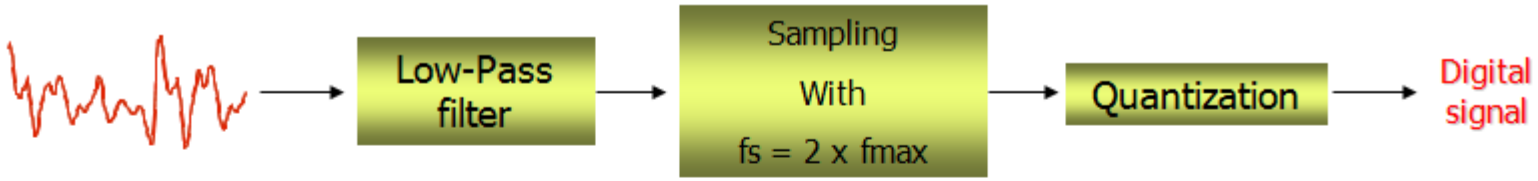
1) CODORUL PCM (Pulse Code Modulation) – schema de codare (G.721)

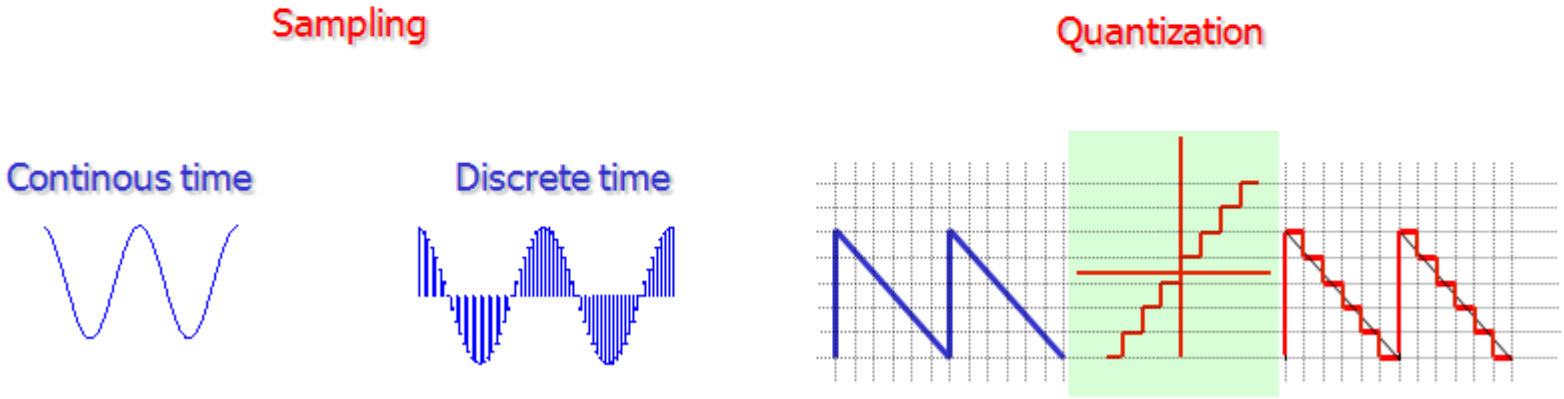
- Schema bloc a sistemului de codare (FTJ, CAN, Codor, …) + Rolul blocurilor



1) Esantionarea

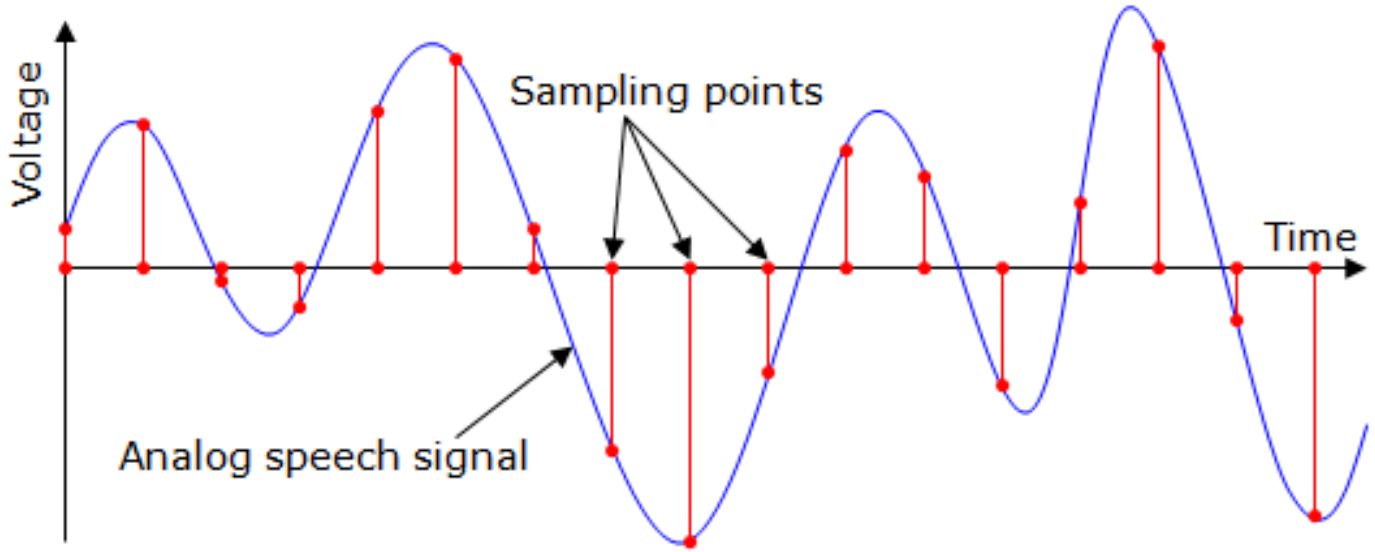
- Necesara in vederea conversiei analog numerice
- Respectarea teoremei esantionarii (Fe > 2 * Fmax)
- Semnale audio (De banda ingusta, 4 KHz, Fe=8 KHz; De banda larga, 8KHz, Fe = 16 KHz)
- Q: Cum influenteaza Fe debitul binar? (capacitatea de stocare)





1) Esantionarea

- Necesara in vederea conversiei analog numerice



1) Esantionare / frecventa Nyquist

Eşantionarea = prelevarea de probe dintr-un semnal la momente de timp decalate intre ele cu T_e – cu frecventa de eşantionare, $f_e=1/T_e$.

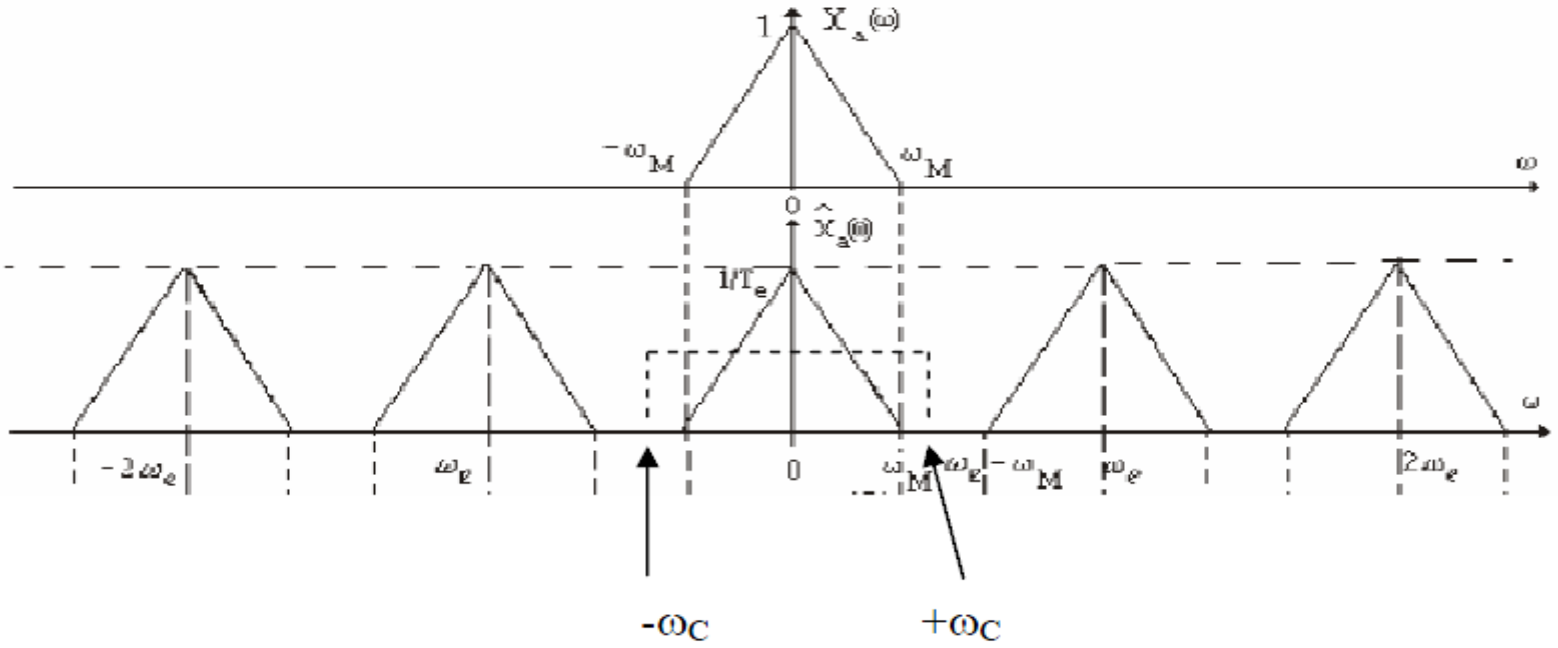
Semnalul eşantionat ideal: $\hat{x}(t)=\sum_{k=-\infty}^{\infty} x(kT_e)\delta(t-kT_e)$.

Spectrul $\hat{X}(\omega)=\frac{1}{T_e}\sum_{k=-\infty}^{\infty} X(\omega-k\omega_e)$

Teorema eşantionării: Un semnal x(t) de energie finita si banda limitata $B=\omega_M$ este unic determinat de mulţimea eşantioanelor sale $\{x(nT_e)|n\in Z\}$ daca $\omega_e \geq 2\omega_M$.

1) Esantionare – spectru semnal esantionat

- Spectru semnal initial
- Spectru semnal esantionat (repetat in jurul Fe)



1) Esantionare – refacere remnal

- Refacere semnal = din semnal esantionat in convolutie cu raspunsul la impuls a filtrului trece jos

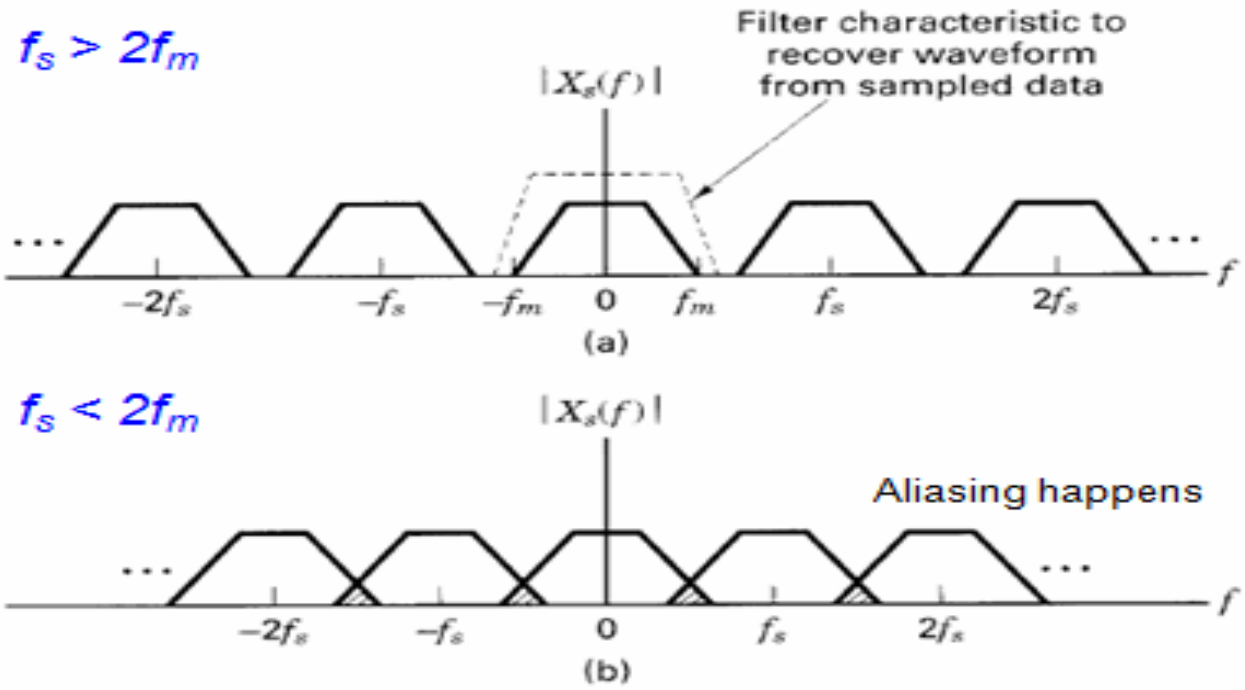
FTJ ideal: $H_r(\omega) = T_e p_{\omega_C}(\omega) \leftrightarrow h_r(t) = T_e \frac{\sin(\omega_C t)}{\pi t}$

Semnalul reconstituit: $x_r(t) = \hat{x}(t) * h_r(t) = \sum_{k=-\infty}^{\infty} \frac{2\omega_C}{\omega_e} x(kT_e) \frac{\sin(\omega_C(t-kT_e))}{\omega_C(t-kT_e)}$

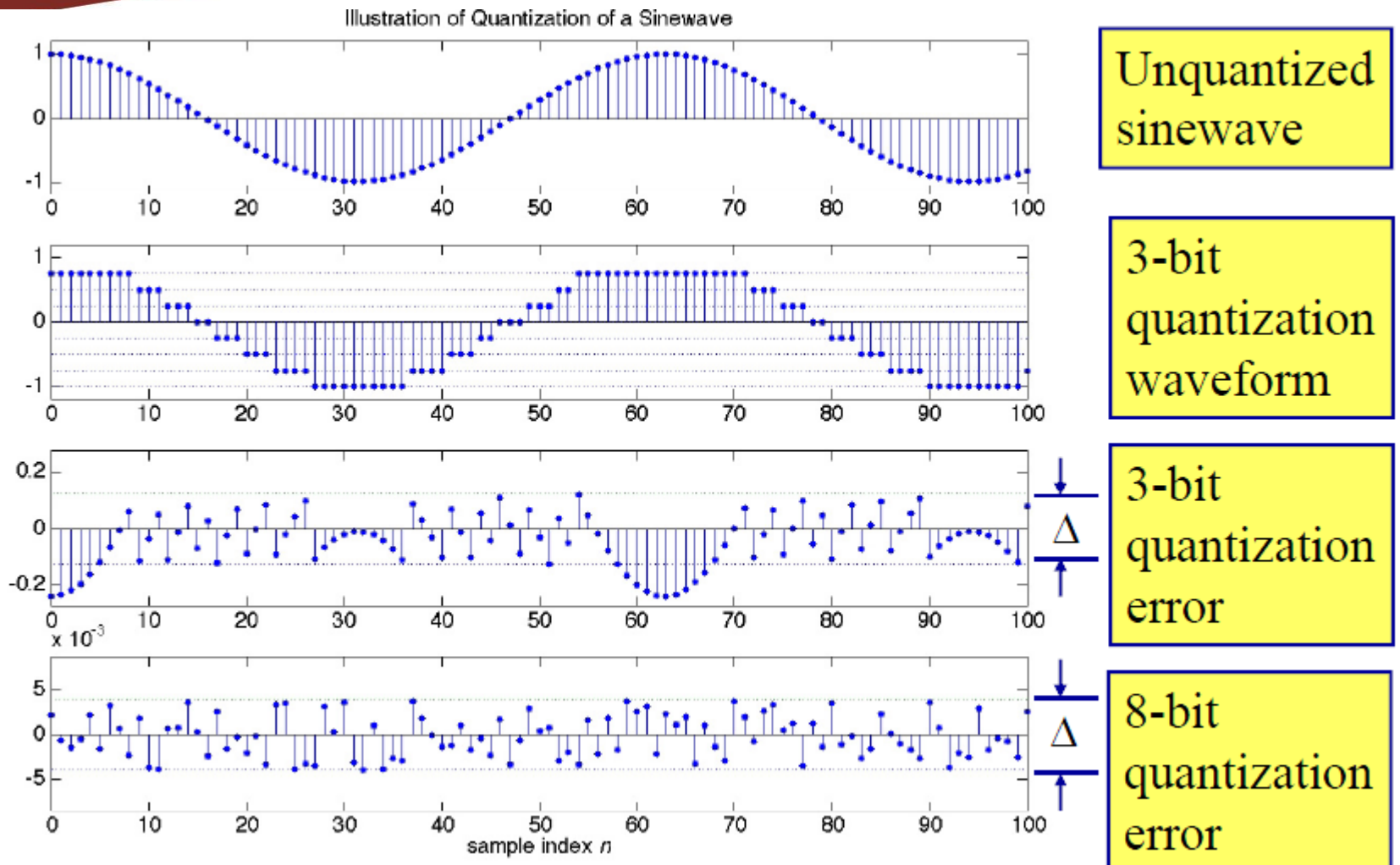
$\omega_C = \omega_M = \frac{\omega_e}{2}$ => Semnalul reconstituit: $x_r(t) = \sum_{k=-\infty}^{\infty} x(kT_e) \frac{\sin\left(\pi\left(\frac{t}{T_e}-k\right)\right)}{\pi\left(\frac{t}{T_e}-k\right)}$

2) Reconstrucție semnal - Efectul de aliere a spectrelor

- Respectarea TE = un semnal poate fi reconstituit din esantioanle sale daca Fe >> 2Fmax
- In caz contrar – aliere spectre

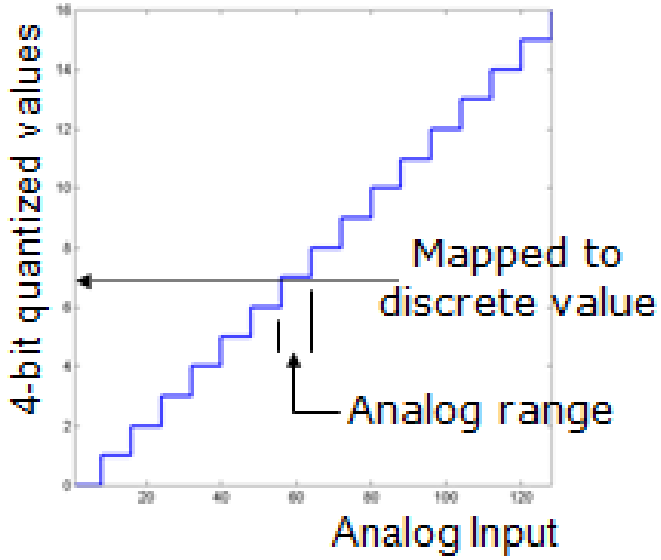


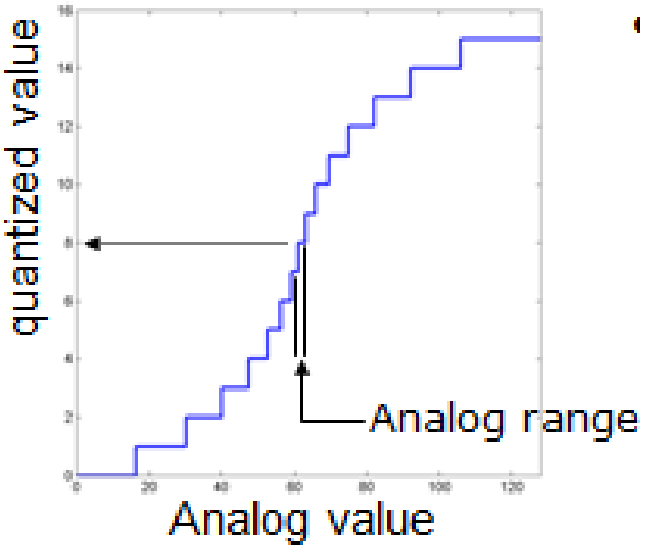
1) CODORUL PCM (Pulse Code Modulation) – determinarea erorilor de cuantizare

- 

3) Cuantizarea semnalului

- Liniara vs neliniara



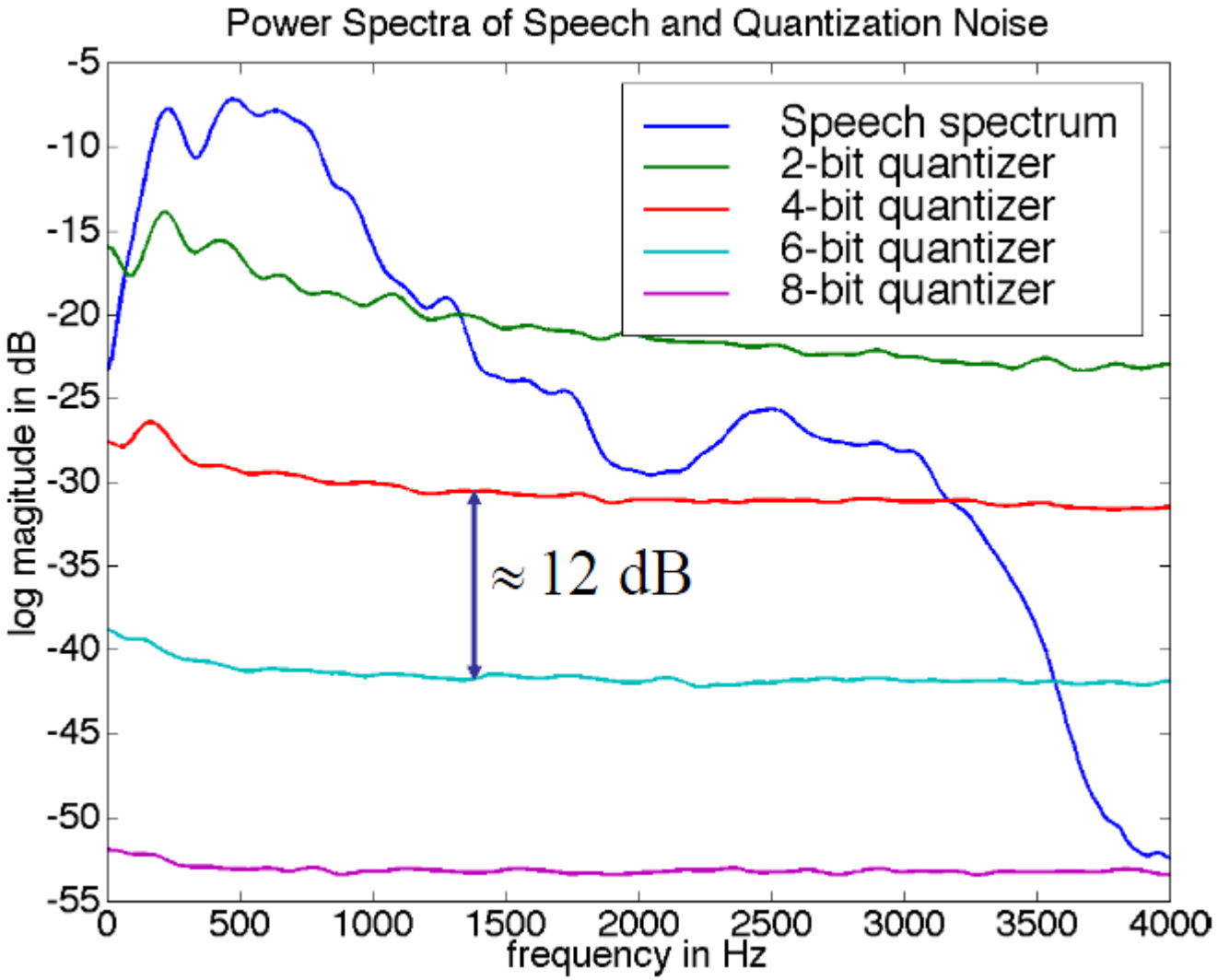


3) Cuantizarea semnalului

- Exercitiu: un semnal vocal cu Avv = 10V este esantionat cu Fe = 10KHz si 8 biti/esantion.
- A) Determinati cat este valoare cuantei cuantizorului?
- B) cat e zgomotul de cuantizare pentru un esantion care are amplitudinea de 10 mv? Dar pentru un esantion care are amplitudinea de 8V
- C) ce solutie propuneti pentru a minimiza erorile de cuantizare? (graphic ampl nel)

3) Cuantizarea semnalului – amplific neliniara, compandare

1) CODORUL PCM (Pulse Code Modulation) – spectru zgomot cuantizare



1) CODORUL PCM (Pulse Code Modulation) – determinarea erorilor de cuantizare

$$\hat{x}(n) = x(n) + e(n) \qquad -\frac{\Delta}{2} \le e(n) \le \frac{\Delta}{2}$$

$$\Delta 2^B = 2\,X_{max}$$

$$\Delta = 2\,X_{max} / 2^B$$

$$SNR = \frac{\sigma_x^2}{\sigma_e^2} = \frac{E\left(x^2(n)\right)}{E\left(e^2(n)\right)} = \frac{\sum_n x^2(n)}{\sum_n e^2(n)}$$

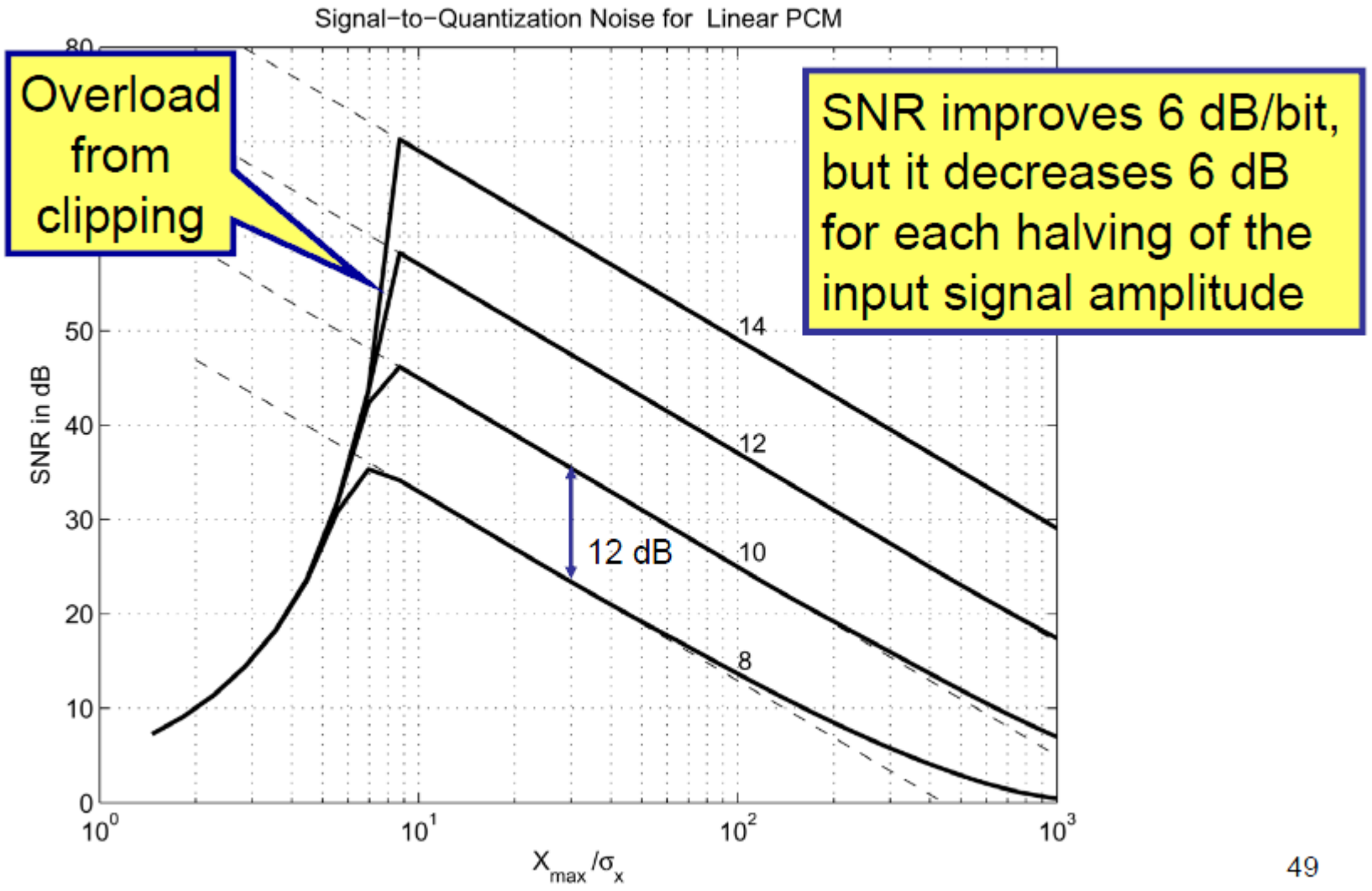
$$\Delta = \frac{2X_{\max}}{2^B} \text{ (uniform quantizer step size)}$$

$$SNR = \frac{(3)2^{2B}}{\left[\frac{X_{\max}}{\sigma_x}\right]^2};\ SNR(dB) = 10\log_{10}\left[\frac{\sigma_x^2}{\sigma_e^2}\right] = 6B + 4.77 - 20\log_{10}\left[\frac{X_{\max}}{\sigma_x}\right]$$

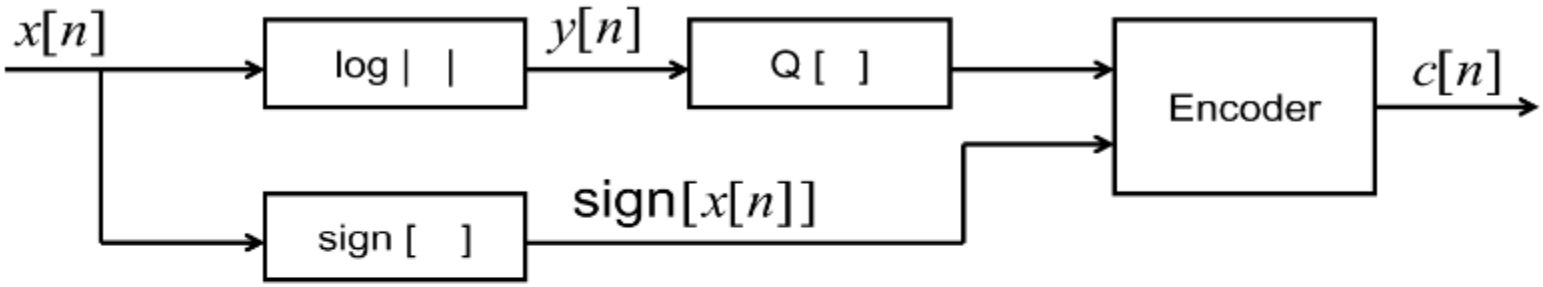
1) CODORUL PCM (Pulse Code Modulation) – determinarea erorilor de cuantizare

Determinati nr de biti la codare pentru a avea un SNR > 35 dB

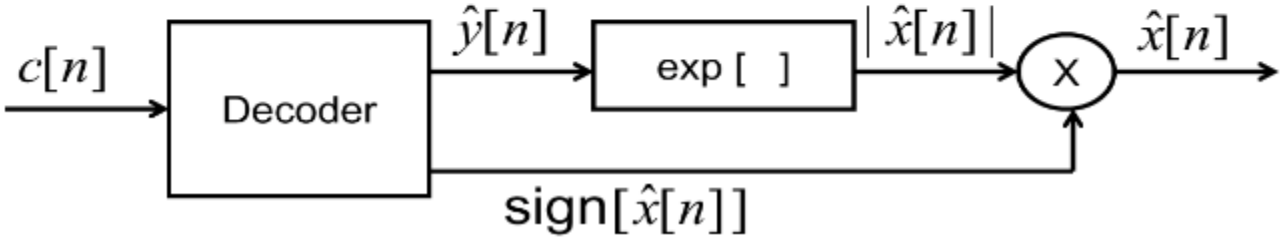
1) CODORUL PCM (Pulse Code Modulation) – SNR variabil in functie de nivelul



1) CODORUL PCM (Pulse Code Modulation) – CUANTIZARE NEUNIFORMA



(a)



1) CODORUL PCM (Pulse Code Modulation) – Legea A / u

$$F(x) = \operatorname{sgn}(x) \begin{cases} \frac{A|x|}{1+\log(A)}, & |x| < \frac{1}{A} \\ \frac{1+\log(A|x|)}{1+\log(A)}, & \frac{1}{A} \le |x| \le 1, \end{cases}$$

where *A* is the compression parameter. In Europe, $A = 87.6$.

A-law expansion is given by the inverse function,

$$F^{-1}(y) = \operatorname{sgn}(y) \begin{cases} \frac{|y|(1+\ln(A))}{A}, & |y| < \frac{1}{1+\ln(A)} \\ \frac{\exp(|y|(1+\ln(A))-1)}{A}, & \frac{1}{1+\ln(A)} \le |y| < 1. \end{cases}$$

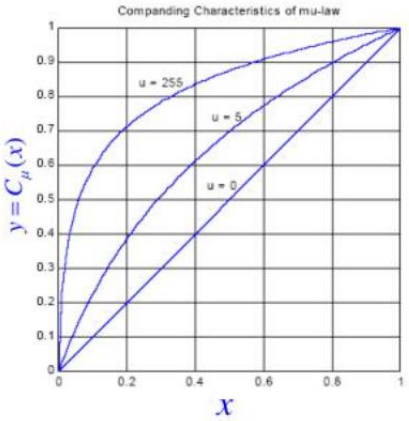
$$F(x) = \operatorname{sgn}(x) \frac{\ln(1+\mu|x|)}{\ln(1+\mu)} \qquad -1 \le x \le 1$$

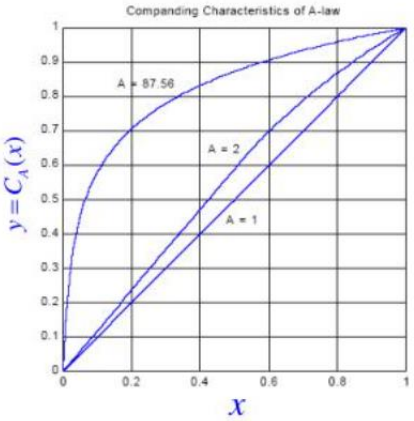
where $\mu = 255$ (8 bits) in the North American and Japanese standards. It is

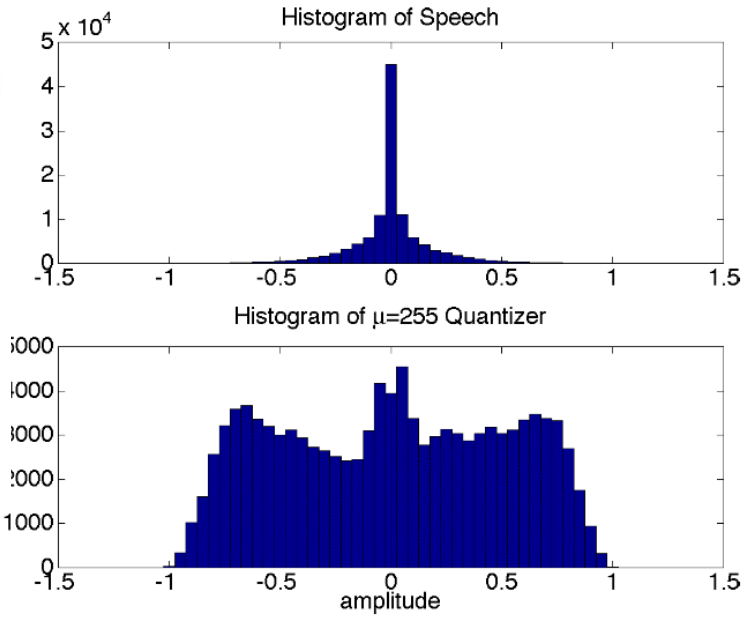
μ-law expansion is then given by the inverse equation:

$$F^{-1}(y) = \operatorname{sgn}(y)(1/\mu)((1+\mu)^{|y|} - 1) \qquad -1 \le y \le 1$$

1) CODORUL PCM (Pulse Code Modulation) – efectul legilor de compresie





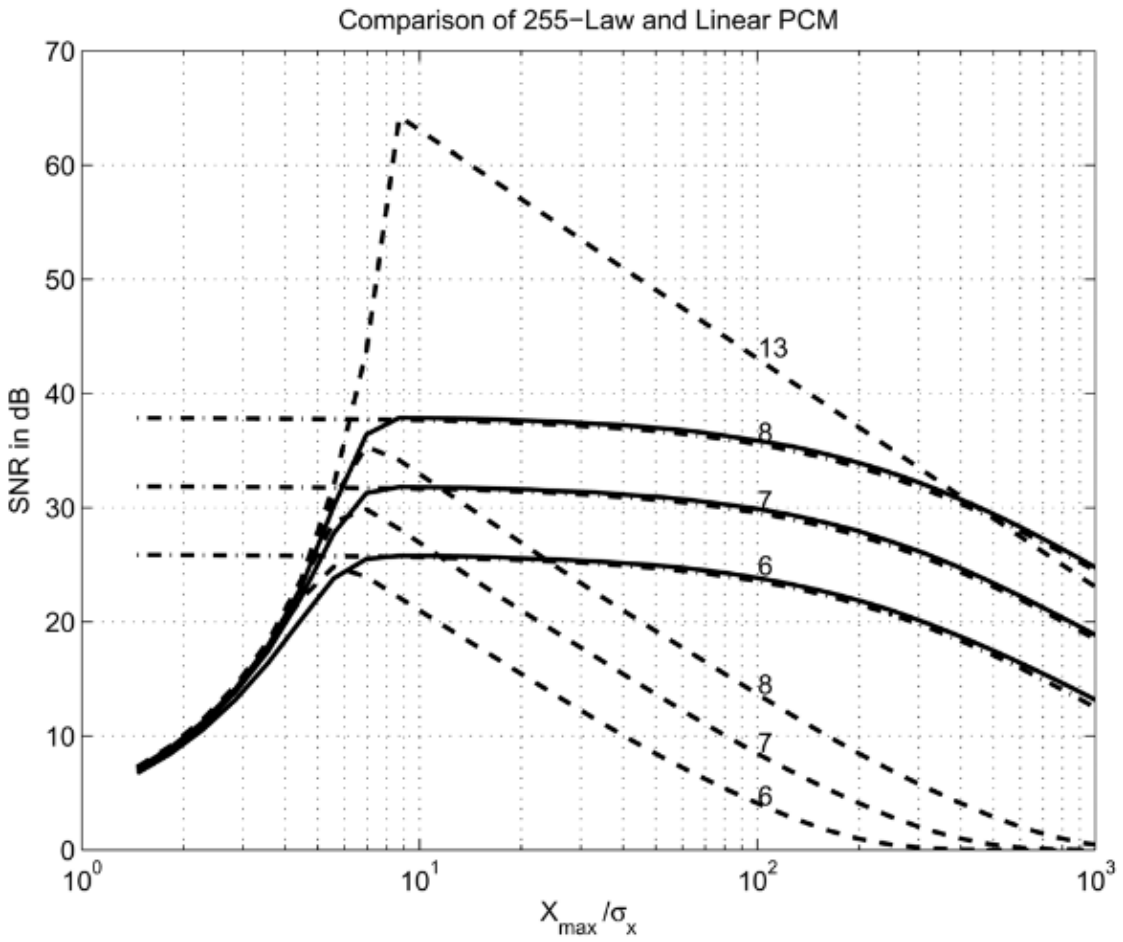


Speech waveform

Output of μ-Law compander

59

1) CODORUL PCM (Pulse Code Modulation) – efectul legilor de compresie



Dashed line – linear (uniform) quantizers with 6, 7, 8 and 13 bit quantizers

Solid line – μ-law quantizers with 6, 7 and 8 bit quantizers (μ=255)

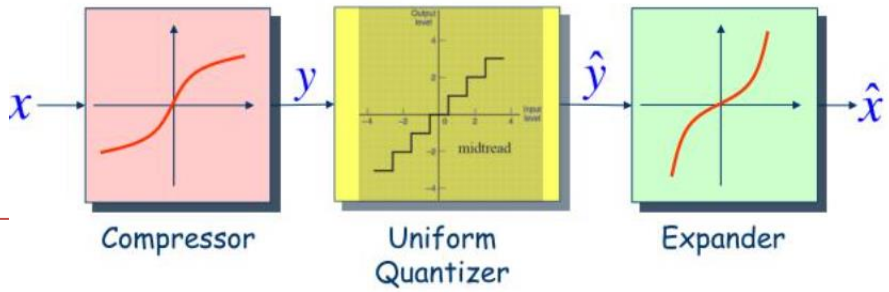
65

1) CODORUL PCM (Pulse Code Modulation) – efectul legilor de compresie

Companding

- The dynamic range of signals is compressed before transmission and is expanded to the original value at the receiver.
- Allowing signals with a large dynamic range to be transmitted over facilities that have a smaller dynamic range capability.
- Companding reduces the noise and crosstalk levels at the receiver.

Companding



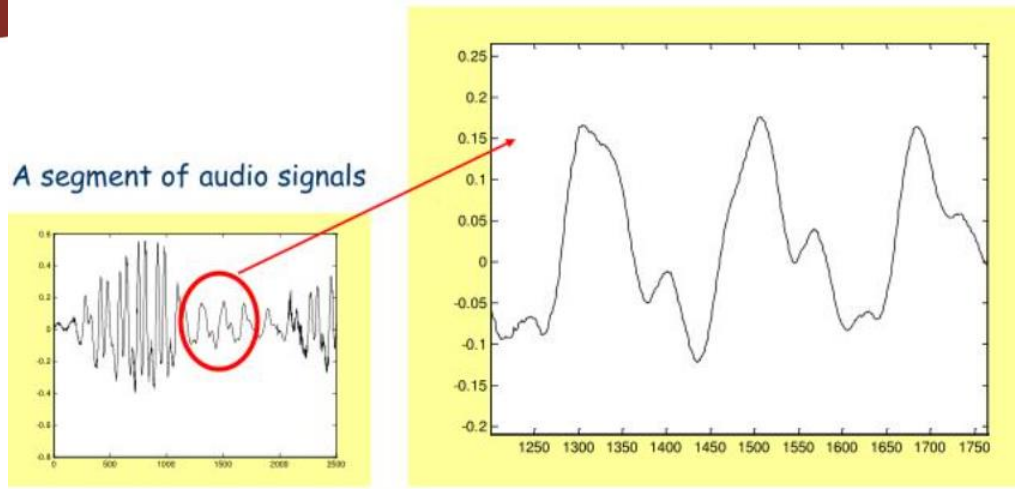
2) CODAREA PCM DE BANDA LARGA

- Vezi document aditional PDF cu WB-PCM….

2) CODAREA PCM DIFERENTIALA (DPCM)

- Principiul codarii – codarea diferentei
- Calculul cresterii SNR pentru acelasi numar de nivele de cuantizare
- Schema de codare / decodare
- Calculul cumularii zgomotului de cuantizare
- Modificarea schemei de codare DPCM
- … utilizarea unui predictor de ordin p

2) CODAREA PCM DIFERENTIALA (DPCM)



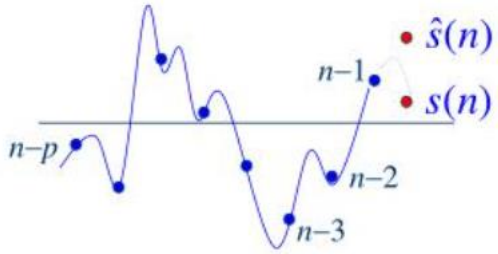
- Adjacent samples exhibit a high degree of correlation.
- Removing this adjacent redundancy before encoding, a more efficient coded signal can be resulted.
- How?
  - Accompanying with prediction (e.g., linear prediction)
  - Encoding prediction error only

2) CODAREA PCM DIFERENTIALA (DPCM)

$$\hat{s}(n) = \sum_{k=1}^{p} a_k s(n-k)$$

$$e(n) = s(n) - \hat{s}(n)$$

$$= s(n) - \sum_{k=1}^{p} a_k s(n-k)$$



$$\mathcal{E}_p = \sum_{n=1}^{N} e^2(n)$$

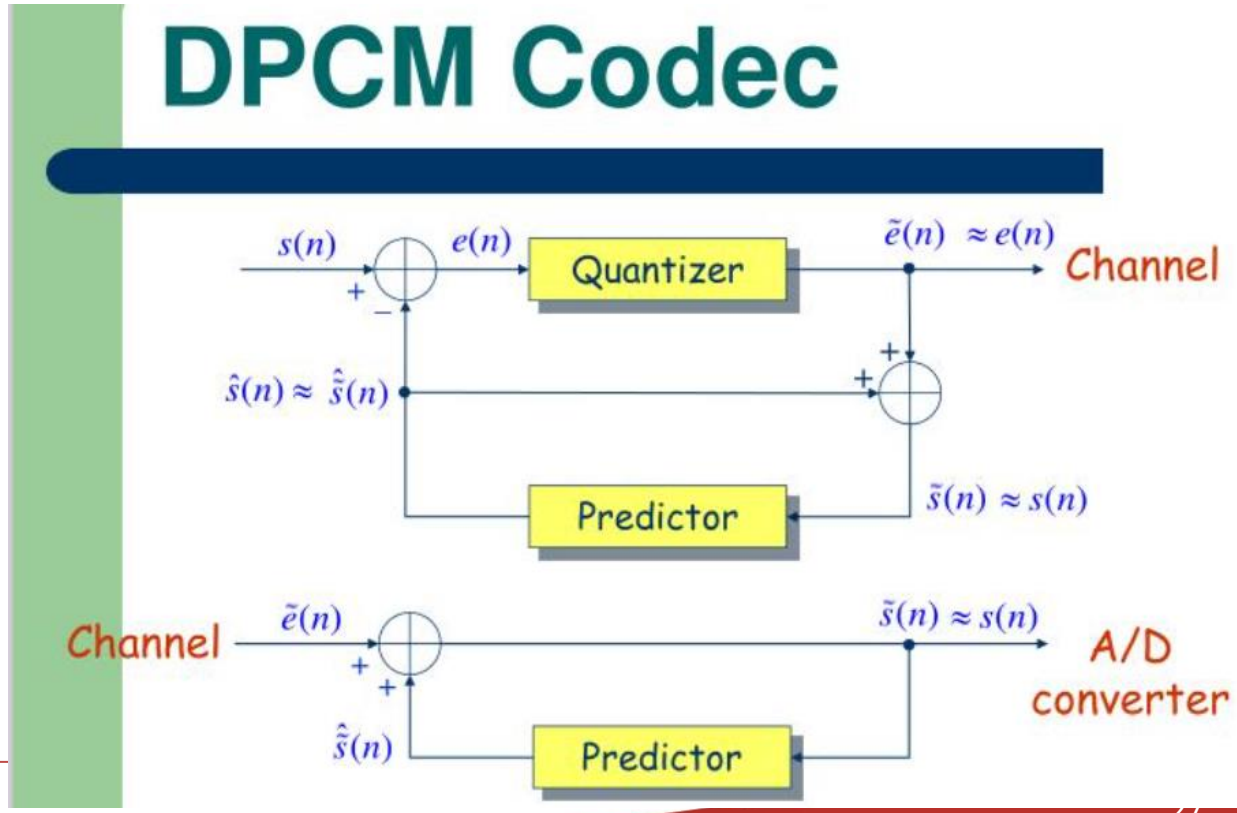
$$\mathbf{a}^* = \arg\min_{\mathbf{a}} \mathcal{E}_p$$

$$\mathbf{a} = (a_1, \ldots, a_p)'$$

$$\hat{s}(n) = \sum_{k=1}^{p} a_k s(n-k)$$

2) CODAREA PCM DIFERENTIALA (DPCM)

Codarea de 4 biti a semnalului diferenta e(n)
➔ debit de 32 Kbps

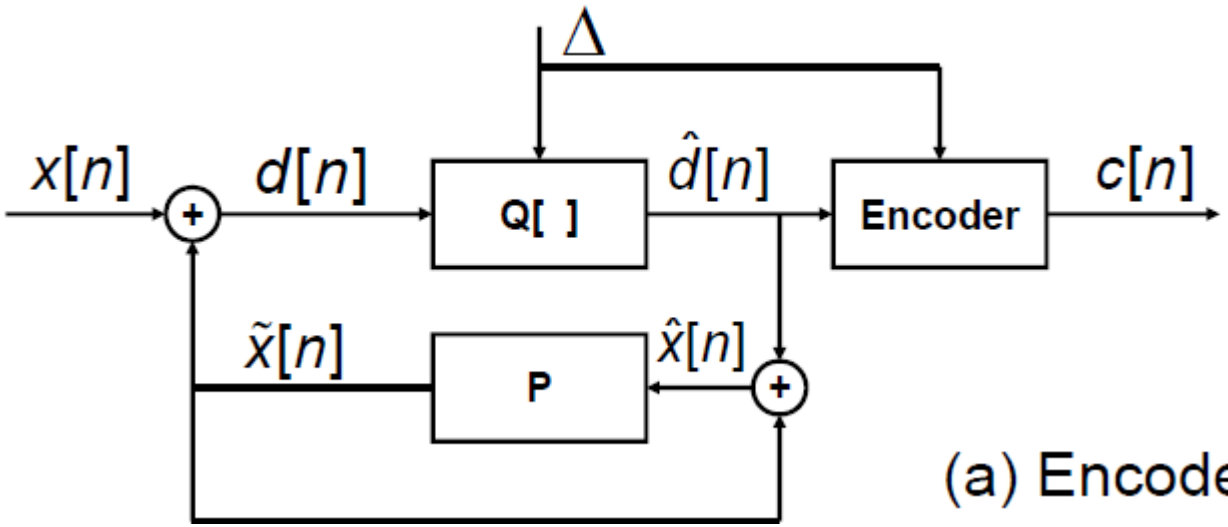


2) CODAREA PCM DIFERENTIALA (DPCM)

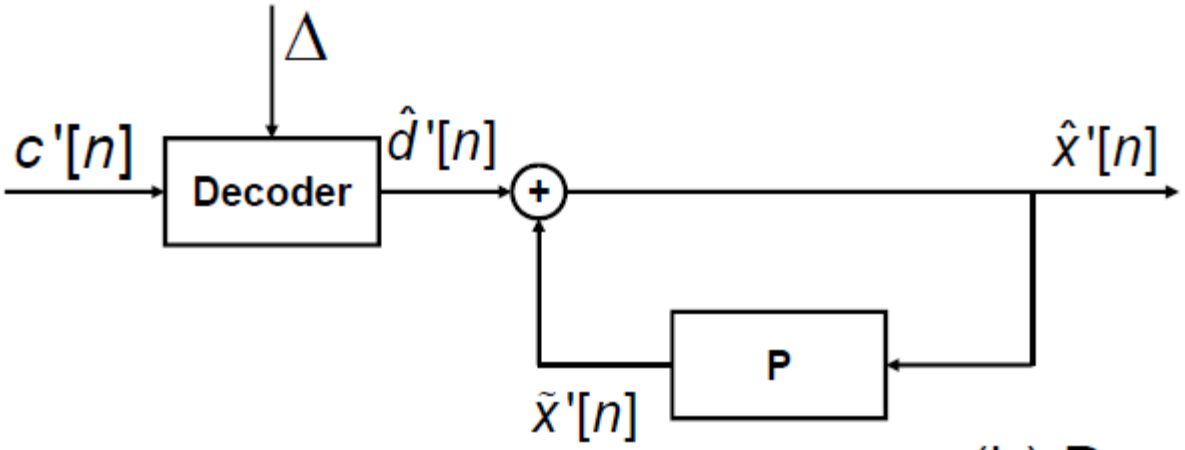
- Exercitiu:
- Cum arata schema de codare pentru un predictor realizat cu un element de intarziere?

