



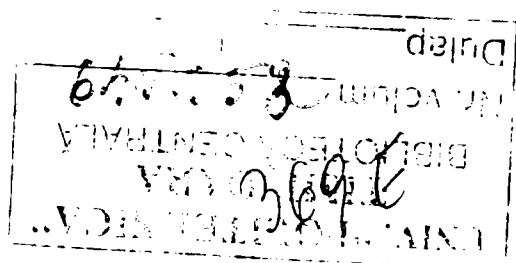
**Balint Cornel**

***Contribuții la cuantizarea vectorială a semnalului vocal,  
cu aplicații la codarea CELP***

**Teză de doctorat**

***Conducător științific:***

***Prof. dr. ing. Miranda Naforniță***



**2005**



# Cuprins

Cap. 1	Introducere	1
1.1	Mecanismul producerii vorbiri naturale	2
1.2	Modelarea mecanismului de producere a vorbirii	3
1.3	Caracteristicile semnalului vocal	5
1.3.1	Caracteristici temporale	5
1.3.2	Caracteristici spectrale	6
1.3.3	Caracteristici lingvistice	7
1.3.4	Caracteristici statistice	8
1.4	Scurt istoric	9
1.5	Performanțele sistemelor de codare a semnalului vocal	10
1.5.1	Calitatea semnalului vocal	10
1.5.2	Rata de bit	12
1.5.3	Complexitatea algoritmică	13
1.5.4	Întârzierea semnalului	13
1.5.5	Banda semnalului	14
1.5.6	Robustețea	14
1.6	Standardizarea în domeniul codării semnalului vocal	14
Cap. 2	Metode de codare a semnalului vocal	16
2.1	Codarea formei de undă	18
2.1.1	Codarea scalară și codarea vectorială	18
2.1.1.1	Codarea scalară	18
2.1.1.2	Codarea vectorială	20
2.1.2	Codarea în domeniul frecvență	20
2.1.2.1	Codarea în subbenzi	20
2.1.2.2	Codarea prin transformări ale semnalului	22
2.1.2.3	Codarea perceptuală a semnalelor audio	23
2.1.3	Codarea semnalului vocal utilizând analiza și sinteza sinusoidală	24
2.1.3.1	Codarea semnalului vocal folosind transformata Fourier pe termen scurt	24
2.1.3.2	Vocoderul de fază	25
2.2	Codarea sursei de semnal	25
2.2.1	Vocoderul canal	26
2.2.2	Vocoderul formantic	26
2.2.3	Vocoderul homomorfic	27
2.2.4	Codorul LPC	28
2.2.4.1	Codorul LPC în buclă deschisă	28
2.2.4.2	Codorul LPC în buclă închisă	30
Cap. 3	Predicția liniară a semnalului vocal	35
3.1	Predicția liniară	35
3.1.1	Ecuțiile de predicție liniară	35
3.1.2	Prelucrarea semnalului pe cadre	37
3.2	Calculul coeficienților de predicție	40
3.2.1	Fereastra de ponderare	40
3.2.2	Metoda autocorelației	41

3.2.3	Metoda covarianței .....	43
3.2.4	Câștigul de predicție .....	43
3.2.5	Predicția liniară a semnalului vocal .....	44
3.3	Reprezentări ale coeficienților de predicție .....	47
3.3.1	Coeficienții de reflexie .....	47
3.3.2	Coeficienții LAR .....	48
3.3.3	Perechile de linii spectrale .....	48
3.4	Codorul CELP .....	49
3.4.1	Codorul și decodorul CELP .....	50
3.4.2	Ponderarea perceptuală a erorii .....	51
3.4.3	Postfiltrarea .....	53
Cap. 4	Cuantizarea vectorială .....	57
4.1	Cuantizarea .....	57
4.2	Principiile cuantizării vectoriale .....	58
4.3	Proprietăți structurale ale cuantizorului vectorial .....	59
4.4	Performanțele cuantizorului vectorial .....	60
4.5	Cuantizarea vectorială după regula celui mai apropiat vecin .....	61
4.6	Condiții de optimalitate pentru cuantizorul vectorial .....	62
4.6.1	Condiția de codor optim pentru un decodor dat .....	62
4.6.2	Condiția de decodor optim pentru un codor dat .....	62
4.6.3	Condiția de margine cu probabilitatea zero .....	63
4.7	Proiectarea unui cuantizor vectorial .....	64
4.7.1	Dicționarul inițial .....	64
4.7.2	Algoritmul Lloyd .....	66
4.7.3	Algoritmul Lloyd generalizat pentru date empirice .....	67
4.7.4	Alte metode pentru construcția unui cuantizor vectorial .....	69
4.7.5	Algoritmul LBG îmbunătățit .....	69
4.8	Metode de cuantizare vectorială .....	71
4.8.1	Metode rapide de codare vectorială nestructurată .....	72
4.8.2	Cuantizarea vectorială structurată .....	74
4.8.2.1	Cuantizarea vectorială în structură arborescentă (TSVQ) .....	74
4.8.2.2	Cuantizarea vectorială cu clasificare (CVQ) .....	74
4.8.2.3	Cuantizarea vectorială cu transformare .....	75
4.8.2.4	VQ de produs (PVQ) .....	76
4.8.2.5	VQ în mai multe etaje MSVQ (Multistage VQ) .....	78
4.8.2.6	VQ ierarhic – multirezoluție .....	78
4.8.2.7	VQ prin interpolare neliniară .....	79
4.8.3	Cuantizarea vectorială predictivă .....	80
4.8.4	Cuantizarea vectorială adaptivă .....	81
4.8.4.1	Modelul cuantizorului adaptiv .....	83
4.8.5	Cuantizarea vectorială cu rată de bit variabilă .....	84
Cap. 5	Calculul și cuantizarea vectorială a coeficienților de predicție .....	86
5.1	Coeficienții de predicție .....	86
5.1.1	Perechile de linii spectrale LSP .....	87
5.1.2	Proprietăți ale perechilor de linii spectrale .....	90
5.2	Calculul LSP .....	91
5.2.1	Calculul LSP în cazul general .....	91
5.2.2	Optimizarea calculului LSP pentru ordinul 10 .....	92
5.3	Interpolarea LSP .....	96
5.4	Cuantizarea vectorială a LSP .....	100
5.4.1	Cuantizarea P-SVQ a coeficienților LSP .....	101



5.4.2 Cuantizarea C-SVQ a coeficienților LSP .....	103
<b>Cap. 6 Modelarea periodicității în codorul CELP</b> .....	106
6.1 Extragerea periodicității din semnalul rezidual.....	107
6.2 Deducerea relațiilor de calcul pentru parametrii LTP .....	112
6.3 Optimizarea filtrului LTP cu un coeficient .....	115
6.3.1 Recircularea excitației prin filtrul LTP.....	115
6.4 Creșterea rezoluției filtrului LTP .....	118
6.4.1 Filtrul LTP cu ordin fracționar .....	118
6.5 Filtrul LTP cu trei coeficienți modificat.....	119
6.5.1 Calculul parametrilor LTP pentru un filtru cu trei coeficienți.....	120
6.5.2 Filtru LTP cu constrângeri cu trei parametri – LTP3 .....	121
6.5.3 Filtru LTP cu constrângeri cu doi parametri – LTP2 .....	122
6.5.4 Răspunsul în frecvență al filtrului LTP .....	123
6.5.5 Stabilitatea filtrului de sinteză LTP .....	123
6.6 Rezultate experimentale .....	125
<b>Cap. 7 Cuantizarea vectorială a semnalului rezidual în codorul CELP</b> .....	129
7.1 Dicționarul fix în codorul CELP .....	129
7.2 Operația de căutare în dicționarul fix .....	130
7.3 Complexitatea căutării în dicționarul fix.....	133
7.4 Reducerea complexității de căutare în dicționarul fix.....	135
7.4.1 Trunchierea răspunsului la impuls al filtrului .....	135
7.4.2 Simplificarea calculului corelației .....	136
7.4.3 Simplificarea calculului energiei .....	137
<b>Cap. 8 Concluzii și contribuții</b> .....	140
8.1 Concluzii .....	140
8.2 Contribuții originale .....	144
<b>Bibliografie</b> .....	147

# Glosar

ACELP	CELP algebric (Algebraic CELP)
AMDF	Funcția diferență de amplitudine medie
AR	Autoregresiv (model autoregresiv)
ARMA	Autoregresiv cu medie alunecătoare (model autoregresiv cu medie alunecătoare)
ASIC	Application Specific Integrated Circuits
CCITT	Comitetul Consultativ Internațional pentru Telefonie și Telegrafie (Comité Consultatif International Téléphonique et Télégraphique)
CELP	Predicție liniară excitată cu coduri (Code Excited Linear Prediction)
CVQ	VQ cu clasificare (Class VQ)
DALT	Test subiectiv de inteligibilitate (Diagnostic Alliteration Test)
DCT	Transformata cosinus discretă
DFT	Transformata Fourier discretă
DPCM	PCM diferențial
DRT	Test subiectiv de inteligibilitate (Diagnostic Rythme Test)
ETSI	Institutul European de Standardizare în Telecomunicații (European Telecommunication Standards Institute)
FFT	Transformata Fourier rapidă
GLA	Algoritmul Lloyd generalizat
GSM	Sistem de comunicații mobile (Global System for Mobile communications)
ISO	Organizația Internațională de Standardizare (International Standardization Organization)
ITU	Uniunea Internațională de Telecomunicații (International Telecommunication Union)
LAR	Raport logaritmic al coeficienților (log area ratio)
LBG	Algoritmul Linde – Buzo – Gray
LPC	Predicție liniară
LSP	Perechi de linii spectrale
LTP	Predicție pe termen lung (Long Term Prediction)
MA	Medie alunecătoare
MELP	CELP cu excitație mixtă (Mixed Excitation LP)
MOS	Scor de opinie medie (Mean Opinion Score)
MSVQ	VQ multietaj (Multistage VQ)
NN	Cel mai apropiat vecin (Nearest Neighbor)
PCM	Pulse Code Modulation
pdf	Funcția densitate de probabilitate
pmf	Funcția densitate de masă
PVQ	VQ de produs (Product VQ)
RPE	Excitație cu impulsuri regulate (Regular Pulse Excitation)
SD	Distorsiune spectrală (Spectral Distortion)
SEGSNR	Raport semnal – zgomot segmental
SNR	Raport semnal – zgomot
STP	Predicție pe termen scurt (Short Term Prediction)
TSVQ	Cuantizor vectorial cu căutare arborescentă (Tree Search VQ)
TTS	Sinteza semnalului vocal pe baza textului (Text to speech)
VQ	Cuantizare (cuantizor) vectorială (Vector Quantization)
VSELP	CELP excitat cu sumă de vectori (Vector Sum Excited LP)
WSTP	Predicție pe termen scurt ponderată (Weighted STP)

## *Lista simbolurilor folosite*

$\alpha, \beta, \gamma, \mu$	Constante
$\delta(n)$	Impuls unitate
$\{a_i\}$	Coeficienți de predicție
$a_i^{(m)}$	Coeficienți de predicție ai modelului de ordin $m$
$A(z)$	Filtru de analiză LPC
<b>a, b, s, <math>\phi</math></b>	Vectori
<b><math>\Phi, M, W</math></b>	Matrici
$\arg(z)$	Argumentul numărului complex $z$
$d(\mathbf{x}, \hat{\mathbf{x}})$	Distanța între vectorii $\mathbf{x}$ și $\hat{\mathbf{x}}$
$D(\cdot)$	Distanța medie
$e(n)$	Eroarea
$E\{ \cdot \}$	Valoare medie
$F_e$	Frecvența de eșantionare
$G, \beta$	Factor de câștig
$h(n)$	Răspunsul la impuls al filtrului $H(z)$
$H(x)$	Entropie
$H(z)$	Filtru de sinteză LPC
$i, j, k$	Indici
<b>I</b>	Matricea unitate
$k(m)$	Coeficienți de reflexie
$L$	Ordinul predicției pe termen lung
$N$	Lungimea unui cadru sau subcadru de semnal
$p$	Ordin de predicție
$R(i)$	Funcția de autocorelație
$r(n)$	Eroare de predicție
$S(j\omega)$	Spectrul de putere al semnalului $s(n)$
$s(n)$	Semnal vocal
$\hat{s}(n)$	Valoarea estimată a semnalului $s(n)$
$S_i(x)$	Funcție selector
<b>T</b>	Perioada
$T_m(\omega)$	Polinom Cebîșev
$w(n)$	Funcție fereastră
$W(z)$	Filtru de ponderare
$x(n)$	Semnal de intrare, în general
<b>x, y, s, e</b>	Vectori
$\hat{x}, \hat{y}$	Valoarea cuantizată a semnalului $x, y$
<b><math>\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{s}}, \hat{\mathbf{e}}</math></b>	Valoarea cuantizată (sau estimată) a vectorului <b>x, y, s, e</b>
$y(n)$	Semnal de ieșire, în general
$z$	Variabila complexă
$\varepsilon$	Eroarea pătratică
$\omega$	Pulsația ( $2\pi f$ )
<b>C</b>	Dicționar (al cuantizorului vectorial)
<b><math>\mathcal{E}, \mathcal{D}</math></b>	Codor vectorial, respectiv decodor asociat
$\mathcal{R}^k$	Spațiul euclidian $k$ – dimensional
$\  \cdot \ $	Norma euclidiană
$( \cdot )^T$	Operația de transpunere a matricii

# Capitolul 1

## Introducere

Vorbirea, ca modalitate de transmitere a informației prin intermediul sunetelor articulate, reprezintă mijlocul primar de comunicare al omului. Prin vorbire, omul s-a diferențiat de toate celelalte viețuitoare, omul fiind singurul care poate să transmită prin voce informații complexe. Comunicarea prin voce reprezintă, fără îndoială, cea mai importantă și cea mai mult folosită cale de comunicare între oameni, dar începe să devină o cale de comunicare tot mai mult utilizată și în dialogul între om și mașină.

Se pot identifica mai multe categorii de aplicații în care semnalul vocal este elementul care urmează a fi transmis sau stocat.

O primă categorie o constituie serviciile de telecomunicații care asigură în principal transmiterea semnalului vocal, folosind rețelele fixe, celulare, mobile, prin satelit, etc., a căror dezvoltare practic exponențială se confruntă cu necesitatea economiei de bandă utilizată, în condițiile asigurării unei calități ridicate a comunicației.

O a doua categorie o constituie acele aplicații care integrează comunicația vocală în cadrul general mai larg al comunicațiilor multimedia și în particular aplicațiile de tip VoIP, în contextul dezvoltării rețelei Internet. În această categorie se încadrează și unele aplicații pentru calculatoarele personale, ca de exemplu programe pentru învățarea limbilor străine sau jocurile.

O a treia categorie cuprinde aplicațiile comunicării vocale în contextul dialogului om – mașină [LPE99], aplicații integrate în tot mai multe produse, ca de exemplu telefonul, automobilul, sau produsele de larg consum din categoria electrocasnice sau jucării.

Toate aceste aplicații presupun existența semnalului vocal în format digital, în vederea procesării, memorării sau transmiterii acestui semnal sub controlul unui software specializat. În acest sens se subînțelege că semnalul a fost eșantionat cu o frecvență corespunzătoare și apoi cuantizat pentru a se obține o formă digitală. Codarea sau compresia semnalului vocal presupune obținerea unei reprezentări compacte a semnalului în vederea stocării sau transmiterii printr-un canal de comunicație.

Dezvoltările metodelor de prelucrare a semnalului vocal au urmărit mai multe scopuri, care ar putea fi sintetizate astfel:

- codarea semnalului vocal (reducerea debitului binar) pentru:
  - stocare,
  - transmisie eficientă (inclusiv criptare),
- sinteza semnalului vocal:
  - cu vocabular redus (dialog om-mașină),
  - cu vocabular nelimitat (sinteza vorbirii pornind de la text – sisteme TTS – *text to speech*),
- recunoașterea vorbirii (cuvinte izolate sau vorbire continuă) și recunoașterea vorbitorului,
- analiza semnalului vocal pentru îmbunătățirea raportului semnal – zgomot.

## 1.1 Mecanismul producerii vorbirii naturale

Comunicarea prin vorbire presupune existența unui lanț de comunicare din care face parte vorbitorul și ascultătorul sau ascultătorii. Procesul comunicării începe prin formularea unei idei care este convertită într-o structură lingvistică folosind cuvinte și reguli gramaticale specifice limbii respective, după care creierul vorbitorului generează o secvență de semnale de comandă pentru mușchii care fac parte din sistemul de producere a vorbirii. Ca rezultat se obține o undă acustică care se propagă în mediul înconjurător (aer) și care este percepută de ascultător. Această undă acustică este percepută și de vorbitor și folosită ca semnal de reacție pentru controlul mecanismului producerii vorbirii.

Vorbirea este produsul final acustic al mișcărilor voluntare tipice ale aparatului respirator și masticator, constituind o activitate motorie produsă, controlată și menținută prin reacția acustică a mecanismului de audiere și prin comanda musculaturii organelor de vorbire de către sistemul nervos central.

Mecanismul de producere a vorbirii naturale este alcătuit din următoarele elemente componente [FLA72]:

- plămâni și mușchii aferenți, care produc fluxul de aer care constituie sursa de energie a procesului de vorbire;
- traheea, laringele și coardele vocale, care, prin vibrație, imprimă o modulație periodică fluxului de aer;
- tractul vocal, care începe la capătul traheii și se termină la buze, formând un tub acustic de secțiune variabilă în timp datorită mișcărilor articulatorii ale velumului, limbii și buzelor;
- tractul nazal, de secțiune constantă în timp, ca o cale suplimentară de transmitere a sunetelor, cuplat cu cel vocal prin velum, cu gradul de cuplare variabil, controlat prin deschiderea velumului.

Energia acustică necesară vorbirii este furnizată de musculatura toracică și abdominală prin mecanismul respirației. Prin expirație, aerul este eliberat din plămâni, trece prin trahee, străbate coardele vocale și apoi pătrunde în faringe. Se pot identifica trei surse principale de excitație vocală:

- **coardele vocale.** Modul de producere a sunetelor corespunzător acestei surse de excitație se numește fonație. La momentul inițial, coardele vocale se consideră lipite; presiunea subglotală crește, forțându-le să se depărteze și aerul trece prin glotă. Ca urmare, scade presiunea subglotală și coardele vocale revin în poziția inițială. Ciclul de relaxare se repetă cu o perioadă mai mică decât perioada proprie de oscilație a coardelor vocale și acestea execută oscilații forțate. Volumul de aer care trece prin glotă variază în timp la fel ca și suprafața deschiderii glotală, după o lege aproximativ triunghiulară, cu un factor de umplere cuprins între 0,3 – 0,7. Forma acestor impulsuri variază în limite largi, chiar și în cazul aceluiași individ. Deschiderea relativ mică a coardelor vocale determină pentru sursa glotală o impedanță acustică relativ mare față de impedanța acustică a tractului vocal, modificările tractului vocal având o influență neglijabilă asupra fluxului glotal. Din punct de vedere electric, glota se poate echivala cu un generator de curent constant care are ca sarcină un circuit variabil în timp. Acest mod de excitație a tractului vocal produce semnalele denumite sonore (voiced), care au un caracter periodic, determinat de caracterul periodic al semnalului de excitație.

- **fluxul turbulent de aer** creat în unele puncte de constricție ale tractului vocal. Acest flux constituie sursa de producere a sunetelor continue, nevoalizate. Fluxul turbulent produce un zgomot acustic care excită necoerent tractul vocal. Deoarece spectrul acestui zgomot acustic este larg și relativ uniform, tractul vocal, prin caracteristica sa de transfer, are influența

determinantă asupra spectrului sunetelor produse în acest mod. Acest mod de excitație a tractului vocal produce semnalele denumite nesonore (unvoiced), care nu prezintă caracterul periodic, fiind mai apropiate de un zgomot.

- **presiunea** creată în unele puncte de închidere a tractului vocal. O reducere bruscă a presiunii prin deschiderea acestor puncte produce o excitație a tractului vocal cu o funcție treaptă de presiune, a cărui spectru este căzător cu frecvența. Excitația tranzitorie poate apare cu sau fără vibrația coardelor vocale, pentru a produce sunete explozive sonore sau surde. Punctele de închidere a tractului vocal pot fi: labial (la nivelul buzelor), dental (la nivelul dinților), sau palatal (la nivelul boltei palatine).

## 1.2 Modelarea mecanismului de producere a vorbirii

Mecanismul producerii vorbirii naturale se poate studia matematic pe baza unor modele mecanice sau electrice. Analiza cu aceste modele este aproximativă, circuitele simulatoare fiind simplificate, dar se pot evidenția astfel mod caracteristicile de bază ale semnalului vocal.

Modelul mecanic al producerii vorbirii, primul din punct de vedere istoric, asimilează tractul vocal cu o succesiune de tuburi și încearcă descrierea propagării undei acustice prin aceste tuburi cu ajutorul unor ecuații diferențiale, bazate pe teoria propagării undelor acustice și pe mecanica fluidelor de vâscozitate redusă (aerul) [FLA72].

Un model electric liniar pentru mecanismul producerii vorbirii [MAT98] se bazează pe următoarele ipoteze simplificatoare:

- sistemul de producere este liniar
- cavitatea bucală este reprezentată de un lanț de tuburi acustice de arii și lungimi diferite
- cavitatea nazală este modelată de un sistem liniar cuplat în paralel cu cavitatea bucală
- cuplajul între sursa de excitație și tractul vocal este neglijabil, ceea ce permite modelarea lor separată.

Sursa de excitație se consideră o succesiune periodică de impulsuri, modelată prin răspunsul unui filtru trece jos de ordinul 2, cu frecvența de tăiere de aproximativ 100 Hz, la o excitație idealizată  $e(n)$  reprezentată de o succesiune de impulsuri unitate:

$$e(n) = \sum_{k=0}^{\infty} \delta(n - kP), \quad (1.1)$$

unde  $P$  este perioada corespunzătoare frecvenței fundamentale, exprimată printr-un număr întreg de perioade de eșantionare.

Funcția de transfer a filtrului trece-jos ce modelează forma impulsurilor glotale are expresia:

$$H_G(z) = \frac{A}{(1 + \alpha z^{-1})(1 + \beta z^{-1})}; \quad A, \alpha, \beta \in \mathbb{R} \quad (1.2)$$

iar răspunsul  $y(n)$  al filtrului este dat de convoluția excitației idealizate cu funcția pondere  $g(n)$ :

$$y(n) = e(n) * g(n) = \sum_{k=0}^{\infty} g(n - kP). \quad (1.3)$$

Pentru sunetele sau fonemele nesonore (fricative sau plozive) sursa de semnal de excitație este un generator de zgomot alb.

Cu aceste considerente, se poate adopta pentru mecanismul producerii vorbirii naturale modelul discret prezentat schematic în figura 1.1.

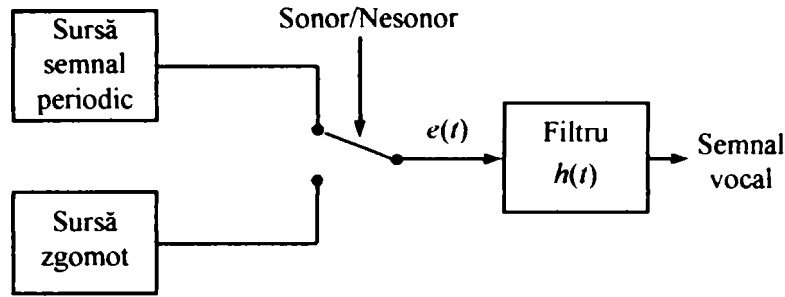


Fig. 1.1. Modelul mecanismului de producere a vorbirii.

Tractul vocal poate fi asimilat cu o succesiune de  $M$  tuburi acustice elementare, propagarea undei acustice prin aceste tuburi fiind modelată prin conectarea în cascadă a  $M$  rezonatori de ordinul 2.

Transmitanța tractului  $H_T(z)$  are forma:

$$H_T(z) = \frac{B}{\prod_{k=1}^M (1 + b_{1k}z^{-1} + b_{2k}z^{-2})}. \quad (1.4)$$

Fie funcția de transfer a unui rezonator de ordinul 2, având polii:

$$z_k = \rho_k \exp(j\theta_k) \text{ și } z_k^* = \rho_k \exp(-j\theta_k),$$

$$H_k(z) = \frac{B_k}{(1 - \rho_k e^{j\theta_k} z^{-1})(1 - \rho_k e^{-j\theta_k} z^{-1})} = \frac{B_k}{1 + b_{1k}z^{-1} + b_{2k}z^{-2}}, \quad (1.5)$$

cu  $b_{1k} = -2\rho_k \cos\theta_k$ ;  $b_{2k} = \rho_k^2$ .

Se poate arăta că  $|H_k(e^{j\omega T})|^2$  are un maxim la frecvența  $\omega_k$  ce satisface relația:

$$\cos(\omega_k T) = \frac{1 + \rho_k^2}{2\rho_k} \cos\theta_k. \quad (1.6)$$

În ipoteza polilor foarte apropiați de cercul unitate rezultă:

$$\cos(\omega_k T) \Big|_{\rho_k \rightarrow 1} = \cos\theta_k,$$

de unde se obține pentru frecvența polului  $k$  valoarea:

$$f_k = \frac{\theta_k}{2\pi T} = \frac{F_e}{2\pi} \theta_k = \frac{F_e}{2\pi} \arccos\left(\frac{-b_{1k}}{2\sqrt{b_{2k}}}\right), \quad (1.7)$$

unde  $F_e$  este frecvența de eșantionare. Maximele  $f_k$  ale fiecărui rezonator digital, date de relația (1.7) corespund cu frecvențele proprii de rezonanță ale tractului vocal.

Porțiunea finală a tractului vocal (regiunea buzelor) poate fi modelată prin transmitanța:

$$H_B(z) = C(1 - z^{-1}); \quad C \in R. \quad (1.8)$$

În final, funcția de transfer globală între excitația idealizată și semnalul vocal va fi:

$$H(z) = H_G(z)H_T(z)H_B(z) = \frac{\sigma(1 - z^{-1})}{(1 + \alpha z^{-1})(1 + \beta z^{-1}) \prod_{k=1}^M (1 + b_{1k}z^{-1} + b_{2k}z^{-2})}. \quad (1.9)$$



Dacă se consideră că unul dintre polii lui  $H_G(z)$  este foarte apropiat de cercul unitate ( $\beta \approx -1$ ), se obține pentru  $H(z)$  forma simplificată:

$$H(z) = \frac{\sigma}{A(z)}, \quad (1.10)$$

$$A(z) = (1 + \alpha z^{-1}) \prod_{k=1}^M (1 + b_{1k} z^{-1} + b_{2k} z^{-2}) = \sum_{i=0}^{2M+1} a(i) z^{-i}; \quad a(0) = 1. \quad (1.11)$$

Un sistem liniar caracterizat printr-o funcție de transfer având numai poli, de forma (1.10), se numește model autoregresiv (AR). Factorul  $\sigma$  se numește câștigul modelului. Notând cu  $e(n)$  excitația și cu  $x(n)$  răspunsul, avem:

$$X(z) = H(z)E(z) = \frac{\sigma}{A(z)} E(z) \quad (1.12)$$

Expresiei (1.12), cu polinomul  $A(z)$  de ordinul  $p = 2M+1$  nefactorizat din (1.11), îi corespunde în timp discret relația de recurență liniară [MAT98]:

$$x(n) + \sum_{i=1}^p a(i)x(n-i) = \sigma e(n) \quad (1.13)$$

de unde se obține pentru eșantionul curent al semnalului vocal valoarea:

$$x(n) = -\sum_{i=1}^p a(i)x(n-i) + \sigma e(n) \quad (1.14)$$

Răspunsul  $h(n)$  la impulsul unitate al modelului AR se obține din (1.14) pentru excitația particulară  $e(n) = \delta(n)$ . Rezultă:

$$h(n) = -\sum_{i=1}^p a(i)h(n-i) + \sigma \delta(n) \quad (1.15)$$

Modelul autoregresiv în timp discret din fig. 1.1 pentru producerea eșantioanelor semnalului vocal stă la baza majorității algoritmilor de prelucrare numerică a vorbirii. Trebuie menționat totuși că modelarea autoregresivă exprimată prin relația de recurență (1.14) nu caracterizează producerea unui semnal vocal decât cu o oarecare aproximație, modelarea AR având anumite limitări, determinate de simplificarea semnalului de excitație pentru model, de neglijarea efectului cuplării tractului nazal în paralel cu cel bucal și de presupunerea că modelul este staționar.

## 1.3 Caracteristicile semnalului vocal

### 1.3.1 Caracteristici temporale

Urmărind variația în timp a semnalului vocal, se constată prezența unor porțiuni de semnal care prezintă un caracter pronunțat periodic (pseudoperiodic), alături de porțiuni de semnal cu caracter apropiat de un zgomot aleator, evidențiindu-se zone de tranziție între cele două categorii [MAR91].

Din punct de vedere fonetic, aceste porțiuni corespund vocalelor, respectiv consoanelor. Astfel, vocalele au un caracter pseudoperiodic, iar consoanele un caracter pronunțat aleator datorită surselor diferite de excitație. Aceste diferențe permit stabilirea duratei sunetelor în procesul vorbirii curente. Durata unui sunet depinde de contextul vorbirii, de poziția sunetului în cuvânt, (inițial, mediu sau final) și de faptul că sunetul este sau nu accentuat.



Durata medie a vocalelor accentuate este 210 ms, iar a vocalelor neaccentuate este de 140 ms. Duratele consoanelor variază în limite mai largi. Astfel, consoanele plozive au o durată medie de 30 ms, iar cele fricative de 140 ms. Durata medie a tuturor consoanelor este de 95 ms (valori medii pentru limba română [MAR91]).

În figura 1.2 este reprezentat în timp semnalul vocal corespunzător cuvântului „transmite”.

Analiza variației amplitudinii semnalului vocal în timp permite aprecierea dinamicii semnalului. Cele mai puternice sunete (strigăte) au o intensitate relativă raportată la valoarea de prag de 94 dB. Vorbirea de tărie normală are în medie 75 dB, iar șoapta 40 dB. Pentru a cuprinde și vârfurile care depășesc media, se poate considera că dinamica vorbirii normale este de aproximativ 40 dB. Amplitudinea semnalului vocal este percepută de auzul uman sub forma tăriei sau intensității sunetului.

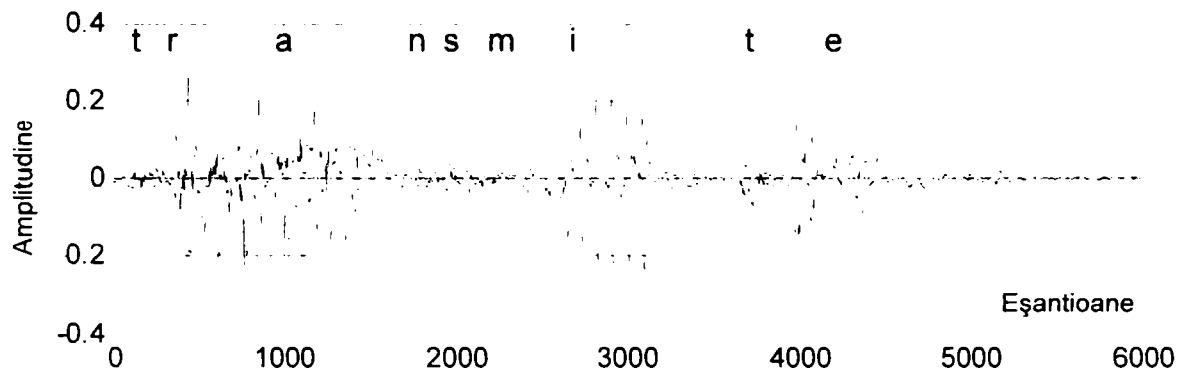


Fig. 1.2 Semnalul vocal corespunzător cuvântului „transmite”

În timpul procesului de vorbire, frecvența și amplitudinea fiecărui fragment din semnalul sonor determinat prin durata sa, variază într-un mod care determină prozodia vorbirii. Această caracteristică este determinată de foarte mulți factori, dintre care un rol deosebit îl au cei legați de tipul frazei sau propoziției (declarativă, exclamativă, interogativă). Cunoașterea prozodiei poate contribui la segmentarea semnalului vocal în cazul recunoașterii vorbirii sau la îmbunătățirea caracterului natural al vorbirii sintetizate.

### 1.3.2 Caracteristici spectrale

Semnalele vocale sunt semnale audio complexe cu energia concentrată în banda 50-8000 Hz. O mărime obiectivă care caracterizează acest semnal în domeniul frecvență este frecvența fundamentală, egală cu inversul perioadei de repetiție a impulsurilor create la nivelul coardelor vocale. Analiza semnalului vocal evidențiază variații ale frecvenței fundamentale în limitele a trei octave, de la 80Hz pentru voci grave până la 500Hz pentru voci de copii. Chiar și pentru același vorbitor, frecvența fundamentală se modifică în funcție de accent, intonație, stare emoțională, etc.

Ca mărime fizică obiectivă, frecvența fundamentală determină înălțimea sunetului. Două sunete de aceeași înălțime și tărie, dar emise de persoane diferite, nu sunt percepute de urechea umană ca fiind identice. Ceea ce le diferențiază este amplitudinea relativă a componentelor spectrale care reprezintă armonicile superioare ale frecvenței fundamentale, diferență percepută de auz sub formă de timbrul vorbirii.

Înălțimea și timbrul sunt elemente caracteristice vorbitorului. Auzul este sensibil la variații ale frecvenței fundamentale, putând sesiza variații ale frecvenței fundamentale sub 1% din valoarea sa absolută.

Studiul distribuției intensității semnalului vocal în funcție de frecvență evidențiază structura formantică, datorată frecvențelor naturale ale tractului vocal. Spectrul  $X(\omega)$  al semnalului vocal poate fi descris de un sistem de rezonatori serie  $T(\omega, \omega_i)$ :

$$X(\omega) = \prod_i AT_i(\omega, \omega_i)G(\omega)R(\omega) \quad (1.16)$$

unde  $G(\omega)$  este spectrul impulsurilor de excitație, iar  $R(\omega)$  este spectrul de radiație; produsul  $G(\omega)R(\omega)$  scade cu frecvența cu 6 dB/octavă.

Pentru vocale se pot evidenția mai mulți formanți, în general între 3 și 5, dar numai primii trei sunt semnificativi pentru inteligibilitatea semnalului vocal, iar formanții 4 și 5 sunt specifici vorbitorului.

Parametrii care caracterizează un formant sunt frecvența formantului, nivelul de intensitate și banda ocupată. Analiza unui mare număr de semnale a permis delimitarea domeniului de frecvență ocupat de primii trei formanți astfel:

- primul formant ocupă banda 200 - 850 Hz,
- al doilea formant ocupă banda 850 - 2350 Hz,
- al treilea formant ocupă banda 2100 - 3000 Hz.

Consoanele nu se pot caracteriza prin formanți deoarece spectrul lor nu prezintă zone de concentrare a energiei clar delimitate.

### 1.3.3 Caracteristici lingvistice

Limba constituie un sistem unitar, caracterizat printr-un fond fonetic, sintactic și semantic, care face posibilă comunicarea vocală între oameni. Limba se manifestă acustic prin secvențe de sunete cu semnificație în limba respectivă [MAR91].

Elementul de bază în studiul limbii este fonemul, definit ca fiind o unitate lingvistică abstractă, invariantă, care se manifestă și se reprezintă printr-un sunet. Totalitatea variantelor și varietăților unui fonem constituie alofonele fonemului respectiv. Fiecare limbă are propriile foneme. În general, fonemul nu poate fi identificat cu o literă, deși între foneme și literele alfabetului există o legătură strânsă.

Analizând relația existentă între foneme și literele alfabetului, limbile se pot împărți în două clase:

- limbi **etimologice**, pentru care fonemul diferă mult de transcrierea sa literală (exemplu: limbile franceză, engleză),
- limbi **fonetice**, pentru care valoarea fonetică este foarte apropiată de transcrierea literală. Din această categorie face parte și limba română.

Cel mai mare număr de foneme dintre limbile de largă circulație îl are limba engleză (42 foneme). Limba română are 31 de foneme, repartizate în două categorii: vocale, care se caracterizează prin faptul că pot forma singure silabe și consoane, care pot forma silabe doar însoțite de un fonem din prima categorie (vocale).

Pentru aprecierea informației purtată de un mesaj, se consideră un mesaj constituit din elemente discrete  $x_i$  aparținând unei mulțimi  $X$ . Dacă probabilitatea de apariție a priori a elementului  $x_i$  este  $p(x_i)$ , selectarea acestui simbol va aduce informația  $I = -\log_2 [p(x_i)]$ . Entropia sursei, exprimată în biți/simbol, este dată de relația [STA98]:

$$H(X) = -\sum_{i=1}^L p(x_i) \log_2 [p(x_i)] \quad (1.17)$$

Pentru fiecare limbă au fost determinate probabilitățile de apariție a fiecărui fonem  $\{p(x_i)\}$  pe baza unor analize statistice ale unui volum mare de texte din limba respectivă. Cu aceste valori ale probabilităților, relația (1.17) dă pentru entropia sursei valori sub 5 biți/simbol ( $H=4,9$  biți/simbol pentru limba engleză,  $H=4,73$  biți/simbol pentru limba franceză). Funcția  $H(X)$  atinge valoarea maximă atunci când fonemele  $x_i$  sunt echiprobabile iar maximul este dat de relația:

$$H_{\max}(X) = \log_2 L \quad (1.18)$$

Cu relația (1.18), entropia maximă este  $H_{max} = 5,39$  biți/simbol pentru limba engleză,  $H_{max} = 5,17$  biți/simbol pentru limba franceză și  $H_{max} = 5,08$  biți/simbol pentru limba română. Se constată astfel că fiecare fonem se poate coda folosind 5 sau maxim 6 biți. Considerând că într-o conversație curentă se pronunță în medie 10 foneme pe secundă, rezultă un debit informațional de 50 – 60 biți pe secundă. Comparând cu debitul de 64 kbps pentru același semnal eșantionat la 8 kHz și cuantizat pe 8 biți, se evidențiază redundanța enormă a mesajului vocal. De fapt, mesajul vocal cuprinde mult mai multă informație decât același mesaj sub formă scrisă, informație referitoare la vorbitor și la starea sa emoțională și psihică, informație care este cuprinsă în intonația și timbrul mesajului vocal.

### 1.3.4 Caracteristici statistice

Semnalul vocal este o realizare a unui proces aleator care, în general, nu este staționar, dar care se poate considera staționar și ergodic pe durate de timp mici, corespunzător unor segmente de semnal de 10 – 50 ms [STA98]. Segmentele de semnal de durată mai mare, de ordinul secundelor sau zecilor de secunde sunt realizări ale unui proces nestaționar.

Corespunzător acestei delimitări, operațiile efectuate asupra segmentelor scurte de semnal (10 – 50 ms) sunt considerate prelucrări pe termen scurt (statistica pe termen scurt), iar prelucrările efectuate pe segmente mai lungi de ordinul secundelor, se consideră prelucrări pe termen lung (statistica pe termen lung).

Pentru stabilirea legii de repartiție a semnalului vocal ar trebui efectuate evaluări la diferite momente  $t_k$  ale densității de repartiție a variabilei aleatoare  $x(\omega, t_k)$ , determinând o secțiune a semnalului aleator la momentul considerat. Se obține în acest fel o densitate de repartiție dependentă de momentul considerat,  $p_x(\xi, t_k)$ . Se constată experimental că densitatea de repartiție  $p_x$  nu depinde de momentul de observare  $t_k$ . Fie  $\{x(n)\}$  eșantioanele unei realizări, din care se consideră eșantioanele cuprinse în intervalul simetric  $[-N, N]$ . Pentru un număr  $\xi$  fixat, se determină numărul de eșantioane  $N_\xi$ , care au valoarea cuprinsă în intervalul  $[\xi - \Delta\xi/2, \xi + \Delta\xi/2]$ , unde  $\Delta\xi$  este o valoare relativ mică față de  $\xi$ . Densitatea de probabilitate pentru realizarea considerată se calculează cu relația:

$$p_x(\xi) = \lim_{\substack{n \rightarrow \infty \\ \Delta\xi \rightarrow 0}} \frac{N_\xi}{2N + 1} \quad (1.19)$$

Funcția (1.19) nu are o expresie analitică ci doar una grafică, obținută experimental, reprezentată, împreună cu câteva densități de repartiție mai cunoscute, în figura 1.3 [STA98].

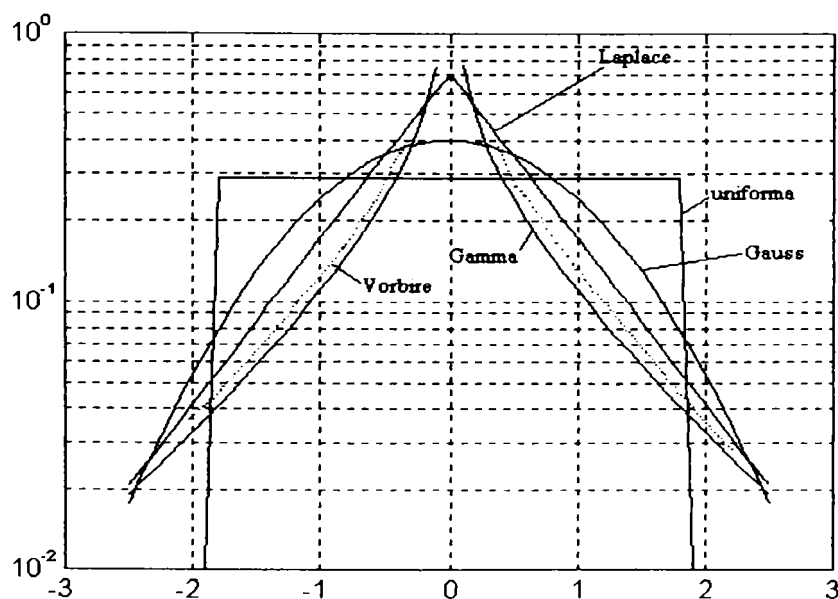


Fig. 1.3. Repartiția statistică a semnalului vocal

Comparând această funcție cu unele legi de repartiție simetrice, de valoare medie nulă, se observă că semnalul vocal se apropie cel mai mult de legea Gamma (1/2), cu expresia analitică [STA98]:

$$p_x(\xi) = \left[ \frac{\sqrt{3}}{8\pi\sigma_x|\xi|} \right]^{1/2} \exp\left[ \frac{-\sqrt{3}|\xi|}{2\sigma_x} \right] \quad (1.20)$$

Valorile mari ale densității de repartiție din vecinătatea lui zero se datorează ponderii mari a pauzelor din cadrul semnalului vocal, când semnalul ia valori foarte apropiate de zero.

Valoarea medie a semnalului vocal, presupus staționar, se poate calcula cu relația:

$$\mu_x = \int_{-\infty}^{+\infty} \xi p_x(\xi) d\xi = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N x(n) \quad (1.21)$$

Pentru semnalul vocal, această medie se presupune nulă și nu conține nici o informație utilă. În cazul în care media temporală a semnalului nu este nulă, această valoare se poate extrage (scădea) din semnal pentru unele simplificări ale calculelor.

## 1.4 Scurt istoric

Înțelegerea mecanismului producerii vorbirii și crearea unor mecanisme care să imite vorbirea naturală au constituit probleme pentru spiritul cercetător al omului încă din antichitate. Dezvoltarea mecanicii și acusticii, a electricității și a prelucrării analogice a semnalelor, iar mai apoi a prelucrării digitale a semnalelor, au reprezentat etape fundamentale pentru cunoașterea mecanismului de producere și transmitere la distanță a semnalului vocal.

Deși se cunosc exemple de „oracole vorbitoare” realizate încă din antichitate și experiențele făcute de Albertus Magnus (1198 – 1280) sau Roger Bacon (1214 – 1294), prima încercare documentată de realizare a unui mecanism care să modeleze mecanismul natural de vorbire este atribuită lui Ch. G. Kratzenstein și datează din 1779, când acesta a reușit construcția unui rezonator acustic similar tractului vocal uman, cu care putea produce cinci vocale diferite. În aceeași perioadă, Wolfgang Ritter von Kempelen face experimentări în acest domeniu și în 1791 prezintă o adevărată mașină vorbitoare (aflată în prezent la Deutsches Museum, München), cu care un operator puțin antrenat putea produce un semnal vocal, publicând în același an o carte (*Mechanismus der menschlichen Sprache nebst Beschreibung einer sprechenden Maschine -- Mecanismul vorbirii umane, urmat de o descriere a unei mașini vorbitoare*) în care își descrie realizarea. Sunt cunoscute și alte realizări îmbunătățite de mașini vorbitoare din secolele XIX – XX, legate de nume cunoscute ale epocii: Herman von Helmholtz, Charles Wheatstone, sau Alexander Graham Bell [WEB1], [WEB2].

Prima încercare de modelare electrică a mecanismului producerii vorbirii aparține lui J. Q. Stewart (1922), care folosește un semnal de excitație periodic și două circuite rezonante pentru a modela tractul vocal, reușind generarea unor vocale izolate. Primul dispozitiv considerat un adevărat sintetizor de vorbire electric este cel realizat de H. Dudley în 1939 și prezentat la New York sub numele de Voder (*Voice Operating Demonstrator*). Dispozitivul lui Dudley este inspirat de vocoderul analogic dezvoltat în aceeași perioadă la Bell Laboratories, destinat transmiterii semnalului vocal la distanță, folosind un canal cu bandă redusă. Dezvoltarea tehnicilor de prelucrare numerică a semnalelor și evoluția spectaculoasă a microprocesoarelor și a procesoarelor de semnal din ultimele două decenii au condus pe de o parte la dezvoltarea unor metode de sinteză a semnalului vocal, ajungându-se în prezent la sinteza directă pe baza textului (TTS), iar pe de altă parte, au condus la elaborarea și standardizarea unor noi metode de codare, de calitate bună, la debit redus, prezentate succint în paragraful 1.6.

## 1.5 Performanțele sistemelor de codare a semnalului vocal

Ansamblul codor – decodor pentru semnalul vocal este evaluat din punct de vedere calitativ pe baza mai multor criterii. Unele dintre aceste criterii sunt obiective și se pot exprima printr-o mărime cu semnificație clar definită, cum ar fi debitul binar exprimat în biți/s, în timp ce altele sunt subiective și dificil de exprimat printr-o mărime fizică, ca de exemplu calitatea semnalului vocal. Criteriile pe baza cărora se apreciază performanțele sistemelor de codare sunt prezentate în continuare.

### 1.5.1 Calitatea semnalului vocal

Prin codarea semnalului vocal se urmărește obținerea unei calități cât mai ridicate, adică a unui nivel cât mai redus al distorsiunilor la un debit binar dat.

Aprecierea obiectivă a calității semnalului vocal obținut după decodarea semnalului codat constituie o problemă dificilă, având în vedere faptul că semnalul este recepționat de mecanismul uman de percepție a sunetelor, cu particularitățile sale specifice, deși fiecare ascultător are o reprezentare intuitivă a noțiunii de calitate a semnalului. Spre exemplu, în [QBC88] sunt prezentate 322 de tipuri de distorsiuni care pot afecta semnalul vocal.

Primele sisteme de codare a semnalului vocal, cu aplicații în domeniul comunicațiilor militare, erau apreciate din punct de vedere a calității printr-un singur criteriu: inteligibilitatea. Odată cu pătrunderea metodelor de codare a semnalului vocal în domeniul telecomunicațiilor și a aplicațiilor destinate unor categorii mai largi de beneficiari, acest criteriu a devenit insuficient. Este de așteptat, din acest punct de vedere, ca semnalul să fie cât mai natural, dar este foarte dificil ca această calitate a semnalului, de a fi „natural” să fie măsurată cu precizie, deoarece naturalețea presupune mai multe aspecte, și orice încercare de a exprima naturalețea printr-un număr va fi asociată cu o „simplificare” inevitabilă a acestei noțiuni complexe.

În domeniul telecomunicațiilor se acceptă o clasificare a semnalului vocal din punct de vedere a calității în patru clase:

- calitate radio,
- calitate CCITT,
- calitate comunicații,
- calitate sintetică.

Calitatea radio presupune un semnal vocal de foarte bună calitate, de categoria semnalelor asigurate de radiodifuziunea MF, având banda de frecvențe la peste 10 kHz și care în comunicațiile digitale se poate obține la un debit binar de cel puțin 64 kbps. Calitatea CCITT se referă la semnalele de calitate bună obținute prin comunicațiile analogice clasice sau la un debit binar de până la 16 kbps. Calitatea comunicații a semnalului vocal acceptă unele pierderi de calitate a semnalului, care rămâne totuși natural și complet inteligibil. Calitatea sintetică caracterizează un semnal care este inteligibil, dar este asociat cu un caracter nenatural, sintetic și uneori nu permite recunoașterea vorbitorului.

Pentru aprecierea calității semnalului se folosesc două categorii de metode:

- metode subiective
- metode obiective.

Metodele subiective presupun aprecierea calității semnalului de către unul sau mai mulți ascultători umani, care atribuie semnalului supus testului un punctaj sau scor. Există metode subiective care apreciază doar inteligibilitatea semnalului și metode subiective care apreciază calitatea în ansamblu a semnalului vocal. Deși metodele subiective par a oferi aprecieri corecte asupra calității semnalului, în concordanță cu percepția umană, aceste metode prezintă totuși dezavantajul că sunt laborioase, necesită timp, condiții deosebite și în unele cazuri un număr relativ mare de ascultători cu un minim de experiență.



- **Teste subiective de inteligibilitate**

Pentru testarea subiectivă a inteligibilității semnalului vocal se utilizează testele DRT (Diagnostic Rythme Test) și DALT (Diagnostic Alliteration Test), care folosesc un vocabular corespunzător ales pentru a testa diferite tipuri de inteligibilitate [QBC88]. Astfel, fiecare test constă din 96 de perechi de cuvinte ușor confundabile, care diferă printr-o singură caracteristică, și care se pronunță izolat. Cuvintele folosite pentru testul DRT diferă doar prin consoana inițială (exemplu: carte – parte), iar cele folosite în testul DALT diferă doar prin final. Testul constă în a asculta câte unul din cuvintele din aceste perechi, pronunțate izolat, într-o ordine aleatoare, iar rezultatul exprimă procentual numărul de cazuri în care cuvintele au fost identificate corect.

- **Teste subiective de calitate**

Testarea subiectivă a calității semnalului în ansamblu, se face cu testul MOS (Mean Opinion Score), ce presupune un număr de 12 – 24 ascultători (32 – 36 în testele CCITT) care sunt familiarizați pentru început cu semnale vocale de diverse calități, după care apreciază semnalul supus testului cu ajutorul unei scări cu 5 trepte de calitate, prezentată în tabelul 1.1. Scorul final rezultă prin medierea calificativelor acordate de toți ascultătorii.

Tabelul 1.1 Scara MOS

Scara MOS	Calitatea semnalului
1	nesatisfăcătoare
2	slabă
3	satisfăcătoare
4	bună
5	excelentă

Calitatea excelentă după scara MOS presupune un semnal complet lipsit de zgomot și care nu poate fi diferențiat prin nimic față de semnalul original. Calitatea nesatisfăcătoare este atribuită unui semnal cu caracter sintetic, puternic afectat de zgomot și cu pierderi de semnal.

Metodele obiective de apreciere a calității semnalului vocal încearcă să cuantizeze această mărime și să o exprime printr-un număr care să reflecte cât mai exact noțiunea intuitivă de calitate [VOR99-1], [VOR99-2], [CBA03-5]. Dintre acestea, cele mai uzuale vor fi prezentate în continuare, grupate în două categorii.

### Măsurarea obiectivă în domeniul timp

- **Raportul semnal – zgomot**

O măsură obiectivă a calității unui semnal vocal codat este raportul semnal – zgomot, exprimat prin raportul dintre puterea semnalului și puterea zgomotului de cuantizare, în decibeli:

$$SNR = 10 \lg \left\{ \frac{\sum_{n=0}^M s^2(n)}{\sum_{n=0}^M (s(n) - \hat{s}(n))^2} \right\} \quad (1.21)$$

unde  $s(n)$  reprezintă eșantioanele semnalului vocal original iar  $\hat{s}(n)$  eșantioanele semnalului reconstituit. Raportul semnal – zgomot ca măsură a calității pentru semnalul vocal are dezavantajul că ar putea fi determinat în principal de segmentele de energie mare din semnal, în timp ce anumite segmente, cu energie mică, dar importante din punct de vedere al calității percepute, să fie subestimate. În acest sens, raportul semnal – zgomot este o mărime care caracterizează pe termen lung calitatea semnalului reconstituit, și care tinde să ascundă anumite imperfecțiuni de moment în reconstrucția semnalului, manifestate mai ales în cazul

semnalelor de nivel mic. Acest dezavantaj este confirmat și de coeficientul de corelație redus (0,24) între  $SNR$  și scorul  $DAM$  [HPL98].

- **Raportul semnal – zgomot segmental**

Se poate defini, pentru a evita acest neajuns, un raport semnal - zgomot pe termen scurt, calculat ca o medie a raportului semnal – zgomot (1.21) pentru fiecare segment de semnal (Segmental SNR - SEGSNR):

$$SEGSNR = \frac{10}{L} \sum_{i=0}^{L-1} \lg \left\{ \frac{\sum_{n=0}^{N-1} s^2(iN + n)}{\sum_{n=0}^{N-1} (s(iN + n) - \hat{s}(iN + n))^2} \right\}, \quad (1.22)$$

unde  $N$  este lungimea segmentului, în eșantioane, iar  $L$  este numărul de segmente considerat.

SEGSNR poate atinge, pe durata anumitor segmente de semnal, valori foarte mari sau dimpotrivă, valori foarte mici, care ar putea afecta pe ansamblu valoarea globală a SEGSNR, fără a reflecta însă realitatea percepută. Astfel, SEGSNR poate lua valori negative mari pe durata unor segmente de liniște, cu energie mică, afectând astfel valoarea medie. Pentru a evita acest lucru, se poate stabili o limită negativă, în domeniul (0 – 20 dB) și o limită pozitivă (uzual 35 dB), înlocuind cu aceste limite valorile care depășesc acest interval.

Raportul semnal – zgomot și raportul semnal – zgomot segmental, ca mărimi obiective, sunt afectate de anumiți factori, ca de exemplu întârzierea și nu pot exprima unele aspecte subiective ale percepției semnalului vocal, justificând introducerea conceptului de entropie perceptuală, exprimat prin rata minimă de bit necesară pentru a transmite semnalul fără ca să se poată percepe o diferență între semnalul original și cel codat [JJH98].

Alte mărimi obiective în domeniul timp care se pot folosi pentru aprecierea calității semnalului sunt câștigul de predicție și energia erorii.

### Măsurarea obiectivă în domeniul frecvență

- **Diferența spectrală ponderată**

Diferența spectrală ponderată este bazată pe un model al sistemului auditiv, care calculează, cu ajutorul unor filtre distanțele între spectrul semnalului original și codat în fiecare bandă critică, ponderate după criteriile perceptuale [ITU98-1], [CBA03-5].

- **Distorsiunea spectrală logaritmică**

Distorsiunea spectrală logaritmică este legată de reprezentarea spectrelor semnalului original și codat cu ajutorul predicției liniare. Dacă se consideră spectrele de putere pentru semnalul original și cel codat,  $S(j\omega)$  și respectiv  $\hat{S}(j\omega)$ , atunci distorsiunea spectrală logaritmică bazată pe norma  $L_p$  este definită prin:

$$d_{SD}^{(p)} = \sqrt[p]{\frac{1}{2\pi} \int_{-\pi}^{\pi} \left| 10 \log_{10} \left[ \frac{S(j\omega)}{\hat{S}(j\omega)} \right] \right|^p d\omega} \quad [\text{dB}] \quad (1.23)$$

Modul de calcul depinde de modul de reprezentare a coeficienților de predicție [HPL98].

### 1.5.2 Rata de bit

Rata de bit este exprimată prin numărul de biți produși într-o secundă de codor, biți care urmează să fie stocați sau transmiși prin canalul de comunicație. Obiectivul codării semnalului vocal este obținerea unei rate de bit cât mai scăzute [WMG96], [COS93]. Cerințele privind rata de bit și calitatea semnalului obținut sunt contradictorii și, deși este dorit un debit binar cât mai mic la o calitate înaltă, se acceptă întotdeauna un compromis.

În multe aplicații codarea semnalului vocal are ca scop transmiterea acestui semnal printr-un canal de comunicație a cărui capacitate poate fi variabilă în timp. În aceste cazuri, este de dorit ca debitul binar al codorului să fie corelat cu capacitatea canalului, astfel încât în fiecare moment să se asigure transmisia la capacitatea maximă oferită de canal. Modificarea debitului binar al codorului, în acest scop, se poate realiza pe două căi.

În prima variantă, denumită **codare multimod**, codarea se poate face la rată variabilă, iar transmițătorul și receptorul negociază o rată de transmisie dependentă de starea canalului, înainte de a începe transmisia propriu-zisă; apoi codarea semnalului se va face la această rată, sau la o valoare cât mai apropiată [BTS02].

A doua variantă o reprezintă **codarea „împachetată”** (embedded), în care caz fluxul de biți care rezultă în urma codării la o rată redusă este împachetat, sau altfel spus, este conținut în fluxul de biți corespunzător codării la rată înaltă. În acest caz, decodorul primește fluxul corespunzător unei anumite calități a semnalului, care este funcție de capacitatea canalului, de natura semnalului sau eventual de opțiunea utilizatorului.

### 1.5.3 Complexitatea algoritmică

Complexitatea algoritmului este strâns legată de noțiunea de operare în timp real a circuitelor de codare și decodare și se poate exprima obiectiv prin numărul de instrucțiuni care trebuie realizate, în milioane de instrucțiuni pe secundă – MIPS. Totuși acest număr este dependent de procesorul pe care este implementat algoritmul. Astfel, pentru același algoritm, numărul de instrucțiuni pe secundă obținut în cazul unei implementări pe un procesor de semnal cu virgulă flotantă trebuie înmulțit cu un factor de 2 – 3 în cazul folosirii unui procesor de semnal lucrând în virgulă fixă. În cele mai multe cazuri, este necesar ca algoritmul să permită codarea în timp real folosind un anumit DSP. Decodarea este aproape întotdeauna de o complexitate mult mai redusă decât codarea, dar în unele aplicații, cum ar fi aplicațiile de tipul difuziune (radio sau TV), sau de redare a sunetelor stocate, este importantă minimizarea complexității decodorului, în timp ce complexitatea codorului nu este atât de importantă. În unele cazuri se impun și cerințe privind limitarea capacității memoriei RAM necesare pentru memorarea unor date folosite în procesul de codare (tabele, dicționare, etc.). În unele aplicații care presupun implementarea algoritmilor de codare pe circuite integrate de tip ASIC (Application Specific Integrated Circuits) complexitatea algoritmului determină complexitatea circuitului, exprimată prin numărul de tranzistori sau numărul de porți echivalente și determină, indirect, consumul de energie al circuitului, esențial în aplicații portabile.

### 1.5.4 Întârzierea semnalului

Este reprezentată de suma tuturor întârzierilor care intervin între semnalul de intrare în codor și semnalul de ieșire din decodor. Există două cauze care produc o întârziere a semnalului de ieșire.

Pe de o parte apare o întârziere, de valoare relativ mică, datorată algoritmului de prelucrare și timpului necesar pentru calcule, iar pe de altă parte, apar o întârziere datorată faptului că majoritatea algoritmilor presupun prelucrarea unui număr oarecare de eșantioane, grupate într-un cadru, întârziere dependentă de mărimea cadrului. Întârzierea semnalului poate lua valori de sub 1 ms pentru algoritmii care prelucrează semnalul eșantion cu eșantion și poate ajunge la valori de 60 – 80 ms pentru algoritmii care prelucrează semnalul pe cadre.

O întârziere prea mare este deranjantă în transmisiunile în timp real, mai ales în cazul unei transmisii duplex și, în unele cazuri, poate conduce la apariția ecoului. Întârzierea nu este foarte importantă în cazul transmisiilor simplex, ca de exemplu radiodifuziunea, și este total irelevantă în cazul unor semnale vocale redade de pe diverse medii de stocare.



### 1.5.5 Banda semnalului

Banda semnalului este un indicator al calității semnalului redat și este aleasă în concordanță cu scopul pentru care este realizată codarea semnalului. Banda tipică pentru transmisiunile de tip telefonic este (300 – 3400) Hz.

Pentru anumite categorii de comunicații speciale (militare) se acceptă o bandă mai redusă, cu condiția asigurării unei inteligibilități corespunzătoare a semnalului.

Codarea de bandă largă necesită o limită superioară a benzii audio de 7 până la 20 kHz, pentru transmisiuni de voce de calitate ridicată sau pentru aplicații multimedia [CBA00].

### 1.5.6 Robustețea

Ansamblul codor – decodor trebuie să prezinte robustețe față de erorile introduse de canalul de transmisiuni, precum și față de zgomotul canalului, fadingul, interferența intersimbol sau pierderile de pachete [GKM99]. Pe de altă parte, acest ansamblu trebuie să fie robust față de codările și decodările succesive, eventual realizate prin metode diferite și față de zgomotele ambientale, pentru semnalele vocale transmise în aplicații reale: zgomotul străzii, zgomot de mașină, de avion, etc.

## 1.6 Standardizarea în domeniul codării semnalului vocal

Standardele folosite pentru codarea semnalului vocal în scopul transmiterii acestuia prin rețeaua telefonică fixă PTSN sunt elaborate de ITU, organism component al ISO. În acest sens, ITU a elaborat o serie de standarde care includ diferite metode de codare a semnalului vocal, de debit redus și foarte redus, ca: G.711 (PCM), G.726 și G.727 (ADPCM), G.728 (LDCELP), G.729 (ACELP), G.723.1 (multimod) și G.722 (codare semnal de bandă 7 kHz) [ITU90], [ITU96-1], [ITU96-2], [JAL03].

ITU a participat și la elaborarea unor standarde privind interoperabilitatea între rețelele de telecomunicații, între care trebuie menționat standardul H.323 care stabilește modul de codare a semnalului audio pentru transmiterea prin Internet, prin serviciul VoIP [ITU98-2].

Pentru estimarea perceptuală obiectivă a calității semnalului vocal codat, ITU a publicat standardele P.861 și P.862 [ITU98-1].

ISO a elaborat standardele de codare din familia MPEG [ISO98-1]. Dintre acestea, standardul MPEG-2 include codarea semnalului audio pe trei nivele de complexitate, codecul audio de nivel 3 fiind binecunoscutul MP3, folosit în mod curent pentru stocarea sau transmiterea muzicii prin Internet. Standardul ISO MPEG-4 include un standard de codare audio structurată, în care semnalul vocal sau audio este codat sub formă de „obiecte”, folosind diferite metode de codare, specificate în preambulul fluxului de date. Pentru codarea semnalului vocal MPEG-4 folosește metodele HVXC [ISO98-4] sau CELP [ISO98-2], iar pentru semnalul audio codarea în domeniul timp – frecvență [ISO98-3].

Standardele pentru codarea semnalului vocal în telefonia celulară în Europa sunt elaborate de ETSI prin subcomitetul GSM. Primul standard pentru telefonia celulară este bazat pe o variantă a codorului CELP (RPE – LTP) [HEL89], [KDS86]. Codoarele actuale folosite în rețeaua celulară europeană GSM sunt bazate pe codecul ACELP, în versiunile GSM îmbunătățit și adaptiv – multirată [JAR97],[GSM97], [GSM98]. Se preconizează ca în viitorul apropiat ETSI și ITU – T să anunțe noi standarde privind codarea semnalului vocal de bandă largă, pentru sistemul GSM 3G.

TIA este un organism similar ETSI, care elaborează standardele pentru codarea semnalului vocal folosite în telefonia celulară în SUA și Japonia, între care menționăm standardul IS54 (VSELP) [GMJ91], precum și standardul TR41 pentru VoIP.

Codarea semnalului vocal pentru telefonia prin Internet (VoIP) a apărut ca urmare a dezvoltării explozive a Internetului. Primul standard în domeniul VoIP l-a constituit recomandarea H.323 elaborată de ITU, care stabilește un protocol pentru comunicații multimedia printr-o rețea LAN cu comutație de pachete [ITU98-2]. Un subset al protocolului H.323 constituie platforma pentru telefonia prin Internet, care recomandă codoarele cu rată ridicată de bit G.711 (voce cu banda 3,4 kHz la debit de 48, 56 sau 64 kbps) și G.722 (voce și muzică cu banda 7 kHz la debit 48, 56 sau 64 kbps), respectiv codoarele cu rată redusă G.728 (16 kbps), G.723 (5,3 și 6,5 kbps) și G.729 (8 kbps) [ITU92], [ITU93], [ITU96-1], [ITU96-2].

O propunere de standard competitiv cu H.323 este RFC 2543, care folosește un protocol denumit SIP (Session Initiation Protocol) și care suportă atât TCP cât și alte protocoale pentru transmisii de date prin Internet [HAN99].

Modul de transmisie cu comutație de pachete folosit în rețeaua Internet determină câteva particularități specifice pentru transmisia semnalului vocal în timp real prin Internet. Congestiile rețelei pot conduce la pierderi de pachete și astfel pot apare artefacte deranjante. Telefonia VoIP de înaltă calitate se bazează pe codarea cu rată variabilă a sursei și a canalului, pe tehnici de mascare a pierderilor de pachete și pe tehnici de eliminare a efectelor jiterului și întârzierilor de pachete [RGR00], [GKN00].

În final trebuie amintit Departamentul apărării (DoD – Department of Defense) al SUA, care, prin Voice Processing Consortium (DDVPC) a elaborat o serie de standarde privind codarea semnalului vocal la rată de bit redusă, cu aplicabilitate în comunicațiile speciale folosite în armată sau alte servicii similare. În această categorie se înscrie primul standard LPC, cunoscut sub numele FS – 1015 precum și versiunile mai noi care includ standardul FS – 1016 (CELP) [CTT91], [CTT94] și MELP [SCC97].

Standardele prezentate sunt sintetizate în tabelul 1.2.

Tabelul 1.2. Standarde în domeniul codării semnalului vocal

Standard	Metoda de codare	Rata bit [kbps]	Calitate (Scor MOS)	Întârziere algoritmică [ms]
ITU-G.711	PCM	64	4.3	0.125
ITU-G.721	ADPCM	32	4.0	0.125
ITU-G.726	VBR-ADPCM	16, 24, 32,40	2.0, 3.2, 4.0, 4.2	0.125
ITU-G.727	Embedded-ADPCM	16, 24, 32,40	—	0.125
ITU-G.728	LD-CELP	16	4.0	0.625
Inmarsat-B	APC	9.6/16	comunicații	20
GSM Full-rate	LTP-RPE	13	3.7	20
GSM-EFR	ACELP	13	4.0	20
Skyphone	BT-MPLPC	8.9	3.5	28
DAMPS Full-rate IS54	VSELP	7.95	3.6	20
ITU-G.729	CSA-CELP	8	4.0	15
IS-96	CELP	1, 2, 4, 8	3.5	-
JDC Japanese	VSELP	6.7	comunicații	20
GSM Half-rate	VSELP	5.6	3.5	24.375
ITU-G.723	A/MP-MLQ CELP	5.27/6.3	comunicații	37.5
DOD FS1016	CELP	4.8	3.0	45
TETRA	ACELP	4.56	comunicații	35
Inmarsat-M	IMBE	4.15	3.4	78.75
JDC Jap. Half-rate	PSI-CELP	3.45	comunicații	40
DOD FS1015	LPC-10	2.4	sintetic	22.5

# Capitolul 2

## Metode de codare a semnalului vocal

Deși metodele de codare a semnalului vocal propuse și dezvoltate în literatura de specialitate sunt foarte numeroase și diverse, este unanim acceptată o clasificare a acestor metode în trei mari categorii, dintre care două sunt de bază și o a treia reprezintă o combinație între primele două:

- *codarea formei de undă*
- *codarea sursei de semnal*
- *codarea hibridă*

Metodele de codare a formei de undă fac abstracție de natura semnalului, fiind, prin urmare, aplicabile oricărui tip de semnal, nu neapărat de vorbire. Comparativ cu a doua categorie, aceste metode sunt în general mai robuste în ceea ce privește semnalul de intrare, în sensul că se comportă bine față de o clasă largă de semnale de intrare, incluzând semnalele audio de tip muzică, dar debitul binar rezultat este mai mare. Ceea ce se codează în acest caz sunt parametrii semnalului ca atare, deosebindu-se, din acest punct de vedere, metode care codează forma de undă în domeniul timp, respectiv metode care codează forma de undă în domeniul frecvență. Codarea se poate face utilizând metode de cuantizare scalară, care codează fiecare eșantion al semnalului vocal în mod independent față de eșantioanele vecine, sau metode de cuantizare vectorială, care codează simultan mai multe eșantioane ale semnalului vocal, grupate într-un vector. În domeniul frecvență, codarea se poate face descompunând semnalul în subbenzi, sau folosind diferite transformări ale semnalului. Codarea poate fi non-parametrică sau parametrică, după cum se aplică direct semnalului vocal sau unei reprezentări parametrice a acestuia, bazate în general pe transformări unitare sau modele ale semnalului. Dintr-un alt punct de vedere, codarea poate fi cu memorie, respectiv fără memorie, dependent de modul în care regulile de codare țin cont sau nu țin cont de valorile trecute ale eșantioanelor de intrare sau ieșire. Metodele de codare scalară în domeniul timp a semnalului vocal cuprind modulația impulsurilor în cod (Pulse code Modulation - PCM), PCM diferențial (DPCM) și modulația delta (Delta Modulation - DM). Metodele de cuantizare vectorială vor fi prezentate pe larg în cap. 4. Dintre metodele de codare a formei de undă în domeniul frecvență se vor prezenta codarea în subbenzi, codarea prin transformare și codarea prin analiză și sinteză sinusoidală. Ca un caz particular al metodelor de codare în subbenzi va fi prezentată codarea perceptuală a semnalelor audio.

Metodele de codare a sursei de semnal realizează o reprezentare parametrică a sursei de semnal vocal, codând acest semnal cu ajutorul parametrilor sursei, sau, cu alte cuvinte, simulând, într-un fel sau altul, mecanismul natural de producere a semnalului vocal. Din acest motiv, codoarele din această categorie sunt desemnate prin termenul generic *vocoder* (voice coder). Față de metodele din prima categorie, această clasă de metode asigură în general un debit binar mai redus, dar performanțele codării se degradează puternic pentru semnale de altă natură decât cele vocale. În cazul acestor metode, pentru etapa de codare se mai folosește și termenul de analiză, deoarece în această etapă, prin analiza semnalului vocal, se extrag

parametri care îl reprezintă, iar pentru etapa de decodare se folosește termenul de sinteză. Analiza presupune prelucrarea semnalului vocal în sensul extragerii parametrilor modelului, care apoi sunt transmiși la decodor. Determinarea parametrilor prin care se reprezintă modelul se poate face în buclă deschisă sau în buclă închisă. În cel de-al doilea caz, codorul include o replică a decodorului, justificând denumirea de analiză-prin-sinteză. Modelarea sursei semnalului vocal stă la baza celor mai cunoscute metode de codare, dintre care unele reprezintă standarde în domeniu. De remarcat că au fost propuse unele modele denumite articulatorii, care modelează direct mecanismul uman de producere a vorbirii naturale, fiecare organ component al acestui mecanism fiind explicit reprezentat în model [SJS87], [FMS92]. Deși aceste modele ar putea produce un semnal vocal natural la un debit binar foarte redus [COS93], există dificultăți în extragerea parametrilor articulatorii ai modelului din secvența de semnal vocal aplicată la intrare. Parametrii care caracterizează modelul, la rândul lor, se pot cuantiza folosind oricare dintre metodele cunoscute, fie scalare, fie vectoriale, sau combinații ale acestora. Dintre metodele de codare a sursei se vor prezenta în continuare patru categorii reprezentative, și anume: vocoderul canal, vocoderul formantic, vocoderul homomorfic și vocoderul cu predicție liniară.

Metodele de codare hibride îmbină eficiența vocoderelor în ceea ce privește rata de bit cu calitatea codorului pentru forma de undă și asigură în acest mod o calitate bună la o rată de bit medie sau mică. Codarea hibridă folosește un model pentru producerea semnalului vocal, la fel cu metodele de codare a sursei, dar semnalul de excitație pentru acest model este codat cu metodele folosite de codoarele de formă de undă. În această categorie se înscriu codoarele care folosesc analiza prin sinteză, din familia vocoderelor LPC, la care semnalul de excitație este reprezentat de impulsuri multiple (MPLP), impulsuri regulate (RPELP), coduri (CELP) sau sumă de vectori (VSELP).

În figura 2.1 se reprezintă situarea celor trei categorii de metode pe o scară logaritmică a debitului binar realizat, indicându-se totodată și calitatea semnalului vocal obținut [KNT87].

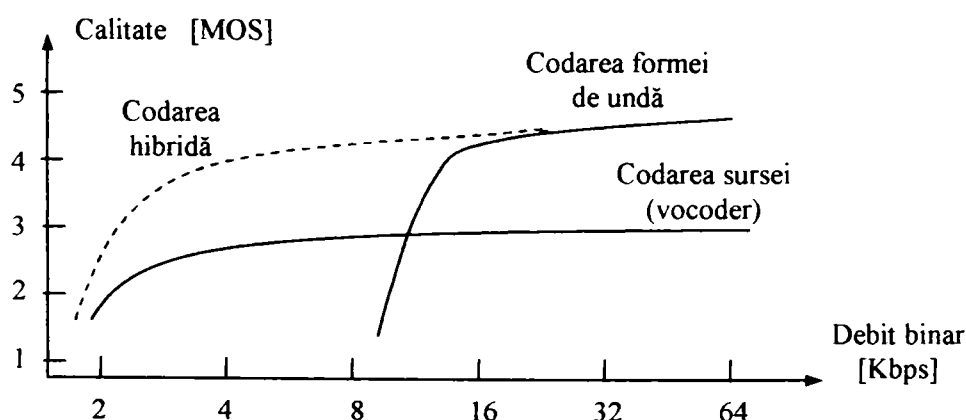
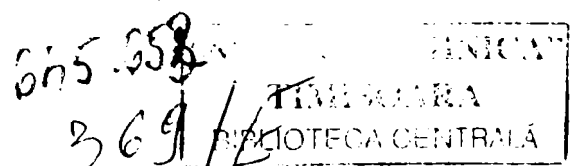


Fig. 2.1. Calitatea subiectivă funcție de debitul binar

Toate metodele de codare a semnalului vocal realizează o compresie a semnalului și din acest punct de vedere sunt cunoscute două categorii de metode de compresie: *compresia cu pierderi* și *compresia fără pierderi*. Aproape toate metodele de codare folosite în mod curent pentru codarea semnalului vocal realizează o compresie cu pierderi. Evident este de dorit ca aceste pierderi să fie cât mai mici, iar efectul lor asupra calității semnalului obținut după decodare să nu fie perceptibil, sau, în cel mai rău caz, să nu fie deranjant. În acest sens, un rol important îl are cunoașterea mecanismului percepției naturale a sunetelor de către sistemul auditiv uman și realizarea codorului după criterii perceptuale.

Sunt cunoscute însă și metode de codare a semnalului vocal care realizează o compresie fără pierderi, cu aplicabilitate la codarea semnalelor audio pentru transmisiuni de înaltă calitate [HRS01].



De remarcat că metodele de compresie fără pierderi, utilizate pentru arhivarea datelor și programelor pentru calculatoarele personale (ARJ, ZIP, RAR), oferă o rată de compresie foarte redusă pentru semnalele vocale sau semnalele audio în general, comparativ cu alte tipuri de date. Astfel, dacă pentru un fișier de tip document (.doc) se poate atinge prin arhivare o rată de compresie de 1:10, pentru un fișier de tip audio (.wav), rata de compresie nu depășește 1:1,2.

În cele ce urmează, se vor trece în revistă metodele de codare a semnalului vocal, prezentându-se principiile care stau la baza acestor metode.

## 2.1 Codarea formei de undă

### 2.1.1 Codarea scalară și codarea vectorială

#### 2.1.1.1 Codarea scalară

Codarea scalară prin metoda PCM uniformă este un proces fără memorie care cuantifică semnalul atribuind fiecărui eșantion o valoare dintr-un set de valori discrete, rezultate prin divizarea intervalului cuprins între o valoare minimă  $S_{\min}$  și o valoare maximă  $S_{\max}$ , în părți egale. Diferența între două nivele adiacente de cuantizare, adică pasul de cuantizare, sau cuanta, este o constantă care se poate exprima prin:

$$\Delta = \frac{S_{\max} - S_{\min}}{2^B - 1}, \quad (2.1)$$

unde  $B$  este numărul de biți utilizați pentru exprimarea valorii numerice a eșantionului.

Această metodă, aplicată unui semnal telefonic de bandă limitată la 3,4 kHz, eșantionat cu 8 kHz și cuantizat cu 8 biți pe eșantion, conduce la un debit binar de 64 kbps, debit considerat drept referință pentru aprecierea compresiei realizate de diferite metode de codare a semnalului.

Raportul semnal zgomot realizat de această metodă de codare poate fi apreciat prin relația:

$$SNR = 6B + K \text{ [dB]}, \quad (2.2)$$

unde  $B$  este numărul de biți pe care se face cuantizarea iar  $K$  este o constantă dependentă de pasul de cuantizare.

Codarea PCM neuniformă, așa cum sugerează și numele, folosește un pas de cuantizare neuniform. În general se folosesc pași mai mici pentru valorile de amplitudine mai mici, care apar frecvent în semnal și pași mai mari pentru valorile care apar mai rar. Modelul codării neuniforme, reprezentat în figura 2.2, constă dintr-o compresie neliniară, urmată de o cuantizare uniformă și apoi de o expandare neliniară cu o funcție inversă celei de compresie.

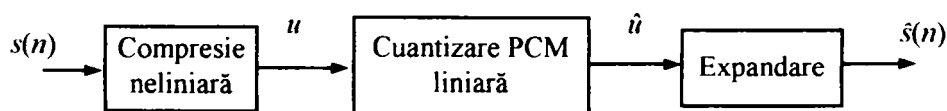


Fig. 2.2 Cuantizarea neliniară

Valorile pașilor de cuantizare pot fi optimizate exploatănd funcția de densitate de probabilitate a semnalului, dacă această funcție se cunoaște, astfel încât, pentru un semnal cu o funcție de repartiție cunoscută, cuantizat pe un număr dat de biți, raportul semnal zgomot datorat zgomotului de cuantizare să fie maxim.



O clasă aparte de cuantizoare PCM neuniforme sunt acelea care folosesc o scară logaritmică pentru pasul de cuantizare (legile de compresie  $\mu$  și  $A$ ).

Cuantizorul PCM neuniform din fig. 2.2 poate fi privit ca un cuantizor uniform care cuantizează semnalul:

$$u = F(s), \tag{2.3}$$

unde  $F$  este o funcție care realizează o compresie neliniară a semnalului de intrare  $s$ .

Funcția  $F(s)$  are, pentru legea  $A$  de compresie, utilizată în Europa, expresia analitică [KNT87]:

$$F(s) = \frac{A|s|}{1 + \ln A} \operatorname{sgn}(s), \tag{2.4}$$

Se poate arăta că pentru nivele mici ale semnalului de intrare, legea  $A$  asigură o creștere a raportului semnal - zgomot cu valoarea  $10 \log[A/(1+\ln A)]$  față de cuantizarea uniformă cu același număr de biți.

Legea  $A$  adoptată în Europa folosește o valoare  $A$  care satisface relația:

$$A/(1+\ln A) = 16, \tag{2.5}$$

de unde se obține pentru  $A$  valoarea  $A = 87,56$  [KNT87].

Pentru legea  $\mu$  de compresie, expresia analitică a funcției  $F$  este [KNT87]:

$$F(s) = s_{\max} \frac{\ln(1 + \mu |s/s_{\max}|)}{\ln(1 + \mu)} \operatorname{sgn}(s). \tag{2.6}$$

Pentru o valoare  $\mu = 255$ , legea de compresie  $\mu$  aproximează legea  $A$ .

Un cuantizor logaritmic pe 8 biți atinge aceleași performanțe din punct de vedere al raportului semnal – zgomot ca și un cuantizor uniform pe 12 biți.

Funcția  $F(s)$  se poate aproxima prin segmente liniare pe porțiuni, simplificând astfel calculele [KNT87].

Un codor scalar mai eficient este codorul PCM diferențial (DPCM). Schemele codorului și decodorului DPCM sunt prezentate în figura 2.3.a, respectiv figura 2.3.b și conțin un predictor pe termen scurt  $P(z)$ , descris de relația:

$$P(z) = \sum_{i=1}^p a_i z^{-i} \tag{2.7}$$

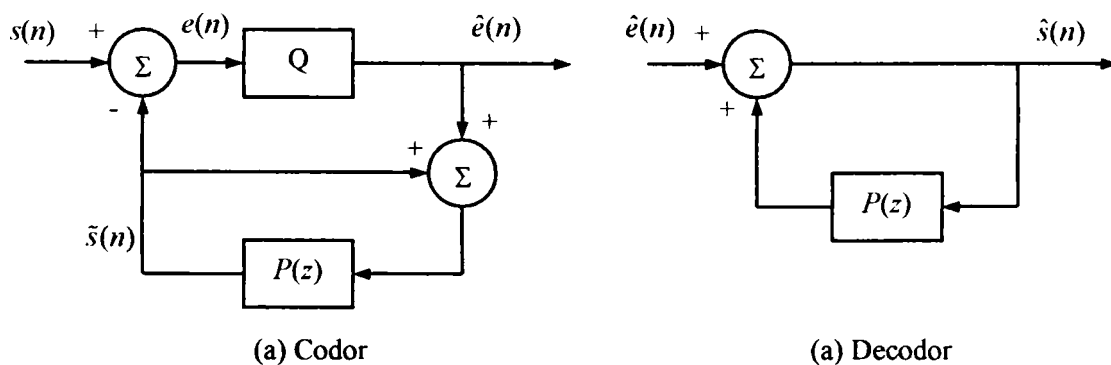


Fig. 2.3 Codorul și decodorul PCM diferențial

Bazându-se pe redundanța existentă în semnalul care urmează să fie codat, în care eșantioanele adiacente sunt puternic corelate, codorul DPCM cuantizează diferența dintre eșantionul curent și valoarea prezisă a acestuia, calculată cu ajutorul predictorului  $P(z)$ , diferență numită reziduu de predicție  $d(n)$ :

$$d(n) = s(n) - \tilde{s}(n). \quad (2.8)$$

Deoarece reziduul de predicție  $d(n)$  are o gamă dinamică mai mică decât  $s(n)$ , pentru cuantizarea cu aceeași eroare vor fi necesari mai puțini biți. Înglobând în codor un decodor, semnalul reconstituit  $\hat{s}(n)$  este cunoscut la codor și este folosit pentru a calcula reziduul de predicție și, deoarece conform schemei din fig. 2.3 se poate scrie:

$$\hat{s}(n) = \hat{d}(n) + \tilde{s}(n). \quad (2.9)$$

rezultă că eroarea de cuantizare făcută la codarea semnalului  $s(n)$  este egală cu eroarea făcută pentru a coda semnalul  $d(n)$ , fără ca precizia predictorului să conteze:

$$e(n) = s(n) - \hat{s}(n) = d(n) - \hat{d}(n). \quad (2.10)$$

Eșantionul  $\tilde{s}(n)$  este estimatul eşantionului curent  $s(n)$  și este obținut din  $p$  valori trecute ale acestuia. Cuantizorul calculează eroarea  $e(n)$  ca diferență dintre eşantionul curent și estimatul său, o cuantizează și transmite valoarea cuantizată  $\hat{e}(n)$  la recepție. Pentru a reconstitui eşantionul curent, eroarea  $e(n)$  este adunată eşantionului estimat.

CCITT a adoptat un ansamblu codor-decodor ADPCM cu debitul binar de 32 kbiți/secundă, prin standardul G.721 [CCI88], realizat după schema bloc prezentată în figura 2.3. După cum se observă din această schemă, decodorul este cuprins în codor.

Predictorul folosit în codorul G.721 este un predictor adaptiv cu 2 poli și 6 zerouri. Semnalul eroare  $e(n)$  este cuantizat pe 4 biți.

Standardul G.721 a fost modificat pentru a se adapta la debite de 24 kbps sau 48 kbps, rezultând standardul G.723 [ITU96-2].

Codorul delta, cunoscut și sub numele de modulator delta, este un caz particular al codorului DPCM, în care diferența între două eşantioane succesive este codată doar pe un bit.

### 2.1.1.2 Codarea vectorială

Compresia semnalului vocal prin codare vectorială [GER93] se obține codând simultan un set de date (eșantioane de semnal) grupate într-un vector. Pentru aceasta semnalul vocal se împarte în cadre și se consideră că eşantioanele dintr-un cadru formează un vector.

Codarea vectorială este relativ complexă față de cea scalară, dar permite obținerea unui semnal de bună calitate la un debit binar redus. Principiile codării vectoriale și aplicațiile acesteia la codarea semnalului vocal vor fi prezentate în cap. 4.

## 2.1.2 Codarea în domeniul frecvență

Aceste tipuri de codificare a formei de undă exploatează redundanța semnalului în domeniul frecvență. Posibilitatea reducerii debitului binar al semnalului constă în structura spectrului de putere pe timp scurt a semnalului vocal și în unele particularități ale urechii umane în perceperea sunetelor.

### 2.1.2.1 Codarea în subbenzi

Codorul în subbenzi divide banda de frecvențe a semnalului de intrare în mai multe subbenzi folosind un banc de filtre trece-bandă [MAT98], [TSA97]. Semnalul de la ieșirea fiecărui filtru este eşantionat și codat; rezultatul codării este multiplexat în vederea transmiterii. La recepție, semnalele sunt demultiplexate, decodate (demodulate), și prin însumare cu ajutorul unor filtre se obține un semnal care aproximează semnalul original. Metoda se pretează la fel de bine la o implementare analogică sau digitală. În cazul realizării digitale a filtrelor, acestea sunt urmate de etaje de decimare la etapa de analiză, respectiv sunt

precedate de etaje de interpolare la sinteză. Un exemplu de codare în subbenzi pentru semnalul vocal este prezentat în figura 2.4.

Reducerea debitului binar al semnalului se bazează pe faptul că fiecare subbandă poate fi codată pe un număr diferit de biți. Astfel, pentru semnalul vocal, benzile de frecvență joasă, care poartă informații referitoare la frecvența fundamentală și la structura formantică se vor coda cu un număr mai mare de biți față de benzile de frecvență înaltă.

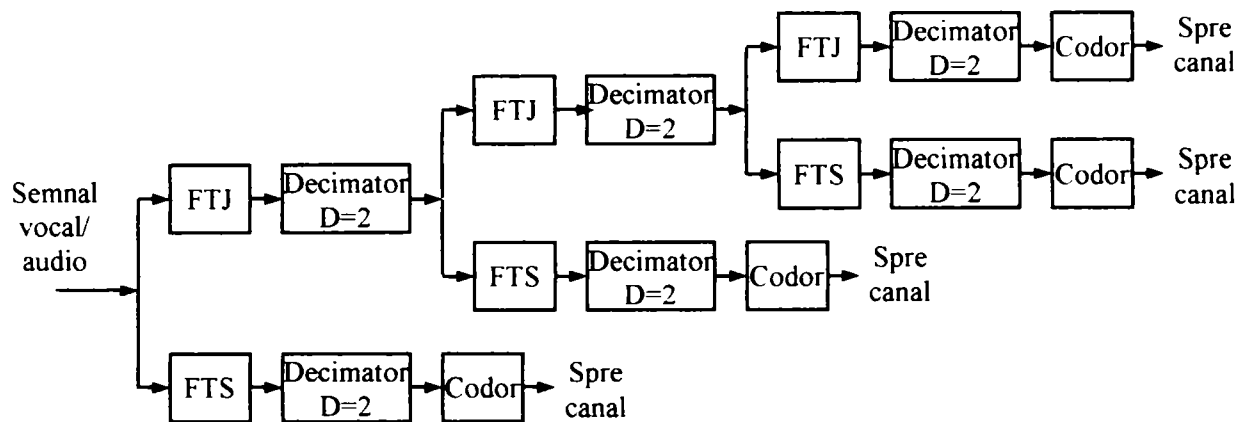


Fig. 2.4 Codarea semnalului vocal în subbenzi

În schema prezentată mai sus, se folosesc mai multe subdivizări în frecvență, care împart de fiecare dată semnalul în două subbenzi de lățime egală. În final, semnalul este împărțit în 4 subbenzi de frecvență cuprinzând în total 4 octave.

Refacerea semnalului codat în subbenzi este procesul invers codării și se face cu o schemă bloc ca cea din figura 2.5, folosind un banc de filtre de sinteză.

Proiectarea adecvată a bancului de filtre este elementul esențial al codorului în subbenzi. Banda de trecere a filtrelor poate fi egală sau inegală. Pentru codarea semnalului vocal este adecvată împărțirea spectrului de frecvență în benzi de câte o octavă. Filtrarea se poate face cel mai bine cu ajutorul filtrelor în oglindă în cuadratură (QMF) [MAT98].

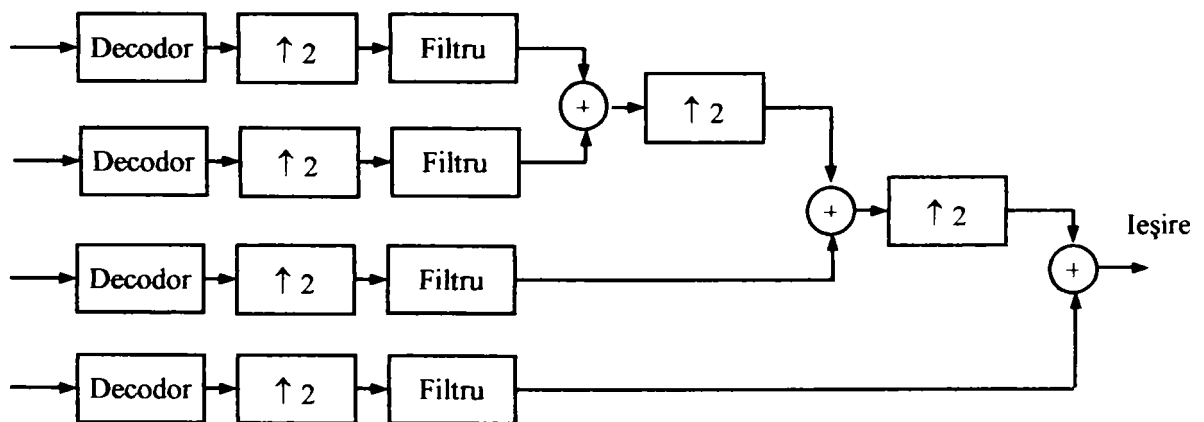


Fig. 2.5 Sinteza semnalelor codate în subbenzi

O variantă de codare în subbenzi este standardizată de CCITT prin standardul G.722 pentru un semnal audio de 7 kHz la un debit binar de 64 kbps pentru teleconferință în rețeaua ISDN. Banda audio este împărțită în două subbenzi. Codarea în cele două subbenzi se face folosind tehnica ADPCM: subbanda de frecvențe joase este codată la 48 kbps, iar subbanda de frecvențe înalte este codată la 16 kbps. Codorul G.722 include o schemă de alocare dinamică a biților și un canal auxiliar de date. Schema bloc a codorului și decodorului în subbenzi CCITT G.722 [ITU93] este reprezentată în figurile 2.6 (codorul) și 2.7 (decodorul).



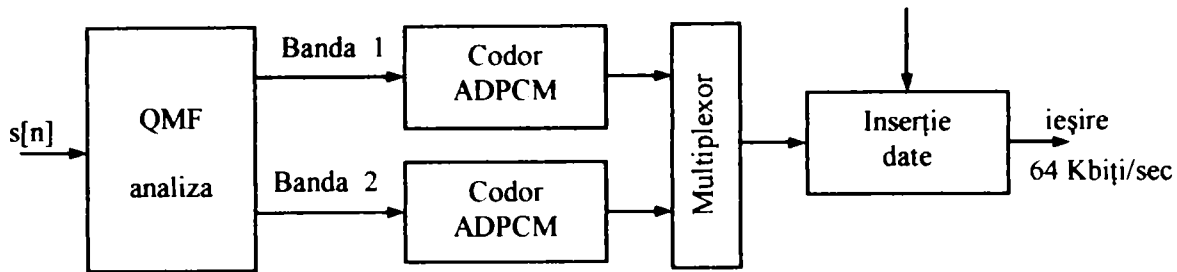


Fig. 2.6 Codorul în subbenzi G.722

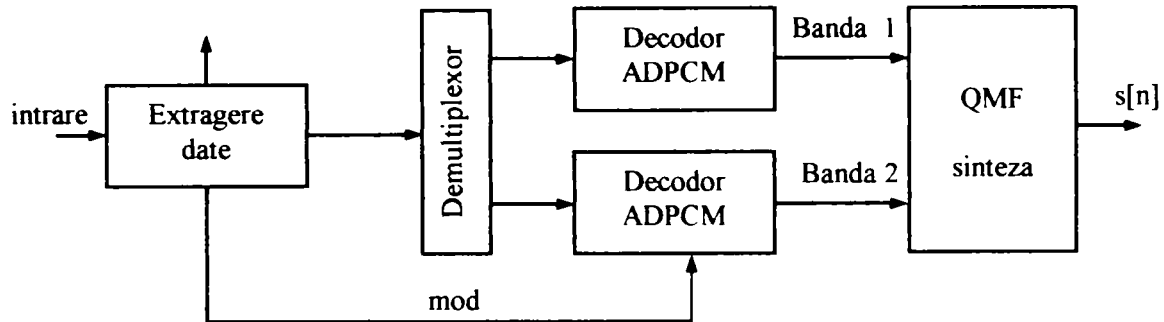


Fig. 2.7 Decodorul în subbenzi G.722

### 2.1.2.2 Codarea prin transformări ale semnalului

Acest tip de codare a semnalului vocal folosește o transformare unitară ale cărei componente sunt codate în vederea transmiterii iar la recepție după decodare se face transformarea inversă pentru refacerea semnalului original [MOR95], [SPA94].

Posibilitatea reducerii debitului binar într-o astfel de codare derivă din proprietatea transformării unitare de a produce componente aproape necorelate, care pot fi codate independent, cu un număr variabil de biți, funcție de importanța lor. În plus, se poate exploata faptul că unele din aceste componente au variații lente în timp.

Semnalul este procesat cadru cu cadru, calculându-se transformarea unitară conform relației:

$$\mathbf{S} = \mathbf{T} \mathbf{s} \quad (2.11)$$

unde  $\mathbf{s}$  este matricea coloană care conține  $N$  eșantioane ale semnalului de intrare,  $\mathbf{S}$  este matricea coloană care conține rezultatul transformării, iar  $\mathbf{T}$  este matricea transformării.

Transformarea inversă poate fi exprimată prin relația:

$$\mathbf{S} = \mathbf{T}^{-1} \mathbf{S} \quad (2.12)$$

Transformarea unitară este caracterizată de proprietatea matricii de transformare:

$$\mathbf{T}^{-1} = \mathbf{T}^H \quad (2.13)$$

unde indicele superior  $^H$  indică operatorul hermitian (complex conjugat și transpus).

Vectorii coloană ai matricii  $\mathbf{T}^H$  sunt vectorii bază ai transformării și vectorii de semnal se pot reconstitui printr-o combinație liniară a vectorilor bază. În cazul cuantizării componentelor transformării, reconstrucția nu se mai poate face exact datorită zgomotului de cuantizare. Se cunosc mai multe transformări discrete care se pot folosi pentru acest tip de codare: transformata Fourier discretă (DFT), transformata cosinus discretă (DCT) [KLO00], transformata Walsh-Hadamard (WHT), transformata Karhunen-Loeve (KLT).

Dintre acestea, transformata Kahunen-Loewe este optimală, în sensul că furnizează componente maxim decorelate pentru orice semnal de intrare, dar volumul mare de calcule necesare limitează folosirea ei.

Transformata Walsh-Hadamard este deosebit de avantajoasă din punct de vedere al simplității calculelor, având în vedere forma particulară a matricii de transformare, care constă numai din valorile +1 și -1.

### 2.1.2.3 Codarea perceptuală a semnalelor audio

Această categorie de metode de codare [PAS00], [NJY01], [NOL99], cunoscută și sub numele de compresie audio, s-a dezvoltat în legătură cu aplicațiile multimedia, în scopul de a obține o reprezentare digitală comprimată a semnalelor audio de înaltă fidelitate, în vederea stocării sau transmiterii. Banda audio acoperită este cuprinsă între 100 Hz și minimum 15 kHz, mult mai largă decât banda semnalului vocal din transmisiunile de tip telefonic. Obiectivul declarat al acestor metode de compresie este de a obține un semnal audio de foarte bună calitate, care să nu poată fi diferențiat de semnalul original.

Codarea propriu-zisă a semnalului se face folosind unii dintre algoritmi prezentați anterior (codarea în subbenzi, codarea prin transformări, etc.), iar cuantizarea rezultatelor se face pe un număr variabil de biți, având în vedere principiile psihoacustice specifice mecanismului auzului uman (codare perceptuală).

Principiile psihoacustice pe care se bazează codarea perceptuală pot fi rezumate astfel:

- pragul absolut de audibilitate al urechii umane reprezintă presiunea sonoră minimă a unei acustice care este percepută de auzul uman, în funcție de frecvență,
- banda critică, definită ca banda de frecvență la care percepția subiectivă a sunetului se modifică abrupt,
- mascarea simultană a sunetelor: un sunet puternic maschează un sunet mai slab,
- mascarea temporală: un sunet puternic, abrupt, introduce regiuni de pre-mascare și de post-mascare în timp, regiuni în care sunetele slabe nu sunt percepute,

Dintre algoritmi de codare perceptuală care s-au impus, cel mai răspândit în aplicațiile multimedia este algoritmul de codare ISO/MPEG de nivel I sau II, prezentat în schema bloc din figura 2.8.

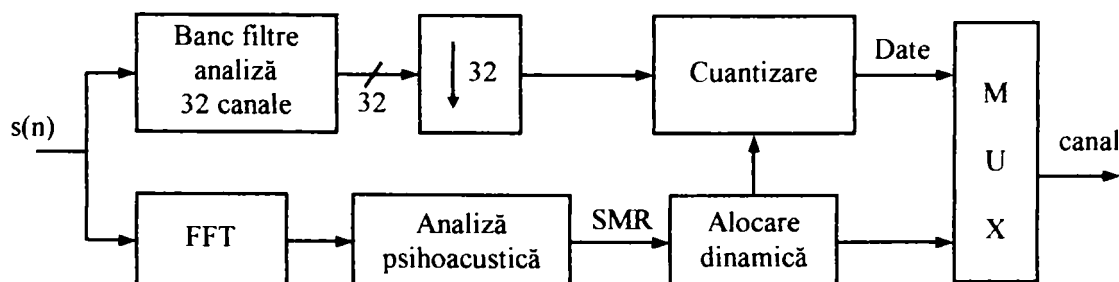


Fig. 2.8 Codorul ISO/MPEG

Semnalul de intrare este descompus în 32 de subbenzi cu un banc de filtre digitale după care semnalul de la ieșirea fiecărui filtru este decimat cu un factor de 32. Benzile acestor filtre sunt alese în concordanță cu banda critică a auzului. Semnalele de la ieșirile acestor filtre sunt cuantizate folosind un număr corespunzător de biți pentru fiecare canal, astfel încât să se asigure în ansamblu o eroare minimă.

Analiza psihoacustică se face asupra transformatei Fourier a semnalului de intrare, care se calculează în 512 puncte pentru standardul ISO/MPEG nivel I, respectiv 1024 puncte pentru nivel II, determinând criteriile de alocare dinamică a biților la cuantizare în concordanță cu principiile codării perceptuale [ISO98-1], [NJY01].

### 2.1.3 Codarea semnalului vocal utilizând analiza și sinteza sinusoidală

#### 2.1.3.1 Codarea semnalului vocal folosind transformata Fourier pe termen scurt

Din punct de vedere istoric, această metodă de codare a semnalului vocal face parte din metodele de pionierat în domeniu. Metoda are la bază presupunerea că semnalul vocal este lent variabil în timp, deci poate fi considerat cvasistaționar pe intervale scurte de timp și astfel poate fi modelat prin spectrul său pe timp scurt [SPA94].

Expresia spectrului semnalului pe timp scurt este dată de relația:

$$S(n, \Omega) = \sum_{m=-\infty}^{\infty} s(m)h(n-m)e^{-j\Omega m} = h(n) * s(n)e^{-j\Omega n} \quad (2.14)$$

unde  $\Omega$  este frecvența normalizată, în radiani, iar  $h(n)$  este fereastra de analiză, care decupează din semnal un interval temporal în care semnalul este cvasistaționar.

Refacerea semnalului se face cu transformarea inversă, dată de relația:

$$h(n-m)s(m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S(n, \Omega)e^{j\Omega m} d\Omega \quad (2.15)$$

Pentru  $n = m$  și  $h(0) = 1$ , transformarea inversă descrisă de relația (2.15) furnizează semnalul original  $s(n)$ .

Rezoluția temporală și spectrală a transformării Fourier pe termen scurt depinde de lungimea și forma ferestrei de analiză. Pentru semnalul vocal, fereastra de analiză va conține (5 – 20) ms de semnal pentru o rezoluție temporală bună, în detrimentul rezoluției spectrale.

Secvența  $h(n)$  poate fi interpretată ca răspunsul la impuls al unui filtru liniar invariant în timp, excitat cu un semnal deplasat în frecvență, interpretare care duce la modelul de calcul cu bancuri de filtre.

Pentru o variabilă discretă în domeniul frecvență,  $\Omega_k = \Delta\Omega_k$ ,  $k = \{1, 2, \dots, N-1\}$ , alegând  $\Delta\Omega$  și  $N$  astfel încât să fie acoperită banda semnalului vocal, expresiile de calcul pentru analiză, respectiv sinteză devin:

$$S(n, \Omega_k) = \sum_{m=-\infty}^{\infty} s(m)h(n-m)e^{-j\Omega_k m} = h(n) * s(n)e^{-j\Omega_k n} \quad (2.16)$$

$$\hat{s}_{STFT}(n) = \sum_{k=1}^{N-1} S(n, \Omega_k)e^{j\Omega_k n} \quad (2.17)$$

unde  $\hat{s}_{STFT}(n)$  este semnalul reconstituit. Dacă  $h(n)$ ,  $\Delta\Omega$  și  $N$  sunt corect alese, reconstrucția se poate face exact, adică:

$$\hat{s}_{STFT}(n) = s(n) \quad (2.18)$$

Un canal care face analiza și sinteza conform acestei metode este prezentat în figura 2.9, iar în figura 2.10 schema este completată cu blocurile care realizează codarea respectiv decodarea în vederea transmiterii.

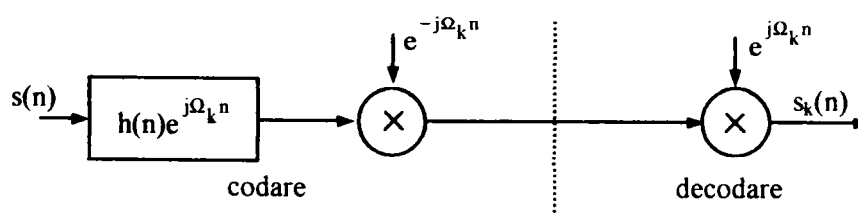


Fig. 2.9 Canal din bancul de filtre de analiză și sinteză sinusoidală

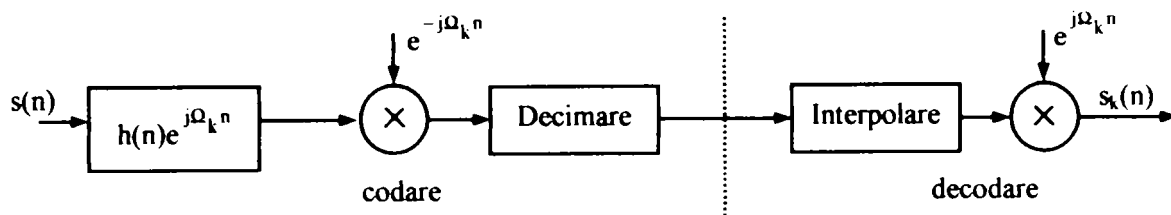


Fig. 2.10 Canal complet din bancul de filtre de analiză și sinteză sinusoidală

### 2.1.3.2 Vocoderul de fază

O metodă care are la bază tot analiza spectrală pe timp scurt realizează transformarea semnalului vocal în informație de amplitudine și fază [SPA94].

Dacă semnalul  $x(t)$  este trecut printr-un banc de filtre trece-bandă, el va putea fi aproximat de suma semnalelor de la ieșirile celor  $n$  filtre:

$$x(t) \cong \sum_n x_n(t). \quad (2.19)$$

Dacă răspunsul la impuls al filtrului trece-bandă  $n$  se exprimă prin relația:

$$g_n(t) = h(t) \cos \omega_n t, \quad (2.20)$$

se poate arăta ca semnalul la ieșirea filtrului  $n$  are forma:

$$x_n(t) = \text{Re} \left[ e^{j\omega_n t} \int_{-\infty}^{\infty} x(\tau) h(t - \tau) e^{-j\omega_n \tau} d\tau \right]. \quad (2.21)$$

Integrala din relația (2.21) reprezintă transformata Fourier pe timp scurt a semnalului de intrare evaluată la frecvența  $\omega_n$ . Dacă valoarea complexă a acestei transformări se notează  $X(\omega_n, t)$ , se poate scrie:

$$X(\omega_n, t) = |X(\omega_n, t)| \cos(\omega_n t + \rho(\omega_n, t)) \quad (2.22)$$

Prin urmare fiecare componentă a semnalului poate fi descrisă prin modulația simultană în amplitudine și fază a purtătoarei  $\cos \omega_n t$ .

Funcțiile de amplitudine pot fi limitate la o bandă de aproximativ 50 Hz fără să se piardă din detaliile semnificative. Funcțiile de fază însă nu sunt limitate și ca atare nu se transmit direct ci se folosesc în acest scop derivatele lor. Vocoderul de fază prezintă avantajul că transmite informații complete despre spectrul semnalului fără a fi necesară estimarea frecvenței fundamentale și nici caracterul sonor/nesonor al semnalului.

## 2.2 Codarea sursei de semnal

Această categorie de metode de codare a semnalului vocal se bazează pe modelul producerii semnalului vocal, fiind codate separat cele două componente ale modelului, respectiv sursa de excitație și tractul vocal. Din acest motiv, metodele din categoria codării sursei de semnal sunt specifice semnalului vocal, iar dispozitivul care realizează codarea se mai numește și **vocoder** (voice coders). Spre deosebire de metodele de codare a formei de undă, codarea sursei de semnal asigură o calitate bună la un debit binar relativ redus, dar performanțele vocoderelor se degradează pentru alte tipuri de semnale decât semnalul vocal. În această categorie de metode care codează sursa de semnal, se încadrează vocoderul cu predicție liniară, care va fi prezentat în detaliu în cap. 3, dedicat predicției liniare, în cele ce urmează fiind prezentate câteva vocodere de pionierat în domeniu.

### 2.2.1 Vocoderul canal

Vocoderul canal propus de Dudley este, din punct de vedere istoric, cea mai veche metodă de analiză și sinteză a semnalului vocal, folosind reprezentarea spectrului semnalului ca produs între caracteristica tractului vocal și spectrul sursei de excitație. Tractul vocal este reprezentat prin înfășurătoarea spectrală a semnalului vocal, extrasă cu un banc de filtre cu 16 – 20 de canale. Numărul de canale reprezintă un compromis între finețea reprezentării și complexitatea codorului.

Schema bloc a vocoderului canal [KNT87] este prezentată în figura 2.11.

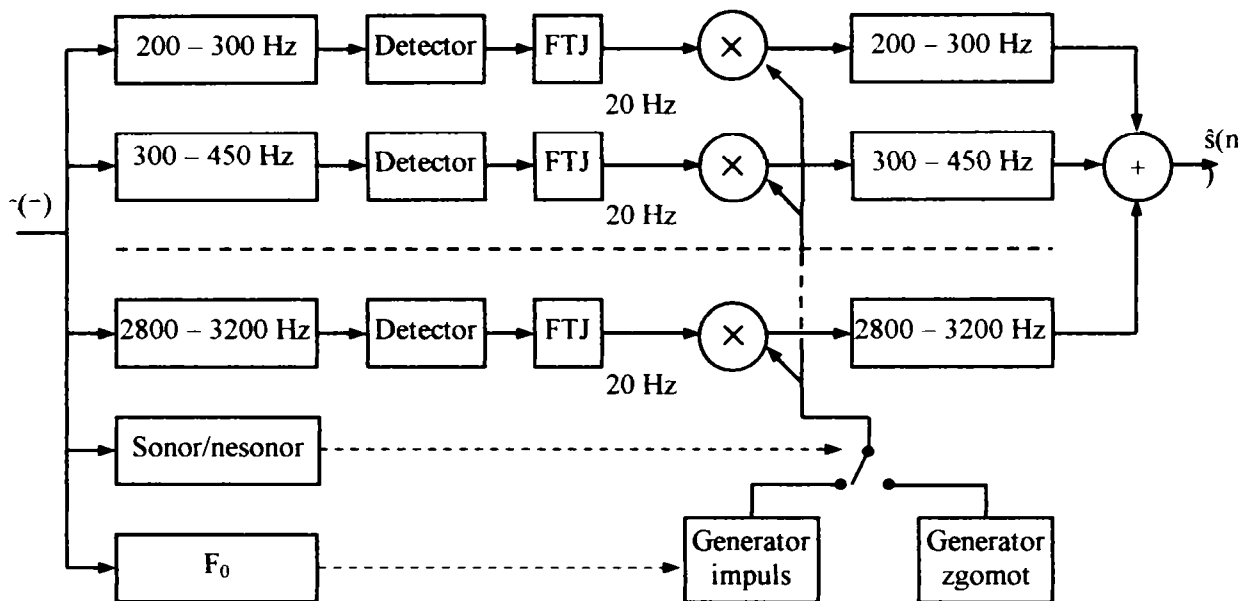


Fig. 2.11 Vocoderul canal

Semnalul de la ieșirea fiecărui filtru este redresat și aplicat unui filtru trece jos, cu banda de trecere de 20 Hz, considerată ca fiind suficientă pentru a reproduce evoluția temporală a anvelopei spectrale a semnalului vocal.