2) CODAREA PCM DIFERENTIALA (DPCM)

2) DPCM cu predictie



(a) Encoder

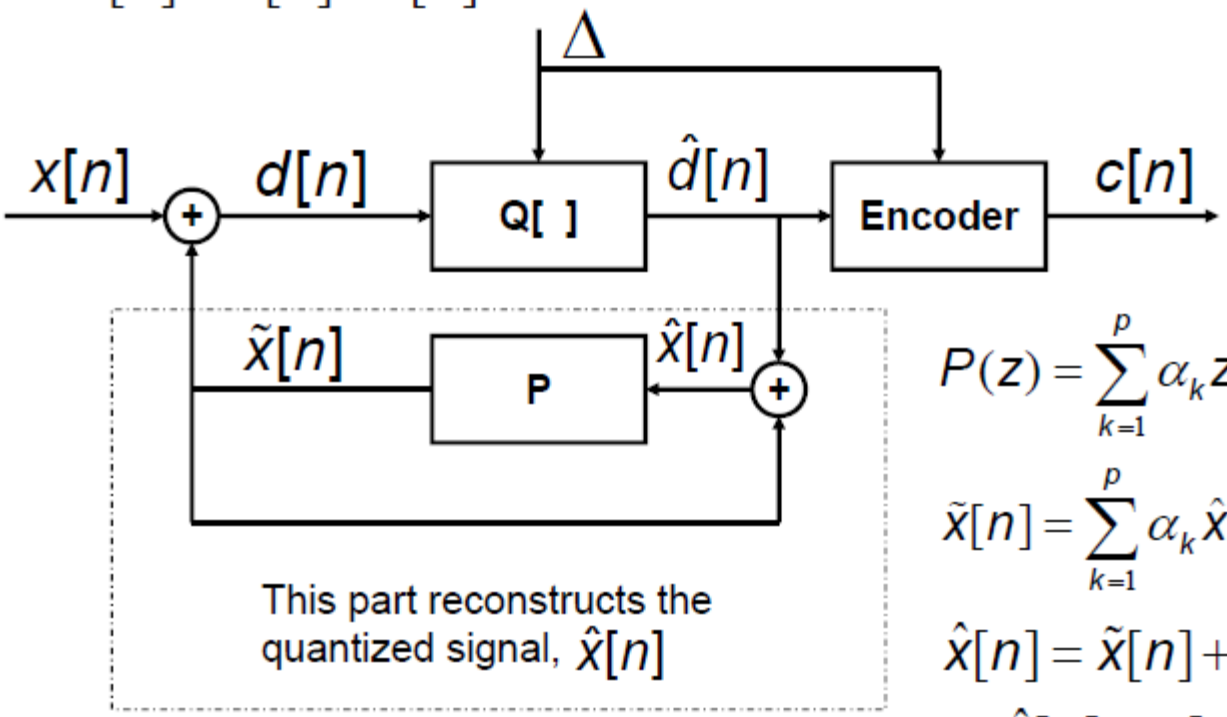


(b) Decoder

2) DPCM predictiv - codorul

$$d[n] = x[n] - \tilde{x}[n]$$

$$\hat{d}[n] = d[n] + e[n]$$



$$P(z) = \sum_{k=1}^{p} \alpha_k z^{-k}$$

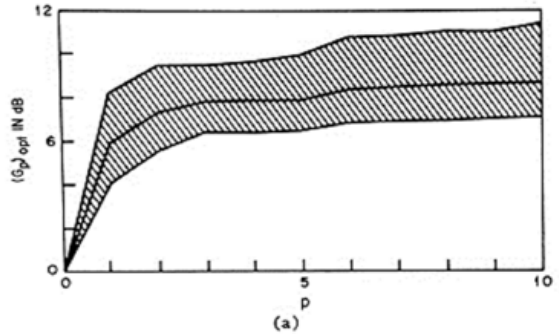
$$\tilde{x}[n] = \sum_{k=1}^{p} \alpha_k \hat{x}[n-k]$$

$$\hat{x}[n] = \tilde{x}[n] + \hat{d}[n]$$

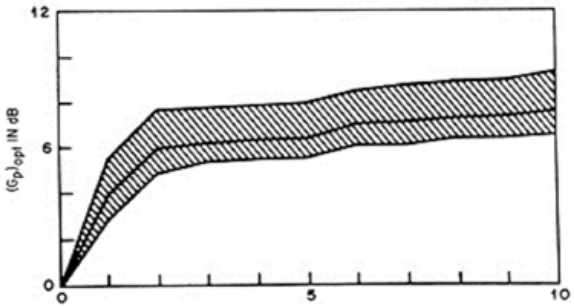
$$\Rightarrow \hat{x}[n] = x[n] + e[n]$$

3) DPCM predictiv - codorul

- Determinarea coeficientilor de predictie
- Cazuri:
  - P = 1
  - P = 2

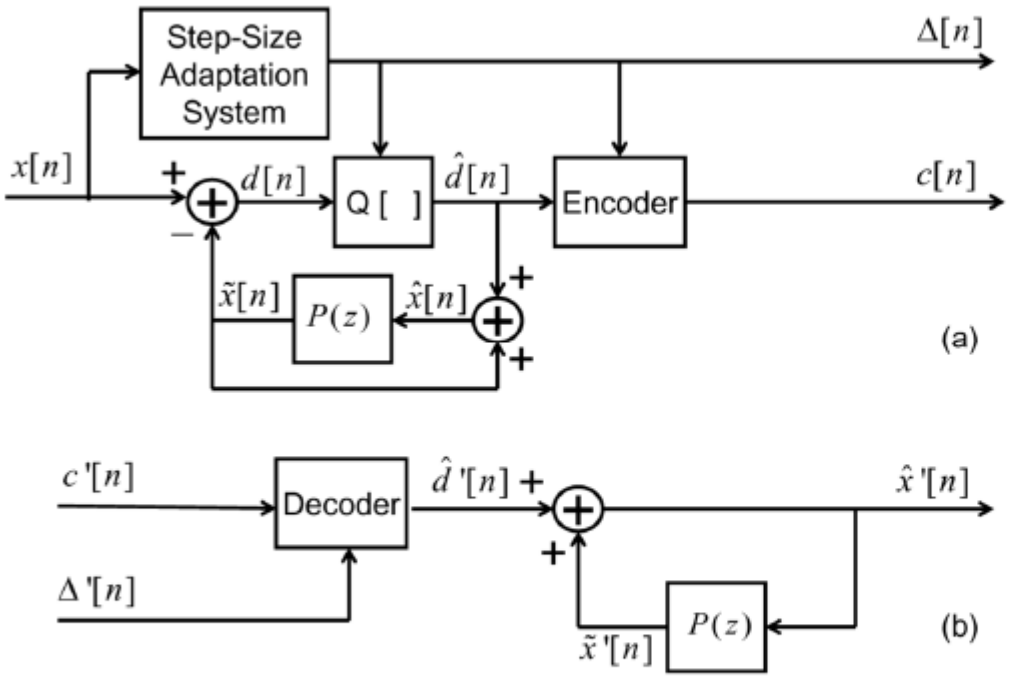


- variation in gain across 4 speakers
- can get about 6 dB improvement in *SNR* => 1 extra bit equivalent in quantization—but at a price of increased complexity in quantization



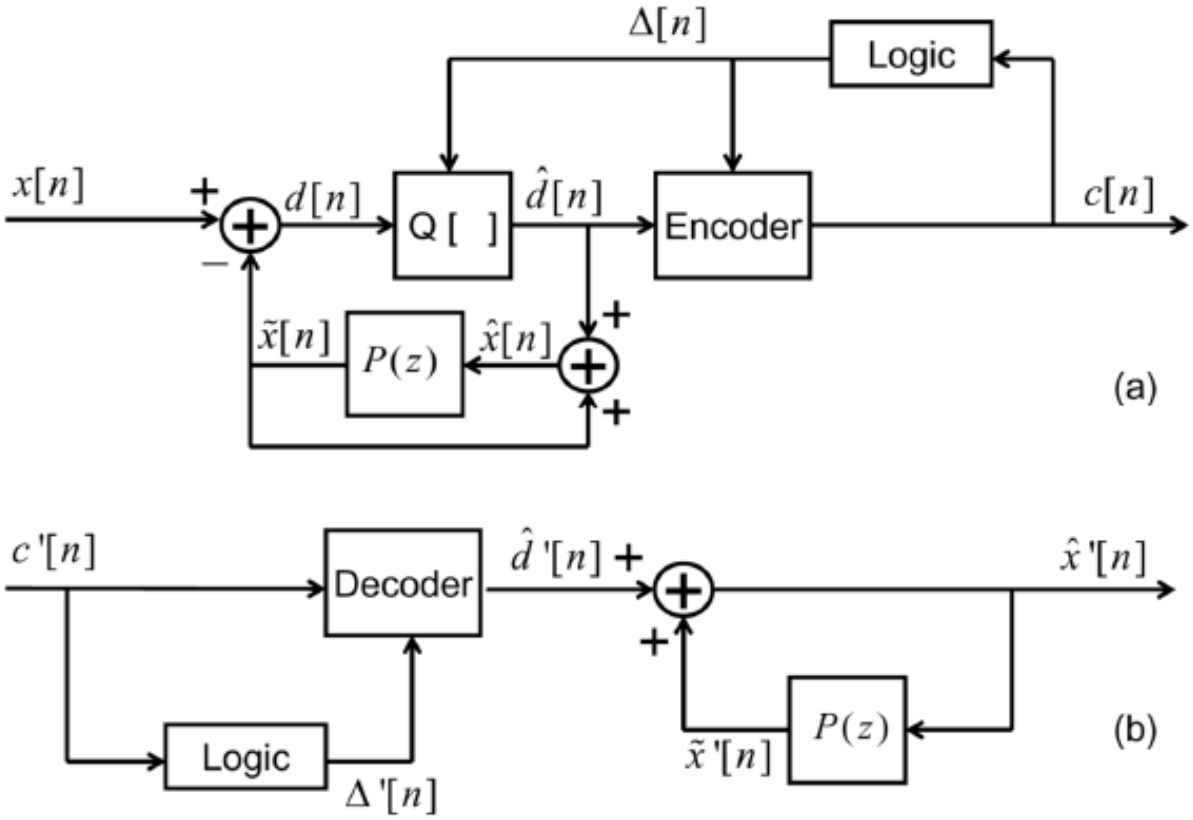
- differential quantization works!!
- gain in *SNR* depends on signal correlations
- fixed predictor cannot be optimum for all speakers and for all speech material

CODAREA ADPCM (G.721)



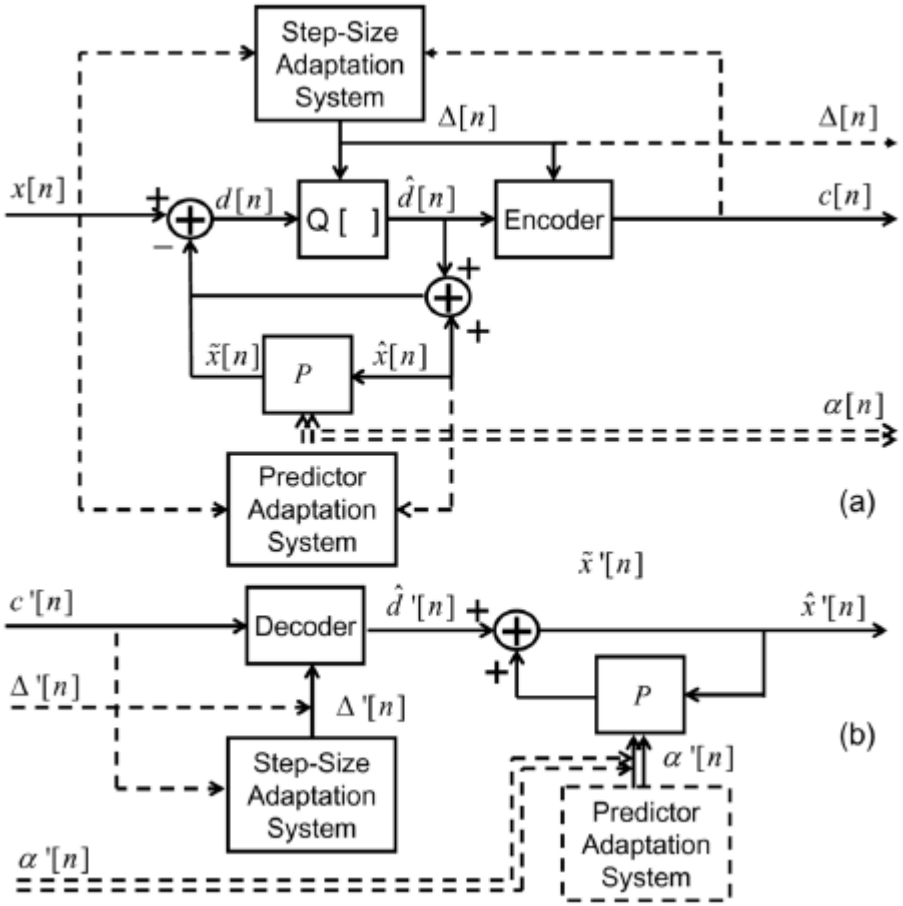
- quantizer step size proportional to variance at quantizer input
- can use *d*[*n*] or *x*[*n*] to control step size
- get 5 dB improvement in *SNR* over *μ*-law non-adaptive PCM
- get 6 dB improvement in *SNR* using differential configuration with fixed prediction => ADPCM is about 10-11 dB *SNR* better than from a fixed quantizer

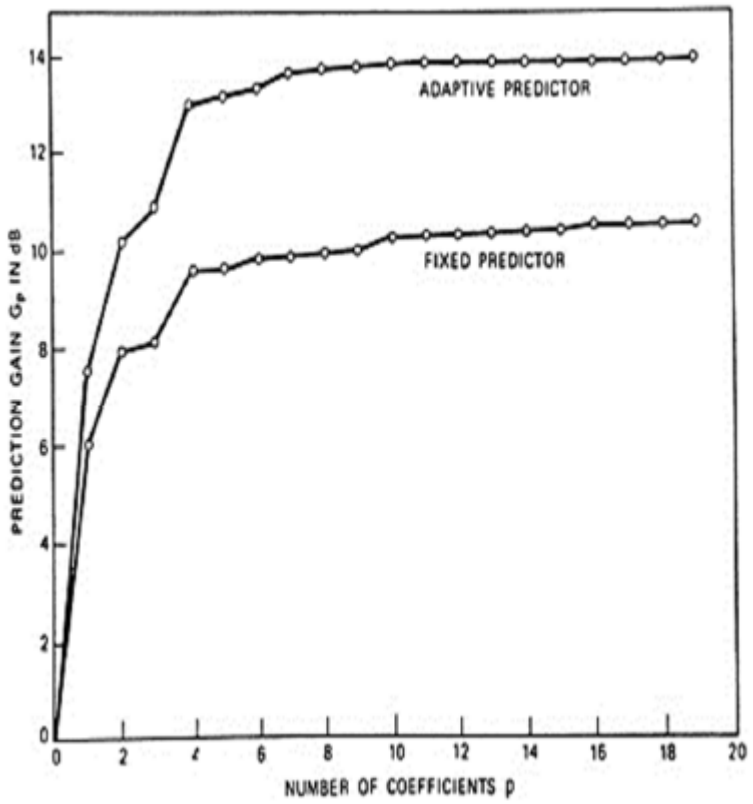
3) ADPCM – cu feedback



can achieve same improvement in *SNR* as feed forward system

3) CODAREA ADPCM – predictie adaptiva

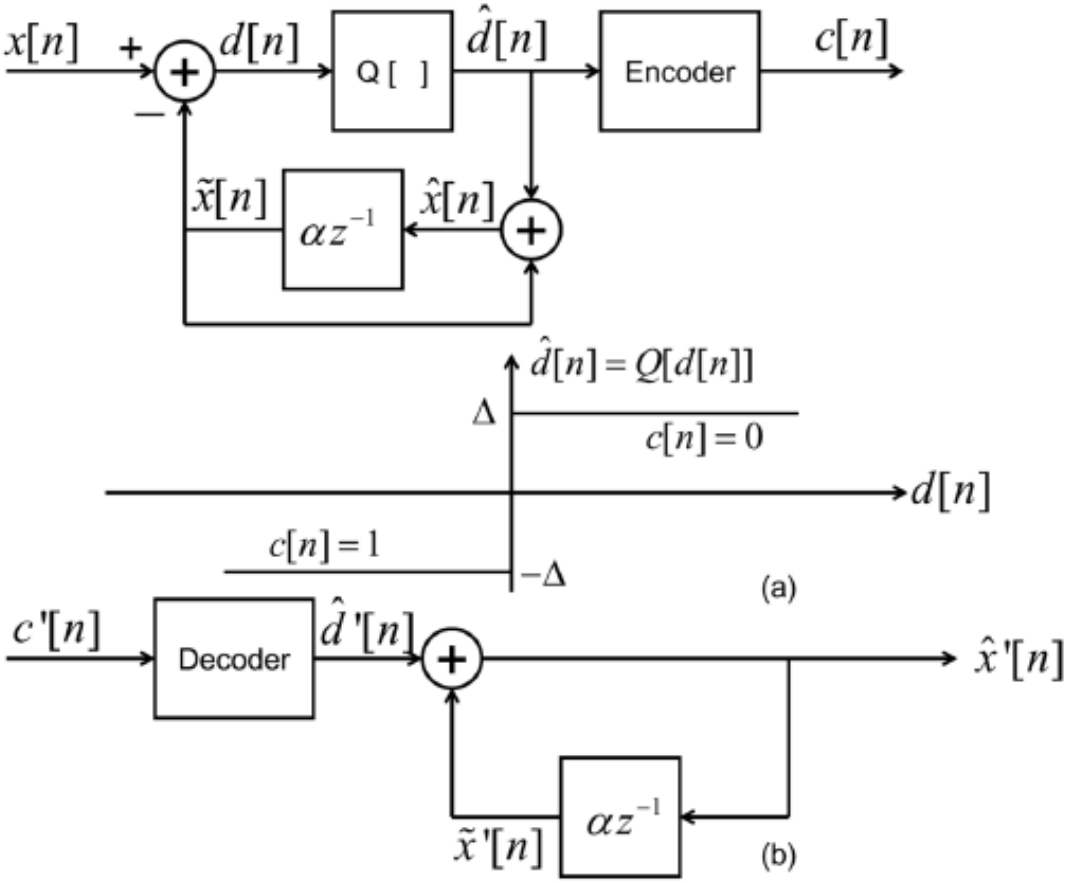




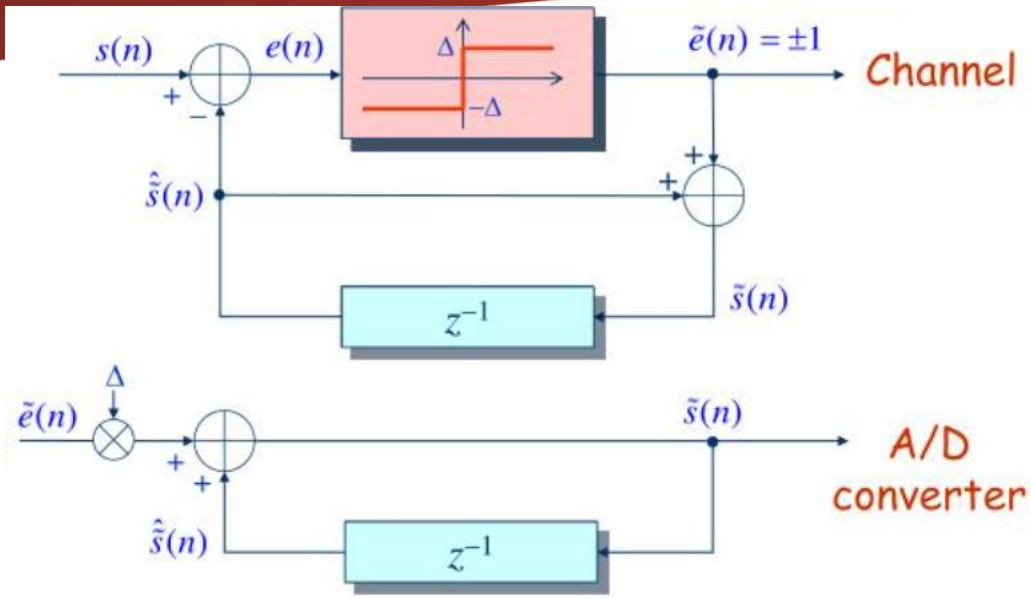
3) ADPCM – detalii de implementare

- Implementarea codorului (blocurile si explicarea acestora)
- Implementarea decodorului
- **… (vezi si documentele pdf anexate: “Codarea ADPCM – G.726)**

5) CODORUL DELTA



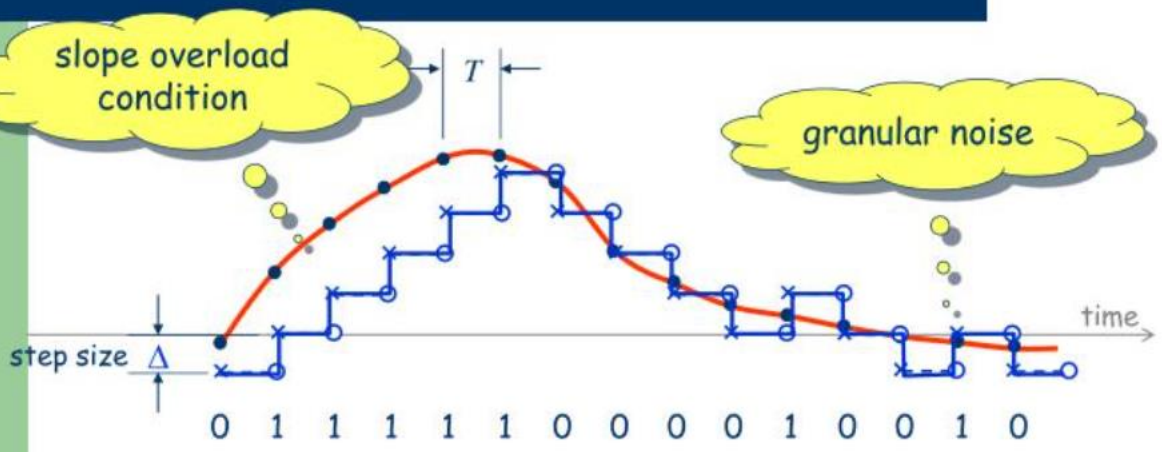
5) CODORUL DELTA



- Simplest form of DPCM
  - The prediction of the next is simply the current
- Sampling rate chosen to be many times (e.g., 5) the Nyquist rate, adjacent samples are quite correlated, i.e., $s(n) \approx s(n-1)$.
  - 1-bit (2-level) quantizer is used
  - Bit-rate = sampling rate

5) CODORUL DELTA

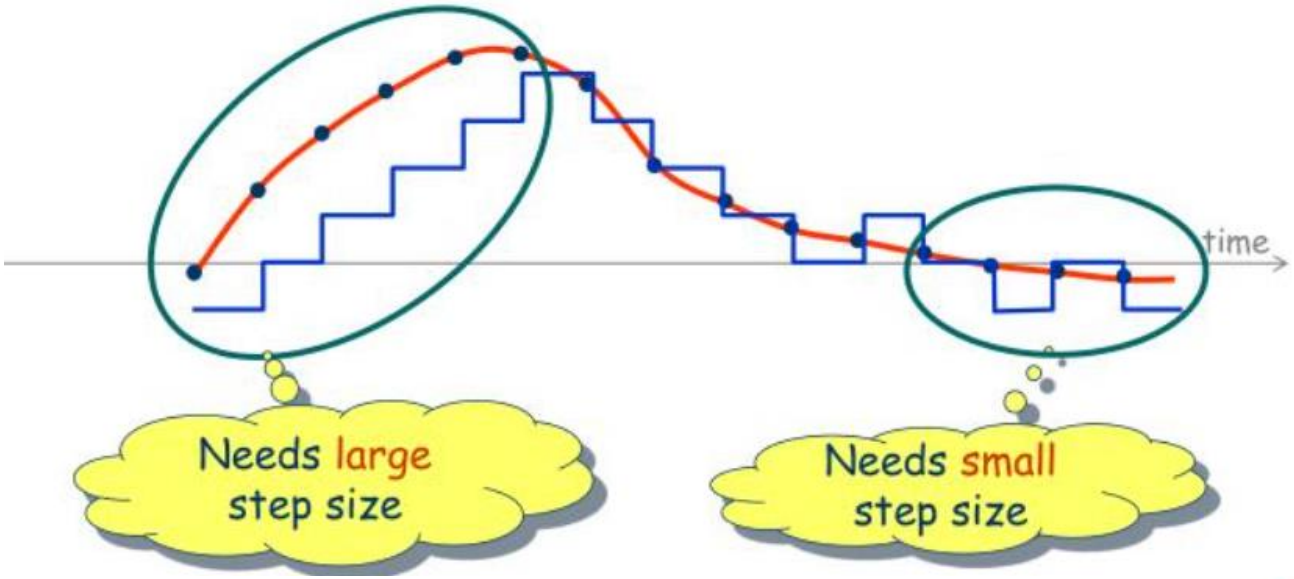
Distortions of DM



code words: $c(n) = \begin{cases} 1 & e(n) = +1 \\ 0 & e(n) = -1 \end{cases}$

5) CODORUL DELTA

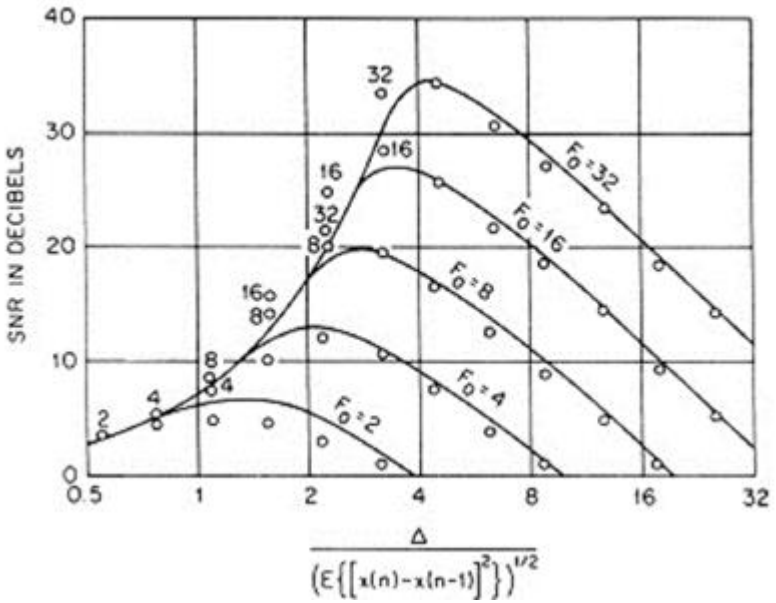
Choosing of Step Size



5) CODORUL DELTA - semnale

- Modalitati de urmarire semnal:
  - Cuanta fixa
  - Cuanta adaptabila

- Zgomotul:
  - de neurmarire
  - de granularitate



$$BR = F_S = 2F_N \cdot F_0$$

6) ADPCM DE BANDA LARGA

- Vezi documente pdf: “Codorul ADPCM de banda larga”

4) IMA - ADPCM

- Schema de prelucrare
- Caracteristici
- … vezi si notite curs.

CONCLUZII - discutii

- recapitulare notiuni
- intrebari

Wideband PCM (banda 8KHz, Fe = 16 KHZ). Low: 64 Kbps, High 16 Kbps

Figure 6-1 shows the high-level block diagram of the encoder. **A pre-processing high-pass filter is applied to the 16-kHz-sampled input signal** $s_{WB}(n)$ **to remove 0-50 Hz** components. The pre-processed signal $\tilde{s}_{WB}(n)$ is divided into 8-kHz-sampled **lower-band and higher-band signals**, $s_{LB}(n)$ and $s_{HB}(n)$, using a 32-tap quadrature mirror filterbank (QMF). The lower-band signal is encoded with an embedded lower-band encoder which generates **G.711-compatible** core bitstream (Layer 0) I_{L0} at **64 kbit/s, and lower-band enhancement (Layer 1) bitstream** I_{L1} **at 16 kbit/s**. The **higher-band signal** is transformed into modified discrete cosine transform (MDCT) domain and the frequency domain coefficients $S_{HB}(k)$ are encoded together with and its normalization factor η_{HB} by the higher-band encoder which generates higher-band enhancement (Layer 2) bitstream I_{L2} at **16 kbit/s.** All bitstreams are multiplexed as a scalable bitstream.

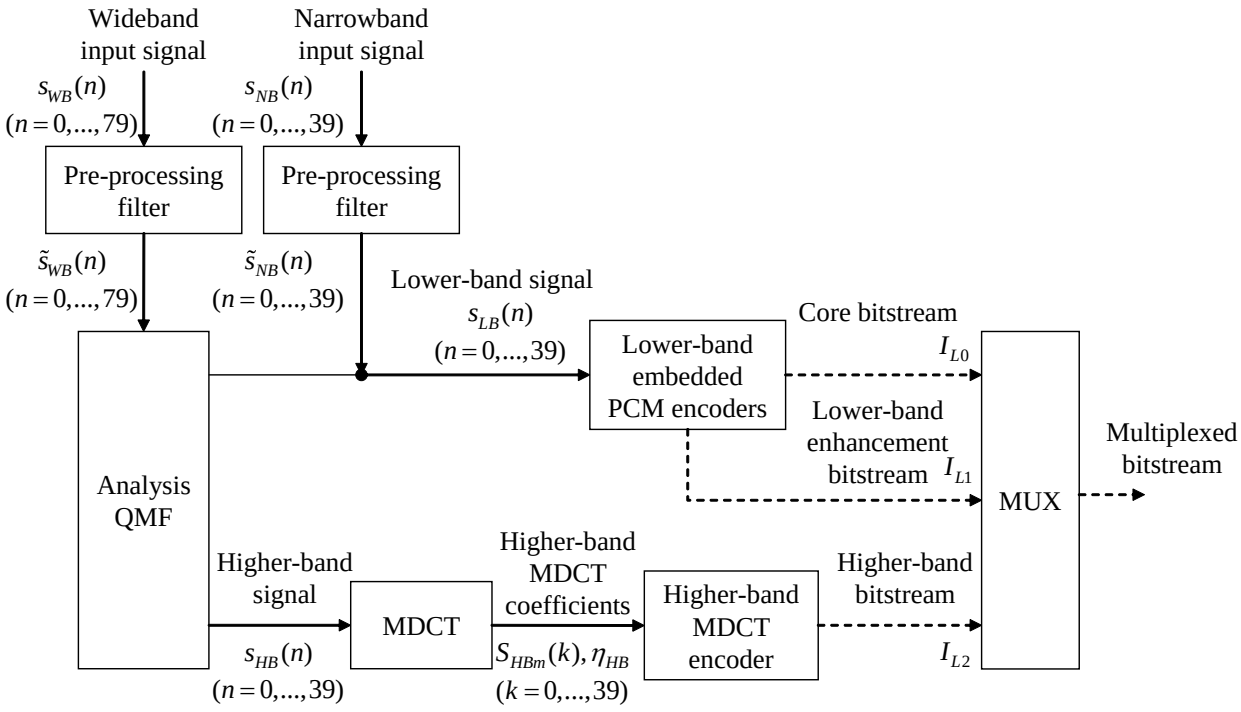


Figure 6-2 shows the high-level block diagram of the decoder. The whole bitstream is de-multiplexed to G.711-compatible core (Layer 0) bitstream I_{L0}, lower-band enhancement (Layer 1) bitstream I_{L1}, and higher-band enhancement (Layer 2) bitstream I_{L2}. Both the Layer 0 and 1 bitstreams are handed to the lower-band decoder. The Layer 2 bitstream is given to the higher-band decoder, and consequently decoded signal in the frequency domain $\hat{S}_{HB}(k)$ is fed to inverse MDCT (iMDCT), together with the normalization factor η_{HB}. By this iMDCT process, the higher-band signal in time domain $\hat{s}_{HB}(n)$ is obtained. To improve the quality under frame erasures due to channel errors such as packet-losses, frame erasure concealment (FERC) algorithms are applied to the lower-band and higher-band signals independently. Although the concealment process is performed independently, the pitch lag T_{LB} estimated in the lower-band FERC is given to higher-band FERC as auxiliary information. The lower- and higher-band signals, $\hat{s}_{LB}(n)$ and $\hat{s}_{HB}(n)$, are combined using a synthesis QMF filterbank to generate a wideband signal $\hat{s}_{QMF}(n)$. Noise gate processing is applied to the QMF output to reduce low-level background noise. At the decoder output, 16-kHz-sampled speech $\hat{s}_{WB}(n)$ (or 8-kHz-sampled speech $\hat{s}_{NB}(n)$) is synthesized.

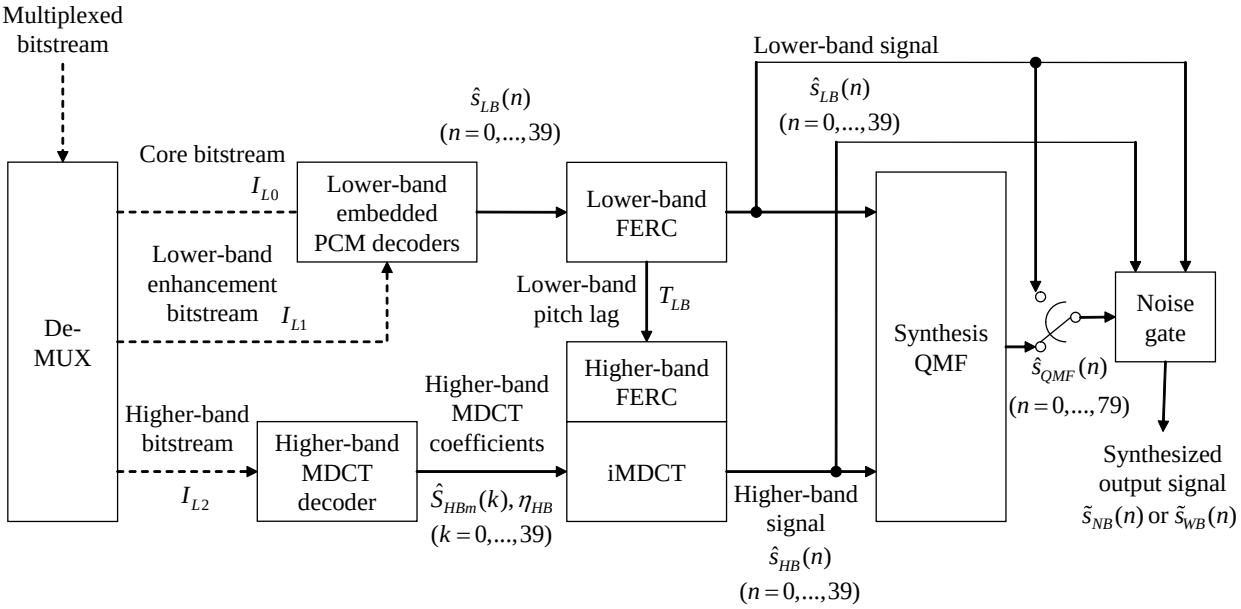
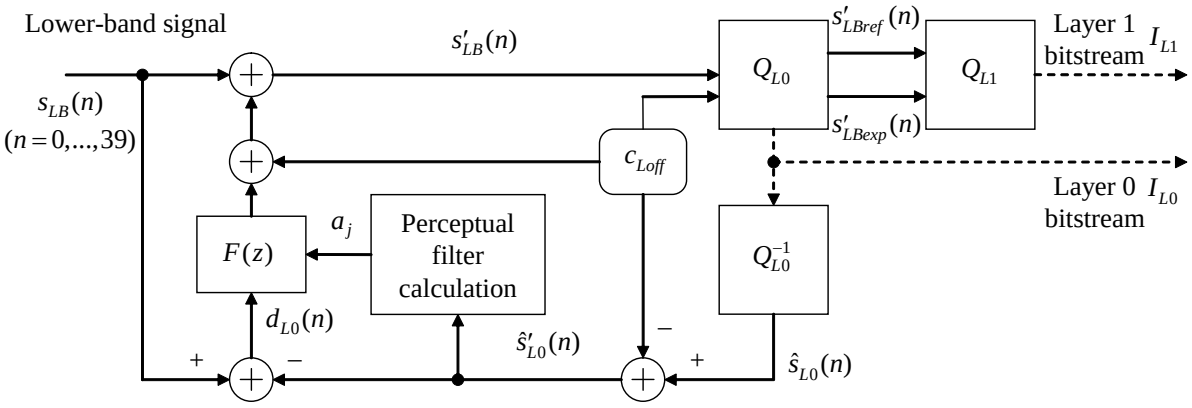


Figure 7-2 – Lower-band encoder with weighted noise feedback loop



3. CODAREA ADPCM G.726 A SEMNALULUI VOCAL

Obiective:

- ***intelegerea necesitatii transmiterii digitale a semnalului vocal***
- ***cunoasterea structurii si functiilor unui codor ADPCM G.726***
- ***studiul performantelor codorului ADPCM G.726 prin implementarea lui practica***

3.1. Introducere

In sistemele de prelucrare digitala a semnalului vocal exista **avantaje** unanim recunoscute:

- stocare simpla
- precizie deosebita
- posibilitati de compresie
- imunitate la erori, etc.

Prelucrarea numerica este de asemenea incurajata de dezvoltarea exploziva a calculatoarelor si a circuitelor integrate digitale. Tehnica digitala se impune nu numai in prelucrarea semnalelor dar si in transmisia lor, iar transmiterea semnalului vocal a fost si ramine un obiectiv important in retelele de telecomunicatii. In acest context telecomunicatiile, in general si telefonia, in special, evolueaza rapid spre digital .

Reprezentarea informatiei analogice (semnal vocal) sau digitale (date) este aleasa astfel incit sa se asigure **optimizarea unei anumite caracteristici a transmisiunii** cum ar fi:

- utilizarea optima a canalului
- compatibilitatea cu un anumit tip de mediu sau echipament
- costuri scazute
- asigurarea secretului transmisiei sau a unui anumit nivel de calitate .

Transmiterea digitala a semnalului vocal prezinta un numar impresionant de avantaje. Cea mai mare parte dintre acestea caracterizeaza transmisiunile digitale in general:

- rezistenta mare la perturbatii (raportul semnal / zgomot necesar la receptie este practic cu 20 de dB mai mic decit in cazul transmisiilor analogice)
- posibilitatea de regenerare a semnalelor (zgomotul nu se mai acumuleaza)
- posibilitatea de protectie la erori prin codarea canalului si de asigurare a secretului transmisiei prin criptare
- multiplexarea (cu diviziune in timp mai ieftina decit cea cu diviziune in frecventa)
- avantaje tehnologice (circuite VLSI dedicate, procesoare de semnal, interfete elctro-optice , etc .) .

Nu cu multi ani in urma principalul dezavantaj al transmisiunilor digitale era reprezentat de cresterea benzii necesare transmisiei. Odata cu dezvoltarea si implementarea algoritmilor evoluati de digitizare a vocii, care asigura o calitate corespunzatoare a vocii la debite de 32 Kbps, 24 Kbps sau chiar la 13, respectiv 6.5 Kbps in telefonia mobila, eficienta spectrala a transmisiunilor digitale a devenit comparabila sau chiar mai buna decit cea a transmisiunilor analogice, cel putin din punctul de vedere al transmisiilor de voce. Fara indoiala ca elaborarea unei tehnici performante de modulatie , pe de o parte, si extinderea impresionanta a utilizarii fibrei optice ca mediu de transmisiune, pe de alta parte, a contribuit esential la aceasta. Ca problema ramine necesitatea asigurarii sincronizarii la diferite nivele, cerinta cu atit mai costisitoare cu cit debitul in retelele de telecomunicatii creste.

Vocea umana este o sursa de natura analogica. Ea este abordata de o maniera statistica fiind modelata prin distributii de variabile aleatoare continue, cu variatii puternice in timp si de la un vorbitor la altul. Semnalele vocale sunt nestationare si continue. Ele pot fi considerate ca si cvasistationare pe intervale de timp scurte. **Spectrul de putere** al semnalului vocal difera in functie de sunet. El este cuprins in general intre 80 de Hz si 12 KHz, densitatea spectrala de putere scazind puternic la frecvente inalte, mai mari de 4 KHz.

EXERCITIU:

Care sunt avantajele prelucrarii digitale in comparatie cu prelucrarea analogica?

EXERCITIU:

Argumentati imunitatea la perturbatii a comunicatiilor digitale in compartie cu comunicatiile analogice. Dati exemple pentru cazul particular a semnalului vocal.

EXERCITIU:

Determinati debitul binar pentru transmisia semnalului vocal, respectiv a semnalelor audio de calitate CD in format PCM si dati solutii pentru liniile de comunicatie care suporta aceste debite.

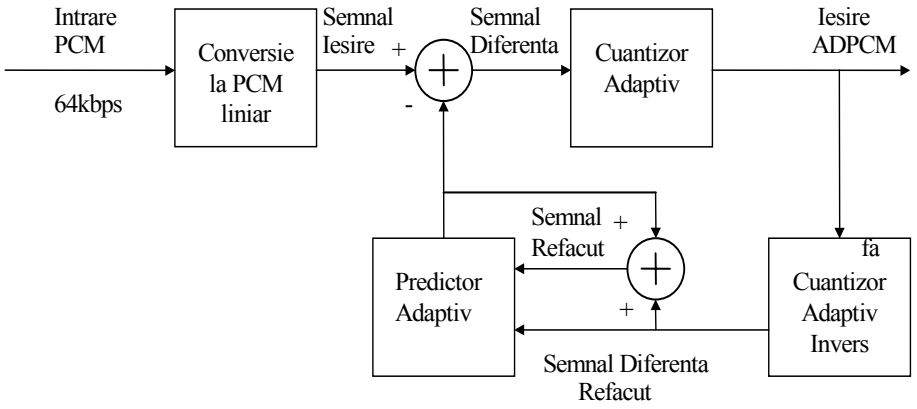
3.2. Sistemul de codare ADPCM G.726

Caracteristicile ce se vor descrie in continuare se recomanda pentru conversia unui canal PCM de 64 de Kbps (legea A sau μ) la/de la un canal de 40, 32, 24 sau 16 kbps. Conversia se aplica debitelor binare PCM folosind o tehnica de **transcodare** ADPCM. Relatia dintre frecventa semnalului si legile de codare - decodare PCM este complet specificata in Recomandarea G.711.

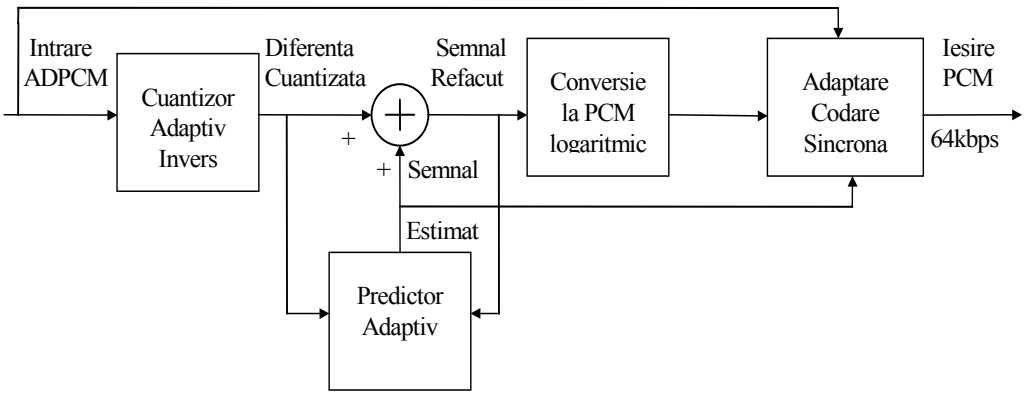
Principala aplicatie a canalelor de 24 si 16 Kbps este folosirea lor ca si canale suplimentare de voce, iar canalele de 32 si 40 Kbps se folosesc pentru transmisii de prin modemuri operind la mai mult de 4800 Kbps .

3.2.1. Codorul ADPCM G.726

Dupa conversia semnalului PCM de intrare (legea A sau μ), in semnal PCM liniar se obtine un semnal diferenta prin scaderea din semnalul de intrare a unui semnal estimat. Aceast semnal diferenta este codat adaptiv. Un cuantizor adaptiv cu 31, 15, 7 sau 4 nivele (in functie de debitul de iesire) este folosit pentru a coda pe 5, 4, 3 sau 2 biti valoarea semnalului diferenta (Figura 3.2.1.)



a) CODORUL



b) DECODORUL

Figura 3.2.1. Scheme de principiu

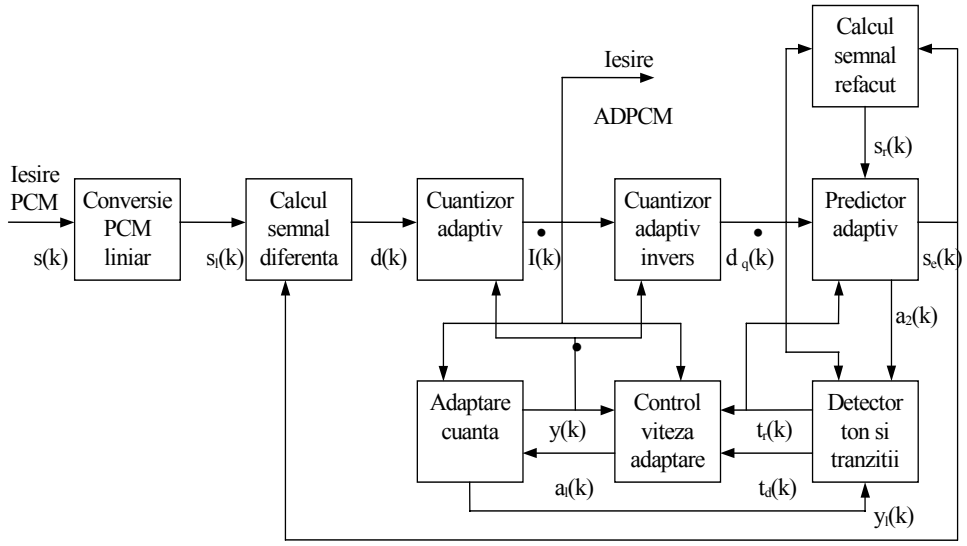


Figura 3.2.2. Codorul - schema bloc

Figura 3.2.2. reprezinta schema bloc a codorului. Pentru variabilele ce vor apare in continuare k este indexul indexul esantionului, iar esantioanele se iau la intervale de 125 μs (f_e =8 kHz).

Conversia formatului PCM de intrare

Blocul *Conversie PCM liniar* converteste semnalul PCM de intrare s(k) din logaritmic (legea A sau μ) in semnal PCM uniform (liniar), $s_l(k)$.

Calcul diferenta

Acest bloc calculeaza semnalul d(k), obtinut ca diferenta dintre semnalul uniform PCM $s_l(k)$ si semnalul estimat $s_e(k)$:

$$d(k) = s_l(k) - s_e(k)$$

Cuantizorul adaptiv

Un cuantizor adaptiv neuniform cu 31, 15, 7 sau 4 nivele este folosit pentru a cuantiza semnalul diferenta d(k) rezultind 40, 32, 24 respectiv 16 Kbps. Inainte de cuantizare d(k) este normalizat (logaritmat in baza 2) si scalat cu y(k) (calculat de blocul de adaptare a cuantei):

$$d_n(k) = \log_2(d_l(k)) - y(k)$$

Operatii la 32 de biti (exemplificare), analog se procedeaza si pentru 40, 24, 16 Kbps)

Pentru reprezentarea nivelului cuantizat $d_n(k)$ se folosesc 4 biti (trei pentru amplitudine si unu pentru semn). Iesirea pe 4 biti a cuantizorului I(k), este chiar semnalul de iesire la 32 kbps; I(k) ia una din cele 15 valori diferite de zero, dar I(k) este semnal de intrare pentru cuantizorul adaptiv invers, pentru blocurile de control a vitezei de adaptare si adaptarea factorului de scala.

Tabelul 3.2.1. Corespondenta valoare diferenta - nivel de cuantizare (32 Kbps)

$d_n(k)$	I I(k) I	$d_{qn}(k)$
(3.12 , +∞)	7	3.32
(2.72 , 3.12)	6	2.91

(2.34 , 2.72)	5	2.52
(1.91 , 2.34)	4	2.13
(1.38 , 1.91)	3	1.66
(0.62 , 1.38)	2	1.05
(-0.98 , 0.62)	1	0.031
(-∞ , -0.98)	0	-∞

Cuantizorul adaptiv invers

Versiunea cuantizata a semnalului diferenta $d_q(k)$ se obtine prin scalare folosind y(k) si diferenta cuantizata normalizata $d_{qn}(k)$, care se gaseste in tabel, si apoi se renunta la domeniul logaritmic :

$$d_q(k) = 2^{dqn(k) + y(k)}$$

Adaptarea factorului de cuantizare (cuantei)

Acest bloc calculeaza y(k), cuanta folosita atit de cuantizor cit si de cuantizorul invers. La intrarea lui avem iesirea cuantizorului I(k) si parametrul de control al vitezei de adaptare a_l. Specifice sunt cele doua moduri de adaptare a cuantei :

- **rapid** , pentru semnale cu fluctuatii mari
- **lent** , pentru semnale cu fluctuatii mici

Viteza de adaptare este cotrolata de o combinatie de factori de scala rapizi si lenti. Factorul de scala rapid (unlocked), $y_u(k)$, este calculat recursiv in domeniu logaritmic in baza doi, din factorul de scala logaritmic rezultat anterior, y(k) :

$$y_u(k) = (1 - 2^{-5}) y(k) + 2^{-5} \text{ WI I(k) I}$$

unde $y_u(k)$ este limitat intre : $1.06 \le y_u(k) \le 10.00$

De exemplu pentru 32 kbps, functia discreta W(I) este definita dupa cum urmeaza :

I I(k)I	7	6	5	4	3	2	1	0
WI I(k) I	70.13	22.19	12.38	7.00	4.00	2.56	1.13	-0.75

Factorul de scala lent (locked) $y_l(k)$ se obtine din $y_u(k)$ printr-o operatie de filtrare trece jos :

$$y_l(k) = (1 - 2^{-6}) y_l(k-1) + 2^{-6} y_u(k)$$

Factorii de scala rapid si lent sunt apoi combinati pentru a obtine factorul de scala rezultant:

$$y(k) = a_l(k) y_u(k-1) + (1- a_l(k)) y_l(k - 1)$$

unde : $0 \le a_l(k) \le 1$

Controlul vitezei de adaptare

Parametrul de control $a_l(k)$ poate lua valori in intervalul (0, 1). El tinde spre unu pentru semnal vocal si tinde spre zero pentru semnal de date. Este o masura a ratei de variatie a valorilor semnalului diferenta. Se calculeza doua variante ale amplitudinii medii a lui I(k) :

$$d_{ms}(k) = (1 - 2^{-5}) d_{ms}(k-1) + 2^{-5} \text{ FI I(k) I}$$

si $$d_{ml}(k) = (1 - 2^{-7}) d_{ml}(k-1) + 2^{-7} \text{ FI I(k) I}$$

Pentru 32de kbps F │I(k)│ este definit ca :

│I(k)│	7	6	5	4	3	2	1	0
F │I(k)│	7	3	1	1	1	0	0	0

Astfel $d_{ms}(k)$ este media pe termen scurt, iar $d_{ml}(k)$ este media pe termen lung a lui F │ I(k)│ .
Folosind aceste doua medii , variabila $a_p(k)$ poate fi definita astfel :

$$a_p(k)=\begin{cases} (1-2^{-4})\,a_p(k-1)+2^{-3}, & \text{daca } |d_{ms}(k)-d_{ml}(k)| \geq 2^{-3}\,d_{ml}(k) \\ (1-2^{-4})\,a_p(k-1)+2^{-3}, & \text{daca } y(k)<3 \\ (1-2^{-4})\,a_p(k-1)+2^{-3}, & \text{daca } t_d(k)=1 \\ 1, & \text{daca } t_r(k)=1 \\ (1-2^{-4})\,a_p(k-1), & \text{altfel} \end{cases}$$

$a_p(k)$ este limitat de $a_l(k)$:

$$a_l(k)=\begin{cases} 1, & \text{daca } a_p(k-1)>1 \\ a_p(k-1), & \text{daca } a_p(k-1)\leq 1 \end{cases}$$

Predictorul adaptiv si calculul semnalului refacut

Principala functie a predictorului adaptiv este de a calcula semnalul estimat $s_e(k)$ din semnalul diferenta cuantizat $d_q(k)$. Se folosesc doua structuri de predictoare adaptive, un predictor de ordinul sase, care modeleaza zerourile si un predictor de ordinul doi care modeleaza polii sistemului care produce semnalul de intrare. Acesta structura dubla ii permite lucreze mai bine cu marea varietate de semnale ce pot fi intilnite. Semnalul estimat este calculat ca:

$$s_e(k)=\sum_{i=1}^{2} a_i(k-1)\,s_r(k-i)+s_{ez}(k)$$

unde $$s_{ez}(k)=\sum_{i=1}^{6} b_i(k-1)\,d_q(k-i)$$

iar semnalul refacut este definit ca:

$$s_r(k-i)=s_e(k-i)+d_q(k-i)$$

Ambele seturi de coeficienti ai predictorului se modifica folosind un algoritm de gradient simplificat. Pentru predictorul de ordinul doi, avem :

$$a_1(k)=(1-2^{-8})\,a_1(k-1)+3\,2^{-8}\,\mathrm{sgn}[p(k)]\,\mathrm{sgn}[p(k-1)]$$
$$a_2(k)=(1-2^{-7})\,a_2(k-1)+2^{-7}\{\mathrm{sgn}[p(k)\mathrm{sgn}[p(k-2)]-f[a_1(k-1)]\mathrm{sgn}[p(k)\,\mathrm{sgn}[p(k-1)]\}$$

unde $p(k)=d_q(k)+s_{ez}(k)$,

$$f(a_1)=\begin{cases} 4a_1, & |a_1|\leq \tfrac{1}{2} \\ 2\,\mathrm{sgn}(a_1), & |a_1|>1/2 \end{cases}$$

si sgn[0]= 1 , cu exceptia lui sgn[p(k-i)] care poate fi 0 doar daca p(k-i) =0 si i = 0 ;
cu constringerile de stabilitate :

$$I\ a_2(k)\ I \le 0.75 \text{ si } I\ a_1(k)\ I \le 1 - 2^{-4} - a_2(k)$$

Daca $t_r(k) = 1$, atunci $a_1(k)$ i=$a_2(k) = 0$.

Pentru predictorul de ordinul sase :

$$b_i(k) = (1 - 2^{-8})\ b_i(k-1) + 2^{-7}\ \text{sgn}[d_q(k)]\ \text{sgn}[d_q(k-i)] ,$$

pentru i = 1 , 2 , ... , 6 .

Detector de ton si tranzitii

Pentru a imbunatati performantele sistemului pentru semnale provenite de la modemuri ce opereaza cu modulatia FSK, se realizeaza un proces de detectie in doi pasi. Mai intii subbanda de semnal este detectata astfel incit cuantizorul sa poata fi adus intr-un mod de adaptare rapid :

$$t_d(k) = \begin{cases} 1 , & a_2(k) < -0.71875 \\ 0 , & \text{altfel} \end{cases}$$

In plus, o tranzitie de la o banda de semnal este definita astfel incit coeficientii predictorului sa poata fi setati in zero si cuantizorul sa poata fi fortat intr-un mod rapid de adaptare

$$t_r(k) = \begin{cases} 1 , & a_2(k) < -0.71875\ [i\ I\ d_q(k)\ I > 24\ 2^{yl\ (k-1)} \\ 0 , & \text{altfel} \end{cases}$$

EXERCITIU: Explicati de ce se prefera codarea adaptiva a *diferentei* esantioanelor consecutive de semnal vocal si nu codarea adaptiva a esantioanelor?

EXERCITIU: Justificati cuantizarea neuniforma a diferentei, observind valorile din tabelul 3.2.1.

Simplificind schema de codare ADPCM G.726 prin eliminarea blocurilor: control viteza de adaptare, detector de ton si tranzitii si adoptind un predictor format dintr-un element de intirziere, se cere implementarea practica a codorului in scopul studiului performantelor pentru debitul de 32 Kbps.

EXERCITIU:

Cum se pot studia performantele pentru debitele de 40, respectiv 16 Kbps, indicind eventualele informatii suplimentare de care aveti nevoie. Indicati care sunt rapoartele semnal/zgomot la codare pentru diferite fisiere de sunet.

EXERCITIU:

Explicati cum are loc controlul vitezei de adaptare a codorului.

EXERCITIU:

Care este efectul limit`rii asimetrice a factorului de control $a_l(k)$ in conditiile in care I(k) ramine constant.

EXERCITIU:

Enuntati functiile predictorului adaptiv si dati schema unui predictor adaptiv cu doi poli si sase zerouri. Cum credeti ca influenteaza ordinul filtrului de predictie calitatea codarii?

3. 2.2. Decodorul ADPCM G.726

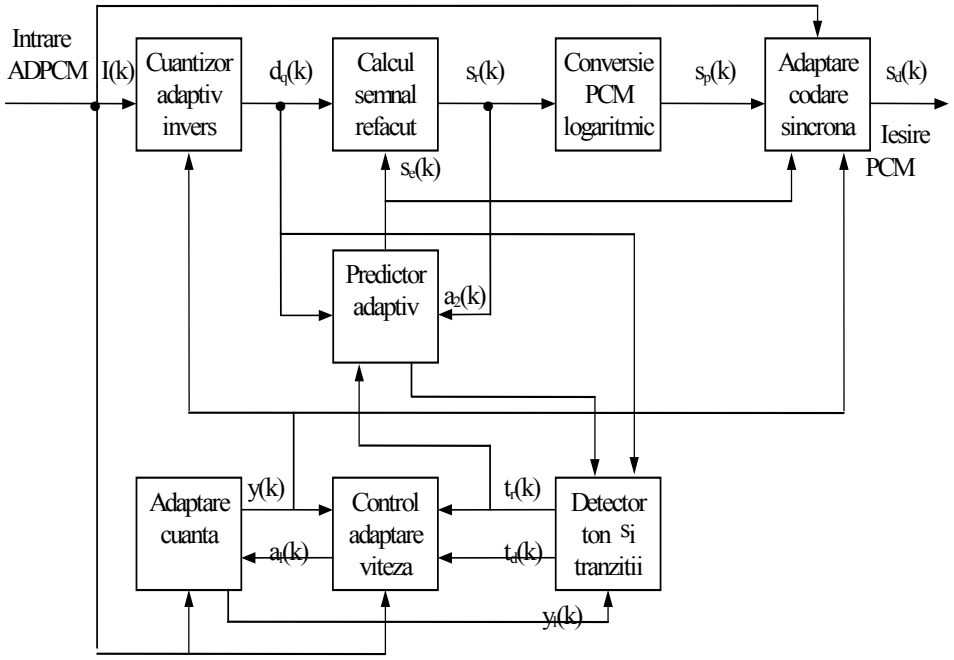


Figura 3.2.3. Decodorul - Schema bloc

O serie de blocuri care intervin in structura decodorului au fost descrise deja mai sus. In continuare se va face descrierea functionala a blocurilor care sunt specifice decodorului.

Conversia in format PCM de iesire

Acest bloc converteste semnalul refacut $s_r(k)$ intr-un semnal PCM logaritmic (legea A sau μ) , $s_p(k)$.

Adaptarea la codarea sincrona

Adaptarea la codarea sincrona previne acumularea distorsiunilor datorate codarii sincrone tandem (ADPCM - PCM - ADPCM , e t c.) si conexiunilor digitale, cind :

1) transmisia semnalelor ADPCM este fara erori

2) debitele ADPCM si PCM nu sunt alterate de dispozitivele digitale de procesare a semnalelor .

EXERCITIU:	Identificati diferentele dintre schema bloc a codorului si schema bloc a decodorului.

EXERCITIU:

Demonstrati ca in schema de codare ADPCM erorile de cuantizare sunt necumulative.

3.3. Implementarea unui codor/decodor ADPCM G.726

Deoarece ADPCM este o tehnica de codare complet digitala este avantajoasa implementarea ei soft. Pachetul de programe propus spre studiu se imparte in doua seturi.

Primul set, format din *codor.cpp* si *decodor.cpp* urmareste implementarea cit mai exacta a algoritmului matematic din Recomandarea G.726.

Al doilea set contine programele *compara.cpp, spectre.m, grafic.m, wavmaker.cpp* care au ca scop analiza rezultatelor experimentale furnizate de programele din primul set, adica analiza performantelor algoritmului G.726. Cele patru programe in C au fost grupate intr-un mediu integrat cu meniuri pentru o interfata mai prietenoasa cu utilizatorul, dar functioneaza si de sine statator.

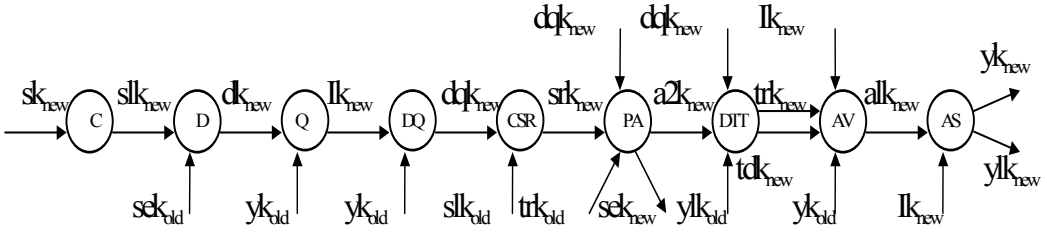


Figura 3.3.1. Organigrama codorului

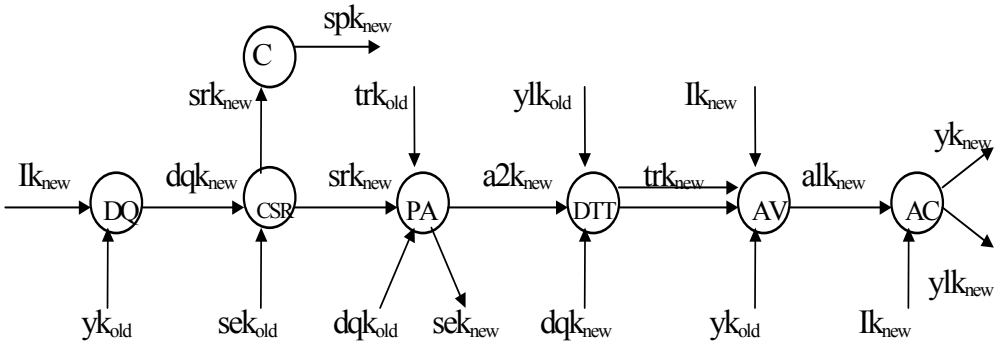


Figura 3.3.2. Organigrama decodorului

<<< CODOR / DECODOR ADPCM >>>

<!> Alegeti debitul la care doriti sa lucreze sistemul .
1. 16 kbps
2. 24 kbps
3. 32 kbps
4. 40 kbps
<?> : **3**

<!> Introduceti numele si extensia fisierului de intrare .
<?> : **noua.wav**
<!> A fost creat fisierul < adpcm.dat > .

<!> Introduceti numele fisierului cu eroarea de predictie .
<?> : **erprn_32.wav**
<!> A fost creat fisierul < erprn_32.wav > .