Sinteza presupune circuite de multiplicare în care se realizează produsul dintre înfășurătoarea spectrală și semnalul de excitație, care poate fi, în funcție de natura semnalului sonor/nesonor, zgomot alb sau impulsuri periodice, urmate de filtre trece bandă identice cu cele de analiză pentru fiecare canal.

### 2.2.2 Vocoderul formantic

Vocoderul formantic este în esență asemănător cu vocoderul canal, cu deosebirea că bancul de filtre are o caracteristică care se adaptează traiectoriei formaților din spectrul semnalului vocal.

Implementările vocoderului formantic [KNT87], [SPA94] folosesc două configurații de filtre: configurația serie sau configurația paralel. În figura 2.12 este reprezentată o schemă de vocoder formantic (partea de sinteză), în configurație serie. Semnalul de excitație este un tren de impulsuri periodice pentru semnalele sonore și un zgomot alb pentru semnalele nesonore. Funcția de transfer pentru semnalele sonore,  $H_v(z)$  este realizată de o cascadă de trei filtre în fig. 2.12 ( $L$  în cazul general), numai cu poli, pentru fiecare filtru fiind precizată frecvența centrală și banda de trecere, rezultând astfel caracteristica:

$$H_v(z) = \prod_{i=1}^L H_i(z) \quad (2.23)$$

Funcția de transfer pentru semnale nesonore  $H_v(z)$  este realizată cu o cascadă de filtre de ordinul doi, cu zerouri la frecvențele  $\Omega_z$  și poli la frecvențele  $\Omega_p$ .

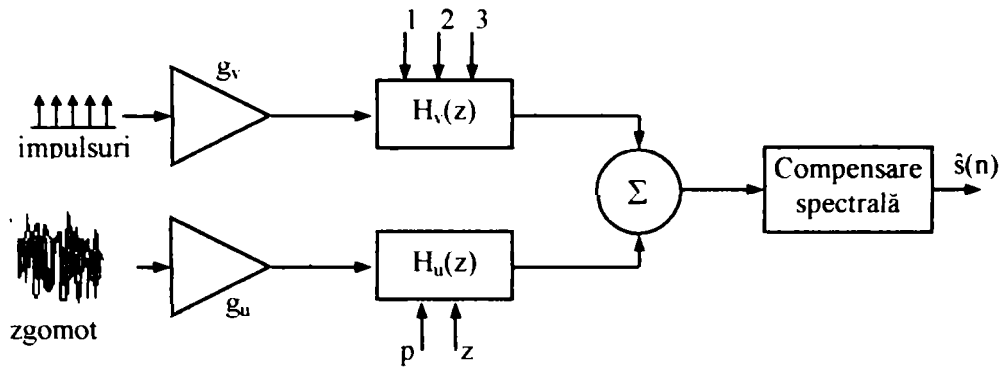


Fig. 2.12 Vocoderul formatic

Blocul de compensare spectrală este un filtru care are o caracteristică fixă ce modelează efectele excitației glotale și efectele radiației semnalului vocal la buze.

### 2.2.3 Vocoderul homomorfic

Metodele de prelucrare homomorfică a semnalelor vocale [KNT87], [SPA94] pot fi folosite pentru a separa din acest semnal informațiile referitoare la tractul vocal și la sursa de excitație, pe baza proprietăților transformatei Fourier și a proprietăților funcției logaritm, conform relațiilor:

$$\log |X(\omega)|^2 = \log [|S(\omega)|^2 |H(\omega)|^2] = \log |S(\omega)|^2 + \log |H(\omega)|^2 \quad (2.24)$$

$$F^{-1} \{ \log |X(\omega)|^2 \} = F^{-1} \{ \log |S(\omega)|^2 \} + F^{-1} \{ \log |H(\omega)|^2 \} \quad (2.25)$$

Un sistem de analiză și sinteză a semnalului vocal utilizând prelucrarea homomorfică este reprezentat în figurile 2.13a și 2.13 b.

În etapa de analiză, transformata Fourier inversă a logaritmului spectrului de amplitudine al semnalului vocal produce cepstrumul acestui semnal,  $Ce(n)$ , din care, folosind o fereastră cepstrală se pot extrage caracteristicile tractului vocal, aflate mai aproape de origine pe axa frecvenței cepstrale (*quefreny*).

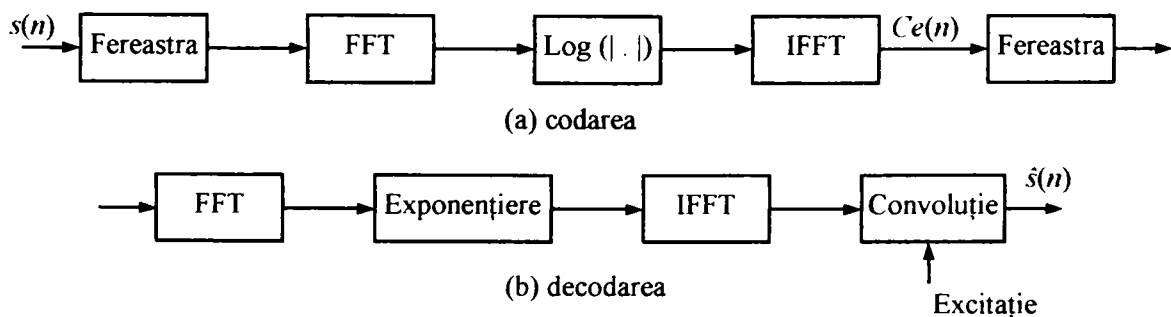


Fig. 2.13 Analiza și sinteza homomorfică

Etapa de sinteză face o prelucrare inversă pentru a reface răspunsul la impuls al tractului vocal. Prin convoluția acestui răspuns cu semnalul de excitație se produce semnalul vocal.



## 2.2.4 Codorul LPC

Modelarea mecanismului de producere a semnalului vocal prin predicție liniară a devenit cea mai utilizată tehnică de codare a semnalului vocal în ultima perioadă, stând la baza unui mare număr de standarde, așa cum rezultă și din tabelul 1.2. Predicția liniară se bazează pe modelul liniar al producerii semnalului vocal, ilustrat în fig. 1.1 și §. 1.2. Modelul LPC de bază consideră semnalul vocal ca fiind rezultat la ieșirea unui sistem liniar invariant în timp descris de o relație de forma (1.10), la a cărei intrare se aplică un semnal de excitație corespunzător. Acest model este complet determinat dacă se cunosc parametrii sistemului liniar și semnalul de excitație. Modul în care se determină acești parametri clasifică codoarele LPC în două clase, care vor fi prezentate în continuare:

- codarea LPC în buclă deschisă,
- codarea LPC în buclă închisă.

### 2.2.4.1 Codorul LPC în buclă deschisă

Codorul LPC în buclă deschisă se bazează pe modelul LPC care va fi prezentat în detaliu în cap. 3, pentru care excitația ideală este eroarea reziduală de predicție. Modul în care se obține semnalul de excitație la recepție conduce la gruparea codoarelor LPC în buclă deschisă în mai multe categorii, dintre care cele mai importante sunt prezentate în continuare.

- **Modelul cu excitație zgomot - impuls**

Acest model evită necesitatea codării și transmiterii erorii reziduale, înlocuind acest semnal de excitație cu unul generat local, la recepție, compus dintr-o secvență de impulsuri periodice pentru sunetele sonore, respectiv un zgomot alb pentru sunetele nesonore.

Parametrii pentru semnalul de excitație al acestui model sunt câștigul, parametrul binar sonor/nesonor și frecvența fundamentală pentru segmentele de vorbire sonore. Câștigul modelului este determinat astfel ca energia segmentului vocal sintetizat să fie egală cu energia segmentului analizat. Determinarea segmentelor de vorbire sonore se poate face prin măsurarea energiei și a numărului mediu de treceri prin zero ținând cont că segmentele de vorbire nesonore sunt asociate cu o energie relativ mică și cu un număr mare de treceri prin zero, iar segmentele sonore au, în general o energie mai mare și un număr mai redus de treceri prin zero. Detecția sonor-nesonor poate fi realizată chiar prin algoritmul de determinare a frecvenței fundamentale pentru segmentele sonore.

Un codor care folosește acest model a devenit standardul federal FS-1015 și este cunoscut ca algoritmul LPC-10, asigurând un debit binar de 2,4 kbps. Schema bloc utilizată [SPA94], CCT86], pentru partea de codare (analiză) este prezentată în figura 2.14 iar pentru partea de decodare (sinteză) în figura 2.15.

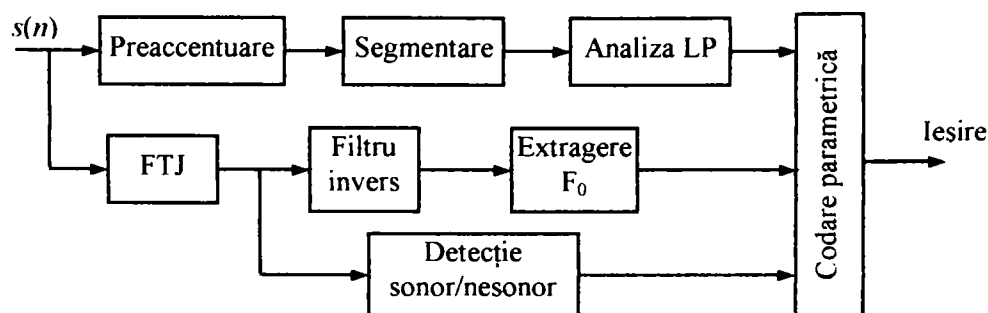


Fig. 2.14 Codorul FS – 1015

Algoritmul LPC-10 constă dintr-un predictor liniar de ordinul 10 pentru a estima tractul vocal. Semnalul vocal este preaccentuat folosind un filtru trece sus de ordinul întâi. Segmentarea este făcută în concordanță cu natura semnalului: pentru semnale sonore, analiza

se face sincron cu frecvența fundamentală, pe durate de timp multipli întregi ai perioadei fundamentale, iar pentru segmentele nesonore, analiza se face pe segmente cu durata fixă de 22,5 msec. Pentru determinarea caracterului sonor/nesonor și a frecvenței fundamentale se folosește numai porțiunea de sub 800 Hz din spectrul semnalului, obținută prin filtrare trecejos. Pentru a îmbunătăți calitatea estimării frecvenței fundamentale, se folosește un algoritm de filtrare inversă de ordinul doi.

Determinarea efectivă a frecvenței fundamentale se face folosind metoda funcției diferență de amplitudine medie (AMDF) aplicată semnalului filtrat. Decizia sonor/nesonor se ia pe baza valorii energiei semnalului și a numărului mediu de treceri prin zero, corelate cu raportul dintre maximum și minimum funcției AMDF. Parametrii rezultați prin analiză sunt codati pe 7 biți. Pentru segmentele sonore se codează toți cei 10 coeficienți de predicție, în timp ce pentru segmentele nesonore se codează doar primii 4 coeficienți.

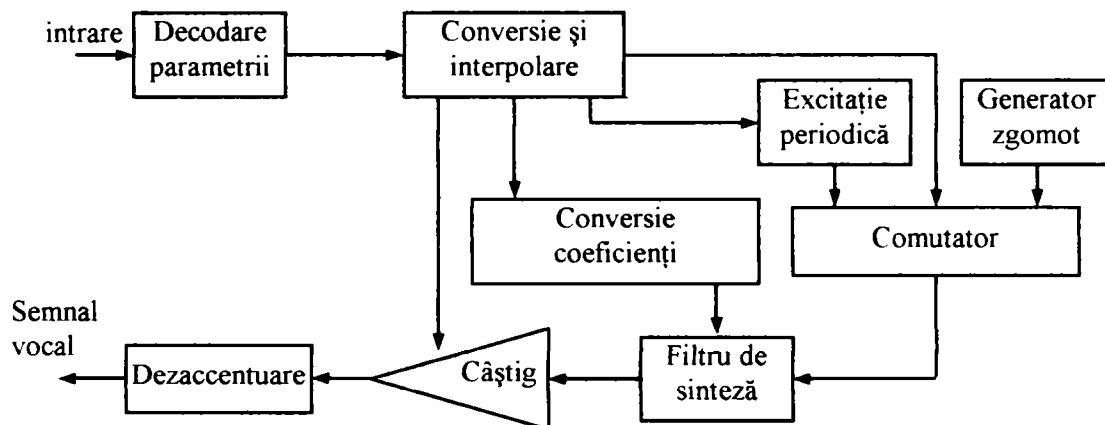


Fig. 2.15 Decodul FS – 1015

Partea de sinteză a algoritmului LPC10 respectă modelul liniar al producerii semnalului vocal, folosind un filtru de sinteză numai cu poli, excitat cu semnal periodic pentru segmentele sonore și cu zgomot pentru segmentele nesonore.

#### • Modelul cu excitație mixtă

Modelul cu excitație mixtă aduce o îmbunătățire în ceea ce privește erorile datorate excitației cu impulsuri de frecvență fixă pentru semnalele sonore ale modelului prezentat anterior, mai ales în cazul tranzițiilor sonor-nesonor [MCB95]. O sursă de excitație mixtă este prezentată în figura 2.16.

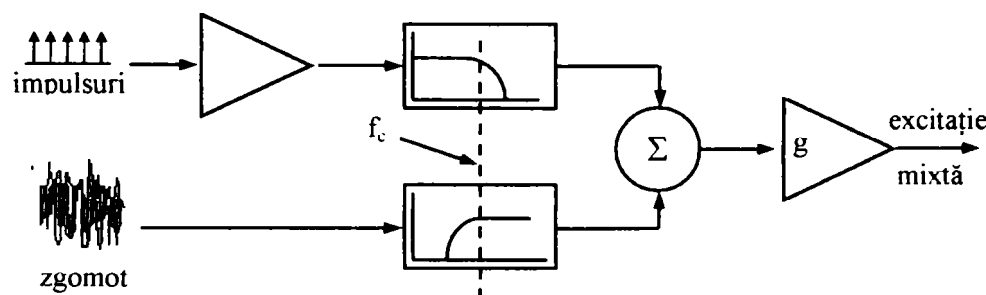


Fig. 2.16 Sursă de excitație mixtă

Trenul de impulsuri periodice excită filtrul de sinteză în domeniul frecvențelor joase iar zgomotul excită filtrul în domeniul frecvențelor înalte. Frecvențele de tăiere ale celor două filtre se modifică simultan și sunt alese cu un algoritm care determină regiunea din spectrul semnalului care este asociată cu o periodicitate. Un model de excitație mai elaborat poate conține filtre care se modifică independent și poate atribui impulsurilor periodice un jitter de poziție pe durata segmentelor de semnal vocal slab periodice.



- **Modelul excitat rezidual**

Deoarece filtrul de predicție liniară lucrează ca un decorelator pe termen scurt, eroarea reziduală de predicție va avea un spectru relativ plat. Eroarea de predicție este semnalul de excitație ideal pentru filtrul de sinteză de tipul numai poli și conține toate informațiile despre semnalul vocal care nu au fost „capturate” de analiza prin predicție liniară (de exemplu informații despre frecvența fundamentală, zerourile datorate tractului nazal).

Codarea semnalului eroare de predicție se poate face folosind un debit binar redus, având în vedere banda de frecvență relativ limitată a acestui semnal (max. 800 Hz) și variația lentă în timp a acestui semnal.

#### 2.2.4.2 Codorul LPC în buclă închisă

Un codor LPC în buclă închisă poate fi caracterizat prin:

- un model de producere a semnalului vocal, care depinde de setul de parametri  $\theta$ :

$$\hat{s}(n) = f(\theta), \quad (2.26)$$

- un număr de  $K$  valori posibile pentru setul de parametri  $\theta: \theta_1, \theta_2, \theta_3, \dots, \theta_K$ ,
- o măsură a erorii  $|E_k|^2$  prin care se compară semnalul original  $s(n)$  cu semnalul refăcut  $\hat{s}(n)$ .

Codorul în buclă închisă găsește setul optim de parametri  $\theta$  prin sintetizarea tuturor celor  $K$  semnale vocale diferite  $\hat{s}_k(n)$ , corespunzătoare tuturor valorilor posibile pentru parametrii  $\theta$  și calcularea erorii  $|E_k|^2$  corespunzătoare, reținând și transmițând la decodor indexul  $k$  care minimizează această măsură a erorii. Deoarece se calculează explicit fiecare valoare  $\hat{s}_k(n)$  și se compară cu semnalul de intrare  $s(n)$ , se justifică denumirea de analiză prin sinteză, sub care este cunoscut acest codor, spre deosebire de codarea în buclă deschisă, unde parametrii codorului se determină pe baza unor relații analitice, fără a calcula explicit  $\hat{s}_k(n)$ .

Această metodă de codare a semnalului vocal în buclă închisă conduce la o complexitate de calcul foarte ridicată, datorată necesității estimării tuturor valorilor posibile  $\hat{s}_k(n)$  și respectiv  $|E_k|^2$ . Metodele de codare LPC în buclă închisă folosite în practică adoptă un compromis pentru a reduce complexitatea, calculând parametrii filtrului liniar al modelului în buclă deschisă, folosind o metodă analitică adecvată, iar semnalul de excitație optim se determină în buclă închisă [RTB90].

Codorul LPC în buclă închisă poate utiliza, în general, trei modele diferite pentru semnalul de excitație:

- excitația multiimpuls,
- excitația cu impulsuri periodice,
- excitația cu vectori sau cod,

care vor fi prezentate pe scurt în continuare.

- **Codorul LPC cu excitație multiimpuls (Multi-pulse Excited LP - MPLP)**

Algoritmul MPLP presupune folosirea unui semnal de excitație care constă într-un tren de impulsuri neuniform distanțate în timp, pentru care se calculează în cursul procesului de analiză atât amplitudinea cât și localizarea în timp, astfel încât eroarea medie pătratică între semnalul vocal de intrare și semnalul sintetizat local să fie minimă. Uzual se folosesc secvențe formate din 4 – 6 impulsuri pentru un cadru de semnal cu durata de până la 5 ms.

Schema bloc a unui sistem de analiză conform algoritmului MPLP este prezentată în figura 2.17.

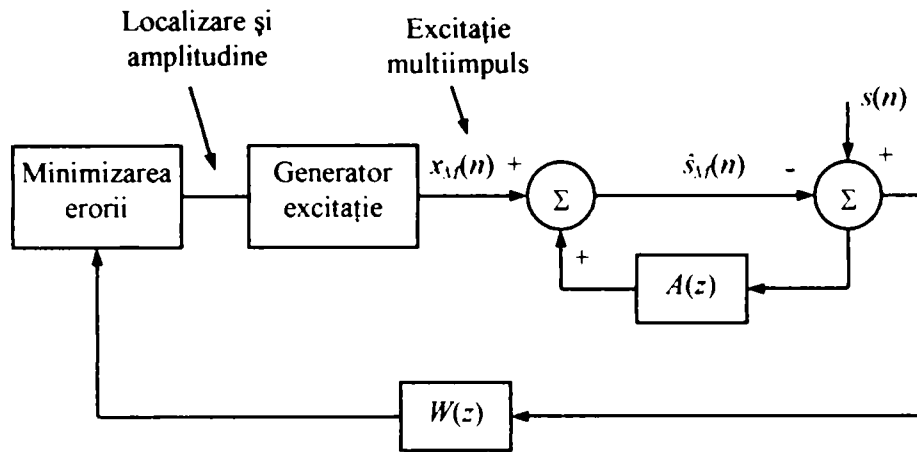


Fig. 2.17 Analiza prin algoritmul MPLP

Eroarea ponderată care urmează să fie minimizată este dată de relația:

$$e_M(n) = (s(n) - \hat{s}_M(n)) * w(n) \quad (2.27)$$

unde indicele  $M$  se referă la multiimpuls iar  $w(n)$  este răspunsul la impuls corespunzător funcției de transfer  $W(z)$ , care realizează o ponderare a erorii în conformitate cu unele criterii perceptuale.

Pentru fiecare impuls din secvența de excitație se determină poziția în timp și amplitudinea, prin minimizarea unei erori de forma (2.27), particularizată pentru impulsul respectiv. Impulsurile care formează o secvență de excitație se determină pe rând, la fiecare iterație determinându-se câte un impuls și după fiecare iterație eroarea este actualizată prin scăderea impulsului astfel determinat din eroarea de predicție.

Un codor pentru semnalul vocal funcționând după algoritmul MPLP a fost propus de British Telecom pentru un serviciu telefonic aeronautic sub numele Skyphone. Predictorul de termen scurt este de ordinul 10 și coeficienții săi sunt determinați prin metoda autocorelației, prin algoritmul Levinson-Durbin și actualizați la fiecare 20 msec.

- **Codorul LPC excitat cu impulsuri periodice (Regular Pulse Excitation - RPE)**

Codoarele RPE [STA98], [GSM97], [GSM98] folosesc ca secvență de excitație tot un impuls multiplu, dar, spre deosebire de codorul LPC excitat multiimpuls, codorul RPE folosește o succesiune de impulsuri uniform distanțate în timp. Prin urmare, secvența de excitație, aleasă dintr-un set de secvențe prestabilite, este determinată prin specificarea locației  $k$  a primului impuls și a distanței dintre două impulsuri succesive, în cazul în care această distanță nu are o valoare fixă. Numărul de impulsuri pentru un cadru de semnal cu durata de 5 ms este de ordinul 10 – 15. Amplitudinea impulsurilor se determină prin rezolvarea unui set de ecuații liniare obținute prin anularea derivatei unei erori similare (2.27), dar este important de remarcat avantajul că distanța în timp între impulsuri nu mai trebuie determinată dacă se stabilește un număr fix de impulsuri de excitație uniform distanțate în timp pentru fiecare cadru de semnal vocal.

Algoritmul RPE a fost ales pentru implementarea unui codor pentru semnalul vocal utilizat în sistemul de telefonie mobilă european GSM. Schema bloc a sistemului de codare și decodare a semnalului vocal conform standardului european GSM RPE-LPT este prezentată simplificat în figura 2.18 [GSM97], [STA98]. Partea de codare, prezentată în figura 2.18a, primește semnalul vocal eșantionat la frecvența de 8 KHz, pentru o bandă de (0,3 – 3,4) KHz și cuantizat liniar pe 13 biți. Prelucrarea se face pe cadre de semnal vocal cu durata de 20 ms, corespunzător cu 160 eșantioane, care sunt segmentate în subcadre de câte 5 ms (40 eșantioane). Etajul de preprocesare realizează o preaccentuare și o compensare a componentei continue.

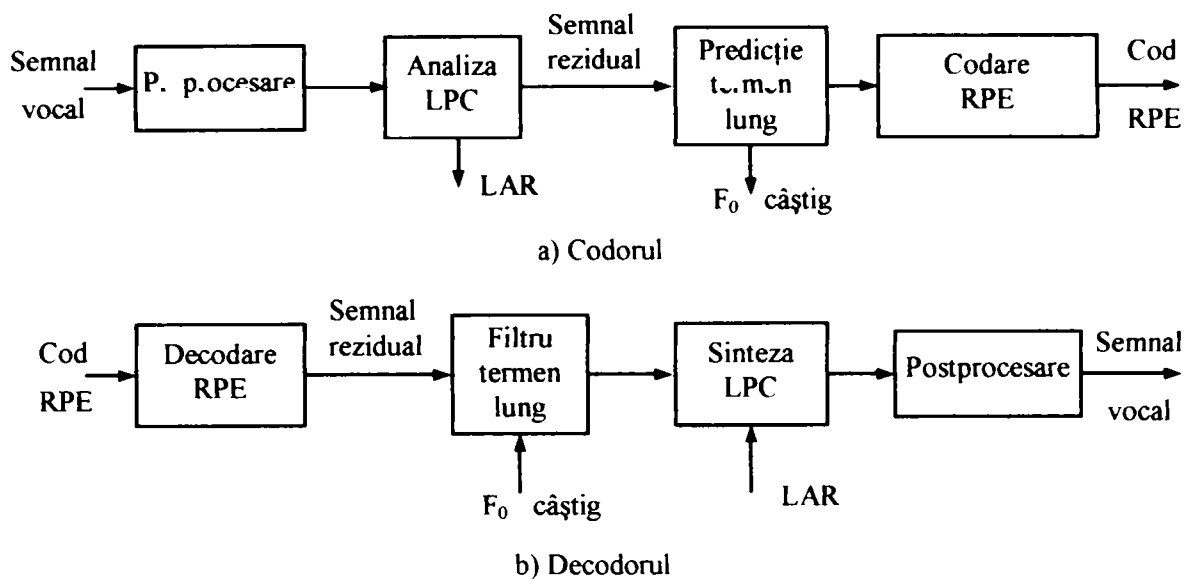


Fig. 2.18 Schema bloc a sistemului RPE – LPT utilizat în standardul GSM

Predictorul pe termen scurt care urmează etajului de preprocesare determină un număr de 8 coeficienți de predicție la fiecare 20 ms, coeficienți care, pentru o cuantizare mai eficientă, sunt codati ca parametri LAR. Pentru cei 8 coeficienți LAR, în ordinea LAR(1)...LAR(8), se alocă un număr variabil de biți pentru codare și anume 6/6/5/5/4/4/3/3.

Blocul de predicție pe termen lung extrage din semnalul rezidual informațiile referitoare la câștig și frecvența fundamentală, informații care sunt codate și transmise prin canalul de comunicație la decodor. Pentru câștig se alocă 2 biți iar pentru ordinul predictorului pe termen lung 7 biți.

Pentru semnalul de excitație corespunzător unui subcadru de semnal vocal de 40 de eșantioane, se folosesc 13 impulsuri uniform distanțate în timp, dispuse sub forma unei „grile”, în care, după fiecare impuls de excitație urmează 3 impulsuri nule. Această grilă poate avea prin urmare 4 poziții distincte în subcadru, poziții care se pot coda pe 2 biți. Pentru a coda amplitudinea fiecărui impuls din grilă se determină impulsul cu amplitudinea maximă și această valoare se codează logaritmice folosind 6 biți, iar pentru celelalte impulsuri se codează pe 3 biți valoarea normalată la amplitudinea maximă.

#### • Codorul LPC excitat cu coduri (Code Excited Linear Prediction - CELP)

Cele două variante de codoare folosind excitația multiimpuls determinată prin metoda analizei prin sinteză, prezentate anterior, realizează un semnal vocal de foarte bună calitate la un debit binar considerat mediu. Păstrarea calității semnalului sintetizat la rate binare mai mici reclamă folosirea unei secvențe de excitație codată mai eficient. Această cerință o îndeplinește codorul CELP, prezentat în figura 2.19, care folosește pentru excitație un cod sau un vector, memorat într-un dicționar de coduri.

Dicționarul de coduri folosit în schema prezentată conține un număr corespunzător de secvențe de excitație, memorate sub forma unor vectori, fiecare vector conținând un număr de eșantioane cu care acoperă o durată de semnal vocal de un subcadru. Vectorul de excitație, înmulțit cu un coeficient de câștig, este trecut prin filtrele de predicție pe termen lung, respectiv pe termen scurt. Semnalul sintetizat prin filtrare este comparat cu semnalul vocal de intrare iar blocul de minimizare a erorii va alege ca optim acel vector de cod pentru care semnalul sintetizat este cel mai apropiat de semnalul de intrare. Atât semnalul sintetizat cât și cel de intrare sunt ponderate perceptual cu funcția  $W(z)$  înainte de a se calcula eroarea.

Procesul de căutare în codorul CELP va fi prezentat în detaliu în cap. 3.

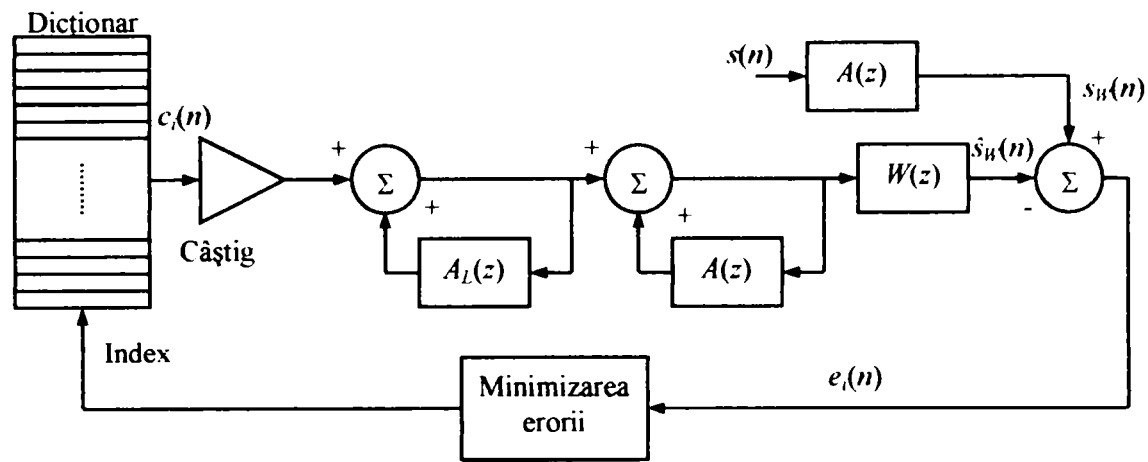


Fig. 2.19 Codorul CELP

Dezavantajul major al codorului CELP este puterea mare de calcul necesară pentru codare și memoria relativ mare pentru dicționarul de coduri. Acest fapt a dus la dezvoltarea unor algoritmi performanți de căutare și a unor dicționare de coduri structurate, în scopul reducerii complexității operațiilor de căutare.

Sunt cunoscute variante de codor CELP, denumite VSELP (Vector Sum Excited CELP), care folosesc mai multe dicționare, semnalul de excitație fiind obținut ca o sumă de vectori din aceste dicționare

Codorul CELP, prin performanțele atinse, s-a impus ca standard pentru anumite tipuri de comunicații, dintre care este cunoscut codorul CELP care constituie standardul federal FS – 1016, cu debit de 4,8 Kbps, adoptat de Departamentul Apărării SUA pentru un sistem de comunicații telefonice cu aplicații militare.

- **Codorul CELP cu întârziere redusă ( Low Delay CELP)**

O caracteristică a codoarelor CELP prezentate anterior este aceea că există o întârziere apreciabilă între semnalul de intrare în codor și semnalul de ieșire din decodor, datorată modului în care se face codarea semnalului, pe cadre, respectiv subcadre, la care se adaugă timpul determinat de algoritmul de căutare în dicționarele de coduri. Această întârziere poate atinge valoarea corespunzătoare a două până la patru cadre de analiză. O valoare tipică pentru întârzierea unui codor CELP cu un cadru de analiză de 20 ms este aproximativ 60 ms.

ITU a luat în considerare posibilitatea standardizării unor algoritmi de codare a semnalului vocal cu întârziere redusă, fiind ales un codor CELP de întârziere redusă (Low-Delay CELP) pentru standardul G.728 [ITU92], [CCL92] [WDG94].

În figura 2.20 este prezentat codorul Low-Delay CELP, iar în figura 2.21 decodorul corespunzător.

Reducerea întârzierii se obține pe două căi.

Pe de o parte, se folosesc pentru reducerea întârzierii cadre de analiză scurte, cu durata de 2,5 ms, divizate în subcadre de 0,625 ms, care corespund unor vectori de excitație scurți, de numai 5 eșantioane.

Pe de altă parte, se folosește la recepție un predictor adaptiv înapoi, care determină parametrii LPC pe baza valorii cuantizate a eșantioanele precedente ale semnalului vocal, eșantioane deja disponibile la recepție.

Codorul Low-Delay CELP nu folosește predicția pe termen lung, acest lucru fiind compensat de utilizarea unui ordin ridicat pentru predictorul pe termen scurt ( $p = 50$ ), pentru a păstra informații despre frecvența fundamentală (pitch). Pentru o îmbunătățire a spectrului semnalului în zona formanților și a frecvenței fundamentale, se folosește o postfiltrare.

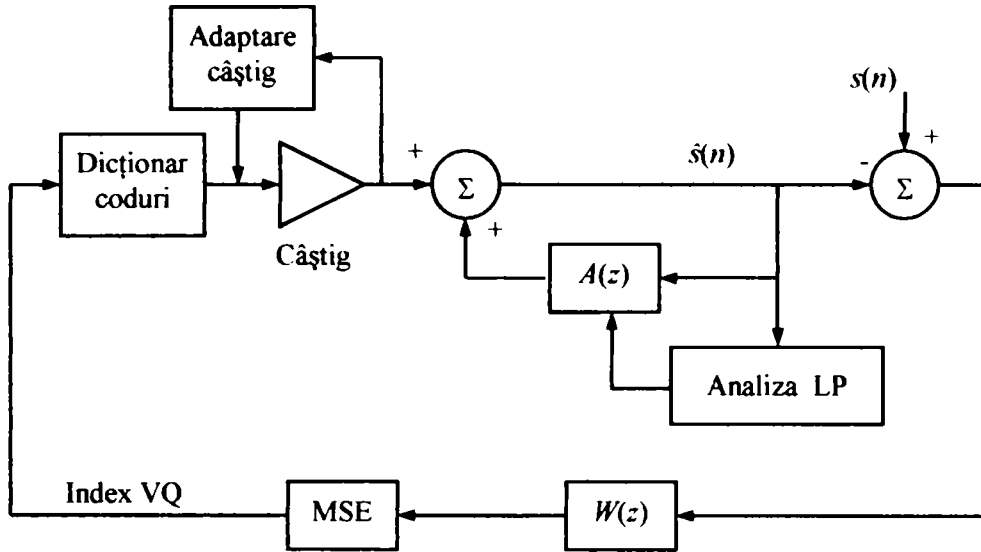


Fig. 2.20 Codorul Low – Delay CELP

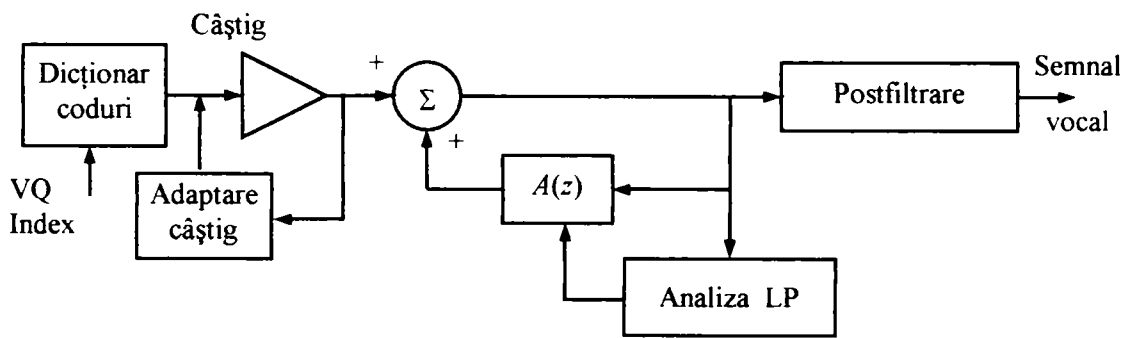


Fig. 2.21 Decodorul Low – Delay CELP

Codorul Low-delay CELP prezentat realizează o întârziere de maximum 2 ms, ceea ce permite și îndeplinirea cerințelor CCITT referitoare la transmiterea informațiilor de semnalizare DTMF și a semnalelor provenite de la modemuri în banda telefonică de bază.

# Capitolul 3

## Predicția liniară a semnalului vocal

### 3.1 Predicția liniară

Predicția liniară este metoda predominant folosită în codarea semnalului vocal la debit binar redus, datorită simplității calculelor și posibilității de reprezentare cu bună precizie a parametrilor caracteristici ai semnalului vocal. Predicția liniară permite atât extragerea redundanței din semnalul vocal, cât și construcția unui model pentru mecanismul de producere a vorbirii naturale.

#### 3.1.1 Ecuațiile de predicție liniară

Conform modelului de predicție liniară, se poate considera că pe durata unui cadru de semnal vocal de  $N$  eșantioane, eșantionul curent  $s(n)$  al semnalului vocal este obținut la ieșirea unui sistem ca urmare a semnalului de excitație  $u(n)$  aplicat la intrare [JMK75], [MHG76]:

$$s(n) = \sum_{k=1}^p a_k s(n-k) + G \sum_{l=0}^q b_l u(n-l), \quad b_0 = 1, \quad (3.1)$$

în care  $\{a_k\}$ ,  $\{b_l\}$  sunt parametrii sistemului și  $G$  are semnificația unui factor de câștig.

Conform relației (3.1), eșantionul curent este „prezis” ca o combinație liniară a unui număr de  $p$  eșantioane de ieșire precedente precum și a eșantionului de intrare curent și a  $q$  eșantioane de intrare precedente.

Transformata  $z$  care caracterizează sistemul descris de (3.1) este:

$$H(z) = \frac{S(z)}{U(z)} = G \frac{1 + \sum_{l=1}^q b_l z^{-l}}{1 - \sum_{k=1}^p a_k z^{-k}} \quad (3.2)$$

și reprezintă modelul general al producerii semnalului vocal, cu poli și zerouri, denumit model ARMA (autoregressive moving average – autoregresiv cu medie alunecătoare) [DPH93], [STA98].

Modelul descris de relația (3.2) prezintă două cazuri particulare de interes:

- dacă  $a_k = 0$  pentru  $k = 1, 2, \dots, p$ ,  $H(z)$  devine un model numai cu zerouri, numit model MA (moving average – model cu media alunecătoare),
- dacă  $b_l = 0$  pentru  $l = 1, 2, \dots, q$ ,  $H(z)$  devine un model numai cu poli sau model autoregresiv AR.

Modelul cu poli (AR) este preferat în modelarea semnalului vocal față de modelul ARMA datorită simplității calculului și datorită corespondenței cu modelul cu tuburi acustice al producerii vorbirii naturale [DPH93], [ATH71].



Conform modelului numai cu poli, eșantionul curent se poate estima ca o combinație liniară a unor eșantioane precedente, la care se adaugă un termen care reprezintă excitația și, în acest caz, rel. (3.1) devine:

$$s(n) = \sum_{k=1}^p a_k s(n-k) + Gu(n). \quad (3.3)$$

Deoarece pentru un sistem AR care produce un semnal de forma (3.3) excitația nu este accesibilă, estimarea parametrilor modelului se poate face numai pe baza observării semnalului de ieșire și, pentru că acest semnal este produs prin recurența (3.3), se poate defini un semnal  $\hat{s}(n)$  care estimează semnalul original  $s(n)$  în forma:

$$\hat{s}(n) = \sum_{k=1}^p a_k s(n-k). \quad (3.4)$$

Eroarea de predicție care se face aproximând  $s(n)$  cu  $\hat{s}(n)$  este:

$$r(n) = s(n) - \hat{s}(n) = s(n) - \sum_{k=1}^p a_k s(n-k). \quad (3.5)$$

Eroarea (3.5) se mai numește și semnal rezidual, și comparând (3.3), (3.4) și (3.5) se constată că eroarea de predicție coincide cu semnalul de excitație înmulțit cu factorul de câștig.

Aplicând transformarea  $z$  în rel. (3.5) rezultă:

$$R(z) = A(z)S(z), \quad (3.6)$$

unde  $R(z)$  este transformata  $z$  a semnalului rezidual, iar

$$A(z) = 1 - \sum_{k=1}^p a_k z^{-k} \quad (3.7)$$

reprezintă filtrul de analiză predictivă.

Filtrul de sinteză cu poli  $H(z)$  de forma:

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 - \sum_{k=1}^p a_k z^{-k}} \quad (3.8)$$

modelează înfășurătoarea spectrului de putere pe termen scurt a semnalului vocal.

Sistemul numai cu poli este un bun estimant al sistemului cu poli și zerouri descris de (3.2). Astfel, un sistem cauzal descris de rel. (3.2) se poate descompune în forma [STA98]:

$$H(z) = G' H_{min}(z) H_{ap}(z), \quad (3.9)$$

unde  $G'$  este un factor de câștig,  $H_{min}(z)$  este funcția de transfer a unui filtru trece – tot de fază minimă, iar  $H_{ap}(z)$  este funcția de transfer a unui filtru numai cu poli de forma (3.7) [OSF89], [DPH93].

Filtrul trece – tot  $H_{min}(z)$  contribuie numai la faza semnalului de ieșire și poate fi omis, având în vedere insensibilitatea relativă a auzului uman la faza semnalului.

Prin urmare, se va reține pentru sistemul de producere a semnalului vocal modelarea numai cu poli, exprimată prin rel. (3.8), în care ordinul de predicție  $p$  are o valoare întregă relativ mică.

Predicția se poate face *înainte* sau *înapoi* [STA98]. Predicția înainte se bazează pe estimarea eșantionului curent folosind un anumit număr de eșantioane precedente ale

semnalului vocal, conform relației (3.3) și astfel eșantionul curent poate fi prezis înainte de a fi cunoscut. Predicția înapoi folosește pentru estimarea eșantionului curent al semnalului eșantioane viitoare ale semnalului vocal astfel că valoarea eșantionului curent este de fapt doar verificată, după ce se cunosc eșantioanele care urmează și, din acest punct de vedere, predicția înapoi nu prezintă interes în prelucrarea semnalului vocal.

În literatură se folosește însă termenul de predicție înapoi și în alt sens, înțelegându-se prin predicție înapoi estimarea valorii eșantionului curent pe baza unor valori trecute ale eșantioanelor cuantizate ale semnalului vocal. Avantajul predicției înapoi, în acest sens al termenului, este acela că parametrii de predicție nu trebuie să mai fie cuantizați și transmiși la decodor, deoarece decodorul dispune de valorile precedente cuantizate (refăcute) și poate calcula local acești coeficienți [VZZ97].

Schema bloc din figura 3.1 ilustrează operațiile de analiză și sinteză a semnalului prin predicție liniară, prin implementarea directă a rel. (3.3) și (3.5).

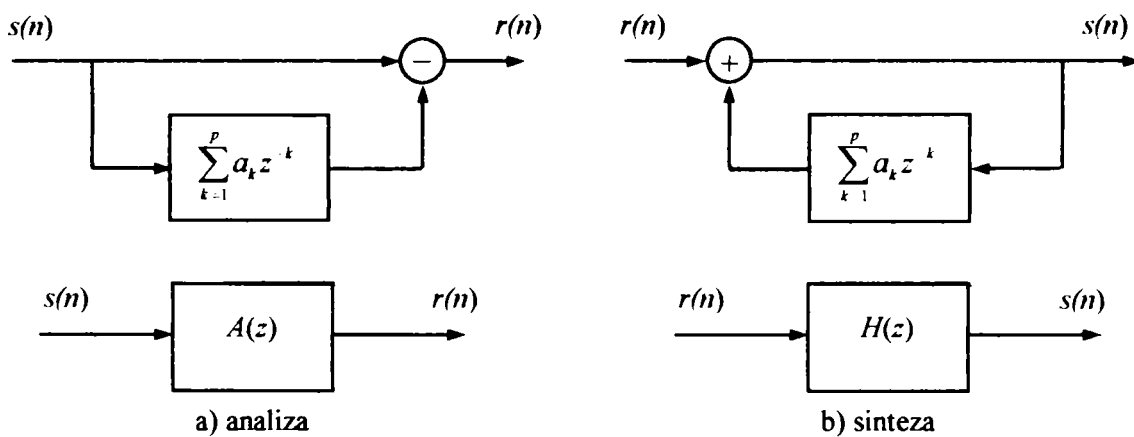


Fig. 3.1. Analiza și sinteza prin predicție liniară

În concluzie, un sistem de codare a semnalului vocal, compus dintr-un codor și un decodor care utilizează predicția liniară, se poate reprezenta schematic ca în figura 3.2.

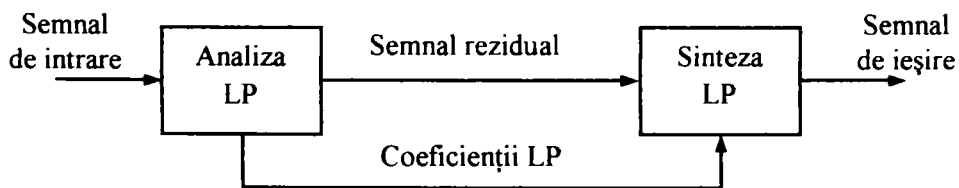


Fig. 3.2. Analiza și sinteza LP

Conform acestei scheme, în urma analizei LP a semnalului de intrare rezultă două categorii de informații care se cuantizează și se transmit la decodor:

- coeficienții de predicție,
- semnalul rezidual de predicție.

Modul în care se cuantizează cele două categorii de informații va fi analizat în detaliu în cap. 5 și respectiv cap. 6 și 7.

### 3.1.2 Prelucrarea semnalului pe cadre

Filtrele  $A(z)$  și  $H(z)$  din fig. 3.1, descrise de rel. (3.7), respectiv (3.8), reprezintă elementele esențiale ale unui sistem de codare a semnalului vocal prin predicție liniară. Pentru implementarea lor se poate folosi fie forma directă, care rezultă din relațiile echivalente în domeniul timp, fie forma latice, care folosește coeficienții de reflexie  $\{k_i\}$  [CHU03-2], a căror semnificație și mod de calcul vor fi prezentate în § 3.3.

Calculul semnalului de la ieșirea acestor filtre se face cadru cu cadru, având în vedere că implementarea practică a codoarelor de semnal vocal bazate pe predicție liniară se face prin divizarea semnalului în cadre.

Astfel, semnalul  $x(n)$ , definit pentru  $-\infty < n < \infty$ , se împarte în cadre conform relației:

$$x_r(n) = \begin{cases} x(n + N), & 0 \leq n \leq N - 1 \\ 0, & \text{în rest} \end{cases}, \quad (3.10)$$

fiecare cadru desemnat prin indicele  $r$  urmând a fi aplicat la intrarea unui filtru.

Divizarea în cadre este ilustrată în figura 3.3. Fiecare cadru constă din  $N$  eșantioane, notate cu indicele  $n$ , care ia valori cuprinse între 0 și  $N - 1$  în interiorul fiecărui cadru.

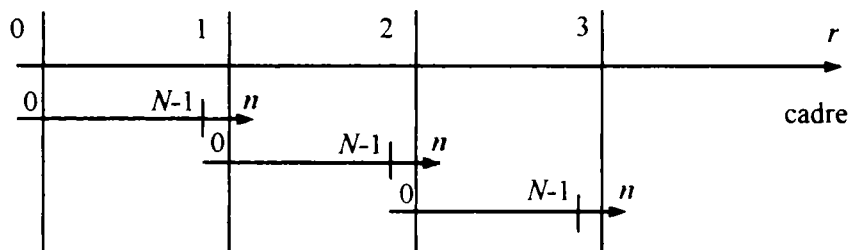


Fig. 3.3. Divizarea semnalului în cadre

Deoarece prelucrarea semnalului prin filtrare se face cadru cu cadru, este necesar să se asigure continuitatea prelucrării, și din acest motiv, pentru fiecare cadru, starea inițială a filtrului trebuie să fie egală cu starea finală a filtrului din cadrul precedent.

Semnalul  $s_r(n)$ , de la ieșirea filtrului, corespunzător cadrului  $r$ , este dat de relațiile:

$$s_r(n) = s_{r-1}(n + N) \quad -P \leq n \leq -1, \quad (3.11)$$

$$s_r(n) = x_r(n) - \sum_{i=1}^p a_i s_r(n - i) \quad 0 \leq n \leq N - 1, \quad (3.12)$$

fiind astfel memorate, conform rel. 3.11, cele  $P$  valori care caracterizează starea filtrului din cadrul precedent  $r - 1$ , pentru a fi folosite ca stare inițială în cadrul curent  $r$ .

O variantă alternativă de a ține cont de starea anterioară a filtrului în operația de filtrare a cadrului curent este reprezentată în figura 3.4 și constă în folosirea a două filtre  $H(z)$  identice [CHU03-2], [OSF89]:

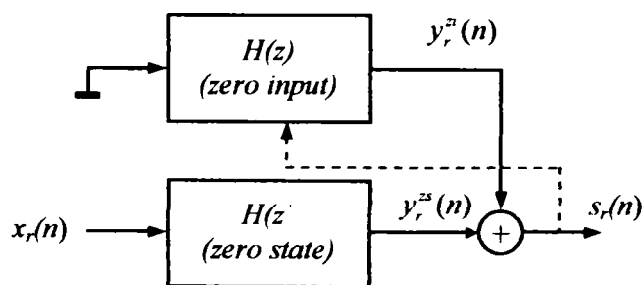


Fig. 3.4. Filtrarea „stare-zero – intrare-zero“

Unul din cele două filtre are semnalul de intrare zero, reprezentat în schema din fig. 3.4 prin conectarea intrării din punct de vedere electric la masă.

Semnalul de la ieșirea acestui filtru, notat cu indicele superior  $zi$  (*zero-input*), este:

$$y_r^{zi}(n) = s_{r-1}(n + N) \quad -P \leq n \leq -1, \quad (3.13)$$

$$y_r^z(n) = -\sum_{i=1}^p a_i y_r^z(n-i) \quad 0 \leq n \leq N-1 \quad (3.14)$$

și după cum este indicat în figură cu linie punctată, starea filtrului pe durata cadrului curent este egală cu semnalul de ieșire din cadrul anterior.

Al doilea filtru are starea inițială zero și cadrul de semnal care urmează a fi filtrat se aplică la intrare, rezultând un semnal de ieșire notat cu indicele superior  $zs$  (*zero-state*):

$$y_r^{zs}(n) = 0 \quad -M \leq n \leq -1, \quad (3.15)$$

$$y_r^{zs}(n) = x_r(n) - \sum_{i=1}^p a_i s_r^{zs}(n-i) \quad 0 \leq n \leq N-1. \quad (3.16)$$

Semnalul de ieșire se obține în final prin însumarea semnalelor de la ieșirile celor două filtre, cu intrarea zero și respectiv cu starea zero:

$$s_r(n) = y_r^z(n) + y_r^{zs}(n) \quad 0 \leq n \leq N-1. \quad (3.17)$$

Deși acest mod de calcul este ceva mai complicat, structura particulară a codorului CELP se pretează la aplicarea acestei variante [HRM01].

În cazul general, semnalul  $y(n)$  de la ieșirea unui filtru cu răspunsul la impuls  $h(n)$ , la a căruia intrare se aplică un cadru din semnalul  $x(n)$  este dat de produsul de convoluție:

$$y(n) = \sum_{k=-\infty}^{\infty} x(k)h(n-k) = \sum_{k=-\infty}^{\infty} x(n-k)h(k). \quad (3.18)$$

și prin urmare, semnalul rezultat  $y(n)$  este mai lung decât cadrul de semnal de intrare  $x(n)$ .

Pentru a păstra lungimi egale pentru cadrele de la intrarea și de la ieșirea filtrului, ar fi necesară trunchierea semnalului de la ieșire.

Dacă se consideră că răspunsul la impuls al filtrului este de lungime  $N$ , egală cu lungimea cadrului de intrare, sau, altfel spus,  $h(n) = 0$  pentru  $n \geq N$ , în cazul particular al filtrului cu starea inițială zero, răspunsul acestui filtru la un cadru de semnal de intrare are aceeași lungime ca și cadrul de intrare și se poate scrie, particularizând produsul de convoluție (3.18), în forma [CHU03-2]:

$$s_r^{zs}(n) = \sum_{k=0}^n x_r(k)h(n-k), \quad 0 \leq n \leq N-1. \quad (3.19)$$

Dezvoltând rel. (3.19), se poate constata că această relație se poate scrie compact în forma [CHU03-2]:

$$\mathbf{s}_r^{zs} = \mathbf{H}\mathbf{x}_r, \quad (3.20)$$

unde  $\mathbf{H}$  este o matrice cu dimensiunea  $N \times N$  care conține răspunsul la impuls al filtrului, în forma:

$$\mathbf{H} = \begin{bmatrix} h(0) & 0 & \dots & 0 \\ h(1) & h(0) & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ h(N-1) & h(N-2) & \dots & h(0) \end{bmatrix}, \quad (3.21)$$

$\mathbf{s}_r^{zs}$  este matricea coloană care conține semnalul de ieșire al filtrului, de forma:

$$\mathbf{s}_r^{zs} = [s_r^{zs}(0) \ s_r^{zs}(1) \ s_r^{zs}(2) \ \dots \ s_r^{zs}(N-1)]^T, \quad (3.22)$$

iar  $\mathbf{x}_r$  este matricea coloană care conține eșantioanele ce compun cadrul de semnal de intrare:

$$\mathbf{x}_r = [x_r(0) \ x_r(1) \ x_r(2) \ \dots \ x_r(N-1)]^T \quad (3.23)$$

Această metodă de calcul a semnalului de ieșire al filtrului este avantajoasă în cazul codorului CELP, deoarece permite aplicarea convoluției recursive dacă secvențele de intrare corespunzătoare unor cadre succesive sunt parțial suprapuse [CHU03-2].

### 3.2 Calculul coeficienților de predicție

Pentru determinarea coeficienților de predicție  $\{a_k\}$  din rel. (3.3), se minimizează energia erorii de predicție (3.5) în raport cu acești coeficienți. Calculul se poate face în două moduri diferite, care conduc la același rezultat [STA98].

#### 3.2.1 Fereastra de ponderare

Deoarece semnalul vocal este un semnal aleator nestaționar, iar tractul vocal se modifică lent în timp, semnalul este divizat în segmente succesive, denumite cadre, pe durata cărora semnalul este considerat staționar. Cadrele se obțin prin multiplicare cu o fereastră de analiză de lungime fixă  $w(n)$ . Lungimea ferestrei de analiză este în general egală cu lungimea cadrului și prin poziționarea ferestrei se poate asigura o suprapunere parțială a cadrelor succesive.

Cea mai simplă fereastră este cea dreptunghiulară, cu lungimea  $N_w$ , dată de relația:

$$w(n) = \begin{cases} 1, & 0 \leq n \leq N_w - 1, \\ 0, & \text{în rest,} \end{cases} \quad (3.24)$$

cu dezavantajul unor variații abrupte la extremități, în domeniul timp, care se manifestă în domeniul frecvență prin lobi laterali pronunțați și printr-un efect denumit „ringing” și care constă dintr-un brum suprapus peste semnalul vocal [PRM92].

Pentru experimentele efectuate în lucrare se folosește o fereastră de tip Hamming, descrisă de relația:

$$w(n) = \begin{cases} 0,54 - 0,46 \cos\left(\frac{2\pi n}{N_w + 1}\right), & 0 \leq n \leq N_w - 1, \\ 0, & \text{în rest.} \end{cases} \quad (3.25)$$

Se pot folosi cu rezultate similare și alte tipuri de ferestre de ponderare, ca Hanning, Blackman sau Kaiser, sau ferestre hibride, care sunt alcătuite din două jumătăți de fereastră cu parametrii diferiți, atribuind astfel o pondere mai mare valorilor mai recente [CHU03-1].

Fereastra de ponderare determină un interval temporal de semnal care se folosește, după cum se va arăta în cele ce urmează, pentru calculul coeficienților de predicție, care, la rândul lor, se vor folosi pentru a calcula valorile semnalului pe durata unui cadru. După modul de amplasare relativă a celor două intervale, reprezentat în figura 3.5, se poate vorbi despre **predicție internă** sau **predicție externă**. În cazul predicției interne, coeficienții de predicție sunt calculați folosind valorile semnalului pe durata unui cadru și apoi sunt folosiți pentru a procesa semnalul din același cadru. În cazul predicției externe, coeficienții calculați astfel se folosesc pentru procesarea cadrului următor, sau, altfel spus, coeficienții de predicție folosiți pe durata unui cadru nu sunt asociați cu valorile semnalului din cadrul respectiv, ci cu valorile din cadrul precedent. Justificarea folosirii predicției externe este dată de faptul că statistica semnalului se modifică relativ lent în timp și pentru cadre de lungime convenabilă se păstrează o bună corelație între cadrele vecine.

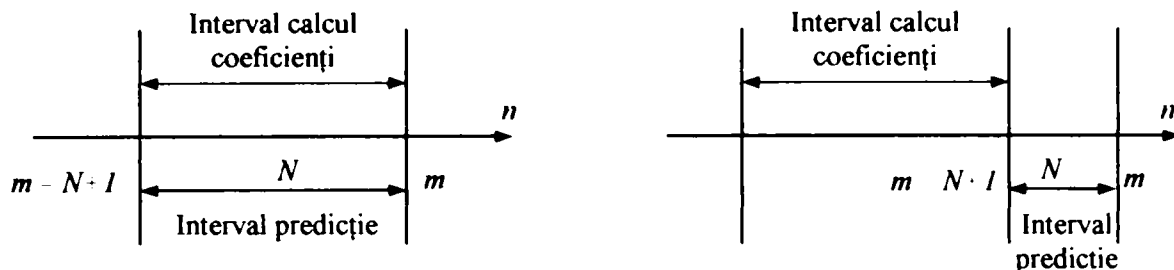


Fig. 3.5. Predicția internă și predicția externă

Predicția internă este folosită în multe scheme de codare, durata tipică a cadrului fiind cuprinsă între 160 – 240 eşantioane. Un cadru de durată mai mare este avantajos deoarece conduce la o rată redusă de actualizare a coeficienților de predicție, rezultând o reducere a volumului de calcule și totodată o reducere a ratei de bit. Ca dezavantaje, pe de o parte, coeficienții de predicție calculați pentru un cadru mai lung pot conduce la un câștig de predicție mai scăzut datorită nestaționarității semnalului, iar pe de altă parte un cadru lung introduce o întârziere mai mare.

Predicția externă este folosită preponderent în codoarele de întârziere redusă, când, datorită cerințelor de întârziere, se impune folosirea unui cadru de predicție de durată mică, o valoare tipică fiind 20 eşantioane. În acest caz, coeficienții de predicție se calculează pe baza eşantioanelor din afara acestui cadru (anterioare) de obicei în mod recursiv.

Există și situații în care predicția nu se încadrează clar în unul din cele două cazuri de predicție internă, respectiv externă prezentate. Este cazul unor scheme de codare în care coeficienții de predicție, calculați pentru două cadre adiacente, se combină într-un mod specific, prin interpolare, pentru a realiza predicția pe durata corespunzătoare acestor cadre.

### 3.2.2 Metoda autocorelației

Semnalul vocal ponderat cu fereastra de ponderare este:

$$s_w(n) = w(n)s(n) \tag{3.26}$$

și în acest caz, energia semnalului rezidual de predicție este:

$$E = \sum_{n=-\infty}^{\infty} e^2(n) = \sum_{n=-\infty}^{\infty} \left( s_w(n) - \sum_{k=1}^p a_k s_w(n-k) \right)^2 \tag{3.27}$$

Valorile  $\{a_k\}$  care minimizează energia erorii (3.27) se pot calcula egalând cu 0 derivatele parțiale ale energiei în raport cu coeficienții  $\{a_k\}$ :

$$\frac{\partial E}{\partial a_k} = 0, \text{ pt. } k = 1, 2, \dots, p, \tag{3.28}$$

ceea ce conduce la un sistem de  $p$  ecuații care are ca necunoscute cei  $p$  coeficienți  $\{a_k\}$ :

$$\sum_{k=1}^p a_k \sum_{n=-\infty}^{\infty} s_w(n-i)s_w(n-k) = \sum_{n=-\infty}^{\infty} s_w(n-i)s_w(n) \text{ pentru } 1 \leq i \leq p. \tag{3.29}$$

În ecuația (3.29) semnalul vocal ponderat cu fereastra  $w(n)$  este nul în afara ferestrei, prin urmare sumele din ecuația (3.29) se calculează numai pe durata  $N_w$  a ferestrei și se poate constata că aceste sume reprezintă funcția de autocorelație a semnalului ponderat  $s_w(n)$ , definită prin:



$$R(i) = \sum_{n=i}^{N_w-1} s_w(n)s_w(n-i), \quad 0 \leq i \leq p. \quad (3.30)$$

Având în vedere că funcția de autocorelație este o funcție pară și substituind (3.30) în (3.29), ecuația (3.29) devine:

$$\sum_{k=1}^p R(|i-k|)a_k = R(i), \quad 1 \leq i \leq p. \quad (3.31)$$

În formă matricială, acest sistem de  $p$  ecuații se poate scrie:

$$\begin{bmatrix} R(0) & R(1) & \dots & R(p-1) \\ R(1) & R(0) & \dots & R(p-2) \\ \dots & \dots & \dots & \dots \\ R(p-1) & R(p-2) & \dots & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_p \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ \dots \\ R(p) \end{bmatrix} \quad (3.32)$$

sau, în formă matricială restrânsă:

$$\mathbf{R}\mathbf{a} = \mathbf{r}. \quad (3.33)$$

Matricea  $\mathbf{R}$  este o matrice Toeplitz, ceea ce permite rezolvarea sistemului (3.33) cu algoritmul Levinson – Durbin sau cu algoritmul Schur, garantând totodată obținerea unui polinom  $A(z)$  de fază minimă și asigurând astfel stabilitatea filtrului de sinteză  $H(z) = 1/A(z)$ . [STA98] [KNT87], [EBR97]

O metodă eficientă de calcul a coeficienților de predicție este algoritmul Levinson – Durbin, care permite rezolvarea unor ecuații de tipul  $\mathbf{A}\mathbf{x} = \mathbf{b}$ , unde  $\mathbf{A}$  este o matrice Toeplitz, iar  $\mathbf{b}$  este un vector care conține anumite elemente din  $\mathbf{A}$ . Sistemul (3.33) îndeplinește aceste condiții și poate fi rezolvat prin algoritmul Levinson – Durbin.

Fie  $a_k^{(m)}$  coeficientul  $a_k$  din iterația  $m$ . Algoritmul Levinson – Durbin rezolvă sistemul (3.33) prin următoarele relații de recurență, pentru  $m = 1, 2, \dots, p$ :

- inițializări:  $m = 0$ ,  $E_0 = R(0)$  și  $a_0 = 0$ .
- pentru  $m = 1, 2, \dots, p$  calculează coeficientul de reflexie:

$$k_m = \frac{1}{E_{m-1}} [R(m) - \sum_{k=1}^{m-1} a_k^{(m-1)} R(m-k)]. \quad (3.34)$$

- calculează coeficienții de predicție pentru predictorul de ordin  $m$ :

$$a_m^{(m)} = k_m, \quad (3.35)$$

$$a_j^{(m)} = a_j^{(m-1)} - k_m a_{j-1}^{(m-1)}, \quad 1 \leq j \leq m-1. \quad (3.36)$$

- calculează eroarea de predicție:

$$E_m = (1 - k_m^2) E_{m-1}. \quad (3.37)$$

- stop dacă  $m = p$ , altfel continuă pentru  $m \leftarrow m + 1$

La fiecare iterație sunt calculați coeficienții de predicție  $a_k^{(m)}$  care reprezintă predictorul liniar optimal de ordin  $m$ . De asemenea, la fiecare iterație, eroarea de predicție este redusă cu factorul  $(1 - k_m^2)$  față de valoarea precedentă. Deoarece eroarea medie pătratică nu poate fi negativă, relația (3.37) conduce la concluzia  $|k_m| \leq 1$ , iar această condiție garantează situarea rădăcinilor polinomului  $A(z)$  în interiorul cercului unitate, asigurând astfel stabilitatea filtrului de sinteză  $H(z)$  [KNT87].

Coeficienții  $k_m$  se numesc coeficienți de corelație parțială și pot fi folosiți direct pentru a implementa o formă latice a filtrului de sinteză.

Din punct de vedere al complexității de calcul, pentru un ordin al predicției  $p = 10$  (valoare tipică pentru un semnal cu banda telefonică 300 – 3400 Hz), algoritmul Levison-Durbin presupune 110 înmulțiri, 100 adunări și 10 împărțiri.

### 3.2.3 Metoda covarianței

Metoda covarianței este asemănătoare cu metoda autocorelației, diferența esențială constând în efectuarea operației de ponderare cu o fereastră adecvată nu direct asupra semnalului vocal de intrare, ci asupra semnalului de eroare de predicție. În acest caz, energia erorii de predicție ponderate este:

$$E = \sum_{n=-\infty}^{\infty} e_w^2(n) = \sum_{n=-\infty}^{\infty} e^2(n)w(n). \quad (3.31)$$

Anulând derivatele parțiale ale energiei de predicție (3.38) în raport cu coeficienții de predicție și ținând cont că însumarea în rel. (3.38) se face doar pe durata ferestrei, termenii din afara ferestrei fiind nuli, se obține un set de  $p$  ecuații liniare cu  $p$  necunoscute de forma:

$$\sum_{k=1}^p \Phi(i, k)a_k = \Phi(i, k), \quad 1 \leq i \leq p, \quad (3.39)$$

unde  $\Phi(i, k)$  este funcția de covarianță, definită prin:

$$\Phi(i, k) = \sum_{n=-\infty}^{\infty} w(n)s(n-i)s(n-k). \quad (3.40)$$

Acest sistem se poate scrie în formă matricială similar cu (3.32):

$$\begin{bmatrix} \phi(1,1) & \phi(1,2) & \phi(1,p) \\ \phi(2,1) & \phi(2,2) & \phi(2,p) \\ \dots & \dots & \dots \\ \phi(p,1) & \phi(p,2) & \phi(p,p) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_p \end{bmatrix} = \begin{bmatrix} \phi(1) \\ \phi(2) \\ \dots \\ \phi(p) \end{bmatrix} \quad (3.41)$$

unde  $\phi(i) = \phi(i, 0)$ ,  $i = 1, 2, \dots, p$ ,

sau în formă matricială restrânsă:

$$\Phi \mathbf{a} = \boldsymbol{\varphi}. \quad (3.42)$$

În acest caz însă, matricea  $\Phi$  din ecuația (3.42) nu mai este matrice Toeplitz, dar este o matrice simetrică și pozitiv definită, iar sistemul (3.42) se poate rezolva prin metoda descompunerii, fără însă ca această metodă să garanteze stabilitatea filtrului de sinteză [KNT87].

### 3.2.4 Câștigul de predicție

Câștigul de predicție pentru un predictor dat este definit prin raportul dintre varianța semnalului de intrare și varianța erorii de predicție, exprimat în decibeli:

$$G = 10 \log_{10} \left( \frac{\sigma_s^2}{\sigma_r^2} \right) = 10 \log_{10} \left( \frac{E\{s^2(n)\}}{E\{r^2(n)\}} \right). \quad (3.43)$$

Câștigul de predicție este o măsură a performanței predictorului, un câștig de predicție mare indicând o eroare de predicție mică, deci un bun predictor.

### 3.2.5 Predicția liniară a semnalului vocal

Predicția liniară, descrisă într-un context general, prezintă câteva particularități specifice atunci când este aplicată semnalului vocal.

Pentru aplicarea predicției liniare se presupune că semnalul vocal satisface modelul AR.

În figura 3.6 s-a reprezentat un fragment de semnal vocal sonor (stânga) și un fragment de semnal nesonor (dreapta). Ambele semnale au durată de 300 eșantioane. Se observă caracterul cvasiperiodic al semnalului sonor, cu o periodicitate estimată de 54 eșantioane și de asemenea amplitudinea mai mare a fragmentului sonor față de cel nesonor.

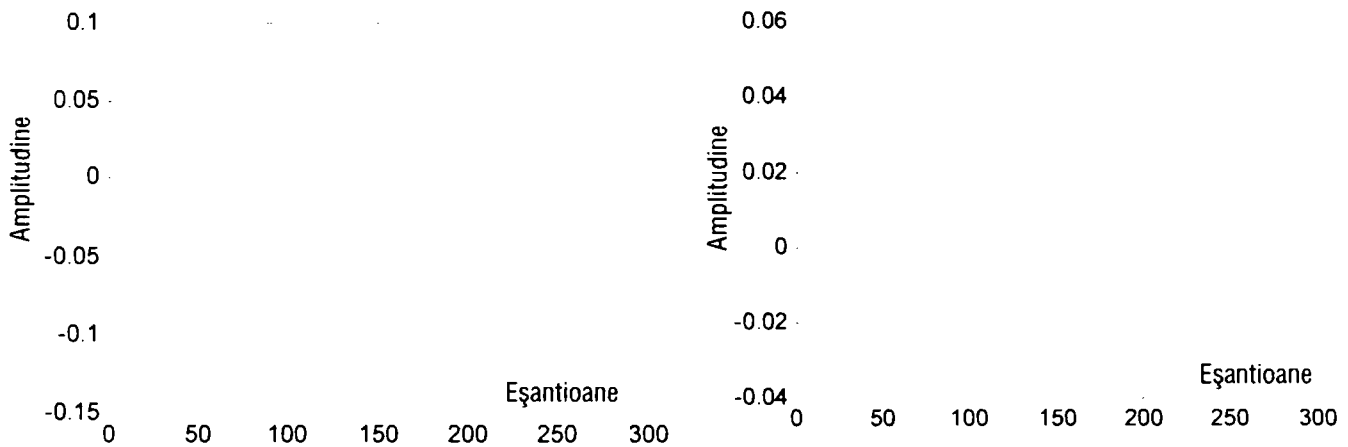


Fig.3.6. Semnal vocal sonor (stânga) și nesonor (dreapta)

În figura 3.7 s-au reprezentat spectrele celor două semnale din fig. 3.6, calculate folosind FFT în 256 puncte (reprezentate cu linie continuă), respectiv estimate prin predicție liniară de ordinul 10 (linie punctată).

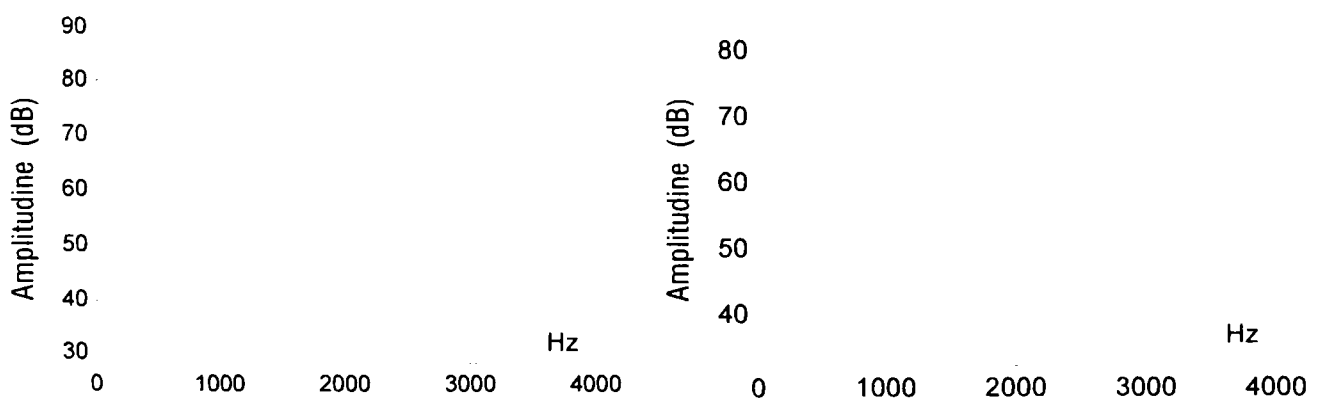


Fig.3.7. Spectrul FFT și LPC pentru semnal vocal sonor (stânga) și nesonor (dreapta)

Pentru semnalul sonor, spectrul evidențiază structura formantică, în timp ce pentru semnalul nesonor spectrul este relativ plat.

Eroarea de predicție este reprezentată în figura 3.8 pentru cazul  $p = 10$  și respectiv în figura 3.9 pentru cazul  $p = 60$ . Pentru calculul erorii de predicție reprezentată în fig 3.8, am folosit pentru calculul coeficienților de predicție algoritmul Levinson – Durbin prezentat anterior, iar pentru implementarea filtrului de analiză facilitățile de filtrare a semnalelor din Matlab [CBA03-1]

Analizând eroarea și câștigul de predicție pentru semnalul vocal, se pot trage câteva concluzii utile pentru realizarea unui codor care folosește predicția liniară [PVT00].

Pentru un ordin dat, câștigul de predicție este mai mare pentru cadrele de semnal sonor decât pentru cele de semnal nesonor, fapt justificat de corelația mai mare care există în cadrele de semnal sonor, în timp ce semnalul nesonor are un caracter mai apropiat de zgomotul alb.

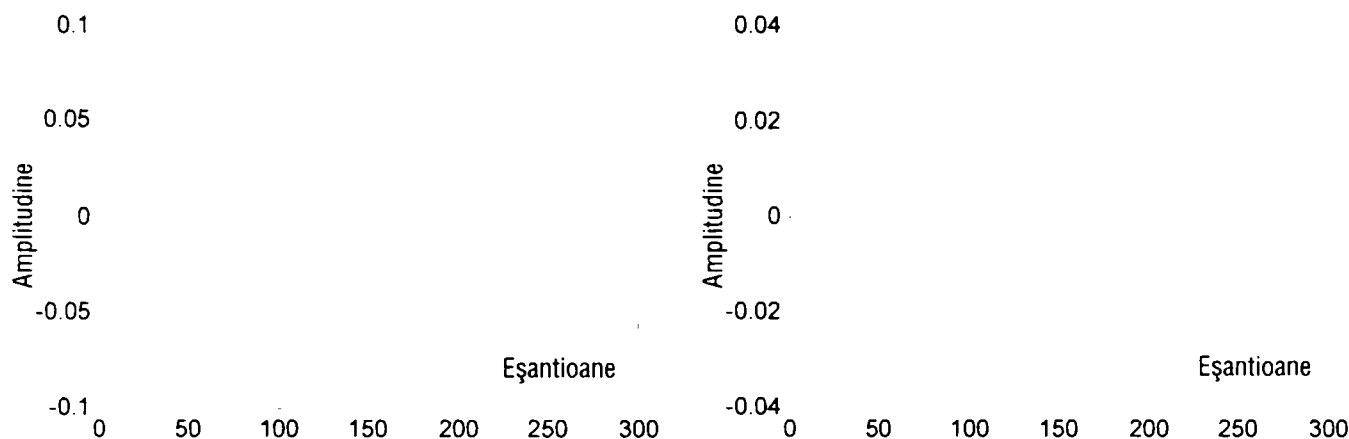


Fig.3.8. Eroarea de predicție pentru  $p = 10$

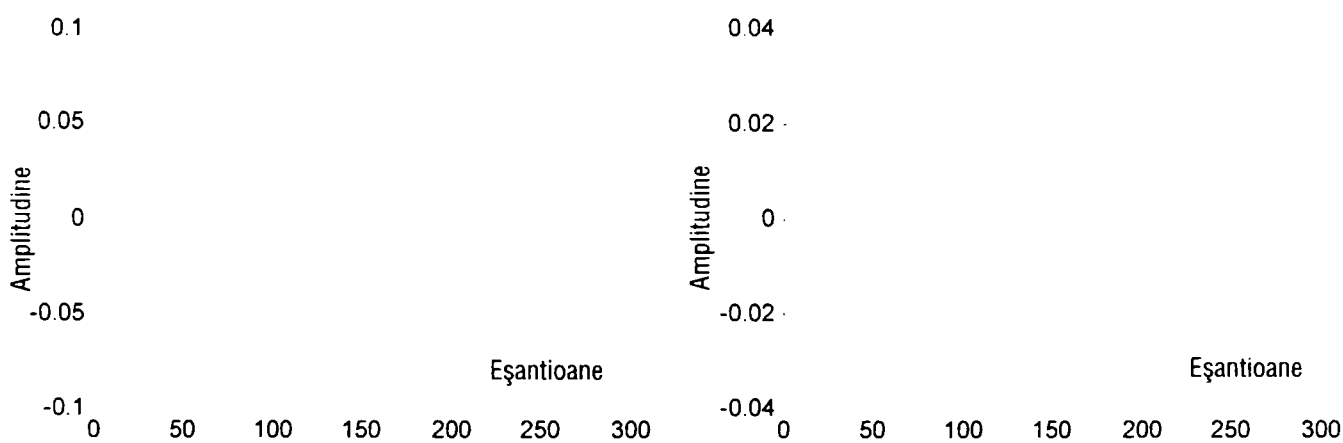


Fig.3.9. Eroarea de predicție pentru  $p = 60$

Câștigul de predicție crește substanțial pentru creșterea ordinului de predicție până la valoarea 10 – 12, după care rămâne aproximativ constant.

Această evoluție justifică alegerea ordinului de predicție  $p = 10$ , folosit în foarte multe scheme de codare a semnalului vocal prin predicție liniară. Aceste considerente sunt valabile pentru un semnal vocal cu banda limitată la 4 kHz, adică eșantionat la 8 kHz. Pentru semnale de bandă mai largă, ordinul de predicție trebuie crescut în mod corespunzător.

Pentru cadrul de semnal sonor, câștigul de predicție va prezenta o creștere semnificativă dacă ordinul de predicție depășește perioada fundamentală, exprimată în multipli ai perioadei de eșantionare. Acest fapt este evidențiat și în figurile 3.8 și 3.9, care arată scăderea erorii de predicție pentru cadrul de semnal sonor, când ordinul de predicție este mai mare (60) și este justificat de faptul că, în acest caz, predicția acoperă și periodicitatea datorată frecvenței fundamentale. În concluzie, pentru o predicție corectă a semnalului vocal sonor, este necesar un predictor de ordin relativ mare, care ar fi dezavantajos din punct de vedere al volumului de calcule. Se poate evita acest lucru, după cum se va arăta ulterior, folosind un predictor de ordin redus și modelând în alt mod periodicitatea caracteristică semnalului vocal sonor.

Codarea semnalului vocal prin predicție liniară trebuie să țină cont și de alte considerente de ordin practic, dintre care câteva vor fi prezentate în continuare.

• **Preaccentuarea frecvențelor înalte**

Analizând spectrul unui semnal vocal, corelat cu analiza mecanismului de producere a semnalului vocal, se constată o cădere a spectrului în zona frecvențelor înalte, datorită efectului produs de radiația sunetului la buze. Amplitudinea redusă a componentelor de frecvență înaltă din spectru conduce la o dinamică ridicată a spectrului, cu consecințe negative asupra analizei semnalului vocal prin predicție liniară. Astfel, pentru a evidenția și componentele spectrale de frecvență ridicată și mică amplitudine, analiza prin predicție liniară trebuie să fie mai precisă, în această zonă a spectrului manifestându-se efectele rotunjirilor, inevitabile într-o implementare reală. În plus, este posibil ca matricea de autocorelație să devină chiar singulară, făcând dificilă rezolvarea ecuațiilor de determinare a coeficienților de predicție. Soluția problemei constă în preaccentuarea semnalului vocal, înainte de prelucrare, folosind un filtru cu funcția de transfer de forma:

$$H(z) = 1 - \alpha z^{-1}, \quad \alpha > 0 \tag{3.44}$$

care prezintă o caracteristică de transfer de tip trece-sus și care va favoriza frecvențele înalte ale spectrului.

Pentru a păstra nealterat spectrul semnalului vocal obținut prin predicție liniară, se aplică acestui semnal o dezaccentuare, în partea de decodor, cu un filtru de tip trece-jos, cu o funcție de transfer inversă celei de preaccentuare:

$$G(z) = \frac{1}{1 - \alpha z^{-1}}. \tag{3.45}$$

O valoare tipică pentru parametrul  $\alpha$  este cuprinsă în intervalul 0,8 – 0,9.

• **Expandarea benzii de trecere**

Analiza prin predicție liniară nu poate estima corect anvelopa spectrală a semnalului vocal pentru semnalele cu frecvența fundamentală de valori reduse, datorită distanței mari între armonicele acestui semnal în domeniul timp și, ca urmare, va rezulta o îngustare artificială a lărgimii de bandă a formanților, care vor fi subestimați, conferind sunetului un caracter metalic. Rezolvarea acestei probleme se poate face prin expansiunea benzii formanților, prin multiplicarea fiecărui coeficient de predicție  $a_k$  cu un factor  $\gamma^k$ , unde  $\gamma$  este un coeficient subunitar, rezultând filtrul de sinteză [CHU03-2]:

$$H'(z) = \frac{1}{A'(z)} = \frac{1}{A(z/\gamma)}, \tag{3.46}$$

având ca efect o deplasare a polilor  $H(z)$  spre interiorul cercului unitate cu factorul  $\gamma$  subunitar și o lărgire a benzii de trecere corespunzătoare cu valoarea  $\Delta B$ , dată de:

$$\Delta B = -\frac{F_s}{\pi} \ln \gamma, \tag{3.47}$$

unde  $F_s$  este frecvența de eșantionare [CHU03-2].

Ca efect pozitiv, sensibilitatea spectrului în jurul vârfurilor spectrale scade și astfel se reduc efectele eventualelor erori de cuantizare a coeficienților de predicție. De asemenea, îndepărtarea polilor de cercul unitate crește stabilitatea filtrului de sinteză și reduce riscul ca polii să ajungă pe cercul unitate sau chiar în afara sa datorită erorilor de cuantizare și transmisie.

Valorile tipice pentru parametrul  $\gamma$  sunt cuprinse între 0,996 și 0,998, ceea ce corespunde unei valori  $\Delta B$  cuprinsă între 10 și 30 Hz, la frecvența de eșantionare de 8 kHz.

- **Corecția prin adăugarea unui zgomot alb**

Corecția prin adăugarea unui zgomot alb semnalului vocal îmbunătățește condițiile de analiză prin predicție liniară, având același efect ca și preaccentuarea frecvențelor înalte, adică reducerea dinamicii spectrale.

Corecția se face prin multiplicarea valorii coeficientului de autocorelație pentru întârziere zero cu o constantă:

$$R(0) \leftarrow \lambda R(0). \quad (3.48)$$

Constanta  $\lambda > 1$  are valori cu puțin mai mari decât unitatea. De exemplu, codorul LD-CELP G.728 folosește  $\lambda = 257/256 = 1,00390625$ , care produce o creștere cu 0,39% a coeficientului  $R(0)$ . Aceasta este echivalent cu a adăuga semnalului vocal inițial, înainte de prelucrare, un zgomot alb având puterea cu 24 dB sub nivelul de putere al semnalului vocal.

- **Netezirea spectrală**

Un efect echivalent cu expandarea benzii de trecere, prezentată anterior, se poate obține prin netezire spectrală [CHU03-2]. Dacă expandarea benzii se face prin operațiuni efectuate asupra coeficienților de predicție, după ce aceștia au fost calculați, netezirea spectrală acționează înaintea calculului coeficienților, prin aplicarea unei ferestre cu formă gaussiană asupra funcției de autocorelație a semnalului. Astfel, dacă funcția de autocorelație a semnalului este  $R(i)$ , după aplicarea ferestrei se obține:

$$R'(i) = R(i) \cdot w(i), \quad i = 0, 1, \dots, p, \quad (3.49)$$

unde fereastra este definită prin:

$$w(i) = e^{-\beta i^2}, \quad (3.50)$$

iar  $\beta$  este o constantă. O valoare tipică pentru  $\beta$  este 0,001.

### 3.3. Reprezentări ale coeficienților de predicție

Coeficienții LPC definiți prin rel. (3.7) nu au proprietăți adecvate pentru a se cuantiza și transmite direct. Deoarece pot avea orice valoare reală, sunt dificil de cuantizat și, în plus, orice eroare de cuantizare sau transmisie se reflectă prin modificări importante ale spectrului. De asemenea, sunt necesare măsuri pentru asigurarea stabilității filtrului de sinteză. Sunt cunoscute și utilizate mai multe reprezentări alternative ale coeficienților de predicție, care prezintă unele avantaje în privința cuantizării [CHU03-2].

#### 3.3.1 Coeficienții de reflexie

Coeficienții de reflexie sunt definiți de rel. (3.34) și constituie o reprezentare alternativă a coeficienților de predicție. După cum rezultă din rel. (3.34) – (3.37) care definesc algoritmul Levinson – Durbin, coeficienții de reflexie rezultă direct prin aplicarea acestui algoritm și au avantajul că se pot folosi direct pentru implementarea unui filtru de sinteză în forma latică [KNT87], [MGH90].

Polinomul  $A(z)$  și coeficienții de reflexie  $k_p$  îndeplinesc următoarea relație de recurență:

$$A_p(z) = A_{p-1}(z) - k_p z^{-p} A_{p-1}(z^{-1}), \quad p = 1, 2, \dots, P. \quad (3.51)$$

în care  $A_p(z)$  este polinomul de predicție de grad  $p$ .



Dacă pentru calculul coeficienților de predicție se folosește metoda covarianței, este necesară conversia coeficienților LPC  $\{a_i\}$  în coeficienții de reflexie  $\{k_i\}$  și verificarea condiției de stabilitate  $|k_i| \leq 1$ .

Coeficienții de reflexie prezintă avantajul de a fi limitați între  $-1$  și  $+1$ , permițând prin aceasta o cuantizare mai eficientă și dacă prin cuantizare și transmisie este păstrată condiția  $|k_i| \leq 1$ , este asigurată stabilitatea filtrului de sinteză tip latice.

### 3.3.2 Coeficienții LAR

Coeficienții LAR (*Log Area Ratio*) sunt definiți prin:

$$K_m = \log \frac{1+k_m}{1-k_m} \quad m = 1, 2, \dots, p, \quad \text{unde } k_m \text{ sunt coeficienții de reflexie.} \quad (3.52)$$

Prin funcția neliniară de conversie se elimină dezavantajele coeficienților de reflexie, care prezintă o sensibilitate neuniformă. Astfel, primul și al doilea coeficient de reflexie sunt foarte apropiați de valorile  $-1$ , respectiv  $+1$  și au sensibilitate ridicată, așa încât mici erori de cuantizare pot conduce la modificări semnificative ale caracteristicii filtrului de sinteză.

Transformarea neliniară în coeficienți LAR definită prin (3.52) elimină acest dezavantaj și poate fi urmată de o cuantizare liniară. Relația (3.52) permite deducerea simplă a transformării inverse, din coeficienți LAR în coeficienți de reflexie.

În unele cazuri, funcția de conversie directă, exprimată de rel. (3.52), precum și funcția de conversie inversă, sunt approximate prin funcții liniare pe porțiuni [ETS97-1].

O altă transformare neliniară care se poate aplica coeficienților de reflexie în scopul cuantizării eficiente este transformarea sinus invers (IS):

$$s_m = \sin^{-1}(k_m), \quad m = 1, 2, \dots, p, \quad (3.53)$$

pentru care revenirea la coeficienții de reflexie se face aplicând funcția sinus direct.

### 3.3.3 Perechile de linii spectrale

Perechile de linii spectrale (LSP sau LSF) (LSP – Line Spectrum Pairs, LSF – Line Spectrum Frequencies), introduse pentru prima dată în [ITA75], constituie o modalitate alternativă de reprezentare a coeficienților de predicție care prezintă proprietăți foarte convenabile pentru cuantizare și pentru testarea stabilității filtrului de sinteză, motiv pentru care au fost preferate în mai multe metode de codare devenite standarde în domeniu, ca de exemplu G.729, G.731 sau FS1016.

Pentru a defini LSP se consideră două polinoame de grad  $p + 1$ , care se construiesc pornind de la polinomul  $A(z)$  care reprezintă filtrul de analiză prin predicție astfel:

$$P(z) = A(z) + z^{-(p+1)} A(z^{-1}), \quad (3.54)$$

$$Q(z) = A(z) - z^{-(p+1)} A(z^{-1}). \quad (3.55)$$

Polinomul  $A(z)$  se poate calcula, dacă se cunosc  $P$  și  $Q$ , prin relația evidentă:

$$A(z) = \frac{P(z) + Q(z)}{2}. \quad (3.49)$$

Conform relațiilor de definiție (3.47) și (3.48), polinomul  $P(z)$  este simetric, iar  $Q(z)$  este antisimetric.

Ambele polinoame au gradul  $p + 1$ , deci fiecare are  $p + 1$  rădăcini. Cele două polinoame  $P$  și  $Q$  au rădăcinile  $z = 1$  și respectiv  $z = -1$ , astfel: valoarea  $z = 1$  este rădăcină pentru

polinomul  $Q(z)$ , iar valoarea  $z = -1$  este rădăcină pentru polinomul  $Q(z)$  dacă  $p$  este impar și rădăcină pentru polinomul  $P(z)$  dacă  $p$  este par.

Celelalte rădăcini ale polinoamelor  $P(z)$  și  $Q(z)$  apar în perechi complex conjugate și sunt situate pe cercul unitate în planul  $z$ .

Aceste proprietăți permit să se reprezinte polinomul  $A(z)$  cu ajutorul a  $p$  valori care reprezintă unghiuri cuprinse între  $0$  și  $\pi$ ,  $0 < \omega_1 < \omega_2 < \dots < \omega_p < \pi$  și care constituie perechile de linii spectrale (LSP) [KLE99], [KWG01], [PAL95].

Frecvențele corespunzătoare, denumite LSF, sunt  $f_i = \omega_i / 2\pi$  și păstrează aceeași relație de ordine, fiind cuprinse între  $0$  și  $0,5$  (frecvențe relative), sau, echivalent, între  $0$  și  $f_c / 2$ .

### 3.4 Codorul CELP

Codorul CELP, prezentat pe scurt în cap. 2 (2.2.4.2), a apărut ca urmare a încercărilor de a îmbunătăți codorul cu predicție liniară. Acronimul CELP – Code Excited Linear Prediction, introdus odată cu principiul analizei prin sinteză [SCA85], sugerează folosirea unui dicționar care conține „coduri” sau vectori pentru a excita filtrul de sinteză LPC.

Codorul CELP folosește modelul predicției liniare pe termen scurt și termen lung pentru producerea semnalului vocal, prezentat în figura 3.10. Conform acestui model, semnalul de excitație este extras dintr-un dicționar de secvențe de excitație printr-un index și după ce este adus la un nivel corespunzător prin înmulțirea cu un factor de câștig  $G$  este filtrat prin două filtre de predicție liniară conectate în cascadă.

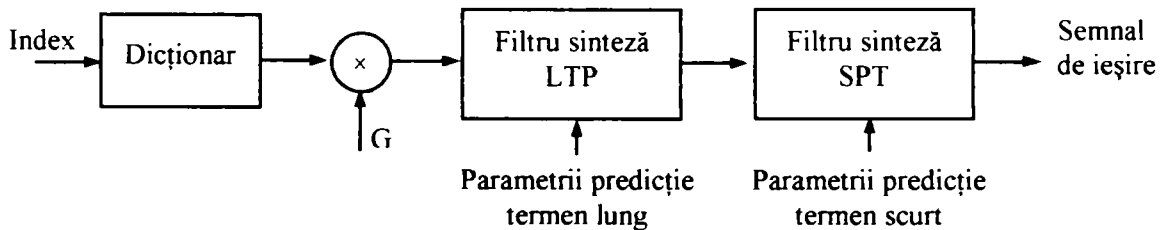


Fig. 3.10. Modelul CELP pentru producerea semnalului vocal

Filtrul de sinteză pe termen lung LTP (Long term prediction) conferă semnalului de excitație caracterul periodic asociat cu frecvența fundamentală a semnalului vocal, iar filtrul de sinteză pe termen scurt STP (Short term prediction) asigură în final anvelopa spectrală a semnalului de ieșire. Justificarea acestor denumiri va fi dată ulterior, odată cu prezentarea în detaliu a acestor operații.

Secvențele de excitație din dicționar pot fi deterministe sau aleatoare, dar pentru simplitate se preferă în cele mai multe cazuri secvențe aleatoare, formate din eșantioane necorelate de zgomot alb. Alegerea secvenței de excitație optimă din dicționar se face în faza de analiză printr-un procedeu în buclă închisă, care alege acea secvență care, filtrată prin cele două filtre de predicție pe termen lung și respectiv scurt, generează semnalul de ieșire cel mai apropiat de semnalul de intrare.

Acest procedeu asigură două avantaje:

- elimină necesitatea clasificării stricte a semnalului vocal în semnal sonor – nesonor, necesară în cazul codoarelor LPC în buclă deschisă, realizate conform modelului prezentat în fig. 1.1 (Cap.1), operație care, în unele cazuri, nu se poate face exact, deoarece semnalul vocal prezintă tranziții, care nu pot fi net încadrate în una din categoriile sonor sau nesonor [NGM01], [QRH93].

- asigură păstrarea unor informații parțiale despre faza semnalului, deoarece comparația între semnalul de ieșire și cel de intrare se face în domeniul timp, spre deosebire de codorul în

buclă deschisă, care nu asigură nici un fel de informație despre faza semnalului, predicția liniară operând în domeniul timp, iar filtrul de predicție fiind de fază minimă.

Principiul codării în buclă închisă este ilustrat în figura 3.11. Parametrii modelului, reprezentați de caracteristicile semnalului de excitație, se pot determina în două moduri conceptuale diferite: în buclă deschisă și în buclă închisă. Modelul în buclă deschisă extrage parametri direct din semnalul de intrare prin operații asupra acestui semnal.

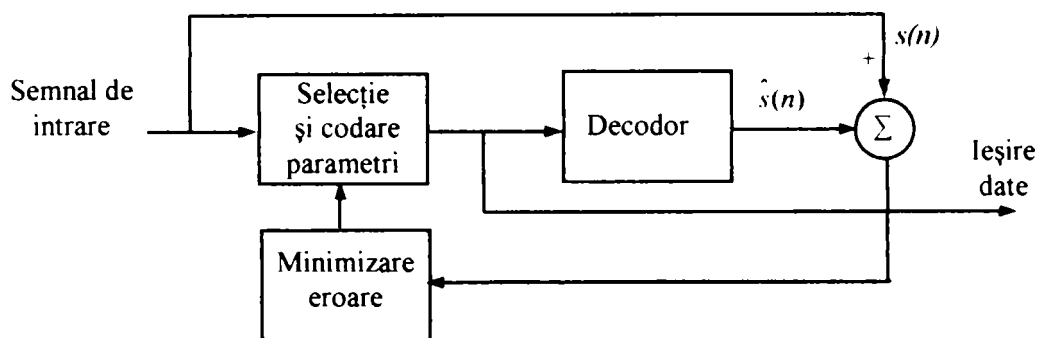


Fig 3.11. Principiul codării în buclă închisă

Modelul în buclă închisă din fig. 3.11 conține o replică a decodurului inclusă în codor, motiv pentru care este denumit și codor cu analiză prin sinteză și determină, prin optimizare, acei parametri pentru model care, aplicați decodurului, produc semnalul de ieșire  $\hat{s}(n)$  care este cel mai apropiat de semnalul vocal de intrare  $s(n)$ .

Nu toți parametri asociați codurului CELP se determină în buclă închisă, datorită complexității de calcul prea mari care ar rezulta. Astfel, parametrii predicției pe termen scurt se determină direct din analiza semnalului vocal de intrare, în buclă deschisă, iar pentru parametrii predicției pe termen lung se determină inițial în buclă deschisă o valoare aproximativă care apoi se ajustează în buclă închisă la valoarea exactă.

### 3.4.1 Codorul și decodorul CELP

Structurile de bază pentru un codor CELP, respectiv pentru un decodor CELP sunt prezentate în figurile 3.12 și respectiv 3.13.

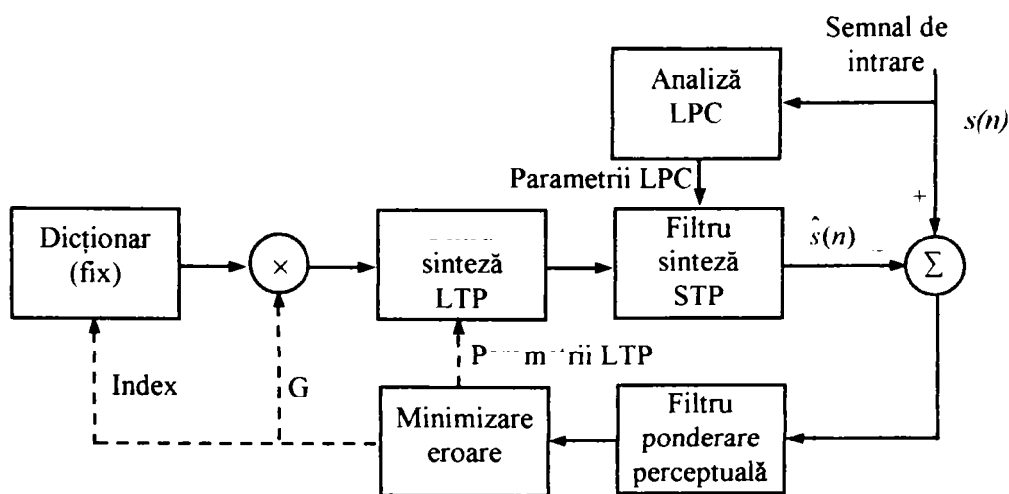


Fig. 3.12. Codorul CELP

Codorul este realizat în buclă închisă, respectând schema de principiu din fig. 3.11. Este evidențiat în schema codurului din fig. 3.12 faptul că semnalul de excitație și câștigul asociat, precum și parametrii predicției pe termen scurt sunt determinate în buclă închisă, iar

parametrii predicției pe termen scurt sunt determinați direct pe baza semnalului vocal de intrare, în buclă deschisă. Se poate recunoaște, inclus în structura codorului, decodorul realizat conform modelului de producere a semnalului vocal din fig. 3.10.

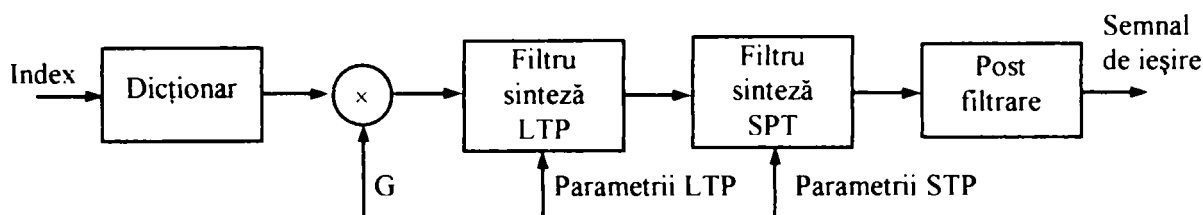


Fig. 3.13. Decodorul CELP

Criteriul folosit pentru compararea semnalului de ieșire din decodorul inclus  $\hat{s}(n)$  și semnalul vocal de intrare  $s(n)$  este eroarea medie pătratică. Ponderarea perceptuală a semnalului de eroare rezultat asigură o îmbunătățire a calității semnalului, prin înglobarea unor criterii perceptuale care țin cont de mecanismul de percepere a sunetului de către sistemul auditiv uman.

Decodorul CELP din fig. 3.13 corespunde schemei generale din fig. 3.10 fiind completat cu un filtru, denumit postfiltru, care asigură o filtrare suplimentară a semnalului de ieșire și care, în combinație cu filtrul de ponderare din codor, asigură o îmbunătățire a semnalului în tot domeniul spectral.

Mărimile de ieșire ale codorului (nefigurate explicit în schema din fig. 3.12) și respectiv cele de intrare ale codorului, sunt grupate în următoarele categorii:

- semnalul de excitație, determinat prin indexul din dicționar și prin câștigul  $G$ ,
- parametrii de predicție pe termen scurt (LPC sau STP),
- parametrii de predicție pe termen lung LTP.

În cele ce urmează, vor fi prezentate pe rând filtrul de ponderare perceptuală din codor, filtrul care asigură postfiltrarea semnalului de ieșire din decodor, precum și problemele legate de determinarea celor trei categorii de mărimi de ieșire ale codorului și cuantizarea acestora în vederea transmisiei spre decodor.

### 3.4.2 Ponderarea perceptuală a erorii

Ponderarea perceptuală a semnalului de eroare, realizată prin filtrul de ponderare perceptuală introdus în figura 3.12 exploatează particularitățile mecanismului de percepție a semnalului vocal de către auzul uman, care este afectat de fenomenul de mascare.

Prin mascare, un sunet mai intens poate, în anumite condiții, să *mascheze* un sunet mai slab care, în acest caz, nu va putea fi perceput ca un sunet distinct.

Mascarea se manifestă în mod *simultan*, în cazul în care cele două semnale se situează simultan în timp și semnalul mascat se află în interiorul unei anumite benzi de frecvență, denumită *bandă critică*, în jurul semnalului mai puternic [NYJ01], [SWK00].

Mascarea se manifestă de asemenea și în domeniul timp, prin fenomenele de *postmascare* și *premascare*, prin care un sunet puternic maschează un sunet mai slab și un anumit timp înainte de apariția sa (premascare) și respectiv după încetarea sa (postmascare) [MOR95], [NJY01].

Din acest punct de vedere, zgomotul de cuantizare va fi mai slab perceput în acele regiuni ale spectrului în care semnalul este intens, datorită mascării și va fi perceput mai accentuat în regiunile spectrului în care semnalul este mai slab. Ca urmare, este posibil ca în acele regiuni în care semnalul vocal este mai intens, și zgomotul să aibă o energie mai mare,

relativ la zgomotul corespunzător regiunilor cu semnal mai puțin intens, fără ca prin aceasta distorsiunea percepută să fie mai mare.

Pentru a ține seama de acest efect, se ponderează semnalul eroare de predicție cu o funcție de forma:

$$W(z) = \frac{A(z/\gamma_1)}{A(z/\gamma_2)}, \quad 0 < \gamma_2 < \gamma_1 \leq 1, \quad (3.57)$$

unde  $A(z)$  reprezintă filtrul de predicție dat de rel. (3.7), și care va atribui ponderi mai importante erorii în acele regiuni ale spectrului în care zgomotul este perceptibil și va atribui ponderi mai reduse erorii în acele regiuni ale spectrului în care zgomotul de cuantizare este mascat de semnalul vocal util de intensitate mai mare. Poziția filtrului de ponderare în codorul CELP este indicată în fig. 3.12.

Valorile  $\gamma_1$  și  $\gamma_2$  pot fi fixe sau adaptive. De cele mai multe ori se alege  $\gamma_1 = 1$  și în acest caz filtrul de ponderare devine:

$$W(z) = \frac{A(z)}{A(z/\gamma)}, \quad 0 < \gamma \leq 1. \quad (3.58)$$

Caracteristica filtrului perceptual din rel. (3.51) este ilustrată în figura 3.14, împreună cu spectrul LPC al semnalului de intrare, corespunzător filtrului de predicție  $A(z)$  și spectrului calculat cu ajutorul FFT.

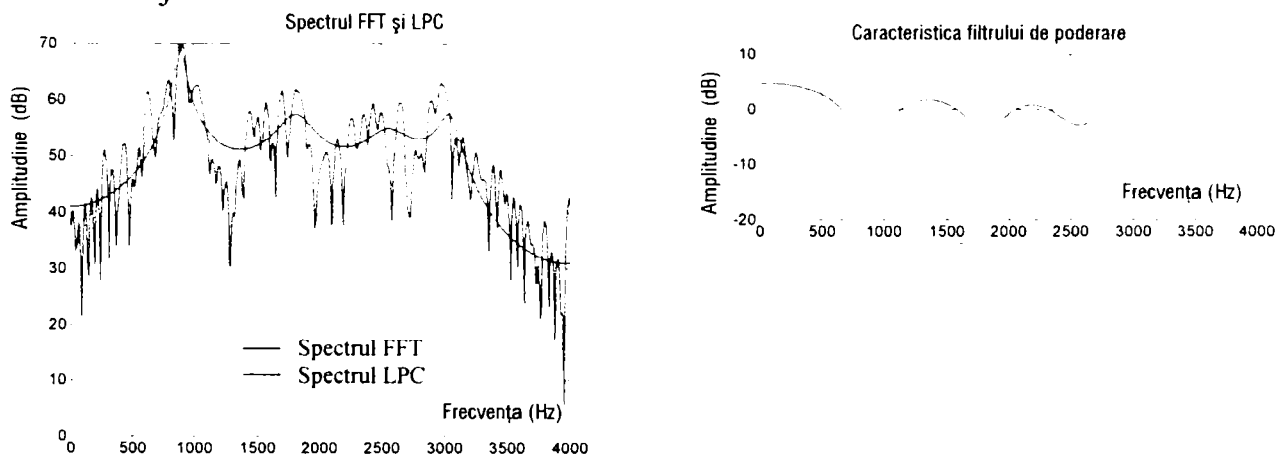


Fig. 3.14. Spectrul semnalului și caracteristica filtrului de ponderare

Se observă că filtrul de ponderare are o caracteristică cu vârfuri mai atenuate față de cea a filtrului  $A(z)$ , datorită deplasării polilor către centrul cercului unitate, proporțional cu factorul  $\gamma$ . În figura 3.15 este reprezentată caracteristica filtrului  $A(z)$  împreună cu caracteristica filtrului de ponderare perceptuală pentru diferite valori ale parametrului  $\gamma$  cuprinse între 0,5 și 0,95.

Gradul de modificare a spectrului este dependent de valoarea parametrului  $\gamma$ . Relația (3.58) arată că pentru  $\gamma \rightarrow 1$ ,  $W(z) \rightarrow 1$  și prin urmare spectrul semnalului de eroare nu este modificat. Pe de altă parte, pentru  $\gamma \rightarrow 0$ ,  $W(z) \rightarrow A(z)$ , adică eroarea este ponderată chiar cu caracteristica filtrului de sinteză. Valoarea optimă pentru  $\gamma$  se alege experimental, prin teste subiective.

O reducere a volumului de calcule necesar pentru ponderarea perceptuală se poate obține ținând cont de liniaritatea operației de filtrare, care poate fi astfel aplicată nu direct semnalului de eroare, ci ponderând cu aceeași funcție de transfer fiecare dintre cele două semnale care prin diferență dau semnalul de eroare, așa cum se ilustrează în figura 3.16, în care s-a reprezentat într-un singur bloc WSTP funcția de sinteză prin predicție liniară și ponderarea perceptuală.

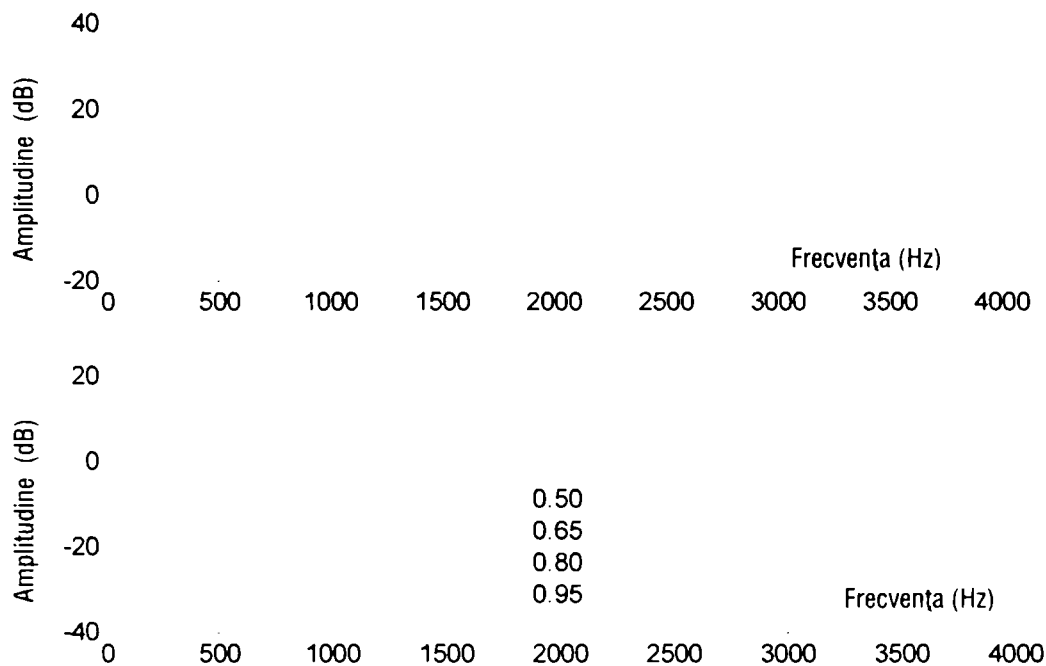


Fig. 3.15. Caracteristica de ponderare funcție de  $\gamma$

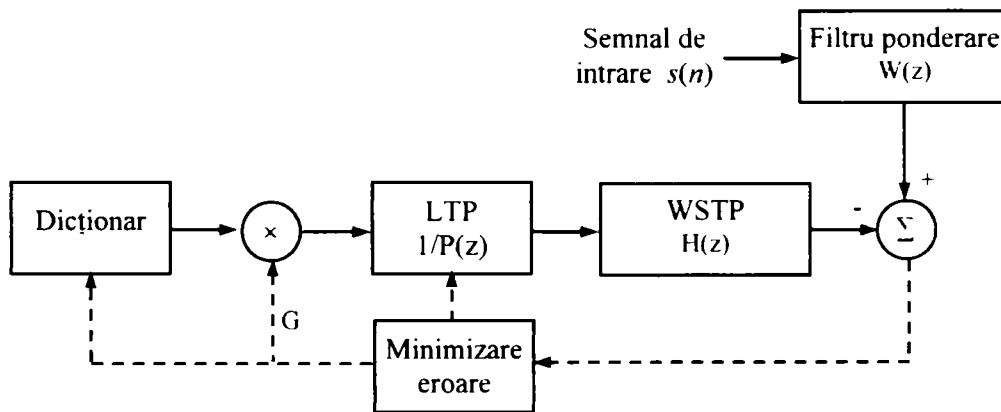


Fig. 3.16. Variantă de filtrare perceptuală

Funcția de transfer  $H(z)$  care rezultă pentru acest bloc este:

$$H(z) = \frac{1}{A(z)} W(z) = \frac{1}{A(z)} \frac{A(z)}{A(z/\gamma)} = \frac{1}{A(z/\gamma)} = \frac{1}{1 - \sum_{k=1}^p a_k \gamma^k z^{-k}} \quad (3.59)$$

Reducerea volumului de calcul rezultă din simplificarea termenului  $A(z)$  în expresia (3.59) și din observația că eroarea ponderată trebuie evaluată pe durata unui cadru de semnal pentru fiecare vector din dicționar, în timp ce, conform schemei din fig. 3.16, ponderarea semnalului de intrare se face o singură dată pentru un cadru de semnal [CHU03-2].