<!> Introduceti numele si extensia fisierului de iesire .
<?> : **noua_32.wav**
<!> A fost creat fisierul < noua_32.wav > .

EXEMPLU: 	*Fisierul audio original este " trei.wav " iar dupa codarea si decodarea ADPCM s-au obtinut fisierele refacute corespunzatoare celor patru debite la care lucreaza sistemul (40, 32, 24, 16 Kbps): " trei_40.wav ", " trei_32.wav ", " trei_24.wav " si " trei_16.wav ". In figura 3.4.2. sunt prezentate formele de unda a semnalului din cele 5 fisiere . Mai jos se prezinta tabelar si grafic rapoartele semnal zgomot calculate pentru cele patru debite .*

Tabelul 3.4.1. Raportul semnal/zgomot pentru diferite debite de codare/decodare

Debit [kbps]	Numar de bi✷i utiliza✷i la codare	RSZG [dB]
40	5	37
32	4	27
24	3	17
16	2	2

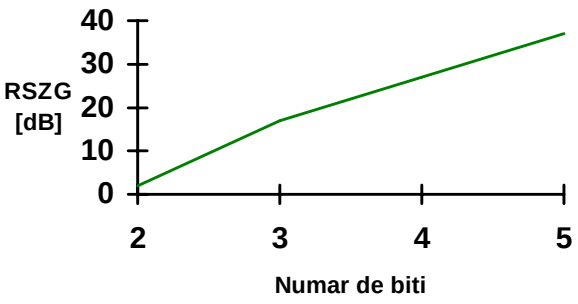


Figura 3.4.1. Dependenta raportului semnal / zgomot de numarul de biti

Din tabelul si graficul de mai sus se pot desprinde doua concluzii :

- raportul semnal / zgomot (RSZG) **creste aproximativ liniar** cu cresterea numarului de biti folositi la codare
- la cresterea cu unu a numarului de biti, raportul semnal / zgomot inregistreaza o **crestere de aproximativ 10 dB**.

In Figura 3.4.2. sunt prezentate comparativ formele de unda ale semnalului original si ale semnalelor refacute pentru diferite debite de codare. La o privire mai atenta se observa ca pe masura ce debitul scade, zgomotul din semnal devine mai vizibil, ajungind in final ca la debitul de 16 kbps semnalul sa fie foarte zgomotos. Acest lucru e confirmat si de rapoartele semnal / zgomot din tabel .

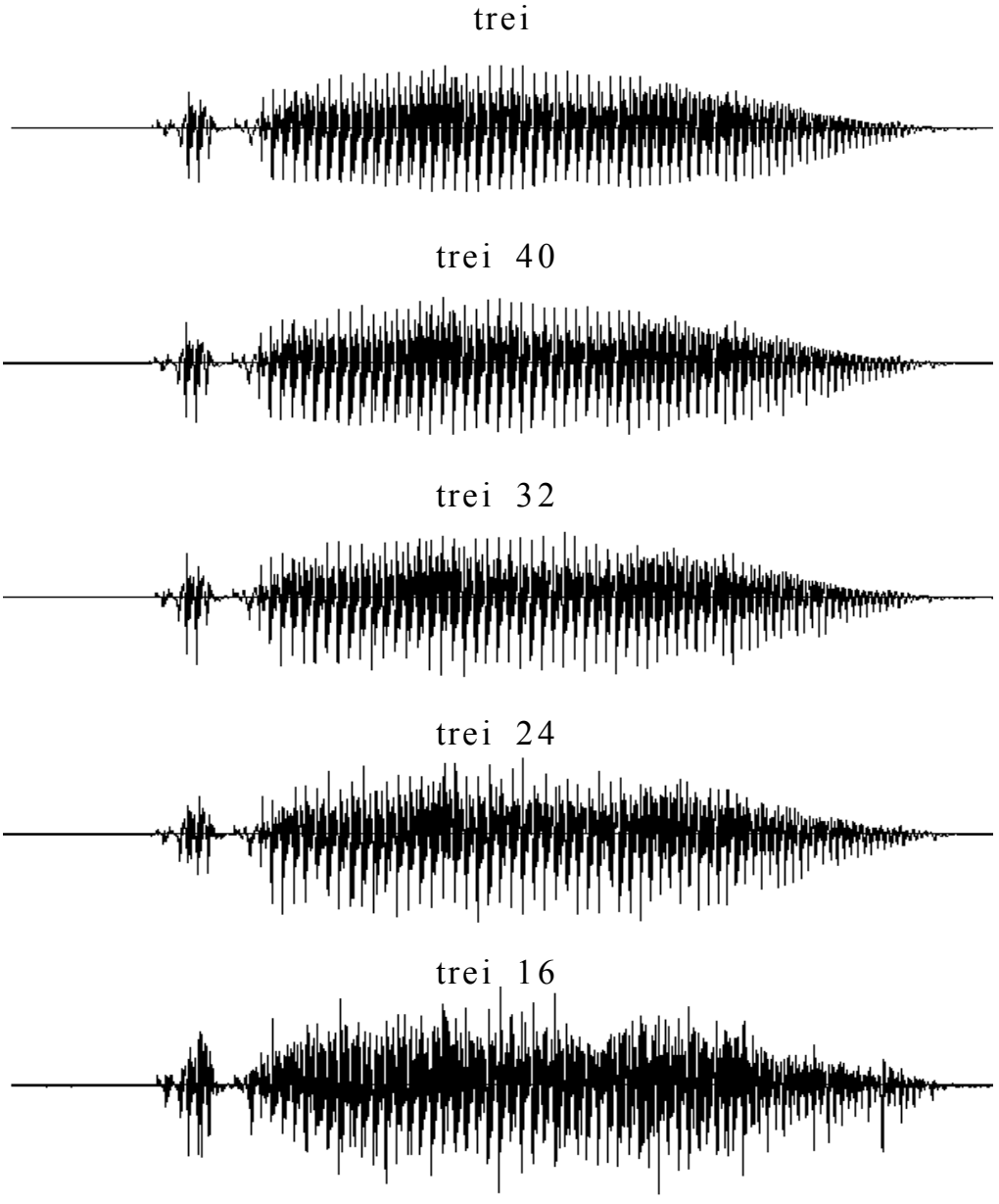


Figura 3.4.2. Formele de unda ale semnalului original si ale semnalelor refacute

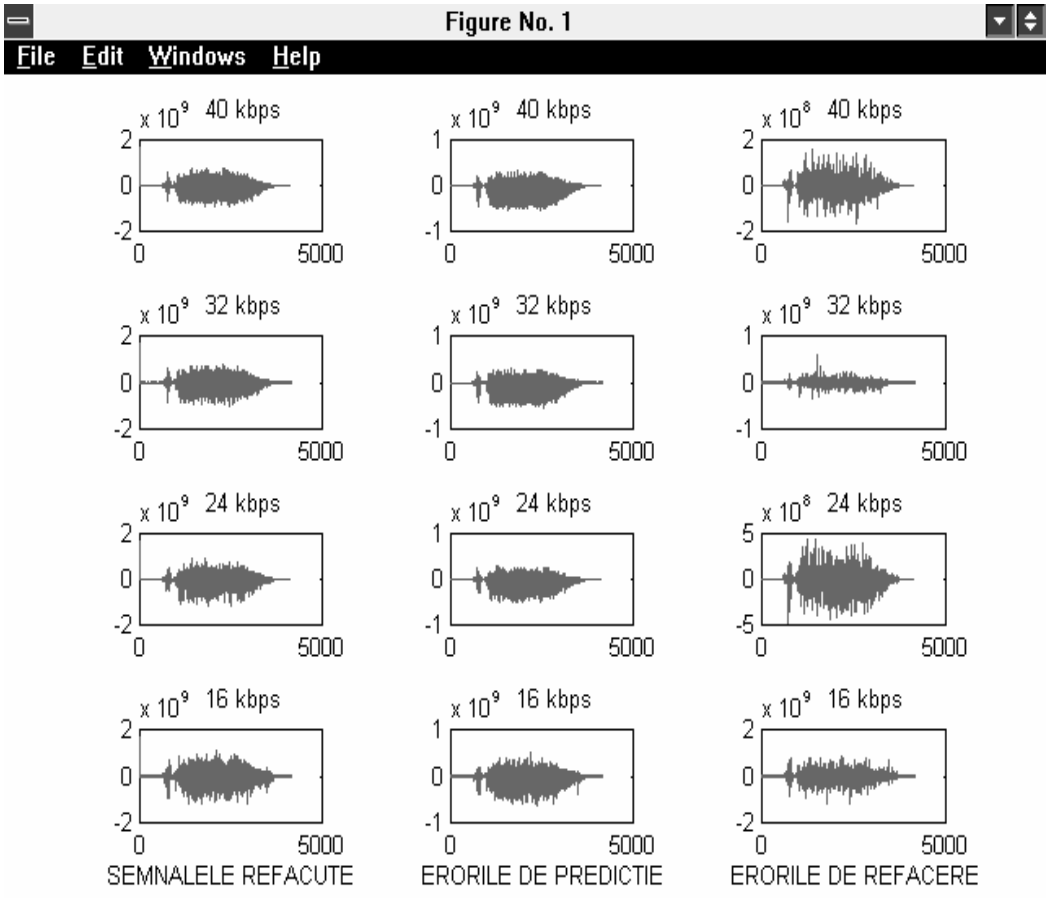


Figura 3.4.3. Comparatie intre semnalul refacut , eroarea de predictie si eroarea de refacere

In Figura 3.4.3. sunt prezentate comparativ semnalele refacute , eroarea de predictie si eroarea de refacere pentru cele patru debite la care poate lucra sistemul . Se observa ca cele doua semnale de eroare seamana foarte bine cu semnalul audio, avind doar o amplitudine mai redusa. La redarea in difuzor a erorii de predictie vom constata ca acesta suna exact ca semnalul original, avind o calitate mai slaba dar totusi neasteptat de buna si o intensitate mai mica, datorata amplitudinii mai mici.

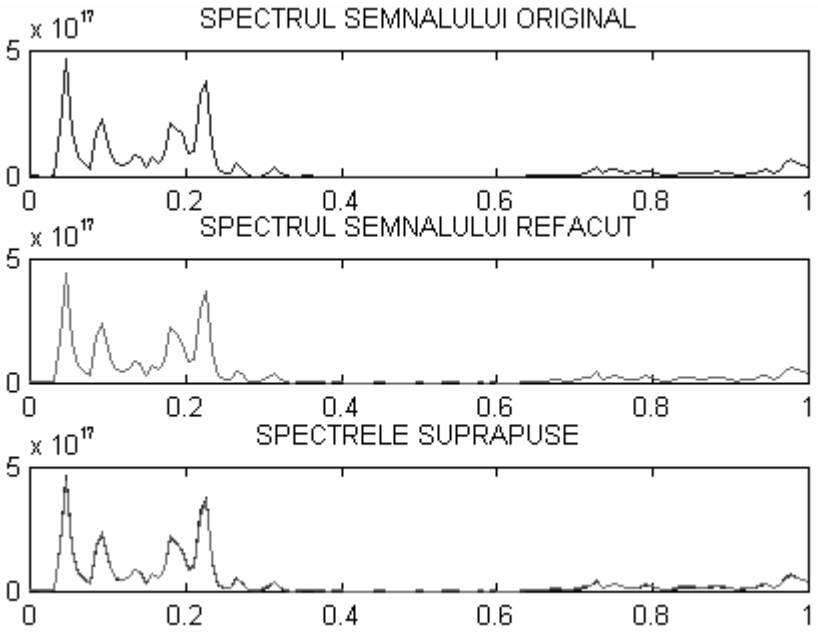


Figura 3.4.4. Spectrele semnalelor

In Figura 3.4.4. sunt prezentate spectrele semnalului *trei* si *trei* refacut la 32 de kbps . Faptul ca cele doua spectre sunt aproape identice ne asigura de faptul ca sistemul de codare decodare functioneaza foarte bine .

EXEMPLU: 	*Vom experimenta cum se comporta sistemul de codare / decodare in cazul in care avem la intrare un fisier de date. Trebuie sa reamintim ca algoritmul implementat trateaza diferit semnalele cu variatie rapida de cele cu variatie lenta. Semnalele de date utilizate sunt periodice, si au un numar oarecare de nivele de amplitudine. Semnalul de date este unipolar periodic cu 4 nivele. Primul nivel are amplitudinea 20000, al doilea 10000, al treilea 5000, si ultimul 0.In continuare sunt prezentate tabelar si grafic rapoartele semnal / zgomot obtinute pentru fiecare fisier comparativ, semnalul original si cele patru semnale refacute.*

Tabelul 3.4.2. Variatia RSZG pentru diferite debite de lucru la codarea datelor

Debit [kbps]	Num◆r de bi✷i utiliza✷i la codare	RSZG [dB]
40	5	31
32	4	27
24	3	15
16	2	13

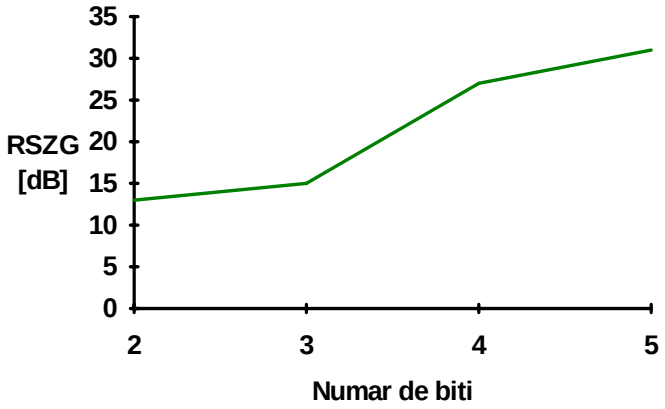


Figura 3.4.5. Variatia grafica a RSZG in functie de debit (numar de nivele de cuantizare)

Primul lucru care se observa este ca variatia raportului semnal zgomot la semnalele de date nu mai este la fel de liniara ca si in cazul semnalelor vocale. La codarea pe 2 sau 3 biti raportul semnal / zgomot se mentine oarecum constant (13 - 15 dB) apoi la 4 biti are o crestere de 12 dB, dupa care la cresterea cu inca unu a numarului de bitI, cresterea RSZG nu mai este asa importanta. La debitul recomandat pentru date de 40 Kbps, semnalul este refacut foarte bine, avind doar un usor zgomot de granularitate. La o privire atenta se poate observa ca valoarea acestui zgomot creste odata cu amplitudinea nivelului de semnal .

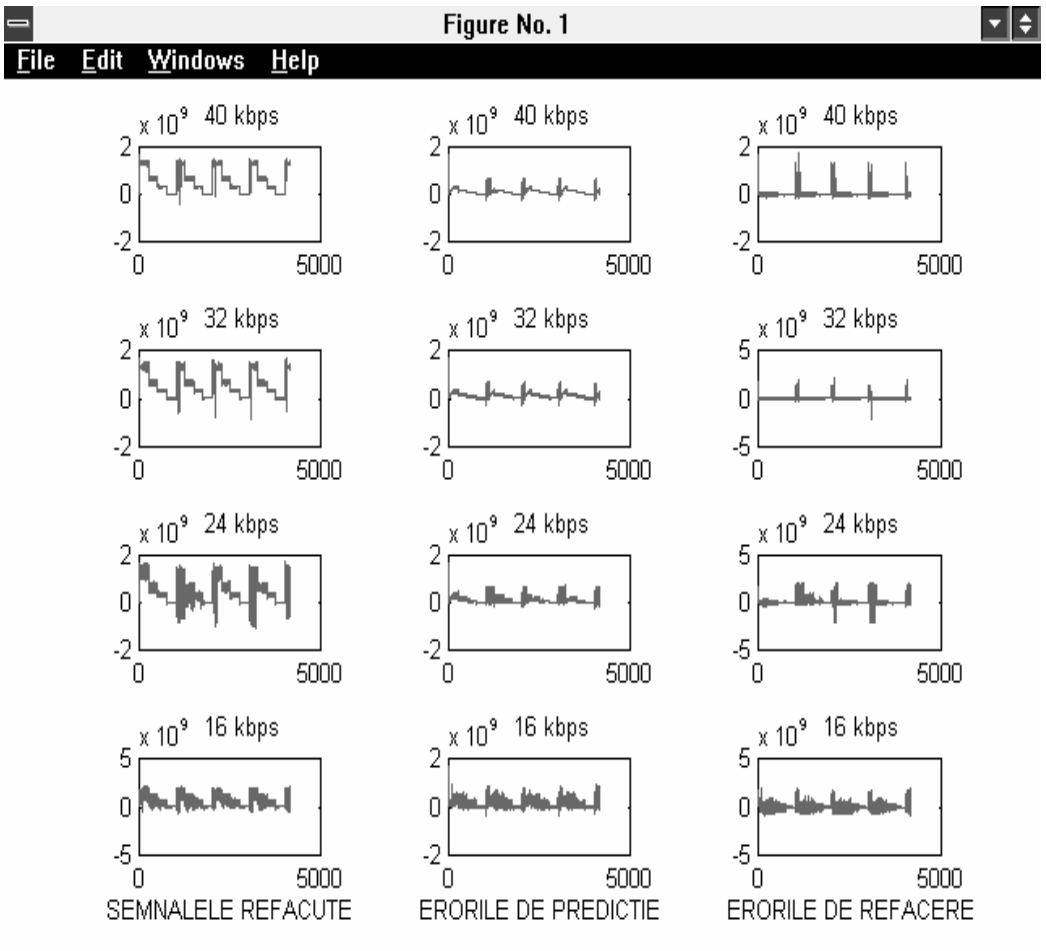


Figura 3.4.6. Comparatie intre semnalele refacute, erorile de predictie si erorile de refacere

EXEMPLU:	*In continuare vom prezenta rezultatele obtinute in urma prelucrarii la debitul de 32 Kbps (recomandat pentru semnal vocal) a 45 de fisiere audio reprezentind vocalele **a,e,i,o,u.** Pentru fiecare vocala prelucrata am luat 9 pronuntii diferite. Mai jos sunt prezentate tabelar rapoartele semna/zgomot obtinute.*

Tabelul 3.4.3. Evaluarea RSZG pentru codarea vocalelor la diferite debite

RAPORT SEMNAL / ZGOMOT [dB]										
Pronuntie	1	2	3	4	5	6	7	8	9	Mediu
a	20	17	17	16	17	18	22	14	14	17.2
e	25	27	25	26	26	24	24	23	27	25.2
i	25	28	25	27	26	27	24	26	25	25.8
o	19	23	15	19	17	22	21	22	18	19.5
u	13	29	20	22	26	15	27	23	26	22.3
RSZG mediu general : 22.04										

Spectrele si formele de unda obtinute pentru citeva vocale:

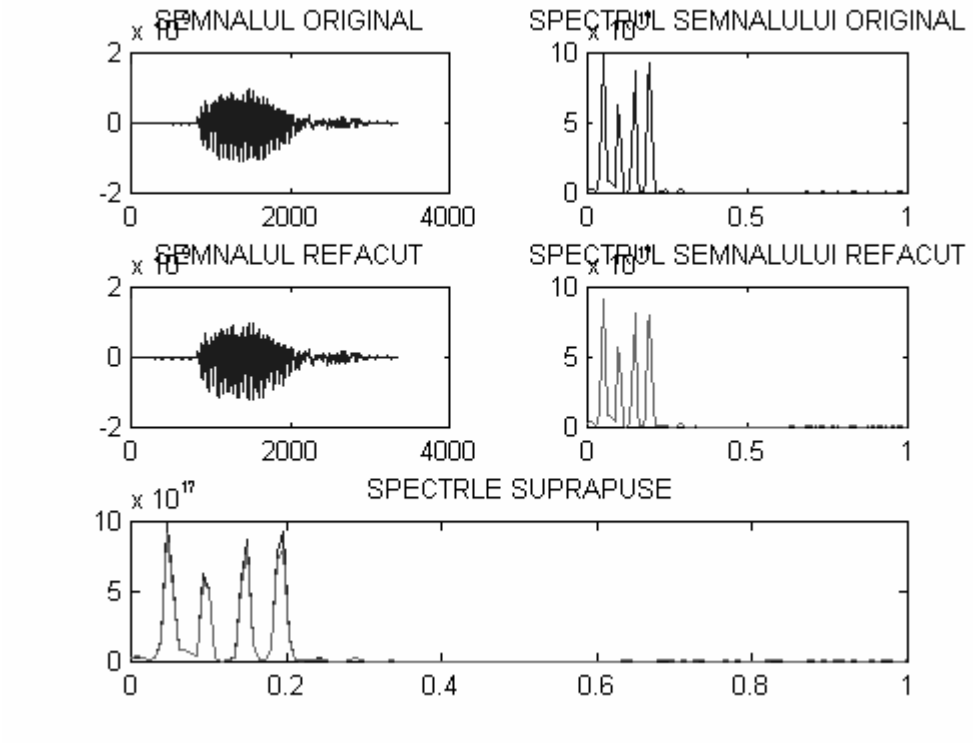


Figura 3.4.7. Spectrul semnalului original si refacut pentru vocala /e/

Algoritmul propus de Recomandarea G.726 se preteaza la implementare soft. Rezultatele obtinute sunt foarte bune. Rapoartele semnal zgomot obtinute pe cele citeva cazuri pe care s-au facut experimentele s-au situat intre 2 si 50 de dB, desigur in functie de debit.

In urma experimentelor se pot trage o serie de concluzii si anume ca raportul semnal zgomot depinde de debitul la care se face prelucrarea . Mai clar se vede acest lucru daca ne raportam la numarul de biti pe care se face codarea. Astfel, se poate spune ca in general la cresterea cu unu a numarului de biti, RSZG creste cu aproximativ 10 dB. De asemenea din graficele obtinute se observa o crestere aproape liniara a RSZG in raport cu cresterea numarului de biti.

In cazul semnalelor de date se obtin de asemenea rezultate bune, in special la debitul de 40 de Kbps. Pentru semnalele de date bipolare pare sa fie general faptul ca nivelele negative sunt refacute mai prost decit cele pozitive.

Trebuie sa remarcam ca atit pentru semnalele de date cit si pentru cele vocale asemanarea dintre spectrul semnalului original si al celui refacut este foarte mare, ceea ce indica o calitate foarte buna a sistemului de codare / decodare.

Foarte interesant este faptul ca erorile de predictie si refacere sunt asemanatoare cu semnalul original, incit daca le redam in difuzor suna la fel ca si semnalul original, avind doar o amplitudine mai mica si putin zgomot.

Concluzii:

- *Codorul ADPCM G.726 asigura codarea adaptiva a diferentei dintre doua esantioane succesive a unui semnal original (format PCM, frecventa de esantionare de 8 KHz, 8 bitI, legea A sau miu). Debitul de iesire poate fi selectat la: 40, 32, 24 sau 16 Kbps.*
- *Debitul de iesire depinde de numarul nivelelor de cuantizare folosite de blocul de cuantizare dupa cum urmeaza:*

  40 Kbps → 5 biti

32 Kbps →4 biti
24 Kbps →3 biti
16 Kbps →2 biti

- *In structura sistemului de codare se identifica blocurile: cuantizare adaptiva a diferentei, adaptare cuanta, control al vitezei de adaptare, predictorul adaptiv, detector de tonuri.*
- *Adaptarea cuantei se face sub comanda blocului de control al vitezei de adaptare si adaptare cuanta. Viteza de adaptare este controlata de o combinatie de factori de adaptare (rapid/lent) prin intermediul unui parametru de control. Adaptarea poate fi:*
  - *rapida - pentru semnale cu fluctuatii de amplitudine mari*
  - *lenta - pentru semnale cu fluctuatii mici*
- *Predictorul are o structura cu 2 poli si sase zerouri, asigurind stabilitatea sistemului in bucla inchisa si calculind un semnal estimat, in scopul minimizarii erorii de predictie. Coeficientii filtrului se modifica in timp pe baza unui algoritm de gradient simplificat.*
- *Codorul contine in structura sa, pe bucla de reactie negativa, elementele specifice ale decodorului. In decodor apare blocul de adaptare la codarea sincrona pentru a preveni erorile de codare sincrona tandem ADPCM - PCM - ADPCM.*
- *Implementarea practica a codorului poate fi realizata in timp real pe procesor de semnal sau chiar pe PC.*
- *Rezultatele experimentale obtinute la codarea cu acest tip de codec pun in evidenta urmatoarele:*
  - *prin prisma rapoartelor semnal zgomot, se recomanda utilizarea dbitului de 40 Kbps pentru transmisii de date, 32 Kbps pentru transmisii de voce, iar 24 Kbps si 16 Kbps pentru canale auxiliare de voce*
  - *pornind de la esantioane de 16 biti se pot obtine 5, 4, 3 sau 2 biti (corespunzator debitelor de 40, 32, 24, 16 Kbps) pentru fiecare esantion, rezultind rapoarte de compresie intre 3 si 8*
  - *RSZG creste liniar cu crsterea numarului de biti*
  - *o crestere cu 1 a numarului de biti conduce la o crestere cu aproximativ 10 dB a RSZG*
  - *eroarea de predictie are caracteristici acustice aproximativ identice cu cele ale semnalului original*
  - *spectrul semnalului refacut este aproape identic cu spectrul semnalului original*
  - *la codarea datelor, variatia RSZG nu mai depinde liniar de numarul de biti.*

4. CODAREA ADPCM PENTRU SEMNALE AUDIO DE BANDA LARGA - G.722

Obiective

- ***identificarea problematicii codarii semnalelor audio de banda larga***
- ***cunoasterea standardului ADPCM G.722***
- ***cunoasterea structurii si functiilor blocurilor componente dintr-un codec G.722***
- ***studiul posibilitatilor de multiplexare voce/date***

4.1. Principiul codarii

Codarea ADPCM a unor **semnale audio de banda larga pentru aplicatii multimedia** pune o serie de probleme. Astfel, datorita benzii semnalului de pina la 7 KHz, frecventa de esantionare standard de 8 kHz este insuficienta.

Standardul G.722 propune **codarea ADPCM in subbenzi**. Ideea este de a imparti banda de 7 - 8 KHz in doua subbenzi, inferioara si superioara, fiecare avind 4 KHz latime. Asupra acestor benzi se poate aplica o codare ADPCM clasica. Frecventa de esantionare in standardul G.722 este egala cu 16 KHz. Sistemul poate avea trei debite diferite de iesire (64 , 56 sau 48 Kbps) in functie de modul de lucru. Se ofera posibilitatea de a transmite pe linga voce, un canal de date de 8 sau 16 Kbps .

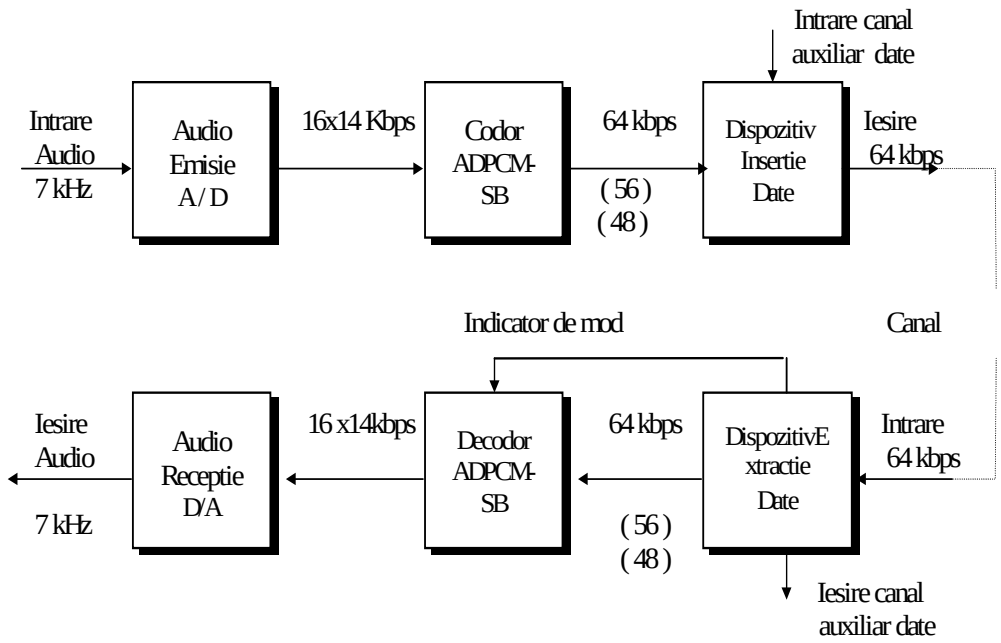


Figura 4.1.1. Schema de principiu a codorului ADPCM G.722.

Daca se transmit date trebuie adaugata informatie suplimentara pentru a preciza modul de lucru. Datele se transmit in cel mai putin semnificativ bit (la debit de 8 Kbps), sau in cei mai putin semnificativi doi biti ai subbenzii inferioare (la debit de 16 kbps).

Tabelul prezinta debitele corespunzatoare celor trei moduri de lucru .

Tabelul 4.1. Debitele transmise in sistemul de codare ADPCM G.722

Mod	Debit semnal audio [Kbps]	Debit canal auxiliar date [Kbps]
1	64	0
2	56	8
3	48	16

EXERCITIU:

Ce functii trebuie sa indeplineasca blocul "audio receptie" din Figura 4.1.1.?

EXERCITIU:

Considerind ca semnalul de intrare este esantionat la 16 KHz pe 14 biti, calculati raportul de compresie asigurat de schema din Figura 4.1.1. in diferitele conditii de lucru.

4.2. Descriere functionala

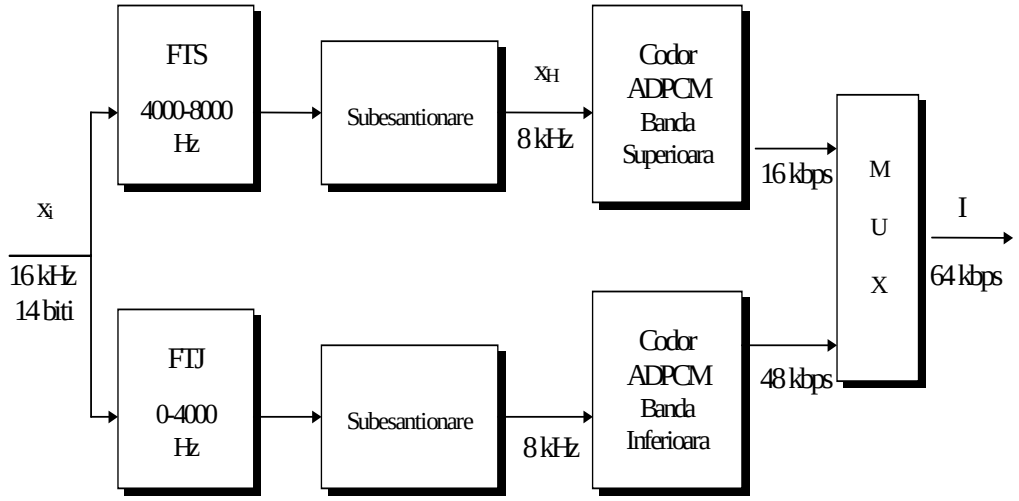


Figura 4.2.1. Codorul - descriere functionala

Codorul ADPCM pentru banda inferioara

EXERCITIU:

Pentru schema din Figura 4.2.1. justificati rolul blocului de subesantionare.

EXERCITIU:

Explicati diferentele de debit binar de la iesirea codorului ADPCM de pe banda inferioara, respectiv superioara.

Semnalul diferenta e_L se obtine prin scaderea semnalului estimat s_L din semnalul de intrare x_L.

$$e_L = x_L - s_L$$

Se observa ca prin suprimarea celor mai putin semnificativi doi biti de la iesirea cuantizorului apare posibilitatea de a insera un flux de date de maxim 2 x 8 = 16 Kbps pe canalul inferior, fara a afecta functionarea corecta a decodorului. Utilizarea unui cuantizor de 60 de nivele in loc de 64 garanteaza indeplinirea conditiei de densitate de amplitudine ceruta de standardul G.802 in toate conditiile si toate modurile de lucru.

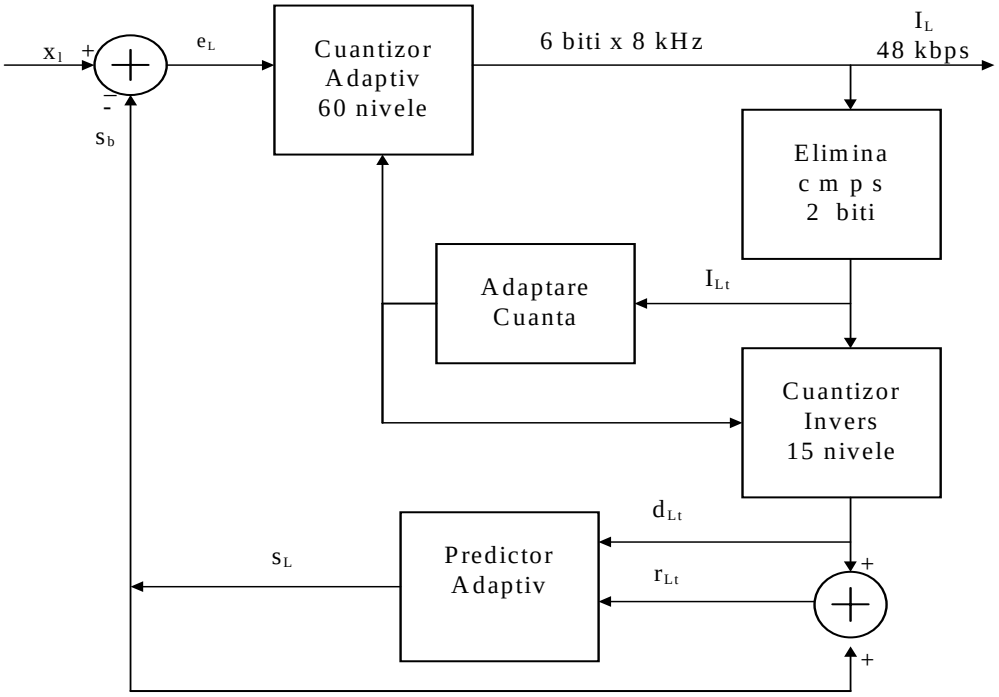


Figura 4.2.2. Codor ADPCM pentru banda inferioara

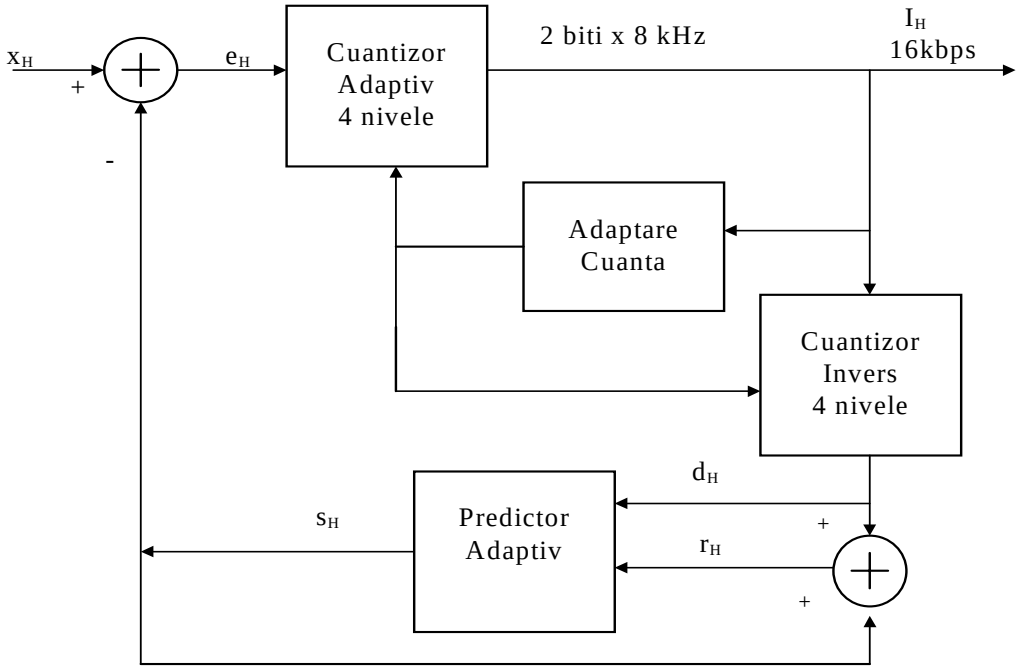


Figura 4.2.3. - Codor ADPCM pentru banda superioara

Debitele de 48, respectiv 16 Kbps se multiplexeaza intr-un debit unic de 64 kbps. Primul bit transmis in linie este I_{H1} .

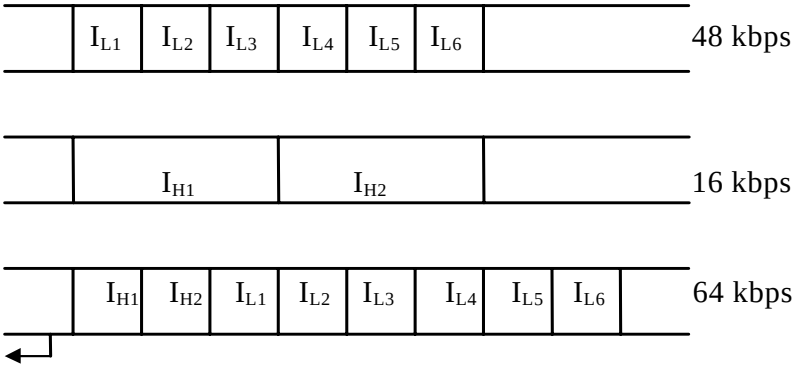


Figura 4.2.4. Multiplexarea celor doua benzi

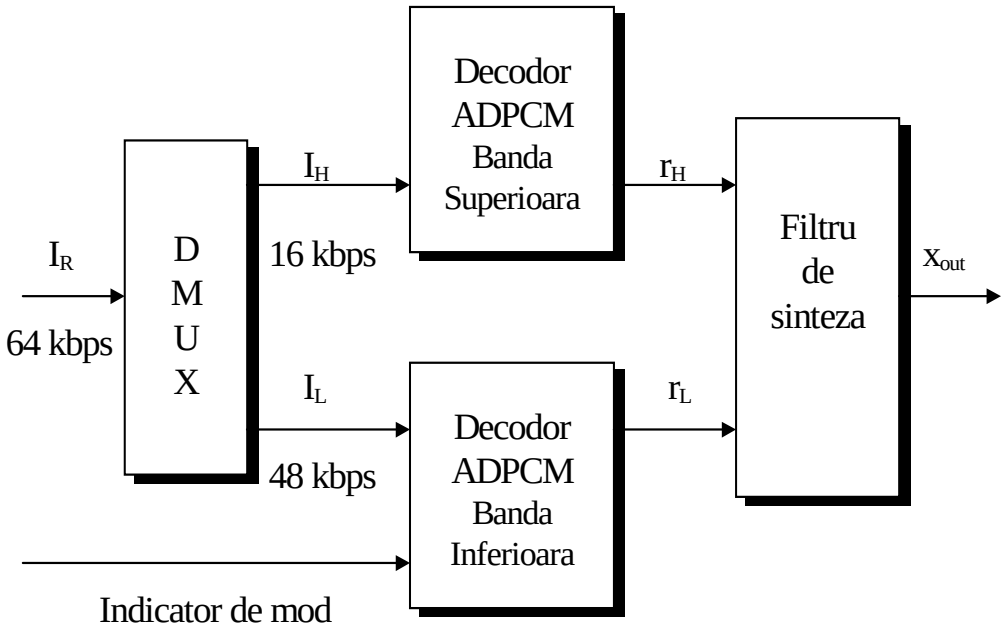


Figura x.6. Decodorul ADPCM G.722

EXERCITIU:	Identificati ce diferente exista intre codarea benzii superioare, respectiv inferioare in G.722

EXERCITIU:

Identificati ce diferente exista intre sistemul de codare G.726 si G.722.

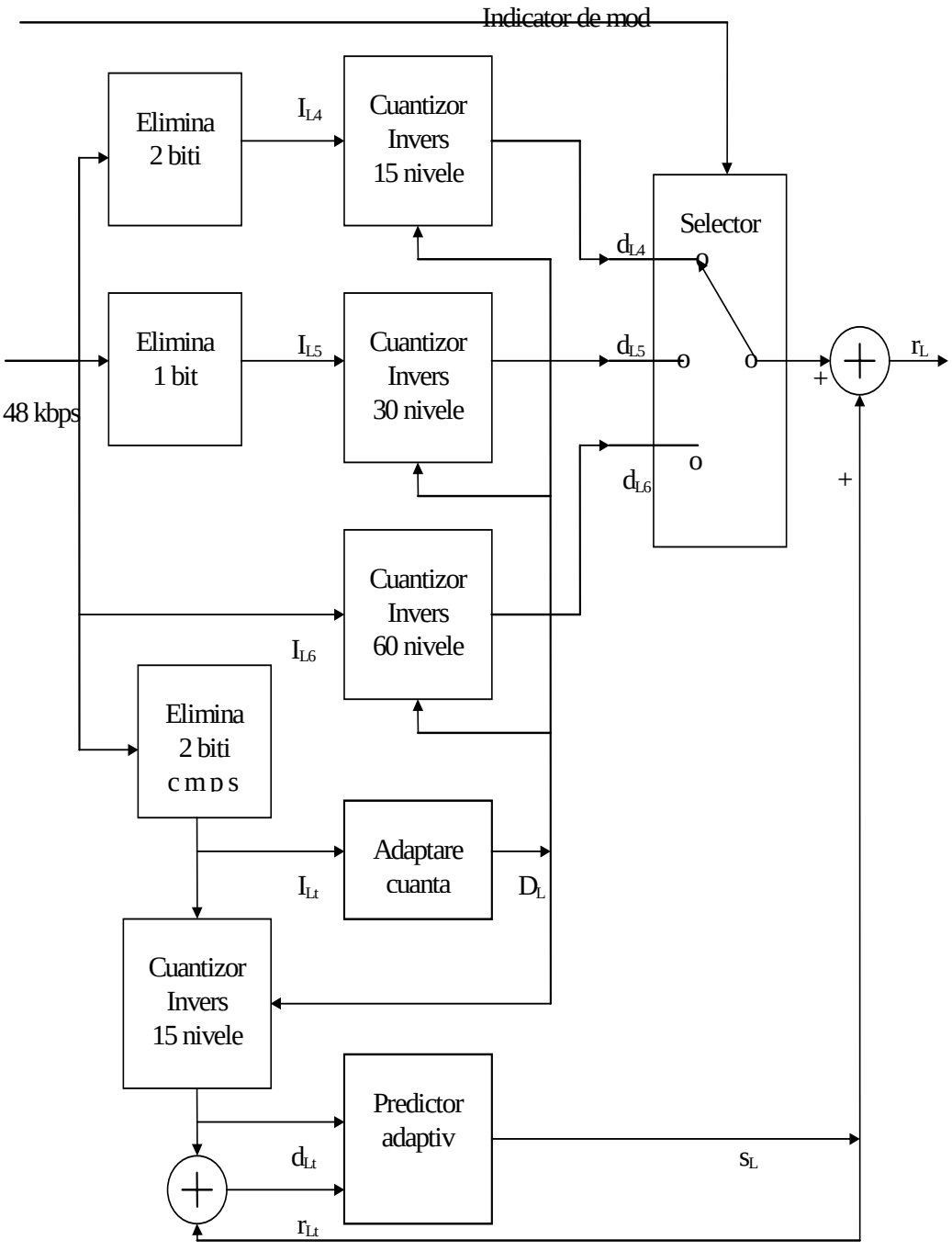


Figura 4.2.5. Decodorul pentru banda inferioara G.722

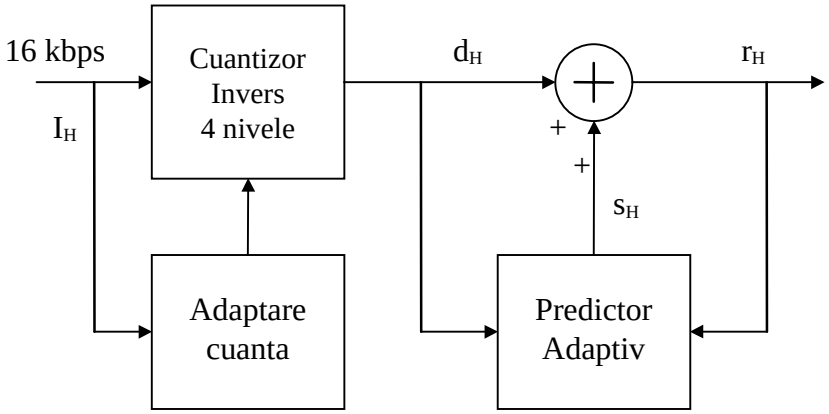


Figura 4.2.6. Decodor pentru banda superioara

Concluzii:

- *codarea ADPCM G.722 se aplica semnalelor audio de banda larga (banda de 7 KHz)*
- *pentru o codare cu raport semnal/zgomot ridicat, G.722 are in vedere codarea ADPCM in subbenzi. Semnalul de intrare cu banda maxima de 8 KHz este filtrat trece sus, respectiv trece jos pentru a obtine doua subbenzi: superioara si inferioara. Subbenzile sunt cuantizate pe 2 biti (subbanda superioara, acolo unde informattia acustica este mai putin relevanta), respectiv 6 biti (subbanda inferioara, care contine informatii importante despre frecventa fundamentala si frecventele formantilor)*
- *debitul sursei vocale fi de 64, 56 si 48 Kbps, in functie de numarul bitilor care se transmit pentru banda inferioara: 6, 5 sau 4. Astfel apare posibilitatea de a transmite pe linga canalul vocal un canal de date de 0, 8 sau 16 Kbps*
- *debitul de la iesirea codecului este de 64 Kbps si se obtine prin multiplexarea fluxului vocal cu fluxul de date*
- *calitatea semnalului audio reconstituit este acceptabila, pentru debitul de 48 Kbps si este buna, respectiv foarte buna pentru debitele de 56 si 64 Kbps.*

Multimedia Communications

Subband Coding

Subband coding

- Subband coding: decompose the input signal into different frequency bands
- After the input is decomposed to its constituents, we can use the coding technique best suited to each constituent to improve the compression performance
- Each component may have different perceptual characteristics
  - Quantization errors that are objectionable in one component may be acceptable in a different component

Subband Coding

- Idea: decompose a signal into components by applying frequency-selective filtering. Then select the best coding technique that best suits each component (subjectively and objectively).
- Example: slow- and fast-varying components.

$$y[n] = (x[n] + x[n-1])/2 \quad z[n] = (x[n] - x[n-1])/2$$

$$\text{The signal can be recovered: } x[n] = y[n] + z[n]$$

$$\text{The filters are: } h[n] = (\delta[n] + \delta[n-1])/2 \quad g[n] = (\delta[n] - \delta[n-1])/2$$

McMaster University

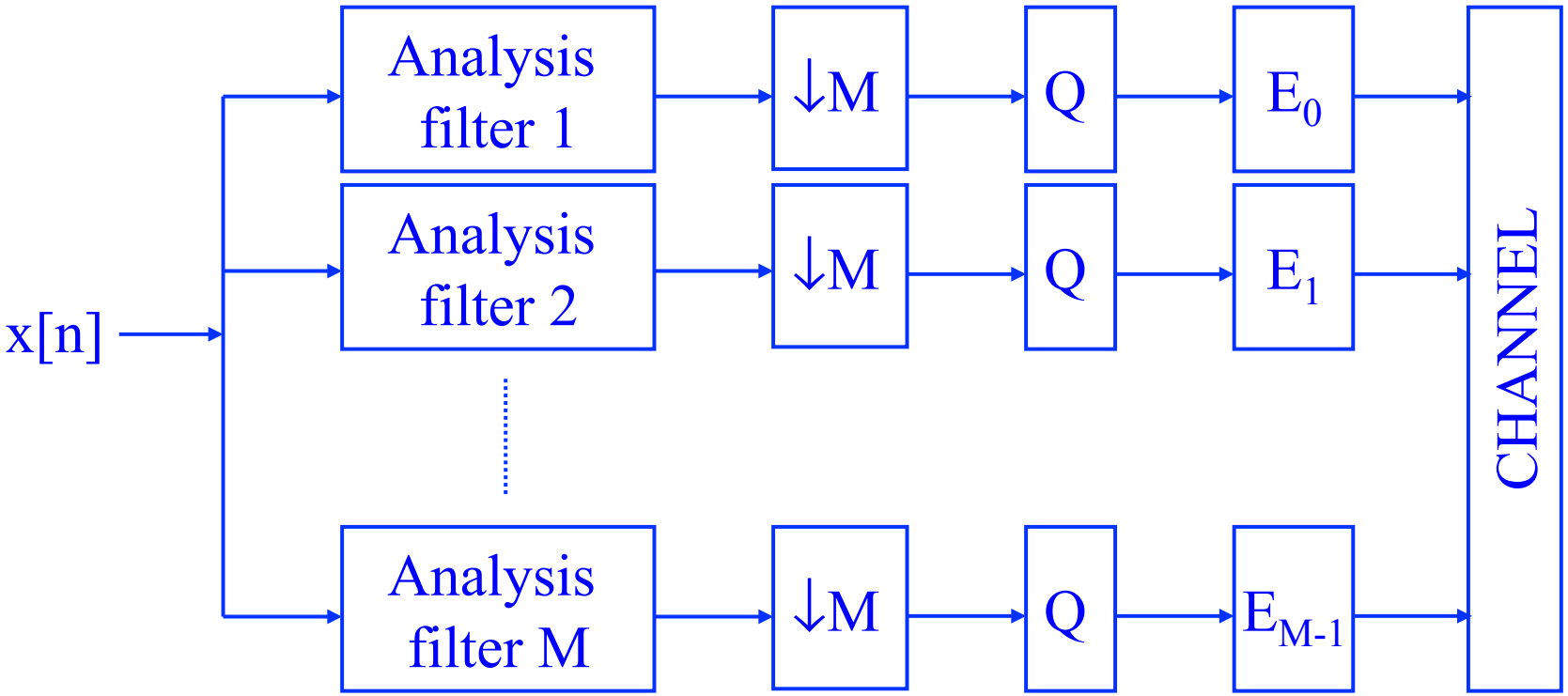
Subband Coding

- If we use the same number of bits for each of y[n] and z[n], we are transmitting twice as many samples, doubling the bit rate.
- We can avoid this by sending every other value of y[n] and z[n] (e.g., even numbered elements)

$$y[2n] = (x[2n] + x[2n-1])/2$$
$$z[2n] = (x[2n] - x[2n-1])/2$$

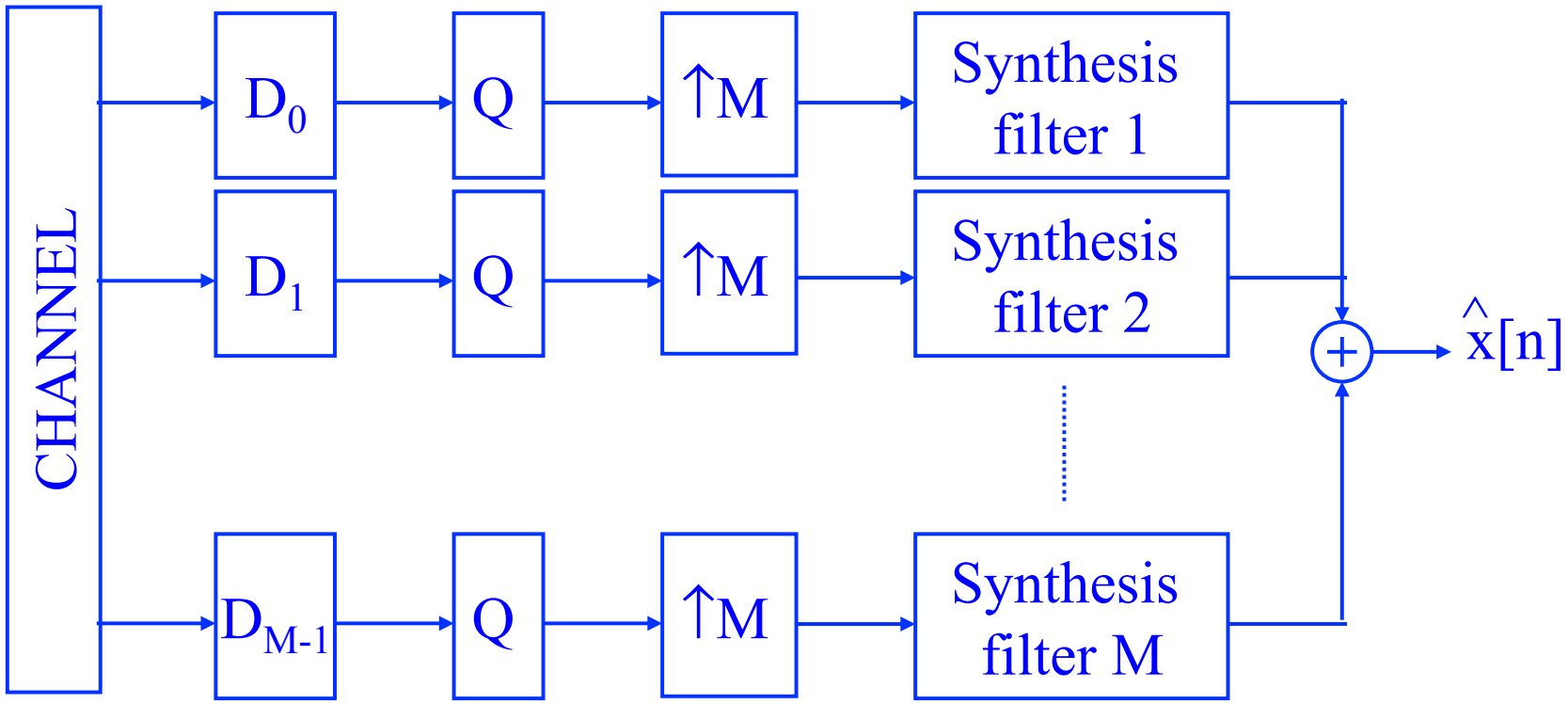
$$x[2n] = y[2n] + z[2n]$$
$$x[2n-1] = y[2n] - z[2n]$$

Subband Encoding



McMaster University

Subband Decoding



Subband Coding

Analysis

- Source output is passed through a bank of filters (analysis filters)
- Analysis filters cover the range of frequencies that make up source output
- Passband of the filters can be non-overlapping or overlapping
- Output of filters are then subsampled (also called decimation or downsampling)
- Justification for subsampling: Nyquist rule (range of frequencies of output of the filter is less than input to the filter)

Subband Coding

Quantization, coding and bit allocation

- Selection of compression scheme and allocation of bits between subbands is important and can have significant impact on the quality of the final reconstruction

Synthesis

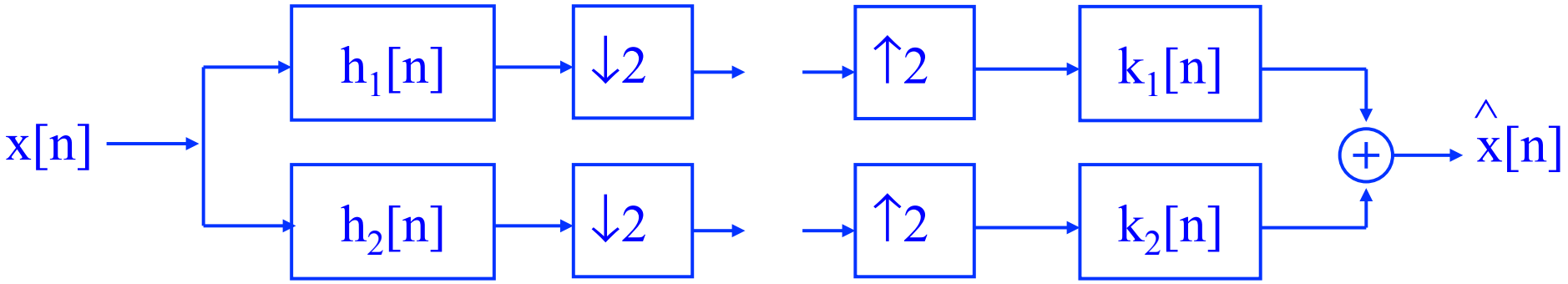
- Encoded samples from each subband are decoded
- Decoded values are then upsampled by inserting an appropriate number of 0s between samples
- Upsampled signals are passed through a bank of reconstruction filters
- Output of reconstruction filters are added to give final output

Subband Coding

- Three major components of subband system are:

1. Analysis and synthesis filters
   - Simple to implement, good separation between frequency bands
2. Bit allocation (quantization)
   - Can have s significant affect on the quality of the reconstruction
3. Encoding scheme
   - Based on the characteristics of each of the subbands, we can use a separate compression scheme
   - Human perception is frequency dependent. We can use this fact to design our compression scheme so that the frequency bands that are most important to perception are reconstructed most accurately

Filter Banks: Two-Band



- For M=2 the filters are easy to analyze.
- Goal:
  - good frequency-domain separation
  - no aliasing terms
  - perfect reconstruction: system is equivalent to a delay

McMaster University

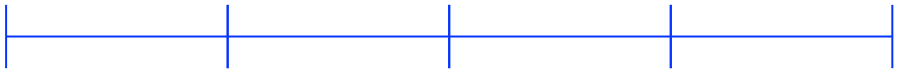
Filter Banks: Two-Band

- Quadrature Mirror Filters (QMF) solution
  - no aliasing, no phase distortion, some magnitude distortion
  - the filters are symmetric and
  - set of filters have been designed by Johnston
  - the decomposition efficiency increases with the length
- Conjugate Quadrature Filters (CQF, Smith-Barnwell) solution
  - perfect reconstruction
  - better frequency characteristics for the same nr. of taps
  - closely related to wavelets

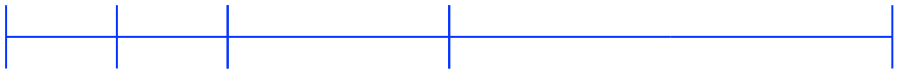
Filter Banks: Tree-Structured

- We can design an M-band filter bank by successively applying 2-band filter banks.