Având în vedere că dicționarul conține în general între 256 și 1024 vectori cu dimensiunea tipică de 40 – 60 eșantioane, reducerea volumului de calcule este semnificativă.

### 3.4.3 Postfiltrarea

Filtrarea perceptuală a semnalului de eroare, prezentată anterior, reduce componentele de zgomot în zonele de semnal cu energie mică, dar accentuează zgomotul în zonele cu energie mare, corespunzătoare formațiilor. Pentru a îmbunătăți calitatea semnalului după decodare, se poate introduce o operație de postfiltrare, cu rolul de a îmbunătăți spectrul



semnalului de ieșire. Răspunsul în frecvență al postfiltrului ideal trebuie să urmeze vârfurile și văile din înfășurătoarea spectrală a semnalului vocal, fără a da însă o atenuare globală a spectrului. În codarea LPC, răspunsul filtrului de sinteză urmărește înfășurătoarea spectrală a semnalului vocal, ceea ce poate conduce la soluția de a realiza postfiltrarea pe baza filtrului de sinteză LPC [SYH02].

• **Postfiltrarea pe termen scurt**

Se consideră filtrul obținut prin expandarea benzii filtrului de sinteză STP:

$$H_1(z) = \frac{1}{1 + \sum_{i=1}^M a_i \alpha^i z^{-i}}, \quad (3.60)$$

în care  $0 < \alpha < 1$  este o constantă [CHU03-2]. Figura 3.17 ilustrează răspunsul acestui filtru pentru diferite valori  $\alpha$ , calculat pe baza aceluiași filtru  $A(z)$  ca și cel din fig. 3.15.

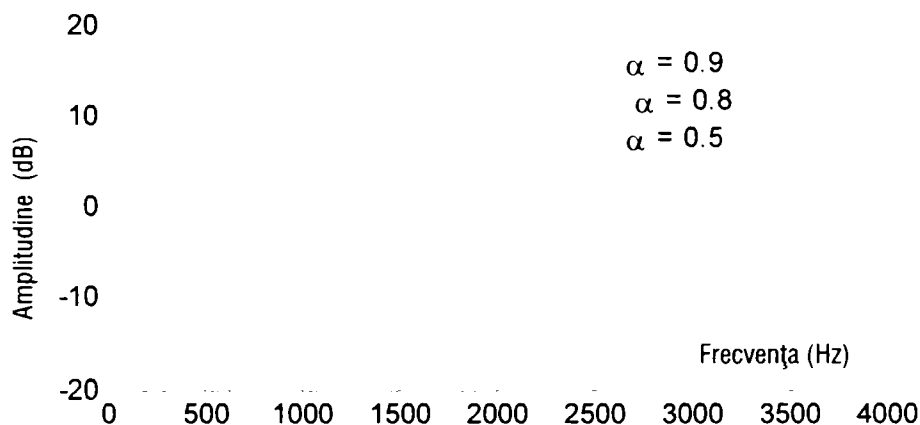


Fig. 3.17. Caracteristica de postfiltrare  $H_1$

Acest filtru va reduce nivelul zgomotului perceput, dar are ca efect negativ o estompare a sunetelor (efect de „surdină”), datorat caracteristicii care are o cădere continuă spre frecvențe înalte, de tip trece-jos.

Pentru a reduce această cădere spectrală, se pot adăuga filtrului (3.60) cu  $M$  poli un număr egal de zerouri, rezultând un filtru de forma:

$$H_2(z) = \frac{1 + \sum_{i=1}^M a_i \beta^i z^{-i}}{1 + \sum_{i=1}^M a_i \alpha^i z^{-i}}, \quad (3.61)$$

în care  $0 < \beta < \alpha < 1$  sunt constante.

Răspunsul în frecvență al filtrului  $H_2(z)$  pe o scară logaritmică poate fi privit ca diferență a două filtre de tip (3.60):

$$20 \log\left(\frac{1}{|1 + \sum_{i=1}^M a_i \alpha^i z^{-i}|}\right) - 20 \log\left(\frac{1}{|1 + \sum_{i=1}^M a_i \beta^i z^{-i}|}\right). \quad (3.62)$$

Valorile  $\alpha$  și  $\beta$  se determină tot pe cale experimentală, prin teste de ascultare.

Deși efectul trece-jos este redus, acest efect se manifestă încă în zona frecvențelor înalte, după cum se observă în figura 3.18, în care s-a reprezentat caracteristica filtrului  $H_1(z)$  pentru  $\alpha = 0,8$  împreună cu filtrul  $H_2(z)$  pentru  $\alpha = 0,8$  și  $\beta = 0,5$ .

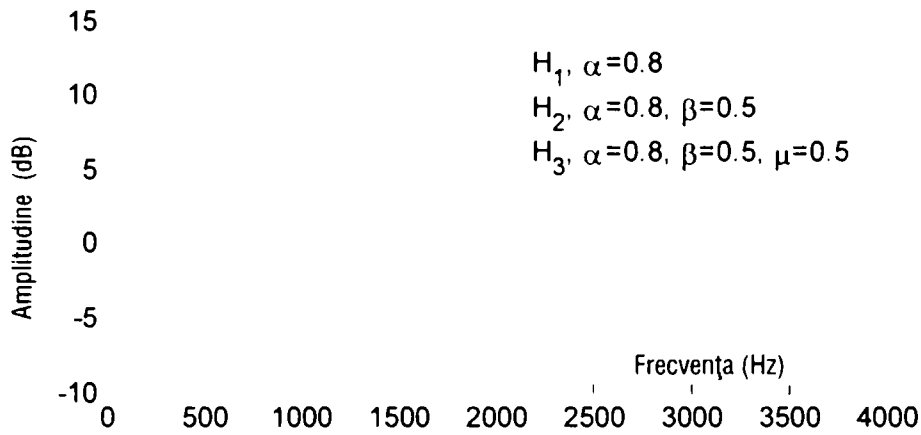


Fig. 3.18. Caracteristicile de postfiltrare  $H_1, H_2, H_3$

O reducere mai accentuată a efectului trece-jos se poate realiza adăugând încă un filtru de ordinul I, de forma  $(1 - \mu z^{-1})$ , similar celui folosit pentru preaccentuarea frecvențelor înalte din semnalul vocal înainte de codare, rezultând în final:

$$H_3(z) = (1 - \mu z^{-1}) \frac{1 + \sum_{i=1}^M a_i \beta^i z^{-i}}{1 + \sum_{i=1}^M a_i \alpha^i z^{-i}} \quad (3.63)$$

Pentru valoarea  $\mu = 0,5$  caracteristica filtrului  $H_3(z)$  este reprezentată, împreună cu  $H_1(z)$  și  $H_2(z)$  în fig. 3.18, în care se observă îmbunătățirea în domeniul frecvențelor înalte și păstrarea formei generale a caracteristicii în zonele formațiilor. Coeficientul  $\mu$  poate să fie fix, caz în care o valoare convenabilă, determinată experimental, este  $\mu = 0,5$ , sau poate fi adaptiv, având în vedere că partea căzătoare a spectrului la frecvențe înalte depinde de semnalul vocal [CHU03-2].

Astfel,  $\mu$  poate fi proporțional cu primul coeficient de reflexie  $k_1$ :

$$\mu = 0,5 k_1 \quad (3.64)$$

care poate fi obținut din coeficienții de predicție  $a_i$ . Pentru semnale sonore, care prezintă o corelație puternică,  $k_1$  este foarte apropiat de unitate, rezultând pentru  $\mu$  valoarea aproximativă 0,5. Pentru semnalele nesonore, care nu prezintă o astfel de corelație,  $k_1$  scade sau chiar devine negativ și frecvențele înalte sunt mai puțin accentuate [CHE95].

• **Controlul automat al câștigului**

Postfiltrarea introduce o alterare a nivelului general al semnalului, care trebuie neutralizată prin includerea, în structura postfiltrului, a unui circuit de control al nivelului semnalului, având schema bloc reprezentată în figura 3.19.

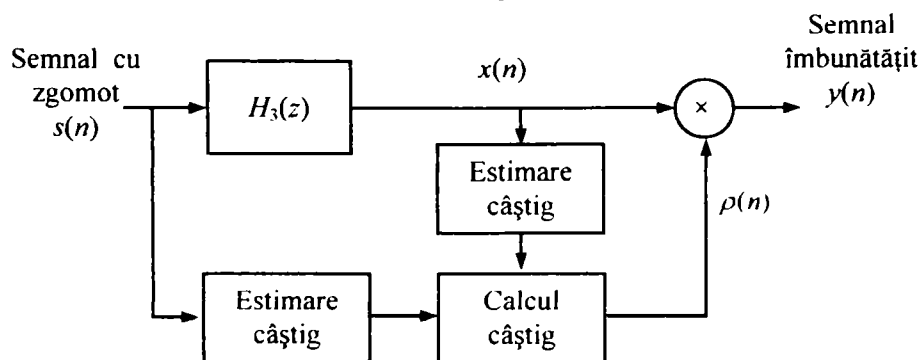


Fig. 3.19. Postfiltrarea adaptivă

Puterea semnalului vocal de intrare cu zgomot  $s(n)$  și a semnalului filtrat  $x(n)$  sunt estimate prin:

$$\sigma_s^2(n) = \zeta \sigma_s^2(n-1) + (1-\zeta)s^2(n), \quad (3.65)$$

și respectiv:

$$\sigma_x^2(n) = \zeta \sigma_x^2(n-1) + (1-\zeta)x^2(n). \quad (3.66)$$

O valoare  $\zeta = 0.99$  asigură un timp de adaptare acceptabil pentru o frecvență de eșantionare de 8 kHz. Pentru valori mai mici ale parametrului  $\zeta$  rezultă un timp de adaptare mai mic, dar semnalul este afectat de distorsiuni de amplitudine inacceptabil de mari.

Estimantul puterii se calculează pentru fiecare eșantion, după care se calculează, pentru fiecare eșantion, câștigul:

$$\rho(n) = \sqrt{\frac{\sigma_s^2(n)}{\sigma_x^2(n)}} \quad (3.67)$$

și, în final, eșantionul semnalului vocal de ieșire îmbunătățit:

$$y(n) = x(n)\rho(n). \quad (3.68)$$

Calculul presupune câte o împărțire și o extragere de radical pentru fiecare eșantion. O simplificare a volumului de operații rezultă dacă, în loc de (3.65) și (3.66) se calculează amplitudinea:

$$\sigma_s(n) = \zeta \sigma_s(n-1) + (1-\zeta)s(n) \quad (3.69)$$

$$\sigma_x(n) = \zeta \sigma_x(n-1) + (1-\zeta)x(n) \quad (3.70)$$

și câștigul corespunzător:

$$\rho(n) = \frac{\sigma_s(n)}{\sigma_x(n)}. \quad (3.71)$$

#### • Postfiltrarea pe termen lung

Postfiltrarea pe termen scurt descrisă anterior operează pe un număr redus de eșantioane, similar cu predicția pe termen scurt. O îmbunătățire a semnalului de ieșire se poate realiza prin postfiltrare pe termen lung, adăugând o componentă care realizează filtrarea pe termen lung, cu o caracteristică de forma:

$$H_l(z) = \frac{1}{1+b} (1 + bz^{-L}), \quad (3.72)$$

care asigură o atenuare a componentelor dintre vârfurile corespunzătoare armonicilor frecvenței fundamentale, îmbunătățind periodicitatea semnalului.

Parametrul  $b$ , cuprins între 0 și 1, stabilește atenuarea filtrului, iar factorul  $1/(1+b)$  asigură o normare a răspunsului astfel ca la vârf atenuarea să fie întotdeauna unitară, iar  $L$  reprezintă perioada corespunzătoare frecvenței fundamentale, exprimată în multipli întregi ai perioadei corespunzătoare frecvenței de eșantionare.

# Capitolul 4

## Cuantizarea vectorială

### 4.1 Cuantizarea

Operația de cuantizare este definită, conform dicționarului, ca operația prin care unei variabile continue (reprezentând o mărime fizică, ca de exemplu amplitudinea unui semnal electric) i se atribuie valori discrete (discontinue), care în multe cazuri sunt multipli întregi ai unei cantități convenabil aleasă. Primul exemplu de cuantizare menționat în literatura de specialitate aparține lui W. F. Shepard, care, în anul 1898, într-o lucrare citată în [GNF98], analizează problema rotunjirilor în legătură cu histogramele.

Pentru a defini un cuantizor scalar, se consideră un set de intervale sau *celule*,  $\mathcal{S} = \{S_i, i \in \mathcal{I}\}$ , împreună cu un set de puncte sau *nivele* de reproducere  $\mathcal{C} = \{y_i, i \in \mathcal{I}\}$ .

Cuantizorul  $q$  definit prin:

$$q(x) = y_i \text{ pentru } x \in S_i. \quad (4.1)$$

realizează operația exprimată prin relația:

$$q(x) = \sum_i y_i 1_{S_i}(x). \quad (4.2)$$

unde funcția indicator  $1_{S_i}(x)$  are valoarea 1 dacă  $x \in S_i$ , și 0 în caz contrar [GNF98].

În această definiție se presupune că  $\mathcal{S}$  este o partiție a axei reale, adică celulele  $S_i$  sunt disjuncte și reuniunea lor acoperă axa reală.

Calitatea unui cuantizor se poate aprecia cu ajutorul unei funcții de cost  $d$ , definind o măsură a distorsiunii  $d(x, \hat{x})$  între mărimea de intrare  $x$  și mărimea cuantizată  $\hat{x}$ .

Este de dorit ca distorsiunea de cuantizare să fie cât mai mică (neglijabilă) și pentru aceasta celulele  $S_i$  care definesc cuantizorul ar trebui să fie cât mai mici și mai multe. Această cerință conduce însă la creșterea numărului de biți necesari pentru a desemna valoarea de ieșire a cuantizorului și din acest motiv, nu se poate obține practic o valoare arbitrar de mică pentru distorsiune, datorită limitărilor impuse de capacitatea de stocare sau de capacitatea canalului de transmisie. Pentru a aprecia un cost al cuantizării din acest punct de vedere, se presupune că prin cuantizare, mărimea de intrare  $x$  se exprimă printr-un indice  $i$ , care arată care nivel de reproducere  $y_i$  se folosește pentru reprezentare. Dacă există  $N$  nivele posibile și se presupune că fiecare nivel se reprezintă prin același număr de biți, se poate defini rata cuantizorului, în biți pe eșantion de intrare, prin numărul de biți necesari pentru a reprezenta numărul  $N$  [BME99]:

$$r = \lceil \log_2 N \rceil. \quad (4.3)$$

În concluzie, scopul cuantizării este acela de a coda semnalul produs de o sursă cu o rată de bit cât mai mică, astfel încât semnalul reprodus să aproximeze cu distorsiuni cât mai mici semnalul original. Este evident că aceste două cerințe sunt contradictorii și în practică trebuie

să se accepte un compromis între cele două criterii de performanță, *distorsiunea* și *rata*, exprimat matematic printr-o *funcție distorsiune-rată*,  $\delta(R)$  funcție care exprimă cea mai mică distorsiune ce se poate atinge pentru o rată dată [GNF98], [PVT00], [KLO99], [GRA95], [BRG71].

Cuantizarea vectorială poate fi considerată ca o generalizare a cuantizării scalare, fiind o extindere a cuantizării scalare la cuantizarea unui vector  $k$ -dimensional, privit ca un set ordonat de numere reale. Cuantizorul vectorial este caracterizat de o partiție a spațiului de intrare  $k$ -dimensional și de un dicționar  $k$ -dimensional care conține vectori de cod de asemenea  $k$ -dimensionali.

Un vector poate fi cuantizat prin identificarea unui vector de cod dintr-o mulțime finită de astfel de vectori cuprinși într-un dicționar. Pentru un vector de intrare, cuantizorul generează adresa sau indicele vectorului de cod care se apropie cel mai mult de vectorul de intrare, în sensul unei măsuri a distorsiunii definite în mod corespunzător. Acest index este transmis, într-o formă adecvată, cel mai adesea binară, spre decodor, care extrage, dintr-o copie locală a dicționarului, vectorul de cod cu care se aproximează vectorul de intrare.

Trebuie precizat că dacă, în mod curent, operația de cuantizare scalară este asociată operației de conversie analog-numerică, prin cuantizare vectorială se înțelege cuantizarea unor valori numerice care corespund unei mărimi analogice deja convertite într-un format numeric.

## 4.2 Principiile cuantizării vectoriale

Un cuantizor vectorial VQ,  $k$  – dimensional și de mărime  $N$  [GER93], [CBA03-3] este o aplicație a unui vector sau a unui punct din spațiul euclidian  $k$ -dimensional  $\mathcal{R}^k$ , pe o mulțime finită  $\mathcal{C}$ , numită dicționar (codebook), și care conține  $N$  vectori, denumiți cuvinte de cod:

$$Q: \mathcal{R}^k \rightarrow \mathcal{C}, \quad \mathcal{C} \in \mathcal{R}^k, \quad (4.4)$$

$$\mathcal{C} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N) \quad \text{și} \quad \mathbf{y}_i \in \mathcal{R}^k \quad \text{pentru orice} \quad i \in \mathcal{I} = \{1, 2, 3, \dots, N\}, \quad (4.5)$$

unde  $\mathcal{I}$  este o mulțime de numere întregi, denumite index.

Cele  $N$  cuvinte de cod care compun dicționarul sunt fiecare vectori în spațiul  $\mathcal{R}^k$ .

Se definește *rezoluția* sau *rata* cuantizorului vectorial prin:

$$r = (\log_2 N) / k, \quad (4.6)$$

reprezentată de numărul de biți pe componentă a vectorului de cod.

Pentru o dimensiune  $k$  a vectorilor de cod dată, rezoluția depinde numai de numărul de cuvinte de cod din dicționar, adică de mărimea dicționarului, și nu de numărul de biți folosiți pentru a reprezenta valoarea numerică a vectorilor de cod din dicționar.

Cuantizorul vectorial realizează o partiție a spațiul  $\mathcal{R}^k$ , prin divizarea acestui spațiu în regiuni  $R_i$ , definite prin relația:

$$R_i = \{ \mathbf{x} \in \mathcal{R}^k : Q(\mathbf{x}) = \mathbf{y}_i \}, i \in \mathcal{I}, \quad (4.7)$$

și asociază fiecărei regiuni un cuvânt de cod din  $\mathcal{C}$ .

Un cuantizor vectorial poate fi descompus în două părți: un codor și un decodor, reprezentate schematic în figura 4.1.

Codorul  $\mathcal{E}$  realizează asocierea unei celule din spațiul  $\mathcal{R}^k$  cu un index  $\mathcal{I}$ , iar decodorul  $\mathcal{D}$  realizează funcția inversă:

$$\mathcal{E}: \mathcal{R}^k \rightarrow \mathcal{I} \quad \text{și} \quad \mathcal{D}: \mathcal{I} \rightarrow \mathcal{R}^k, \quad (4.8)$$

și, prin urmare, operația de cuantizare vectorială se poate reprezenta ca o succesiune a celor două operații de codare și decodare:

$$\hat{\mathbf{x}} = Q(\mathbf{x}) = D(\mathcal{E}(\mathbf{x})), \quad (4.9)$$

unde  $\hat{\mathbf{x}}$  este valoarea rezultată în urma operațiilor de codare și decodare succesivă a variabilei de intrare  $\mathbf{x}$ .

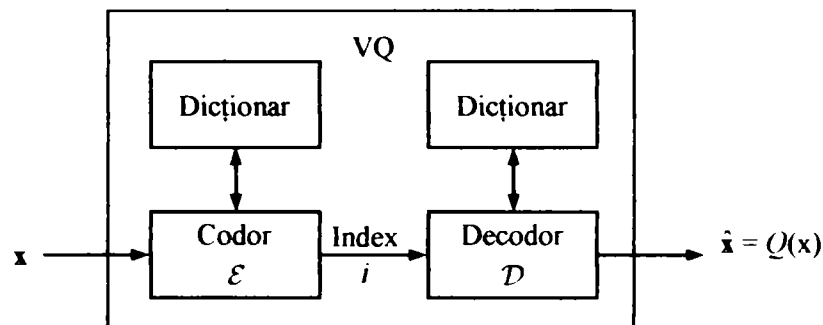


Fig. 4.1. Codorul și decodorul vectorial

După cum rezultă din figura 4.1, atât codorul cât și decodorul dispun fiecare de câte o copie a dicționarului.

Un cuantizor vectorial nu este doar o simplă generalizare a unui cuantizor scalar, ci este de fapt cea mai bună metodă posibilă de cuantizare [GER93], [GNF98].

Această afirmație este sprijinită de următoarea:

*Teoremă:* Pentru orice sistem de codare dat, care codifică un vector printr-un cuvânt binar selectat din  $N$  cuvinte posibile și reconstruiește din acest cuvânt binar un vector care aproximează vectorul de intrare, există un cuantizor vectorial cu un dicționar de dimensiunea  $N$  care are exact aceleași performanțe, adică pentru orice vector de intrare produce un vector de ieșire identic cu cel produs de sistemul de codare considerat.

O demonstrație intuitivă este dată în [GER93].

### 4.3 Proprietăți structurale ale cuantizorului vectorial

Intuitiv, funcția elementară a codorului dintr-un cuantizor vectorial constă în a analiza fiecare vector de intrare și a decide în care regiune a spațiului  $R^k$  se încadrează acesta, generând un index  $i$  care identifică această regiune. Funcția îndeplinită de decodor este aceea de a genera vectorul  $\mathbf{y}_i$  pe baza acestui indice. Structura primară a unui cuantizor vectorial, rezultată din această reprezentare intuitivă este indicată în figura 4.2 [GER93].

Modelarea codorului vectorial se poate face cu ajutorul unei funcții selector  $S_i(\mathbf{x})$ , definită prin relația:

$$S_i(\mathbf{x}) = \begin{cases} 1 & \text{pentru } \mathbf{x} \in R_i \\ 0 & \text{pentru } \mathbf{x} \notin R_i, \end{cases} \quad (4.10)$$

și care este un indicator al apartenenței unui vector la o regiune  $R_i$  a spațiului  $R^k$ .

Operația de decodare, inversă codării vectoriale, poate fi reprezentată prin relația:

$$\hat{\mathbf{x}} = Q(\mathbf{x}) = \sum_{i=1}^N \mathbf{y}_i S_i(\mathbf{x}), \quad (4.11)$$

în care, pentru orice valoare a vectorului de intrare  $\mathbf{x}$ , doar un termen din sumă este diferit de zero, deoarece numai una din valorile  $S_i(\mathbf{x})$  este diferită de zero. În figura 4.2 operația de multiplicare din relația (4.11) poate fi interpretată ca făcând parte din dicționar, care memorează valoarea cuvântului de cod  $\mathbf{y}_i$ .



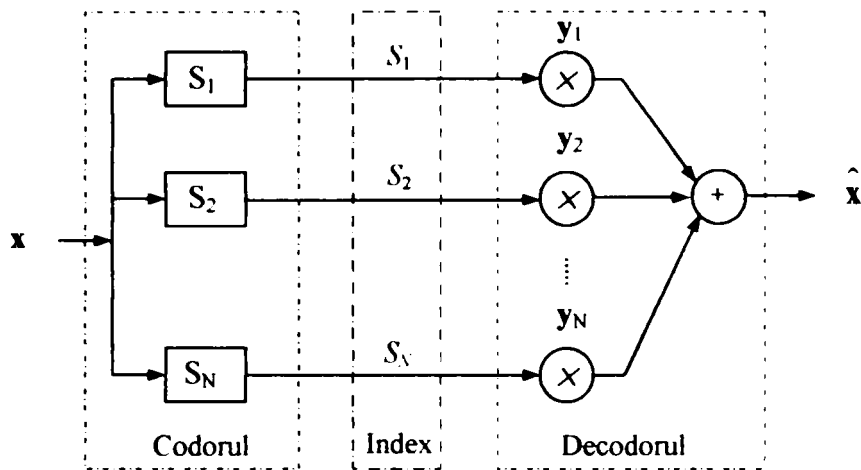


Fig. 4.2. Structura primară a cuantizorului vectorial

Codorul și decodorul pot fi modelate separat, folosind pentru fiecare o structură primară corespunzătoare, rezultată din schema generală din figura 4.2.

Astfel,  $A$  reprezintă generatorul de index, care realizează funcția:

$$A: \mathcal{B}_N \rightarrow \mathcal{I}, \tag{4.12}$$

și  $A^{-1}$  reprezintă decodorul de index, care realizează funcția inversă:

$$A^{-1}: \mathcal{I} \rightarrow \mathcal{B}_N, \tag{4.13}$$

unde  $\mathcal{B}_N$  este un cuvânt binar de forma  $b=(b_1, b_2, \dots, b_N)$ , cu elementele  $b_n \in \{0,1\}$  și cu restricția ca  $b$  să aibă exact o singură componentă cu valoarea diferită de zero.

#### 4.4 Performanțele cuantizorului vectorial

Performanțele unui cuantizor vectorial se pot estima prin măsura distorsiunii  $d$ , definită ca o funcție de cost, nenegativă  $d(\mathbf{x}, \hat{\mathbf{x}})$ , asociată vectorului de intrare  $\mathbf{x}$  și vectorului reconstituit prin cuantizare  $\hat{\mathbf{x}}$ . Definind astfel distorsiunea, se poate aprecia performanța generală a unui cuantizor vectorial prin distorsiunea medie între semnalul de intrare și cel estimat [GER93]:

$$D = E \{d(\mathbf{x}, \hat{\mathbf{x}})\} \tag{4.14}$$

Este evident de dorit ca distorsiunea medie să fie cât mai mică.

Funcția de cost nenegativă,  $d(\mathbf{x}, \hat{\mathbf{x}})$  asociată vectorului de intrare  $\mathbf{x}$  și vectorului reconstituit  $\hat{\mathbf{x}}$ , care definește eroarea sau distorsiunea de cuantizare și permite aprecierea performanțelor cuantizorului, poate avea diferite forme, dintre care câteva cazuri particulare, care prezintă interes practic sunt următoarele [GER93], [RMG90], [ZBL94], [VLL00], [VKB03], [HCH02]:

*eroarea medie pătratică sau distanța euclidiană*, definită prin relația:

$$d(\mathbf{x}, \hat{\mathbf{x}}) = \|\mathbf{x} - \hat{\mathbf{x}}\|^2 = \sum_{i=1}^k (x_i - \hat{x}_i)^2. \tag{4.15}$$

*distorsiunea* definită de o funcție de tip:

$$d(\mathbf{x}, \hat{\mathbf{x}}) = \sum_{i=1}^k d_m(x_i, \hat{x}_i). \tag{4.16}$$

în care  $d_m(x, \hat{x})$  este o funcție aditivă, în cele mai multe cazuri de forma:

$$d_m(x, \hat{x}) = |x - \hat{x}|^m, \quad m > 0. \quad (4.17)$$

eroarea pătratică ponderată:  $d(\mathbf{x}, \hat{\mathbf{x}}) = (\mathbf{x} - \hat{\mathbf{x}})^T \mathbf{W}(\mathbf{x} - \hat{\mathbf{x}})$

$$d(\mathbf{x}, \hat{\mathbf{x}}) = (\mathbf{x} - \hat{\mathbf{x}})^T \mathbf{W}(\mathbf{x} - \hat{\mathbf{x}}), \quad (4.18)$$

unde  $\mathbf{W}$  este o matrice pătrată de ponderare, simetrică și pozitiv definită. Un caz particular este acela în care matricea  $\mathbf{W}$  este o matrice diagonală, cu elementele  $w_{ii} > 0$ :

$$d(\mathbf{x}, \hat{\mathbf{x}}) = \sum_{i=1}^k w_{ii} (x_i - \hat{x}_i)^2, \quad (4.19)$$

care permite atribuirea de ponderi diferite diverselor componente ale vectorului. Eroarea ponderată poate fi interpretată ca o eroare neponderată între vectorii  $\mathbf{x}' = \mathbf{A}\mathbf{x}$  și  $\hat{\mathbf{x}}' = \mathbf{A}\hat{\mathbf{x}}$ , unde matricea  $\mathbf{A}$  se obține prin factorizarea  $\mathbf{W} = \mathbf{A}^T \mathbf{A}$ .

Măsurile  $d(\mathbf{x}, \mathbf{y})$  ale distorsiunii prezentate sunt simetrice față de argumentele  $\mathbf{x}$  și  $\mathbf{y}$ . În unele cazuri, se poate alege o matrice  $\mathbf{W}$  nesimetrică, ca de exemplu:

$$\mathbf{W}(\mathbf{x}) = \|\mathbf{x}\|^{-2} \mathbf{I}, \quad (4.20)$$

cu  $\mathbf{I}$  matricea identitate (unitate) și în acest caz distorsiunea (4.19) devine:

$$d(\mathbf{x}, \hat{\mathbf{x}}) = \frac{\|\mathbf{x} - \hat{\mathbf{x}}\|^2}{\|\mathbf{x}\|^2}, \quad (4.21)$$

care atribuie o valoare cu atât mai mare distorsiunii, cu cât semnalul este mai mic [GER93].

eroarea maximă sau norma  $l_\infty$ , definită prin:

$$d_{\max}(\mathbf{x}, \hat{\mathbf{x}}) = \max_i |x_i - \hat{x}_i|. \quad (4.22)$$

## 4.5 Cuantizarea vectorială după regula celui mai apropiat vecin

Cuantizarea vectorială după regula „celui mai apropiat vecin” (nearest neighbour - NN), se caracterizează prin aceea că partiția spațiului de intrare este complet determinată prin cunoașterea dicționarului și a măsurii pentru distorsiune. Acest cuantizor se mai numește cuantizor Voronoi și o reprezentare grafică intuitivă pentru cazul bidimensional ( $k = 2$ ) este ilustrată în figura 4.3. Pentru un dicționar dat, cuantizorul NN este un cuantizor local optimal în sensul minimizării distorsiunii medii [GER93].

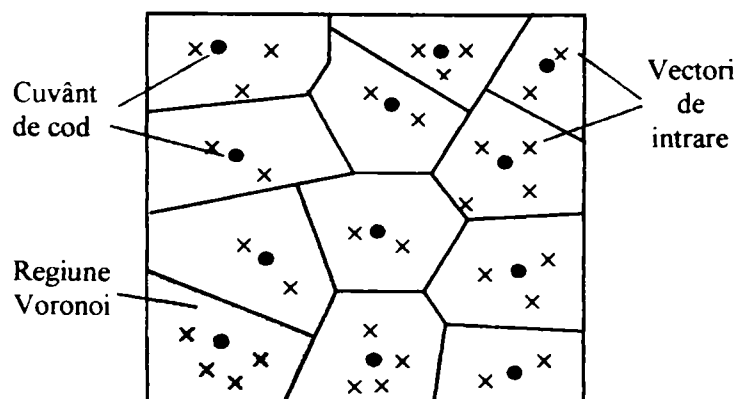


Fig. 4.3. Cuantizarea vectorială NN

Algoritmul de codare pentru cuantizorul NN se poate rezuma astfel:

*Pasul 1. Se fac inițializările:  $d = d_0, j = 1$  și  $i = 1$ .*

*Pasul 2. Se calculează  $D_j = d(x, y_j)$ .*

*Pasul 3. Dacă  $D_j = d$ , se pune  $D_j \rightarrow d$  și  $j \rightarrow i$ .*

*Pasul 4. Dacă  $j = N$  se pune  $j = 1 \rightarrow j$  și se revine la pasul 2.*

*Pasul 5. Dacă  $j = N$ , stop. Rezultatul este indexul  $i$ .*

## 4.6 Condiții de optimalitate pentru cuantizorul vectorial

Proiectarea unui cuantizor vectorial presupune:

- determinarea unui dicționar (care definește de asemenea și decodorul),
- determinarea unei partiții a spațiului de intrare, sau o regulă de codare, care definește decodorul,

astfel încât să se obțină un maxim pentru o măsură a performanței definită conform §4.4.

Pentru a determina condiția necesară pentru ca VQ să fie optimal în sensul minimizării distorsiunii medii  $d$ , se va presupune că dimensiunea  $N$  a dicționarului este dată, că distribuția statistică a vectorilor de intrare  $\mathbf{x}$  este cunoscută și de asemenea că este stabilită o măsură particulară pentru distorsiune  $d(\cdot, \cdot)$ . Condiția de optimalitate se va stabili pe rând, găsind mai întâi condiția ca un codor să fie optimal pentru un decodor dat și apoi condiția ca un decodor să fie optimal pentru un codor dat [CBA03-3].

### 4.6.1 Condiția de codor optim pentru un decodor dat

Fiind dat decodorul, înseamnă că este cunoscut dicționarul  $C = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\}$  prin care acesta este definit. Partiția optimă a spațiului de intrare, care definește codorul, va fi aceea care satisface:

*Condiția de cel mai apropiat vecin (condiția NN - nearest neighbor):*

$$R_i \subset \{\mathbf{x} : d(\mathbf{x}, \mathbf{y}_i) \leq d(\mathbf{x}, \mathbf{y}_j), \forall j\} \quad (4.23)$$

adică:

$$Q(\mathbf{x}) = \mathbf{y}_i \text{ dacă și numai dacă } d(\mathbf{x}, \mathbf{y}_i) \leq d(\mathbf{x}, \mathbf{y}_j), \forall j \quad (4.24)$$

și prin urmare:

$$d(\mathbf{x}, Q(\mathbf{x})) = \min_{\mathbf{y}_i \in C} d(\mathbf{x}, \mathbf{y}_i) \quad (4.25)$$

Condiția NN impune ca dacă  $\mathbf{x}$  are ca cel mai apropiat unic vecin pe  $\mathbf{y}_i$ , atunci  $\mathbf{x}$  se va asocia cu partiția  $R_i$  și  $i$  se va atribui valoarea  $\mathbf{y}_i$ . Dacă însă  $\mathbf{x}$  are distanța egală față de doi sau mai mulți vectori de cod, atunci  $\mathbf{x}$  nu se poate coda în mod unic și se va adopta convenția de a se atribui lui  $\mathbf{x}$  vectorul de cod cu cel mai mic indice.

### 4.6.2 Condiția de decodor optim pentru un codor dat

În continuare se va analiza în ce caz se obține un dicționar optim dacă este dată partiția spațiului de intrare. În acest caz vectorul de cod asociat cu fiecare partiție a spațiului conduce la noțiunea de centroid.

Se definește centroidul  $cent(R)$ , pentru orice set  $R \in \mathcal{R}^k$ , ca fiind acel vector  $\mathbf{y}$  care minimizează distorsiunea medie dintre vectorii  $\mathbf{X}$  din  $R$  și  $\mathbf{y}$ :

$$\mathbf{y}^* = \text{cent}(R) \text{ dacă } E\{d(\mathbf{X}, \mathbf{y}^*) | \mathbf{X} \in R\} \leq E\{d(\mathbf{X}, \mathbf{y}) | \mathbf{X} \in R\}, \forall \mathbf{y} \in \mathcal{R}^k. \quad (4.26)$$

Prin această definiție, centroidul este asimilat unei reprezentări naturale a „centrului” regiunii  $R$ , cu distribuția de probabilitate asociată. De remarcat că pentru unele măsuri ale distorsiunii  $d(\cdot, \cdot)$  este posibil ca centroidul să nu fie unic determinat [GER93]. Dacă regiunea  $R$  are probabilitatea zero, centroidul acestei regiuni se poate defini în mod arbitrar.

Pentru cazul în care ca măsură a distorsiunii se folosește eroarea medie pătratică, centroidul exprimat prin relația (4.26) devine:

$$\text{cent}(R) = E\{\mathbf{X} | \mathbf{X} \in R\} \quad (4.27)$$

regăsindu-se noțiunea de centroid din mecanică.

Având definite condiția de cel mai apropiat vecin NN și noțiunea de centroid, se poate afirma că un dicționar este optim pentru o partiție dată [GER93], [CBA03-3], dacă îndeplinește:

*Condiția de centroid:*

Pentru partiția spațiului de intrare:

$$\{R_i; i = 1, 2, \dots, N\}, \quad (4.28)$$

vectorii de cod optimali sunt:

$$\mathbf{y}_i = \text{cent}(R_i) \quad (4.29)$$

### 4.6.3 Condiția de margine cu probabilitatea zero

Condițiile de cel mai apropiat vecin și de centroid sunt condiții necesare pentru ca un codor vectorial să fie optimal, dar nu sunt și suficiente. În cele mai multe cazuri, aceste condiții sunt completate de o a treia condiție și anume condiția de margini cu probabilitate zero. Pentru a introduce această condiție, se presupune un cuantizor vectorial care satisface condițiile de cel mai apropiat vecin și de centroid.

Fie  $\{\mathbf{y}_i\}$  setul de vectori care constituie dicționarul,  $\{R_j\}$  celulele în care este partiționat spațiul de intrare și  $\{R_j^i\}$  definit prin:

$$R_j^i = \{\mathbf{x} : d(\mathbf{x}, \mathbf{y}_j) \leq d(\mathbf{x}, \mathbf{y}_i), \text{ pentru orice } i \in \mathcal{I}\} \quad (4.30)$$

$R_j^i$  reprezintă, conform acestei definiții, celulele  $R_j$  completate cu punctele limită care le mărginesc. Este evident că  $R_j \subset R_j^i$ .

Denumim setul de puncte care aparțin  $R_j^i$ , dar nu aparțin  $R_j$  ca fiind marginile setului  $R_j$  (boundary):

$$B_j = R_j^i - R_j \quad (4.31)$$

Prin urmare, marginile constau din puncte care nu aparțin regiunii  $R_j$ , dar sunt egal depărtate de  $\mathbf{y}_j$  și de un alt punct  $\mathbf{y}_i$  oarecare și, prin urmare, nu au un unic cel mai apropiat vecin și deci condiția de cel mai apropiat vecin nu se poate aplica. Rezultă o condiție suplimentară față de cele enunțate, necesară pentru ca un codor vectorial să fie optim pentru o sursă cu o distribuție dată și anume:

*Condiția de margini cu probabilitate zero:*

$$P\left(\bigcup_{j=1}^N B_j\right) = 0, \quad (4.32)$$

adică punctele de margine să aibă probabilitatea de apariție zero:

$$P(x : d(x, y_i) = d(x, y_j)) = 0, \text{ pentru orice } i \neq j \quad (4.33)$$

Cu acesta, se poate afirma că un cuantizor vectorial care satisface cele trei condiții necesare, enunțate anterior, este un cuantizor vectorial optim [GER93].

Noțiunea de optim se poate defini mai precis, având în vedere că distorsiunea  $D$ , care trebuie minimizată, este o funcție atât de partițiile spațiului de intrare cât și de dicționar. Deoarece condiția de cel mai apropiat vecin este satisfăcută, fiind necesară pentru optimalitate, se presupune că această condiție este îndeplinită pentru orice dicționar  $C$ . Astfel, distorsiunea  $D$ , ca măsură a performanței, va depinde doar de dicționar, devenind o funcție de  $kN$  variabile reale, iar noțiunea de codor optimal se poate defini pe baza acestei funcții.

Un codor vectorial este *local optimal*, dacă orice perturbație a vectorilor de cod nu va afecta distorsiunea  $D$  (se presupune că o astfel de perturbație va modifica corespunzător și partițiile construite pe baza condiției de cel mai apropiat vecin).

Un codor vectorial este *global optimal* dacă nu există un alt dicționar, care să conducă la o distorsiune  $D$  mai mică.

Nu există o metodă teoretică de a trata problema unui codor global optimal, după cum nu există o metodă teoretică de a construi un codor care să garanteze optimalitatea în sens global [GER93]. Un codor care satisface cele trei condiții impuse este însă, în cele mai multe cazuri, optimal în sens local. În unele cazuri particulare însă, un codor optimal în sens local poate fi foarte departe de un codor optimal în sens global [GER93].

## 4.7 Proiectarea unui cuantizor vectorial

Cele trei condiții necesare pentru ca un codor vectorial să fie optimal constituie punctul de plecare pentru a îmbunătăți un cuantizor dat, printr-o procedură iterativă.

Iterația începe cu un cuantizor dat, definit prin dicționarul său și prin partiția corespunzătoare a spațiului de intrare, optimală pentru dicționarul respectiv, care respectă principiul celui mai apropiat vecin. Pentru această partiție, se caută un nou dicționar, construit după principiul centroidului, care va fi optimal pentru partiția dată. Acest nou dicționar, împreună cu noua partiție pe care o determină, conform principiului celui mai apropiat vecin, constituie un nou codor vectorial, cu o distorsiune care este mai mică sau cel mult egală cu cea asigurată de cuantizorul inițial.

Deși nu întotdeauna se poate ajunge la un optim global prin această metodă care încearcă îmbunătățirea dicționarului cu ajutorul partiției și apoi a partiției cu ajutorul dicționarului, totuși, aplicarea repetată a pașilor de îmbunătățire a cuantizorului, prezentați mai sus, constituie un algoritm iterativ, care reduce sau, în cel mai defavorabil caz, lasă nemodificată distorsiunea cuantizorului, adică îl îmbunătățește sau îl lasă nemodificat. Ca rezultat al acestui algoritm, se obțin o serie de cuantizoare cu performanțe de distorsiune din ce în ce mai bune, care asigură cel puțin un optim local.

### 4.7.1 Dicționarul inițial

Cuantizorul inițial, care va fi îmbunătățit conform algoritmului anterior, este dat de obicei prin dicționarul său  $C$  și, din acest punct de vedere, se poate considera că acesta este un dicționar inițial. Problema stabilirii acestui dicționar inițial este esențială în proiectarea unui cuantizor vectorial. Se prezintă în continuare câteva metode de construcție a dicționarului inițial.

### **Codarea aleatoare**

Codarea aleatoare ar putea fi cea mai simplă cale de a obține dicționarul inițial. Pentru aceasta, dicționarul inițial este format din  $N$  vectori care sunt aleși aleator, respectând distribuția sursei ce urmează a fi codată. Dacă această distribuție nu se respectă, atunci îmbunătățirea prin metoda iterativă prezentată anterior nu va avea succes, deoarece probabil vor fi unii vectori de cod din dicționarul inițial cărora nu li se vor asocia niciodată vectori de intrare și cuantizorul nu va fi optimal nici măcar în sens local.

Cei  $N$  vectori aleatori pot fi luați chiar din semnalul ce se va cuantiza sau dintr-un semnal similar, care constituie semnalul de antrenare, fiind primii  $N$  vectori consecutivi din acest semnal. În cazul în care semnalul este puternic corelat, se pot alege  $N$  vectori care să nu fie consecutivi în semnalul de intrare, de exemplu vectori situați la distanța  $K$  unul de altul, unde  $K$  este un număr natural fixat. Trebuie precizat că termenul de codare aleatoare se referă exclusiv la alegerea dicționarului inițial, dar odată ce acesta a fost stabilit, codarea se face deterministic, în sensul celui mai apropiat vecin.

### **Codarea prin eliminare (Pruning)**

Termenul pruning semnifică curățarea ramurilor uscate și în acest sens, în metoda codării prin eliminare, toți vectorii din setul de vectori de antrenare reprezintă candidați la vectorii de cod care vor constitui dicționarul și din acest set se elimină vectori pe rând, în mod selectiv, până se ajunge la numărul stabilit de vectori de cod în dicționar.

Se începe cu primul vector din setul de antrenare, care se consideră ca fiind primul vector din dicționar. În continuare, se calculează distorsiunea între acest vector de cod și următorul vector din setul de antrenare; dacă distorsiunea este mai mică decât o valoare impusă, se continuă cu următorul vector din setul de antrenare, iar dacă distorsiunea depășește valoarea impusă, se adaugă acest vector la dicționar ca un nou cuvânt de cod. Este posibil ca pentru anumite seturi de vectori de antrenare, dicționarul rezultat să aibă un număr mai mic de vectori decât cel dorit. În acest caz, se reia procedura după ce se reduce cu o cantitate corespunzătoare valoarea impusă pentru distorsiune.

### **Gruparea (PNN – pairwise NN)**

Gruparea PNN este asemănătoare cu metoda pruning, folosind, în mod similar, întreg setul de vectori de antrenare și ajungând la dicționarul de mărime  $N$ , dar, spre deosebire de metoda anterioară, vectorii de cod pot fi diferiți de vectorii din setul de antrenare.

Se presupune că secvența de antrenare conține  $L$  vectori; fiecare dintre aceștia este considerat ca formând o regiune (cluster) separată, care conține un singur vector. Scopul final este de a grupa mai mulți vectori într-un cluster, sau grup, până se ajunge la numărul dorit de grupuri  $N$ . Dicționarul va conține în final centroizii acestor cluster, iar spațiul de intrare este partiționat în numărul  $N$  dorit de partiții sau celule. Este posibil însă ca aceste partiții să nu respecte condiția de cel mai apropiat vecin, dar în final aceste partiții se pot înlocui cu unele care respectă această condiție.

Se începe prin a calcula distorsiunea între toate perechile de vectori din setul de antrenament. Cei doi vectori care au cea mai mică distorsiune se asociază într-un cluster, care se reprezintă prin centroidul său. Rezultă  $L - 1$  cluster, unul format din cei doi vectori menționați, iar celelalte formate fiecare din câte un singur vector. În continuare, la fiecare etapă câte un vector se asociază unui cluster și numărul acestora scade.

Se presupune că la un moment dat există  $K$  cluster, unde  $N < K \leq L - 1$  și se urmărește asocierea a două cluster, pentru a obține  $K - 1$  cluster. În acest scop, pentru fiecare set de două cluster din cele  $K$  se calculează cu cât ar crește distorsiunea dacă cele două cluster s-ar uni și centroizii lor s-ar înlocui cu un singur centroid, calculat corespunzător și se va face



această grupare pentru acele două cluster care au cel mai mic efect asupra creșterii distorsiunii.

### Codul de produs

Un cod de produs se poate folosi ca punct de plecare pentru un dicționar inițial. Dacă  $q(x)$  este un cuantizor scalar, atunci produsul cartezian a  $k$  cuantizori scalari:

$$Q(x_0, x_1, \dots, x_{k-1}) = (q(x_0), q(x_1), \dots, q(x_{k-1})) \quad (4.34)$$

este un cuantizor vectorial. Un dicționar de mărime  $N = 2^{kR}$ , cu  $R$  (rezoluția) exprimată printr-un număr întreg, rezultă ca produs a  $k$  cuantizori scalari, cu  $2^R$  cuvinte de cod fiecare.

### Divizarea (Splitting)

Divizarea (splitting) este o tehnică de dezvoltare a unui dicționar din aproape în aproape, până se ajunge la dimensiunea dorită. Construcția dicționarului se face pornind de la un dicționar de un singur cuvânt, care este chiar centroidul întregului set de antrenare. Fie  $y_0$  acest cuvânt de cod. Pornind de la acest cuvânt, se construiesc două noi cuvinte de cod:  $y_0$  și  $y_0 + \varepsilon$ , unde  $\varepsilon$  este un vector cu o normă euclidiană suficient de mică. Mărimea  $\varepsilon$  se poate alege în mai multe moduri, una din posibilități fiind aceea de a stabili  $\varepsilon$  proporțional cu un vector a cărui componentă  $i$  este egală cu deviația standard a vectorului  $i$  din setul de antrenare. În acest mod, se poate trece de la un cuantizor cu  $r$  cuvinte de cod, la unul cu  $r + 1$  cuvinte, până se ajunge la dimensiunea dorită.

Obținerea unui cuantizor vectorial optim presupune existența unui set de vectori de antrenare,  $T$ , care conține un număr suficient de mare de vectori cu o distribuție apropiată de a vectorilor care urmează a fi codati.

### 4.7.2 Algoritmul Lloyd

Algoritmul Lloyd este bazat pe ideile expuse privind îmbunătățirea unui cuantizor vectorial prin modificarea succesivă a dicționarului și a partiției. Pentru început se presupune cunoscută funcția de distribuție de probabilitate *pdf* a sursei.

În acest caz, se poate enunța:

*Iterația Lloyd pentru îmbunătățirea unui dicționar dat:*

- a) Fiind dat un dicționar

$$C_m = \{y_i; i = 1, 2, \dots, N\}, \quad (4.35)$$

se calculează partiția optimală, folosind regula celui mai apropiat vecin:

$$R_i = \{x: d(x, y_i) < d(x, y_j)\}; \text{ pentru orice } i \neq j \quad (4.36)$$

- b) Cu ajutorul regulii centroidului, se calculează

$$C_{m+1} = \{cent(R_i); i = 1, 2, \dots, N\} \quad (4.37)$$

care constituie un dicționar optimal pentru partiția calculată.

Fiecare aplicare a iterației Lloyd reduce sau păstrează nemodificată distorsiunea totală [GER93].

Nu se garantează însă obținerea unui cuantizor optimal și nici măcar convergența spre un cuantizor local optimal. Este posibil ca în anumite situații să apară celule goale, dacă dicționarul  $C_m$  nu este bine ales și în acest caz, în etapa b) nu se poate calcula centroidul.

### 4.7.3 Algoritmul Lloyd generalizat (GLA sau LBG) pentru date empirice

Aplicarea iterației Lloyd în forma enunțată presupune cunoașterea funcției densitate de probabilitate a sursei și geometria celulelor, pentru a putea calcula centroidul, conform relației (4.26). Evaluarea unei integrale multiple pe o regiune de formă oarecare este dificilă dacă nu chiar imposibilă în formă analitică. Mai mult, în unele situații practice nu este disponibilă o descriere analitică a funcției densitate de probabilitate a sursei. În locul acesteia se poate totuși folosi, cu bune rezultate, un set de vectori de intrare obținut dintr-o observare empirică a sursei ce urmează a se coda [GER93], [CRL92], [JLP99], [CBA04-1].

Se presupune că este disponibil un set de  $M$  vectori obținut dintr-o observare a semnalului ce se va cuantiza, numit set de antrenare:

$$\mathcal{T} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_M\}, \quad (4.38)$$

Se definește raportul de antrenare prin:

$$\beta = M / N, \quad (4.39)$$

unde  $N$  este dimensiunea dicționarului.

Parametrul  $\beta$  arată în ce măsură setul de antrenare poate descrie funcția de distribuție de probabilitate a sursei ce se va coda.

Pe baza setului de antrenare, se definește un vector aleator  $\mathbf{U}$ , atribuind funcției de probabilitate de masă valoarea  $1/M$  pentru fiecare valoare  $\mathbf{v}_i$  din setul de antrenare  $\mathcal{T}$ . Deși pentru o variabilă discretă nu există *pdf*, pentru variabila  $\mathbf{U}$  există și este definită funcția de distribuție cumulativă. Funcția de distribuție cumulativă *cdf*, empirică, pentru  $\mathbf{U}$  este [GER93]:

$$F^{(M)}(\mathbf{x}) = \frac{1}{M} \sum_{i=1}^M \hat{u}(\mathbf{x} - \mathbf{v}_i), \quad (4.40)$$

unde  $\hat{u}(\mathbf{w})$  este vectorul treaptă unitate, care are valoarea unu dacă toate componentele vectorului  $\mathbf{w}$  sunt nenegative și valoarea zero în caz contrar.

Legea numerelor mari asigură convergența funcției de distribuție cumulativă empirică a setului de antrenare către funcția distribuție cumulativă a variabilei aleatoare de intrare cu probabilitate unu, dacă dimensiunea setului de antrenare  $M$  tinde spre infinit. Astfel, un cuantizor vectorial de dimensiune  $N$ , optimal pentru un set de antrenare suficient de mare, este aproape optimal pentru vectorul de intrare curent.

Centroidul unei celule  $R_j$  este definit prin:

$$\mathbf{Y}_j = E\{\mathbf{X} | \mathbf{X} \in R_j\} = \frac{E\{\mathbf{X}S_j(\mathbf{X})\}}{E\{S_j(\mathbf{X})\}}, \quad (4.41)$$

unde  $S_j$  este funcția selector definită de (4.10).  
care devine:

$$\mathbf{Y}_j = \frac{\frac{1}{M} \sum_{i=1}^M \mathbf{v}_i S_j(\mathbf{v}_i)}{\frac{1}{M} \sum_{i=1}^M S_j(\mathbf{v}_i)} \quad \text{pentru } j = 1, 2, 3, \dots, N, \quad (4.42)$$

unde indexul  $i$  indică vectorii de antrenare iar indexul  $j$  indică celulele partiției.

Se observă din nou că centroidul calculat conform relației (4.42) este nedefinit pentru regiunile  $R_j$  cu probabilitatea zero, adică celulele goale, în care nu există puncte de antrenare.

Condiția de cel mai apropiat vecin va stabili partiția spațiului de antrenare:

$$R_j = \{\mathbf{v} \in T: \|\mathbf{v} - \mathbf{y}_j\| < \|\mathbf{v} - \mathbf{y}_i\| \text{ pentru orice } i\}. \quad (4.43)$$

Distorsiunea medie pentru cazul discret se poate exprima prin relația:

$$D = \frac{1}{M} \sum_{i=1}^M \sum_{j=1}^N d(\mathbf{v}_i, \mathbf{y}_j) S_j(\mathbf{v}_i) \quad (4.44)$$

Cu acestea, iterația Lloyd se poate aplica unui vector de antrenare pentru a obține un cuantizor vectorial local optimal pentru acest vector și, în concluzie, dacă se obține un cuantizor vectorial optimal, de mărime  $N$ , pentru un set de antrenare suficient de mare, acest cuantizor va fi aproape optimal pentru un set de intrare reprezentat de un vector aleator.

*Iterația Lloyd pentru date empirice:*

a) fiind dat un dicționar  $C_m = \{\mathbf{y}_i; i = 1, 2, \dots, N\}$  se partiționează setul de antrenare folosind regula celui mai apropiat vecin:  $R_i = \{x \in T : d(\mathbf{x}, \mathbf{y}_i) \leq d(\mathbf{x}, \mathbf{y}_j)\}$

b) pentru partiția astfel obținută, se calculează centroidul pentru fiecare regiune, rezultând un nou dicționar  $C_{m+1} = \{cent(R_i); i = 1, 2, \dots, N\}$

În cazul în care în etapa a) rezultă o celulă goală, se alege o metodă empirică pentru a atribui acestei celule un vector de cod, în locul centroidului, care în acest caz nu se poate calcula.

Cu acestea, se poate enunța algoritmul complet pentru a obține un cuantizor vectorial cu performanța dată prin distorsiunea  $D$ .

*Algoritmul Lloyd generalizat (GL):*

1. se începe cu un dicționar inițial  $C_1$ ,  
se atribuie variabilei  $m$  valoarea 1,
2. fiind dat dicționarul  $C_m$ , se aplică iterația Lloyd, rezultând un nou dicționar, mai bun,  $C_{m+1}$ ,
3. se calculează distorsiunea  $D$  pentru noul dicționar. Dacă distorsiunea a scăzut cu mai puțin decât o valoare impusă față de ultima iterație, atunci algoritmul se oprește. Dacă nu, se face  $m = m + 1$  și se reia pasul 2.

Se pot folosi mai multe criterii de oprire pentru algoritmul GL:

- atingerea valorii impuse pentru distorsiunea  $D$ , dar trebuie precizat că acest lucru nu este garantat de algoritmul GL, depinzând de secvența de antrenare, de dicționarul inițial și mai ales de mărimea  $N$  a dicționarului,
- scăderea distorsiunii față de valoarea avută la pasul anterior, eventual exprimată printr-un raport, în forma:  $(D_m - D_{m+1}) / D_m$ .

În toate situațiile este util să se prevadă oprirea algoritmului după un număr suficient de mare de iterații, chiar dacă nu s-a atins criteriul de distorsiune, pentru a evita un calcul la infinit, mai ales în cazul în care criteriul stabilit nu se poate atinge.

Aplicarea succesivă a iterației Lloyd pentru îmbunătățirea unui dicționar este cunoscută sub numele de algoritmul Lloyd generalizat (GLA) sau algoritmul LBG (Linde – Buzo – Gray) [LBG80]. Pentru un set de antrenare finit, algoritmul GLA produce întotdeauna o secvență de cuantizori vectoriali a căror distorsiune medie converge către o valoare limită după un număr finit de iterații [GER93].

Algoritmul GL trebuie inițializat cu un dicționar care se poate construi prin oricare din metodele prezentate anterior. Dată fiind proprietatea acestui algoritm de a îmbunătăți performanțele unui dicționar dat, orice alt algoritm de construcție a unui dicționar poate fi

punctul de plecare pentru algoritmul GL, și se poate folosi pentru a obține un dicționar inițial. Aplicarea algoritmului GL poate aduce în acest caz o îmbunătățire a acestui dicționar, într-o măsură mai mare sau mai mică, dependentă de eficiența algoritmului folosit pentru generarea dicționarului inițial.

Trebuie precizat că, deoarece fiecare iterație a algoritmului GL produce doar modificări locale în dicționar, dicționarul final nu diferă esențial de cel inițial și, odată ce dicționarul inițial a fost ales, algoritmul conduce la un dicționar optimal care este cel mai apropiat de cel inițial, din spațiul dicționarelor posibile, dar nu conduce la dicționarul global optimal.

#### 4.7.4 Alte metode pentru construcția unui cuantizor vectorial

##### *Relaxarea stochastică*

Această metodă introduce un factor aleator în algoritmul GLA, ceea ce permite în cele mai multe cazuri să se evite oprirea algoritmului la un minim local și să se reducă sau chiar să se elimine dependența soluției finale de dicționarul inițial.

Relaxarea stochastică presupune că fiecare iterație prin care se încearcă obținerea unui minim al funcției de cost se face aplicând o perturbare aleatoare a stării (a dicționarului). Convergența este asigurată prin reducerea progresivă a perturbației aplicate.

Primele încercări în acest sens au folosit un zgomot aleator care este adăugat vectorilor din setul de antrenare înainte de a aplica algoritmul GL, sau, într-o variantă a acestei metode, adăugat fiecărei componente a fiecărui vector de cod. Zgomotul are media nulă și varianța zgomotului este redusă la fiecare iterație.

##### *Învățarea Kohonen*

O altă metodă de a construi dicționarul este bazată pe rețele neuronale artificiale, fiind cunoscută sub numele de învățare Kohonen [WNF94] sau învățare competitivă [BHT97]. Pentru fiecare vector  $\mathbf{X}(n)$  din setul de învățare, se calculează un nou cuvânt de cod definit prin relația:

$$\mathbf{y}_i(n) = \mathbf{y}_i(n-1) + \alpha_i N_i(j, n) [\mathbf{X}(n) - \mathbf{y}_i(n-1)] \quad (4.45)$$

unde  $\alpha_i(n)$  este rata de învățare, reprezentată de o funcție care scade de la 1 la 0 pe măsură ce învățarea progresează și  $N_i(j, n)$  este o funcție de vecinătate, care exprimă relația de vecinătate a vectorilor din dicționar cu noul vector  $\mathbf{X}(n)$ .

În literatură se prezintă și alte metode de construcție a unui dicționar pentru un cuantizor vectorial: inițializarea aleatoare, relaxarea stimulată, despicierea dicționarului [GER93].

#### 4.7.5 Algoritmul LBG îmbunătățit

Dicționarul obținut prin algoritmul LBG depinde într-o măsură foarte mare de modul de alegere al dicționarului inițial. Așa cum s-a precizat, algoritmul LBG conduce, în general, la un optim local, dar pot exista cazuri, ca de exemplu la alegerea aleatoare a dicționarului inițial, în care acest optim local este departe de a fi acceptabil.

Pentru a obține un dicționar eficient și în astfel de cazuri, s-a elaborat un algoritm LBG îmbunătățit [FTZ97], [CBA04-2]. Două astfel de situații dezavantajoase sunt ilustrate în figura 4.4. a), iar în figura 4.4. b) este reprezentată situația rezultată după aplicarea algoritmului LBG îmbunătățit.

În figura 4.4. a) este reprezentată o distribuție posibilă a vectorilor de intrare și un vector de cod inițial, notat cu numărul 4. Acest vector de cod va determina întotdeauna o

celulă goală și nu va fi folosit niciodată pentru a coda un vector de intrare. Această situație se poate evidenția ușor în cazul aplicării algoritmului LBG și se pot aplica diferite procedee pentru a deplasa acest vector de cod într-o regiune utilă [GER93].

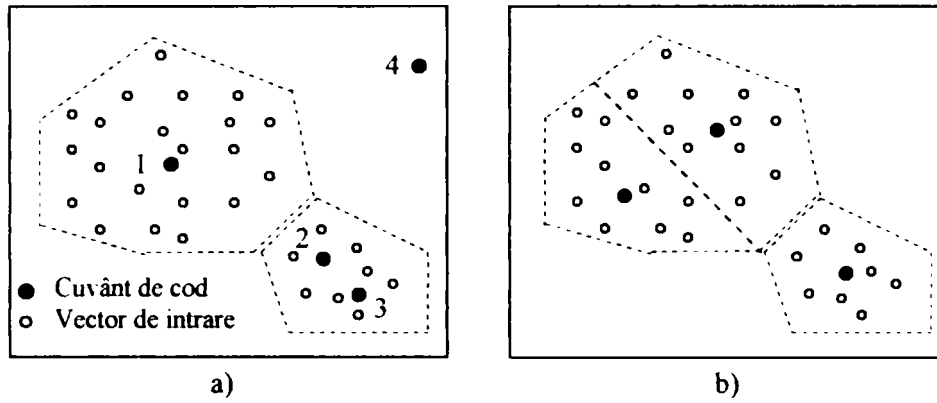


Fig. 4.4. O distribuție posibilă a vectorilor de cod

O a doua situație, pe care însă algoritmul LBG nu o poate evidenția și nici elimina, este aceea reprezentată în figura a) prin vectorii de cod cu numerele 2 și 3 asociați unei regiuni de mici dimensiuni și vectorul de cod 1 asociat cu o regiune relativ mare.

Prin aplicarea algoritmului LBG, regiunea asociată vectorilor de cod 2 și 3 se va divide în două celule de mici dimensiuni, în care cuantizarea se va face cu o eroare foarte mică în timp ce regiunea asociată vectorului de cod 1 va rămâne aproape neschimbată, contribuind, prin eroarea de cuantizare mare care rezultă, la degradarea erorii medii a cuantizorului.

Algoritmul LBG îmbunătățit are la bază o teoremă enunțată și demonstrată în [CSQ95], conform căreia, într-un cuantizor vectorial, fiecare celulă contribuie cu o distorsiune egală la distorsiunea cuantizorului.

Se calculează distorsiunea medie pentru întreg cuantizorul prin:

$$D_m = \frac{1}{N} \sum_{i=1}^N D_i \tag{4.46}$$

unde  $D_i$  este distorsiunea introdusă de celula  $i$  și se definește, pentru fiecare celulă, o măsură a utilității  $U_i$ , prin:

$$U_i = \frac{D_i}{D_m} \tag{4.47}$$

Pentru un cuantizor optim, conform teoremei enunțate anterior, ar trebui ca  $D_m = D_i$ , adică  $U_i = 1$ . Egalizarea distorsiunilor pentru fiecare celulă se face în două etape:

- pentru fiecare celulă cu  $U_i < 1$  se încearcă pe rând concatenarea cu una din celulele vecine, obținând astfel o celulă cu un indice de utilitate mai mare,
- simultan se încearcă divizarea fiecărei celule cu utilitate mare în două celule mai mici, cu indice de utilitate mai mic, adică mai apropiat de unu.

Aceste două operații sunt echivalente cu a muta un cuvânt de cod dintr-o celulă în alta, așa cum se arată în figura 4.4.b), numărul de cuvinte de cod rămânând neschimbat.

Operația de divizare a unei celule, respectiv a vectorului de cod asociat, se face conform celor ilustrate în figura 4.5.

Celula supusă operației de divizare se consideră încadrată într-un hiper cub  $k$  – dimensional definit de:

$$I = [x_{1m}, x_{1M}] \times [x_{2m}, x_{2M}] \times \dots \times [x_{km}, x_{kM}] \tag{4.48}$$

în care  $x_{im}$  și  $x_{iM}$  sunt coordonatele minime și maxime după dimensiunea  $i$  a celulei.



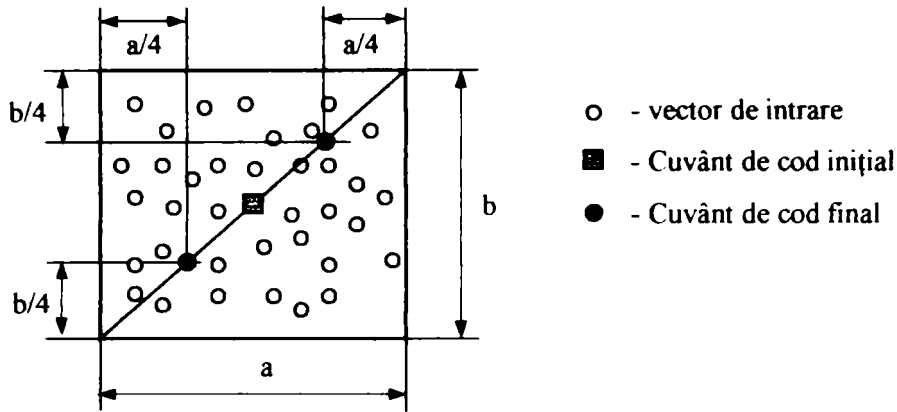


Fig. 4.5. Amplasarea noilor vectori de cod

Cele două noi cuvinte de cod sunt plasate pe diagonala principală a hiper-cubului I, la distanță egală de centrul hiper-cubului.

După ce se realizează divizarea vectorului de cod în două componente conform metodei ilustrate în figura 4.5, se aplică algoritmul LBG clasic (limitat la un număr redus de iterații, respectiv la un prag de modificare a distorsiunii relativ ridicat), pentru a ajusta poziția vectorilor de cod în noile celule.

Se calculează distorsiunea înainte de a executa deplasarea vectorului de cod:

$$D_{old} = D(\{Y, C\}) \quad (4.49)$$

și apoi noua valoare a distorsiunii, pentru fiecare posibilitate de deplasare, considerând diferite celule cu indicele de utilitate mai mare decât 1:

$$D_{new} = D(\{Y', C'\}) \quad (4.50)$$

unde  $C'$  este noul dicționar rezultat prin deplasarea vectorului de cod, și reținând cea mai mică distorsiune  $D_{new}$ .

## 4.8 Metode de cuantizare vectorială

Deși cuantizarea vectorială cu un dicționar proiectat în mod optimal pentru o anumită distribuție de probabilitate care caracterizează vectorii de intrare pare a fi cea mai bună metodă de cuantizare, utilizarea practică a cuantizării vectoriale, așa cum a fost prezentată în capitolul 1, poate ridica în unele situații probleme datorită următoarelor cauze:

- dimensiunea memoriei  $M$  necesare pentru păstrarea dicționarului și
- complexitatea calculelor, exprimată prin numărul de operații aritmetice  $A$  necesare pentru căutarea vectorului de cod în dicționar.

Pentru un cuantizor cu rata  $r$ , exprimată în biți pe componentă a vectorilor de cod, cu un dicționar de dimensiune  $N$  și cu vectori de dimensiune  $k$ , complexitatea se poate exprima prin:

$$A = kN = k2^{rk} \quad (4.51)$$

Valoarea din expresia (4.51) reprezintă atât necesarul de memorie, exprimat în componente ale vectorilor de cod, cât și complexitatea de calcul, exprimată în număr de operații și, după cum rezultă din această relație, ambele mărimi cresc exponențial odată cu creșterea dimensiunii cuantizorului [MNH94]. Un cuantizor vectorial se poate caracteriza complet prin setul  $(r, D, A, M)$ , unde pe lângă memoria  $M$  și complexitatea de calcul  $A$  se consideră și rata  $r$  și distorsiunea  $D$ . Pentru un semnal eșantionat cu frecvența  $f_s = 8$  kHz și un



cuantizor cu rezoluția  $r = 1$  bit/eșantion, rezultă o rată de 8 kbps. Grupând  $k$  eșantioane succesive într-un vector, va rezulta un dicționar de mărime  $N = 2^k$  și deci necesarul de memorie și complexitatea vor fi egale cu  $k2^k$ . Numărul de operații necesare în unitatea de timp pentru căutarea în dicționar este [CLR91]:

$$O = Nf_s = 2^k f_s \quad (4.52)$$

iar inversul acestui număr exprimă timpul disponibil pentru o operație, echivalent cu durata unui ciclu al procesorului.

În tabelul 4.1 [GER93] sunt indicate complexitatea și durata unui ciclu procesor pentru diferite valori  $k$  ale dimensiunii vectorilor de cod, calculate conform relației (4.52). După cum se observă din tabelul 4.1, limita procesoarelor actuale este atinsă pentru valori ale parametrului  $k = 15 \div 17$ . Reducerea dimensiunilor vectorilor de cod însă, în cele mai multe cazuri, împiedică exploatarea unor dependențe statistice care pot exista între vectorii ce urmează a fi codati.

Tabelul 4.1 Complexitatea cuantizorului vectorial funcție de dimensiune

Dimensiune $k$	Complexitate	Viteza procesor	Durata unui ciclu
1	2	16 kHz	62,5 $\mu$ s
4	64	128 kHz	8 $\mu$ s
6	2048	2 MHz	0,5 $\mu$ s
10	10240	8 MHz	125 ns
12	49152	33 MHz	31 ns
20	$10^{13}$	8 GHz	125 ps
40	$10^{25}$	$8,7 \times 10^9$ GHz	0,1 ps

Rezolvarea problemei complexității se poate aborda în două moduri diferite [GER93]:

- aplicarea unor metode rapide de codare vectorială nestructurată,
- structurarea cuantizorului în scopul reducerii complexității

Pe de altă parte, s-au dezvoltat noi categorii de cuantizoare vectoriale, care sunt [GER93]:

- *predictive*, în sensul că dispun de memorie și folosesc, pentru cuantizarea vectorului curent informații despre vectorii precedenți (sau viitori),
- *adaptive*, în sensul că se adaptează, într-un mod specific, caracteristicilor semnalului care se cuantizează,
- *cu rată variabilă*, adică cuvântul de cod care descrie ieșirea cuantizorului poate avea lungime variabilă, dependentă de semnalul de intrare,

precum și unele combinații între aceste categorii.

În cele ce urmează se vor prezenta unele dintre cele mai importante tehnici de cuantizare vectorială din fiecare categorie din cele menționate mai sus, indicând pentru fiecare caracteristicile fundamentale.

#### 4.8.1 Metode rapide de codare vectorială nestructurată

Pentru reducerea complexității operațiilor de căutare în codorul vectorial s-au dezvoltat metode de cuantizare la care dicționarul nu este structurat, dar se aplică unele artificii cu scopul simplificării operațiilor de căutare a celui mai apropiat vecin pentru un vector de intrare dat [ARM93]. Unele dintre aceste metode se pot numi suboptimale [GER93], deoarece nu găsesc exact cel mai apropiat vecin, ci unul dintre vecinii cei mai apropiați. Față de metoda de căutare exhaustivă, la care timpul de căutare este fix și independent de vectorul de intrare, aceste metode se diferențiază prin timpul variabil de căutare.

- **Metoda distanțelor parțiale**

Pentru un vector de intrare  $\mathbf{x}$ ,  $k$ -dimensional, se execută o procedură de căutare secvențială și după ce s-au testat  $m$  vectori de cod, se reține distorsiunea (distanța) minimă  $D$  găsită până în acest moment, respectiv indexul  $l$  corespunzător.

Pentru a testa vectorul de cod  $m+1$ , se inițializează un contor  $r$  și pentru fiecare valoare  $r = 1, 2, \dots, k$  se calculează distanțele parțiale  $D_r = \sum_{i=1}^r (x_i - c_i)^2$ . Dacă  $D_r > D$ , se respinge vectorul de cod testat și se trece la vectorul de cod următor. Dacă  $D_r < D$  se incrementează  $r$ , se calculează noua distanță parțială  $D_r$  și se reia testul. Dacă  $D_r > D$  și s-a ajuns la valoarea  $r = k$ , se înlocuiește  $D$  cu  $D_r$  și  $l$  cu  $m+1$  și se trece la următorul vector de intrare [CGB84].

Acest algoritm conduce întotdeauna la vectorul de cod cel mai apropiat și în cel mai defavorabil caz, timpul de căutare este același ca și în cazul căutării secvențiale. Este de așteptat însă ca acest cel mai defavorabil caz să nu apară pentru fiecare vector de intrare și se pot obține reduceri ale timpului de căutare de până la patru ori față de căutarea secvențială.

- **Metoda proiecțiilor**

Metoda proiecțiilor este o particularizare a unei metode mai generale de reducere a complexității codorului, bazată pe cunoașterea formei geometrice a celulelor Voronoi și memorarea acestor forme într-o bază de date care se va folosi în procesul de codare pentru încadrarea vectorului de intrare curent în celula corespunzătoare [RKP99].

Deoarece forma celulelor Voronoi este de cele mai multe ori neregulată, metoda proiecțiilor încadrează fiecare celulă în cel mai mic domeniu  $k$ -dimensional limitat de  $2k$  hiperplane perpendiculare pe axele de coordonate, descris de valorile minime și maxime după fiecare axă de coordonate. Prin aceasta, pe fiecare axă de coordonate se determină  $2N - 1$  intervale, care se memorează. Procedura de căutare folosește aceste tabele pentru o cuantizare scalară a fiecărei componente a vectorului de intrare, determinând un set de celule candidate și, pe măsură ce se cuantizează componentele succesive, se restrânge domeniul celulelor candidate prin intersecția domeniului determinat anterior cu domeniul corespunzător componentei curente. În final, în interiorul setului de celule candidate se face o căutare secvențială pe tot domeniul, pentru a găsi vectorul de cod [CGR86].

- **Metoda arborelui  $K - d$**

Un arbore  $K - d$  este un arbore binar, ale cărui noduri terminale sunt vectorii de cod, și care este asociat cu hiperplane de decizie în fiecare nod, unde fiecare hiperplan este ortogonal pe una din axele de coordonate ale spațiului  $k$ -dimensional. În fiecare nod al arborelui se examinează o singură componentă a vectorului de intrare, care se încadrează de o parte sau de alta a hiperplanului respectiv. Arborele  $K - d$  este foarte eficient, dar necesită o procedură de construcție adecvată [EQZ89], [MFG90]. Metoda proiecțiilor prezentată anterior este un caz particular al arborelui  $K - d$ .

- **Metoda inegalității triunghiului**

Această metodă se bazează pe relația de inegalitate care are loc între laturile unui triunghi, aplicată la măsura distorsiunii. În particular metoda este adecvată pentru distanța euclidiană ca măsură a distorsiunii. Metoda inegalității folosește puncte de referință (*ancore*), pentru care sunt calculate și memorate distanțele până la fiecare vector de cod. Pentru fiecare vector de intrare se calculează mai întâi distanțele până la aceste puncte ancoră și prin comparație se elimină un mare număr de vectori candidați [ORC91]. Pentru a ilustra ideea, se presupune un singur punct ancoră  $\mathbf{a}$  și se consideră cunoscute distanțele  $d_i = \|\mathbf{c}_i - \mathbf{a}\|$ , pentru toate valorile  $i = 1, 2, \dots, N$ . Fie  $\mathbf{c}_1$  un vector de cod candidat. Se calculează distanța de la

vectorul de intrare  $\mathbf{x}$  și ancora  $\mathbf{a}$  și la vectorul candidat  $\mathbf{c}_j$  și se exprimă inegalitatea triunghiului în forma:  $\|\mathbf{x} - \mathbf{c}_j\| > d_j - \|\mathbf{x} - \mathbf{a}\|$

Dacă  $d_j > \|\mathbf{x} - \mathbf{c}_j\| + \|\mathbf{x} - \mathbf{a}\|$ , rezultă că și  $\|\mathbf{x} - \mathbf{c}_j\| > \|\mathbf{x} - \mathbf{c}_1\|$ , ceea ce conduce la eliminarea vectorului de cod  $\mathbf{c}_j$  din lista candidaților.

## 4.8.2 Cuantizarea vectorială structurată

### 4.8.2.1 Cuantizarea vectorială în structură arborescentă (TSVQ)

Organizarea dicționarului unui cuantizor vectorial în structură arborescentă (TSVQ – tree – search VQ) permite căutarea în etaje succesive, eliminând în fiecare etapă de căutare un număr relativ mare de vectori candidați. Într-un arbore de ordinul  $m$  ( $m$ -ar), reprezentat în figura 4.6, vectorul de intrare este comparat în fiecare etaj cu  $m$  vectori care reprezintă nodurile arborelui. Cel mai apropiat vector, în sensul distorsiunii minime, va determina calea de urmat în arbore. La fiecare etaj, numărul de vectori candidați este redus de  $m$  ori. Pentru un dicționar cu dimensiunea  $N = m^d$  arborele asociat este de ordinul  $m$  și are adâncimea  $d$ . Dacă  $m = 2$ , arborele este binar [LHM87].

Codarea vectorială arborescentă afectează doar tehnica de căutare folosită la codare, decodorul neavând nevoie decât de cei  $N = m^d$  vectori de cod de pe ultimul nivel, la fel cu o codare vectorială convențională. Arborii prin intermediul cărora se face cuantizarea pot fi echilibrați, adică toate ramurile arborelui au aceeași lungime, sau neechilibrați, dacă arborele are ramuri cu lungimi diferite. În prima variantă, rezultă un număr fix de biți pentru orice vector de intrare (rată fixă). A doua variantă oferă o posibilitate de codare cu număr variabil de biți, utilă în unele aplicații în care se cere o eroare uniformă de cuantizare sau în cazul cuantizării la rată variabilă.

Pentru construcția unui codor vectorial cu structură arborescentă se folosește algoritmul Lloyd generalizat (GLA) pentru a obține etajele succesive cu ajutorul unui set de vectori de antrenare  $\mathcal{T}$ . Pentru cazul arborilor nesimetrice s-au dezvoltat algoritmi de tip greedy în care, la fiecare pas, se despică un singur nod, și anume acela pentru care rezultă cea mai mare îmbunătățire a măsurii distorsiunii pe ansamblul arborelui [WDG94].

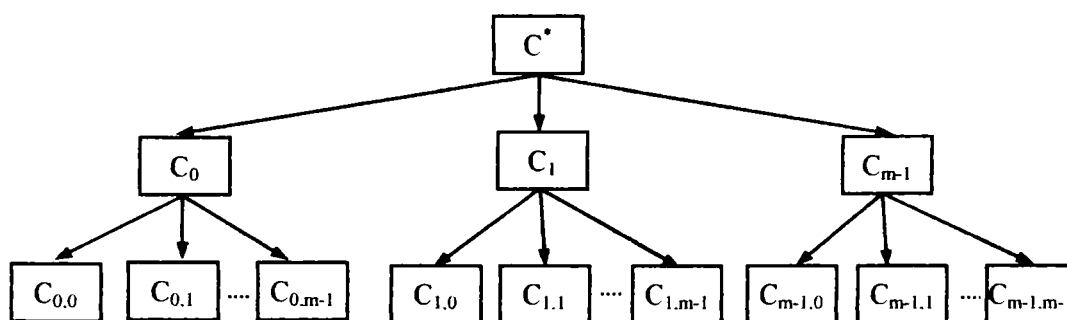


Fig. 4.6. Cuantizarea vectorială arborescentă

### 4.8.2.2 Cuantizarea vectorială cu clasificare (CVQ)

Cuantizarea vectorială cu clasificare (CVQ – Class VQ) este similară codării arborescente cu un singur etaj, dar criteriul de selecție pentru una din cele  $m$  ramuri îl constituie o caracteristică comună unei clase de vectori de intrare. În baza acestei caracteristici, un clasificator adecvat va clasifica vectorul de intrare într-o clasă  $C'_i$ , asociată cu un subset particular  $C_i$  al dicționarului, în care se face apoi căutarea finală pentru a găsi cea mai bună aproximare a vectorului de intrare [CBA04-3]. Cuantizorul CVQ ilustrat în figura

4.7 generează un index  $i$  întreg în domeniul  $1...m$ , care identifică unul din subdicționările  $C_i$ , la care se adaugă  $\log_2 \|C_i\|$  biți pentru a specifica vectorul ales din acest subdicționar.

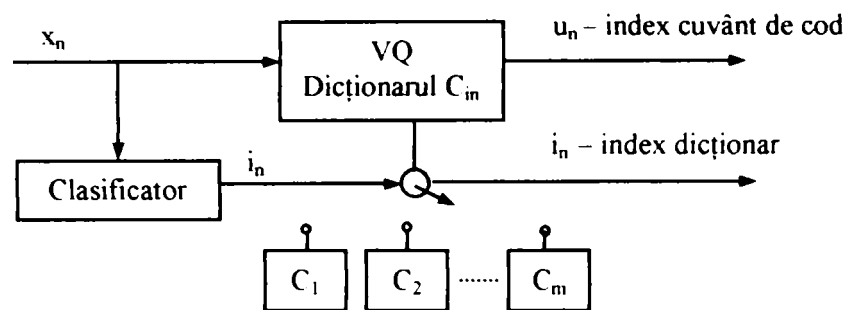


Fig. 4.7. Cuantizarea vectorială cu clasificare

Eficiența CVQ este direct determinată de alegerea corectă a criteriului după care se face clasificarea vectorilor de intrare și de proiectarea clasificatorului. Clasificatorul se poate realiza în multe moduri și poate fi bazat pe anumite caracteristici specifice ale semnalului de intrare sau pe considerente euristice, adoptate în faza de proiectare, în funcție de setul de vectori ce urmează a fi codat. Clasificatorul se poate proiecta euristic sau se poate folosi un algoritm GLA și un set de antrenare cu caracteristicile alese, extras din setul de vectori de intrare.

Odată ales clasificatorul, subseturile de dicționare se pot proiecta folosind algoritmul clasic. Pentru antrenarea subdicționarelor, se trece setul de vectori de antrenare  $\mathcal{T}$  prin clasificatorul ales, rezultând  $m$  subseturi de vectori de antrenare  $\mathcal{T}_j$ , care se folosesc pentru a antrena cele  $m$  subdicționare  $C_i$ . Dimensiunea  $m_j$  a fiecărui subdicționar  $C_i$  se stabilește tot pe criterii euristice, sau pe baza unui algoritm de optimizare a alocării biților, ținând cont de dimensiunea finală totală  $N$  dorită pentru dicționar. Dimensiunea  $m_j$  se poate alege și astfel încât să fie proporțională cu fracțiunea  $\mathcal{T}_j$  a subsetului corespunzător de antrenare.

#### 4.8.2.3 Cuantizarea vectorială cu transformare

Codorul vectorial cu transformare, ilustrat în figura 4.8, aplică la intrarea unui cuantizor vectorial nu vectorul semnal de intrare  $\mathbf{X}$  ci vectorul  $\mathbf{Y}$  rezultat în urma unei transformări liniare ortogonale  $\mathbf{T}$  a vectorului  $\mathbf{X}$ . Ieșirea  $\hat{\mathbf{Y}}$  a codorului vectorial este supusă transformării inverse, pentru a se obține o aproximare  $\hat{\mathbf{X}}$  a semnalului original  $\mathbf{X}$ .

Se observă că, în cazul în care dimensiunea VQ este aceeași cu dimensiunea transformării  $\mathbf{T}$ , vectorii de cod optimali pentru structura din figura 4.8 reprezintă transformata  $\mathbf{TX}$  a vectorilor de cod optimali pentru dicționarul ce s-ar folosi pentru codarea directă a vectorului  $\mathbf{X}$  fără a folosi nici o transformare. Aplicarea transformării are ca efect compactarea informației conținută în vectorul de intrare astfel încât rezultă o reducere a dimensiunii vectorilor care vor fi codati și în final o reducere a complexității codorului.

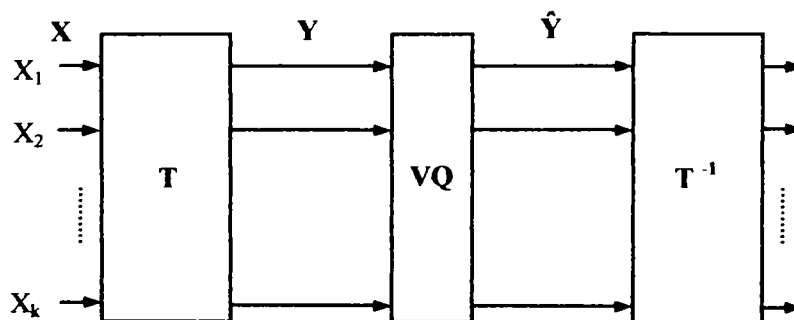


Fig. 4.8. Cuantizarea vectorială prin transformare

Este cunoscută aplicarea transformării cosinus discrete (DCT), folosită în special pentru codarea semnalului de imagine. Aplicații la cuantizarea semnalului vocal sunt prezentate în [LVC96], [CLS01]. De asemenea, aplicarea unei transformări potrivit alese permite reducerea prin trunchiere a vectorului care se va coda, scăzând complexitatea codării. Câteva variante ale codării vectoriale prin transformare prezintă un interes major:

- **TVQ cu partiționare**, în care rezultatul transformării este partiționat în mai multe subseturi care sunt codate fiecare separat folosind codarea vectorială.
- **TVQ în subbenzi**, în care transformarea se face în domeniul frecvență, împărțind semnalul în subbenzi convenabil alese cu ajutorul unor filtre, uzual filtre în cuadratură în oglindă. [ALS84].
- **TVQ cu transformări wavelet**, în care transformarea  $T$  folosită este o transformare de tip wavelet [CGT90], realizând o descompunere a unor clase de semnale într-o sumă ponderată a unor funcții bază. Spre deosebire de transformata Fourier clasică, transformata wavelet folosește pentru obținerea funcțiilor bază o singură funcție, care este scalată și translatată corespunzător. Aceasta permite reconstituirea semnalului la rezoluții multiple. Câteva aplicații ale cuantizării vectoriale prin transformări wavelet sunt prezentate în [VJL91].

#### 4.8.2.4 VQ de produs (PVQ)

Cuantizarea vectorială de produs (PVQ – product VQ) se bazează pe ideea de a reduce complexitatea unui cuantizor vectorial prin descompunerea vectorului care urmează a fi cuantizat în subvectori de dimensiune mai mică și de a cuantiza apoi acești subvectori în mod independent, fiecare cu propriul dicționar, urmând ca la decodare să se reconstituie vectorul original prin concatenarea subvectorilor decodați.

Se prezintă în continuare trei variante particulare pentru PVQ

- **VQ partiționate**

Pentru a reduce dimensiunea dicționarului și volumul de calcule pentru căutare, se poate face o partiționare a vectorului de intrare în doi sau mai mulți subvectori. Astfel, vectorul de intrare  $\mathbf{x} = (x_1, x_2, \dots, x_k)$  se partiționează, de exemplu, în două componente astfel:  $\mathbf{x}_a = (x_1, x_2, \dots, x_i)$ , un vector de dimensiunea  $i < k$  și  $\mathbf{x}_b = (x_{i+1}, x_{i+2}, \dots, x_k)$ , un vector de dimensiunea  $k - i$ . Eroarea medie pătratică se poate exprima prin relația:

$$\|\mathbf{x} - \hat{\mathbf{x}}\|^2 = \|\mathbf{x}_a - \hat{\mathbf{x}}_a\|^2 + \|\mathbf{x}_b - \hat{\mathbf{x}}_b\|^2 \quad (4.53)$$

Codorul va căuta cea mai potrivită valoare pentru  $\mathbf{x}_a$  în dicționarul  $\mathcal{C}_a$  și respectiv pentru  $\mathbf{x}_b$  în  $\mathcal{C}_b$ , în mod independent. Mai mult, construcția optimă a dicționarelor se poate face folosind două seturi de antrenare,  $\mathcal{T}_a$  și  $\mathcal{T}_b$  formate din vectori de dimensiune  $i$  și  $k - i$ , obținuți prin partiționarea setului inițial de antrenare  $\mathcal{T}$ .

- **VQ cu extragerea mediei (mean removed VQ)**

În unele cazuri, valoarea medie a componentelor unui vector poate diferi mult de la un vector la altul și această valoare medie poate fi privită la rândul său ca o variabilă aleatoare. Anumite categorii de semnale au doar componente nenegative și prin urmare au o medie nenulă (ca de exemplu semnalele de imagine). În acest context, prin medie se înțelege media aritmetică a componentelor vectorului și media  $m$  a unui vector  $\mathbf{X}$  este un scalar.

Vectorul  $\mathbf{X}$  poate fi reprezentat ca o sumă dintre vectorul medie și vectorul rezidual:

$$\mathbf{X} = m\mathbf{I} + \mathbf{R} \quad (4.54)$$



rezultând o descompunere naturală a vectorului original în două caracteristici separate: media, ca nivel de bază și reziduul, ca formă a vectorului axat pe valoarea medie.

Cuantizarea separată a acestor două elemente, realizată cu schema din figura 4.9, reprezintă cuantizarea vectorială de tip medie – rezidual MRVQ (mean – residual VQ).

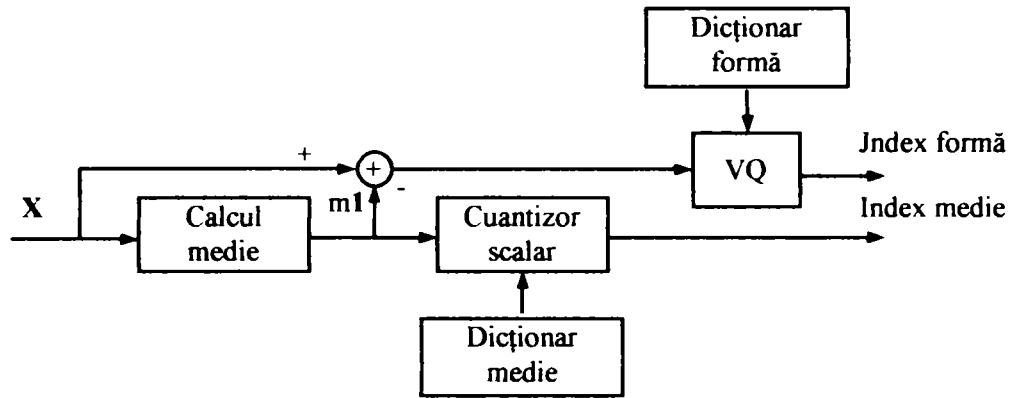


Fig. 4.9. Cuantizarea vectorială cu extragerea mediei

• **VQ formă – câștig (shape – gain VQ)**

O altă variantă de descompunere a unui vector se bazează pe extragerea valorii medii din fiecare componentă a vectorului [BGG80]. Această cantitate este denumită *câștig* și are rolul unui factor de normare a vectorului ce urmează a fi cuantizat. Vectorul normat reprezintă *forma* (*shape*). Această variantă de cuantizare se bazează pe faptul că un vector cu o anumită formă poate să apară în secvența de vectori de intrare cu diferite valori ale câștigului, ca de exemplu în semnalul audio în general sau în semnalul vocal în particular.

Codorul vectorial realizat conform metodei formă-câștig este reprezentat în figura 4.10.

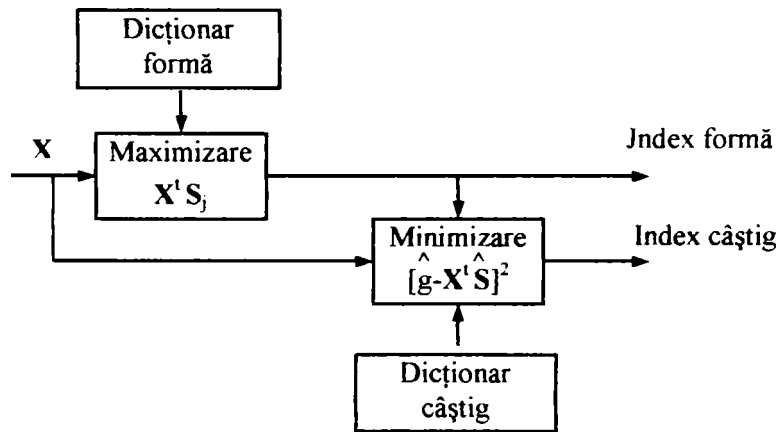


Fig. 4.10. VQ tip formă – câștig

Câștigul  $g$  al unui vector  $\mathbf{X}$  este o variabilă aleatoare definită de relația:

$$g = \|\mathbf{X}\| = \sqrt{\sum_{i=1}^k X_i^2} \tag{4.55}$$

Forma  $\mathbf{S}$  a vectorului  $\mathbf{X}$  caracterizat de un câștig  $g$  nenul este:

$$\mathbf{S} = \frac{\mathbf{X}}{g} \tag{4.56}$$

astfel încât pentru orice  $\mathbf{S}$ ,  $\|\mathbf{S}\|=1$ . Pentru a construi un cuantizor optimal, se exprimă distorsiunea cuantizorului prin:



$$d(\mathbf{X}, \hat{\mathbf{g}}\hat{\mathbf{S}}) = \|\mathbf{X} - \hat{\mathbf{g}}\hat{\mathbf{S}}\|^2 \quad (4.57)$$

unde  $\hat{\mathbf{g}}$  și  $\hat{\mathbf{S}}$  reprezintă valorile cuantizate ale câștigului, respectiv vectorului formă. Expresia (4.57) trebuie minimizată în raport cu  $\hat{\mathbf{g}}$  și  $\hat{\mathbf{S}}$ , ceea ce se poate face în două etape succesive.

#### 4.8.2.5 VQ în mai multe etaje MSVQ (Multistage VQ)

Dacă vectorul care urmează să fie cuantizat nu prezintă variații semnificative ale mediei sau câștigului, se poate aplica o variantă de cuantizare vectorială denumită cuantizare în mai multe etaje (cascadă) [BBM93]. Această tehnică, denumită cuantizarea rezidului, divide operația de cuantizare în mai multe etape succesive, așa cum este ilustrat în figura 4.11 pentru cazul a două etaje. Într-o primă etapă, cuantizorul  $Q_1$  din figura 4.11 realizează o cuantizare brută, folosind un dicționar de dimensiuni relativ mici, și eroarea rezultată este aplicată unui al doilea etaj de cuantizare  $Q_2$ , care realizează o cuantizare mai fină.

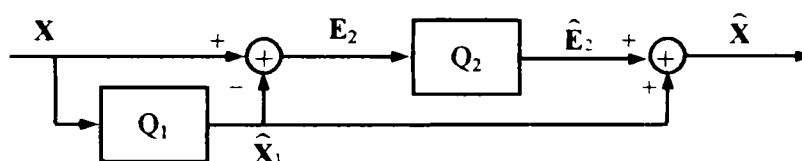


Fig. 4.11. Cuantizarea vectorială în două etaje

Cuantizorul cu două etaje din fig. 4.11 poate fi privit ca un caz particular al cuantizorului TSVQ cu două nivele. Deoarece cuvântul de cod rezultat  $I_1 I_2 \dots I_m$  este obținut prin concatenarea cuvintelor de cod obținute în fiecare etaj, se poate considera că acesta reprezintă un cod de produs, în care dicționarul echivalent de produs corespunde produsului cartezian  $C_1 \times C_2 \times \dots \times C_m$ , unde  $C_i$  este dicționarul folosit în etajul  $i$ , iar funcția  $g$  de compunere din decodor este o simplă sumă algebrică.

- **VQ arborescentă în mai multe etaje**

După cum s-a arătat, MSVQ poate fi privit ca un caz particular al TSVQ. În cazul TSVQ numărul dicționarelor crește exponențial cu numărul de etaje (nivele), în timp ce în cazul MSVQ numărul dicționarelor crește liniar cu numărul de etaje. Este posibil să se combine cele două metode într-una hibridă, care să permită un compromis între numărul de etaje și numărul de dicționare folosite. De exemplu, structura MSVQ este practică pentru un număr de două sau cel mult trei etaje, dar acest număr poate fi insuficient pentru a asigura o complexitate rezonabilă, datorită dimensiunilor mari ale dicționarelor, impuse de considerente de calitate. O soluție hibridă constă în a folosi în fiecare etaj al MSVQ câte un cuantizor TSVQ (în mod frecvent unul binar), obținând un cuantizor hibrid multietaj – arborescent (MSTVQ). O aplicație a acestei metode pentru codarea semnalului vocal de bandă largă este prezentată în [WYG90].

#### 4.8.2.6 VQ ierarhic – multirezoluție

Există situații în care se urmărește cuantizarea unui semnal care prezintă un număr foarte mare de eșantioane succesive puternic corelate între ele astfel încât gruparea acestora într-un vector ar putea fi eficient exploatată folosind cuantizarea vectorială. Este cazul semnalului vocal, pentru care eșantioanele succesive sunt statistic dependente pe o durată de 50 – 100 ms (până la 800 eșantioane pentru  $f_e = 8$  kHz), corespunzător unei unități fonetice. O situație similară se întâlnește frecvent și în cazul semnalelor de imagine. O soluție constă în a extrage din vectorul de mari dimensiuni, numit din acest motiv supervector, un vector de

dimensiuni acceptabile, care să reprezinte o caracteristică (**F** - feature) a supervectorului, care să țină cont de interdependența dintre componentele supervectorului. Caracteristica **F** se cuantizează vectorial cu una din metodele clasice, generând un index **I** și în același timp, caracteristica (sau valoarea sa cuantizată  $\hat{F}$ ) se utilizează pentru a obține, prin reducerea interdependențelor dintre componente, un nou supervector redus **Y**, care apoi se descompune în vectorii  $Y_i$ , de dimensiuni mai mici.

În continuare, fiecare dintre vectorii **Y**, poate fi codat cu aceeași schemă de bază folosită pentru supervectorul inițial, rezultând astfel o structură VQ ierarhică. Această schemă de cuantizare ierarhică poate fi considerată o schemă de cuantizare vectorială generală, care include ca și cazuri particulare cuantizările de tip formă - câștig, cu extragerea mediei, multietaj, precum și cele cu organizare arborescentă.

Pentru cuantizarea semnalelor audio și de vorbire, se poate alege drept caracteristică **F** câștigul, datorită dinamicii mari a acestor semnale. Pentru semnalele de imagine, caracteristica adecvată este media aritmetică a elementelor vectorilor. Este posibil ca vectorul caracteristic să fie doar o variantă subeșantionată a vectorului original și în acest caz schema devine o schemă multirezoluție, deoarece diferitele etaje procesează semnale cu diferite grade de subeșantionare. În acest caz, se face o subeșantionare a semnalului, urmată de cuantizarea vectorială, iar rezultatul se interpolează pentru a se reface semnalul original, rezultând o schemă de cuantizare vectorială interpolativă (IVQ).

#### 4.8.2.7 VQ prin interpolare neliniară

Cuantizarea prin interpolare poate fi considerată ca un caz particular al unei scheme mai generale de cuantizare prezentate în figura 4.12. Blocul de extragere a caracteristicii FE (FE - feature extractor) extrage din vectorul **X**, *n*-dimensional, un vector caracteristic **U**, *k*-dimensional. Eficiența cuantizării vectoriale este condiționată de îndeplinirea condiției  $k \cdot n$ . Codorul  $\mathcal{E}$  și decodorul  $\mathcal{D}$  folosesc fiecare câte un dicționar  $\mathcal{C}$  pentru a obține vectorul cuantizat  $\hat{U}$ , din care se calculează estimatul  $\hat{X} = g(\hat{U})$ . Estimarea se poate face atât printr-o funcție liniară, cât și printr-o funcție neliniară.

Deoarece interpolarea se face asupra unui set finit de vectori, compus din cei *N* vectori de cod (*k*-dimensionali) din dicționarul  $\mathcal{C}$ , este posibil ca operația de interpolare să se facă în faza de construcție a dicționarului, astfel încât la decodare să se folosească un dicționar  $\mathcal{C}^*$  care să conțină deja cei *N* vectori (*n*-dimensionali) interpolați, combinând interpolarea și decodarea într-un singur bloc, denumit decodor cu interpolare.

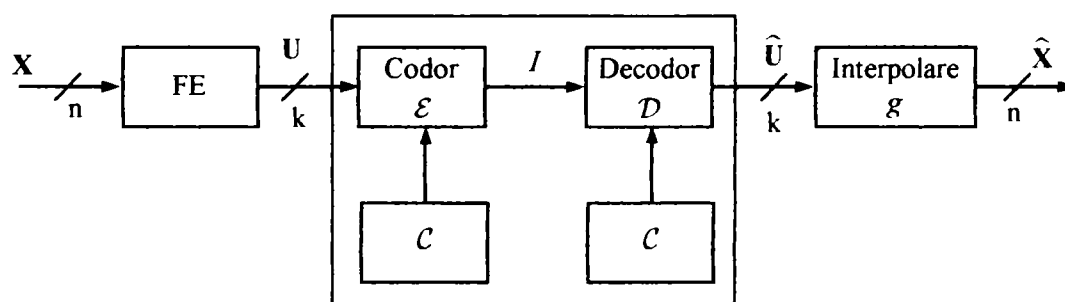


Fig. 4.12. VQ cu interpolare neliniară

În acest caz, se reduce complexitatea calculului, deoarece operația de interpolare nu se mai face în timp real, fiind înglobată deja în structura dicționarului  $\mathcal{C}^*$ , rezultând dicționare diferite, ca mărime și conținut, pentru codor și decodor.

### 4.8.3 Cuantizarea vectorială predictivă

Metodele de cuantizare prezentate se referă la cuantizarea, la un moment dat, a unui singur vector extras dintr-un semnal, adică la un cuantizor fără memorie, care codează fiecare vector în mod independent față de alți vectori.

În mod frecvent, se urmărește cuantizarea unei secvențe de vectori consecutivi, care în cele mai multe cazuri nu sunt statistic independenți. Exploatarea acestei dependențe între vectorii succesivi poate fi făcută de un cuantizor cu memorie, iar un mod eficient de a încorpora memoria în cuantizarea vectorială este utilizarea cuantizorului predictiv (PVQ), ca o extensie la cazul vectorial al cuantizorului predictiv scalar.

Structurile de bază pentru codorul și decodorul vectorial predictiv sunt prezentate în figurile 4.13a și 4.13b. În schema din figura 4.13a, estimatul  $\tilde{\mathbf{X}}_n$  al vectorului de intrare  $\mathbf{X}_n$ , este calculat de către blocul predictor vectorial, pe baza realizărilor trecute ale vectorului reconstituit  $\{\hat{\mathbf{X}}_k, ; k < n\}$ .

Se presupune un predictor de ordin  $m$  finit, adică un predictor care consideră doar cele mai recente  $m$  valori trecute ale vectorului de intrare:

$$\tilde{\mathbf{X}}_n = P(\hat{\mathbf{X}}_{n-1}, \hat{\mathbf{X}}_{n-2}, \dots, \hat{\mathbf{X}}_{n-m}) \tag{4.58}$$

Estimatul  $\tilde{\mathbf{X}}_n$  este scăzut din vectorul de intrare pentru a obține un vector diferență:

$$\mathbf{e}_n = \mathbf{X}_n - \tilde{\mathbf{X}}_n \tag{4.59}$$

care este codat vectorial rezultând

$$\hat{\mathbf{e}}_n = Q(\mathbf{e}_n) \tag{4.60}$$

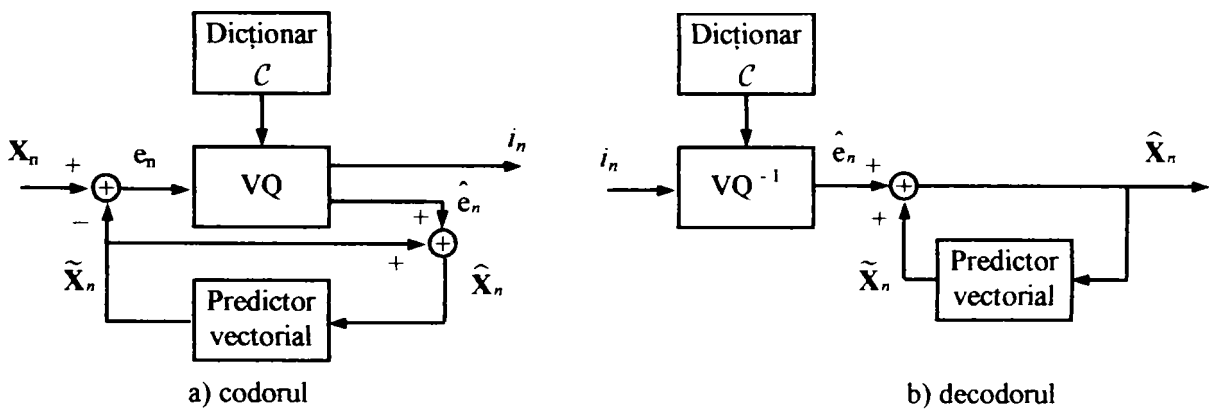


Fig. 4.13. Cuantizorul vectorial predictiv

Vectorul reconstituit prin decodare este deci:

$$\hat{\mathbf{X}}_n = \hat{\mathbf{e}}_n + \tilde{\mathbf{X}}_n \tag{4.61}$$

Se poate observa că pentru un predictor trivial, care produce la ieșire vectorul  $\mathbf{0}$  pentru orice vector de intrare, cuantizorul predictiv se reduce la cuantizorul vectorial simplu care, din acest punct de vedere, poate fi considerat un caz particular al PVQ, cu predictor de ordinul 0.

Este evidentă egalitatea:

$$\mathbf{X}_n - \hat{\mathbf{X}}_n = (\mathbf{X}_n - \tilde{\mathbf{X}}_n) - (\hat{\mathbf{X}}_n - \tilde{\mathbf{X}}_n) = \mathbf{e}_n - \hat{\mathbf{e}}_n \tag{4.62}$$

care arată că eroarea totală a cuantizorului vectorial predictiv este egală cu eroarea de cuantizare a vectorului diferență în cuantizorul vectorial și nu depinde de natura predictorului sau de „calitatea” acestuia.

Se poate defini un câștig al predicției în buclă închisă prin raportul:

$$G_c = \frac{E[\|\mathbf{X}_n\|^2]}{E[\|\mathbf{e}_n\|^2]} \quad (4.63)$$

și un câștig al codării prin:

$$G_Q = \frac{E[\|\mathbf{e}_n\|^2]}{E[\|\mathbf{e}_n - \hat{\mathbf{e}}_n\|^2]} \quad (4.64)$$

Proiectarea unui cuantizor vectorial predictiv presupune definirea predictorului  $P$  și a dicționarului care determină VQ. Există două modalități de a construi un VQ predictiv.

Cea mai simplă cale de proiectare a PVQ este metoda în buclă deschisă, în care predictorul și cuantizorul sunt proiectate separat. Predictorul este construit pe baza statisticii semnalului  $\mathbf{X}_n$ , folosind un set de observații empirice care constituie o secvență de antrenare, sau se poate optimiza pentru un canal de comunicație dat [LIN00].

După ce predictorul  $P$  este determinat, se proiectează cuantizorul vectorial, pe baza semnalului rezidual de predicție

$$\mathbf{W}_n = \mathbf{X}_n - P(\mathbf{X}_{n-1}, \mathbf{X}_{n-2}, \dots, \mathbf{X}_{n-m}) \quad (4.65)$$

Folosind o secvență de antrenare  $\{\mathbf{x}_n, n = -m, \dots, 1, 2, \dots, L\}$  se poate genera secvența de antrenare pentru vectorul eroare de predicție în buclă deschisă  $\{\mathbf{w}_n, n = 1, 2, \dots, L\}$  care se poate folosi pentru antrenarea VQ cu un algoritm de tip Lloyd generalizat.

În schema din fig. 4.13.a) predictorul  $P$  operează în buclă închisă asupra semnalului cuantizat  $\hat{\mathbf{X}}_n$  și deoarece predictorul a fost proiectat pe baza semnalului original  $\mathbf{X}_n$ , necuantizat, se poate aprecia că predictorul nu este optimal, dar va asigura performanțe apropiate de optim, datorită diferențelor mici între semnalul original și cel cuantizat.

O a doua modalitate de proiectare a unui VQ predictiv este metoda în buclă închisă care folosește un algoritm Lloyd, dar nu asigură optimalitatea caracteristică acestui algoritm din cauza existenței buclei închise [CGR86]. Predictorul este realizat, ca și în primul caz, în buclă deschisă, folosind un set de date de intrare ca set de antrenare și rămâne neschimbat. În continuare, sistemul în buclă închisă din fig. 4.13a, folosind un dicționar inițial, codează setul de antrenare, calculând vectorii reziduali și grupându-i în clustere asociate cu vectorii de cod. Pentru fiecare cluster este apoi calculat centroidul și construit un nou dicționar cu acești centroizi, înlocuind dicționarul inițial. Procedeeul se repetă până la atingerea performanței dorite.