- Example: uniform filter bank decomposition



- Example: octave-band filter bank decomposition

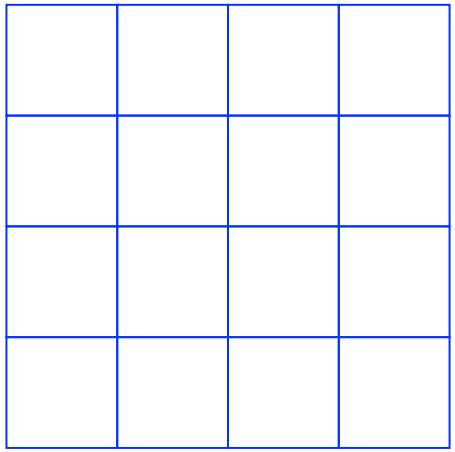


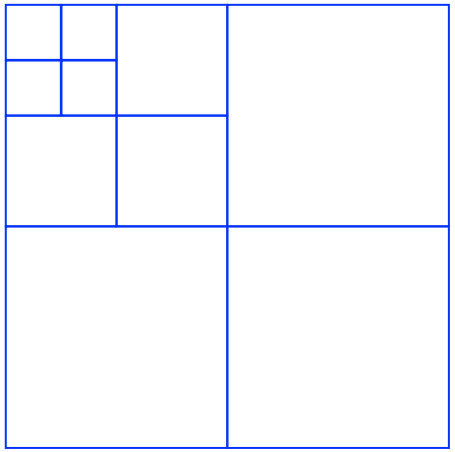
McMaster University

2-Dimensional Filter Banks

- Most 2-D filter banks are obtained by applying 1-D decompositions separably.

Uniform decomposition Octave-tree decomposition





McMaster University

Bit allocation

- Once we have separated the source into subbands, we need to decide how much of the coding resources should be used to encode each subband
- B_T: total bits to distribute among M subbands
- R: average rate in bits per sample for the overall system
- R_k: average rate for subband k
- We assume that we have the rate-distortion function for each band
- We want to find R_k such that $R = \frac{1}{M}\sum R_k$

and the reconstruction error is minimized.

McMaster University

Bit allocation

- Where on the rate distortion curve for each subband should we operate to minimize the average distortion?
- $J_k=D_k+\lambda R_k$
- D_k: distortion for the kth subband
- R_k: rate for the kth subband
- λ: Lagrangian parameter, specifies the tradeoff between rate and distortion
- Primary interested in minimizing the distortion: λ small
- Primary interested in minimizing the rate: λ large
- The value of D_k and R_k that minimize J_k occur where the slope of the rate-distortion curve is λ.

Bit allocation

- What should the value of λ be and how should it change between subbands?
- Fact: we would like to allocate bits in such a way that any increase in any rates in any subbands will have the same impact on the distortion
- Why: because if the above it not true we can take the bits off the subband whose rate reduction has less effect on the distortion and assign it to other subbands
- We pick R_k in such a way that the slope of the rate distortion functions for different subbands are the same

Bit allocation

- Given a set of rate-distortion functions and a value of λ, we can set the rates R_k, and compute the average rate.
- If it satisfies our constraint on the total rate we stop, otherwise we modify λ until we get a set of rates that satisfy our rate constraint
- Generally we do not have the rate-distortion function.
- We can use the operational rate-distortion curves.
- Operational: particular type of encoder operating on specific type of sources
  - Exp: pdf-optimized non-uniform quantizer with entropy coding
- If operational curve is available for a limited number of points we can estimate the other points or use curve fitting

G. 722

- ITU recommendation G. 722: a technique for wideband coding of speech based on subband coding
- Objective: high-quality speech at 64 kbps.
- Recommendation has two other modes that code the input at 56 and 48 kbps (to leave some bandwidth for auxiliary channel)
- Speech is first filtered to 7kHz to prevent aliasing then sampled at 16,000 samples per second.
- Each sample is encoded using a 14-bit uniform quantizer.
- This 14-bit input is passed through a bank of two 24-coefficient FIR filter.

G. 722

- Low-pass filter passes all frequency components in the range of 0 to 4 kHz.
- High-pass filter passes all remaining frequencies.
- The output of filters is downsampled by a factor of two.
- Downsampled sequences are encoded using adaptive differential PCM (ADPCM) system.
- ADPCM system that encodes the downsampled output of the low-frequency filter uses 6 bits per sample with the option of dropping 1 or 2 least significant bits (to provide room for auxiliary channels)
- Output of high-pass filter is encoded using 2 bits per sample.

MPEG audio

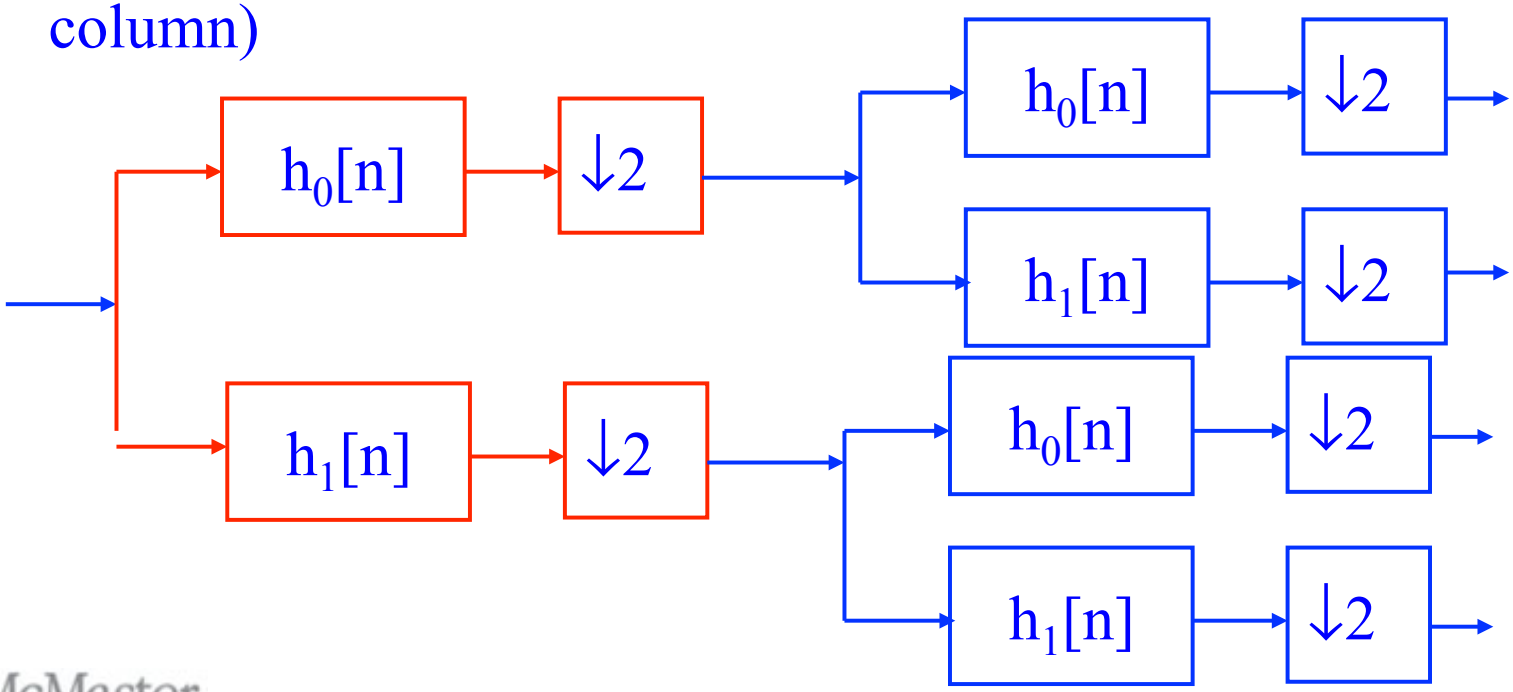
- MPEG has proposed an audio scheme that is based on subband coding.
- MPEG has proposed three coding schemes: Layer1, Layer2, Layer 3
- Coders are upward compatible: a layer N decoder is able to decode bitstream generated by the layer N-1 encoder
- Layer1 and 2 coders use a bank of 32 filters.
- Sampling frequencies are 32,000, 44,100, and 48,000.
- Each subband is quantized with a variable number of bits.
- The number of bits assigned to each subband is determined by a psycho-acoustic model that uses the masking property of the human ears.

MPEG audio

- If we have a large amplitude signal at one frequency it affects the audibility of signals at other frequencies.
- A loud signal at one frequency may make quantization at other frequencies inaudible.
- If we have a large signal in one of the subbands, we can tolerate more quantization error in the neighboring subbands and use fewer bits.

Image compression

- In most cases for subband coding of 2-D signals, we use separable filters.
- If the filters are separable, the 2-D filtering can be implemented as 2, 1-D filtering (filter each row and then each column)

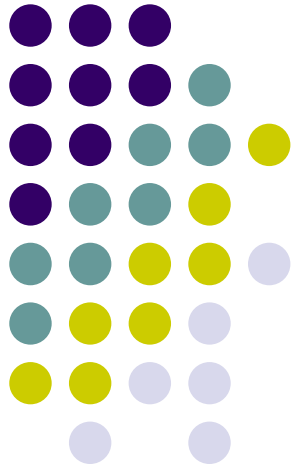


McMaster University

Image compression

- Question: when filtering the image pixels close to the borders what the past values of the signal are assumed to be?
  - Zero: not the best option
  - Reflect the values of pixels at the boundary: 6 9 5 4 7 2 is expanded to 9 6 6 9 5 4 7 2
- Once we decomposed an image into subbands, we need to find the best encoding scheme to use with each subband
- DPCM for the low-low band and scalar quantization for the other bands are common approaches.

Codarea sinusoidala a semnalului vocal



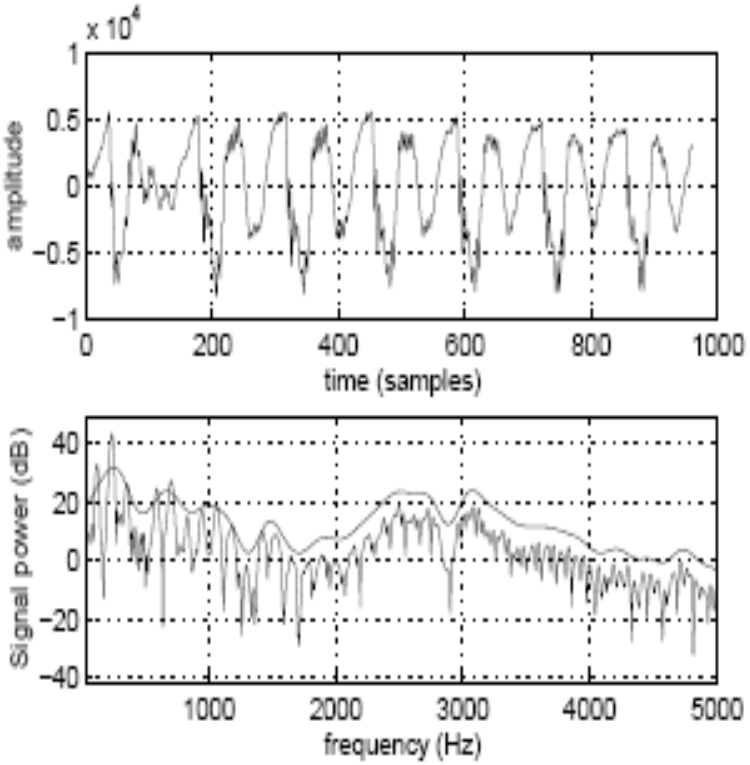
Codarea sinusoidala

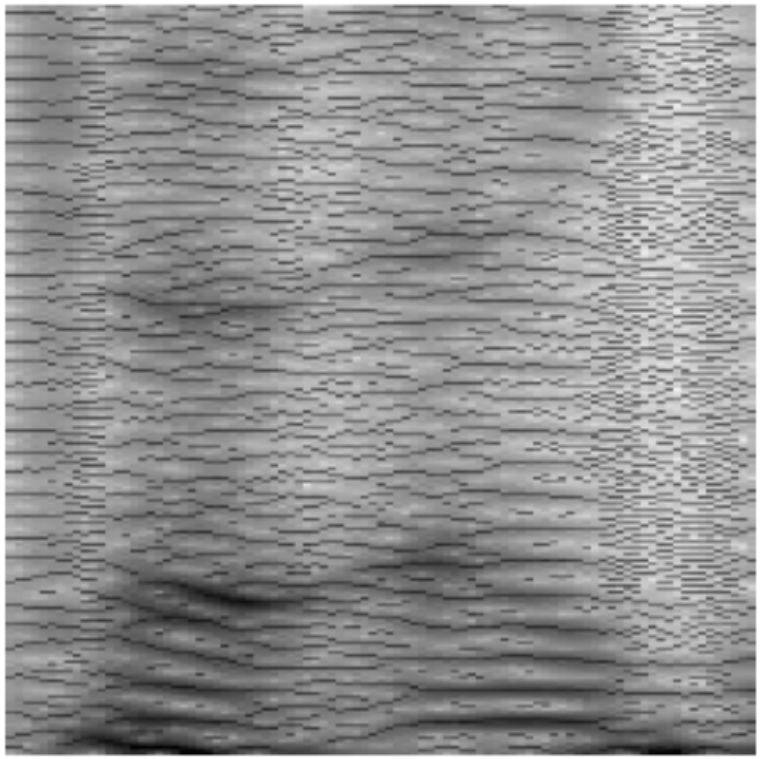
- Codarea sinusoidala urmareste generalizarea modelului excitatiei glotale, inloc sa utilizeze impulsuri sau secvente aleatoare, se presupune ca excitatia este compusa din componente sinusoidale de amplitudini, frecvente sau faze particulare.
- Functia excitatie este adesea reprezentata prin intermediul unui tren de impulsuri pe durata zonelor vocalizate, unde distanta dintre impulsuri este chiar “pitch”,si ca zgomot pe zone nevocalizate.Alternativa acestui model este inlocuirea cu o suma de sinusi. Motivatia reprezentarii sinusoidale este ca excitatia, unde este perfect periodica, poate fi inlocuita cu o serie de componente Fourier, unde fiecare componenta din serie corespunde unui sinus. Mai general, sinusii vor fi inlocuiti cu armonice care apar cand periodicitatea nu e exacta sau cand excitatie este nevocala. La trecerea formei sinusoidale ce reprezinta excitatia, prin tractul vocal rezulta o reprezentare sinusoidala pentru forma de unda a semnalului vocal data de :

$$s(n) = \sum_{\ell=1}^{L} A_\ell \cos(\omega_\ell n + \phi_\ell)$$

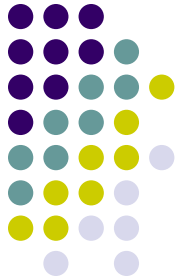
- Unde A_ℓ si ϕ_ℓ reprezinta ampltitudinea si faza pentru fiecare componenta sinusoidala asociata cu frecventa ω_ℓ si L este numarul de forme sinusoidale.

Codarea sinusoidala





Estimarea parametrilor vocali sinusoidali



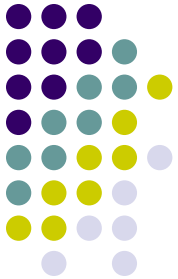
- Problema analizei si sintezei este de a lua forma de unda vocala, de a extrage parametrii ce reprezinta portiuni cvasistationare si utilizarea lor sau a variantei lor codate pentru a reconstrui o aproximare care sa fie cat mai aproape de forma originala. Daca forma de unda vocala este reprezentata de un numar arbitrar de sinusi, problema estimarii parametrilor, desi usor de rezolvat, duce la rezultate care nu ai nici o importanta fizica. In consecinta, abordarea se bazeaza pe observarea ca atunci cand semnalul vocal este perfect periodic, parametrii sinusilor corespund Transformatei Fourier pe termen scurt. In acest caz avem:

$$s(n) = \sum_{\ell=1}^{L} A_\ell cos(n\ell\omega_0 + \phi_\ell) \quad (2)$$

- In care frecventele sinusoidelor sunt multipli ai frecventei fundamentale, iar amplitudinile si fazele sunt date de STFT. Daca STFT este data de:

$$S(\omega) = \sum_{n=-N/2}^{N/2} s(n) exp(-jn\omega) \quad (3)$$

Estimarea parametrilor vocali sinusoidali



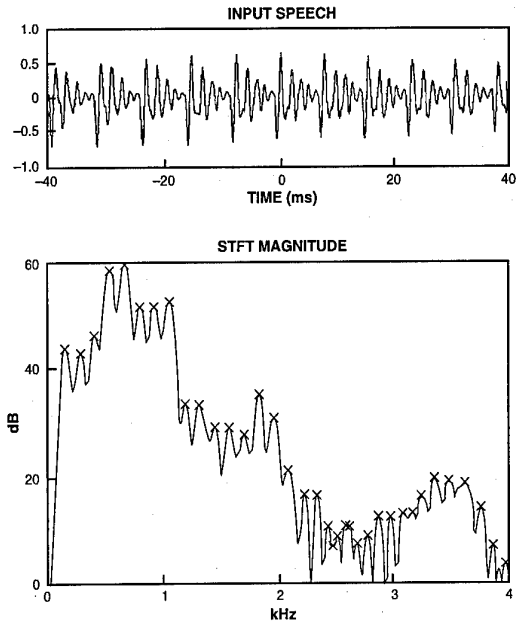
- Atunci analiza Fourier da urmatorii parametrii estimati :
- amplitudinea: $A_\ell = |S(\ell\omega_0)|$ (4)
- faza: $\phi_\ell = arg S(\ell\omega_0)$ (5)

- Magnitudinea STFT va avea valori la multiplii de ω_0.
- Cand semnalul vocal nu este perfect vocal, vor exista valori insa nu neaparat la armonice. In acest caz faza formei de unda sinusoidale va fi calculata pentru partea reala si imaginara a STFT.
- Daca se utilizeaza o fereastra de analiza de tip Hamming, atunci o data calculata latimea ferestrei pentru un cadru particular, avem:

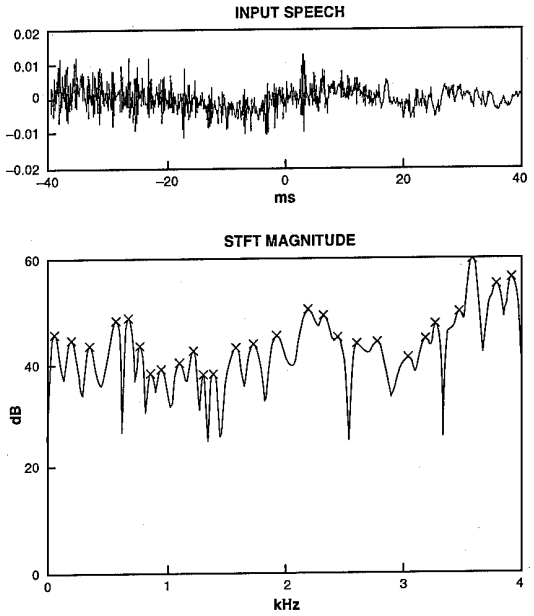
$$\sum_{n=-N/2}^{N/2} w(n) = 1 \quad (6)$$

Estimarea parametrilor vocali sinusoidali

- fereastra adaptata Hamming



Exemplu pentru forma de unda vocalizata



Exemplu pentru forma de unda nevocalizata

Sinteza

- Daca amplitudinea, frecventa si faza care sunt estimate pentru un cadru k sunt notate cu : $(A_\ell^k, \omega_\ell^k, \phi_\ell^k)$, atunci semnalul sintetizat pentru cadru poate fi calculat astfel:

$$\hat{s}^k(n) = \sum_{\ell=1}^{L^k} A_\ell^k cos[n\omega_\ell^k + \theta_\ell^k] \qquad (7)$$

- Forma de unda sintetizata se obtine aplicand ecuatia de mai sus pentru cadrele k-1 si k pentru a genera , $\hat{s}^{k-1}(n)$ respectiv $\hat{s}^k(n)$ care sunt ponderate … si adunate:

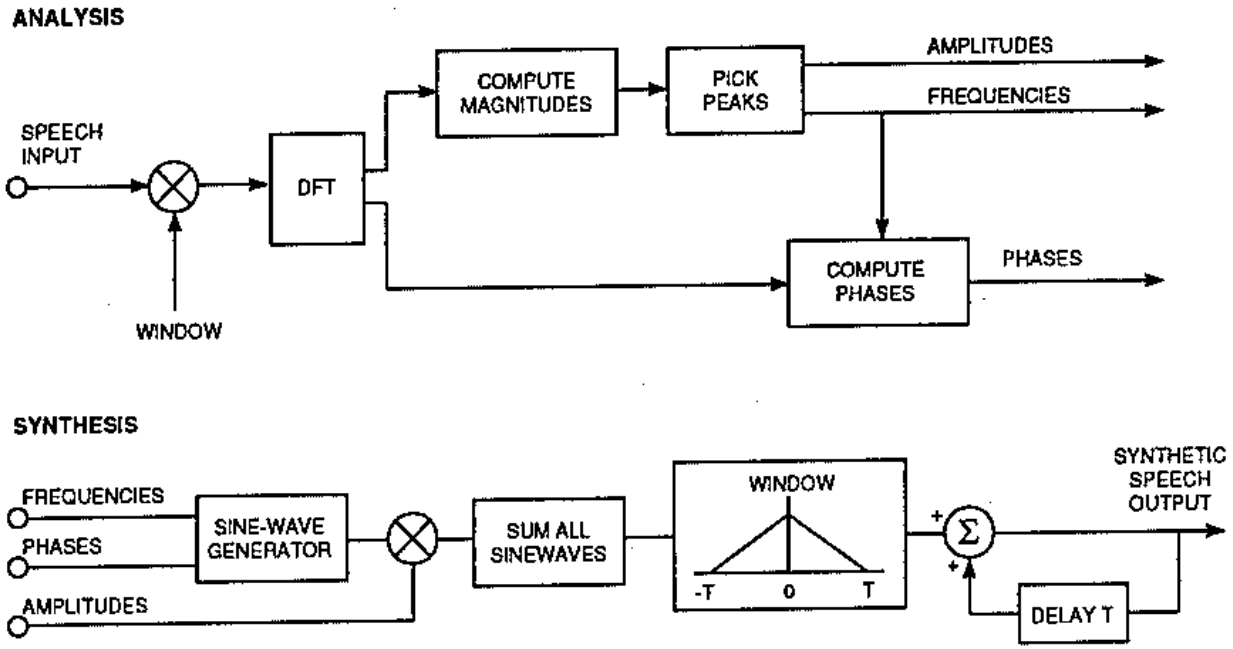
$$\hat{s}(n) = w_s(n)\hat{s}^{k-1}(n) + w_s(n-T)\hat{s}^k(n-T) \quad (8)$$

- Unde $w_s(n)$ e fereastra de sinteza pentru .. si adunare si respecta formula:

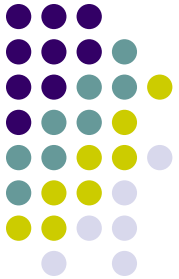
$$w_s(n) + w_s(n-T) = 1 \quad (9)$$

Sinteza

- **O posibila schema bloc pentru analiza si sinteza semnalului vocal prin analiza sinusoidala este:**



Estimarea parametrilor pentru modelul sinusoidal armonic



- Primul pas in procedura de analiza, este de a presupune ca frame-ul de intrare din semnalul vocal a fost deja analizat in termeni de componente sinusoidale, cu tehnica descrisa mai sus. Atunci s(n) este:

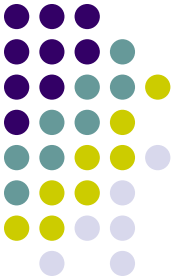
$$s(n) = \sum_{\ell=1}^{L} A_\ell exp[j(n\omega_\ell + \theta_\ell)] \quad (10)$$

Unde, $\{A_\ell, \omega_\ell, \theta_\ell\}_{\ell=1}^{L}$, reprezinta amplitudinea, frecventa si faza celor L sinusi masurati. Scopul este de a incerca sa reprezentam acesti sinusi cu o alta forma de unda pentru care toate frecventele sa fie armonice. Aceasta forma de unda poate fi modelata astfel:

$$\hat{s}(n;\omega_0,\phi) = \sum_{k=1}^{K(\omega_0)} \bar{A}(k\omega_0) exp[j(nk\omega_0 + \phi_k)] \quad (11)$$

Unde: $\omega_0 = 2\pi f_0 / f_s$ este frecventa fundamentala normalizata, $K(\omega_0)$ este numarul de armonici din banda semnalului vocal, $\bar{A}(\omega)$e anvelopa tractului vocal, si $\phi = (\phi_1, \phi_2, \cdots, \phi_{K(\omega_0)})$sunt fazele armonicelor. De acum inainte ω_0 va fi numita “pitch”, desi in zone de semnal nevocal terminologia nu are nici o semnificatie. Este de dorit sa se estimeze frecventa fundamentala si fazele, astfel incat semnalul $\hat{s}(n)$ sa fie cat mai aproape de cel real $s(n)$.

Estimarea parametrilor pentru modelul sinusoidal armonic



- Un criteriu de estimare este cautarea minimelor in eroarea medie patratica:

$$\epsilon(\omega_0, \phi) = \frac{1}{N+1} \sum_{n=-N/2}^{N/2} |s(n) - \hat{s}(n; \omega_0, \phi)|^2 \qquad (12)$$

- care se poate rescrie astfel:

$$\epsilon(\omega_0, \phi) = \frac{1}{N+1} \sum_{n=-N/2}^{N/2} \{|s(n)|^2 - 2Re[s(n)\hat{s}^*(n; \omega_0, \phi)] + |\hat{s}(n; \omega_0, \phi)|^2\} \qquad (13)$$

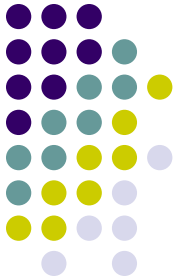
- Primul termen din formula de mai sus repezinta puterea semnalului si este independent de parametrii necunoscuti:

$$P_s = \frac{1}{N+1} \sum_{n=-N/2}^{N/2} |s(n)|^2 \qquad (14)$$

- Inlocuind ecuatia (11) in cel de al doilea termen al ecuatiei (13) avem relatia:

$$\sum_{n=-N/2}^{N/2} s(n)\hat{s}^*(n; \omega_0, \phi) = \sum_{k=1}^{K(\omega_0)} \bar{A}(k\omega_0) exp(-j\phi_k) \sum_{n=-N/2}^{N/2} s(n) exp(-jnk\omega_0) \qquad (15)$$

Estimarea parametrilor pentru modelul sinusoidal armonic



- Inlocuind ecuatia (11) in cel de al treilea termen al ecuatiei (13) avem relatia:

$$\frac{1}{N+1}\sum_{n=-N/2}^{N/2}|\hat{s}(n;\omega_0,\phi)|^2 \simeq \sum_{k=1}^{K(\omega_0)}\bar{A}^2(k\omega_0) \quad (15)$$

- Unde aproximarea este valida daca fereastra de analiza satisface conditia:

$$(N+1) \gg 2\pi/\omega_0$$

- Aceasta conditie presupune ca perioada fundamentala, pitch, a fost deja calculata.

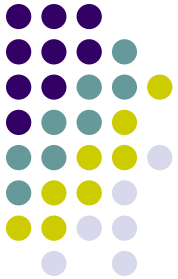
- Daca

$$S(\omega) = \frac{1}{N+1}\sum_{n=-N/2}^{N/2} s(n)exp(-jn\omega) \quad (17)$$

este STFT a semnalului de intrare si utilizand ecuatia (16), atunci expresia pentru eroarea medie patratica din ecuatia (13) devine:

$$\epsilon(\omega_0,\phi) = P_s - 2Re\{\sum_{k=1}^{K(\omega_0)}\bar{A}(k\omega_0)exp(-j\phi_k)S(k\omega_0)\} + \sum_{k=1}^{K(\omega_0)}\bar{A}^2(k\omega_0) \quad (18)$$

Estimarea parametrilor pentru modelul sinusoidal armonic



- Intrucat fazele afecteaza numai al doilea termen al ecuatiei (18), eroarea medie de predictie va fi minimizata alegand:

$$\hat{\phi}_k = arg[S(k\omega_0)] \quad (19)$$

- Si minimizarea va fi :

$$\epsilon(\omega_0) = P_s - 2\sum_{k=1}^{K(\omega_0)} \bar{A}(k\omega_0)|S(k\omega_0)| + \sum_{k=1}^{K(\omega_0)} \bar{A}^2(k\omega_0) \quad (20)$$

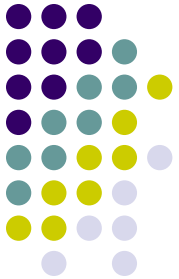
- Necunoasterea perioadei fundamentale afecteaza numai al doilea si al treilea termen din ecuatia de mai sus care poate fi adusa la forma redusa:

$$\rho(\omega_0) = \sum_{k=1}^{K(\omega_0)} \bar{A}(k\omega_0)[|S(k\omega_0)| - \frac{1}{2}\bar{A}(k\omega_0)] \quad (21)$$

- Si eroarea medie patratica poate fi exprimata prin:

$$\epsilon(\omega_0) = P_s - 2\rho(\omega_0) \quad (22)$$

Estimarea parametrilor pentru modelul sinusoidal armonic



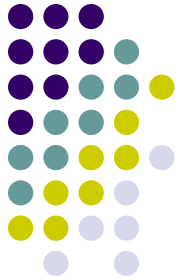
- Intrucat primul termen este o valoare constanta cunoscuta valoarea minima a erorii medii patratice se obtine prin maximizarea lui $\rho(\omega_0)$.
- Este util ca mai departe sa se utilizeze reprezentarea sinusoidala a semnalului vocal de intrare. Substituind reprezentarea din ecuatia (10) in ecuatia (14), atunci puterea devine:

$$P_s = \sum_{\ell=1}^{L} A_\ell^2 \quad (23)$$

- Iar, ecuatia (17) a STFT este: $S(\omega) = \sum_{\ell=1}^{L} A_\ell exp(j\theta_\ell)\ sinc(\omega_\ell - \omega)$ (24)

- unde $sinc(x) = \frac{1}{N+1} \sum_{n=-N/2}^{N/2} exp(jnx) = \frac{sin[(N+1)\frac{x}{2}]}{(N+1)sin(\frac{x}{2})}$ (25)
- Amplitudinea lui STFT poate fi aproximata prin : $|S(\omega)| \approx \sum_{\ell=1}^{L} A_\ell D(\omega_\ell - \omega)$ (26)
- Unde $D(x) = \begin{cases} sinc(x) & \text{if } |x| \le \frac{2\pi}{N+1} \\ 0 & otherwise \end{cases}$ (27)

Estimarea parametrilor pentru modelul sinusoidal armonic



- Criteriul de optimizare devine:

$$\rho(\omega_0) = \sum_{k=1}^{K(\omega_0)} \bar{A}(k\omega_0)[\sum_{\ell=1}^{L} A_\ell D(\omega_\ell - k\omega_0) - \frac{1}{2}\bar{A}(k\omega_0)] \quad (28)$$

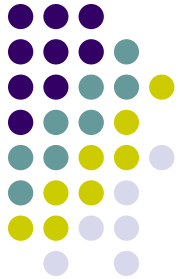
- Pentru a intelege mai bine semnificatia acestui criteriu, se presupune ca intrarea vocii este periodica avand frecventa pitch ω^*. Atunci $\omega_\ell = \ell\omega^*, \bar{A}_\ell = \bar{A}(\ell\omega^*)$ si : $\rho(\omega^*) = \frac{1}{2}\sum_{k=1}^{K(\omega^*)}[\bar{A}(k\omega^*)]^2$ (29)

  cand ω_0 corespunde submultimilor perioadei fundamentale, primul termen din ecuatia (27) ramane neschimbat, deoarece $D(\omega_\ell - k\omega_0) = 0$ pentru toti submultiplii. Dar termenul al doilea, deoarece este anvelopa si este tot timpul diferit de 0 va creste pentru submultiplii lui ω^* .Ca si urmare:

$$\rho(\frac{\omega^*}{m}) < \rho(\omega^*) \quad m = 2, 3, \cdots \quad (30)$$

- Ceea ce inseamna ca criteriul de optimizare duce la estimarea unui pitch clar.

Rezolutia perioadei fundamentala – adaptive



- In formularea de mai sus fereastra de analiza e fixata la N+1 esantioane. Aceasta inseamna ca lobul principal al functiei sinc este fixa pentru toate valorile lui pitch. Aceasta este contrar faptului ca urechea este mai putin toleranta la diferente mari pentru domenii de frecvente inalte, fata de frecventele joase. Acest efect poate fi pus in evidenta prin definirea functiei distanta $D(x)$ pentru lobul k armonic sa fie:

$$D(\omega - k\omega_0) = \frac{sin[2\pi(\frac{\omega - k\omega_0}{\omega_0})]}{2\pi(\frac{\omega - k\omega_0}{\omega_0})} \quad \text{for all } |\omega - k\omega_0| \leq \frac{\omega_0}{2} \quad (31)$$

- Si sa fie zero in rest. In acest fel rezolutia este foarte mare la valori mici a lui pitch, in contrast cu valorle mari unde rezolutia este mica.

5. CODAREA SEMNALELOR AUDIO IN STANDARDUL MPEG PENTRU APLICATII MULTIMEDIA

Obiective:

- ***cunoasterea caracteristicilor principalelor sisteme de codare a semnalelor audio de inalta calitate***
- ***intelegerea tipurilor de mascare a componentelor spectrale si a caracteristicilor psiho-acustice utilizate in codarea MPEG***
- ***prezentarea sistemelor de codare MPEG-1 si MPEG-2 la nivel de scheme bloc***
- ***cunoasterea succesiunii prelucrarilor semnalului vocal in MPEG-1 si MPEG-2***

5.1. Citeva standarde ITU - ISO G.xxx pentru codarea semnalelor audio

Cele mai importante standarde ITU - T din grupul G.xxx pentru codarea semnalului vocal sunt redate in tabelul de mai jos.

Tabelul 5.1.1. Standarde din grupul G.xxx pentru codarea semnalului vocal

Denumire	Banda semnal [Hz]	Frecventa esantionare [KHz]	Debit [Kbps]
G.711	200 - 3200	8	64
G.722	50 - 7000	16	64
G.721	200 - 3200	8	32
G.726	200 - 3200	8	16
G.723	200 - 3200	8	5.3 si 6.3

Tabelul 5.1.2. Caracteristici ale principalelor sisteme de codare a semnalelor audio de inalta calitate

Mediu de stocare	Banda semnal [Hz]	Frecventa esantionare [KHz]	Debit binar stereo [Kbps]	Codare
CD	20	44.1	1411.2	PCM
DAT	16	32.2	1024.0	PCM
ProDAT	16	44.1	1411.2	PCM
DAT	16	48.0	1536.0	PCM
DCC	20	44.1	384	PASC
MD	22	44.1	292	ATRAC

Prescurtarile reprezinta :

CD - Compact Disc ,
DAT - Digital Audio Tape ,
DCC - Digital Compact Cassette,
MD - Mini Disc.

De remarcat ca pentru inregistrarea pe suport magnetic debitul binar creste de 2-3 ori prin adaugarea de informatie suplimentara pentru corectia erorilor.

5.2 Caracteristici psihoacustice aplicate in codarea MPEG a semnalelor audio

Standardul **MPEG** (Motion Picture Expert Group) foloseste in procesul de codare a semnalelor audio caracteristici psihoacustice de audibilitate determinate in laborator pe un mare numar de subiecti. Caracteristica principala este ca in spectrul semnalelor audio exista componente care nu sunt percepute de urechea umana. De exemplu, o componenta spectrala de amplitudine mare face ne-audibile componentele de frecventa apropiata, care au amplitudini mai mici ca un anumit prag. Acest efect se numeste **efect de mascare**.

Efectul de mascare se manifesta in trei ipostaze:

- mascarea componentelor spectrale
- mascarea temporala
- pragul absolut de mascare .

a) Mascarea componentelor spectrale (MCS) se produce atunci cind componentele spectrale din jurul frecventei cu amplitudine dominanta nu depasesc un anumit prag de mascare. Pragul de mascare depinde de frecventa, amplitudinea si durata in timp a tonului dominant .

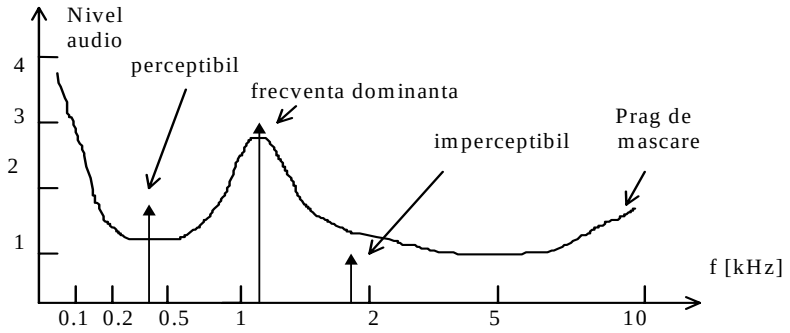


Figura 5.2.1. Pragul de mascare si efectul asupra componentelor spectrale

b) Mascarea temporala (MT) consta in persistenta efectului de mascare o anumita perioada Δt (in general de ordinul a 200-500 ms) chiar si dupa disparitia brusca a frecventei dominante. In intervalul Δt pragul de mascare scade. Practic, frecventa dominanta nu scade brusc, ci in timp si in consecinta pragul de mascare scade pina la o anumita valoare.

c) Pragul absolut de mascare (PAM) este valoarea minima a amplitudinii, pentru care componentele spectrale sunt audibile

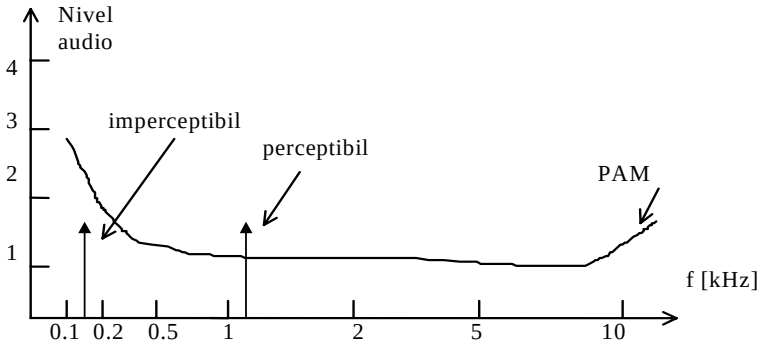


Figura 5.2.2. Pragul absolut de mascare

De remarcat ca in spectrul semnalelor audio exista o multitudine de frecvente dominante si ca atare pragul de mascare global, la un moment dat, se va determina in functie de pragul de mascare, mascarea temporala si pragul absolut de mascare, pentru fiecare componenta in parte. Din Figura 5.2.3 se observa ca pragul de mascare global depinde de amplitudinile si frecventele componentelor spectrale dominante .

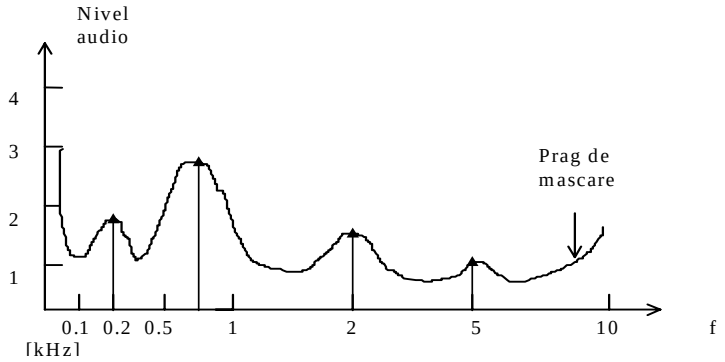


Figura 5.2.3. Pragul de mascare global

EXERCITIU:

Ce standarde G.xxx prevad codarea semnalului vocal la un debit sub 32 Kbps?

EXERCITIU:

Care este debitul binar stereo pentru DAT, ProDAT si MD?

EXERCITIU:

Ce este efectul de mascare si cum se manifesta?
Explicati notiunea de prag absolut de mascare

5.3 Sistemul de codare/decodare MPEG-1

In prima etapa a implementarii MPEG-1 (ISO / IEC 11172-3) realiza codarea a doua canale stereo la debitul de 256 Kbps si la o calitate apropiata de CD. Standardul nu precizeaza codorul ci doar decodorul cu scopul de a asigura compatibilitatea cu dezvoltarile tehnologice viitoare, dar prevede 3 nivele (Layers) de codare: Layer I, Layer II si Layer III asociate cu calitatea si debitul binar dorit, nivele intelese ca trei algoritmi distincti.

Caracteristici ale codarii MPEG-1 :

- semnalul de intrare poate fi :
  - 1 canal audio mono
  - 1 canal audio stereo (in scopul reducerii debitului, se pot coda simultan mai multe canale individuale exploatind corelatia dintre ele)
  - 2 canale independente (eventual bilingv)
- frecventa de esantionare poate fi selectata la 32, 44.1 sau 48 kHz
- debitul binar

  - fix si precizat discret in intervalul 32-448 Kbps, in functie de nivel (I, II sau III)
  - fix, dar neprecizat in intervalul 32-448 Kbps
  - pseudovariabil (valabil doar pentru Layer III)

5.3.1 Structura codorului si decodorului MPEG-1

In Figura 5.3.1. se prezinta schema bloc a unui sistem de codare / decodare MPEG-1, nivelele I si II. Blocurile componente impreuna cu functiile realizate sunt urmatoarele :

- **Bancul de 32 de filtre trece banda** genereaza 32 de canale spectrale (c_i ; i = 1..32) uniform distribuite pe tot spectrul semnalului de intrare. Largimea de banda a unui filtru este f_e / 2 x 32, unde f_e este frecventa de esantionare a semnalului original.

- **Subesantionarea** cu 32 are rolul de a reduce frecventa de esantionare pe fiecare canal c_i in concordanta cu teorema esantionarii. La o banda de f_e /2 x 32, noua frecventa de esantionare va fi f_e / 32.

- Secvente de cite 12 esantioane consecutive din fiecare subbanda subesantionta, cs_I sunt **grupate in blocuri B_i** , care corespund la 12 x 32 = 384 esantioane de intrare.

- Pe baza unei tabele cu 63 de valori blocul de scalare determina pentru fiecare subbanda **factorii de scalare FS_i** (subunitari), astfel incit toate esantioanele din cadrul unui bloc sunt normalizate la valoarea maxima.

- In blocurile de cuantizare si codare se produce **codarea fiecarei subbenzi** c_I , pe un numar variabil de biti R_i, in concordanta cu importanta sa in spectru. Numarul nivelelor de cuantizare este determinat de *modelul psihoacustic* .

- **Obiectivul modelului psihoacustic** este estimarea unui parametru numit *raport semnal - prag de mascare* (RSPM [dB]) pe baza caruia sa se poata determina numarul nivelelor de cuantizare necesare in codare. Desi standardul nu precizeaza algoritmul de lucru al acestui bloc, un exemplu poate fi urmarit in Figura 5.3.2.

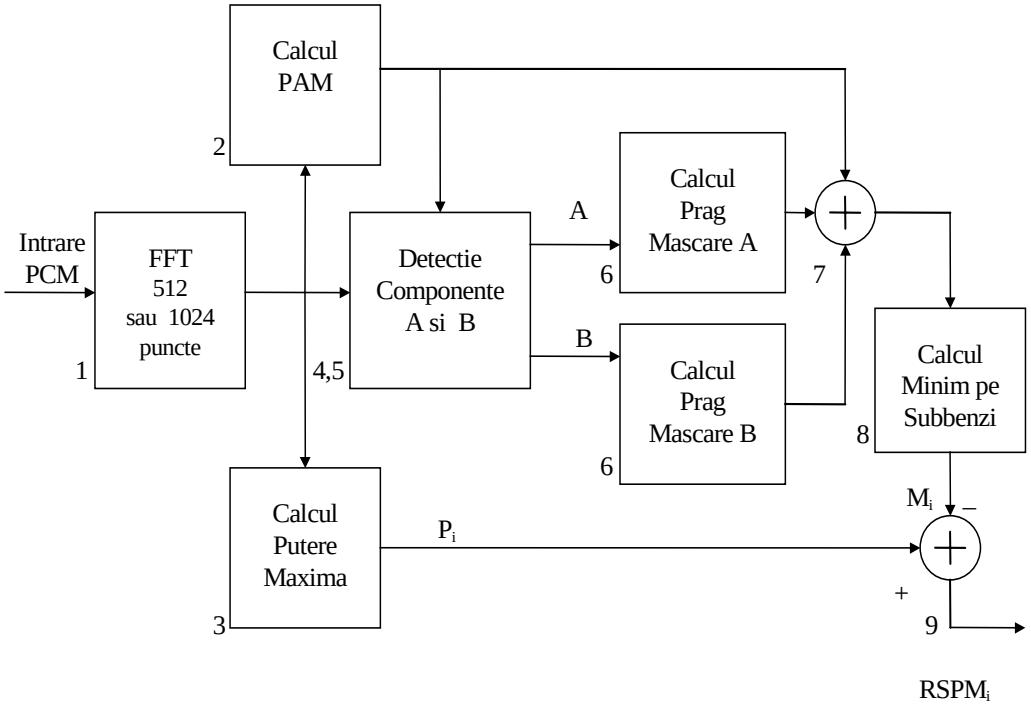


Figura 5.3.2. Model psihoacustic

Algoritmul prezentat in Figura 5.3.2. are 9 pasi:

P1: Pentru fiecare grup de 384 esantioane de intrare (ce corespunde unui bloc B_i) se calculeaza densitatea spectrala prin intermediul FFT in 512 (Nivel I) sau 1024 (Nivel II) puncte, cu fereastra Hamming. Rezolutia spectrala este de 8 puncte (Nivel I), respectiv 16 puncte (Nivel II) pentru fiecare subbanda.

P2 : Pe baza spectrului se calculeaza puterea maxima P_i

P3 : Se determina, pe baza de tabele, Pragul Absolut de Mascare (PAM) pentru fiecare componenta din FFT (valorile depind de calitatea codarii: pentru debite mai mari de 96 Kbps PAM scade cu aproximativ 12 dB).

P4: Separarea componentelor A si B din spectru. Componentele de tip A si B se caracterizeaza prin maxime locale in vecinatatea carora se gasesc (intr-o banda predefinita) una-doua componente spectrale semnificative. Benzile predefinite sunt mai inguste decit subbenzile, pentru frecvente joase si mai largi decit subbenzile pentru frecvente inalte. Prin eliminarea din spectru a componentelor de tip A, ramin componentele de tip B. Acestea se caracterizeaza prin componenta mai semnificativa din fiecare banda a unui banc de filtre neuniform benzi critice).

P5 : Eliminarea componentelor A si B nerelevante. Exista doua etape:
- se elimina atit componentele A cit si B care sunt sub PAM

- pentru doua componente de tip A apropiate in frecventa (Δf) se elimina componenta de amplitudine mai mica .

P6: Se calculeaza, pe considerente experimentale, pragul de mascare pentru componentele A si B ca o functie de amplitudine si frecventa. Se observa ca pentru o aceeasi amplitudine a componentelor A si B, efectul de mascare este mai pronuntat pentru componentele de tip A.

P7 : Se calculeaza pragul de mascare global, ca o suma a pragurilor de mascare absolute, de tip A, respectiv B.

P8 : Pentru fiecare subbanda, corespunzatoare bancului de 32 de filtre, se determina minimul M_i din functia prag de mascare .

P9 : Caculul raportului semnal-prag de mascare pentru fiecare subbanda i : $RSPM_i$ [dB] = P_i [dB] - M_i [dB]
Acest parametru este o masura a gradului in care se produce mascarea in fiecare subbanda si ca atare poate fi folosit in a aloca un numar mai mare de biti la codare, pentru subbenzile mai importante (P_i - M_i , mare), sau un numar mai mare de biti pentru subbenzile nerelevante (P_i -M_i , mic).

- In conditiile de mai sus este posibila o alocare dinamica si adaptiva a numarului de biti de cod, R_i , pentru subbenzile componente. Alocarea este diferita pentru Nivel I si Nivel II.
- Datele obtinute de la sistemul de codare formeaza un cadru, care este precedat de un header de 32 de biti.

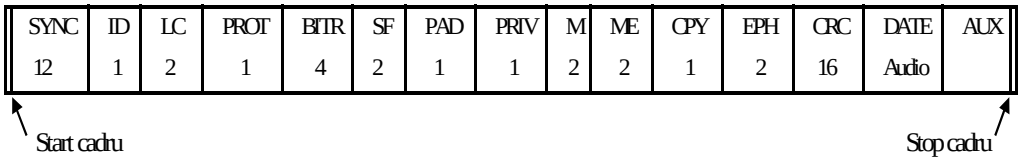


Figura 5.3.3. Structura cadrului MPEG-1

Semnificatia cimpurilor de biti:
SYNC - Syncword - 12 biti toti ‘1’ pentru sincronizare
ID - defineste algoritmul: ID = 1 (MPEG-1), ID = 0 (MPEG-2)
LC - Layer Code ;
PROT - Protection Bit; PROT = 0 (CRC este prezent) ; PROT = 1 (nu exista CRC)
BITR - Bit Rate Index - indica debitul binar
SF - Sampling Frequency, indica frecventa de esantionare ; depinde de ID
PAD - Padding Bit (PAD = 1, arata ca in cadru se gasesc biti suplimentari pentru a controla debitul binar)
PRIV - Private Bit, pentru uz privat
M - Mode, conform cu tabelul de mai jos :

Tabelul 5.3.1. Moduri de lucru codate de bitul M din cadrl MPEG-1

Mode	Tipuri de semnale de intrare / iesire
00	Stereo
01	Intensity-Stereo (Layer I , II) sau MS-Stereo (Layer III)
10	Canale independente
11	1 canal mono

Trebuie retinut ca in **modul stereo** canalele fiind codate separat este posibil ca pentru anumite cadre calitatea redarii sa fie slaba pentru un anumit debit dat. Pentru a elimina acest neajuns, modul Intesity-Stereo exploateza caracteristica prin care la frecvente medii si inalte urechea este mai sensibila la diferenta de amplitudine de pe cele doua canale decit la continutul spectral.

Astfel, pentru a folosi eficient numarul de biti alocati unui cadru, la frecvente medii si inalte se aduna semnalele pentru a transmite un singur semnal. Numarul subbenzii de la care se face adunarea este specificat prin variabila *mode extension* din header. Reconstituirea efectului stereo are loc prin codarea si transmiterea a doi factori de scala pe fiecare canal. La receptie are loc ponderarea semnalului audio cu acesti factori, generind doua amplitudini care sunt controlate separat. Informatia specifica modului *Intensity-Stereo* se include in fluxul de date audio in primii doi octeti .

ME - Mode extension; utilizat pentru M = 01 pentru a indica benzile codate cu Intensity-Stereo sau normal-Stereo (Nivelele I si II) sau ce tip de codare stereo se aplica (la Nivel III).
CPY - Copyright (1 - protejat la copiere)
O /C - Original / Copy (1 - indica originalul)
EPH - Emphasis , indica tipul filtrului de accentuare
CRC - Cyclic Redundancy Code, calculat pentru header in scopul de a detecta eventuale erori
AUX- Auxilliary Bit, biti suplimentari

5.3.2 Nivelul I de codare (Layer I)

Algoritmul de codare introduce ideea selectarii unui cuantizor ditr-un grup de 15 cuantizoare cu scopul de a obtine un raport zgomot-prag de mascare cit mai uniform in toate benzile, la un debit de 384 Kbps si calitate apropiata de CD. Numarul de biti pe care se face codarea blocurilor (c_i) poate fi de la 0 la 15.
Daca raportul zgomot-prag de mascare (RZPM [dB]) este negativ, atunci zgomotul de cuantizare va fi imperceptibil.

Se poate scrie ca :

$$RZPM_i\,[dB] = Z_i\,[dB] - RZPM_i\,[dB]\ ,$$

unde :
$RZPM_i$ - rapotrul zgomot de cuantizare - prag de mascare
RSZ_i - raportul semnal - zgomot de cuantizare
$RSPM_i$ - raportul semnal - prag de mascare

Procedura de alocare a bitilor este iterativa, pornind cu o alocare de R_i =0 biti si crescind apoi numarul de biti pe care se face codarea pina cind se obtine calitatea dorita.

EXERCITIU: 	Evidentiati principalele caracteristici ale codarii MPEG-1

EXERCITIU:

Ce se asigura prin gruparea in blocuri a secventelor de 12 esantioane obtinute de la iesirea subesantionatorului?

EXERCITIU:

Ce prelucrari au loc in modulul de modelare psiho-acustica si in ce scop?

EXERCITIU:

Descrieti fluxul prelucrarilor pentru obtinerea raportului semnal/prag de mascare in blocul de modelare psiho-acustica.

EXERCITIU:

Ce este specific codarii in modul "stereo"? Dar in modul "intensity stereo"?

5.3.3 Nivelul II de codare (Layer II)

Pentru reducerea debitului binar de la 384 Kbps la 256 Kbps, cit prevede Nivelul II de codare, algoritmul MPEG-1 Layer II face apel la reducerea redundantei care apare in semnalul audio. In cazul codarii MPEG-1 aceasta redundanta se manifesta prin :

- corelatia inalta dintre factorii de scalare din blocuri succesive de esantioane ale aceleiasi subbenzi

- numarul nejustificat de mare al nivelelor de cuantizare pentru benzile de frcventa inalta

La modul concret, aceasta redundanta este rezolvata in procesul de codare prin mai multe masuri care conduc la reducerea debitului binar :

[1] Se grupeaza cite 3 bocuri pe fiecare subbanda si se codeaza ca un singur bloc
a) daca factorii de scalare din cele trei blocuri succesive nu difera prea mult se transmite un singur factor de scalare
b) daca 2 din 3 factori de scalare sunt asemanatori se transmit 2 factori de scalare
c) daca factorii de scalare difeira mult intre ei se transmit toti trei
Observatie : In plus se mai transmite un parametru SCFSI (Scale Factor Side Information) pentru a putea reconstitui valorile factorilor de scalare la receptie.

2. Se reduce numarul nivelelor de cuantizare pentru frecvente medii si inalte (gama dinamica a semnalului in aceste domenii este mica si un numar mare de nivele de cuantizare nu produce un efect acustic important).

3. Se grupeaza cite trei esantioane de semnal din fiecare bloc intr-un triplet. Tripletul se codeaza cu un cod unic de 5, 7 sau 10 biti .

EXERCITIU: 	Cum este tratata redundanta din semnalul vocal in cazul codarii Layer II?

5.3.4 Nivelul III de codare (Layer III)

Nivelul III de codare are la baza structura Nivelelor I si II , dar introduce citeva elemente noi pentru a creste rezolutia la frecvente joase si pentru a elimina variatiile nedorite ale pragului de mascare la aceste frecvente. Debitul binar asigurat este de 128 kbp, dar nu la calitate de CD. Cresterea rezolutiei in frecventa este asigurata de prelucrarea fiecarei subbenzi printr-o TCDM (Transformata Cosinus Discreta Modificata). Din fiecare subbanda sunt generate 6 sub-subbenzi (in cazul TCDM pe termen scurt) sau 18 sub-subbenzi (in cazul TCDM pe termen lung). Pe termen scurt se obtine o mai buna rezolutie temporala, iar pe termen lung o mai buna rezolutie spectrala.

De exemplu, pentru portiuni de semnal cu variatie lenta (frecvente joase) este necesara o crestere a rezolutiei la frecvente joase, deci o ferestra de analiza mai mare. Pentru primele doua subbenzi se va folosi TCDM pe termen lung, iar pentru celelalte 30, TCDM pe termen scurt. Comutarea ferestrei de analiza (scurt /lung, lung/scurt) are loc intr-un bloc de adaptare a ferestrei de analiza pentru tranzitii scurt-lung, respectiv lung-scurt.

Caracteristicile esentiale ale algoritmului sunt :

a) utilizarea TCD cu fereastra de analiza variabila pentru a forma din subbenzi sub-subbenzi, in scopul cresterii rezolutiei spectrale.

b) gruparea in ordinea frecventei a sub-subbenzilor in granule dupa regula repetitiva :

- un esantion din fiecare sub-subbanda cu fereastra de analiza pe termen lung
- 3 esantioane din fiecare sub-subbanda cu fereastra de analiza pe termen scurt pentru a forma un grup de 576 de esantioane. Aceasta grupare garanteaza amplitudinea descrescatoare a esantioanelor. Doua granule formeaza un cadru audio care corespunde la 1452 esantioane de intrare.

c) gruparea adaptiva a esantioanelor din fiecare granula cu scopul de a reduce numarul de biti de codare pentru factorii de scalare din fiecare subbanda. Daca doi factori de scalare succesivi sunt identici, al doilea nu se mai transmite dar se transmite informatie auxiliara. Factorii de scalare se grupeaza in clase care apoi se codeaza.

d) cuantizarea neuniforma a esantioanelor din granule prin aplicarea transformatei $T(f) = f^{0.75}$ asupra componentelor spectrale. Exista 5 cuantizori, corespunzatori la 5 domenii de variatie ale esantioanelor:

- primii 3 codeaza esantioanele mari prin gruparea cite doi si codarea cu cod Huffman
- al patrulea are 3 nivele de cuantizare (+/- 1 , 0) si codeaza un grup de 4 esantioane
- al cincelea corespunde frecventei inalte si codeaza tot timpul 0. Aceasta informatie nu se mai transmite .

e) introducerea unui buffer pentru a adapta debitul binar variabil de la iesirea codorului Huffman la debitul fix al canalului de iesire. Prin reactie negativa se evita golirea sau umplerea bufferului .

Observatie : Bufferul asigura si codarea de inalta calitate a portiunilor de semnal rapid variabile (frecvente mari). Efectul este numit corectie " pre-echo ".

f) headerul nu se transmite la inceputul cadrului audio de lungime variabila, ci acolo unde este nevoie . Un pointer indica pozitia exacta a datelor (main-data begin).

Observatii :

- Pentru modul stereo, acest algoritm permite codarea sumei, respectiv diferentei celor doua canale, bazinduse pe faptul ca diferenta semnalelor poate fi codata cu un numar mai mic de biti. Astfel debitul binar scade. Acest mod de codare se numeste MS_Stereo (M = Middle : left + right , S = Side : left + right). La receptie semnalul original se obtine prin adunarea, respectiv scaderea semnalelor " Middle " , respectiv " Side ".
- Exista situatii in care se poate utiliza atit MS_Stereo cit si Intesity_Stereo simultan, cazuri in care modul MS_Stereo se aplica doar subbenzilor de frecvente joase, acolo unde Intensity_Stereo nu se aplica.

EXERCITIU:

Explicati cum opereaza blocul Transformata Cosinus Discreta Modificata pentru a asigura rezolutia in frecventa pentru Layer III

EXERCITIU:

Explicati rolul bufferului din schema bloc a sistemului de codare MPEG-1 Layer III.

EXERCITIU:

Cum motivati scaderea debitului binar la codarea stereo?

5.4 Sistemul de codare/decodare MPEG-2

MPEG-2 este o extensie a sistemului MPEG-1 pentru codarea canalelor stereo. Frecventa de esantionare poate fi scazuta la 16, 22.5 sau 24 kHz, iar debitul binar poate fi chiar 8 Kbps, fapt semnalat prin pozitionarea in zero a bitului ID din header.

Trasatura esentiala a MPEG-2 este definirea si utilizarea unui standard pentru compresia a doua canale stereo:

- stinga (L- left),
- dreapta (R - right) , pentru compatibilitate cu MPEG-1
  si inca 4 canale aditionale:
- central (C - front Center),
- auxiliar stinga (LS - side / rear Left Surround) ,
- auxiliar dreapta (RS - side / rear Right Suround)
- un canal optional de inalta fidelitate la frecvente joase (LFE - Low Frequency Enhancement).

Compatibilitatea cu MPEG-1 se realizeaza prin codarea canalelor L si R ca in MPEG-1, iar a canalelor aditionale in cimpul de date auxiliar ce urmeaza dupa datele audio in canalul MPEG-1. Fluxul binar rezultat se numeste multicanal audio.

In continuare se prezinta principalele caracteristici ale codarii MPEG-2 :

1. Gruparea subbenzilor:

- este o metoda prin care are loc codarea unui grup de subbenzi in locul subbenzilor sau sub-subbenzilor individuale, cu csopul de a reduce debitul binar prin analiza dinamicii esantioanelor din subbenzile adiacente .

Tabel 5.4.1. Gruparea subbenzilor la MPEG-2 . Layer I si Layer II

Grup (Layer I si II)	Subbenzi in grup
0	0
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8-9
9	10-11
10	12-15
11	15-31

2. Transformarea matriceala :

- prin aplicarea unei transformari matriceale M (de ordin 5 x 5) asupra multicanalului audio (L, C, R, L , RS) se obtine canalul de transmisie :

(L_o , R_o , T_2 , T_3 , T_4) i (L , C , R , LS , RS) x M ,

in care fiecare componenta L_o , R_o , T_2 , T_3 sau T_4 este o combinatie liniara a canalelor L, C, R, LS, RS.

Aceasta transformare produce cresterea calitatii la decodarea cu un decodor MPEG-1, pentru ca in componentele L_o si R_o se vor regasi informatii din celelalte canale. In plus, componentele T_2 , T_3 si T_4 pot fi comprimate, daca se tine seama de corelatia cu L_o si R_o. Din pacate, acest tip de transformare poate conduce la o mai slaba calitate a decodarii MPEG-2. Transformarea M poate fi definita pentru toate subbenzile sau pentru fiecare grup de subbenzi si modificata, eventual dinamic, pentru fiecare cadru.

3. Aplicarea modului Intensity-Stereo :

- se face intr-o varianta adaptata la cerintele MPEG-2, prin eliminarea bitilor de cod corespunzatori unui grup de subbenzi si inlocuirea lor cu biti de cod ai unui alt grup (de fapt acestia nu se mai transmit cu scopul de a scadea debitul binar). Pentru controlul independent al amplitudinii sunt atasati doar factorii de scala .

4. Predictia adaptiva multicanal :

- un nivel de compresie suplimentar al canalelor T_2, T_3 si T_4 se poate realiza prin codarea predictiva a primelor 8 grupuri de subbenzi (in Layer I si II)

ale acestor canale, pe baza selectiei predictorului cel mai eficient dintr-un grup de 6 predictori. Desi predictorii au structura fixa, predictia este intr-un fel adaptiva din cauza posibilitatii alegerii predictorului din grupul de 6 predictori .

5. Codarea falsa a canalului central :
- este o metoda de compresie a canalului central prin care subbenzile de ordin mai mare ca 11 nu se transmit, iar la decodare, in locul lor se insereaza 0. La receptie se percep frecventele inalte doar pe celelalte canale.

6. Canale multilingve :
- in MPEG-2 pot exista pina la 7 canale mono intr-un canal multilingv si care sa contina fiecare :
  - semnal vocal in diferite limbi
  - comentarii audio pentru persoane cu handicap vizual
  - semnal audio de calitate pentru persoane cu handicap auditiv
  - comentarii cu scop educational , e t c .

La rindul lui un canal multilingv poate fi :
- nindependent
- multiplexat intr-un program multicanal

Observatii :
a) Sistemul MPEG permite pina la 32 de canale intr-un program, fiind in principiu posibila existenta a 224 canale multilingve intr-un program MPEG-2 .

b) Pastrarea compatibilitatii cu MPEG-1 (MPEG-2 BC, Backward Compatible) limiteaza calitatea semnalului receptionat, chiar daca are loc o marire a debitului binar. Exista deja citiva algoritmi pentru codarea celor 5 canale la calitate foarte buna (de exemplu algoritmul AC-3 , propus de Dolby , sau algoritmul PAC , propus de AT & T).

EXERCITIU: 	Cum este asigurata compatibilitatea sistemelor MPEG-1 si MPEG-2?

EXERCITIU: 	Care este efectul gruparii subbenzilor la codarea MPEG-2?

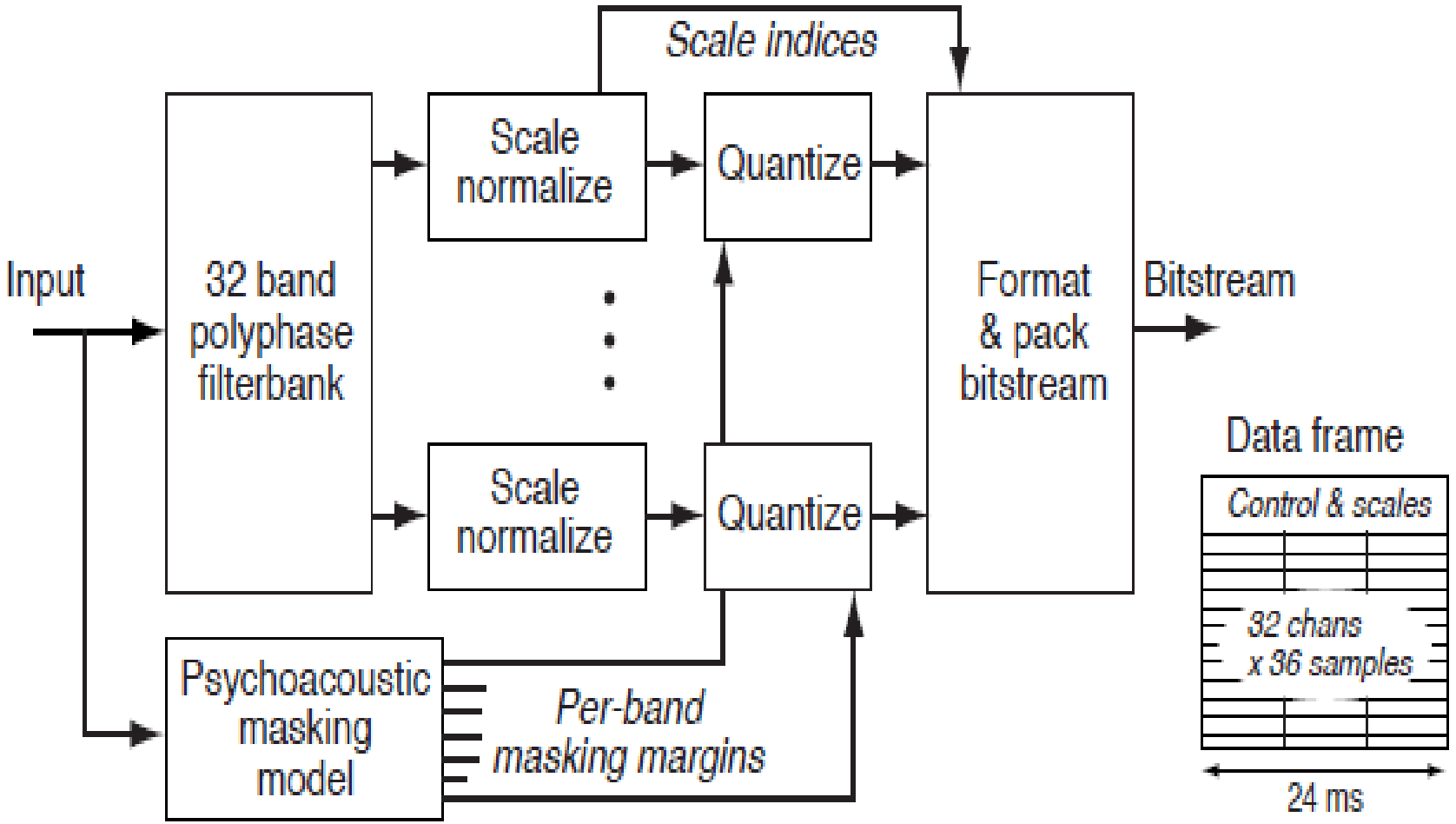
EXERCITIU: 	Explicati urmatoarele notiuni: - predictie adaptiva multicanal - codare falsa a canalului central

Concluzii:

- *Codarea MPEG a semnalelor audio exploateaza caracteristici psiho-acustice intr-un sistem de codare in subbenzi. In spectrul semnalelor audio exista componente spectrale care nu sunt auzite, din cauza ca sunt in vecinatatea unor compenente de frecventa apropiata si de amplitudine mult mai mare. Acest efect este numit efect de mascare.*
- *Efectul de mascare se manifesta in 3 ipostaze: mascarea componentelor spectrale, mascarea temporala si pragul absolut de mascare.*
- *MPEG-1 este un standard care permite codarea canalelor stereo la debite date de nivelul de codare. Nivelele de codare se numesc Layer I, Layer II, Layer III.*
- *Schema MPEG-1 este in esenta un codor in subbenzi dotat cu un bloc de procesare auxialiar: blocul de modelare psiho-acustica, prin intermediul caruia se aloca dinamic numarul de biti pentru fiecare subbanda, dupa ce in prealabil s-a calculat pragul de mascare.*
- *MPEG-1 Layer I are in vedere obtinerea unui raport zgomot/prag de mascare cit mai uniform in toate subbenzile prin selectarea comandata a unui cuantizor dintr-un grup de 15 cuantizori.*
- *MPEG-1 Layer II reduce debitul binar de la 384 Kbps la 256 Kbps prin eliminarea redundantei din semnalul audio folosind tehnici precum: codarea unor grupe de blocuri din fiecare subbnada, reducerea numarului nivelelor de cuantizare pentru frecvente medii si inalte, codarea unor triplete de esantioane.*
- *MPEG-1 Layer III utilizeaza in structura codorului un bloc de Transformata Cosinus Modificata ce opereaza comandat pe termen scurt sau lung in vederea imbunatatirii rezolutiei temporale (pentru semnale cu variatie lenta) sau spectrale (pentru semnale cu variate rapida).*
- *MPEG-2 genereaza un Multicanal Audio folosind urmatoarele idei: gruparea subbenzilor, aplicarea unor transformari matriceale multicanalului audio, utilizarea predictiei adaptive multicanal sau codarea falsa a canalului central. Se pastreaza compatibilitatea cu MPEG-1.*

CODAREA MPEG

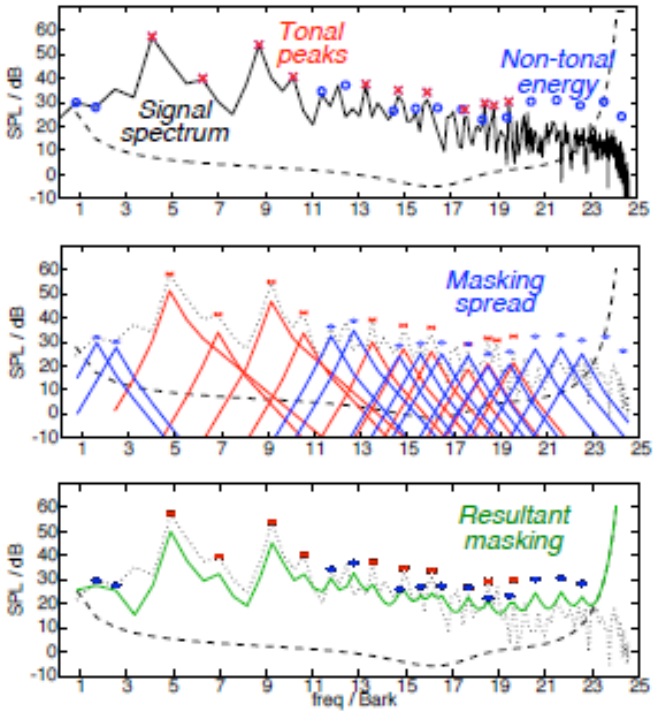
MPEG1 & MPEG 2 // Layer III


Scale indices
Input
32 band polyphase filterbank
Scale normalize
Quantize
Scale normalize
Quantize
Format & pack bitstream
Bitstream
Psychoacoustic masking model
Per-band masking margins
Data frame
Control & scales
32 chans x 36 samples
24 ms

MPEG Psychoacoustic model

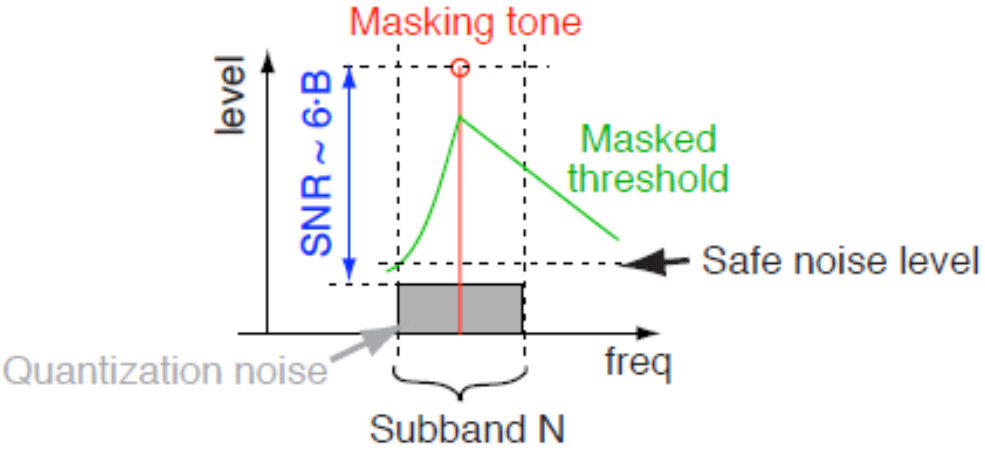
- Based on simultaneous masking experiments
- Difficulties:
  - noise energy masks ∼10 dB better than tones
  - masking level nonlinear in frequency & intensity
  - complex, dynamic sounds not well understood

- Procedure
  - pick 'tonal peaks' in NB FFT spectrum
  - remaining energy → 'noisy' peaks
  - apply nonlinear 'spreading function'
  - sum all masking & threshold in power domain



MPEG Bit allocation

- Result of psychoacoustic model is maximum tolerable noise per subband

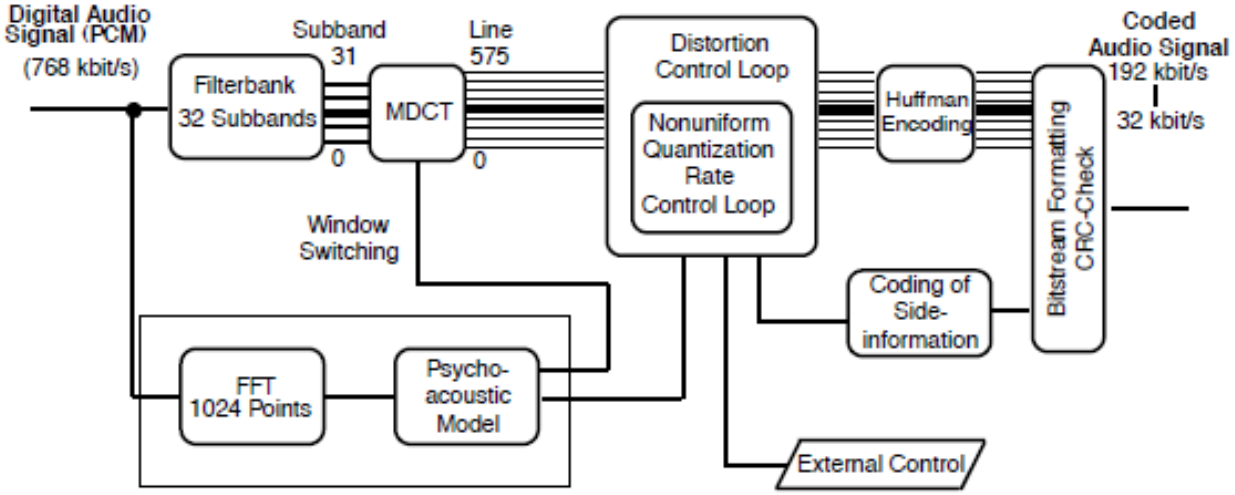


  - safe noise level → required SNR → bits B
- Bit allocation procedure (fixed bit rate):
  - pick channel with worst noise-masker ratio
  - improve its quantization by one step
  - repeat while more bits available for this frame
- Bands with no signal above masking curve can be skipped

MPEG 2, Layer III

MPEG Audio Layer III

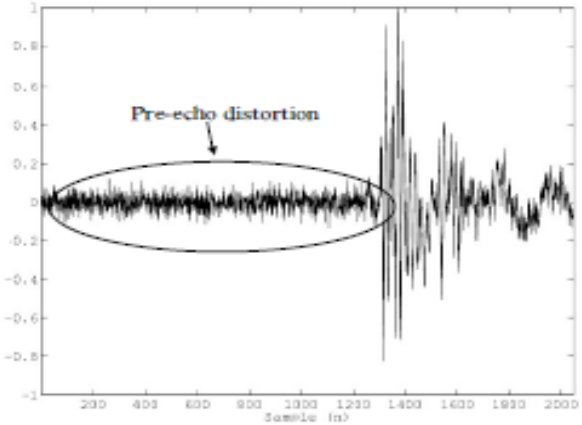
- 'Transform coder' on top of 'subband coder'



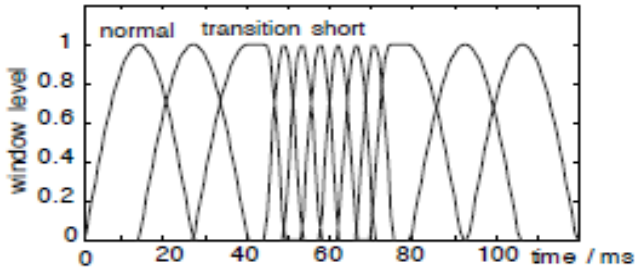
- Blocks of 36 subband time-domain samples become 18 pairs of frequency-domain samples
  - more redundancy in spectral domain
  - finer control e.g. of aliasing, masking
  - scale factors now in band-blocks
- Fixed Huffman tables optimized for audio data
- Power-law quantizer

Adaptive time window

- Time window relies on temporal masking
  - single quantization level over 8-24 ms window
- 'Nightmare' scenario:

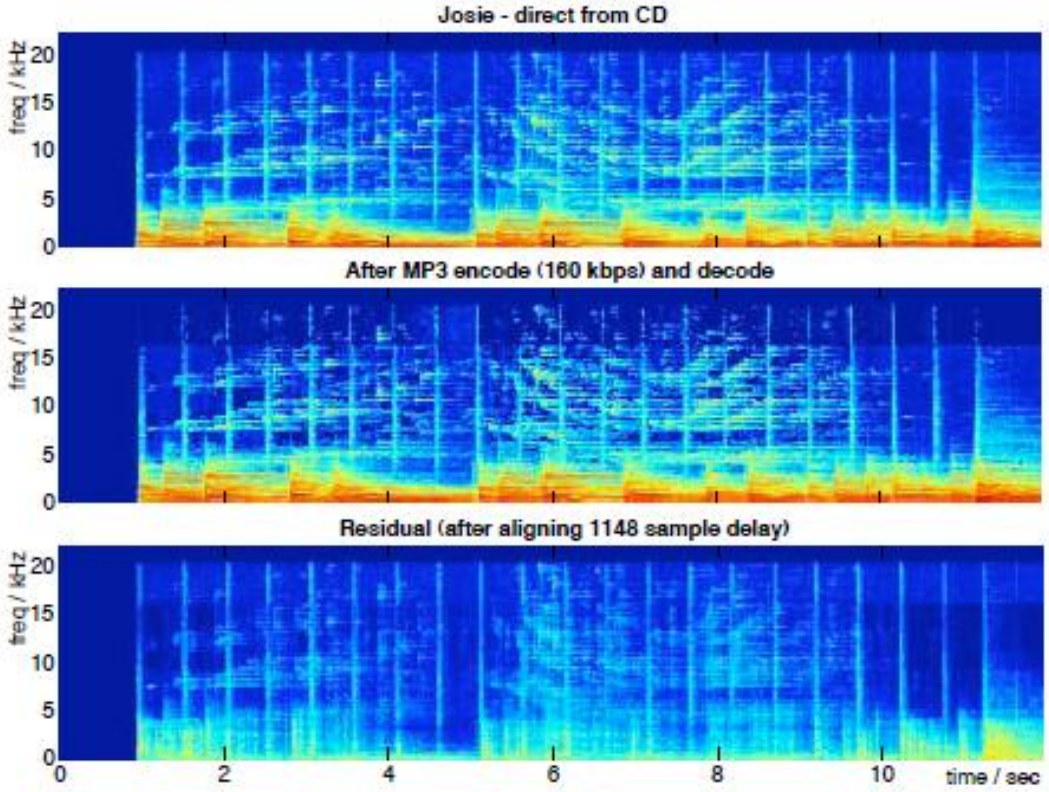


  - 'backward masking' saves in most cases
- Adaptive switching of time window:



The effects of MP3

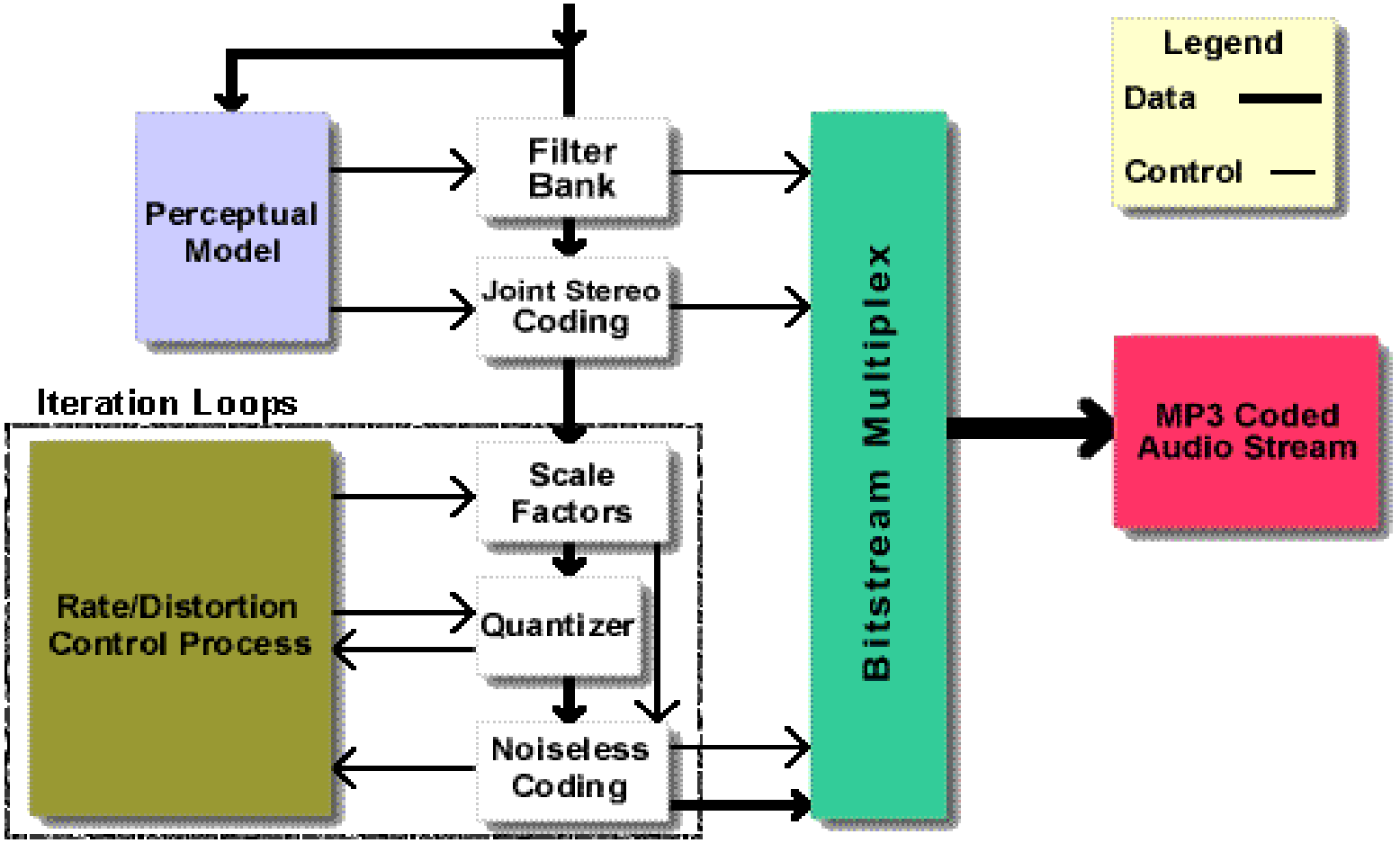
- Before & after:



  - chop off high frequency (above 16 kHz)
  - occasional other time-frequency 'holes'
  - quantization noise under signal

OVERVIEW

Structure


Legend
Data
Control
Perceptual Model
Filter Bank
Joint Stereo Coding
Iteration Loops
Scale Factors
Rate/Distortion Control Process
Quantizer
Noiseless Coding
Bitstream Multiplex
MP3 Coded Audio Stream

[4]

MPEG Audio Input

Data

- Input sampled at 16, 22.05, 24, 32, 44.1, or 48kHz
- Up to 5 channels
- Ancillary Data

Settings

- Set to a desired output bitrate down to 8kbits/sec
- Choose level (layer) of compression desired
- Optional CRC

Layers

- Layer I: > 128kbits/sec per channel. Phillips' Digital Compact Cassette (192kbits/sec)
- Layer II: ≈ 128 kbits/sec. Video CD
- Layer III = MP3: ≈ 64 kbits/sec

Filter Bank

- All layers use same filter bank
- Filters break input signal evenly into 32 bands
- Lossy filter with some subband overlap

Filter Bank

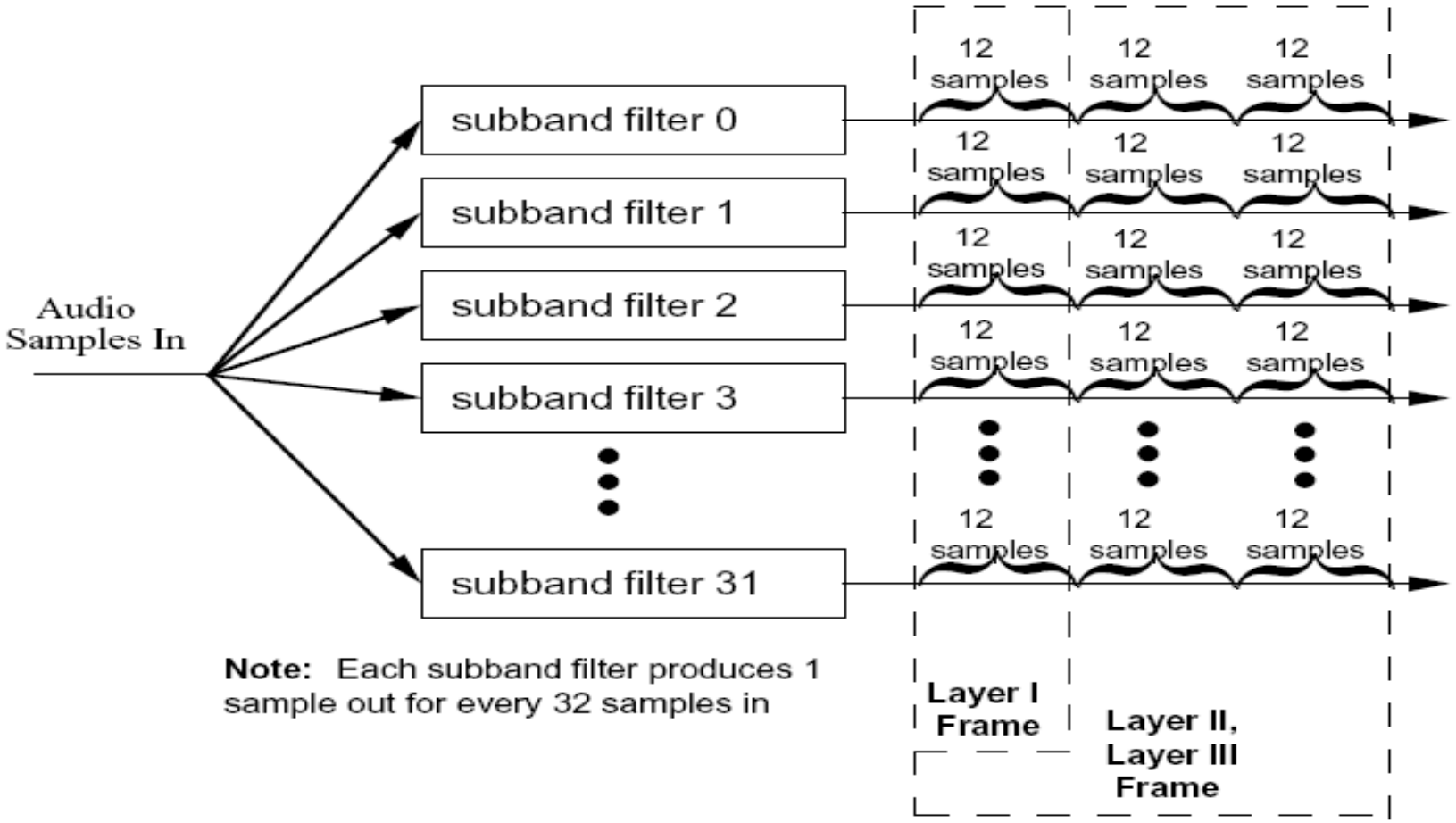
$$s_t[i] = \sum_{k=0}^{63} \sum_{j=0}^{7} M[i][k] * (C[k+64j] * x[k+64j])$$

where:

i is the subband index and ranges from 0 to 31,
$s_t[i]$ is the filter output sample for subband i at time t, where t is an integer multiple of 32 audio sample intervals,
C[n] is one of 512 coefficients of the analysis window defined in the standard,
x[n] is an audio input sample read from a 512 sample buffer, and
$M[i][k] = \cos\left[\frac{(2*i+1)*(k-16)*\pi}{64}\right]$ are the analysis matrix coefficients.

[1]

Filter Bank



[1]

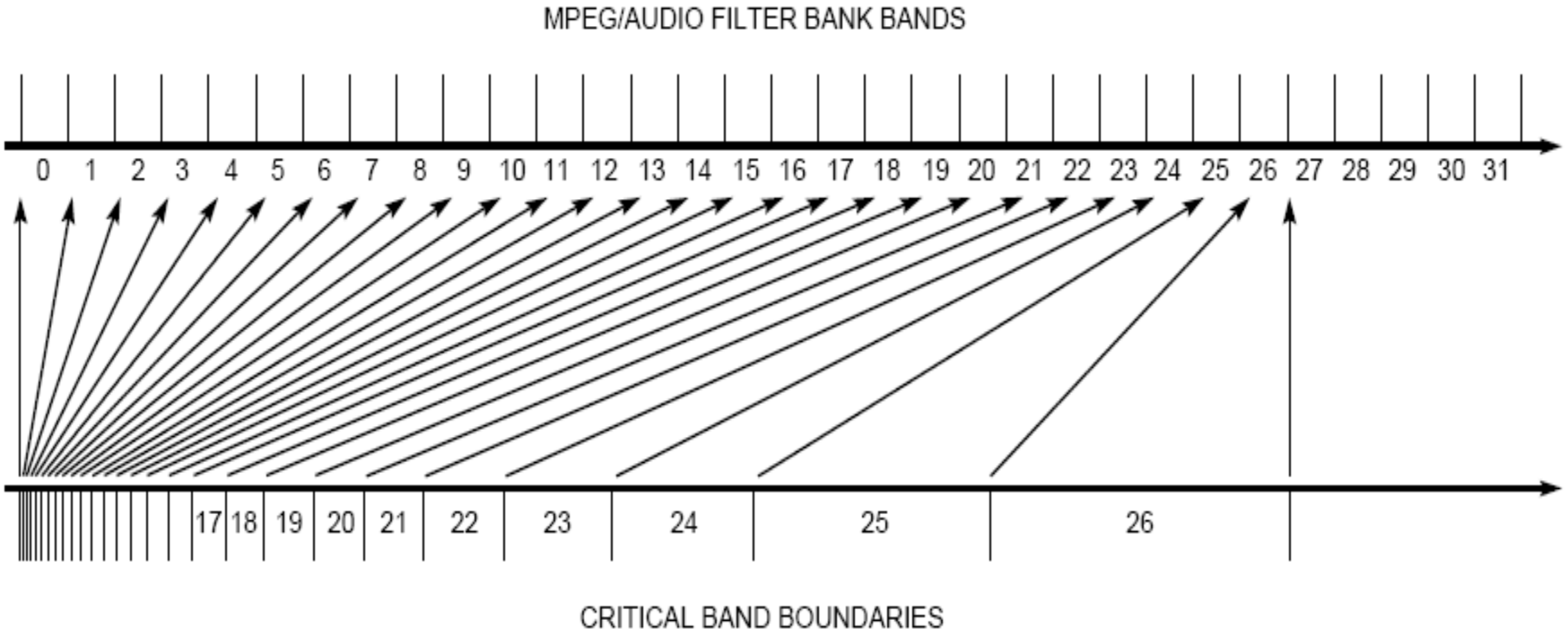
Layer III

- Adds MDCT on each subband to increase the resolution inside each frequency
- Specifies 18 or 6 size MDCT block lengths to better differentiate stationary or transient signals
- Adds Alias Reduction caused by overlapping subbands in filter bank

Psychoacoustic Model

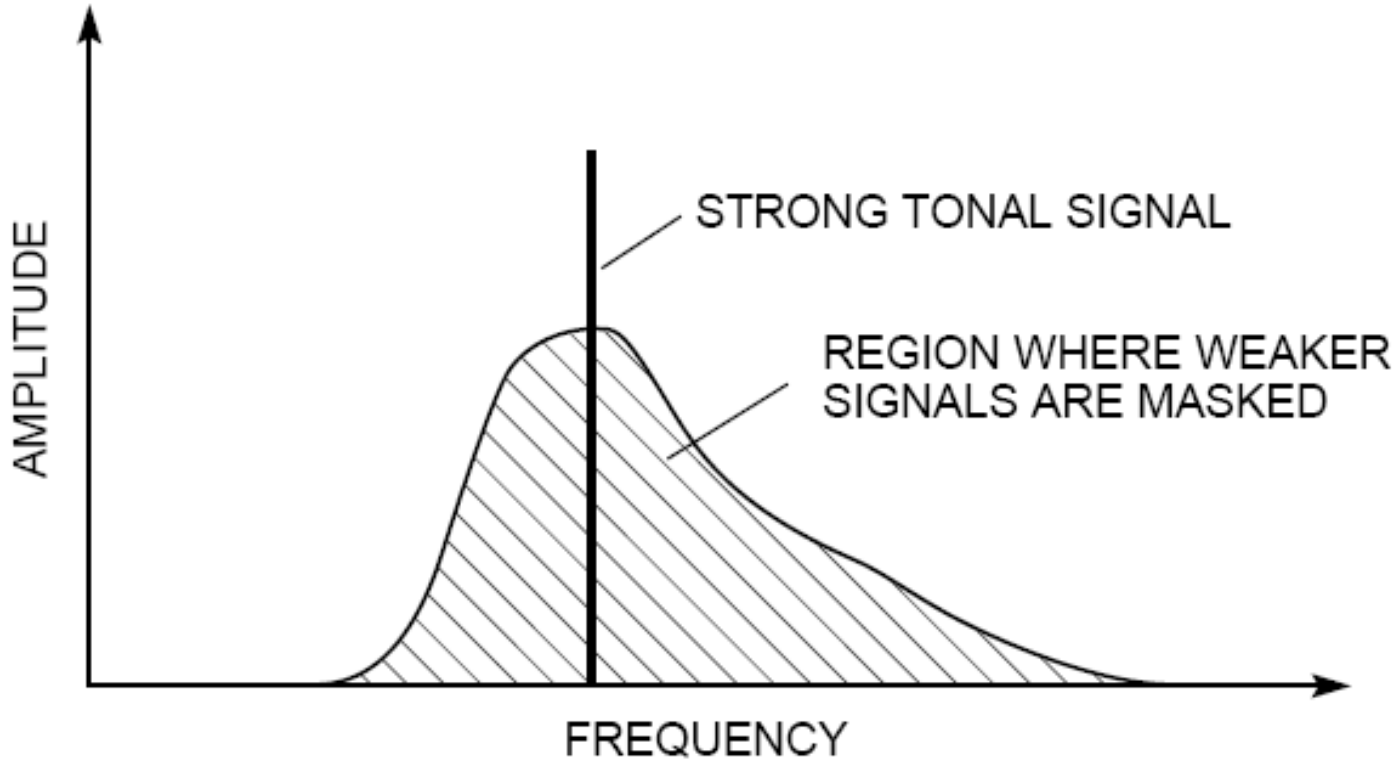
- Critical Region Step – combine/split subbands based on empirical model of human hearing
- Masking Step – eliminate sounds that would be overridden by louder tones

Critical Bands



[1]

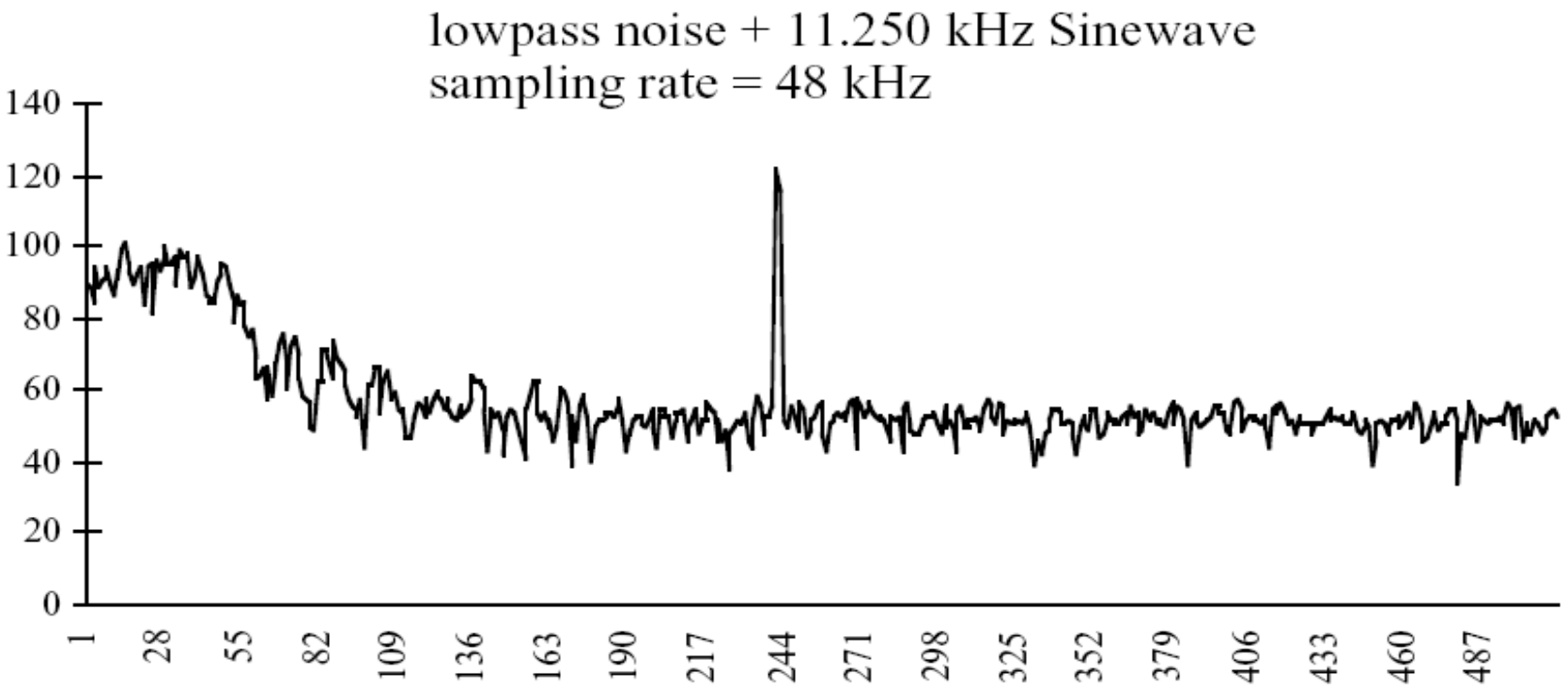
Masking


AMPLITUDE
STRONG TONAL SIGNAL
REGION WHERE WEAKER
SIGNALS ARE MASKED
FREQUENCY

[1]

Masking

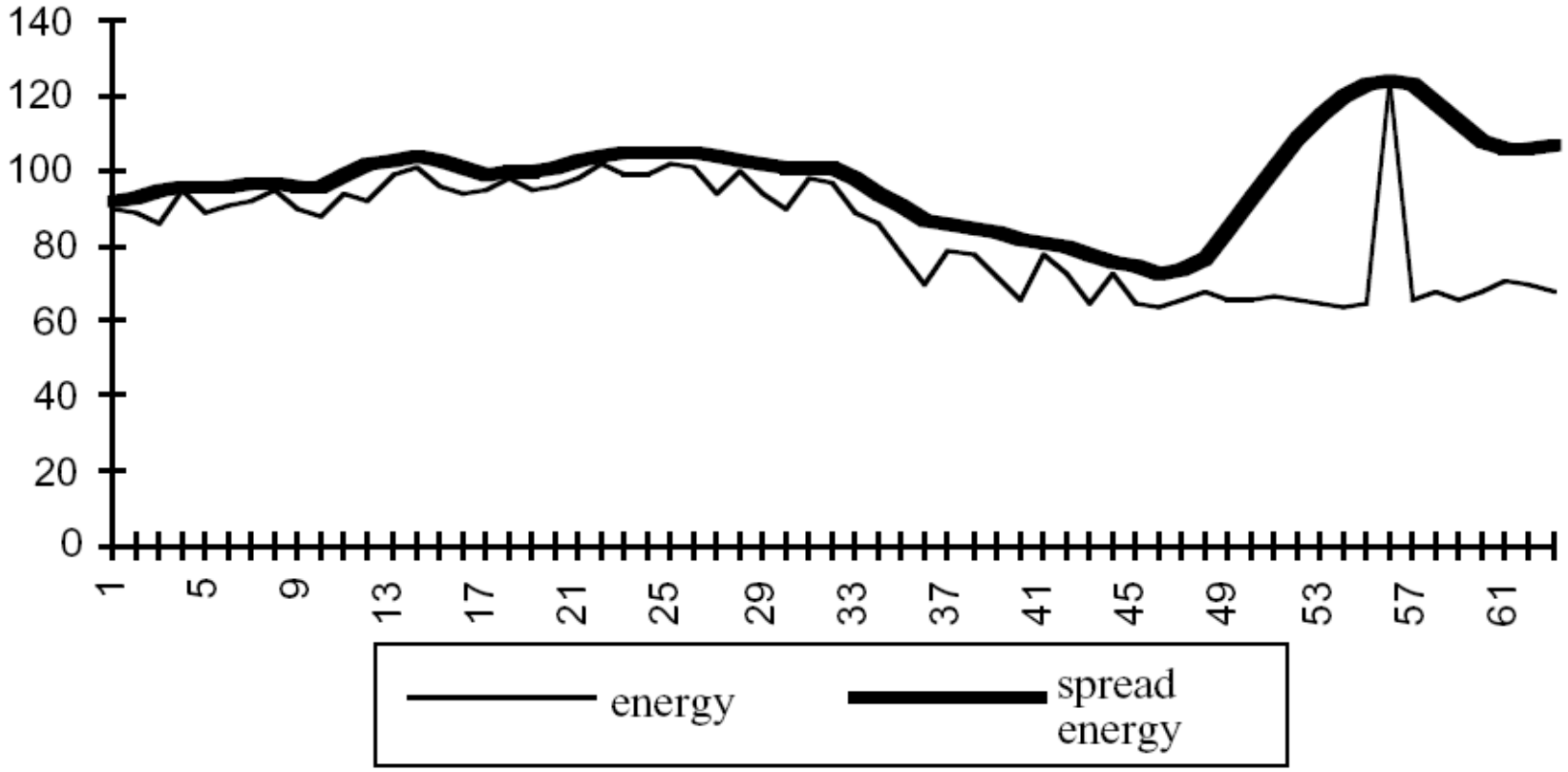
The original signal



[1]

Masking

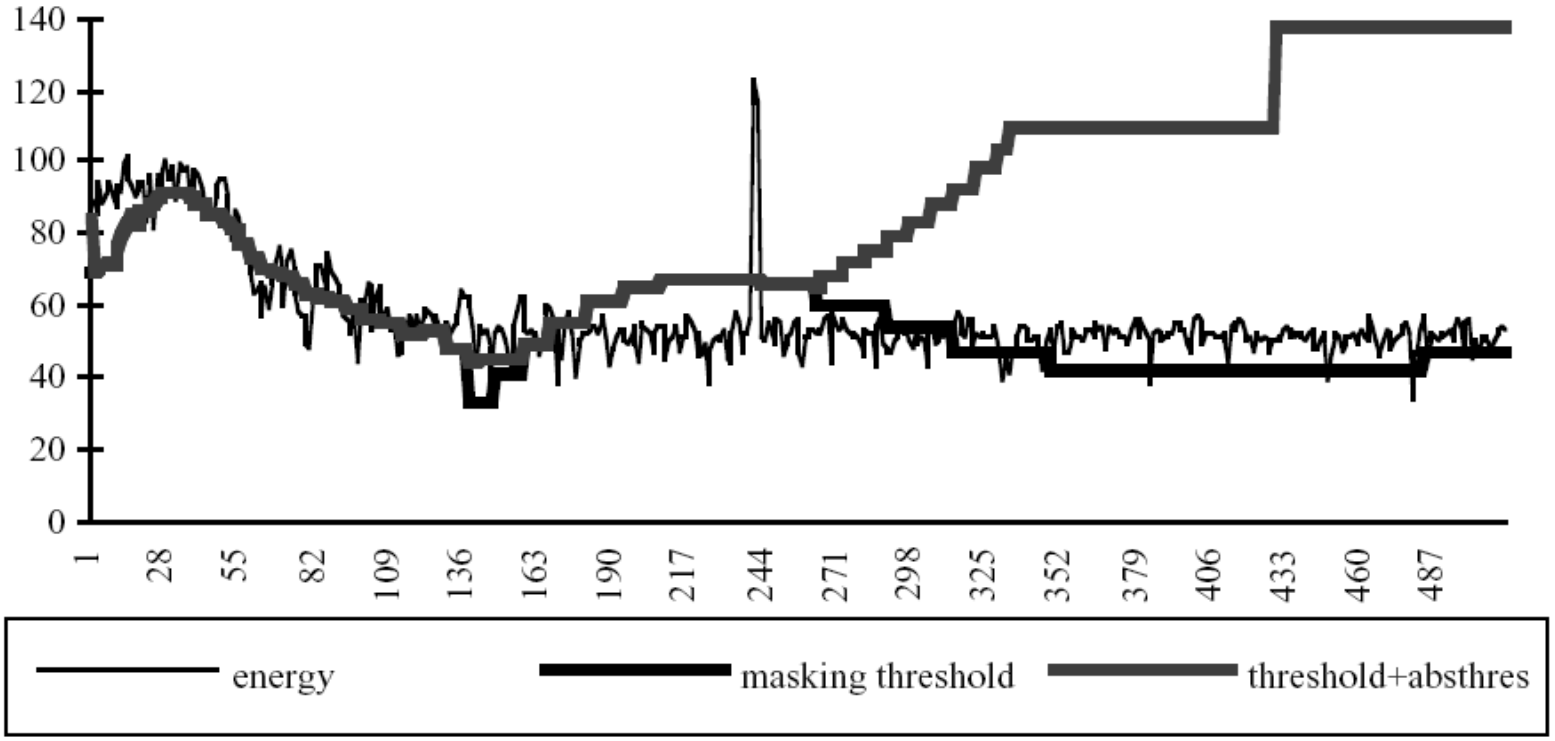
Smooth the signal using a spreading function



[1]

Masking

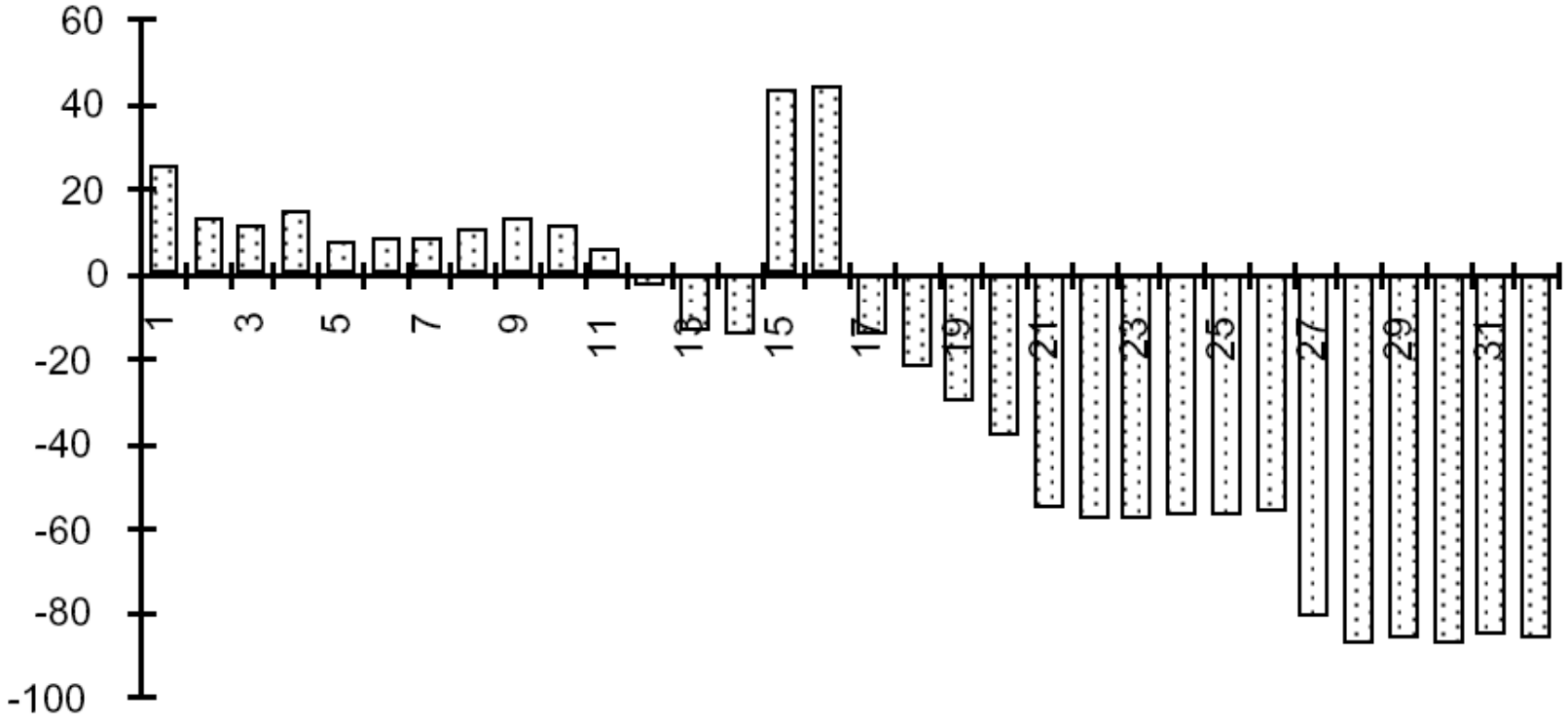
Calculate a threshold using previous windows & quantization level



[1]

Masking

Find the ratio of the signal to the mask (SMR)



[1]

Bit Allocation (Scaling)

- Mask to Noise Ratio for band i MNR[i] = SNR[i] – SMR[i]
- Signal to Noise Ratio is a function that defines how much noise is introduced given a particular bit allocation
- Allocate bits to lowest MNR
- Recompute MNR and repeat

The Bit Reservoir

- A frame with little audio interest may require few bits to encode
- A frame with substantial audio interest may require more bits to encode
- Allow frames to give to or take from a reservoir

Entropy Coding

- Reorder Coefficients
- Define similar segments
- Encode each segment using Huffman

Entropy Coding

Reorder Coefficients

- Predefined ordering by frequency
- High frequencies tend to be runs of zeros
- Low frequencies tend to be large values

Entropy Coding

Define Similar Segments

- Region 1 = from front, a run of all zeros of even length as.
- Region 2 = from region 1 end, a run of length % 4 = 0 over {–1, 0, 1}
- Region 3 = The remaining

Entropy Coding

Huffman Encode

- Don’t encode region 1, we can deduce its size.
- Huffman Encode region 2, 4 values at a time
- Subdivide Region 3 into 3 more even segments each with its own Huffman tree

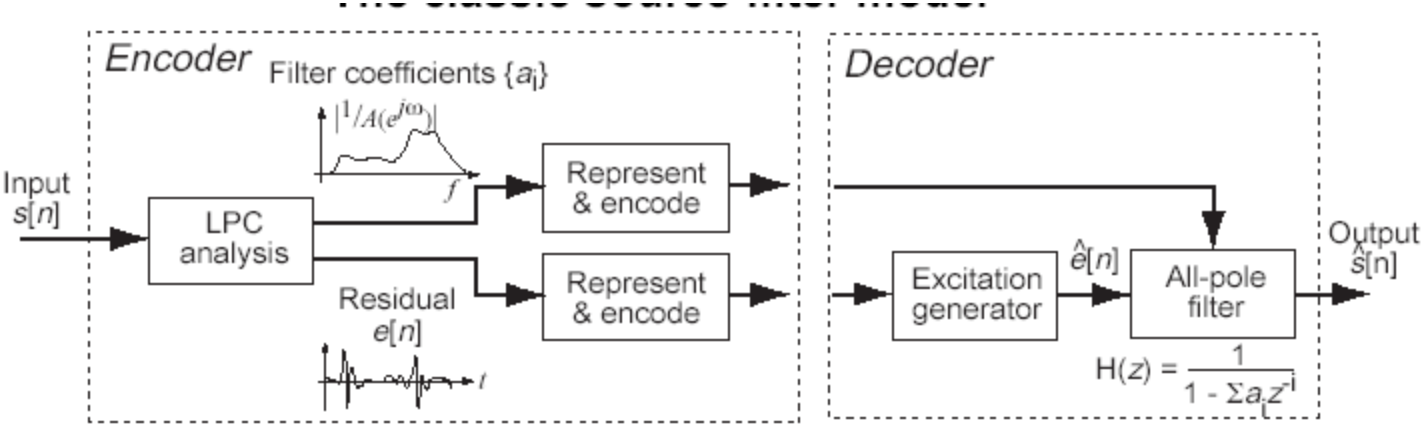
References

- [1] D. Pan, “A Tutorial on MPEG/Audio Compression”, *IEEE Multimedia Journal*, 1995.
- [2] http://en.wikipedia.org/wiki/MP3
- [3] Compression Quality (biased) http://www.iis.fraunhofer.de/amm/techinf/layer3/index.html#2
- [4] Overview http://www.iis.fraunhofer.de/amm/techinf/layer3/layer3_block.gif

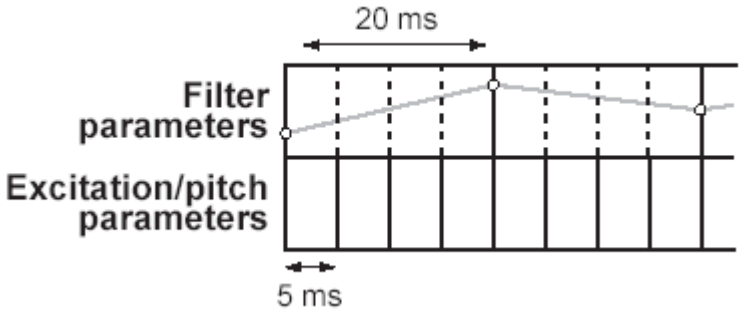
CODAREA PARAMETRICA

LPC encoding

- The classic source-filter model



- Compression gains:
  - filter parameters are ~slowly changing
  - excitation can be represented many ways



Linear Predictive Code

- Model speech production system as an auto-regressive model:

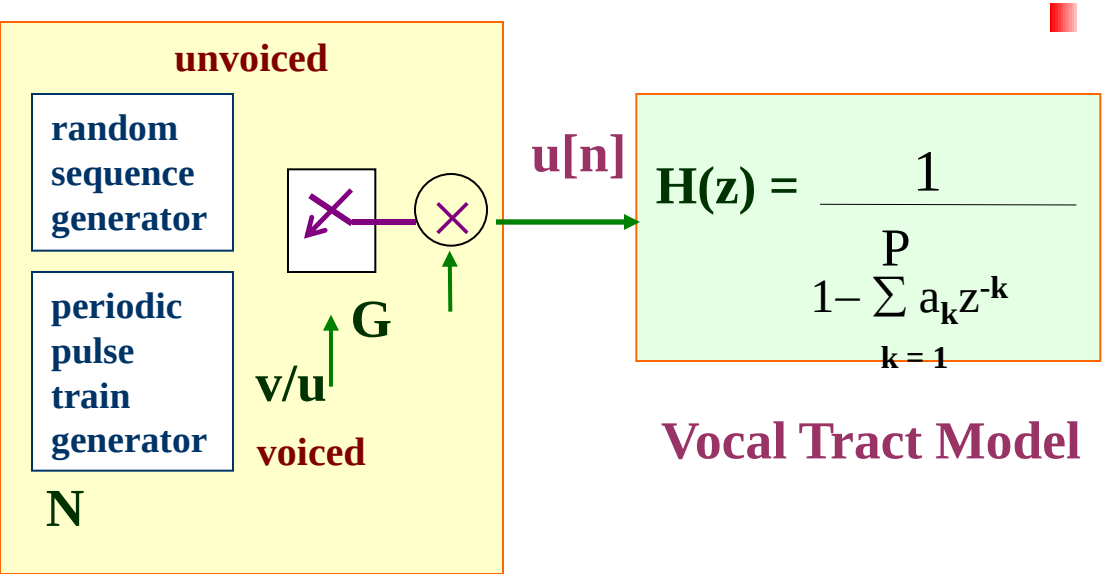
$$s(n) = \sum_{k=1}^{p} a(k)s(n-k) + e(n)$$

- Model parameters are computed for speech segment (~30 ms).
- Parameters {a(k); k=1:p} are found by solving a Toeplitz system of equations.