Ambele variante folosesc un predictor care, odată determinat, nu se modifică pe durata antrenării VQ. Este posibil ca, după stabilirea dicționarului, să se îmbunătățească predictorul pentru un dicționar dat. Pentru aceasta, folosind predictorul existent, se produce o secvență de vectori reziduali, care poate fi folosită ca distribuție empirică pentru determinarea unui nou predictor, optimal. Operațiile de îmbunătățire a predictorului pentru un VQ dat și de îmbunătățire a dicționarului VQ pentru un predictor dat pot fi repetate alternativ de mai multe ori pentru a rezulta un VQ predictiv cât mai bun [CGR86], [WSS85], [WNF94], [KRO03].

#### 4.8.4 Cuantizarea vectorială adaptivă

Se poate îmbunătăți performanța codării dacă cuantizorul se poate adapta, într-un anumit sens, astfel încât să urmărească caracteristicile statistice locale ale vectorului ce urmează a fi codat, prin observarea, directă sau indirectă, a vectorilor dintr-o vecinătate a vectorului curent. Un cuantizor vectorial este adaptiv dacă dicționarul sau regulile de codare se modifică astfel încât să corespundă statisticilor locale ale secvenței de vectori de intrare.

Secvența de vectori asupra căreia se aplică cuantizarea adaptivă poate fi o secvență naturală de vectori, ca de exemplu în cuantizarea semnalului vocal sau a semnalelor audio în

general, sau a semnalelor de imagine. În alte cazuri, vectorii nu formează o secvență naturală, dar sunt asociați și referiți ca făcând parte dintr-o secvență. Cele mai multe scheme adaptive divid secvența de vectori de intrare în blocuri de lungime fixă, denumite cadre. Pentru fiecare cadru, cuantizorul extrage unele caracteristici esențiale, pe care le folosește pentru a se adapta în mod specific la cadrul curent. Metodele de cuantizare cu constrângeri structurale prezentate anterior nu sunt adaptive în sensul precizat, deoarece se referă la cuantizarea unui singur vector, fără a ține cont de secvența de vectori vecini, trecuți sau viitori. Cuantizarea predictivă și cuantizorul cu stări finite pot fi considerate adaptive în sensul definiției anterioare, dar uzual noțiunea de „adaptiv” se referă la sisteme care se modifică relativ lent față de rata vectorilor de intrare, comparativ cu sistemele care se modifică semnificativ de la un vector de intrare la altul și care sunt denumite uneori și recursive. Nu întotdeauna se poate evidenția o limită netă între cuantizarea vectorială adaptivă și cea recursivă. În figura 4.14 se prezintă o clasificare generală a metodelor de cuantizare vectorială adaptivă întâlnite în literatură.

Se disting două categorii mari, care se diferențiază prin dicționarele folosite: *dicționar cunoscut apriori* și *dicționar inductiv*.

Cuantizarea cu *dicționar universal cunoscut apriori* poate fi bazată pe *contextul local*, dacă presupune identificarea unor intervale de timp pe care procesul poate fi considerat staționar și adaptarea cuantizorului la caracteristicile locale ale semnalului. Adaptarea poate fi *înainte* sau *înapoi* relativ la vectorul curent. În această categorie se pot include cuantizarea cu adaptarea mediei și adaptarea câștigului [GER93], metodele de cuantizare cu excitare prin vectori (CELP), prezentate în cap. 3, precum și metodele de cuantizare cu stări finite [GER93] și cele predictive, prezentate anterior. Cuantizarea cu *decizie de prag* folosește un dicționar local, care este actualizat (modificat) numai dacă la cuantizarea unui vector nu se poate asigura distorsiunea minimă impusă [FAH97], [ZZW96]. Dicționarul folosit poate fi *explicit* cunoscut la decodor sau poate fi *implicit* definit ca o funcție de vectorii codați, ca de exemplu cuantizarea vectorială latice [GER93].

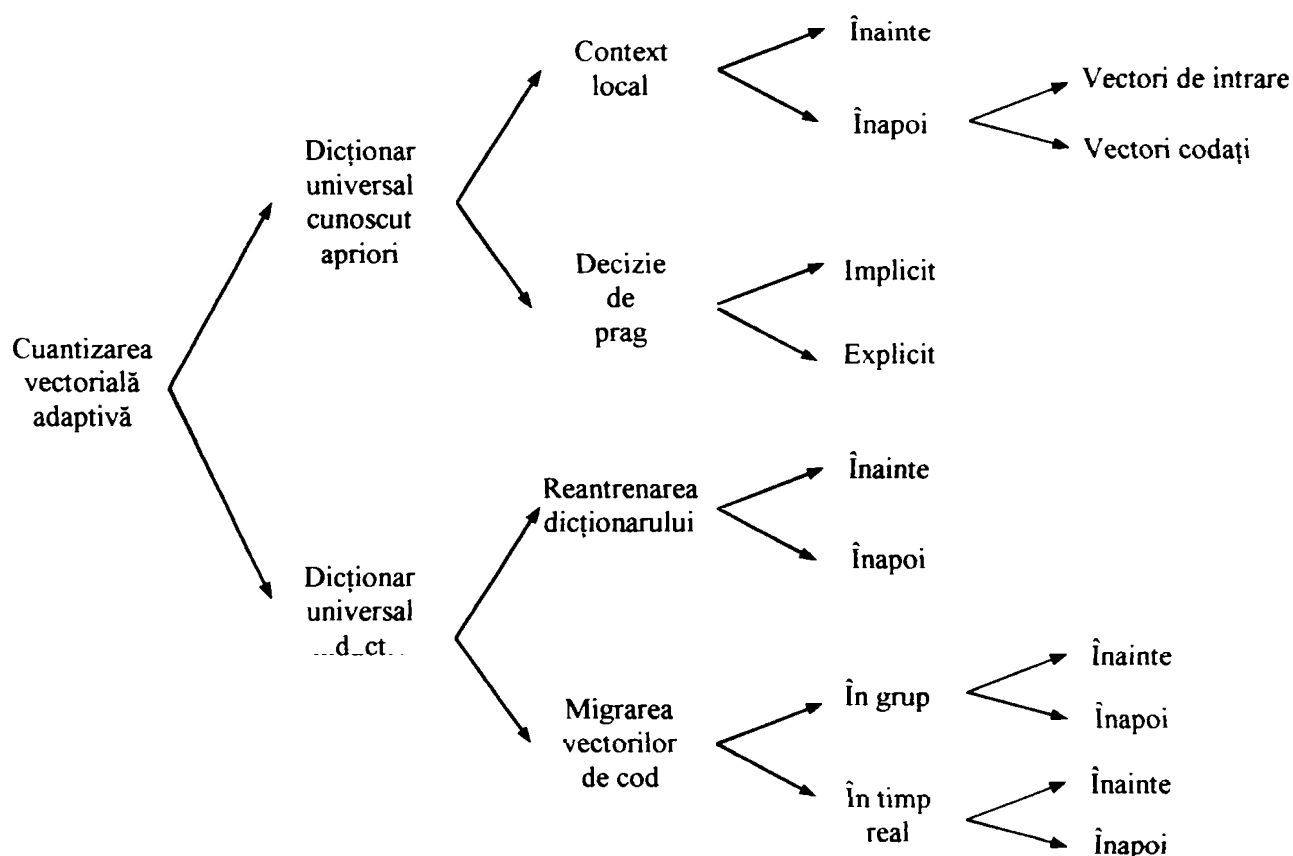


Fig. 4.14. Clasificarea metodelor de cuantizare vectorială adaptivă



Cuantizarea cu *dictionar universal inductiv* se bazează pe algoritmi de antrenare a unui dictionar prezentați în cap. 1, astfel încât să se asigure cel mai bun dictionar pentru cuantizarea semnalului. În cazul reantrenării dictionarului [WGS91], acesta este periodic complet reînnoit prin reantrenare, operație care se poate face pe baza unei secvențe de vectori trecută sau viitoare relativ la vectorul curent. Migrarea vectorilor de cod se poate aplica dacă statistica semnalului se modifică foarte lent, astfel încât la intervale corespunzător alese dictionarul este reactualizat prin antrenare, folosind un set de antrenare local și un algoritm de antrenare din cele prezentate, sau este reactualizat în timp real, concomitent cu procesul de codare.

Se poate vorbi de două categorii distincte de adaptare:

- adaptarea înainte (forward) sau adaptarea la ieșire;
- adaptarea înapoi (backward) sau adaptarea la intrare.

Adaptarea înapoi presupune folosirea unor informații despre vectorii deja codați pentru a coda vectorul curent, exploatând informațiile despre acești vectori în scopul codării mai eficiente a vectorului curent. Nu este necesară transmiterea nici unei informații suplimentare în afara vectorilor anteriori, care sunt deja cunoscuți atât la codare cât și la decodare.

Adaptarea înainte presupune codarea vectorului curent folosind unele informații extrase din secvența viitoare de vectori, relativă la vectorul curent. Dezavantajul este că aceste informații trebuie extrase și transmise decodorului alături de informațiile despre vectorul curent, astfel încât decodarea să se poată face corect, deoarece decodorul încă nu cunoaște secvența de vectori viitori. La momentul de timp la care se codează vectorul curent este necesar să se cunoască deja secvența viitoare de vectori, pentru a fi posibilă extragerea acestor informații. Pe de o parte, prin aceasta se introduce o întârziere la codare, deoarece codarea unui vector nu este posibilă înainte de a cunoaște un anumit număr de vectori viitori relativ la vectorul curent. Pe de altă parte, informațiile extrase din secvența viitoare de vectori nu se pot obține la decodare, deoarece această secvență nu este încă cunoscută și de aceea, aceste informații trebuiesc codate și transmise separat la decodor, ceea ce complică în general schema și crește rata totală de biți.

Unele din variantele de cuantizare adaptivă prezentate în fig. 4.14 se caracterizează prin rată de bit variabilă, datorită faptului că în anumite momente de timp este necesar să se transmită la decodor informații suplimentare privind dictionarul.

#### 4.8.4.1 Modelul cuantizorului adaptiv

Un model matematic al cuantizorului adaptiv este reprezentat schematic în figura 4.15.

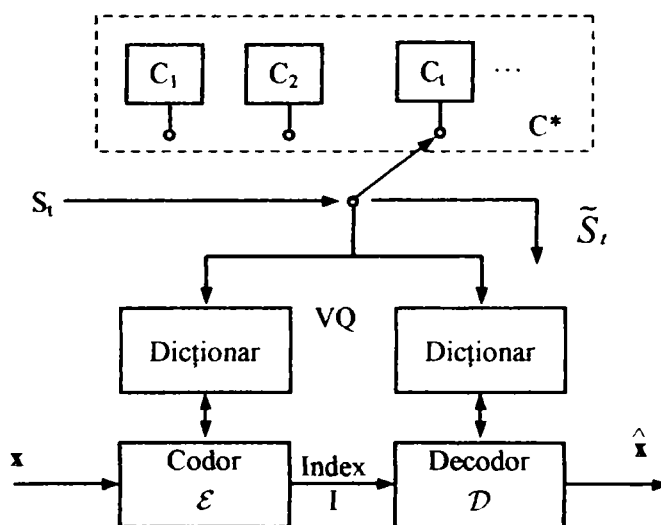


Fig. 4.15. Modelul general al cuantizorului vectorial adaptiv



Modelul din fig. 4.15 presupune existența unui dicționar universal,  $\mathcal{C}^* \subseteq \mathcal{R}^k$ , de mărime corespunzătoare, eventual chiar infinită, care este fix pentru orice moment de timp  $t$  și a unor dicționare momentane,  $\mathcal{C}_t \subset \mathcal{C}^*$  [CBA03-4].

Cuantizorul vectorial adaptiv va realiza în acest caz o asociere între un vector din spațiul euclidian  $k$  – dimensional și dicționarul activ în acel moment:

$$Q_t: \mathcal{R}^k \rightarrow \mathcal{C}_t \quad (4.66)$$

Presupunerea că există un dicționar de mărime suficientă (chiar infinit) asigură existența unui vector de cod optimal pentru orice vector de intrare. Funcția selector  $S_t$  realizează alegerea dicționarului local  $\mathcal{C}_t$  care se va folosi la momentul de timp  $t$ .

Dicționarul universal folosit în cuantizorul vectorial poate fi inclus explicit în cuantizor, caz în care trebuie să fie de mărime finită, sau poate să nu fie inclus implicit în cuantizor, caz în care poate fi și de mărime infinită. În cel mai general caz, dicționarul universal poate cuprinde întreg spațiul  $\mathcal{R}^k$ .

Se cunosc în literatură două modele de dicționar universal [RRG98], [ZBL94], [CBA03-4]:

a) *Dicționarul universal cunoscut apriori*, care este disponibil simultan atât la codare, cât și la decodare,

b) *Dicționarul universal inductiv* are la bază ideea că nici codorul nici decodorul nu cunosc apriori dicționarul, ci îl construiesc pe parcursul codării [GMY85].

Actualizarea dicționarului se poate face în diferite moduri specifice cuantizării vectoriale sau folosind rețele neuronale (învățarea Kohonen).

#### 4.8.5 Cuantizarea vectorială cu rată de bit variabilă

Toate schemele de cuantizare prezentate până acum sunt scheme cu rată de bit constantă, adică generează un număr fix de biți pentru fiecare simbol (vector) de intrare, sau, într-o formulare echivalentă, generează un număr constant de biți pe unitatea de timp.

Ideea de bază a cuantizării cu rată de bit variabilă constă în a aloca cuvinte de cod scurte pentru cuantizarea vectorilor de intrare cu probabilitate mare de apariție și cuvinte de cod mai lungi pentru cuantizarea vectorilor de intrare cu probabilitate mică de apariție. De exemplu, pentru cuantizarea semnalului vocal, se vor aloca mai mulți biți pentru semnalele sonore, cu spectru bogat, și mai puțini biți pentru semnalele nesonore sau pauze, cu un spectru mai apropiat de zgomot. Cuantizarea la rate de bit variabile include, ca un caz particular, cuantizarea la rată fixă de bit și, din acest motiv, este de așteptat ca să se reducă constrângerile impuse asupra cuantizorului și să se obțină performanțe generale mai bune.

Multe dintre metodele de cuantizare prezentate se pretează la o adaptare în vederea obținerii unei rate variabile, ca o generalizare. Pentru semnalele care au anumite caracteristici puternic variabile în timp (de exemplu semnalul vocal), se poate aplica o metodă de cuantizare din cele prezentate, care să conducă la o rată variabilă ca o consecință a ratelor diferite ce se pot aplica la diferite momente de timp [MCG97], [CLS01], [CBA03-2], [HPG99], [DRG96].

Există însă și metode directe de cuantizare vectorială cu rată variabilă.

Astfel, considerând o secvență de valori scalare, aceste valori se pot grupa astfel încât să formeze vectori, dar, ținând cont de caracteristicile locale ale semnalului, se vor stabili diferite dimensiuni  $k_i$  pentru vectorii  $\mathbf{x}_i$ , rezultând o secvență de vectori cu dimensiune variabilă. Această secvență poate fi descrisă de perechi  $\{(k_i, \mathbf{x}_i)\}$  și dacă se presupune că dimensiunile posibile ale vectorilor din secvență fac parte dintr-un set de valori întregi  $K = \{m_1, m_2, \dots, m_l\}$ , atunci această secvență de vectori de dimensiune variabilă poate fi cuantizată cu un set de

cuantizoare vectoriale, folosind regula clasică a celui mai apropiat vecin (NN) și un set de dicționare, câte unul pentru fiecare dimensiune  $m \in K$  și pentru fiecare dicționar, o măsură a distorsiunii  $d(x, \hat{x}, m)$ , dependentă de  $m$ . Setul de dicționare va cuprinde  $L$  dicționare, de forma  $C_j = \{c_{j,i}, i=1, 2, \dots, N_j\}$  pentru  $j=1, 2, \dots, L$ , iar codorul va genera un index care va identifica dicționarul folosit precum și un index care va identifica cuvântul ales din dicționarul respectiv. Această metodă asigură o rată variabilă chiar în cazul particular în care toate dicționarele au aceeași dimensiune, datorită dimensiunii diferite a vectorilor de intrare.

Se pot alege în principiu diferite metode particulare, cu constrângeri, pentru fiecare dimensiune de vector. Cuantizarea arborescentă, datorită folosirii aproximărilor succesive, se pretează la o modificare în sensul obținerii unei rate variabile.

Un optim, prin care să se asigure distorsiunea minimă pentru o rată de bit impusă, se poate obține prin algoritmul BFOS, bazat pe metodele de alocare optimă a biților [GER93].

În literatură se prezintă și alte metode de cuantizare vectorială, dintre care se pot aminti:

- cuantizorul vectorial cu stări finite [YHF93],
- cuantizorul vectorial trellis [SCH98], [MFG90], [RMG90], [MFS93], [MTF93], [NSM93],
- cuantizorul vectorial latice [CSL82-1], [CSL82-2],

precum și diferite combinații între metodele prezentate anterior, dezvoltate pentru aplicații specifice, cu scopul obținerii unei cuantizări optime pentru o anumită categorie de semnale reprezentate prin vectori [CLG89], [LIN00], [HSK00].

## Capitolul 5

# Calculul și cuantizarea vectorială a coeficienților de predicție

### 5.1 Coeficienții de predicție

În codorul CELP, ca de altfel în orice tip de codor care folosește predicția liniară, coeficienții  $\{a_k\}$  ai filtrului de predicție descris de rel. (3.7) (Cap.3) sunt estimați pentru fiecare cadru de semnal prin una din metodele prezentate și apoi sunt cuantizați și transmiși la decodor.

Coeficienții filtrului de predicție se determină prin tehnicile de predicție liniară devenite clasice: metoda corelației (algoritmul Levinson – Durbin) și metoda covarianței (algoritmul Schur) [STA98], [KNT87], [PAK95].

Actualizarea coeficienților de predicție se face periodic, ținând cont că, pentru a reprezenta o bună aproximare a spectrului, pe această perioadă parametrii tractului vocal trebuie să fie aproximativ neschimbați.

În acest scop, semnalul vocal se împarte în cadre și se calculează coeficienții de predicție o dată pentru fiecare cadru, considerând semnalul staționar pe durata unui cadru. Durata tipică pentru un cadru este cuprinsă între 10 și 30 ms, durată care corespunde cu un număr de 80 – 240 eșantioane, pentru frecvența de eșantionare de 8 kHz. Semnalul de intrare corespunzător unui cadru este ponderat cu o funcție fereastră care are rolul de a atenua efectul ce apare la marginile cadrului.

Parametrii de predicție se calculează, conform celor prezentate în cap. 3, considerând că eșantionul curent este o combinație liniară a unui număr relativ redus de eșantioane precedente (8 – 16), justificând în acest fel denumirea de *predicție pe termen scurt* STP (short term prediction).

Pentru a modela cât mai exact comportarea în timp a tractului vocal, care își modifică lent parametrii de la un cadru la altul, pe când parametrii de predicție prezintă salturi de la un cadru la altul, se aplică o tehnică de interpolare. Fiecare cadru este împărțit în acest scop în mai multe subcadre, uzual între 2 și 4. În fiecare subcadru se folosesc valori diferite ale coeficienților de predicție calculate fie prin interpolare, fie ca o medie ponderată a valorilor din cadrul curent și a valorilor din cadrul precedent, ponderea valorilor din cadrul precedent fiind mare la începutul cadrului curent și scăzând progresiv, astfel încât în ultimul subcadru să se folosească doar valorile din cadrul curent [[ETS97-1]. În caz contrar, modificările bruște ale valorilor coeficienților LP din cadre vecine sunt percepute în semnalul de ieșire ca un brum (denumit în literatură *ringing*) cu frecvența egală cu  $F_c / N$ , unde  $F_c$  este frecvența de eșantionare și  $N$  este numărul de eșantioane dintr-un cadru.

Numărul  $p$  al coeficienților de predicție utilizați determină precizia de reprezentare a spectrului. Uzual,  $p$  este cuprins între 8 și 16 și este corelat cu numărul de formanți prezenți în spectrul semnalului. Se consideră practic aproximativ un formant pentru fiecare 1kHz lărgime de bandă a semnalului și respectiv câte 2 – 3 coeficienți pentru fiecare formant. Rezultă astfel 8 – 12 coeficienți de predicție pentru un semnal cu banda limitată la 4 kHz (banda telefonică).

Mai multe standarde de codare [ETS97-1], [ETS97-2], [ITU96-1] au adoptat valoarea  $p = 10$ . Această valoare va fi folosită în toate experimentările care urmează.

În figura 5.1 se prezintă forma de variație în timp a semnalului vocal corespunzător cuvântului „five” și alăturat, pentru cadrul de semnal cuprins între eșantioanele 1500 și 1800, spectrul calculat cu transformata Fourier rapidă FFT, suprapus peste reprezentarea grafică a caracteristicii filtrului de predicție  $A(z)$ , observându-se cum aproximează filtrul de predicție spectrul semnalului.

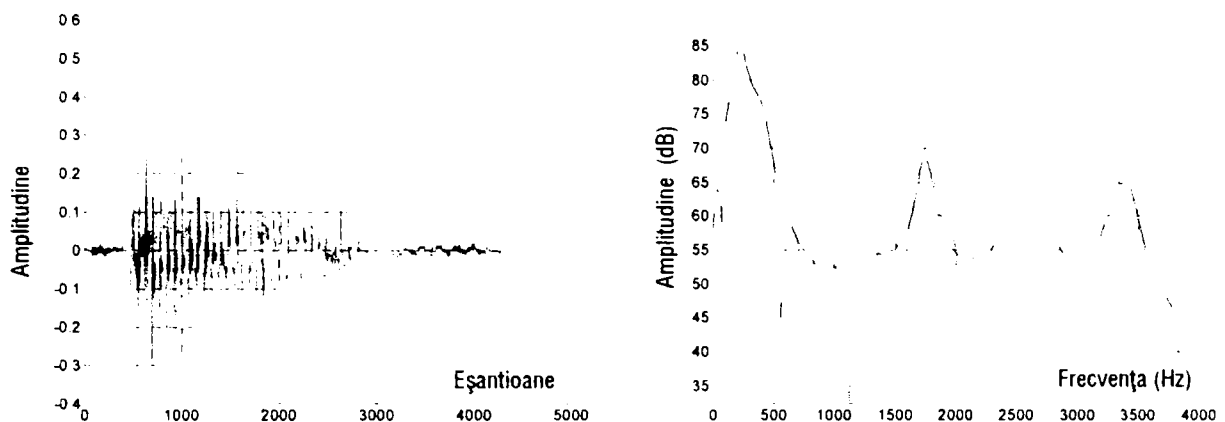


Fig. 5.1. Semnalul vocal și spectrele FFT și LPC

Coeficienții LPC din rel. 3.6 nu au proprietăți adecvate pentru a se cuantiza și transmite direct. În cap. 3 au fost prezentate pe scurt câteva reprezentări alternative ale coeficienților de predicție. Datorită proprietăților favorabile, am ales perechile de linii spectrale LSP, pentru care am elaborat o metodă eficientă de calcul și două variante de cuantizare vectorială.

### 5.1.1 Perechile de linii spectrale LSP

Pentru a defini LSP, considerăm polinomul de predicție liniară de ordinul  $M$ :

$$A_M(z) = 1 + \sum_{k=1}^M a_k z^{-k}, \quad (5.1)$$

cu coeficienții de predicție  $a_k$ ,  $k = 1, 2, \dots, M$ , și construim două noi polinoame de grad  $M+1$  astfel:

$$F_1(z) = A_M(z) + z^{-(M+1)} A_M(z^{-1}), \quad (5.2)$$

$$F_2(z) = A_M(z) - z^{-(M+1)} A_M(z^{-1}). \quad (5.3)$$

Aceste două polinoame sunt determinate de  $M+1$  coeficienți. Dezvoltând expresiile (5.2) și (5.3) obținem:

$$F_1(z) = 1 + (a_1 + a_M)z^{-1} + (a_2 + a_{M-1})z^{-2} + \dots + (a_1 + a_M)z^{-M} + z^{-M+1} \quad (5.4)$$

$$F_2(z) = 1 + (a_1 - a_M)z^{-1} + (a_2 - a_{M-1})z^{-2} + \dots - (a_1 - a_M)z^{-M} - z^{-M+1}. \quad (5.5)$$

Notăm coeficienții polinoamelor  $F_1(z)$  și  $F_2(z)$  cu  $f_i^1$  și respectiv  $f_i^2$ , indicele superior indicând polinomul căruia îi aparțin, și astfel putem scrie:

$$F_1(z) = \sum_{i=0}^{M+1} f_i^1 z^{-i}, \quad (5.6)$$

$$F_2(z) = \sum_{i=0}^{M+1} f_i^2 z^{-i}. \quad (5.7)$$

Comparând (5.4) și (5.5) cu (5.6) și (5.7) se pot scrie următoarele relații pentru calculul coeficienților polinoamelor  $F_1(z)$  și  $F_2(z)$ :

$$f_0^1 = f_{M+1}^1 = 1, \quad (5.8)$$

$$f_i^1 = f_{M-i+1}^1 = a_i + a_{M-i+1}, \quad (5.9)$$

$$f_0^2 = -f_{M+1}^2 = 1, \quad (5.10)$$

$$f_i^2 = -f_{M-i+1}^2 = a_i - a_{M-i+1}, \quad (5.11)$$

pentru  $i = 1, \dots, M$ .

Structura coeficienților justifică denumirea de polinom simetric pentru  $F_1(z)$  și polinom antisimetric pentru  $F_2(z)$ .

Fiecare din cele două polinoame are în total  $M+1$  rădăcini și în funcție de  $M$ , cele două polinoame au câte un zero trivial de valoare  $z = \pm 1$ , astfel:

- $M$  par:  $F_1(z)$  are rădăcina  $z = -1$ ,  $F_2(z)$  are rădăcina  $z = 1$ ,
- $M$  impar:  $F_2(z)$  are rădăcina  $z = \pm 1$ .

Aceste zerouri nu prezintă interes practic și se pot elimina, pentru a scădea gradul polinoamelor, prin împărțirea acestora cu o cantitate adecvată astfel:

$$P(z) = \frac{F_1(z)}{1+z^{-1}}, \quad Q(z) = \frac{F_2(z)}{1-z^{-1}}, \quad \text{pentru } M \text{ par}, \quad (5.12)$$

$$P(z) = F_1(z), \quad Q(z) = \frac{F_2(z)}{1-z^{-2}}, \quad \text{pentru } M \text{ impar}. \quad (5.13)$$

Pentru cazul  $M$  par, cele două polinoame din rel. (5.12) se pot scrie astfel:

$$P(z) = \sum_{i=1}^M p_i z^{-i} \quad (5.14)$$

$$Q(z) = \sum_{i=1}^M q_i z^{-i} \quad (5.15)$$

și efectuând împărțirile conform rel. (5.12), se pot deduce următoarele relații pentru calculul coeficienților polinoamelor (5.14):

$$p_0 = 1, \quad (5.16)$$

$$p_i = f_i^1 - p_{i-1}, \quad \text{pentru } i = 1, 2 \dots M, \quad (5.17)$$

și respectiv (5.15):

$$q_0 = 1, \quad (5.18)$$

$$q_i = f_i^2 + q_{i-1}, \quad \text{pentru } i = 1, 2 \dots M, \quad (5.19)$$

Pentru cazul  $M$  impar, cele două polinoame din rel. (5.13) se pot scrie astfel:

$$P(z) = \sum_{i=1}^{M+1} p_i z^{-i} = F_1(z) = \sum_{i=0}^{M+1} f_i^1 z^{-i}, \quad (5.20)$$

$$Q(z) = \sum_{i=1}^{M-1} q_i z^{-i} \quad (5.21)$$

și, în mod similar, se pot deduce expresiile pentru coeficienții polinoamelor (5.20):

$$p_i = f_i^1, \text{ pentru } i = 0, 1, 2, \dots, M-1, \quad (5.22)$$

și respectiv (5.21):

$$q_0 = 1, \quad (5.23)$$

$$q_1 = f_1^2, \quad (5.24)$$

$$q_i = f_i^2 + q_{i-2} \text{ pentru } i = 2, 3, \dots, M-1. \quad (5.25)$$

Atât  $P(z)$  cât și  $Q(z)$  sunt polinoame de ordin par, indiferent de valoarea lui  $M$ , astfel că se poate nota cu  $2M_1$  gradul polinomului  $P(z)$  și respectiv cu  $2M_2$  gradul polinomului  $Q(z)$ , unde:

$$M_1 = M_2 = M/2, \text{ pentru } M \text{ par}, \quad (5.26)$$

$$M_1 = (M+1)/2, \quad M_2 = (M-1)/2, \text{ pentru } M \text{ impar}, \quad (5.27)$$

Se poate demonstra că cele două polinoame  $P(z)$  și  $Q(z)$  au toate rădăcinile de modul egal cu 1, adică situate pe cercul unitate [KLE99], [CHU03-2]. Deoarece cele două polinoame au coeficienți reali, rădăcinile vor forma perechi complex conjugate. Astfel,  $P(z)$  va avea  $M_1$  perechi de rădăcini complex conjugate, iar  $Q(z)$  va avea  $M_2$  perechi de rădăcini complex conjugate, toate situate pe cercul unitate.

Evaluând cele două polinoame pe cercul unitate, adică pentru  $z = e^{j\omega}$ , rezultă:

$$\begin{aligned} P(e^{j\omega}) &= \sum_{i=0}^{2M_1} p_i e^{-ij\omega} = \\ &= e^{-jM_1\omega} (p_0 e^{j\omega M_1} + p_1 e^{j\omega(M_1-1)} + p_2 e^{j\omega(M_1-2)} + \dots + p_{2M_1} e^{-j\omega M_1}) = \\ &= e^{-jM_1\omega} [2p_0 \cos(M_1\omega) + 2p_1 \cos((M_1-1)\omega) + \dots + 2p_{M_1-1} \cos \omega + p_{M_1}] \end{aligned} \quad (5.28)$$

$$\begin{aligned} Q(e^{j\omega}) &= \sum_{i=0}^{2M_2} q_i e^{-ij\omega} = \\ &= e^{-jM_2\omega} (q_0 e^{j\omega M_2} + q_1 e^{j\omega(M_2-1)} + q_2 e^{j\omega(M_2-2)} + \dots + q_{2M_2} e^{-j\omega M_2}) = \\ &= e^{-jM_2\omega} [2q_0 \cos(M_2\omega) + 2q_1 \cos((M_2-1)\omega) + \dots + 2q_{M_2-1} \cos \omega + q_{M_2}]. \end{aligned} \quad (5.29)$$

Perechile de linii spectrale LSP sunt definite ca acele valori ale frecvențelor care corespund zerourilor polinoamelor  $P(z)$  și  $Q(z)$  situate în semiplanul superior al planului  $z$ :

$$\{\omega \mid P(e^{j\omega}) = 0 \text{ sau } Q(e^{j\omega}) = 0; \quad 0 < \omega < \pi\}, \quad (5.30)$$

adică sunt rădăcinile polinoamelor:

$$P_0(\omega) = 2 \cos(M_1\omega) + 2p_1 \cos((M_1-1)\omega) + \dots + 2p_{M_1-1} \cos \omega + p_{M_1} \quad (5.31)$$

$$Q_0(\omega) = 2 \cos(M_2\omega) + 2q_1 \cos((M_2-1)\omega) + \dots + 2q_{M_2-1} \cos \omega + q_{M_2}. \quad (5.32)$$

unde s-a ținut seama că  $p_0 = q_0 = 1$ , conform (5.16) și (5.18), respectiv (5.22) și (5.23).

În concluzie, este posibil să se reprezinte polinomul  $A(z)$  cu ajutorul a  $M$  valori care reprezintă unghiuri cuprinse între 0 și  $\pi$ ,  $0 < \omega_1 < \omega_2 < \dots < \omega_M < \pi$ , și care se numesc perechi de



linii spectrale LSP. Valorile LSP sunt asociate cu rădăcinile polinoamelor  $P(z)$  și  $Q(z)$  prin relația:  $\omega_i = \arg(z_i)$ .

Frecvențele corespunzătoare LSP sunt denumite LSF și se pot calcula prin:  $f_i = \omega_i / 2\pi$ .

LSF păstrează aceeași relație de ordine ca și LSP, fiind cuprinse în intervalul  $0 - 0,5$ , exprimat în frecvențe relative față de frecvența de eșantionare. Pentru o frecvență de eșantionare de 8 kHz, LSF sunt cuprinse între 0 și 4 kHz.

### 5.1.2 Proprietăți ale perechilor de linii spectrale

Perechile de linii spectrale LSP au câteva proprietăți utile [SMF93], [EPB95], [KLE99], [KSL00], [KBA03], [BAK03], [KHL00], dintre care cele mai importante sunt următoarele:

*Proprietatea 1:* Rădăcinile polinoamelor  $P(z)$  și  $Q(z)$  se află situate pe cercul unitate și numărul rădăcinilor situate în intervalul deschis  $0 < \arg z < \pi$  este  $M$ .

*Proprietatea 2:* Rădăcinile polinomului  $P(z)$  sunt intercalate pe cercul unitate cu rădăcinile polinomului  $Q(z)$ .

*Proprietatea 3:* Polinomul  $A(z)$  este de fază minimă dacă și numai dacă sunt satisfăcute proprietățile 1 și 2.

*Proprietatea 4:* Pentru orice  $i = 1, \dots, M$ , parametrul LSP  $\omega_i$  al modelului de ordin  $M$  este situat în intervalul deschis  $(\omega_{M-1,i-1}, \omega_{M-1,i})$ , sau altfel spus, parametrii LSP ai modelului de ordin  $M-1$  sunt intercalați pe cercul unitate cu parametrii modelului de ordin  $M$ .

Proprietatea 4 justifică observația că nici una din valorile parametrilor LSP ale modelului de ordin  $M$  nu coincide cu nici una dintre valorile LSP ale modelului de ordin inferior  $M-1$ .

Primele trei proprietăți sunt esențiale pentru cuantizare. Astfel, prima proprietate garantează existența LSP. Modul efectiv de calcul va fi prezentat ulterior, în cadrul acestui capitol.

A doua proprietate stabilește o relație de ordine între cele  $M$  valori care reprezintă parametrii LSP, și anume:

$$0 < \omega_1 < \omega_2 < \dots < \omega_M < \pi. \quad (5.33)$$

Această proprietate este în concordanță cu semnificația fizică a frecvențelor LSF asociate LSP, care reprezintă frecvențele naturale de rezonanță ale tractului vocal.

Pentru modelul electric al tractului vocal, LSF reprezintă frecvențele de rezonanță în gol și respectiv în scurtcircuit.

Din acest motiv, este evidentă relația de ordine și de limitare din cauze naturale ale acestor frecvențe. Distribuția fiecărei valori LSF este restrânsă la un domeniu limitat și nu se poate extinde la întreg domeniul de valori, ceea ce conduce la posibilitatea cuantizării valorilor LSF sau LSP folosind un număr mai mic de biți.

Proprietatea de ordine exprimată prin (5.33) asociată cu proprietatea 3 oferă o metodă eficientă de verificare a stabilității filtrului de sinteză LPC obținut la decodor din parametrii cuantizați.

Filtrul de sinteză este stabil dacă este de fază minimă, adică dacă se păstrează, și după cuantizare, aceeași relație de ordine (5.33) ca și înainte de cuantizare.

Parametrii LSP prezintă selectivitate spectrală, adică o perturbare a valorii parametrului  $\omega_i$  va afecta spectrul doar într-un domeniu învecinat frecvenței  $f_i$  [CHU03-2]

Aceste motive au făcut ca LSP să fie alese pentru a reprezenta coeficienții de predicție.

În figura 5.2 sunt reprezentate valorile LSF suprapuse peste spectrul LPC al unui cadru de semnal.

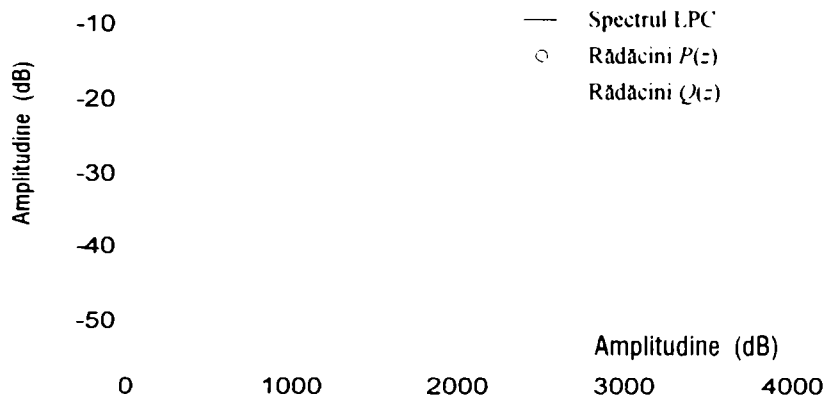


Fig. 5.2. Parametrii LSF și spectrul LPC

## 5.2 Calculul LSP

### 5.2.1 Calculul LSP în cazul general

Pentru calculul efectiv al LSP în cazul general, cunoscând coeficienții de predicție  $\{a_k\}$  determinați printr-o metodă adecvată, se calculează coeficienții (5.8) – (5.11) și apoi coeficienții (5.16) – (5.19) dacă  $M$  este par, respectiv coeficienții (5.22) – (5.25) dacă  $M$  este impar, obținând cele două ecuații de forma (5.31) și (5.32):

Se cunosc mai multe metode pentru a rezolva ecuațiile (5.31) și (5.32).

Este avantajos să se transforme aceste două expresii care conțin funcția trigonometrică cosinus pentru unghiul  $\omega$  și multiplii săi în expresii algebrice care să nu conțină funcții trigonometrice. Pentru aceasta, se poate face substituția:

$$x = \cos(\omega) \quad (5.34)$$

și având în vedere polinoamele Cebîșev de forma  $\cos(m\omega) = T_m(x)$ , care satisfac relația de recurență:

$$T_n(x) = \cos(n\omega) = 2xT_{n-1}(x) - T_{n-2}(x), \quad (5.35)$$

expresiile (5.31) și (5.32) devin expresii algebrice în  $x$  [KAR86], [JRW99].

Pentru rezolvarea ecuațiilor care rezultă, se pot folosi două metode diferite:

- Evaluarea polinoamelor (5.31) și (5.32) pentru valori ale argumentului între 0 și  $\pi$ , cu un pas adecvat, urmărind schimbările de semn, până se găsesc toate cele  $M_1$  rădăcini ale (5.31) și cele  $M_2$  rădăcini ale (5.32),
- Evaluarea polinoamelor (5.31) și (5.32) pentru valori ale argumentului între 0 și  $\pi$ , cu un pas adecvat, urmărind schimbările de semn; după fiecare rădăcină găsită se divide polinomul cu  $x - x_i$  scăzând cu o unitate gradul acestuia, iar când se ajunge la gradul 4, se pot aplica formule algebrice, deoarece este sigur că polinomul are rădăcini reale distincte [WUC97].

Pasul de evaluare determină precizia de calcul a rădăcinilor polinoamelor, adică a LSP. Acest pas se alege în concordanță cu precizia operației de cunantizare a LSP, fiind inutilă o precizie de calcul ridicată, care conduce la un volum ridicat de calcule, pentru ca apoi precizia să se piardă printr-o cuantizare a valorilor calculate cu un număr relativ redus de biți.

Dacă sunt cunoscute valorile cuantizate ale parametrilor LSP, este posibil să se reducă volumul de calcule, efectuând evaluarea polinoamelor numai pentru valorile cuprinse în tabela de cuantizare.

### 5.2.2 Optimizarea calculului LSP pentru ordinul 10

Pentru cazul  $M = 10$ , folosind relația de recurență (5.35), primele 5 polinoame Cebîșev sunt:

$$T_0(x) = \cos(0) = 1, \quad (5.36)$$

$$T_1(x) = \cos(\omega) = x, \quad (5.37)$$

$$T_2(x) = \cos(2\omega) = 2x^2 - 1, \quad (5.38)$$

$$T_3(x) = \cos(3\omega) = 4x^3 - 3x, \quad (5.39)$$

$$T_4(x) = \cos(4\omega) = 8x^4 - 8x^2 + 1, \quad (5.40)$$

$$T_5(x) = \cos(5\omega) = 16x^5 - 20x^3 + 5x. \quad (5.41)$$

și înlocuind (5.36) – (5.41) în (5.31) și (5.32) am obținut pentru polinoamele  $P$  și  $Q$  forma:

$$P_{10}(x) = 32x^5 + 16p_1x^4 + (8p_2 - 40)x^3 + (4p_3 - 16p_1)x^2 + (10 - 6p_2 + 2p_4)x + (2p_1 - 2p_3 + p_5), \quad (5.42)$$

$$Q_{10}(x) = 32x^5 + 16q_1x^4 + (8q_2 - 40)x^3 + (4q_3 - 16q_1)x^2 + (10 - 6q_2 + 2q_4)x + (2q_1 - 2q_3 + q_5). \quad (5.43)$$

Pentru a rezolva ecuațiile  $P_{10}(x) = 0$  și  $Q_{10}(x) = 0$  se poate aplica metoda menționată anterior, de a evalua cele două polinoame pentru a evidenția schimbarea de semn.

O reducere a volumului de calcul este posibilă dacă s-ar putea izola rădăcinile prin încadrare în intervale închise, sugerată de proprietatea de ordine a LSP.

O ilustrare intuitivă a acestei idei este dată în figura 5.3.

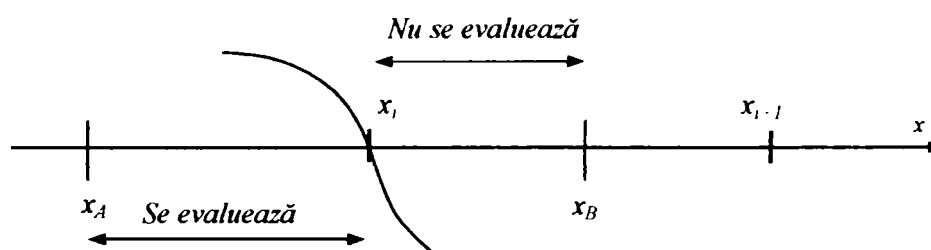


Fig. 5.3. Delimitarea rădăcinilor

Dacă se caută rădăcinile unui polinom prin evaluarea sa pe axa reală, și se cunoaște faptul că rădăcina  $x_i$  este cuprinsă în intervalul deschis  $x_{i-1} - x_B$ , atunci se evaluează polinomul începând din punctul  $x_A$ , până la prima schimbare de semn, care determină rădăcina  $x_i$ . Pentru a determina rădăcina  $x_{i+1}$ , polinomul se va evalua începând cu punctul  $x_B$ , eliminând din operația de evaluare intervalul  $x_i - x_B$ .

Pentru cazul polinoamelor  $P(x)$  și  $Q(x)$  este posibil să se aplice această metodă de reducere a volumului de calcule.

În figura 5.4 am reprezentat pentru un caz real evoluția celor două polinoame pentru  $M = 10$ , care, în acest caz sunt fiecare de gradul 5 și deci au câte 5 rădăcini reale și distincte în intervalul  $0 \div \pi$ .

Se pot identifica pe axa  $x$  un număr de cinci intervale, determinate de cele patru puncte în care  $P(x) = Q(x)$ . Fiecare interval conține doar o rădăcină a polinomului  $P(x)$  și o rădăcină a polinomului  $Q(x)$ .

Ținând cont că intervalele menționate sunt delimitate de punctele în care  $P(x) = Q(x)$ , ele vor fi delimitate de rădăcinile ecuației  $P(x) - Q(x) = 0$  și extremele intervalului  $0 \div \pi$  care cuprinde rădăcinile polinoamelor  $P(x)$  și  $Q(x)$ .

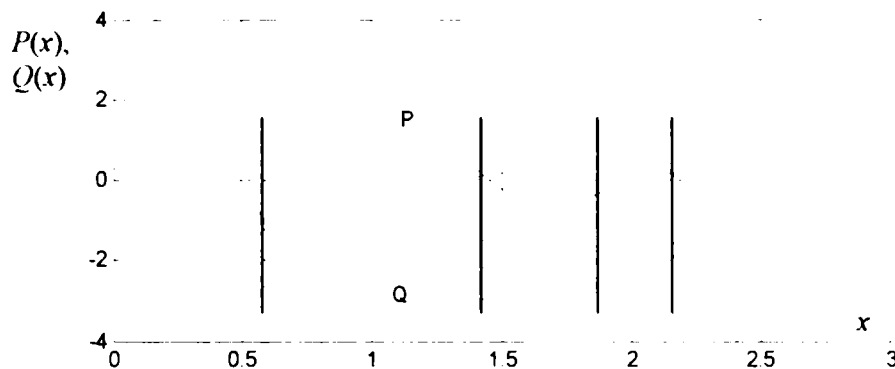


Fig. 5.4. Rădăcinile polinoamelor  $P$  și  $Q$

Am demonstrat această afirmație intuitivă într-un mod riguros [CBA05], pentru cazul general al predicției de ordin  $M$ . Pentru aceasta, am avut în vedere relațiile de recurență pe care le îndeplinesc polinomul de predicție  $A(z)$  și polinomul reciproc  $B(z)$ , relații care stau la baza algoritmului Levinson – Durbin:

$$A_0(z) = B_0(z) = 1, \tag{5.44}$$

$$A_m(z) = A_{m-1}(z) - k_m z^{-1} B_{m-1}(z), \tag{5.45}$$

$$B_m(z) = k_m A_{m-1}(z) + z^{-1} B_{m-1}(z), \tag{5.46}$$

pentru  $m = 1, \dots, M$ , unde  $B_m(z) = z^{-m} A_m(z)$  este polinomul reciproc al  $A(z)$ , indicele  $m$  indică ordinul predictorului iar  $k_m$  este coeficientul de reflexie, care are proprietatea de a avea amplitudinea limitată în intervalul deschis  $(-1, +1)$  dacă  $A(z)$  este de fază minimă [KNT87], [PRM92].

Diferența  $P_M(z) - Q_M(z)$  se poate calcula ținând cont de modul în care au fost definite aceste polinoame prin rel. (5.12) și (5.13), precum și de rel. (5.2) și (5.3). Grupând convenabil termenii, rezultă, pentru  $M$  par:

$$P_M(z) - Q_M(z) = 2 \frac{-z^{-1}(1 - k_M)F_2^{(M-1)}(z)}{1 - z^{-2}} = -2z^{-1}(1 - k_M)Q_{M-1}(z) \tag{5.47}$$

unde cu  $F_2^{(M-1)}(z)$  s-a notat funcția de forma (5.3), pentru ordinul de predicție  $M - 1$ .

Deoarece  $k_M$  este mai mic decât 1, rădăcinile ecuației  $P_M(z) - Q_M(z) = 0$  sunt date de rădăcinile ecuației  $Q_{M-1}(z) = 0$ . Această ecuație are  $\lfloor (M - 1)/2 \rfloor$  rădăcini reale și distincte, care determină  $\lfloor (M - 1)/2 \rfloor + 1$  intervale pe axa  $x$ , și conform proprietății 4 enunțate, în fiecare din aceste intervale se va găsi numai câte o rădăcină a polinomului  $P(z)$  și una a polinomului  $Q(z)$ .

Particularizând (5.47) pentru  $M = 10$  și revenind la necunoscuta  $x$  prin schimbarea de variabilă (5.34), se poate afirma că ecuația  $P_{10}(x) - Q_{10}(x) = 0$  are aceleași rădăcini ca și ecuația  $Q_9(x) = 0$ , care, fiind de gradul 4, are 4 rădăcini.

Pentru a verifica experimental această metodă de calcul, am evaluat polinoamele  $P(z)$  și  $Q(z)$  numai pentru valorile cuantizate ale parametrilor LSP, folosind în acest scop valorile cuantizate conform standardului FS 1016, prezentate în tabelul 5.1 (valori normate raportate la frecvența de eșantionare  $F_e = 8000$  Hz).

Tabelul 5.1. Cuantizarea LSP în standardul FS 1016

Index	LSP 1	LSP 2	LSP 3	LSP 4	LSP 5	LSP 6	LSP 7	LSP 8	LSP 9	LSP 10
0	0.0125	0.0262	0.0525	0.0775	0.1250	0.1838	0.2250	0.2781	0.3450	0.3988
1	0.0213	0.0294	0.0575	0.0825	0.1312	0.1962	0.2350	0.3000	0.3600	0.4088
2	0.0281	0.0331	0.0625	0.0900	0.1412	0.2112	0.2450	0.3156	0.3750	0.4188
3	0.0312	0.0369	0.0675	0.0994	0.1512	0.2288	0.2625	0.3312	0.3875	0.4275
4	0.0350	0.0406	0.0731	0.1100	0.1606	0.2500	0.2875	0.3500	0.4000	0.4362
5	0.0425	0.0450	0.0800	0.1212	0.1688	0.2750	0.3100	0.3688	0.4138	0.4488
6	0.0525	0.0500	0.0881	0.1350	0.1788	0.3000	0.3375	0.3938	0.4288	0.4638
7	0.0625	0.0550	0.0969	0.1462	0.1888	0.3250	0.3625	0.4188	0.4438	0.4788
8		0.0600	0.1062	0.1588	0.1988					
9		0.0650	0.1188	0.1713	0.2088					
10		0.0700	0.1312	0.1838	0.2188					
11		0.0762	0.1438	0.1962	0.2312					
12		0.0838	0.1562	0.2088	0.2438					
13		0.0925	0.1688	0.2212	0.2562					
14		0.1013	0.1812	0.2338	0.2688					
15		0.1100	0.1938	0.2462	0.2812					

Am implementat în Matlab metoda de reducere a volumului de calcul pe care am propus-o și în acest scop am realizat o subrutină de rezolvare a ecuației  $Q_9(x) = 0$ , având în vedere că are întotdeauna 4 rădăcini reale și diferite, precum și o subrutină de ordonare a acestor rădăcini în ordine crescătoare, pentru a determina cele 5 intervale. Față de evaluarea în toate punctele din tabelul 5.1, a rezultat o reducere a numărului de evaluări necesare de aprox. 32%, măsurată prin timpul de execuție.

Pentru a îmbunătăți precizia de determinare a unei valori LSP prin evaluarea polinomului  $P(x)$  în valorile cuantizate, în figura 5.5 am reprezentat evoluția acestui polinom în jurul rădăcinii exacte  $x_i$ , încadrată între două valori consecutive din șirul de valori cuantizate  $\hat{x}_{i-1}$  și  $\hat{x}_i$ .

Prin evaluarea polinomului în punctele succesive care reprezintă valorile cuantizate, se atribuie rădăcinii  $x_i$  valoarea  $\hat{x}_i$  la prima schimbare de semn, adică atunci când  $\text{sign}(P(\hat{x}_{i-1})) \neq \text{sign}(P(\hat{x}_i))$ . Este posibil ca prin convenție să se atribuie rădăcinii  $x_i$  valoarea  $\hat{x}_{i-1}$ , adică valoarea cuantizată imediat mai mică.

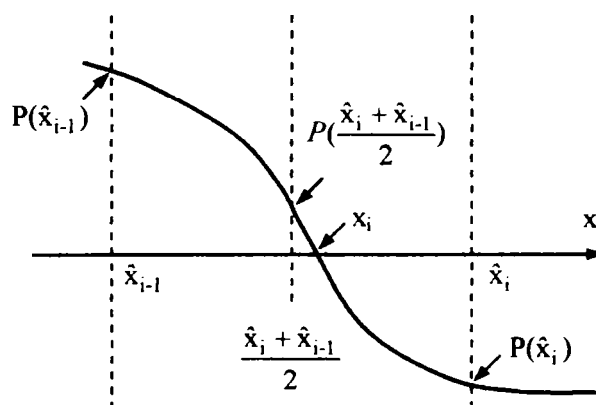


Fig. 5.5. Creșterea preciziei la evaluarea în punctele de cuantizare

Erorile făcute atribuind rădăcinii  $x_i$  valoarea cuantizată  $\hat{x}_i$  sau valoarea cuantizată  $\hat{x}_{i-1}$ , nu sunt egale și depind de forma de variație a polinomului  $P(x)$  în jurul valorii  $x_i$ .

Pentru a alege valoarea care asigură eroarea minimă se pot aplica criteriile de corecție suplimentare față de schimbarea de semn a polinomului și convenția de a atribui valoarea mai mică ( $\hat{x}_{i-1}$ ) sau mai mare ( $\hat{x}_i$ ).

Astfel, în standardul CELP FS 1016 se compară valoarea polinomului  $P(x)$  în punctele de cuantizare și atribuie valoarea  $\hat{x}_i$  pentru rădăcina  $x_i$ , dacă  $|P(\hat{x}_{i-1})| > |P(\hat{x}_i)|$ , iar în caz contrar atribuie acestei rădăcini valoarea  $\hat{x}_{i-1}$ .

Pentru a crește precizia, am evaluat polinomul la mijlocul intervalului format de cele două valori cuantizate  $\hat{x}_{i-1}$  și  $\hat{x}_i$  și am definit următoarea regulă de alocare simplă:

Se atribuie rădăcinii  $x_i$ , valoarea  $\hat{x}_i$  dacă

$$\text{sign}(P(\hat{x}_{i-1})) \neq \text{sign}(P(\hat{x}_i)) \text{ și } \text{sign}(P((\hat{x}_{i-1} + \hat{x}_i)/2)) \neq \text{sign}(P(\hat{x}_i)),$$

altfel se atribuie rădăcinii  $x_i$  valoarea  $\hat{x}_{i-1}$ .

În figura 5.6 este reprezentată situația care poate apare la determinarea LSP prin metoda evaluării polinoamelor în punctele de cuantizare, în cazul în care sunt considerate două valori LSP consecutive, prima, notată  $x_i$ , fiind rădăcină pentru  $P(x)$  și a doua, notată  $x_j$ , fiind rădăcină pentru  $Q(x)$ . Valorile cuantizate pentru rădăcinile celor două polinoame au fost notate cu indicele superior  $P$ , respectiv  $Q$ . În fig. 5.6 s-a ținut cont de faptul că pasul de cuantizare este în mod uzual diferit pentru cele două LSP consecutive.