- Transfer function

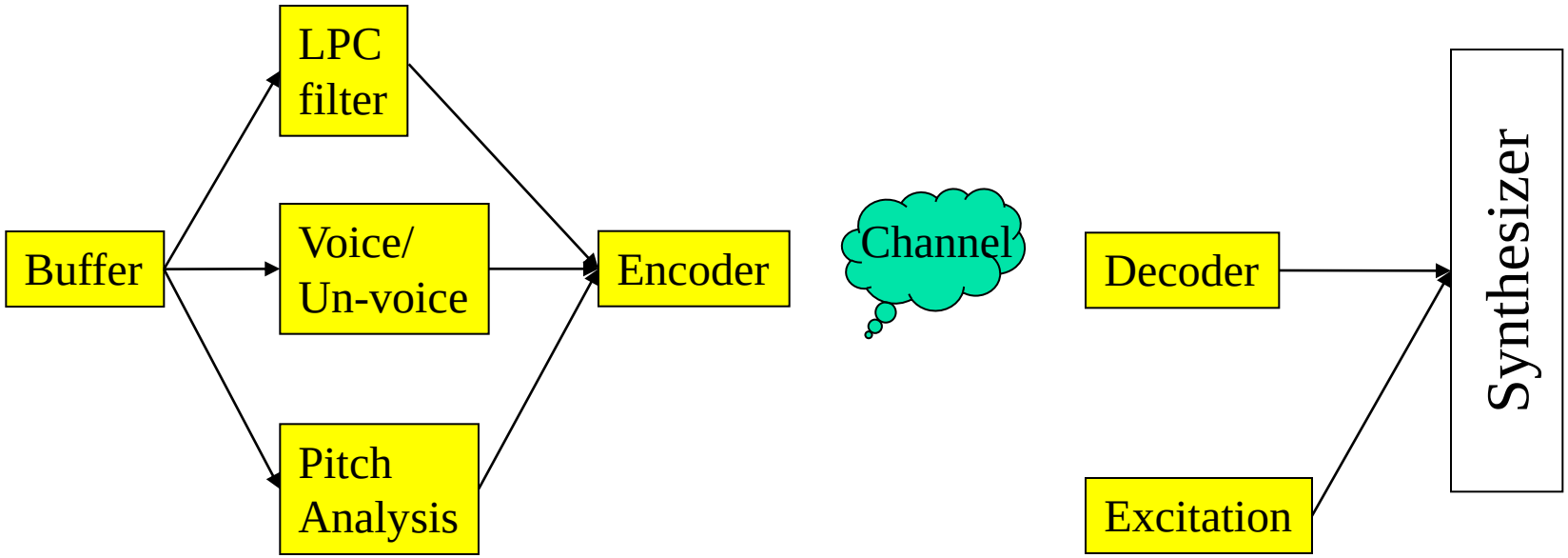
$$H(z) = \frac{S(z)}{E(z)} = \frac{G}{1 - \sum_{k=1}^{p} a(k)z^{-k}}$$

- To encode speech, one may transmit the quantized parameters {a(k)} and G or equivalent parameter set.
- The model order is 8-10 in most speech coding standards.

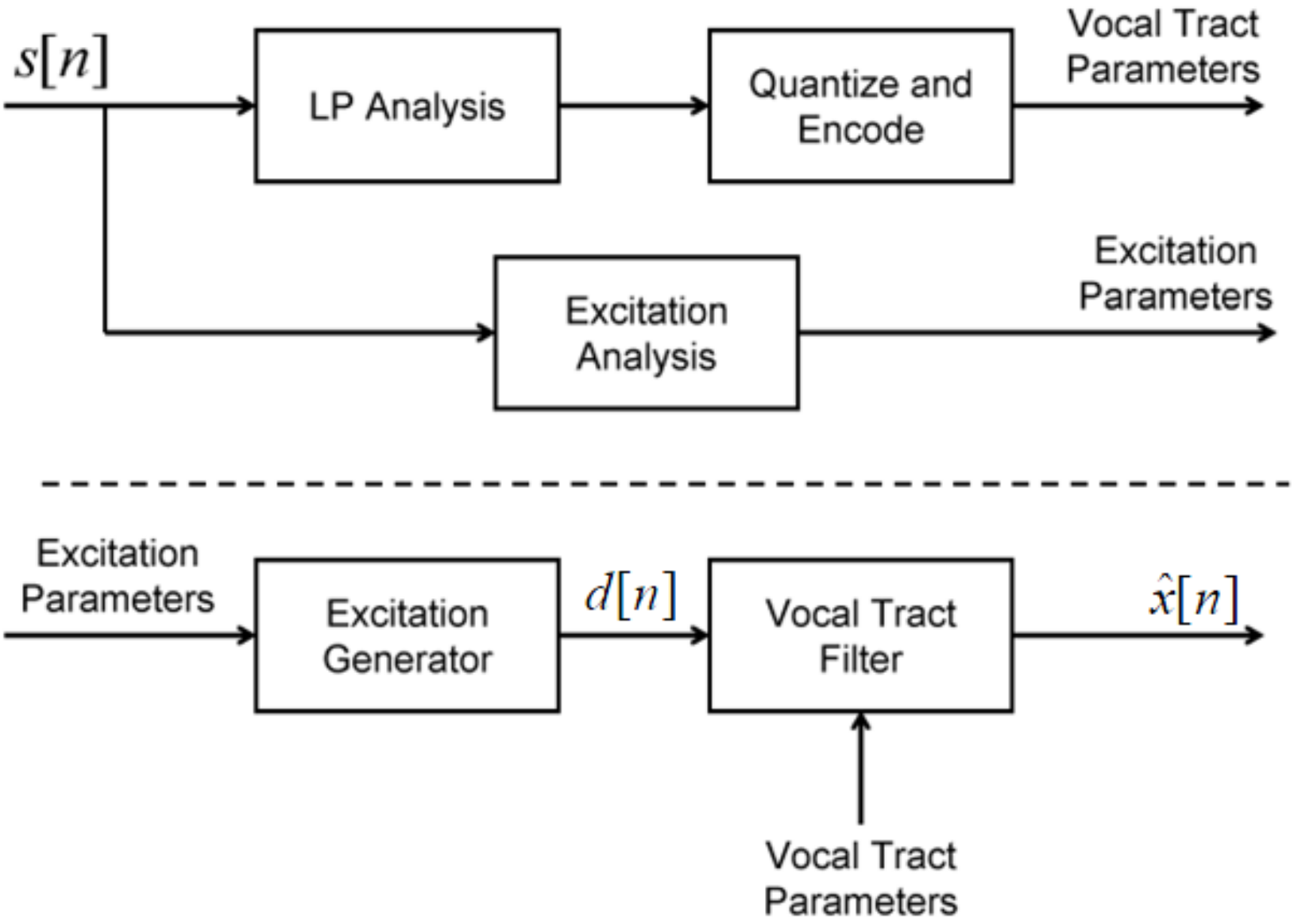


Vocal Tract Model

LPC Speech Coder



Using LP in Speech Coding



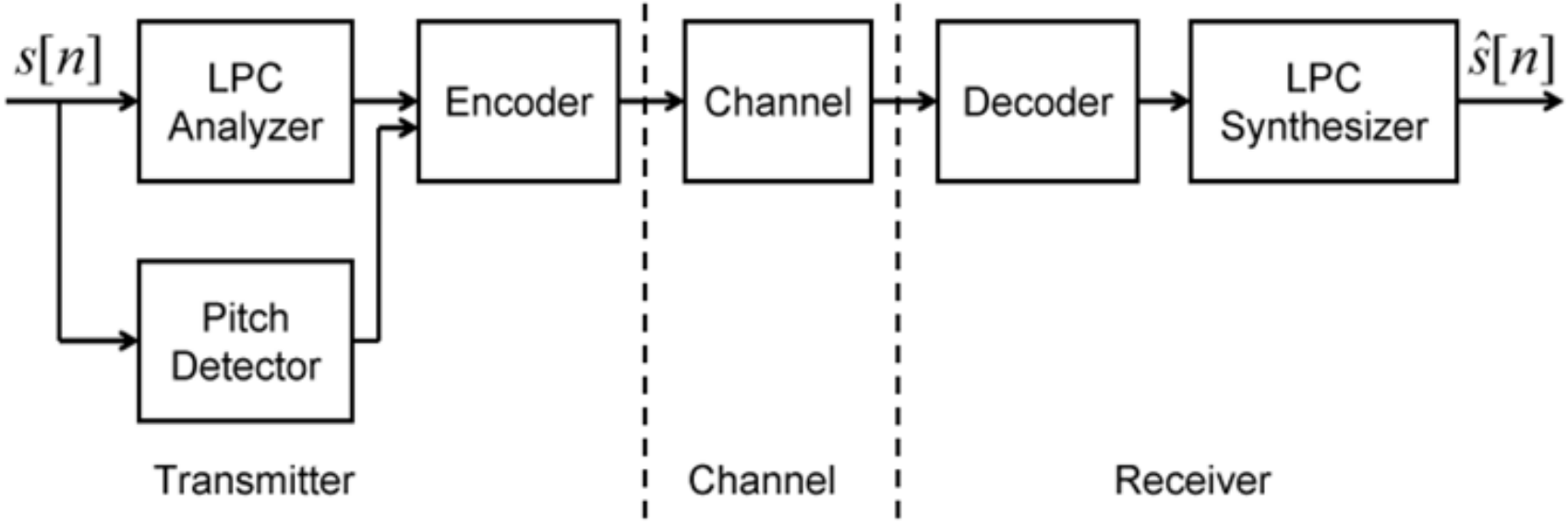
Model-Based Coding

- assume we model the vocal tract transfer function as

$$H(z) = \frac{X(z)}{S(z)} = \frac{G}{A(z)} = \frac{G}{1 - P(z)}$$

$$P(z) = \sum_{k=1}^{p} a_k z^{-k}$$

- LPC coder => 100 frames/sec, 13 parameters/frame (p = 10 LPC coefficients, pitch period, voicing decision, gain) => 1300 parameters/second for coding $\leftrightarrow$ versus 8000 samples/sec for the waveform



LPC Parameter Quantization

- don't use predictor coefficients (large dynamic range, can become unstable when quantized) => use LPC poles, PARCOR coefficients, etc.
- code LP parameters optimally using estimated pdf's f each parameter

1.	V/UV-1 bit	100 bps
2.	Pitch Period-6 bits (uniform)	600 bps
3.	Gain-5 bits (non-uniform)	500 bps
4.	LPC poles-10 bits (non-uniform)-5 bits for BW and 5 bits for CF of each of 6 poles	6000 bps
Total required bit rate		7200 bps

• no loss in quality from uncoded synthesis (but there is a loss from original speech quality)

• quality limited by simple impulse/noise excitation model

S5

S5-synthetic

LPC Coding Refinements

1. log coding of pitch period and gain
2. use of PARCOR coefficients ($|k_i|<1$) => log area ratios $g_i=\log(A_{i+1}/A_i)$—almost uniform pdf with small spectral sensitivity => 5-6 bits for coding

- can achieve 4800 bps with almost same quality as 7200 bps system above
- can achieve 2400 bps with 20 msec frames => 50 frames/sec

LPC-10 Vocoder

LPC-10 Vocoder

- U.S. Government standard
  - covariance LP analysis (10th-order)
  - AMDF pitch detector (see Chapter 4)
- Bit rate

Frame rate = 44.44 frames/sec

param.	$k_1 - k_4$	$k_5 - k_8$	k_9	k_{10}	pitch	ampl.	sync.	Total
# bits	5 ea.	4 ea.	3	2	7	5	1	54

Bit rate = 2400 bits/sec

LPC-Based Speech Coders

- the key problems with speech coders based on all-pole linear prediction models
  - inadequacy of the basic source/filter speech production model
  - idealization of source as either pulse train or random noise
  - lack of accounting for parameter correlation using a one-dimensional scalar quantization method => aided greatly by using VQ methods

Schema de codare

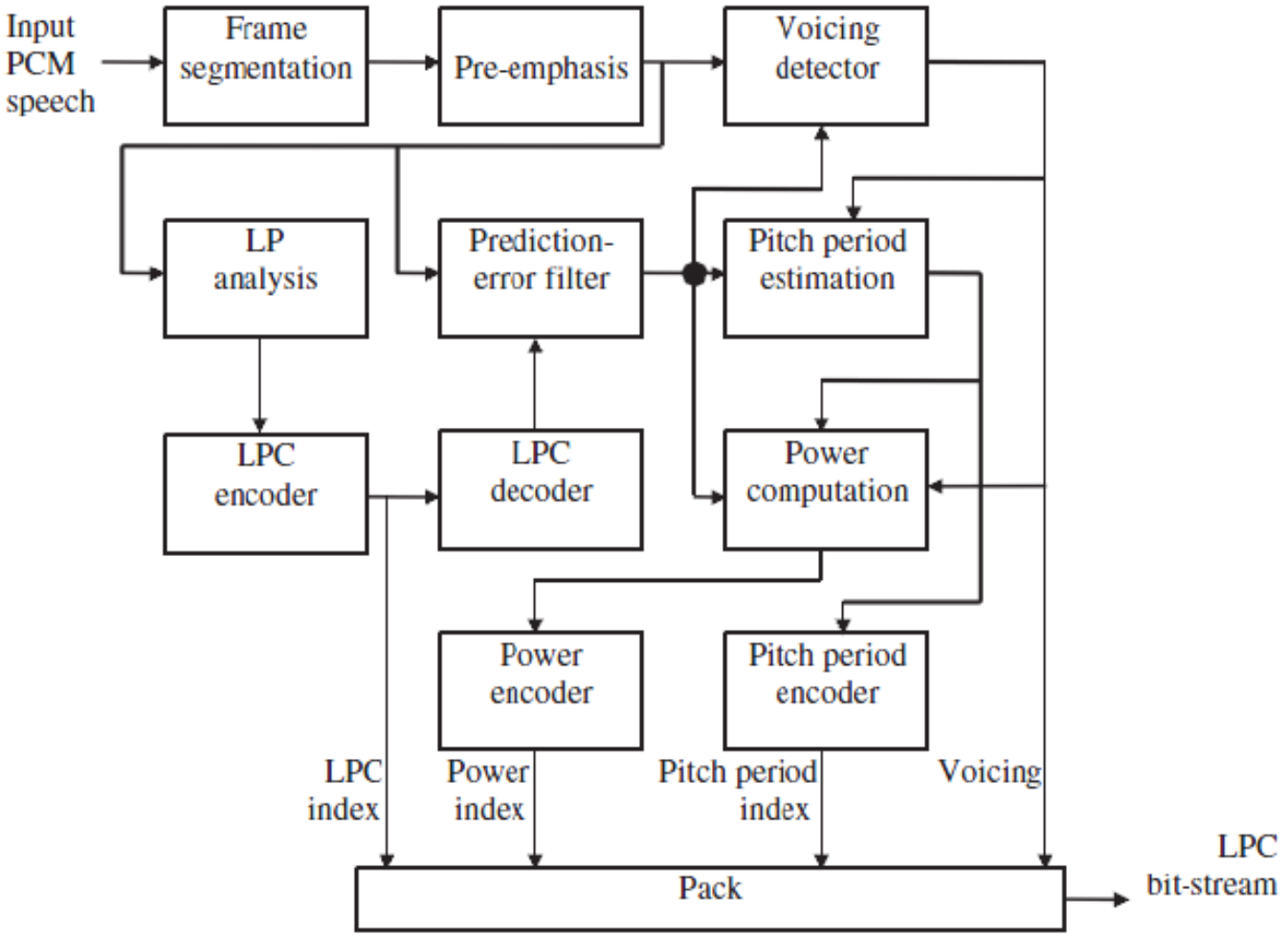


Figure 9.6 Block diagram of the LPC encoder.

Schema de decodare

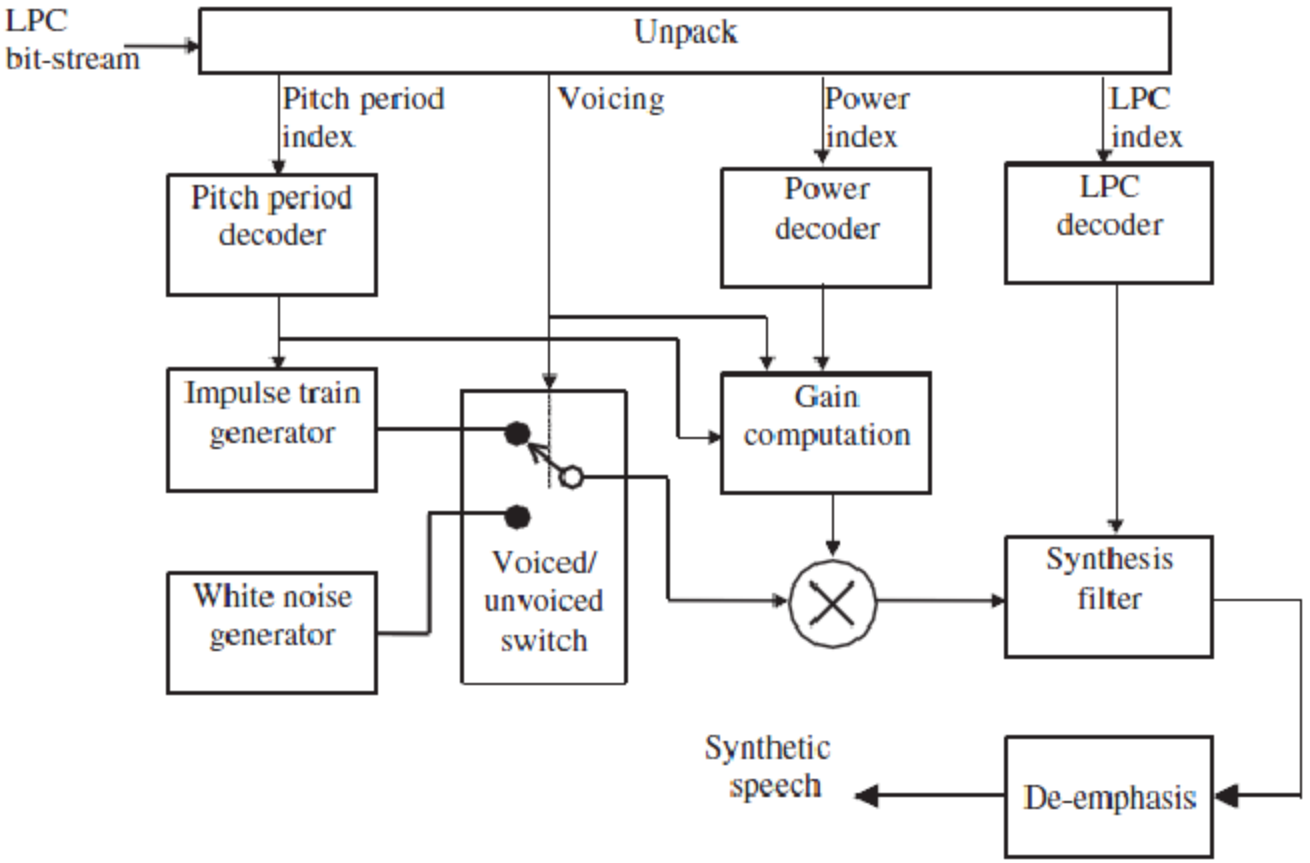
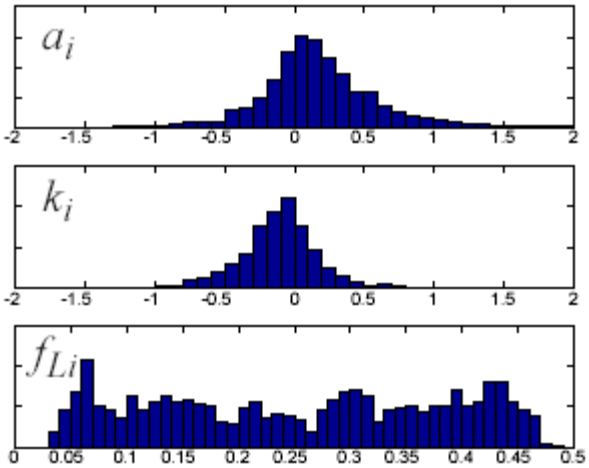


Figure 9.8 Block diagram of the LPC decoder.

Encoding LPC filter parameters

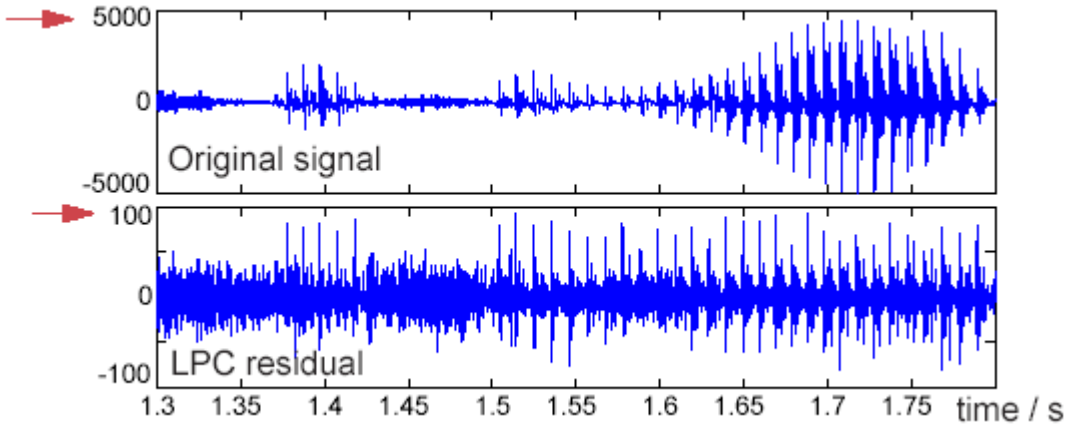
- For ‘communications quality’:
  - 8 kHz sampling (4 kHz bandwidth)
  - ~10th order LPC (up to 5 pole pairs)
  - update every 20-30 ms → 300 - 500 param/s
- Representation & quantization
  - {*ai*} - poor distribution, can’t interpolate
  - reflection coefficients {*ki*}: guaranteed stable
  - log area ratios (LAR) - stable



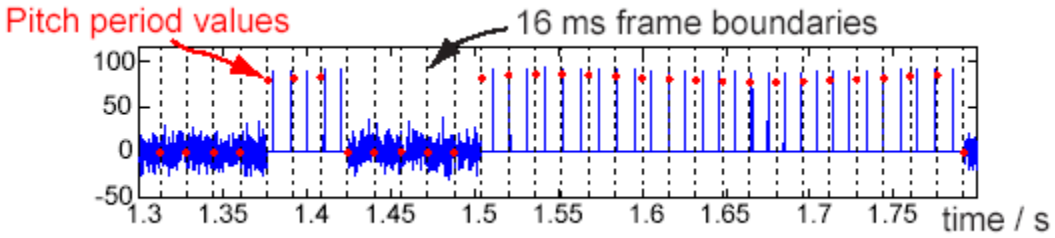
- Bit allocation (filter):
  - GSM (13 kbps):
    8 LARs x 3-6 bits / 20 ms = 1.8 Kbps

Excitation

- Excitation as LPC residual is already better than raw signal:
  - save several bits/sample, still > 32 Kbps



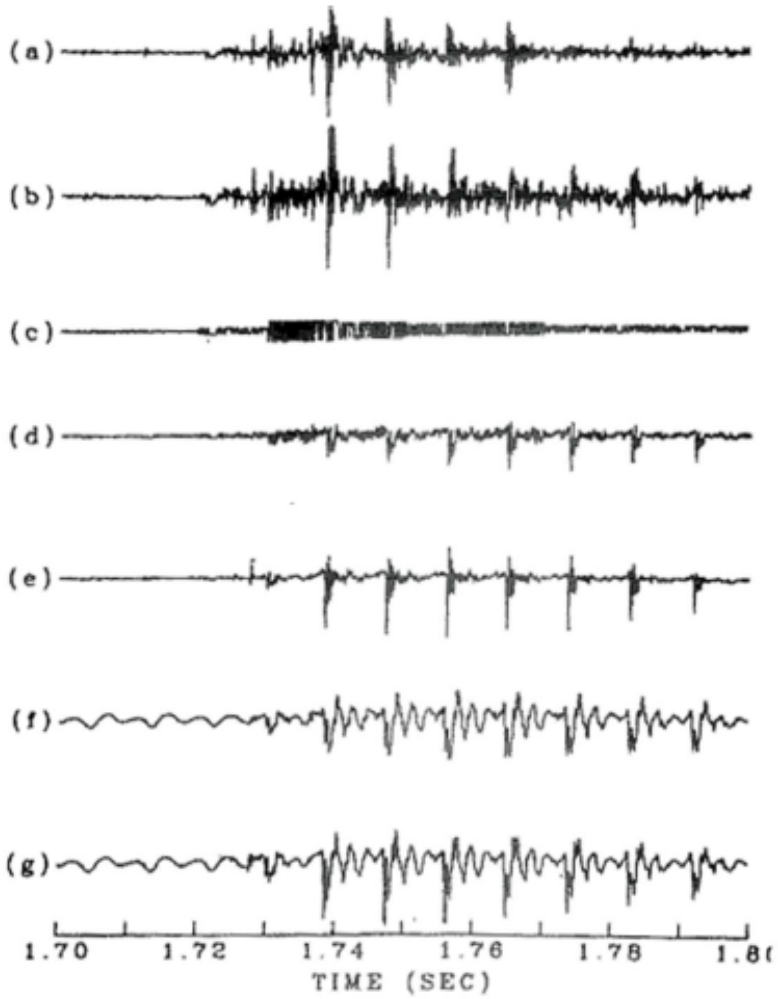
- Crude model: U/V flag + pitch period
  - ~ 7 bits / 5 ms = 1.4 Kbps → LPC10 @ 2.4 Kbps



Bit Rate for LP Coding

- speech and residual sampling rate: F_s=8 kHz
- LP analysis frame rate: $F_\Delta = F_P$ = 50-100 frames/sec
- quantizer stepsize: 6 bits/frame
- predictor parameters:
  - *M* (pitch period): 7 bits/frame
  - pitch predictor coefficients: 13 bits/frame
  - vocal tract predictor coefficients: PARCORs 16-20, 46-50 bits/frame
- prediction residual: 1-3 bits/sample
- total bit rate:
  - $BR = 72 * F_P + F_s$ (minimum)

Two-Level (B=1 bit) Quantizer



Prediction residual

Quantizer input

Quantizer output

Reconstructed pitch

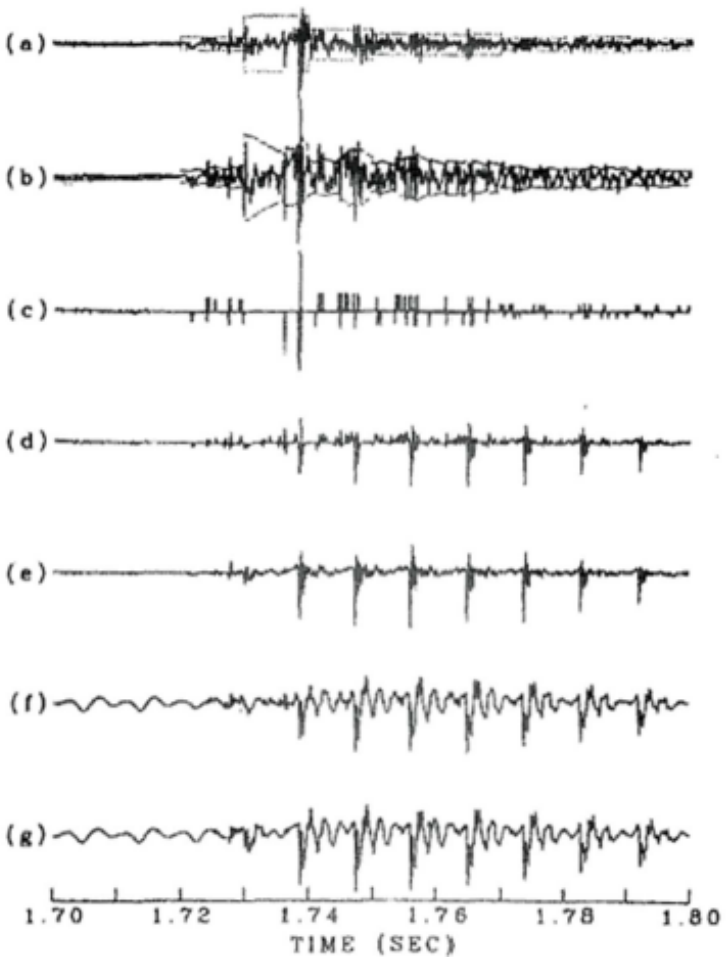
Original pitch residual

Reconstructed speech

Original speech

26

Three-Level Center-Clipped Quantizer



Prediction residual

Quantizer input

Quantizer output

Reconstructed pitch

Original pitch residual

Reconstructed speech

Original speech

27

ANALYSIS BY SYNTHESIS
Codarea folosind Analiza prin Sinteza

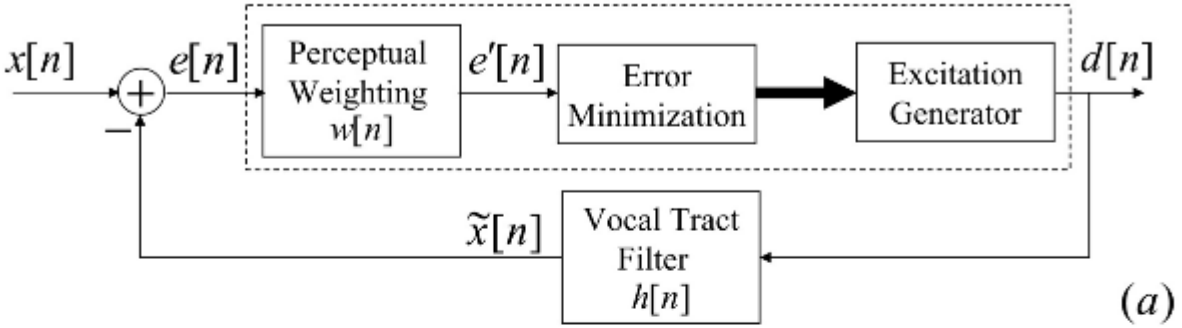
A-b-S Speech Coding

- The key to reducing the data rate of a closed-loop adaptive predictive coder was to force the coded difference signal (the input/excitation to the vocal tract model) to be more easily represented at low data rates while maintaining very high quality at the output of the decoder synthesizer

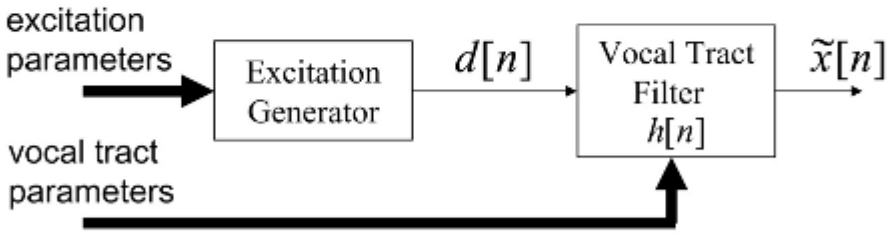
- **Goal**: find a representation of the excitation for the vocal tract filter that produces high quality synthetic output, while maintaining a structured representation that makes it easy to code the excitation at low data rates
- **Solution**: use a set of basis functions which allow you to iteratively build up an optimal excitation function is stages, by adding a new basis function at each iteration in the A-b-S process

Analysis by Synthesis (AbS)

A-b-S Speech Coding



(a)



(b)

Replace quantizer for generating excitation signal with an optimization process (denoted as Error Minimization above) whereby the excitation signal, $d[n]$ is constructed based on minimization of the mean-squared value of the synthesis error, $d[n]=x[n]-\tilde{x}[n]$; utilizes Perceptual Weighting filter.

31

- Basic operation of each loop of closed-loop A-b-S system:
  1. at the beginning of each loop (and only once each loop), the speech signal, $x[n]$, is used to generate an optimum p^{th} order LPC filter of the form:

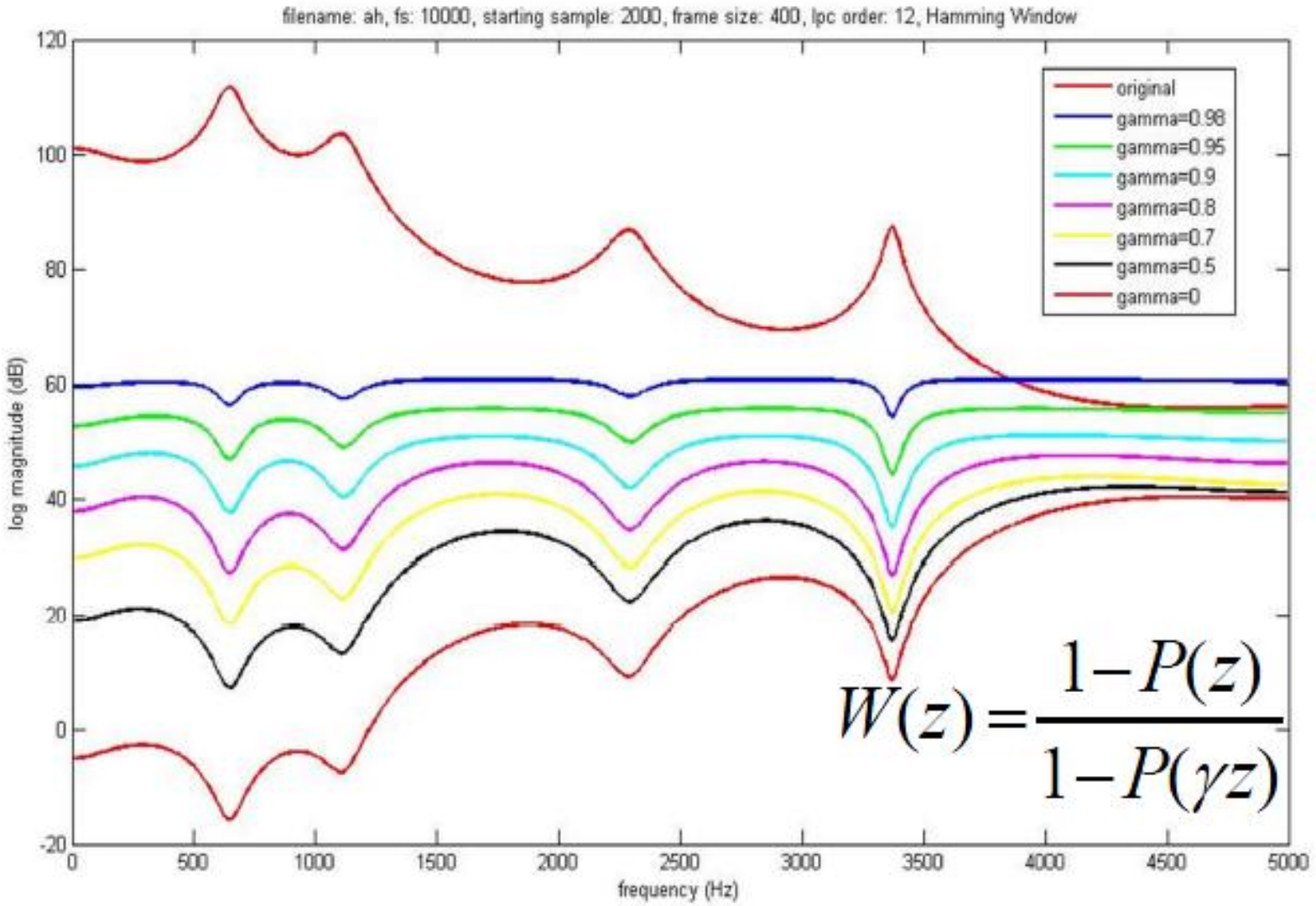
$$H(z) = \frac{1}{1 - P(z)} = \frac{1}{\sum_{i=1}^{p} \alpha_i z^{-1}}$$

  2. the difference signal, $d[n] = x[n] - \tilde{x}[n]$, based on an initial estimate of the speech signal, $\tilde{x}[n]$, is perceptually weighted by a speech-adaptive filter of the form:

$$W(z) = \frac{1 - P(z)}{1 - P(\gamma z)}$$ (see next vugraph)

  3. the error minimization box and the excitation generator create a sequence of error signals that iteratively (once per loop) improve the match to the weighted error signal
  4. the resulting excitation signal, $d[n]$, which is an improved estimate of the actual LPC prediction error signal for each loop iteration, is used to excite the LPC filter and the loop processing is iterated until the resulting error signal meets some criterion for stopping the closed-loop iterations.

Perceptual Weighting Function

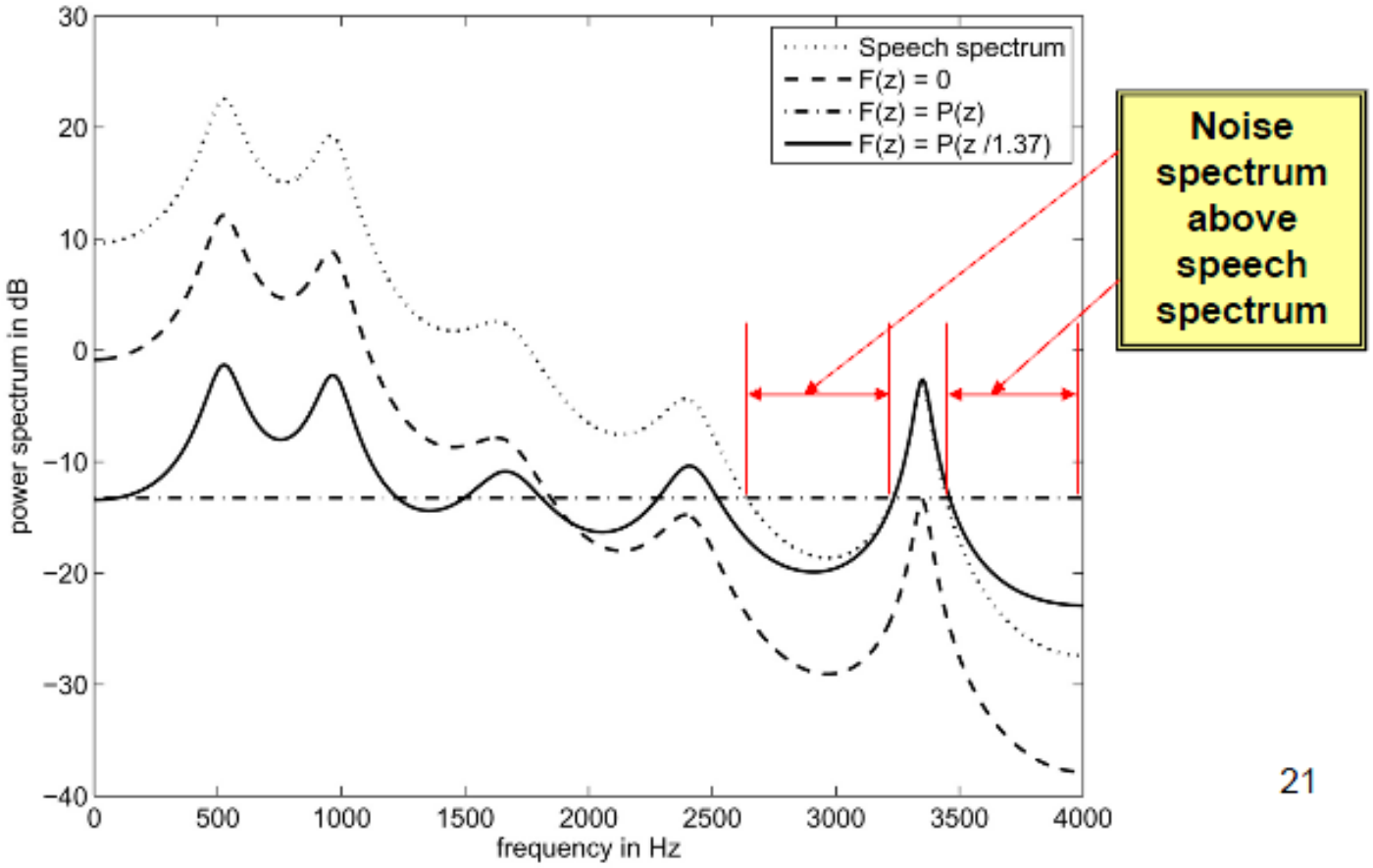


As γ approaches 1, weighting is flat; as γ approaches 0, weighting becomes inverse frequency response of vocal tract.

Noise Shaping Filter

If we assume that the quantization noise has a flat spectrum with noise power of $\sigma_{e'}^2$, then the power spectrum of the shaped noise is of the form:

$$P_{e'}(e^{j2\pi F/F_S}) = \left|\frac{1-F(e^{j2\pi F/F_S})}{1-P(e^{j2\pi F/F_S})}\right| \sigma_{e'}^2$$



21

Implementation of A-B-S Speech Coding

- Assume we are given a set of Q basis functions of the form:

$$\Im_\gamma = \{f_1[n], f_2[n], ..., f_Q[n]\}, \quad 0 \le n \le L-1$$

and each basis function is 0 outside the defining interval.

- At each iteration of the A-b-S loop, we select the basis function from $\Im_\gamma$ that maximally reduces the perceptually weighted mean-squared error, E:

$$E = \sum_{n=0}^{L-1} \left[\left(x[n] - d[n] * h[n] \right) * w[n] \right]^2$$

where $h[n]$ and $w[n]$ are the VT and perceptual weighting filters.

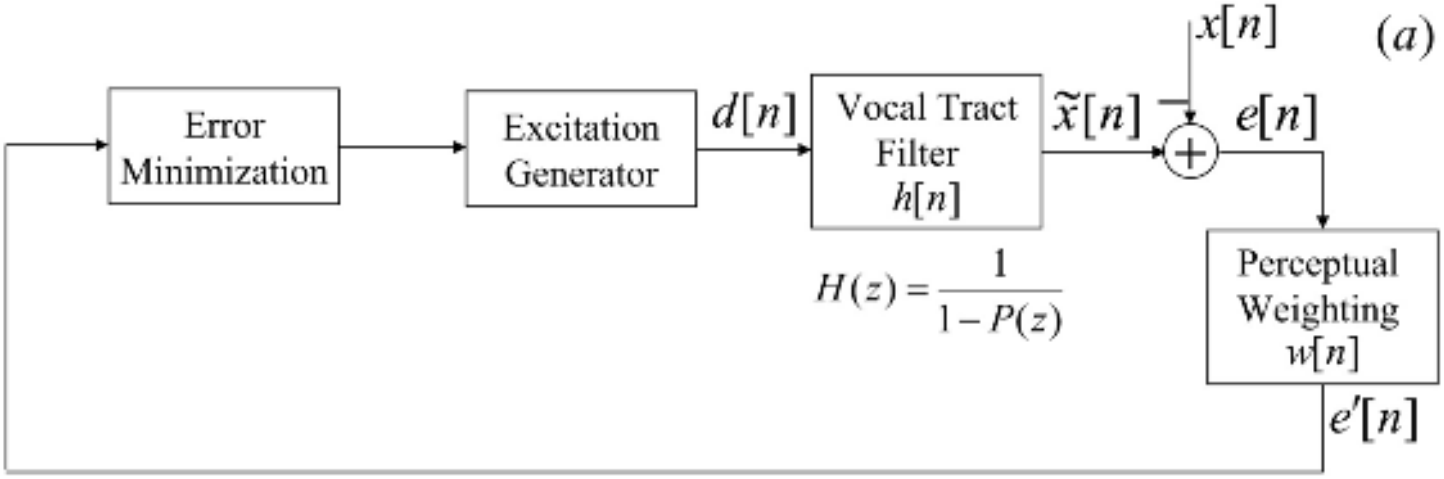
- We denote the optimal basis function at the k^{th} iteration as $f_{\gamma_k}[n]$, giving the excitation signal $d_k[n] = \beta_k f_{\gamma_k}[n]$ where β_k is the optimal weighting coefficient for basis function $f_{\gamma_k}[n]$.

- The A-b-S iteration continues until the perceptually weighted error falls below some desired threshold, or until a maximum number of iterations, N, is reached, giving the final excitation signal, $d[n]$, as:

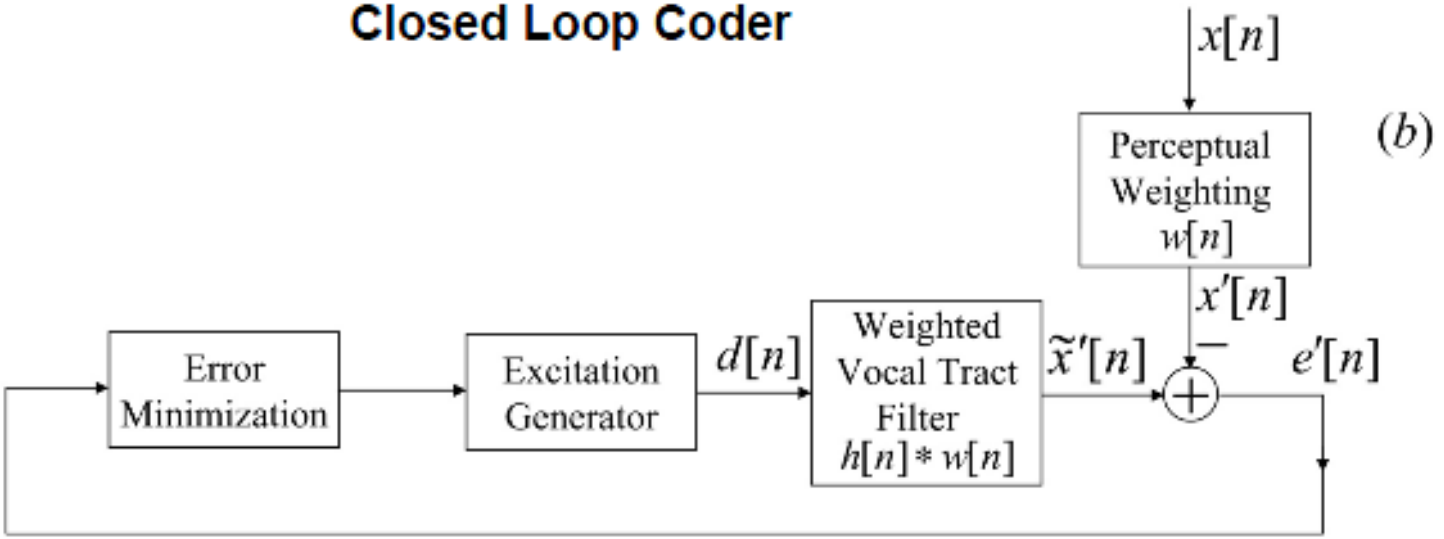
$$d[n] = \sum_{k=1}^{N} \beta_k f_{\gamma_k}[n]$$

36

Implementation of A-B-S Speech Coding



Closed Loop Coder



Reformulated Closed Loop Coder

Analysis-by-Synthesis Coding

- Multipulse linear predictive coding (MPLPC)

$$f_\gamma[n] = \delta[n-\gamma] \qquad 0 \le \gamma \le Q-1 = L-1$$

B. S. Atal and J. R. Remde, "A new model of LPC excitation…," *Proc. IEEE Conf. Acoustics, Speech and Signal Proc.*, 1982.

- Code-excited linear predictive coding (CELP)

$f_\gamma[n]$ = vector of white Gaussian noise, $1 \le \gamma \le Q = 2^M$

M. R. Schroeder and B. S. Atal, "Code-excited linear prediction (CELP)," *Proc. IEEE Conf. Acoustics, Speech and Signal Proc.*, 1985.

- Self-excited linear predictive vocoder (SEV)

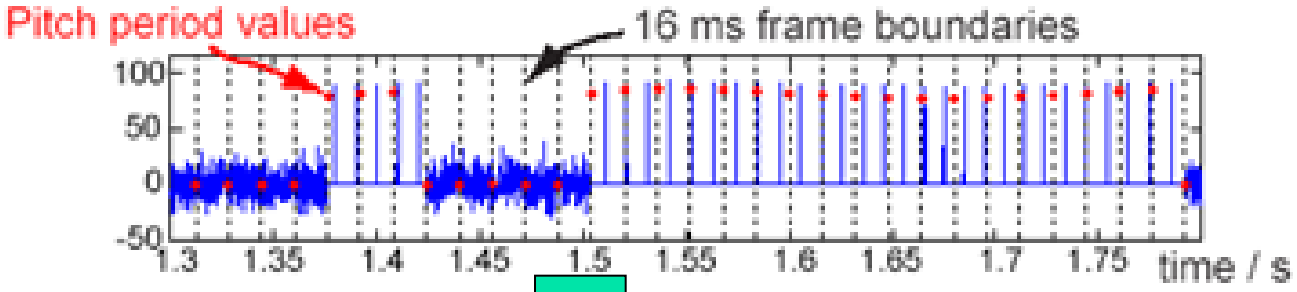
$f_\gamma[n] = d[n-\gamma], \Gamma_1 \le \gamma \le \Gamma_2$ – shifted versions of previous excitation source

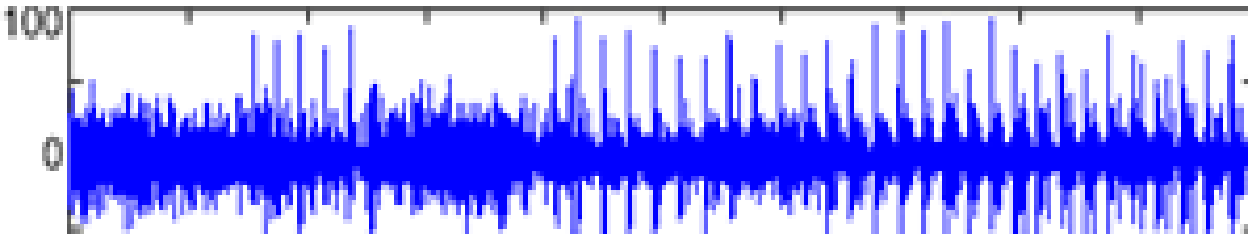
R. C. Rose and T. P. Barnwell, "The self-excited vocoder," *Proc. IEEE Conf. Acoustics, Speech and Signal Proc.*, 1986.

CODAREA MPE (Multiple Pulse Excitation)

MULTIPULSE EXCITATION VOCODER

Source signal is modelled by MULPIPLE PULSES
-Position
-Amplitude (difficult tasks)

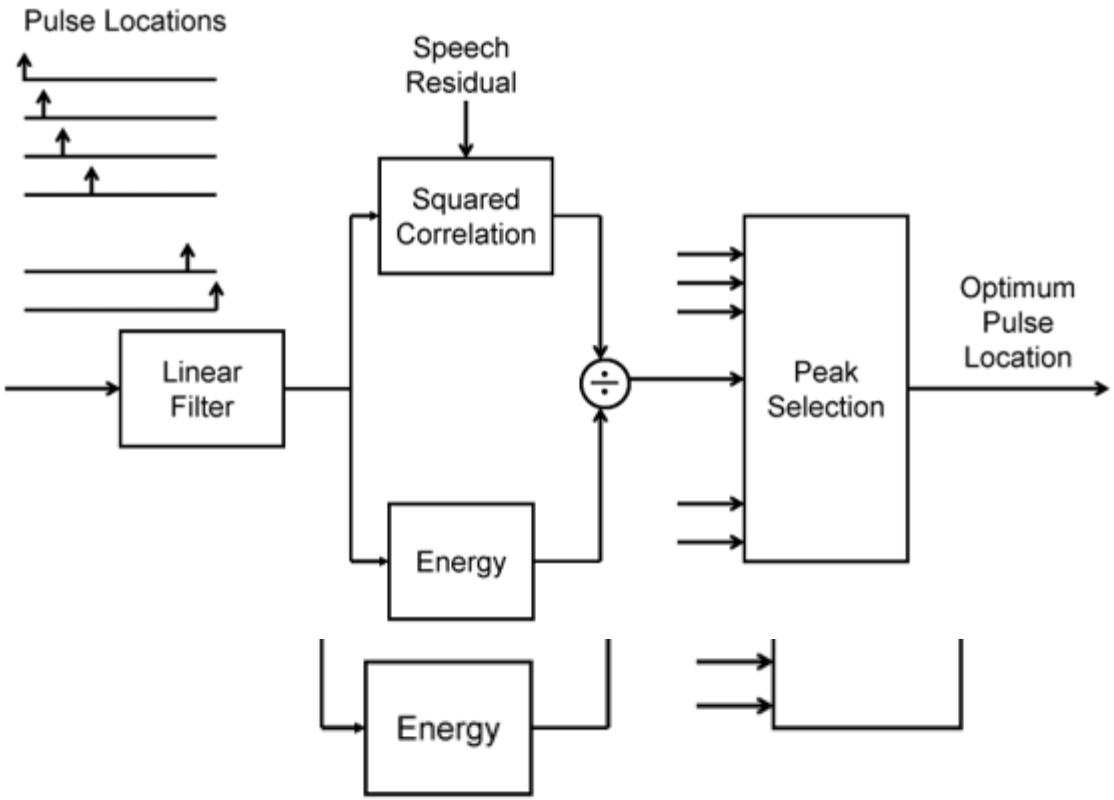




MPE – pulse locations

- Multipulse uses impulses as the basis functions; thus the basic error minimization reduces to:

$$E = \sum_{n=0}^{L-1} \left(x[n] - \sum_{k=1}^{N} \beta_k h[n-\gamma_k] \right)^2$$



44

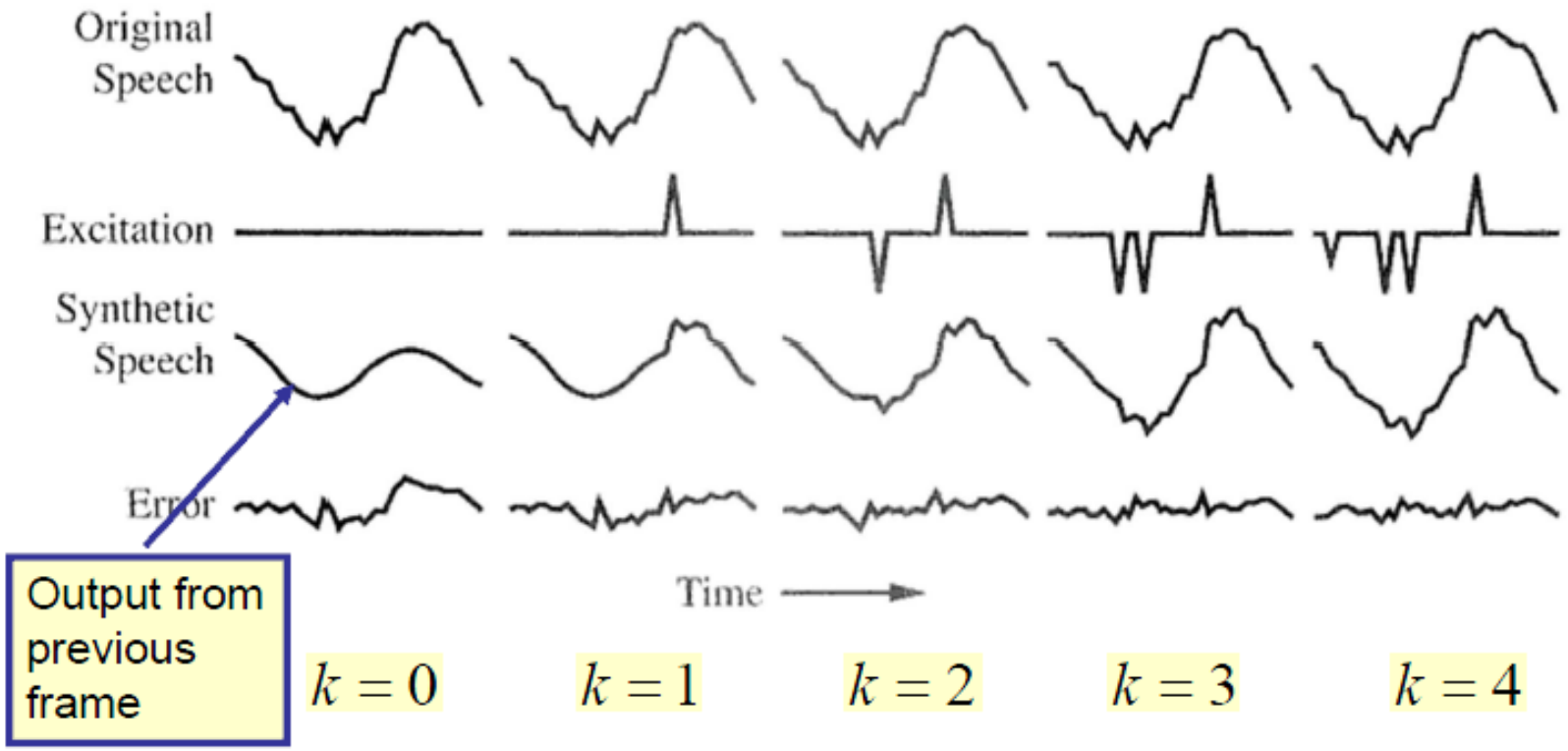
44

Iterative Solution for Multipulse

1. find best β_1 and γ_1 for single pulse solution
2. subtract out the effect of this impulse from the speech waveform and repeat the process
3. do this until desired minimum error is obtained
   - 8 impulses each 10 msec gives synthetic speech that is perceptually close to the original

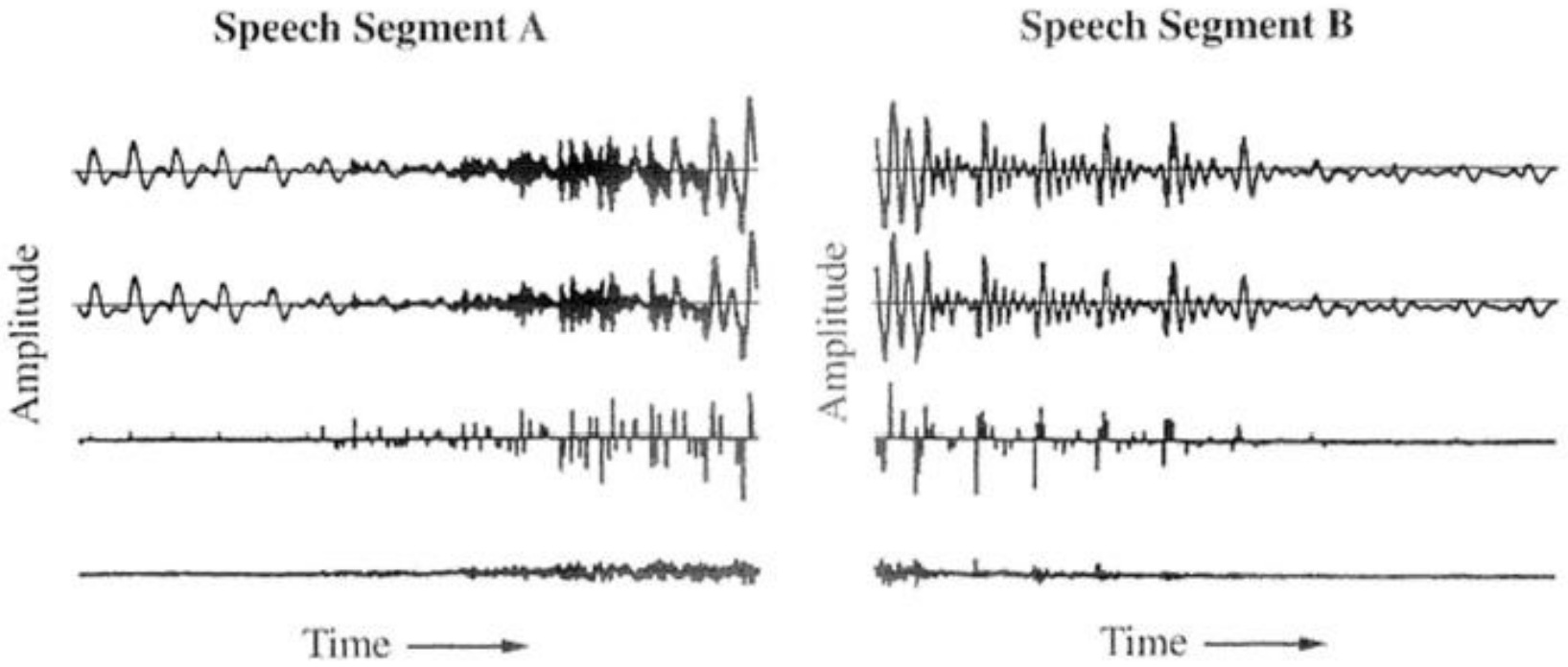
MPE – encoder

Multipulse Analysis



MPE – encoder

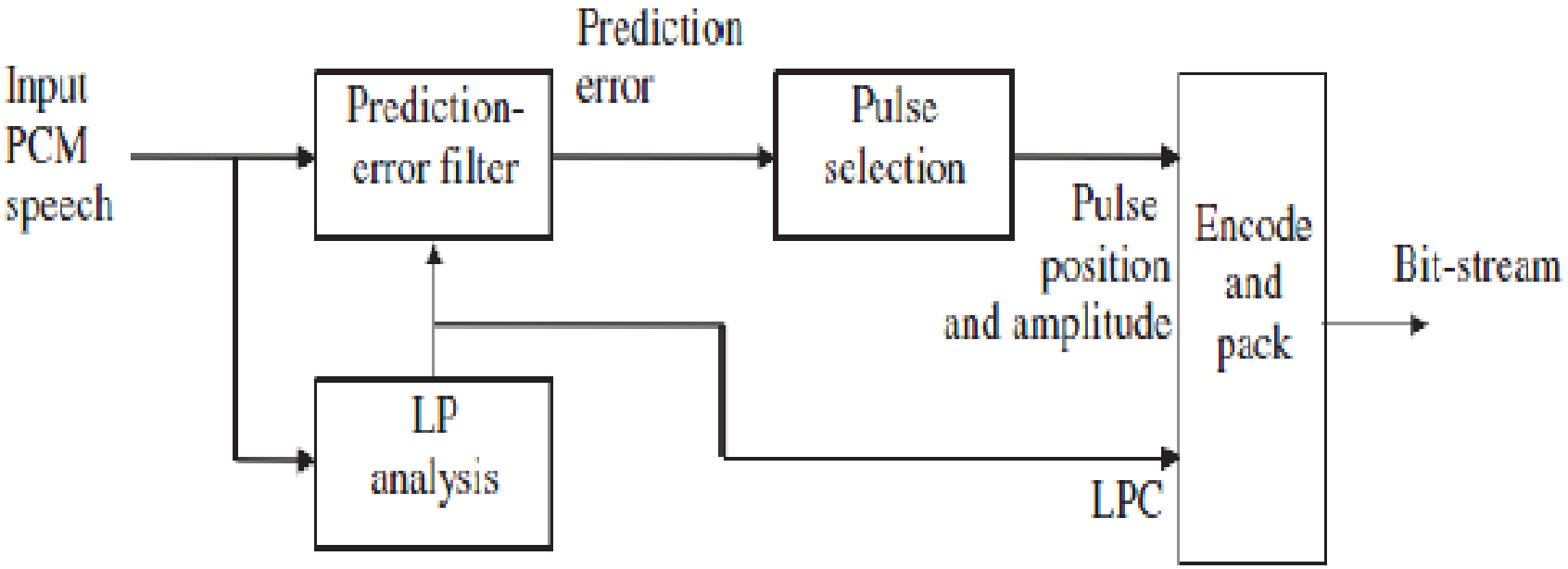
Examples of Multipulse LPC



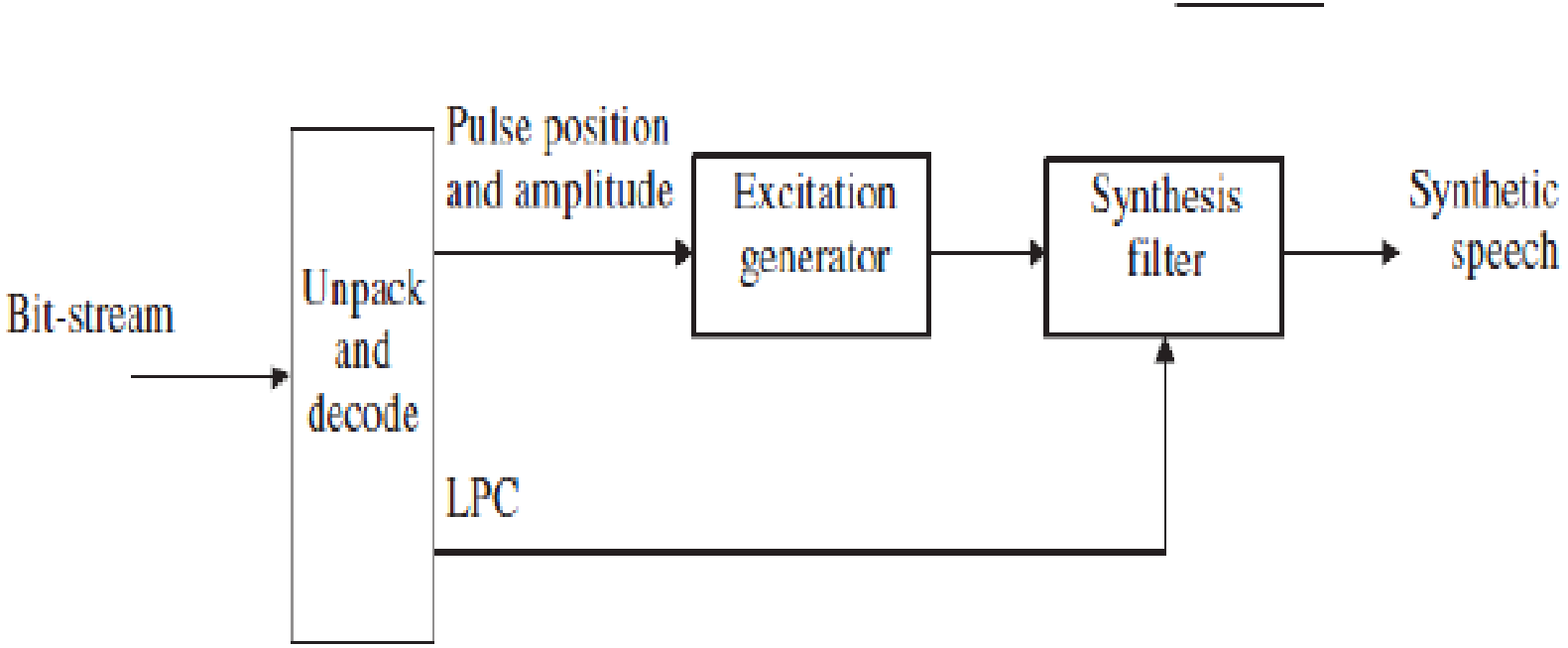
Coding of MP-LPC

- 8 impulses per 10 msec => 800 impulses/sec X 9 bits/impulse => 7200 bps
- need 2400 bps for $A(z)$ => total bit rate of 9600 bps
- code pulse locations differentially ($\Delta_i = N_i - N_{i-1}$) to reduce range of variable
- amplitudes normalized to reduce dynamic range

MPE – encoder


Input
PCM
speech
Prediction-
error filter
Prediction
error
Pulse
selection
Pulse
position
and amplitude
Encode
and
pack
Bit-stream
LP
analysis
LPC

MPE decoder

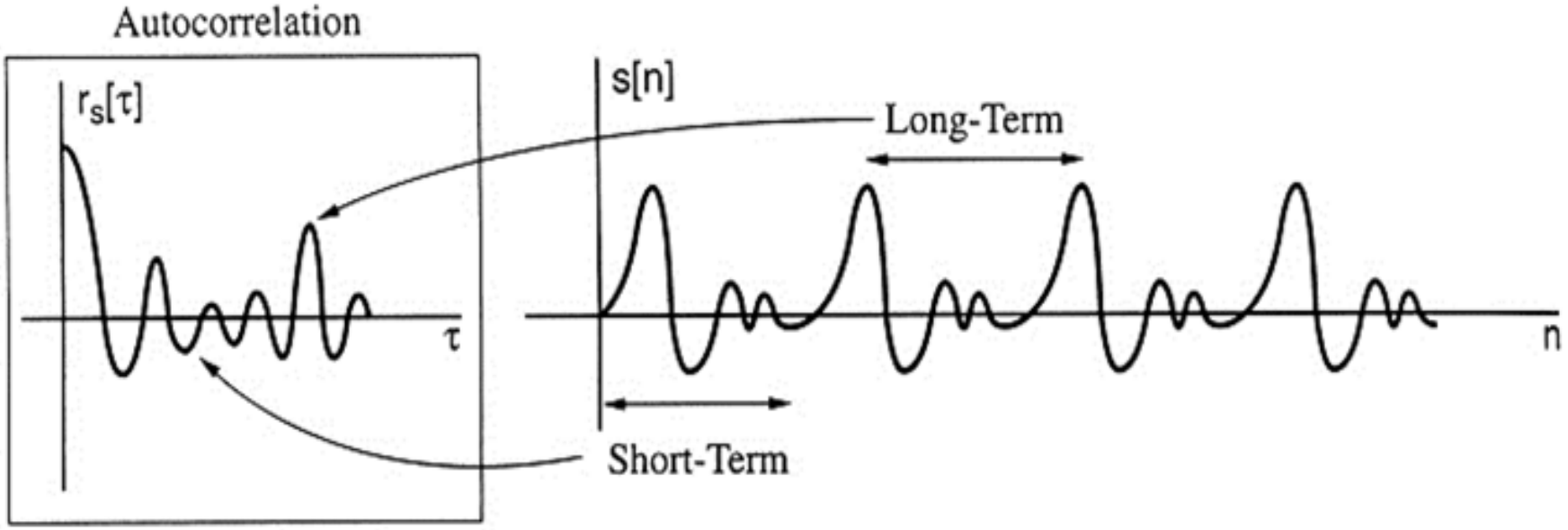

Bit-stream
Unpack and decode
Pulse position and amplitude
LPC
Excitation generator
Synthesis filter
Synthetic speech

MPLPC with LT Prediction

- basic idea is that primary pitch pulses are correlated and predictable over consecutive pitch periods, i.e.,

$$s[n] \approx s[n-M]$$

- break correlation of speech into short term component (used to provide spectral estimates) and long term component (used to provide pitch pulse estimates)
- first remove short-term correlation by short-term prediction, followed by removing long-term correlation by long-term predictions



Short Term Prediction Error Filter

- prediction error filter

$$\hat{A}(z) = 1 - P(z) = 1 - \sum_{k=1}^{p} \alpha_k z^{-k}$$

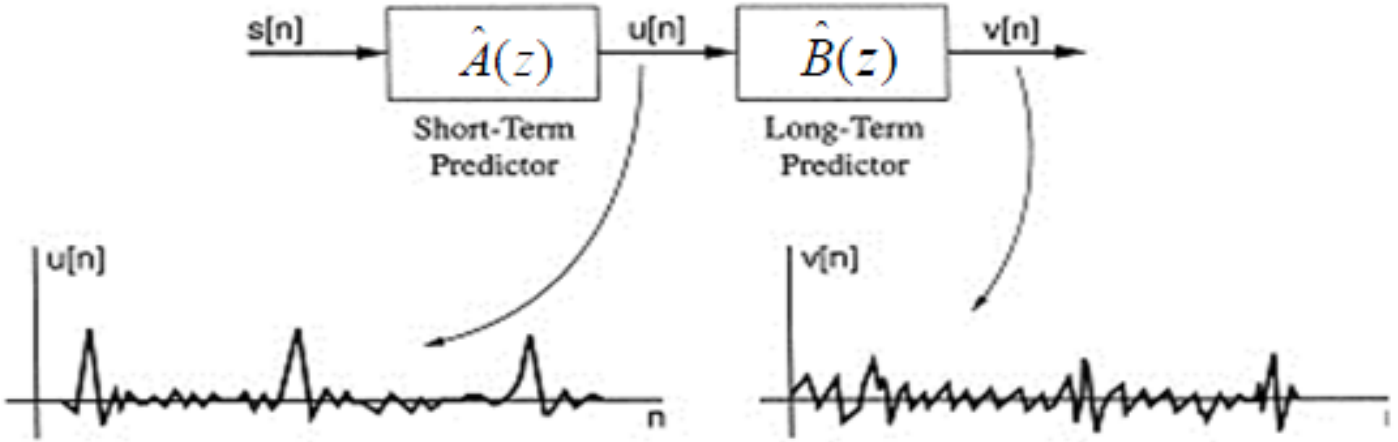
- short term residual, $u(n)$, includes primary pitch pulses that can be removed by long-term predictor of the form

$$\hat{B}(z) = 1 - bz^{-M}$$

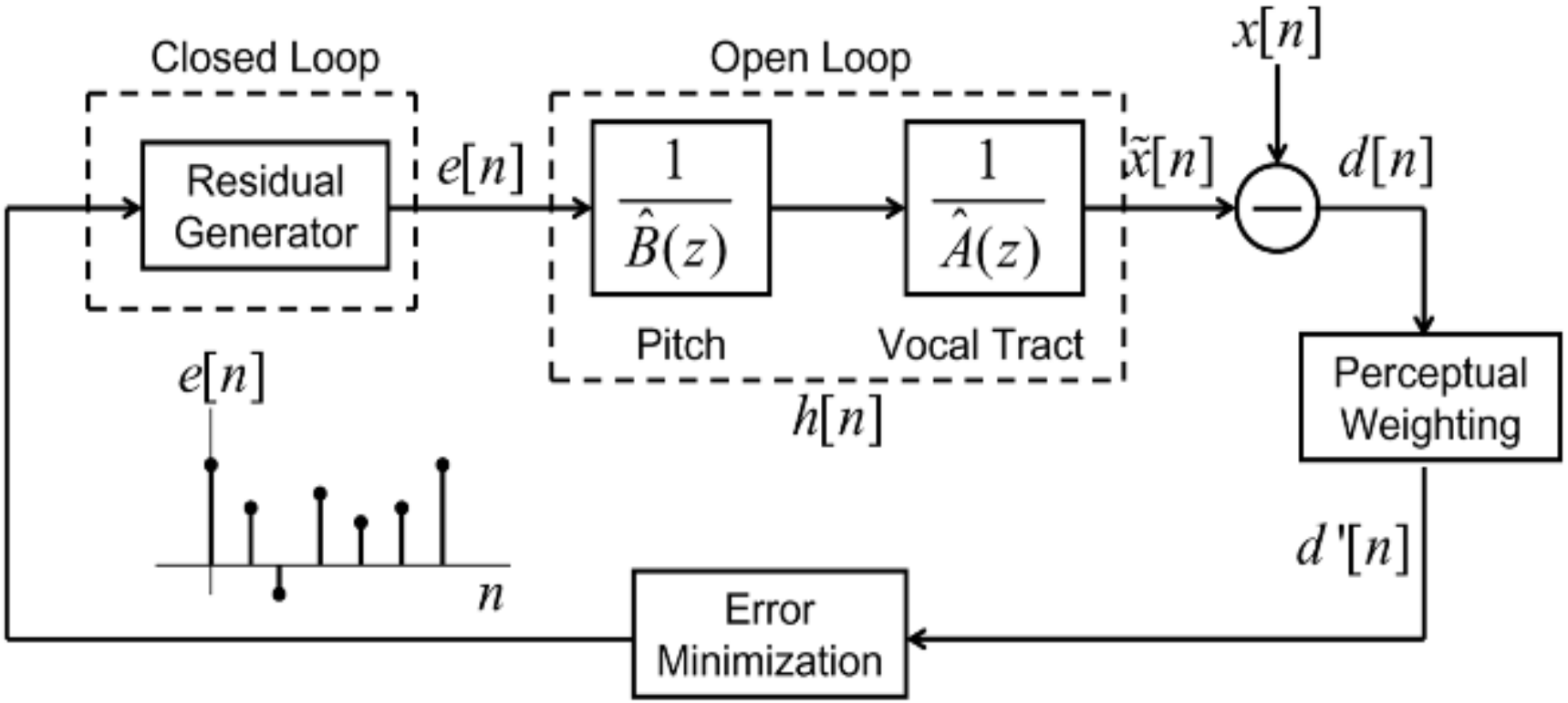
- giving

$$v(n) = u(n) - bu(n - M)$$

- with fewer large pulses to code than in $u(n)$



Analysis-by-Synthesis



- impulses selected to represent the output of the long term predictor, rather than the output of the short term predictor
- most impulses still come in the vicinity of the primary pitch pulse

=> result is high quality speech coding at 8-9.6 Kbps

Applications of Speech Coders

- network-64 Kbps PCM (8 kHz sampling rate, 8-bit log quantization)
- international-32 Kbps ADPCM
- teleconferencing-16 Kbps LD-CELP
- wireless-13, 8, 6.7, 4 Kbps CELP-based coders
- secure telephony-4.8, 2.4 Kbps LPC-based coders (MELP)
- VoIP-8 Kbps CELP-based coder
- storage for voice mail, answering machines, announcements-16 Kbps LC-CELP

Speech Coder Attributes

- bit rate-2400 to 128,000 bps
- quality-subjective (MOS), objective (*SNR*, intelligibility)
- complexity-memory, processor
- delay-echo, reverberation; block coding delay, processing delay, multiplexing delay, transmission delay-~100 msec
- telephone bandwidth-200-3200 Hz, 8kHz sampling rate
- wideband speech-50-7000 Hz, 16 kHz sampling rate

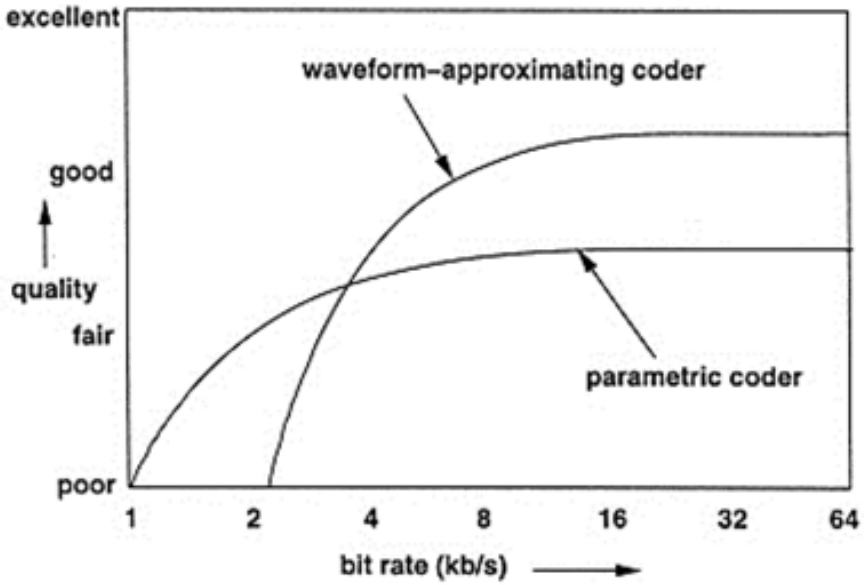
Network Speech Coding Standards

Coder	*Type*	*Rate*	*Usage*
G.711	companded PCM	64 Kbps	toll
G.726/727	ADPCM	16-40 Kbps	toll
G.722	SBC/ADPCM	48, 56,64 Kbps	wideband
G.728	LD-CELP	16 Kbps	toll
G.729A	CS-ACELP	8 Kbps	toll
G.723.1	MPC-MLQ & ACELP	6.3/5.3 Kbps	toll

93

Speech Coding Quality Evaluation

- 2 types of coders
  - waveform approximating-PCM, DPCM, ADPCM-coders which produce a reconstructed signal which converges toward the original signal with decreasing quantization error
  - parametric coders (model-based)-SBC, MP-LPC, LPC. MB-LPC, CELP-coders which produce a reconstructed signal which does not converge to the original signal with decreasing quantization error



- **waveform coder converges to quality of original speech**
- **parametric coder converges to model-constrained maximum quality (due to the model inaccuracy of representing speech)**

97

Factors on Speech Coding Quality

- **talker and language dependency** - especially for parametric coders that estimate pitch that is highly variable across men, women and children; language dependency related to sounds of the language (e.g., clicks) that are not well reproduced by model-based coders
- **signal levels** - most waveform coders designed for speech levels normalized to a maximum level; when actual samples are lower than this level, the coder is not operating at full efficiency causing loss of quality
- **background noise** - including babble, car and street noise, music and interfering talkers; levels of background noise varies, making optimal coding based on clean speech problematic
- **multiple encodings** - tandem encodings in a multi-link communication system, teleconferencing with multiple encoders
- **channel errors** - especially an issue for cellular communications; errors either random or bursty (fades)-redundancy methods oftern used
- **non-speech sounds** - e.g., music on hold, dtmf tones; sounds that are poorly coded by the system

98

Measures of Speech Coder Quality

$$SNR = 10\log_{10}\frac{\sum_{n=0}^{N-1}[s[n]]^2}{\sum_{n=0}^{N-1}[s[n]-\hat{s}[n]]^2}, \quad \text{over whole signal}$$

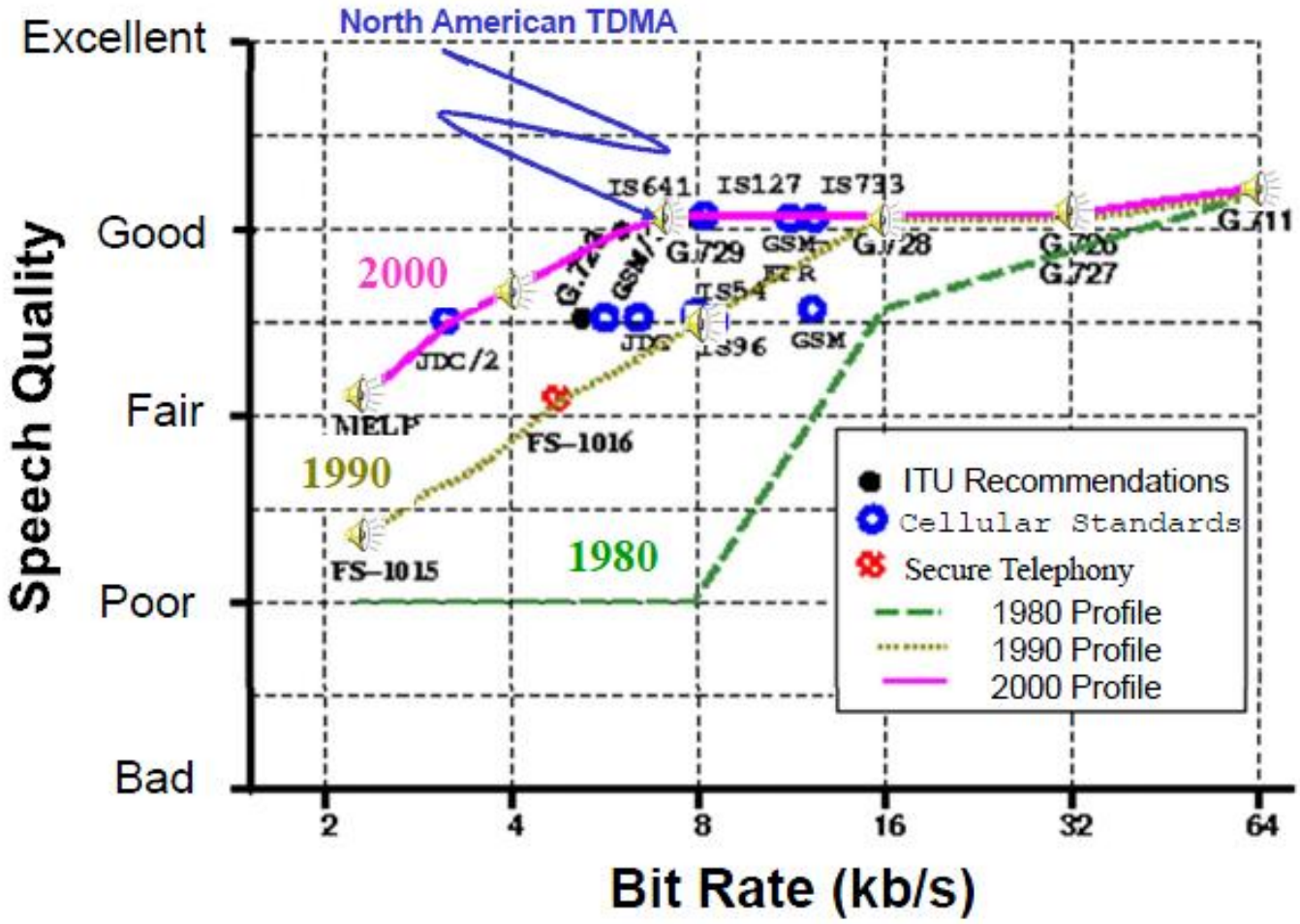
$$SNR_{seg} = \frac{1}{K}\sum_{k=1}^{K} SNR_k \quad \text{over frames of 10-20 msec}$$

- good primarily for waveform coders

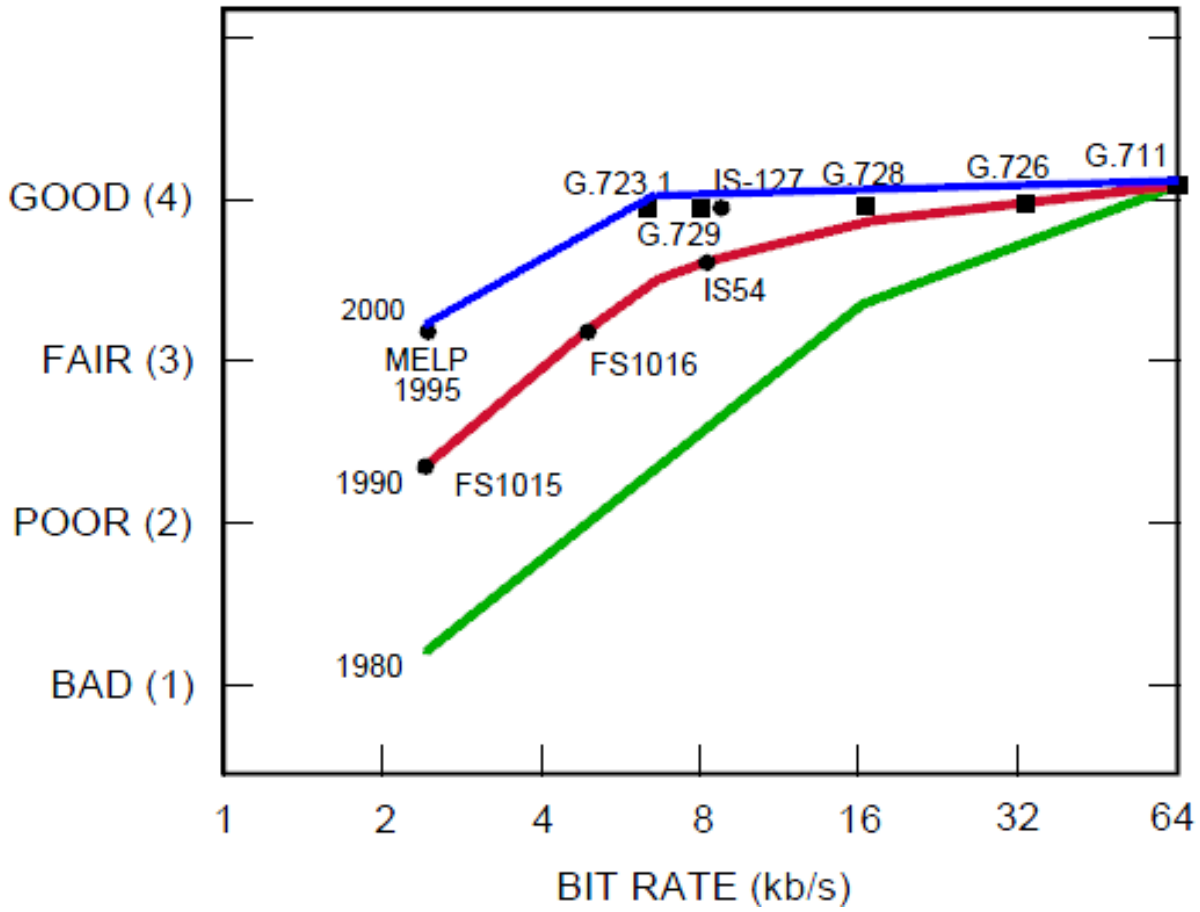
Measures of Speech Coder Quality

- Intelligibility-Diagnostic Rhyme Test (DRT)
  - compare words that differ in leading consonant
  - identify spoken word as one of a pair of choices
  - high scores (~90%) obtained for all coders above 4 Kbps
- Subjective Quality-Mean Opinion Score (MOS)
  - 5 excellent quality
  - 4 good quality
  - 3 fair quality
  - 2 poor quality
  - 1 bad quality
- MOS scores for high quality wideband speech (~4.5) and for high quality telephone bandwidth speech (~4.1)

Evolution of Speech Coder Performance



Speech Coder Subjective Quality



MPE decoder

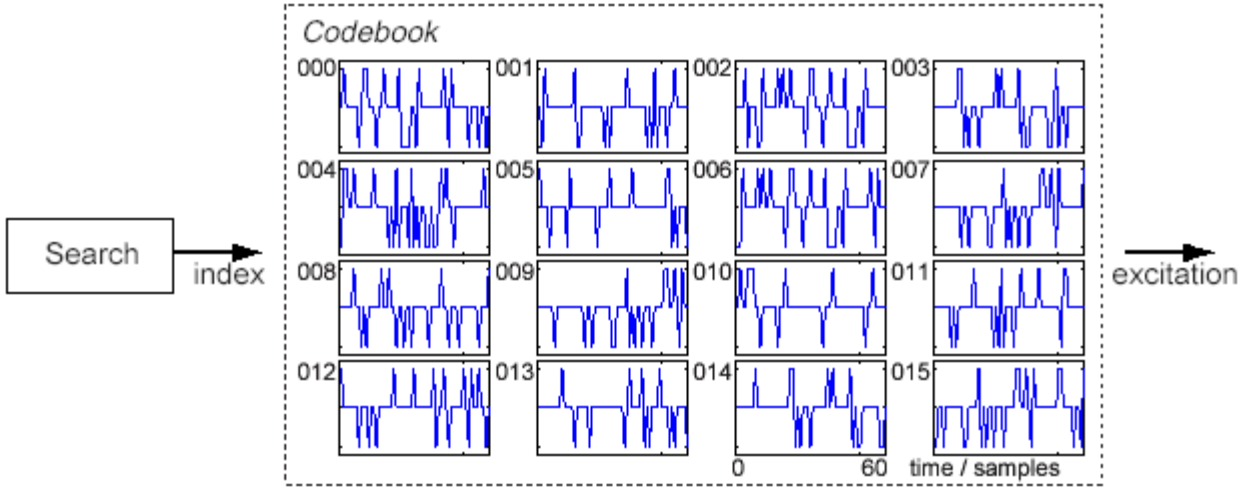
MPE decoder

CELP

- **Code excited linear predictive** (CELP) speech coding.
- White noise input does not give satisfactory results:
  - the residue sequence still contains important information for speech synthesis
  - it is necessary to send the residue to receiving end too.
- To save space, use vector quantization (VQ) technique to encode the residue sequence
  - Hence the name “code excited”.
- In CELP, each code book is a linear vector containing 0 or ±1
  - each code word length is 60 samples
  - successive code words are overlapped by 58 samples
  - a linear search is performed to find the best code words as input to the LPC model.

CELP

- Represent excitation with codebook
  e.g. 512 sparse excitation vectors
  - linear search for minimum weighted error?

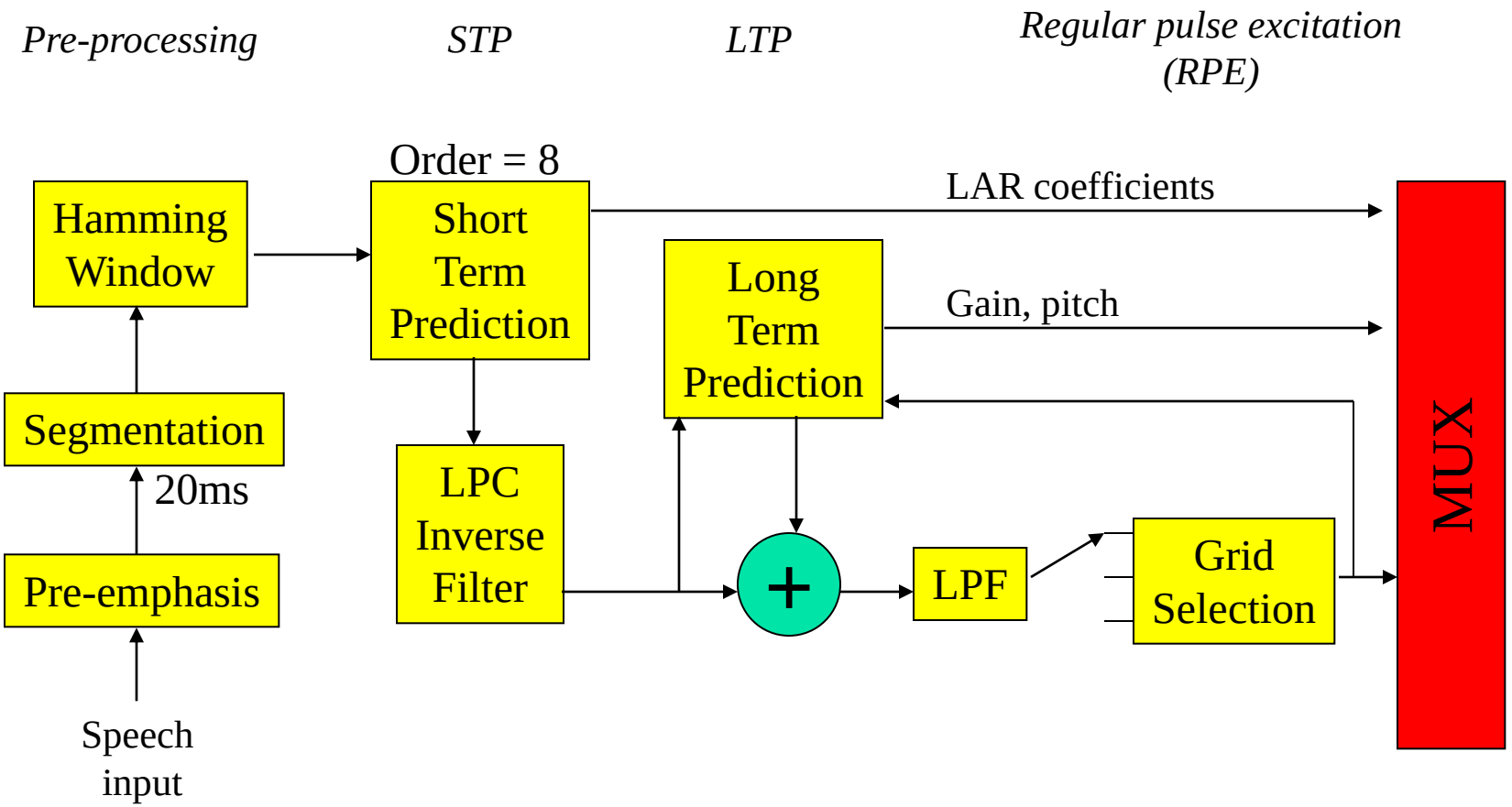


- **FS1016 4.8 Kbps CELP (30ms frame = 144 bits):**

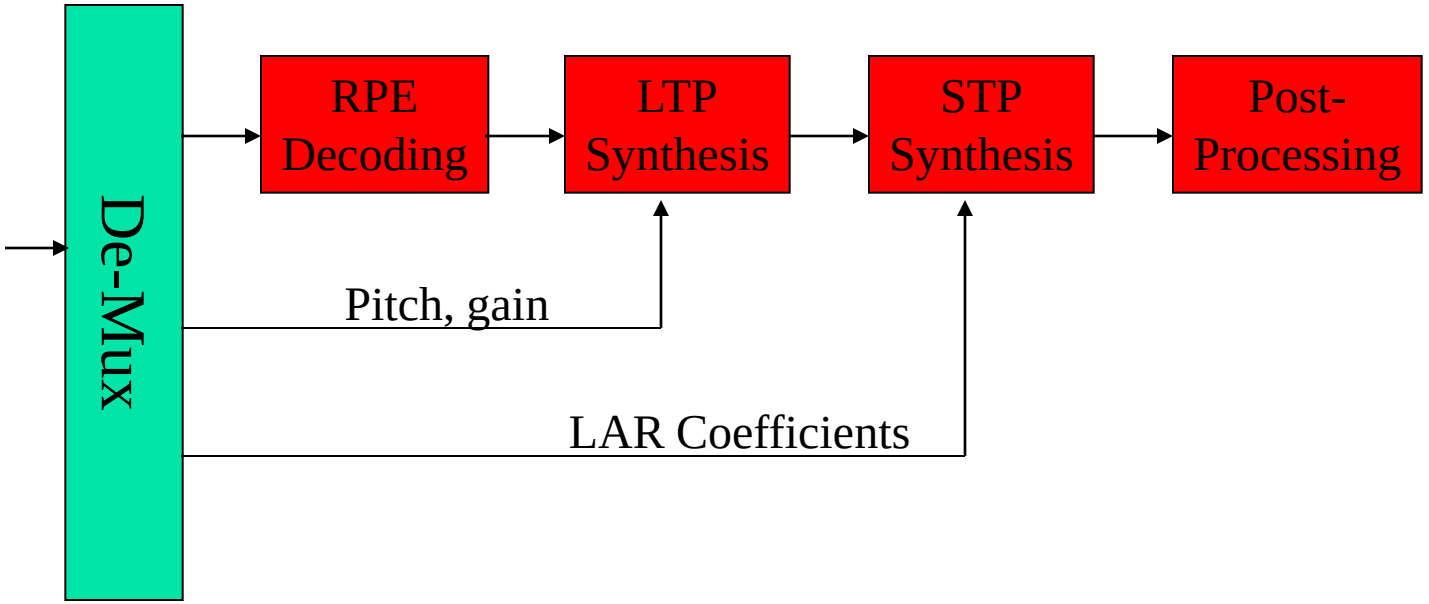
10 LSPs	4x4 + 6x3 bits =	34 bits
Pitch delay	4 x 7 bits =	28 bits
Pitch gain	4 x 5 bits =	20 bits
Codebk index	4 x 9 bits =	36 bits
Codebk gain	4 x 5 bits =	20 bits

138 bits

GSM Speech Encoder



GSM Decoding

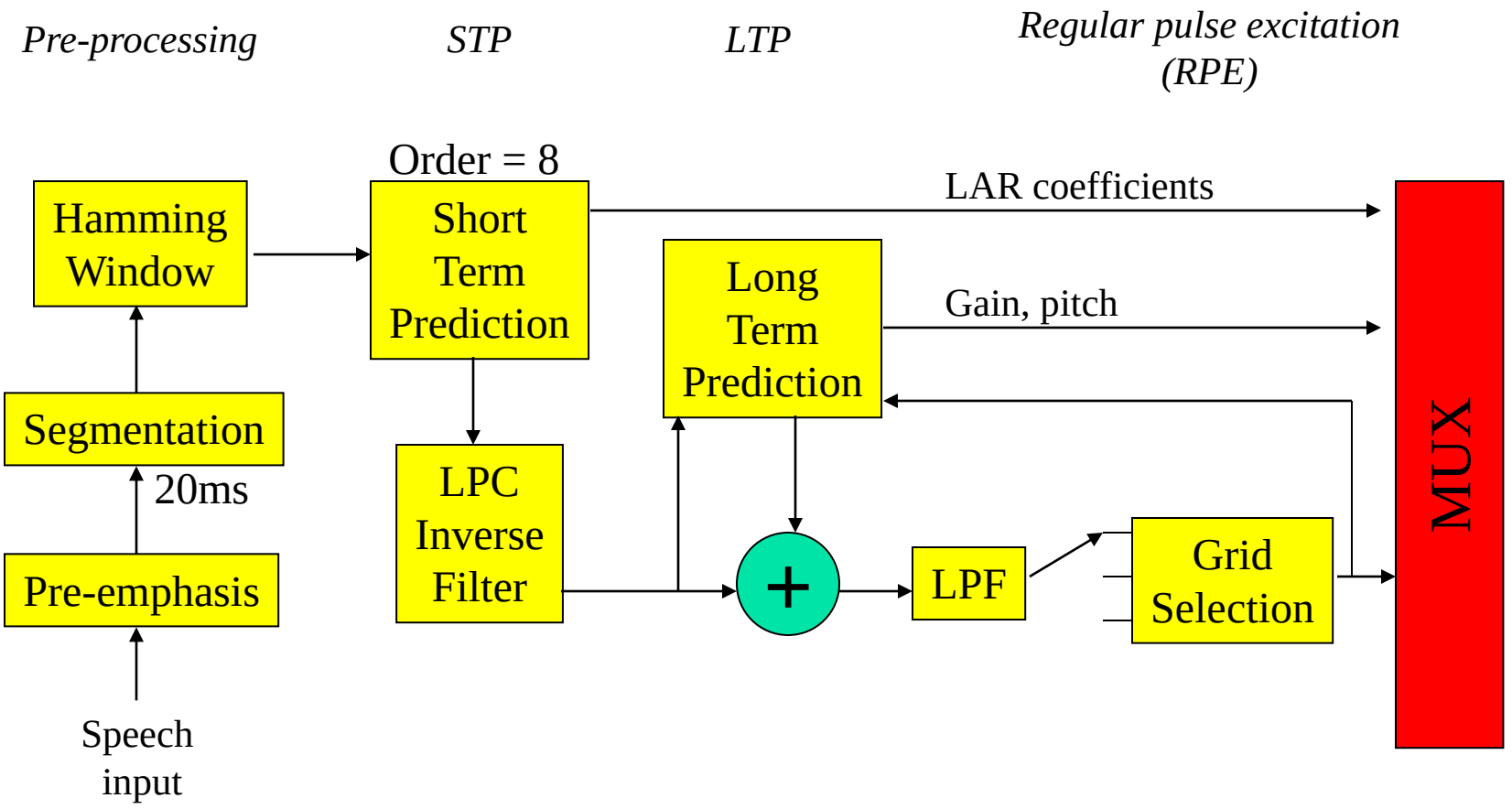


Implementation Issues

- Tasks:
  - LPC analysis filter to calculate the coefficients
  - Long term prediction for pitch analysis need to find delay D and gain
  - VQ search during CELP encoding – Most time consuming
  - FIR filtering for pre- and post processing
- Often implemented in DSP chips for embedded applications (e.g. cell phone).
- The parameter quantization part needs bit-level operation.

CODAREA SEMNALULUI VOCAL IN STANDARDUL GSM

GSM Speech Encoder



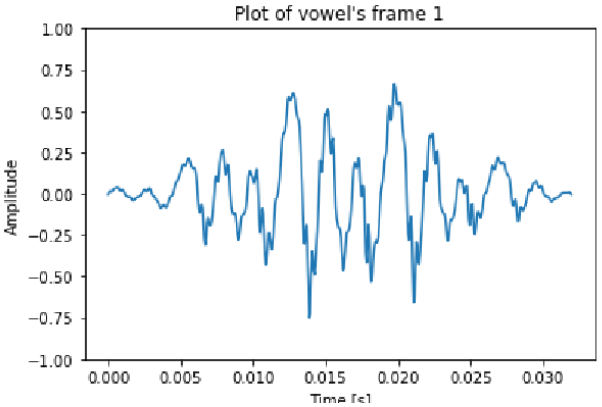
Etape de prelucrare – cadre si subcadre

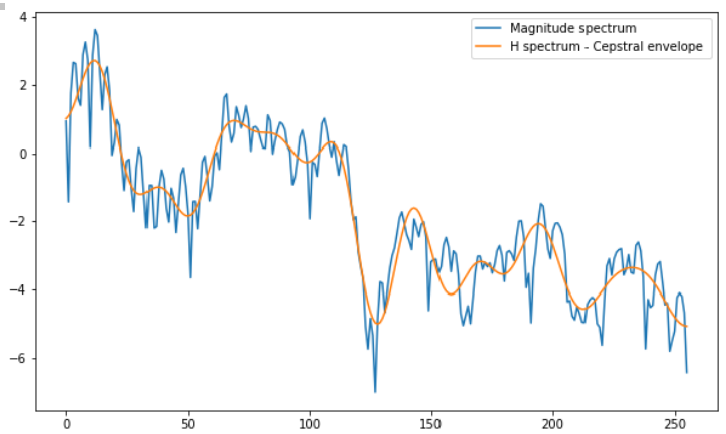
- Semnal intrare
  - Banda 0 – 4KHz
  - Fe = 8KHz, 8 biti / esantion
  - Debit = 64 Kbps
- Pre-accentuare:
  - H(z) = 1 – 0,98*Z^(-1)
  - Pre-accentuare a frecventelor inalte
- Impartire in cadre si subcadre:
  - Fereastra Hamming pe 160 de esantioane (Fe = 8KHz, 20 ms)
  - 1 cadru, se imparte in 4 subcadre de cate 40 de esantioane

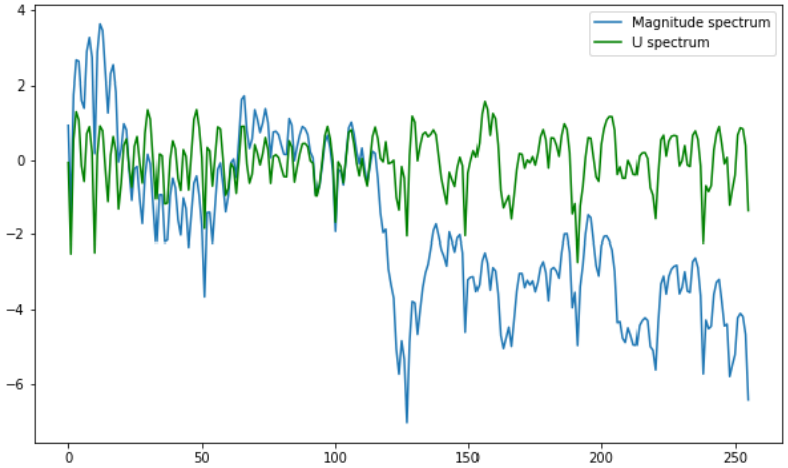
Etape de prelucrare – (I) LPC (coeficientii LAR)

- LPC (Analiza pe termen scurt – Short Term Prediction)
  - Se aplica la nivel de cadru de 160 de esantioane
  - Ordin de predictie p =8
  - Se extrag 8 coeficienti de predictie a1, … a8
  - Se convertesc in coeficienti LAR1, LAR2, … LAR8
  - Se codeaza astfel: LAR8-7 (6 biti), LAR6-5 (5 biti), LAR4-3 (4 biti), LAR2-1 (3 biti). Total 36 biti.
  - Efect: este codata anvelopa spectrala H(z)
  - Din semnalul S(z) se va putea determina sursa (sau eroarea de predictie), prin filtrare inversa:
  - S(z) = E(z) x H(z) ➔ E(z) = S(z) * (1/ H(z))

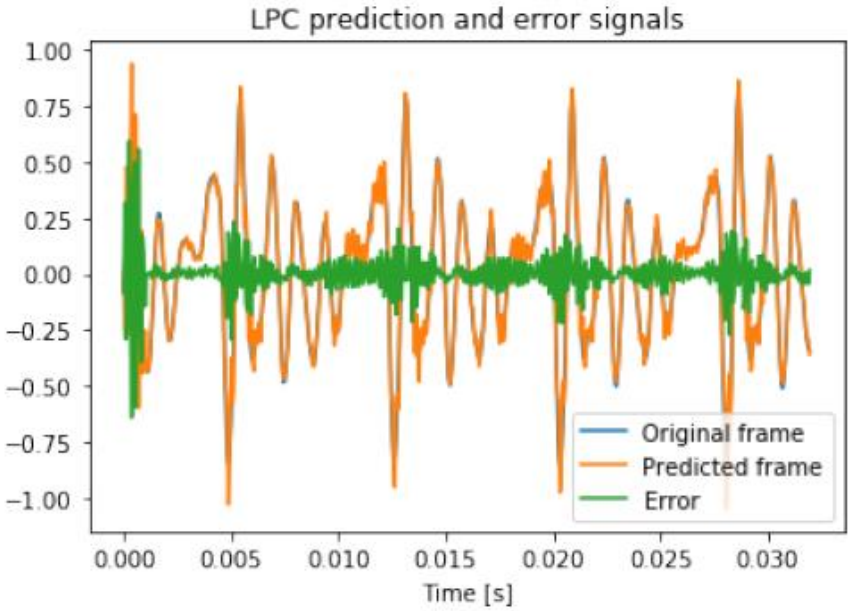
Etape – semnal, spectru LPC, spectru sursa

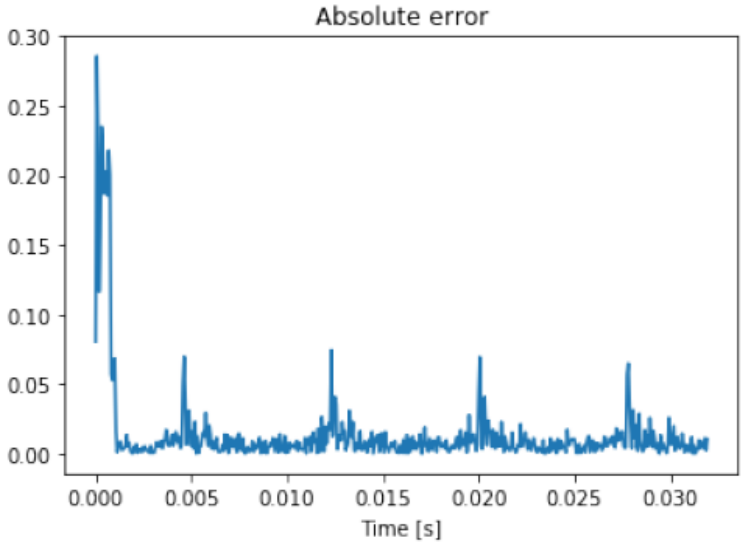






Etape – semnal, eroare de predictie



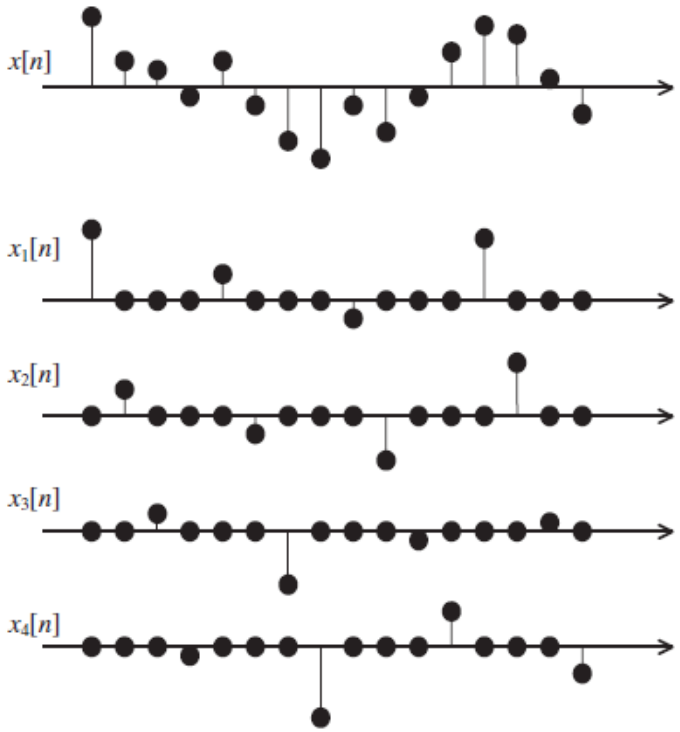


Etape de prelucrare – (II) analiza LTP

- LTP (Long Term Prediction – Analiza pe Termen Lung)
  - Se aplica pe fiecare SUBCADRU de cate 40 de esantioare (+ urmatoarele 2 subcadre, pentru a putea determina T0)
  - Se aplica asupra semnalului eroare de predictie pentru a elimina redundanta datorata lui T0
  - La nivel de subcadru se determina (Ga, Ta) , estimarile pentru amplitudinea si durata perioadei fundamentale
  - Ga si Ta sunt parametrii filtrului pe termen lung LTP(z) = 1-Ga * Z^(-Ta)
  - Se codeaza acesti parametri pe 2 biti (Ga), respectiv 7 biti (Ta)
  - 4 x (2 +7) = 36 biti /cadru

Etape – (III) codare sursa

- Codarea sursei
  - Se filtreaza invers eroarea de predictie
  - Se obtine un semnal de tip zgomot
  - Se codeaza subcadrele acestuia
  - CODAREA = cu impulsuri regulate (RPE – Regular Pulse Excitation)



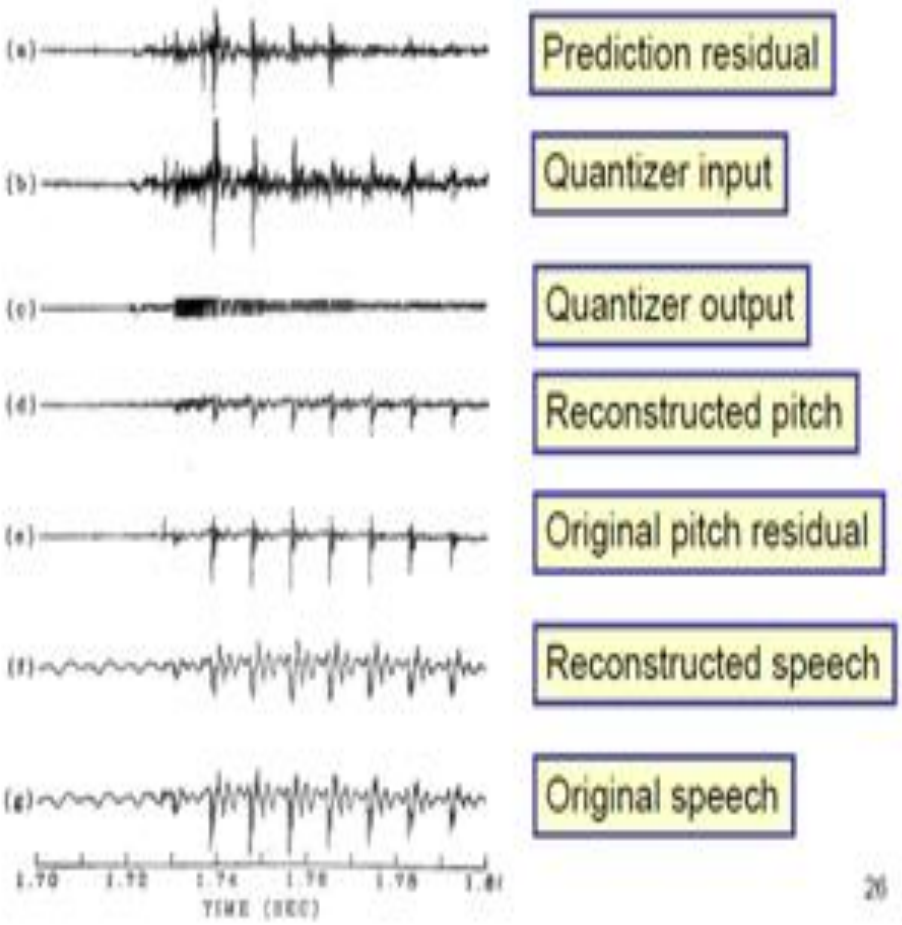
Etape de prelucrare

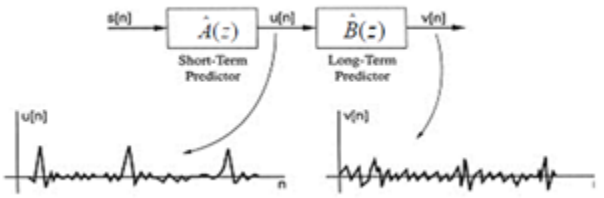
- Codarea sursei
  - Grila RPE
  - Codare: 13 impulsuri (factori de scalare pe 6 biti, codare impuls: 3 biti) + faza impulsurilor (2 biti)
  - Subcadru: 2 + 6 +13*3 = 47 biti
  - 4 subcadre = 4 *47 = 188 biti

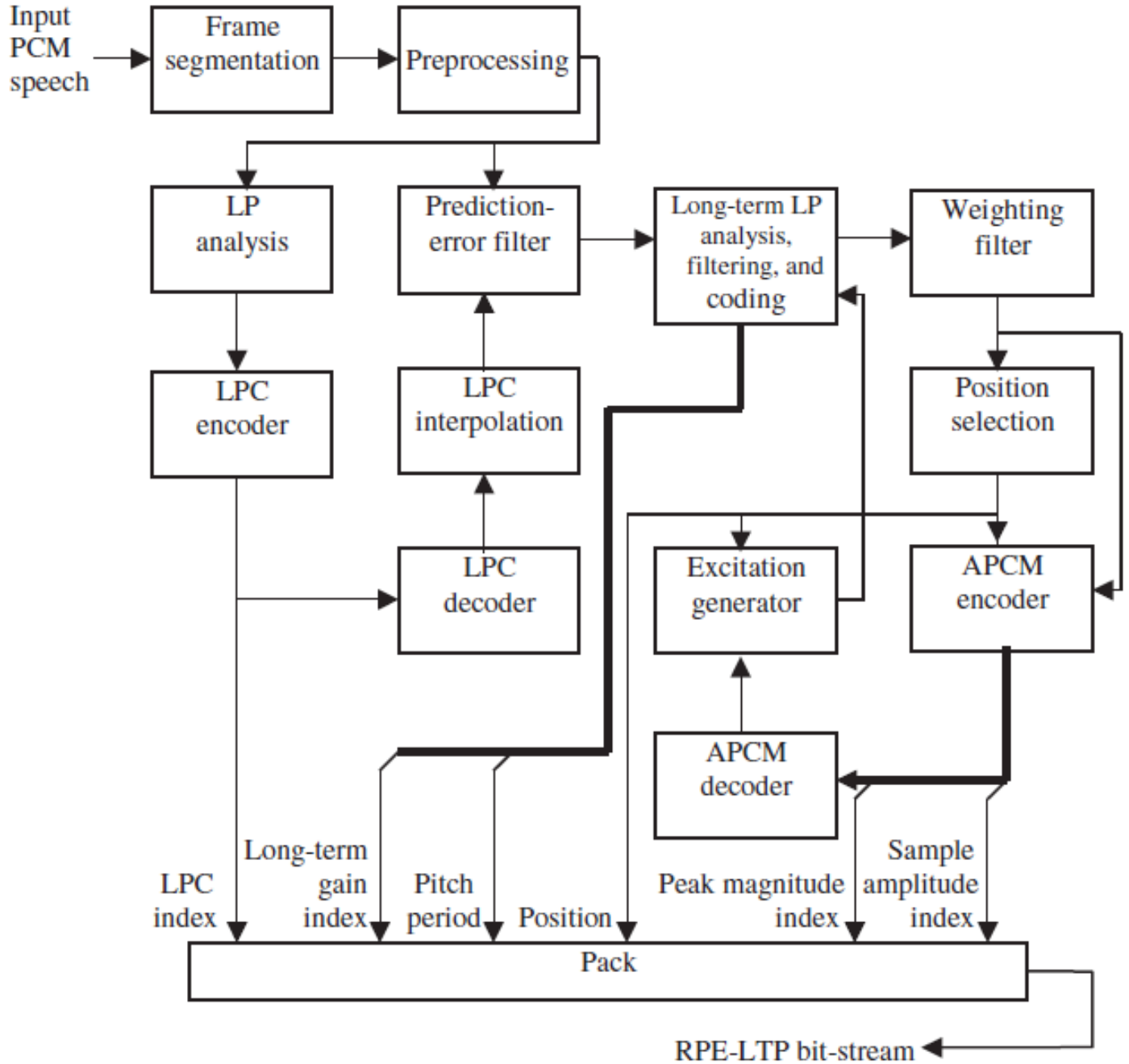
- CODURI
  - LPC: 8 coeficienti LAR: 36 biti
  - LTP: 4 * (2+7) = 36 biti
  - Grila RPE: 188 biti

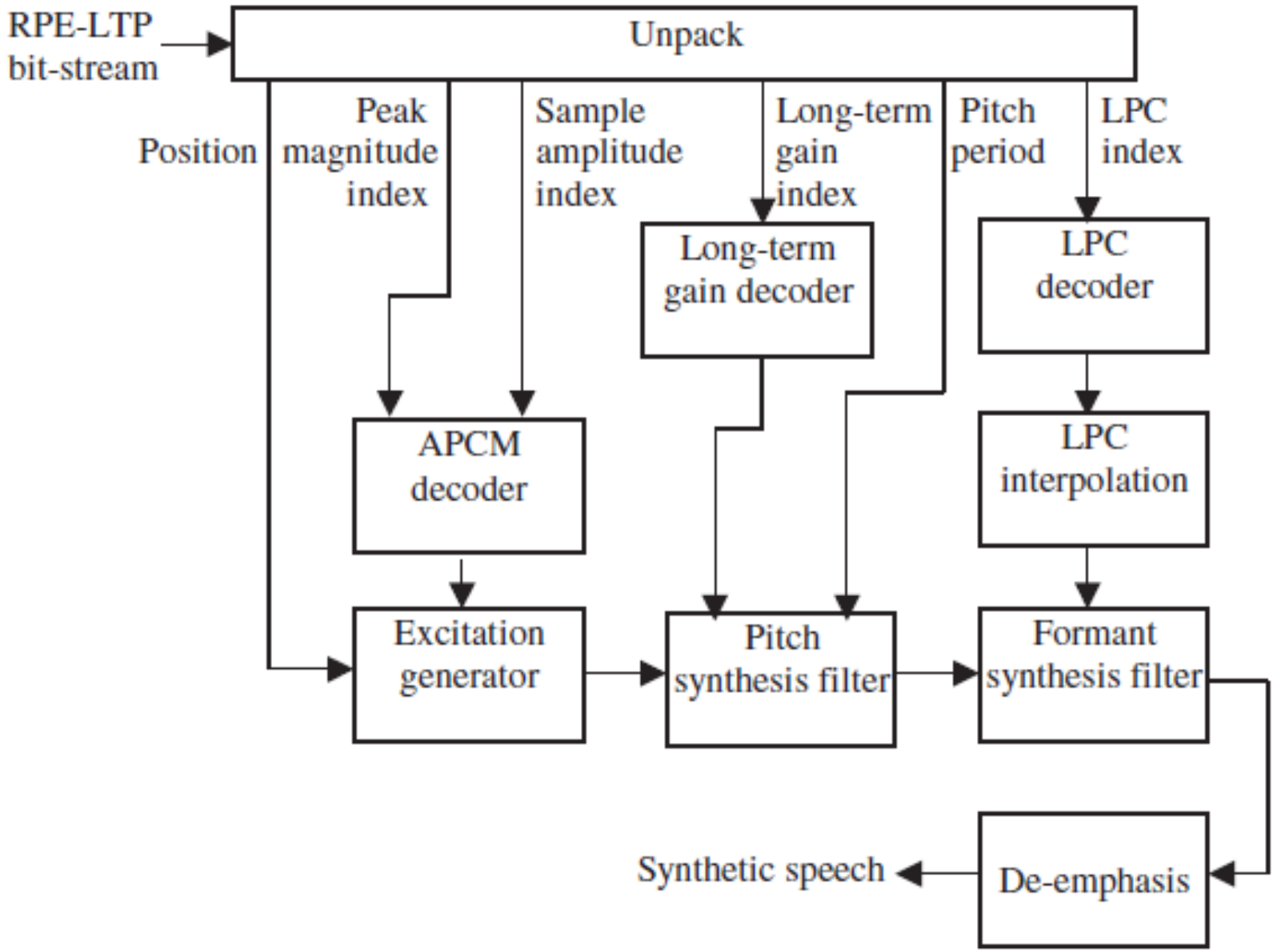
  - TOTAL : 260 biti
  - Debit 260 biti/20ms = 13Kbs

Etape de prelucrare – codare/decodare



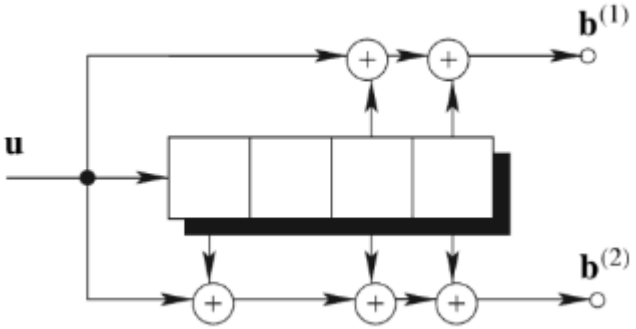



Input PCM speech
Frame segmentation
Preprocessing
LP analysis
Prediction-error filter
Long-term LP analysis, filtering, and coding
Weighting filter
LPC encoder
LPC interpolation
Position selection
LPC decoder
Excitation generator
APCM encoder
APCM decoder
LPC index
Long-term gain index
Pitch period
Position
Peak magnitude index
Sample amplitude index
Pack
RPE-LTP bit-stream


RPE-LTP
bit-stream
Unpack
Position
Peak
magnitude
index
Sample
amplitude
index
Long-term
gain
index
Pitch
period
LPC
index
Long-term
gain decoder
LPC
decoder
APCM
decoder
LPC
interpolation
Excitation
generator
Pitch
synthesis filter
Formant
synthesis filter
Synthetic speech
De-emphasis

Protectia la erori a datelor de voce in GSM

- 3 categorii de biti:
  - Foarte importanti: 50 biti (codare CRC 50 +3 = 53)
  - Importanti: 132 biti (codare convolutionala) : 132 + 53 + 4 de zero = 189 ➔ 378
  - Mai putin importanti: 78 biti



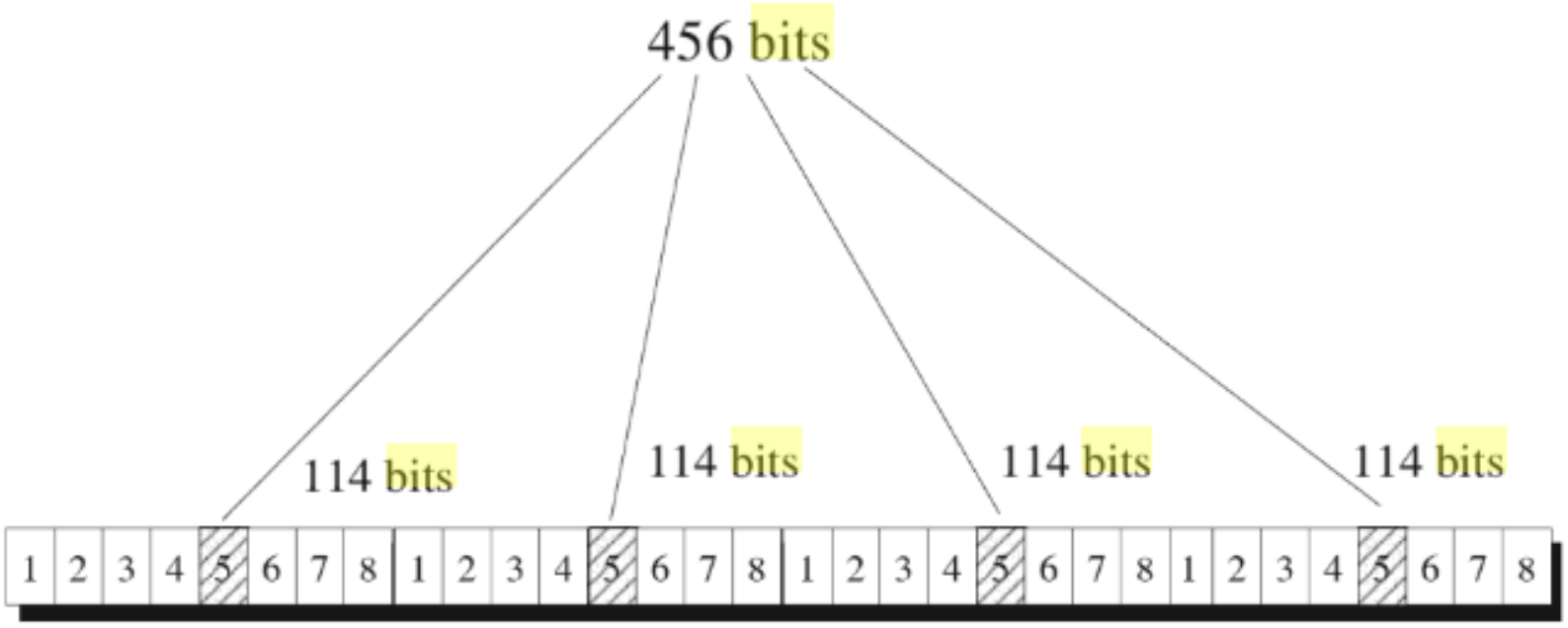
  - TOTAL codati = 456 biti

- Intreteserea celor 456 biti: matrice de 8 linii si 57 de coloane

- In structura unui burst GSM se impacheteaza 2 x 57 biti = 114 biti

Strucutura burst date

Normal burst structure

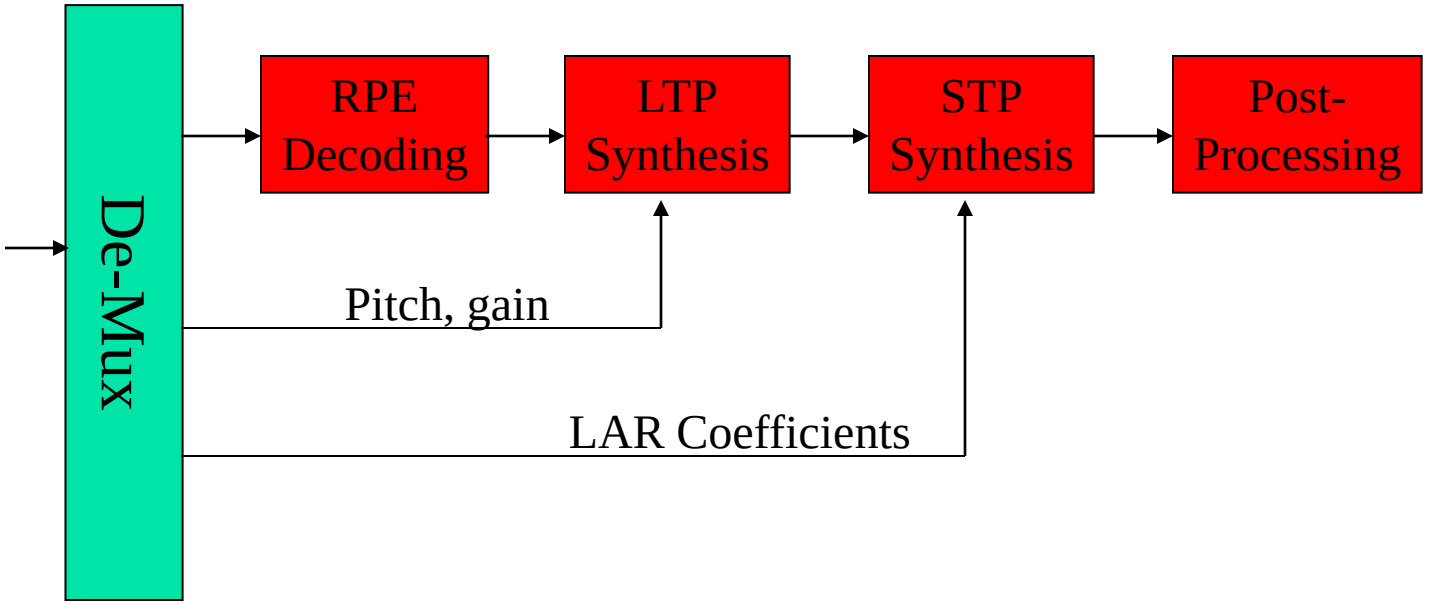


Etape de prelucrare

- 1 burst eronat = 114 biti imposiil de decodat ➔ 114 / 456 = 25% erori
- Inaceptabil

- SOLUTIE
  - Inca un nivel de intretesere la nivel de intercadru
  - Intr-un burst GSM – se include date din cadre diferite ==\> injumatatirea ratei erorilor

GSM Decoding



Implementation Issues

- Tasks:
  - LPC analysis filter to calculate the coefficients
  - Long term prediction for pitch analysis need to find delay D and gain
  - VQ search during CELP encoding – Most time consuming
  - FIR filtering for pre- and post processing
- Often implemented in DSP chips for embedded applications (e.g. cell phone).
- The parameter quantization part needs bit-level operation.

X. CODAREA PARAMETRICA

Scopul:

- extragerea unui set de parametrii din fiecare cadru de semnal vocal conform cu modelul linear separabil

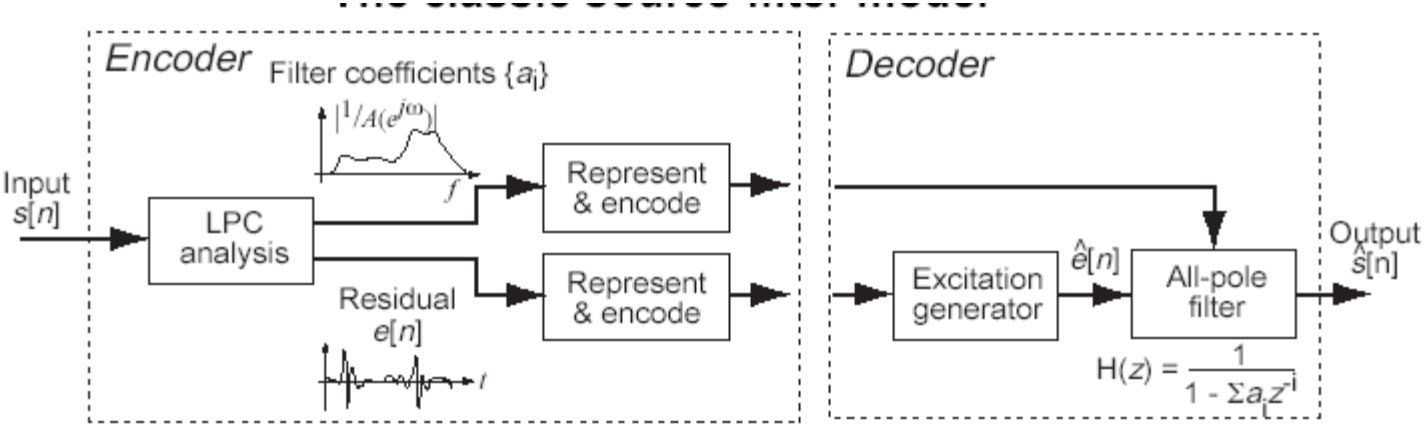
-

acesti parametri sunt folositi in codarea parametica

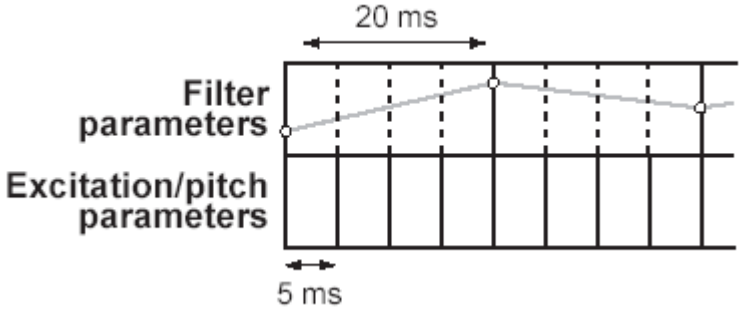
Principiul

LPC encoding

- The classic source-filter model



- Compression gains:
  - filter parameters are ~slowly changing
  - excitation can be represented many ways



Linear Predictive Code

- Model speech production system as an auto-regressive model:

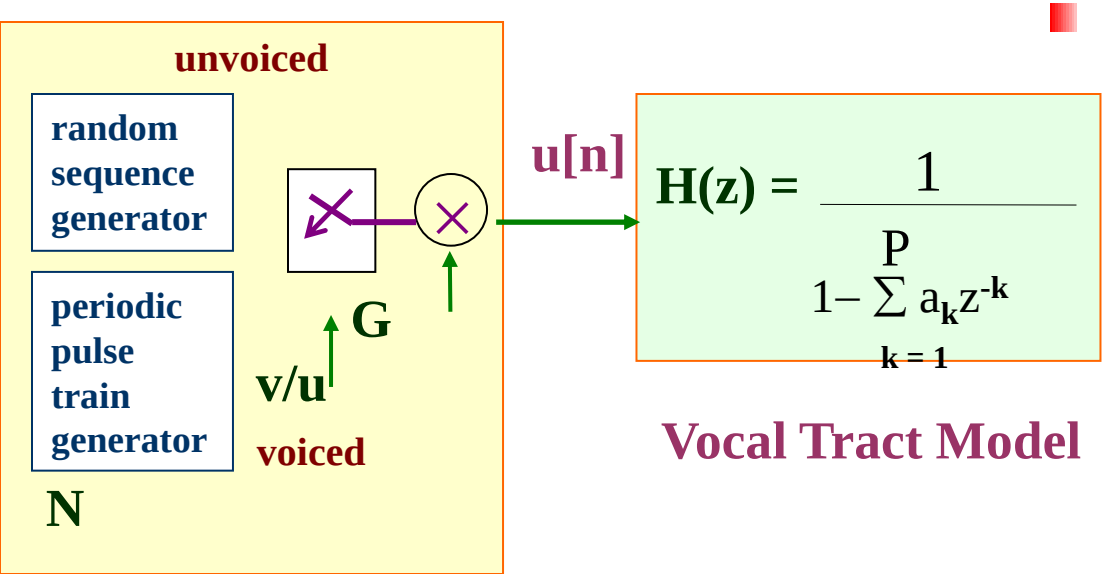
$$s(n) = \sum_{k=1}^{p} a(k)s(n-k) + e(n)$$

- Model parameters are computed for speech segment (~30 ms).
- Parameters {a(k); k=1:p} are found by solving a Toeplitz system of equations.

- Transfer function

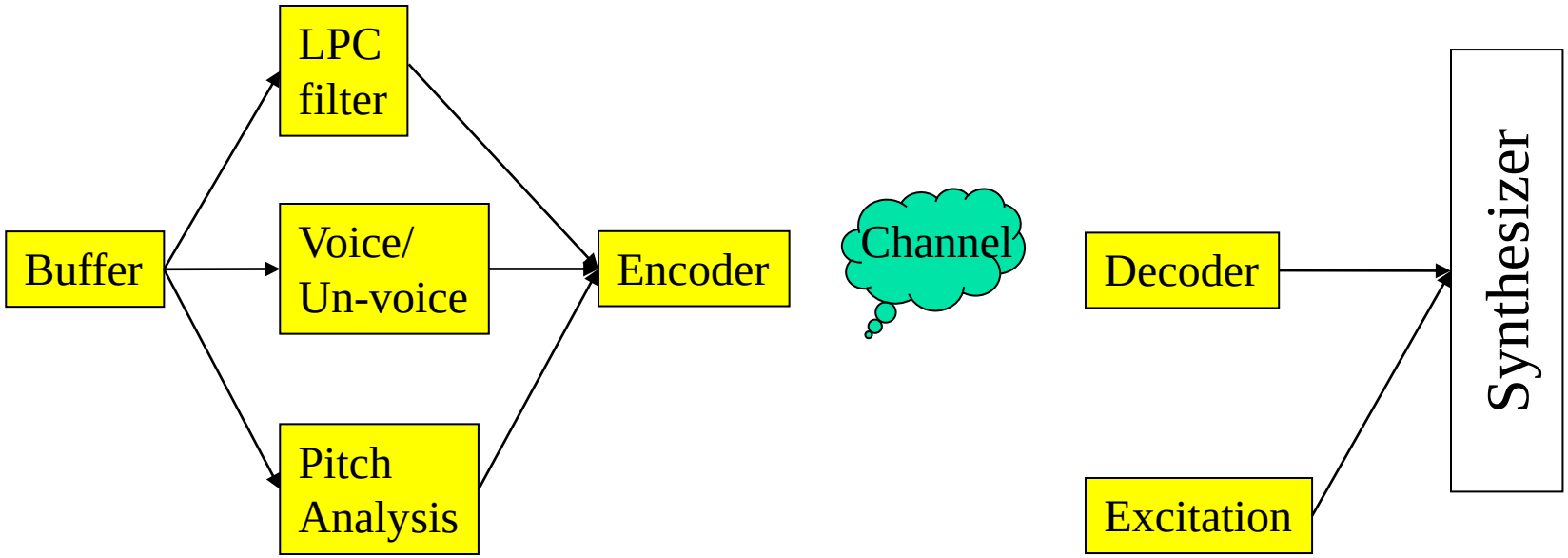
$$H(z) = \frac{S(z)}{E(z)} = \frac{G}{1 - \sum_{k=1}^{p} a(k)z^{-k}}$$

- To encode speech, one may transmit the quantized parameters {a(k)} and G or equivalent parameter set.
- The model order is 8-10 in most speech coding standards.

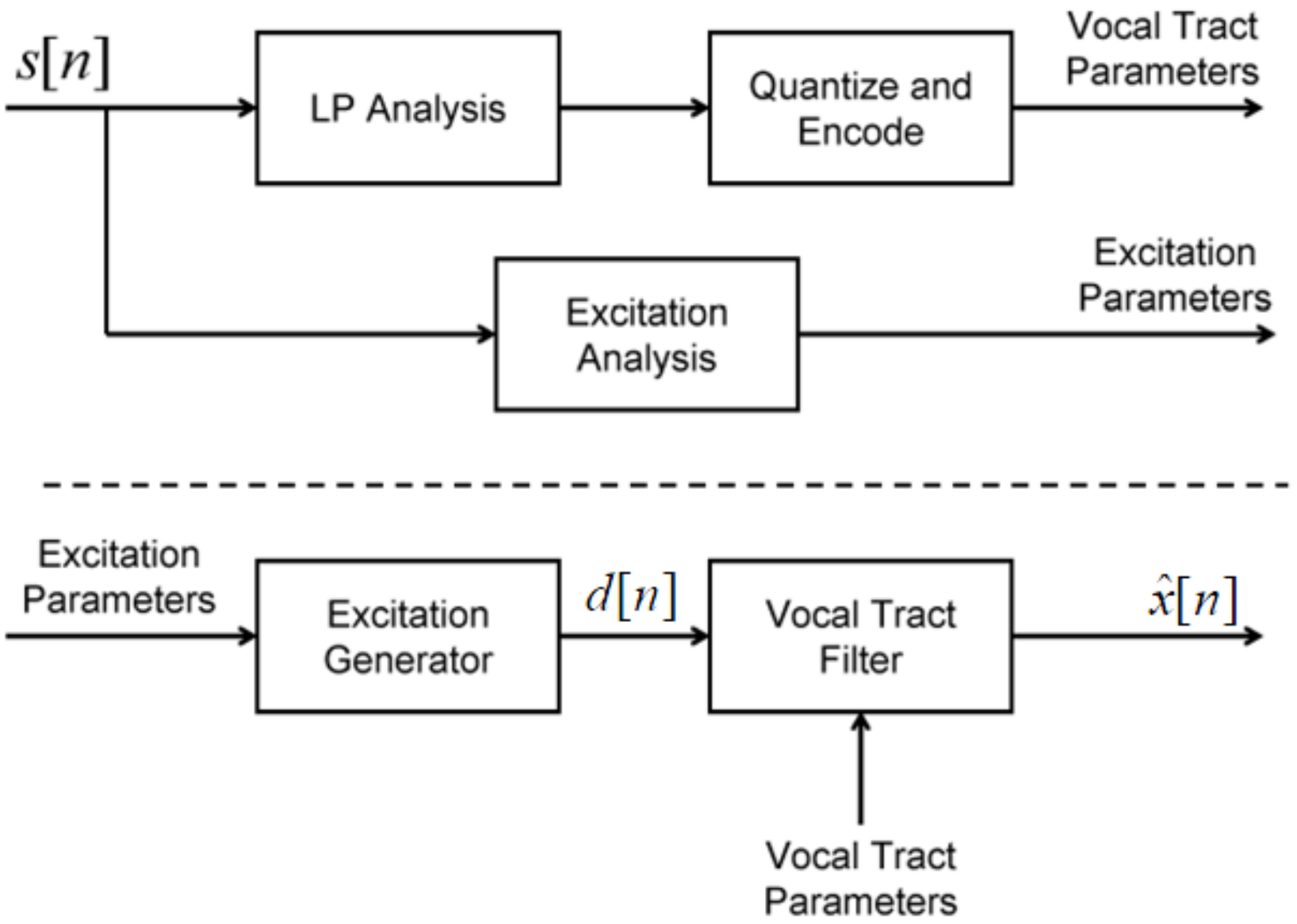


Vocal Tract Model

LPC Speech Coder



Using LP in Speech Coding



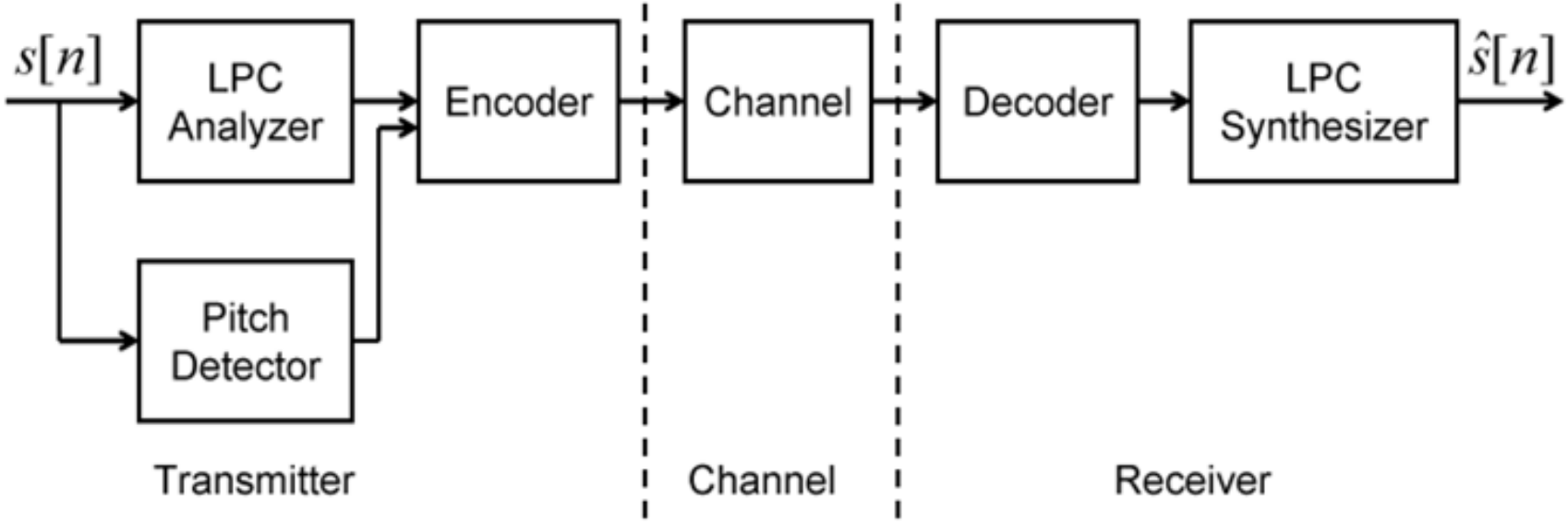
Model-Based Coding

- assume we model the vocal tract transfer function as

$$H(z) = \frac{X(z)}{S(z)} = \frac{G}{A(z)} = \frac{G}{1 - P(z)}$$

$$P(z) = \sum_{k=1}^{p} a_k z^{-k}$$

- LPC coder => 100 frames/sec, 13 parameters/frame (p = 10 LPC coefficients, pitch period, voicing decision, gain) => 1300 parameters/second for coding $\leftrightarrow$ versus 8000 samples/sec for the waveform



LPC Parameter Quantization

- don't use predictor coefficients (large dynamic range, can become unstable when quantized) => use LPC poles, PARCOR coefficients, etc.
- code LP parameters optimally using estimated pdf's f each parameter

1.	V/UV-1 bit	100 bps
2.	Pitch Period-6 bits (uniform)	600 bps
3.	Gain-5 bits (non-uniform)	500 bps
4.	LPC poles-10 bits (non-uniform)-5 bits for BW and 5 bits for CF of each of 6 poles	6000 bps
Total required bit rate		7200 bps

• no loss in quality from uncoded synthesis (but there is a loss from original speech quality)

• quality limited by simple impulse/noise excitation model

S5

S5-synthetic

LPC Coding Refinements

1. log coding of pitch period and gain
2. use of PARCOR coefficients ($|k_i|<1$) => log area ratios $g_i=\log(A_{i+1}/A_i)$—almost uniform pdf with small spectral sensitivity => 5-6 bits for coding

- can achieve 4800 bps with almost same quality as 7200 bps system above
- can achieve 2400 bps with 20 msec frames => 50 frames/sec

LPC-10 Vocoder

LPC-10 Vocoder

- U.S. Government standard
  - covariance LP analysis (10th-order)
  - AMDF pitch detector (see Chapter 4)
- Bit rate

Frame rate = 44.44 frames/sec

param.	$k_1 - k_4$	$k_5 - k_8$	k_9	k_{10}	pitch	ampl.	sync.	Total
# bits	5 ea.	4 ea.	3	2	7	5	1	54

Bit rate = 2400 bits/sec

LPC-Based Speech Coders

- the key problems with speech coders based on all-pole linear prediction models
  - inadequacy of the basic source/filter speech production model
  - idealization of source as either pulse train or random noise
  - lack of accounting for parameter correlation using a one-dimensional scalar quantization method => aided greatly by using VQ methods

Schema de codare

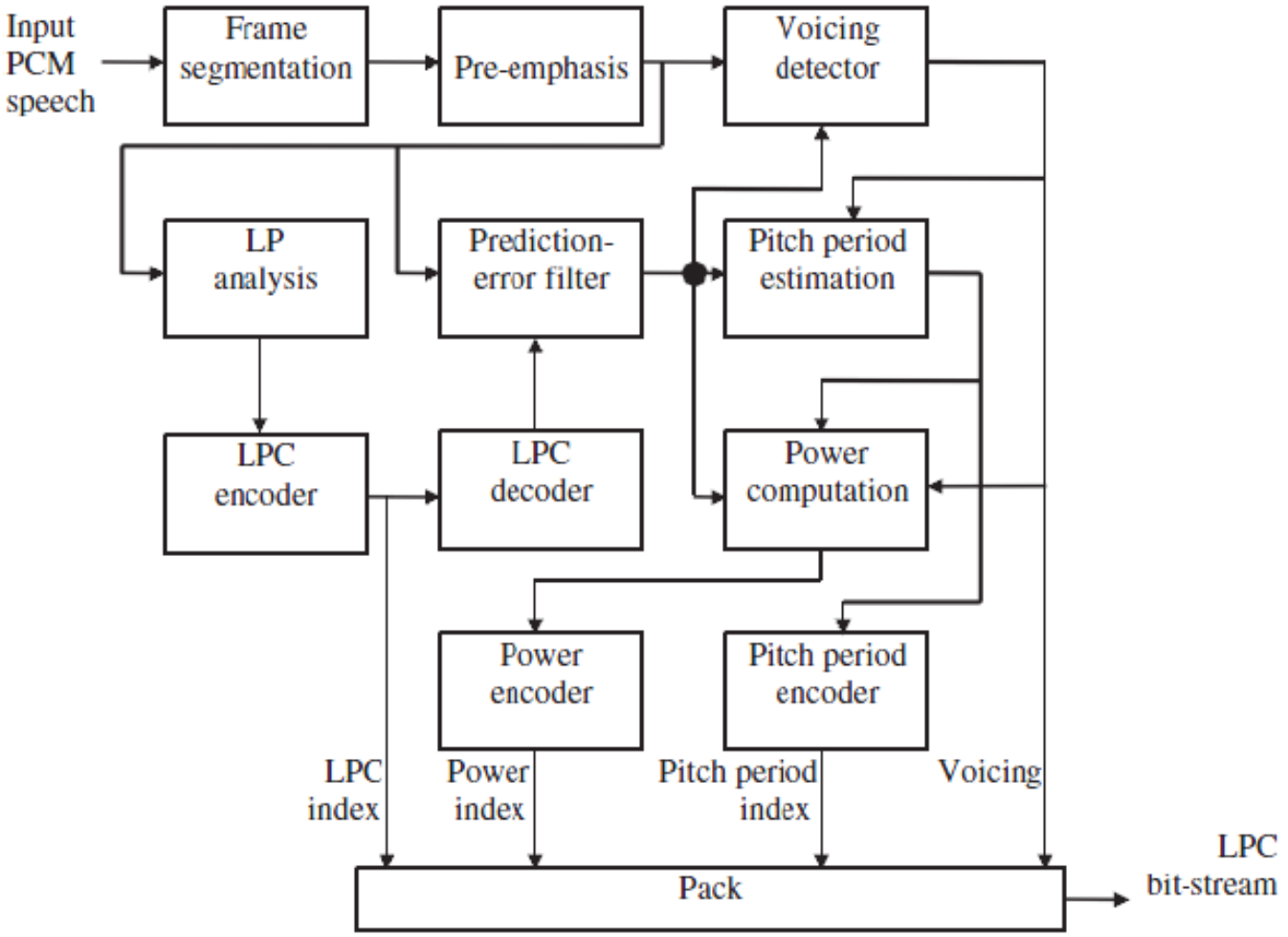


Figure 9.6 Block diagram of the LPC encoder.

Schema de decodare

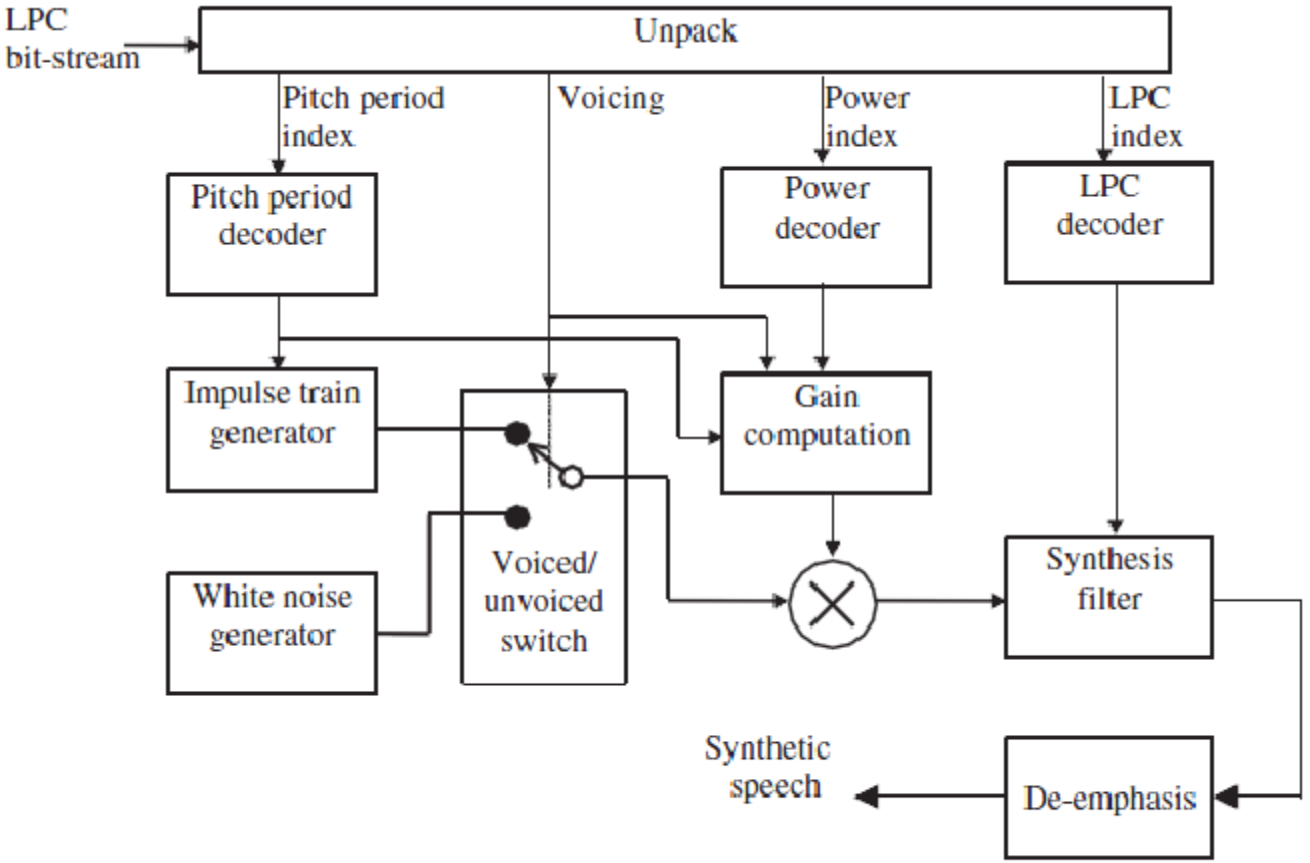
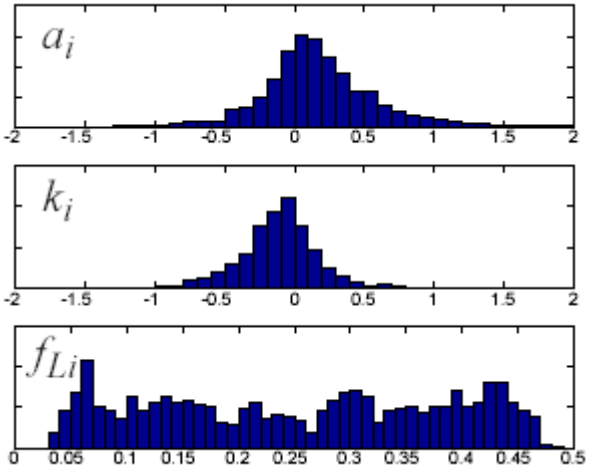


Figure 9.8 Block diagram of the LPC decoder.

Encoding LPC filter parameters

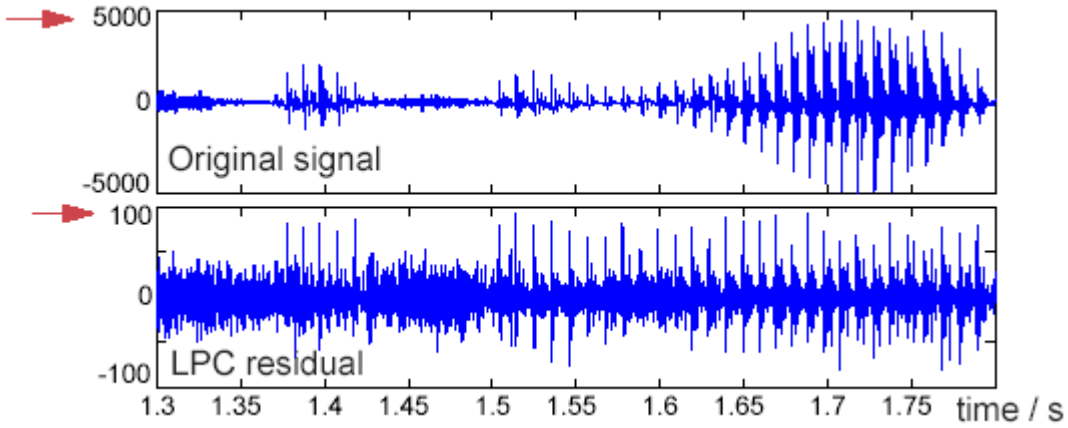
- For ‘communications quality’:
  - 8 kHz sampling (4 kHz bandwidth)
  - ~10th order LPC (up to 5 pole pairs)
  - update every 20-30 ms → 300 - 500 param/s
- Representation & quantization
  - {*ai*} - poor distribution, can’t interpolate
  - reflection coefficients {*ki*}: guaranteed stable
  - log area ratios (LAR) - stable



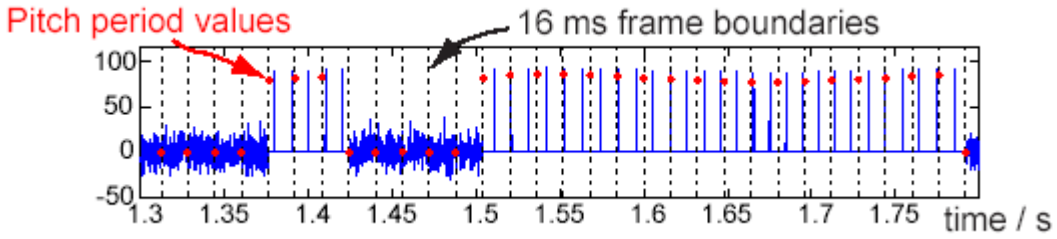
- Bit allocation (filter):
  - GSM (13 kbps):
    8 LARs x 3-6 bits / 20 ms = 1.8 Kbps

Excitation

- Excitation as LPC residual is already better than raw signal:
  - save several bits/sample, still > 32 Kbps



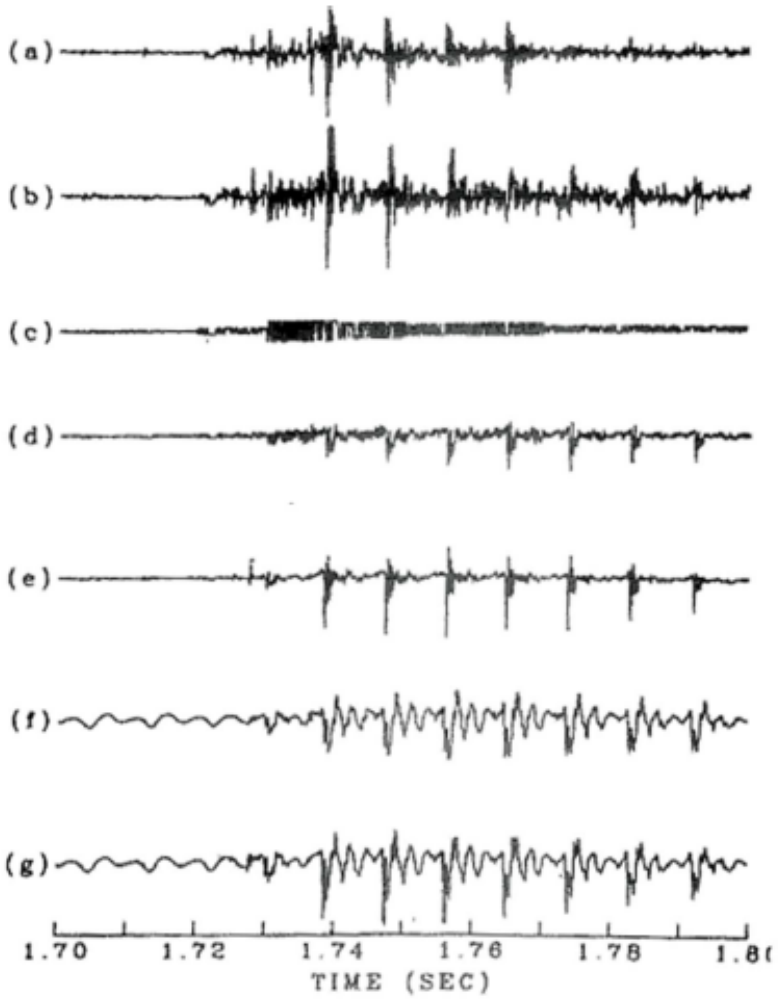
- Crude model: U/V flag + pitch period
  - ~ 7 bits / 5 ms = 1.4 Kbps → LPC10 @ 2.4 Kbps



Bit Rate for LP Coding

- speech and residual sampling rate: F_s=8 kHz
- LP analysis frame rate: $F_\Delta = F_P$ = 50-100 frames/sec
- quantizer stepsize: 6 bits/frame
- predictor parameters:
  - *M* (pitch period): 7 bits/frame
  - pitch predictor coefficients: 13 bits/frame
  - vocal tract predictor coefficients: PARCORs 16-20, 46-50 bits/frame
- prediction residual: 1-3 bits/sample
- total bit rate:
  - $BR = 72 * F_P + F_s$ (minimum)

Two-Level (B=1 bit) Quantizer



Prediction residual

Quantizer input

Quantizer output

Reconstructed pitch

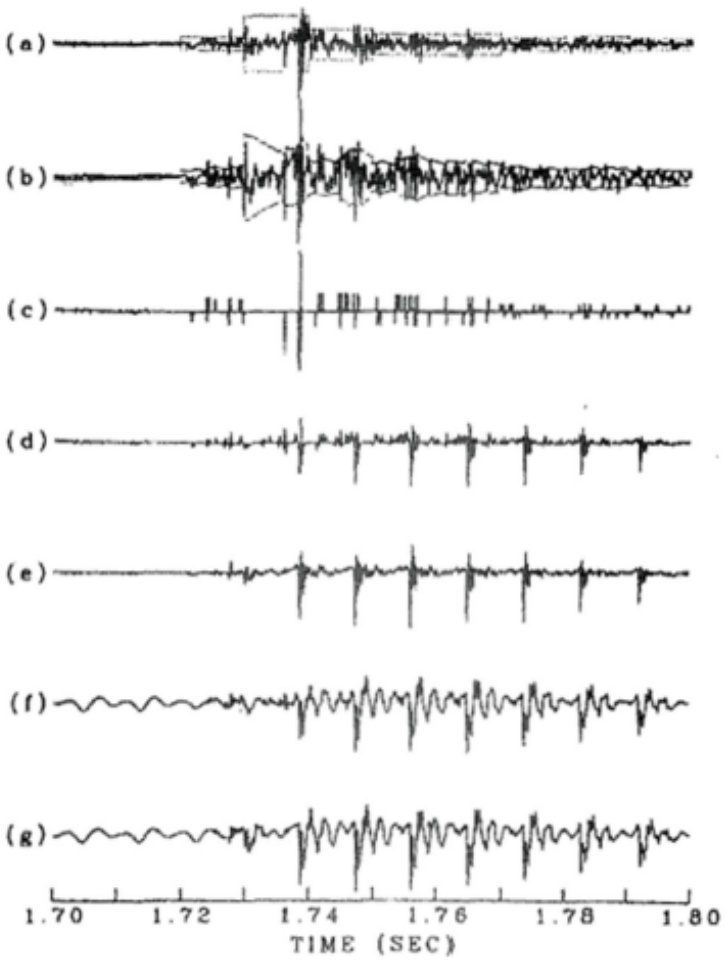
Original pitch residual

Reconstructed speech

Original speech

26

Three-Level Center-Clipped Quantizer



Prediction residual

Quantizer input

Quantizer output

Reconstructed pitch

Original pitch residual

Reconstructed speech

Original speech

27

CURS
RECAPITULATIV
TCSV

Sinteza cursuri
Intrebari si raspunsuri

- Introducere

- Bibliografia
- Aplicatii multimedia care folosesc compresia semnalului vocal
- Atributele sistemelor de compresie de voce
- Calitatea semnalelor (vezi notite curs)
  - Obiectiv – SNR
  - Subiectiv – MOS
- Principalele standarde de compresie de semnal vocal

– Codarea PCM, WPCM, DPCM

- Codarea PCM
  - Esantionare
  - Cuantizare (uniforma, neuniforma)
  - Zgomotul de cuantizare – uniformizarea
  - Legile de compresie “ A” si “miu”
- PCM de banda larga
  - Schema bloc si componentele sale
  - Low band
  - High band
  - Rolul TCD

– Codarea ADPCM

- ADPCM de banda ingusta (0-4KHz)
  - Principiul
  - Schema codorului / decodorului
  - Rolul blocurilor componente
- ADPCM de banda larga (0-8 KHz)
  - Principiul
  - Atribute codare banda inferioara, superioara
  - Schema codorului, decodorului - diferente

– Codarea in sub-benzi

- Schema de principiu
  - Rolul blocurilor
  - Proiectarea filtrelor trece banda
  - Rolul sub-esantionarii
- Aplicatie software

– Codarea sinusoidala

- Principiul codarii sinusoidale
- Schema de procesare
- Discutii privind codarea componentelor sinusoidale

– Codarea MPEG

- Efectul de mascare – explicare
- Modelul psihoacustic - rol
- MPEG, Layer I, II, III
  - Caracteristici
  - Schema de procesare (FTB, sub-esantionare, grupare in blocuri, factorii de scalare)
  - Diferente intre diferitele layere
  - Caracteristici esentiale la MP3 (DCT, fereastra)

Codarea folosind analiza prin sinteza (paramatrica). Codarea in GSM

- Principiul codarii folosind analiza prin sinteza
  - O singura sursa de excitatie
  - Sinteza la emisie
  - Filtrul e ponderare perceptuala
- Diferite abordari: MPE, RPE-LTP, CELP
- Codorul CELP
- Codorul MELP, MBE
- Codarea in GSM
  - Schema
  - Mod de alocare biti

– Compresia prin Cuantizare vectoriala (VQ – Vector Quantization). Transformata W

- Compresie prin cuantizare vectoriala (VQ – Vector Quantization)
  - Principiul
  - Algoritmi de dreare a dictionarului (Lloyd, k-means, LBG)
- Compresia prin Transformata Wavelet
  - Definire, diferente fata de FFT
  - CWT, DWT – coeficienti de aproximare, detaliu
  - Aplicare la compresie (sub prag, anulare coeficienti)

– Recapitulare

– Sinteza cursului
– Elaborare intrebari recapitulative
– Structura examen