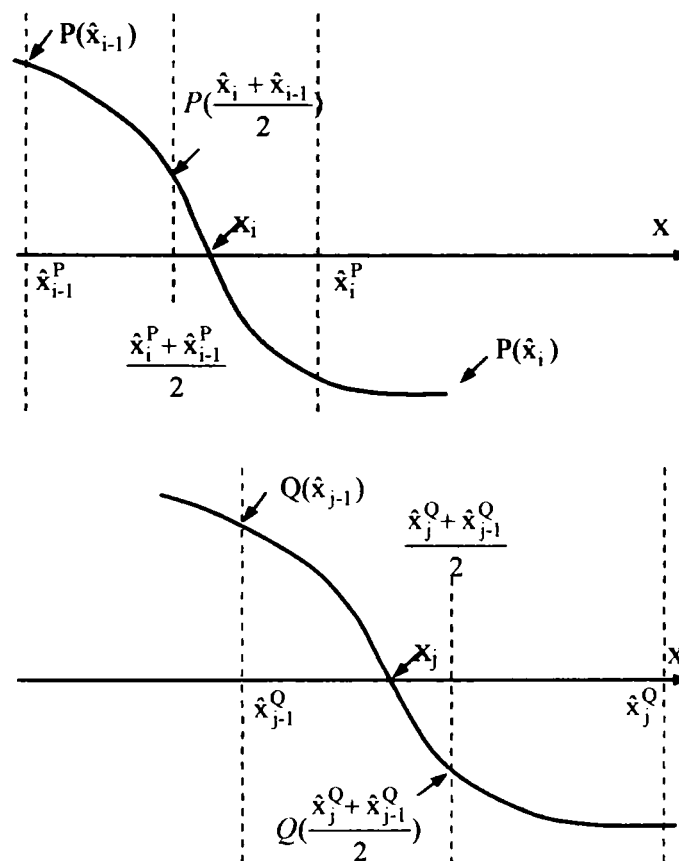


Fig. 5.6. Cuantizarea a două LSP consecutive

Aplicând metoda propusă anterior pentru cazul din fig. 5.6, se va atribui valoarea  $\hat{x}_i^P$  lui  $x_i$  și valoarea  $\hat{x}_{j-1}^Q$  lui  $x_j$ , rezultând o inversare a ordinii LSP, deoarece rădăcinile exacte sunt  $x_i < x_j$ , iar după cuantizare rezultă  $\hat{x}_i^P > \hat{x}_{j-1}^Q$ .

Sunt posibile două soluții simple:

- micșorarea valorii cuantizate pentru LSP mai mică,
- mărirea valorii cuantizate pentru LSP mai mare.



Standardul CELP FS 1016 aplică următoarea corecție:

Dacă  $|P(\hat{x}_{i-1})| + |Q(\hat{x}_{j-1})| > |P(\hat{x}_i)| + |P(\hat{x}_j)|$ , se atribuie  $[x_i, x_j] = [\hat{x}_i, \hat{x}_j]$ ,

altfel se atribuie  $[x_i, x_j] = [\hat{x}_{i-1}, \hat{x}_{j-1}]$

Similar cu situația cuantizării unei valori individuale LSP, pentru a îmbunătăți precizia în cazul cuantizării LSP în general, și pentru a corecta inversarea ordinii LSP consecutive prin cuantizare, am definit următoarea regulă de alocare pentru perechi, care ține seama de valorile celor două polinoame evaluate la mijlocul intervalului de cuantizare:

Dacă  $|P((\hat{x}_{i-1} + \hat{x}_i)/2)| > |Q((\hat{x}_{j-1} + \hat{x}_j)/2)|$ , se alocă  $[x_i, x_j] = [\hat{x}_i, \hat{x}_j]$ ,

în caz contrar se alocă  $[x_i, x_j] = [\hat{x}_{i-1}, \hat{x}_{j-1}]$ .

Pentru a verifica metodele de creștere a preciziei pe care le-am propus, am realizat calculul și cuantizarea în mai multe variante. Rezultatele obținute sunt prezentate în tabelul 5.2, rezultând că metoda de corecție pentru cuantizarea perechilor consecutive pe care am propus-o aduce o îmbunătățire a distorsiunii spectrale cu 0,08 dB față de varianta fără nici o corecție.

Tabelul 5.2. Rezultate comparative pentru metodele de cuantizare

Metoda de corecție	SD [dB]	2÷4 dB [%]	> 4 dB [%]
FS 1016, fără corecție	1,598	14,10	0,322
Corecția FS 1016	1,566	13,78	0,275
Corecția simplă	1,546	12,49	0,196
Perechi, scăderea celei mai mici LSP	1,618	13,29	0,274
Perechi, creșterea celei mai mari LSP	1,634	13,83	0,277
Perechi, cu corecția propusă	1,518	12,21	0,183

În tabel au fost avute în vedere mai multe variante, dintre care primele trei variante constau în calculul și cuantizarea unei singure valori LSP, în mod individual, iar următoarele trei constau în calculul și cuantizarea unor valori pentru LSP consecutive, în perechi:

- fără nici o îmbunătățire, alocând valoarea cea mai mare conform fig. 5.5,
- folosind regula de alocare conform FS 1016,
- aplicând corecția simplă pe care am propus-o, conform fig. 5.5,
- corecție pentru perechi, prin scăderea valorii primei LSP, cea mai mică,
- corecție pentru perechi, prin creșterea valorii celei de-a doua LSP, cea mai mare,
- corecție pentru perechi, prin metoda pe care am propus-o.

### 5.3 Interpolarea LSP

Codorul CELP folosește uzual pentru analiză cadre cu durata de 20 ÷ 30 ms, compuse din 160 ÷ 240 eșantioane la frecvența de eșantionare 8 kHz. Pe durata unui astfel de cadru pot apare frecvent modificări ale înfășurătorii spectrale care conduc la salturi bruște ale parametrilor LPC între două cadre vecine. Aceste salturi se manifestă negativ în calitatea semnalului obținut prin decodare. Soluția acestei probleme ar putea fi reducerea dimensiunii cadrului și astfel diferențele între cadrele vecine se vor reduce.

O soluție care nu necesită reducerea duratei cadrului și deci actualizarea mai frecventă a parametrilor LPC, cu efectul negativ al creșterii debitului binar, o constituie interpolarea parametrilor LPC.

În acest scop, pe durata unui cadru se consideră un număr convenabil de subcadre, iar parametrii LPC pentru aceste subcadre se obțin prin interpolarea parametrilor din cele două cadre vecine.

O variantă de dispunere a ferestrei de analiză față de cadrul folosit la decodare este reprezentată în figura. 5.7. Pe durata corespunzătoare fiecărei ferestre de analiză se calculează un set de parametri LPC, care se folosesc pe o durată echivalentă cu două cadre la decodare.

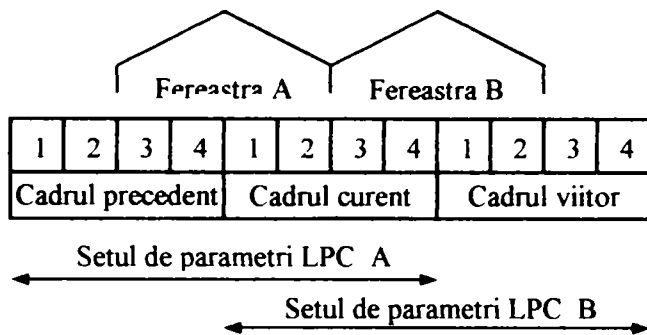


Fig. 5.7. Interpolarea parametrilor LPC

Ponderea celor două seturi de parametri LPC pentru fiecare subcadru, pentru două codoare CELP standardizate, care folosesc patru subcadre pentru fiecare cadru, este prezentată în tabelul 5.3 [CHU03-2].

Tabelul 5.3 Coeficienții de ponderare pe subcadre

	TIA IS 54 VSELP		FS 1016	
	LPC A	LPC B	LPC A	LPC B
Subcadrul 1	0,75	0,25	7/8	1/8
Subcadrul 2	0,5	0,5	5/8	3/8
Subcadrul 3	0,25	0,75	3/8	5/8
Subcadrul 4	0	1	1/8	7/8

Considerând pentru generalitate, un număr de S subcadre pentru un cadru, interpolarea liniară a parametrilor LPC  $a_i$  se poate realiza simplu, conform relației:

$$a_i = \begin{cases} \mu + a_{(i/S)}, & \text{dacă } i \bmod S = 0 \\ \mu + \alpha a_{\lfloor i/S \rfloor} + (1 - \alpha) a_{\lfloor i/S \rfloor + 1}, & \text{dacă } i \bmod S \neq 0 \end{cases} \quad (5.48)$$

unde  $\mu$  are semnificația unui offset.

Coeficientul de interpolare  $\alpha$  variază liniar în intervalul de interpolare și este dat de relația:

$$\alpha = \frac{S - i \bmod S}{S} \quad (5.49)$$

Parametrii  $a_i$  din relația (5.48) pot fi coeficienții de predicție sau oricare dintre reprezentările alternative ale acestora.

O problemă importantă în cazul interpolării o constituie păstrarea stabilității filtrului de sinteză realizat cu valorile interpolate. Din acest punct de vedere nu toate modurile de reprezentare a coeficienților de predicție se comportă la fel.

Analizând interpolarea liniară a paramerilor LSP, am enunțat următoarea proprietate a acestora:

Valorile LSP obținute prin interpolarea liniară a două seturi de parametri LSP conduc la un filtru de sinteză stabil.

Demonstrația este imediată.

Fie două seturi de parametri pentru două cadre consecutive, din care se consideră doar doi parametri consecutivi, notați  $\omega_k, \omega_{k+1}$  și respectiv  $\omega'_k, \omega'_{k+1}$ , care satisfac relațiile de ordine  $\omega_k < \omega_{k+1}$  și  $\omega'_k < \omega'_{k+1}$  și deci corespund unor filtre stabile. Prin interpolare cu un coeficient  $\alpha$  se obțin două noi valori:  $\alpha\omega_k + (1-\alpha)\omega'_k$  și respectiv  $\alpha\omega_{k+1} + (1-\alpha)\omega'_{k+1}$ .

Aceste valori satisfac relația de ordine:

$$\alpha\omega_k + (1-\alpha)\omega'_k < \alpha\omega_{k+1} + (1-\alpha)\omega'_{k+1}. \quad (5.50)$$

Relația (5.50) se poate scrie în forma:

$$\alpha(\omega_k - \omega_{k+1}) < (1-\alpha)(\omega'_{k+1} - \omega'_k) \quad (5.51)$$

care este evidentă, termenul din stânga fiind negativ, iar cel din dreapta pozitiv, având în vedere relația de ordine între valorile neinterpolate ale parametrilor.

Pentru a aprecia performanțele privind interpolarea, am făcut un experiment în care am măsurat distorsiunea spectrală obținută pentru interpolarea a 4 subcadre într-un cadru de 240 eșantioane, pentru trei reprezentări diferite ale parametrilor LPC. Rezultatele, indicate în tabelul 5.4, arată că LSP au cele mai bune proprietăți de interpolare, comparativ cu coeficienții de reflexie și coeficienții LAR.

Tabelul 5.4 Performanțele de interpolare

Parametru	SD [dB]	2 – 4 dB	> 4 dB
LSP	1,32	15,6	3,6
Coeficienții de reflexie	1,57	15,2	5,8
LAR	1,86	14,8	5,5

Este cunoscută interpolarea coeficienților de predicție liniară folosind ca element de interpolare energia cadrului de semnal [EPB98]. Am încercat o nouă variantă de interpolare, în care interpolarea să nu fie făcută folosind direct energia cadrului, ci o valoare ponderată a acestei energii.

Pentru aceasta, am considerat  $R_k^{(1)}$  și  $R_k^{(2)}$  valorile normate ale funcției de autocorelație pentru două cadre consecutive și  $E^i R_k^{(i)}$  funcția de autocorelație pentru subcadrul interpolat, ponderată cu energia subcadrului, astfel:

$$E^i R_k^{(i)} = \alpha E_1^\gamma R_k^{(1)} + (1-\alpha) E_2^\gamma R_k^{(2)}, \quad (5.52)$$

în care exponentul  $i$  pentru energia subcadrului este o valoare necunoscută, iar exponentul  $\gamma$  pentru energia celor două cadre vecine este un parametru subunitar.

Pentru a găsi valoarea  $i$  am rescris rel. (5.52) pentru  $k = 0$ :

$$E^i R_0^{(i)} = \alpha E_1^\gamma R_0^{(1)} + (1-\alpha) E_2^\gamma R_0^{(2)}, \quad (5.53)$$

și deoarece  $R_0 = 1$ , valorile fiind normate:

$$E^i = \alpha E_1^\gamma + (1-\alpha) E_2^\gamma. \quad (5.54)$$

Înlocuind această valoare în (5.52):

$$R_k^{(i)} = \frac{\alpha E_1^\gamma}{\alpha E_1^\gamma + (1-\alpha) E_2^\gamma} R_k^{(1)} + \frac{(1-\alpha) E_2^\gamma}{\alpha E_1^\gamma + (1-\alpha) E_2^\gamma} R_k^{(2)}. \quad (5.55)$$

Am notat

$$\frac{\alpha E_1^\gamma}{\alpha E_1^\gamma + (1-\alpha)E_2^\gamma} = \beta, \quad (5.56)$$

cu semnificația unui coeficient de interpolare generalizat, cu ajutorul căruia relația de interpolare devine:

$$a_m^{(i)} = \beta a_m^{(1)} + (1-\beta)a_m^{(2)} \quad (5.57)$$

Parametrul  $\gamma$  permite stabilirea unei ponderi cu care energia cadrului participă la interpolare. Astfel:

- pentru  $\gamma = 0$ ,  $\beta = \alpha$  și interpolarea nu ține seama de energie
- pentru  $\gamma = 1$ ,  $\beta = \frac{\alpha E_1}{\alpha E_1 + (1-\alpha)E_2}$  și factorul de interpolare este proporțional cu energia neponderată a cadrelor.

Este de așteptat ca ponderarea cu energia cadrelor, prin atribuirea unei importanțe mai mari cadrului cu energie mai mare, să aducă o îmbunătățire față de interpolarea simplă.

Pentru a verifica performanțele interpolării LSP în funcție de energia cadrelor, am comparat câștigul de predicție și distorsiunea spectrală pentru cadrele interpolate, considerând pentru parametrul  $\gamma$  diferite valori în intervalul  $0 \div 1$ . Rezultatul este prezentat în tabelul 5.5.

Pentru valoarea  $\gamma = 0,1$  am obținut un minim al distorsiunii spectrale, cu 0,04 dB (3%) mai mic față de cazul interpolării fără a ține seama de energia cadrelor ( $\gamma = 0$ ).

Câștigul de predicție are o variație similară cu distorsiunea spectrală, obținând un maxim pentru  $\gamma = 0 \div 0,2$ , creșterea procentuală fiind de același ordin de mărime.

O îmbunătățire semnificativă am constatat însă în ceea ce privește procentul cadrelor pentru care distorsiunea spectrală este cuprinsă între  $2 \div 4$  dB, care a scăzut cu 0,8% (ceea ce reprezintă o scădere de 5% față de valoarea obținută fără a ține cont de energie), precum și în ceea ce privește procentul cadrelor cu distorsiunea spectrală peste 4 dB, care a scăzut cu 0,9%, ceea ce reprezintă o scădere cu 25% față de situația în care nu se ține seama de energie la interpolare.

Tabelul 5.5 Distorsiunea spectrală funcție de parametrul  $\gamma$

$\gamma$	SD [dB]	2 ÷ 4 dB [%]	> 4 dB [%]
0,0	1,32	15,6	3,6
0,1	1,28	14,8	2,7
0,2	1,33	15,1	3,1
0,3	1,35	15,2	3,4
0,4	1,38	15,5	3,3
0,5	1,42	15,6	3,5
0,6	1,46	15,9	3,7
0,7	1,51	16,2	3,8
0,8	1,55	16,5	3,8
0,9	1,57	16,6	3,9
1,0	1,59	16,6	4,2

Îmbunătățirea relativă foarte redusă a distorsiunii spectrale se explică prin faptul că această distorsiune este mediată pe toate cadrele.

Scăderea semnificativă a procentului cadrelor cu distorsiunea spectrală peste 4 dB sugerează faptul că interpolarea subcadrelor ținând cont de energia cadrelor vecine se face mai precis pentru acele porțiuni de semnal care prezintă variații mari ale energiei de la un

cadru la altul. Această afirmație este în concordanță și cu relația (5.56), care conduce la  $\beta = \alpha$ , dacă  $E_1 = E_2$ , indiferent de valoarea exponentului  $\gamma$ .

## 5.4 Cuantizarea vectorială a LSP

Pentru a analiza posibilitățile de cuantizare vectorială a LSP și efectele cuantizării acestor parametri asupra semnalului vocal obținut în urma codării CELP, se propune modelul reprezentat în figura 5.8.

Conform acestui model, asupra semnalului de intrare se execută analiza LPC pe cadre ponderate în prealabil cu o funcție fereastră. Se determină coeficienții polinomului  $A(z)$  și apoi se convertesc acești coeficienți în parametri LSP. Se calculează semnalul rezidual, care se folosește direct, necuantizat, ca semnal de intrare în filtrul de sinteză. Parametrii LSP se cuantizează și valorile rezultate se folosesc pentru a construi filtrul de sinteză cu care se obține semnalul de ieșire.

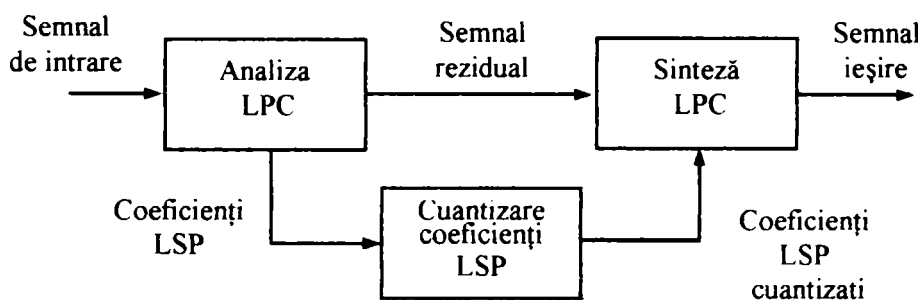


Fig. 5.8. Model pentru studiul cuantizării LSP

Am transpus în Matlab modelul din fig. 5.8, împreună cu modelul care permite studiul cuantizării semnalului rezidual, iar principalele funcții care au fost implementate sunt prezentate în cap. 8 [CBA03-1].

Semnalele de intrare provin de la o sursă de semnal de bună calitate (radio FM) și au fost înregistrate sub formă de fișier .wav, mono, cu frecvența de eșantionare 8 kHz și 16 biți pe eșantion. Nu a fost folosită nici o filtrare sau accentuare a unor porțiuni din spectru. După înregistrare, semnalul care este destinat obținerii vectorilor de antrenare a fost prelucrat cu un program de editare audio, eliminându-se cât mai mult posibil din perioadele de liniște. Au fost obținute astfel un număr de aproximativ 40.000 cadre de semnal, corepunzând cu un total de aproximativ 20 minute de semnal. Semnalele folosite ca secvențe de test nu au fost prelucrate în sensul eliminării perioadelor de liniște.

Ca măsură a performanței cuantizoarelor folosite pentru LSP se folosește distorsiunea spectrală medie, definită de [GDM76], [LRT97]:

$$SD = \sqrt{\frac{1}{f_{\max} - f_{\min}} \int_{f_{\min}}^{f_{\max}} [10 \log_{10}(P_i(f)) - 10 \log_{10}(\hat{P}_i(f))]^2 df} \quad (5.58)$$

în care  $P_i(f)$  și  $\hat{P}_i(f)$  reprezintă spectrele LPC pentru cadrul  $i$ , date de :

$$P_i(f) = 1 / |A_i(e^{j2\pi f / F_s})|^2 \quad (5.59)$$

$$\hat{P}_i(f) = 1 / |\hat{A}_i(e^{j2\pi f / F_s})|^2 \quad (5.60)$$

iar  $F_s$  este frecvența de eșantionare și  $A_i(z)$ , respectiv  $\hat{A}_i(z)$  sunt polinoamele LPC original și cuantizat.



Noțiunea de *cuantizare transparentă* a parametrilor spectrali [PAT93] definește calitatea pentru care cuantizarea parametrilor spectrali nu introduce distorsiuni audibile în semnalul vocal. Dacă cuantizarea parametrilor spectrali este transparentă nu se poate sesiza nici o diferență calitativă, prin teste subiective de audibilitate, între semnalul vocal obținut cu modelul reprezentat în fig. 5.8 și semnalul vocal obținut cu același model, dar la care parametrii spectrali folosiți în blocul de sinteză sunt direct cei obținuți la analiză, adică necuantizați.

Se consideră că se atinge cuantizarea transparentă dacă distorsiunea spectrală medie SD nu depășește 1 dB, și în plus, procentul cadrelor în care SD se încadrează între 2 și 4 dB nu depășește 2% și nu există cadre în care SD să depășească 4 dB.

Tehnicile de cuantizare a parametrilor de predicție prezentate în literatură acoperă o gamă foarte largă de metode, care pot fi metode scalare sau vectoriale.

O primă categorie de metode o reprezintă metodele scalare, care cuantizează fiecare parametru în mod independent [PAK95]. În această categorie se poate menționa metoda folosită în standardul GSM [ETS97-1], care cuantizează scalar, neliniar, cu ajutorul unor tabele, parametrii LAR, sau metoda similară folosită de standardul FS 1016, care codează scalar neliniar parametrii LSF.

Metodele de cuantizare vectorială directă sunt dificil de aplicat, deoarece o cuantizare transparentă presupune un număr de 24 – 30 biți/cadru, adică un dicționar de mărimea  $2^{24} - 2^{30}$ , făcând aproape imposibilă căutarea în timp real.

Majoritatea metodelor de cuantizare vectorială folosesc divizarea vectorului parametri de predicție în subvectori, (uzual 3) și apoi diferite tehnici de cuantizare pentru acești subvectori, care acoperă un spectru larg de metode dintre cele prezentate în cap. 4, începând cu metodele arborescente [KAT94], [KKK90] [BBM93], cuantizarea trellis [LSH03] [MFS93], metode de cuantizare scalar – vectoriale, predicție neliniară, și mergând până la metode care combină codarea sursei cu codarea canalului [GPK91], [SOJ93], [LCH94], [RSS95], [LCK96], [LAK00], [PRP01], [XAD96], [HAG96], [XCP99], [SRO03], [BBM93], [NTL01], metode adaptive [FBF95] sau predictive [LWO01], [CBA04-3], [CBA04-5].

O serie de metode sunt folosite pentru a exploata corelația relativ mare care există între coeficienții de predicție din cadre vecine [LWC95], realizând cuantizarea predictivă, sau predictivă cu medie alunecătoare (MA) [ECT94], [JWG96], [ESJ99], [SAA00], [SHD01].

O altă categorie de metode exploatează diferențele mari existente între cadrele de semnal sonor și cele de semnal nesonor, precum și cerințele mai reduse privind calitatea cuantizării în cadrele de semnal nesonor, realizând o cuantizare cu rată variabilă [CSK99].

De remarcat că în cazul folosirii LSP, ca reprezentare a parametrilor LPC, se pot calcula, cuantiza și transmite fie valorile  $\omega_i$ , ca soluții ale unor ecuații de forma (5.31), (5.32), sau frecvențele asociate (LSF), fie valorile  $x_i$ , ca soluții ale unor ecuații de forma (5.42), (5.43), fără a mai efectua transformarea  $\omega = \arccos x$ , inversă relației de substituție (5.34). Se economisește în al doilea caz calculul unei funcții trigonometrice, iar rezultatele în ceea ce privește performanțele cuantizării și interpolării sunt identice.

#### 5.4.1 Cuantizarea P-SVQ a coeficienților LSP

În figura 5.9 am reprezentat un fragment de semnal vocal (fig. 5.9. a) și evoluția parametrilor LSF corespunzători (fig. 5.9. b), iar în fig. 5.9. c), pentru mai multă claritate, am reprezentat, pentru același semnal, evoluția primilor 4 parametri LSP, pentru un număr de aproximativ 380 cadre de semnal.

Analizând parametrii LSP reprezentați în fig. 5.9, se observă o corelație între valorile LSP din cadre consecutive. Această corelație este mai puternică pentru cadrele cu caracter pronunțat periodic (sonore) și mai redusă în rest (cadrele nesonore și tranziții).



Corelația dintre parametrii LSP din cadre succesive rezultă și din datele din tabelul 5.6, în care am calculat coeficienții de corelație între parametrii LSP din cadrul curent și parametrii LSP din 10 cadre precedente. Prima coloană din tabel conține LSP din cadrul curent, iar pe coloanele următoare sunt parametrii LSP din cadrele precedente.

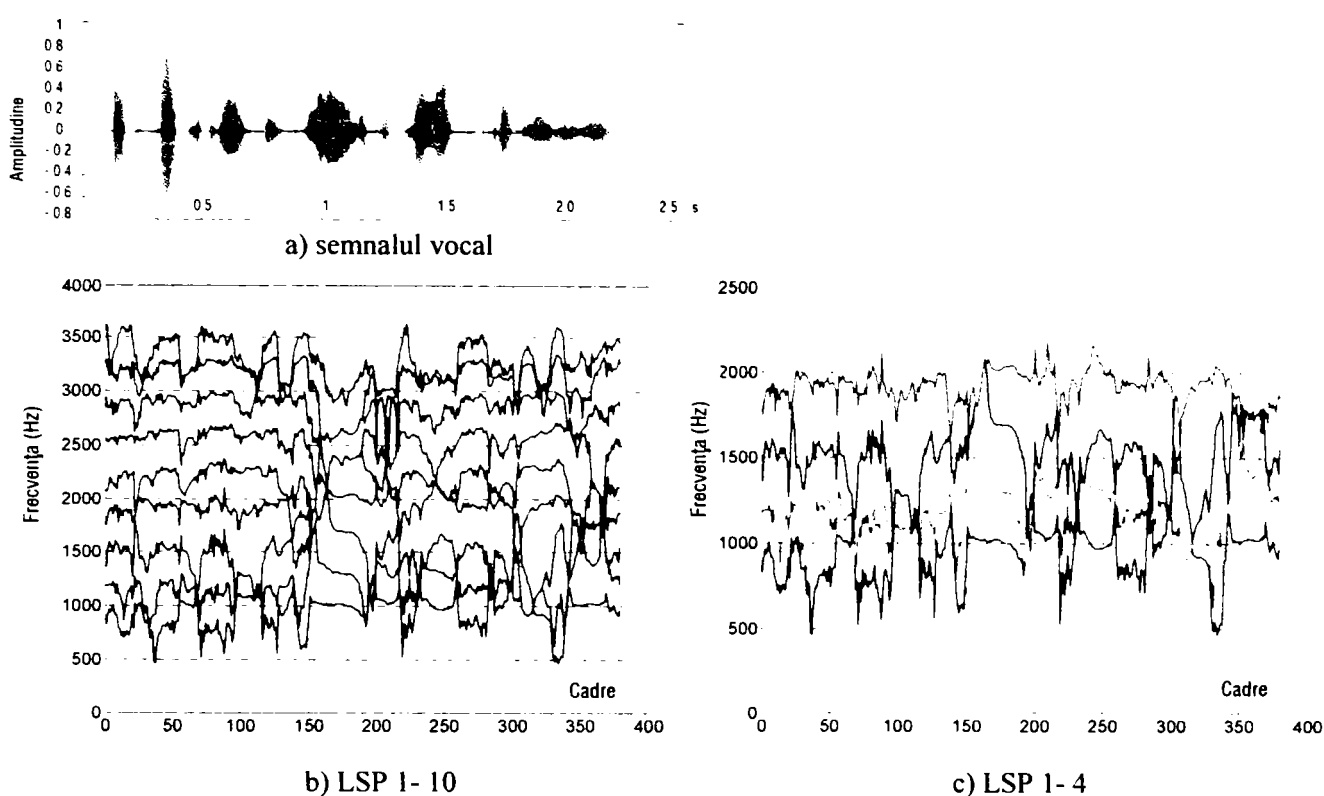


Fig. 5.9. Evoluția parametrilor LSP pentru un segment de semnal

Tabelul 5.6. Coeficienții de corelație între LSP din cadre succesive

LSP nr.	Cadre precedente									
	1	2	3	4	5	6	7	8	9	10
1	0.90	0.85	0.76	0.69	0.61	0.56	0.48	0.37	0.23	0.11
2	0.90	0.77	0.63	0.54	0.47	0.38	0.32	0.30	0.28	0.20
3	0.93	0.79	0.71	0.62	0.50	0.44	0.35	0.23	0.22	0.18
4	0.91	0.83	0.73	0.64	0.55	0.48	0.43	0.21	0.20	0.14
5	0.96	0.87	0.81	0.75	0.66	0.60	0.53	0.21	0.18	0.12
6	0.95	0.85	0.76	0.68	0.61	0.55	0.48	0.20	0.15	0.10
7	0.92	0.82	0.76	0.65	0.59	0.50	0.42	0.18	0.14	0.10
8	0.91	0.82	0.72	0.64	0.56	0.48	0.42	0.14	0.11	-0.08
9	0.89	0.74	0.65	0.56	0.47	0.42	0.36	0.11	0.10	-0.05
10	0.80	0.70	0.59	0.50	0.44	0.38	0.33	0.10	-0.02	-0.03

Această corelație este exploatată de cuantizorul din figura 5.10, care este o combinație între cuantizorul predictiv (P) și cuantizorul partiționat (split VQ – SVQ), prezentate în cap. 4, și pe care l-am denumit **P-SVQ**.

Pentru a reduce fenomenul de propagare a erorilor și pentru a ține seama de faptul că în unele cazuri corelația între parametrii LSP din cadre succesive nu este prea mare, în schema din fig. 5.10 am propus următoarea tehnică de cuantizare [CBA04-3]:

- pentru cadrele pare, parametrii LSP se cuantizează direct folosind cuantizorul SVQ,

- pentru cadrele impare, predictorul din figură estimează o valoare  $\hat{x}$  care se scade din vectorul  $x$  și se cuantizează rezultatul, folosind tot o schemă de tip SVQ.

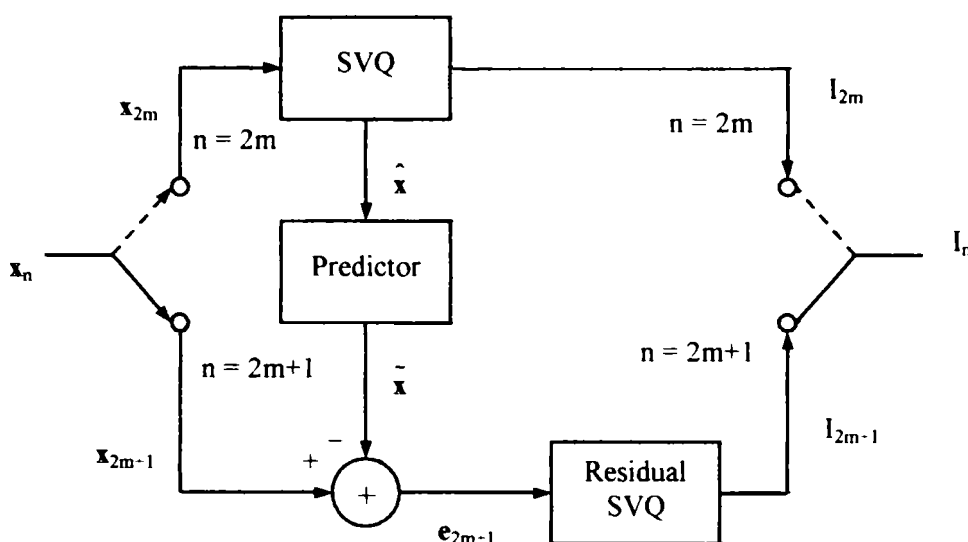


Fig. 5.10 Cuantizarea P-SVQ a LSP

Împărțirea vectorului LSP s-a făcut în două moduri diferite:

- în doi vectori de câte 5 componente fiecare (descompunere 5,5), notată (5,5)P-SVQ,
- în trei vectori de câte 3, 3 și respectiv 4 componente fiecare (descompunere 3,3,4), notată (3,3,4)P-SVQ.

Predictorul folosit este realizat pe baza datelor experimentale, conform celor prezentate în cap. 4.

În tabelul 5.7 am prezentat pentru acest cuantizor distorsiunea spectrală medie SD și numărul de cadre (exprimat în procente din total cadre) care au distorsiunea spectrală cuprinsă între 2 – 4 dB, respectiv mai mare decât 4 dB, pentru trei variante de alocare a biților.

Pentru varianta (5,5) P-SVQ, biții au fost distribuiți egal pentru cele două grupe de vectori, iar pentru varianta (3,3,4) P-SVQ biții au fost repartizați astfel:

- 9, 9, 9 (total 27),
- 8, 8, 8 (total 24),
- 7, 8, 7 (total 22),
- 6, 7, 7 (total 20).

Tabelul 5.7. Distorsiunea spectrală pentru cuantizorul P-SVQ

Biți/cadru	(5,5) P-SVQ			(3,3,4) P-SVQ		
	SD [dB]	2 – 4 dB [%]	> 4 dB [%]	SD [dB]	2 – 4 dB [%]	> 4 dB [%]
27	1,02	2,10	0,01	1,16	2,11	0,02
24	1,17	2,41	0,03	1,22	2,40	0,03
22	1,32	2,66	0,02	1,38	2,97	0,04
20	1,48	6,43	0,04	1,54	6,48	0,04

### 5.4.2 Cuantizarea C-SVQ a coeficienților LSP

O a doua variantă pe care am experimentat-o pentru cuantizare a parametrilor LSP este o combinație între SVQ și cuantizarea cu clasificare (C-VQ – prezentată în cap. 4), pe care am denumit-o în continuare C-SVQ [CBA04-3].

În această variantă vectorul LSP cu 10 componente a fost descompus în 3 subvectori, de câte 3, 3 și respectiv 4 componente fiecare. Pentru fiecare din cei 3 subvectori rezultați s-a folosit o cuantizare cu clasificare, în scopul reducerii complexității operațiilor de căutare.

Am încercat 3 clasificatori diferiți, care vor fi descriși în continuare.

În varianta C-SVQ am folosit pentru fiecare subvector câte un dicționar de 512 cuvinte ( $2^9$ ), rezutând astfel în total un număr de 27 biți pentru cuantizarea LSP pentru un cadru de semnal. Nu au existat în cazul C-SVQ cadre cu distorsiunea spectrală peste 4dB și din acest motiv coloana respectivă nu a mai fost evidențiată în tabelele care prezintă rezultatele experimentale.

- **Clasificatorul vectorial** împarte dicționarul în  $K$  clase, care sunt definite de  $K$  vectori, fiecare dintre acești vectori fiind centroidul clasei respective.

Clasificarea vectorului de intrare în una din cele  $K$  clase se face calculând distanța dintre acest vector și fiecare din cei  $K$  vectori care definesc clasele. Din cele  $K$  clase existente, se rețin  $M$ , alese în ordinea crescătoare a acestor distanțe. În aceste  $M$  clase se face apoi o căutare completă pentru cel mai apropiat vecin.

Rezultatele privind distorsiunea spectrală medie pe care am obținut-o cu clasificatorul vectorial sunt prezentate în tabelul 5.8.

Tabelul 5.8. Distorsiunea spectrală pentru clasificatorul vectorial

Număr clase $K$	Număr clase selectate $M$							
	1		3		5		10	
	SD [dB]	2 – 4 dB [%]	SD [dB]	2 – 4 dB [%]	SD [dB]	2 – 4 dB [%]	SD [dB]	2 – 4 dB [%]
16	0,91	2,02	0,88	1,34	0,82	1,22	-	-
32	0,93	1,92	0,90	1,40	0,85	1,30	0,82	1,02
64	0,94	1,79	0,91	1,78	0,88	1,43	0,80	1,28

Dacă pentru o cuantizare cu un dicționar care conține  $N$  vectori  $k$  – dimensionali sunt necesare pentru căutarea în dicționar un număr de operații proporțional cu  $Nk$ , pentru varianta cu clasificare sunt necesare doar  $Kk$  operații pentru etapa de clasificare și  $(MN K)k$  operații pentru căutarea finală.

Rezultă astfel o reducere a numărului de operații de căutare în raport de  $(KN - MK)$ . Acest raport de reducere a complexității pentru variantele experimentate a fost cuprins între 2,6 și 12 ori.

- **Clasificatorul bazat pe media vectorilor** folosește media celor 3, respectiv 4 subvectori pentru a realiza clasificarea lor, realizând în acest scop o împărțire a dicționarului în regiuni prin hiperplane paralele. O ilustrare intuitivă a claselor este făcută în figura 5.11 pentru cazul bidimensional. După clasificare, se procedează ca în prima variantă, reținând primele  $M$  clase, în ordinea crescătoare distorsiunilor spectrale și apoi cautând doar în interiorul acestor clase.

Pentru clasificare, dicționarul este structurat în ordinea mediilor vectorilor de cod și împărțit în  $K$  clase, iar media vectorului de intrare este un scalar care se cuantizează cu cea mai apropiată valoare a mediei vectorilor de cod, folosind un tabel care memorează aceste medii.

Pentru a aprecia complexitatea acestei scheme, se observă că pentru clasificare sunt necesare  $\log(K)$  operații care se execută pentru fiecare din cei 3 subvectori, iar căutarea ulterioară în cele  $M$  clase selectate necesită același volum de operații ca în cazul anterior.

În tabelul 5.9 sunt prezentate rezultatele experimentale pentru această variantă de clasificator.

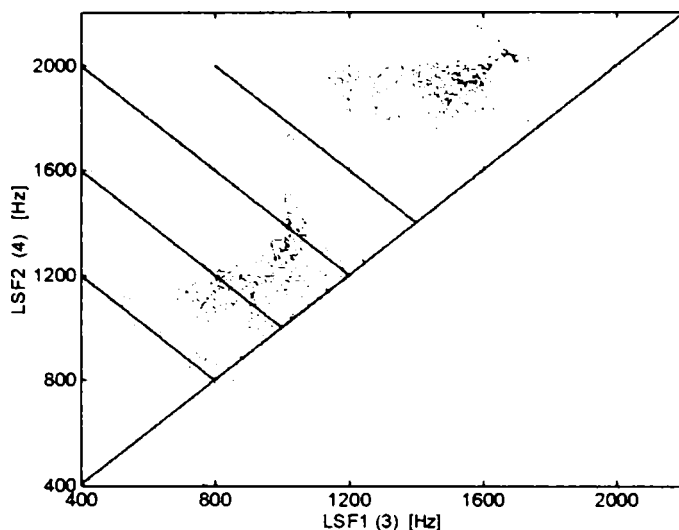


Fig. 5.11. Clasificatorul valoare medie

Tabelul 5.9. Distorsiunea spectrală pentru clasificatorul valoare medie

Număr clase $K$	Număr clase selectate $M$							
	1		3		5		10	
	SD [dB]	2 – 4 dB [%]	SD [dB]	2 – 4 dB [%]	SD [dB]	2 – 4 dB [%]	SD [dB]	2 – 4 dB [%]
16	0,98	2,12	0,89	1,38	0,85	1,26	-	-
32	0,95	1,94	0,93	1,43	0,87	1,34	0,84	1,12
64	0,95	1,93	0,94	1,80	0,91	1,47	0,86	1,34

• **Clasificatorul bazat pe valoarea mediană** folosește pentru clasificarea unui vector  $k$  – dimensional valoarea componentei cu indicele  $k/2$  (rotunjit la întregul imediat următor pentru  $k$  impar). Astfel, pentru subvectorii cu 3, respectiv 4 componente, clasificatorul este valoarea componentei a doua.

În acest caz, spațiul LSP reprezentat în figura 5.11 este împărțit în clase prin hiperplane perpendiculare pe axa componentei cu indicele  $k/2$ .

Complexitatea de calcul este aceeași ca și pentru cazul anterior bazat pe valoarea medie, deoarece clasificarea presupune și în această variantă căutarea unei valori din  $K$  posibilități.

Față de primele două variante de clasificator, la care valoarea folosită pentru clasificare trebuie memorată, crescând necesarul de memorie, clasificatorul bazat pe valoarea mediană nu necesită memorie suplimentară, deoarece valoarea folosită pentru clasificare face parte din dicționar. Distorsiunea spectrală medie obținută experimental pentru acest tip de clasificator este prezentată în tabelul 5.10.

Tabelul 5.10. Distorsiunea spectrală pentru clasificatorul valoare mediană

Număr clase $K$	Număr clase selectate $M$							
	1		3		5		10	
	SD [dB]	2 – 4 dB [%]	SD [dB]	2 – 4 dB [%]	SD [dB]	2 – 4 dB [%]	SD [dB]	2 – 4 dB [%]
16	1,12	2,14	0,99	1,40	0,93	1,29	-	-
32	1,02	1,97	0,96	1,49	0,92	1,38	0,90	1,22
64	0,98	1,98	0,95	1,89	0,91	1,55	0,89	1,39

Analizând datele din aceste tabele, se constată eficiența cuantizorului cu clasificare vectorială, care asigură cuantizarea transparentă la 27 biți/cadru în aproape toate variantele încercate (cu excepția clasificării după valoarea mediană – tabelul 5.10, pentru cazul folosirii unei singure clase) și totodată necesită un volum redus de calcule.

## Capitolul 6

### Modelarea periodicității în codorul CELP

Semnalul eroare de predicție sau semnalul rezidual exprimat prin relația (3.5) (Cap. 3) ar trebui să fie, în cazul ideal, un zgomot alb, deoarece filtrul de predicție constituie un filtru de albire pentru semnalul vocal, ca de altfel pentru orice alt tip de semnal care corespunde modelului AR [STA98]. Acest lucru nu este însă în totalitate exact în practică, din două motive.

Pe de o parte, datorită nestaționarității, semnalul vocal nu corespunde decât aproximativ, pe intervale scurte de timp, modelului AR și astfel semnalul rezidual încă mai conține informații caracteristice semnalului vocal, informații care trebuie cuantizate într-un mod adecvat și transmise la decodor.

Pe de altă parte, predicția pe termen scurt se face luând în calcul un număr relativ redus ( $10 \div 12$ ) de eșantioane ale semnalului vocal, și din acest motiv nu poate extrage din semnalul vocal informația legată de periodicitatea semnalului, corespunzătoare frecvenței fundamentale.

Pentru această periodicitate se folosește în literatura de limbă engleză termenul *pitch*, care indică perioada corespunzătoare frecvenței fundamentale, exprimată în multipli ai perioadei de eșantionare. Având în vedere că această periodicitate va fi modelată în codorul CELP prin predicția pe termen lung LTP, se va folosi în continuare termenul *ordinul LTP* [STA98], justificat de faptul că ordinul filtrului LTP are această valoare.

În figura 6.1 s-au reprezentat două fragmente de semnal vocal, împreună cu semnalul rezidual corespunzător unei predicții de ordinul 10.

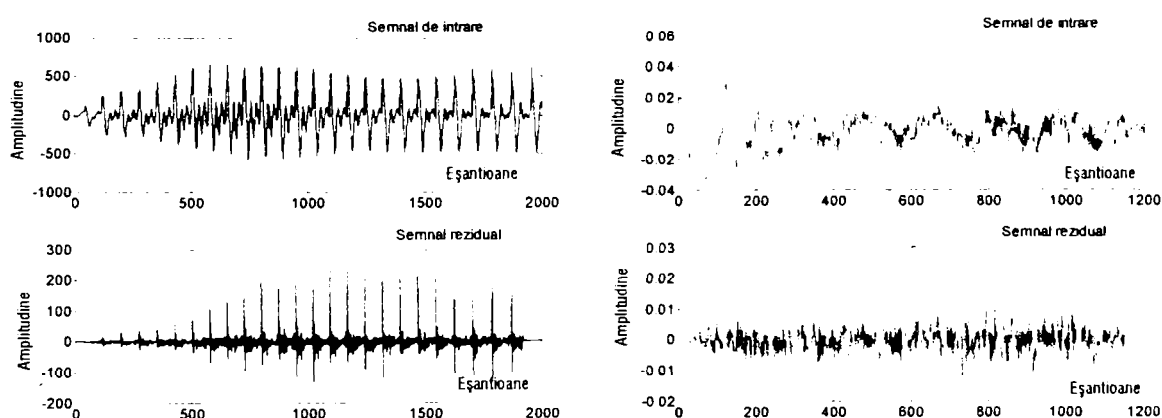


Fig. 6.1. Semnalul vocal și semnalul rezidual

Analizând semnalul rezidual reprezentat în figura 6.1, am constatat că pentru cadrele de semnal care se caracterizează printr-o periodicitate pronunțată, denumite cadre sonore (*voiced*), semnalul rezidual are același pronunțat caracter periodic ca și semnalul de intrare.

Semnalele care nu au acest caracter periodic, denumite semnale nesonore (*unvoiced*), produc un semnal rezidual care este foarte apropiat de zgomotul gaussian.

În figura 6.2 s-a reprezentat evoluția câștigului de predicție pentru un cadru de semnal sonor, respectiv pentru un cadru de semnal nesonor, în funcție de ordinul de predicție. Calculul câștigului a fost făcut pentru fiecare valoare a ordinului de predicție între 2 și 100, iar cadrele au fost extrase din semnalul reprezentat în fig. 6.1.

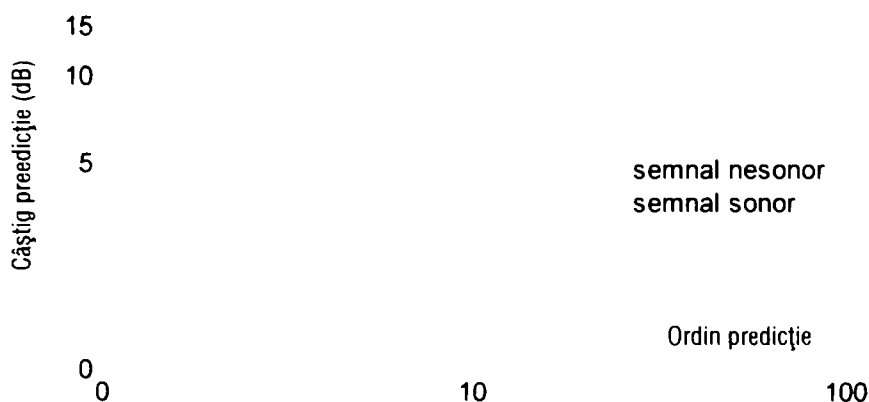


Fig. 6.2. Câștig de predicție în funcție de ordinul de predicție

Analizând fig. 6.2, se constată o creștere semnificativă a câștigului odată cu creșterea ordinului de predicție până la  $10 \div 12$ , atât pentru semnalul sonor cât și pentru cel nesonor. Creșterea ordinului de predicție peste această valoare nu mai aduce modificări semnificative pentru semnalul nesonor și câștigul rămâne aproximativ constant. Pentru cadrul sonor însă, după ce câștigul rămâne aproximativ constant pentru un domeniu larg al ordinului, se mai constată o creștere semnificativă pentru o anumită valoare a ordinului de predicție (62 în fig. 6.1), după care rămâne în continuare constant. Această valoare este egală cu ordinul LTP pentru cadrul respectiv de semnal și în aceste condiții predicția liniară a inclus în model și informațiile referitoare la periodicitatea semnalului.

A alege însă un ordin de predicție suficient de mare pentru a cuprinde și cel mai defavorabil caz pentru ordinul LTP (frecvența fundamentală cea mai mică) nu se justifică datorită volumului foarte mare de calcule și a debitului binar rezultat în urma analizei LPC cu un astfel de ordin. (ordinul LTP 100 corespunde unei voci grave, cu frecvența fundamentală de 80 Hz).

Soluția o constituie utilizarea a două circuite de predicție:

- un predictor pe termen lung (LTP),
- un predictor pe termen scurt (STP).

Analiza STP extrage înfășurătoarea spectrală a semnalului vocal și calculează un prim reziduu de predicție. Analiza LTP, care se face asupra acestui reziduu, va extrage informațiile despre periodicitatea care încă mai este păstrată, și va calcula un al doilea reziduu de predicție. Modelarea periodicității cu ajutorul filtrului LTP este prezentată pe larg în cadrul acestui capitol.

## 6.1 Extragerea periodicității din semnalul rezidual

Caracterul periodic al semnalului rezidual obținut prin prelucrarea pe termen scurt STP se modelează cu ajutorul filtrului de predicție pe termen lung LTP care face parte din codorul reprezentat în fig. 3.10 (Cap. 3).

Denumirea de predicție pe termen lung LTP este justificată de faptul că pentru modelarea caracterului periodic al semnalului rezidual se face o predicție pe un număr mult mai mare de eșantioane, respectiv pe o durată de timp mai mare, comparabilă cu durata unui cadru de semnal, față de predicția pe termen scurt folosită pentru determinarea coeficienților LPC, care se face pe un număr de  $10 \div 12$  eșantioane.



Un model simplu care să extragă periodicitatea semnalului rezidual este reprezentat de un predictor de forma:

$$P'(z) = \beta z^{-L}, \quad (6.1)$$

căruia îi corespunde un filtru de predicție pe termen lung STP de forma:

$$P(z) = 1 - \beta z^{-L} \quad (6.2)$$

și care, în domeniul timp, extrage valoarea eșantionului curent și valoarea eșantionului decalat în timp cu  $L$  poziții și ponderată cu factorul  $\beta$ .

Filtrului de analiză (6.2) îi corespunde, la decodare, filtrul de sinteză:

$$\frac{1}{P(z)} = \frac{1}{1 - \beta z^{-L}}, \quad (6.3)$$

unde  $\beta$  este coeficientul de câștig al LTP și întârzierea  $L$  corespunde periodicității semnalului rezidual.

Coeficientul de câștig  $\beta$  este o măsură a gradului în care se manifestă periodicitatea semnalului, o valoare mai apropiată de unitate indicând un semnal cu un caracter periodic mai pronunțat. Analiza LTP nu are efect pentru semnalele care nu au caracter periodic, coeficientul de câștig fiind pentru aceste semnale apropiat de 0.

Pentru semnalele sonore, în funcție de vorbitor, valoarea întârzierii  $L$  este cuprinsă între  $2 \div 20$  ms. Pentru a permite, spre exemplu, cuantizarea parametrului  $L$  pe 7 biți, așa cum se aplică în cazul codorului CELP FS – 1016, se poate alege pentru  $L$  domeniul de valori  $20 \div 147$ , care cuprinde 128 valori întregi pentru întârziere.

La o frecvență de eșantionare de 8 kHz (perioada  $125\mu\text{s}$ ) se pot cuantiza astfel valori cuprinse între 2,50 ms ( $20 \times 0,125$ ) și 18,575 ms ( $147 \times 0,125$ ), care corespund unei frecvențe fundamentale cuprinse între 53,8 Hz și 400 Hz. Pentru coeficientul de câștig  $\beta$  se pot alocă 3 sau 4 biți.

Actualizarea valorilor întârzierii  $L$  trebuie să se facă mai des decât actualizarea parametrilor LPC, deoarece ordinul LTP, care este o măsură a frecvenței fundamentale, se modifică relativ mult pe durata unui cadru folosit pentru analiza SPT.

Ordinul LTP se actualizează la fiecare subcadru, o valoare tipică fiind actualizarea la fiecare 5 ms. Rezultă astfel o rată de bit necesară pentru codarea periodicității semnalului cuprinsă între 2.000 ÷ 2.200 biți/s.

Se pot folosi și predictorii cu mai mult decât un coeficient, ca de exemplu predictorul cu trei coeficienți, dat de relația [ITU96-1], [ATA82]:

$$H(z) = 1 - \sum_{i=-1}^1 \beta_i z^{-L+i}. \quad (6.4)$$

Parametrii filtrului LTP se pot determina în două moduri diferite și anume:

- în buclă deschisă,
- în buclă închisă.

Cele două metode se pot combina, pentru a reduce volumul de calcul necesar, făcând o primă estimare a acestor parametri în buclă deschisă, după care se caută valoarea finală prin calcul în buclă închisă, în jurul valorilor estimate inițial [CBA03-2].

#### • Determinarea parametrilor LTP în buclă deschisă

Determinarea parametrilor LTP în buclă deschisă folosește modelul din figura 6.3.

Parametrii LTP se determină în sensul minimizării semnalului rezidual LTP pe durata unui cadru compus din  $N$  eșantioane, similar cu determinarea parametrilor LPC.

Semnalul rezidual care rezultă în urma analizei LTP este dat de relația:

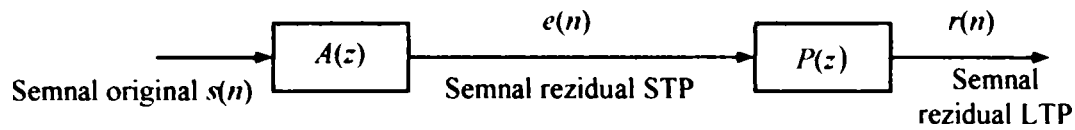


Fig. 6.3. Calculul LTP în buclă deschisă

$$r(n) = e(n) - \beta e(n-L) \quad (6.5)$$

și se minimizează media sa pătratică:

$$\varepsilon = \sum_{n=0}^{N-1} (r(n))^2 = \sum_{n=0}^{N-1} [e(n) - \beta e(n-L)]^2, \quad (6.6)$$

calculând derivata acestei expresii în raport cu coeficientul  $\beta$  și impunând condiția  $\partial E / \partial \beta = 0$ .

Rezultă valoarea optimă pentru câștigul de predicție LTP:

$$\beta = \frac{\sum_{n=0}^{N-1} e(n)e(n-L)}{\sum_{n=0}^{N-1} [e(n-L)]^2} \quad (6.7)$$

și înlocuind această valoare în rel. (6.6) rezultă expresia erorii:

$$\varepsilon = \sum_{n=0}^{N-1} (e(n))^2 - \frac{[\sum_{n=0}^{N-1} e(n)e(n-L)]^2}{\sum_{n=0}^{N-1} [e(n-L)]^2}. \quad (6.8)$$

Minimizarea expresiei (6.8) presupune găsirea valorii  $L$  care maximizează termenul al doilea al expresiei, deoarece primul termen este o constantă.

O variantă de căutare a parametrului  $L$  în buclă deschisă folosește calculul funcției de autocorelație a semnalului de intrare ponderat, care prezintă maxime pentru valoarea  $L$  a întârzierii și pentru multiplii întregi ai acesteia. O altă măsură a ordinului  $L$  o poate constitui funcția diferență de amplitudine medie AMDF [STA98], [CHU03-2] sau se poate folosi un criteriu mai eficient de determinare a ordinului  $L$ , care să țină seama atât de autocorelație cât și de AMDF [SKB01]. Alte metode sunt prezentate în [JSH01], [ASS99], [MGR99].

S-a constatat experimental că determinarea periodicității semnalului rezidual în buclă deschisă, deși se poate face cu bună precizie, conduce la un semnal sintetizat de calitate mai slabă față de determinarea în buclă închisă, care, așa cum s-a precizat, determină acei parametri ai semnalului de excitație și filtrelor de sinteză care asigură semnalul de ieșire cel mai apropiat de semnalul de intrare [CHU03-2].

#### • Calculul parametrilor LTP în buclă închisă

Determinarea parametrilor LTP se poate face și în buclă închisă, conform modelului prezentat în schema din figura 6.4. În acest caz, se caută valorile  $L$  și  $\beta$  care minimizează eroarea între semnalul de excitație provenit din dicționar și filtrat cu filtrul de sinteză LTP cu acești parametri și semnalul de intrare ponderat.

În figura 6.4 s-a reprezentat partea din bucla CELP implicată în determinarea parametrilor LTP, prin minimizarea erorii pătratice ponderate:

$$\varepsilon = \sum_{n=0}^{N-1} [s_w(n) - \hat{s}_w(n)]^2, \tag{6.9}$$

căutând un set de valori optime pentru parametrii de sinteză  $i$ ,  $G$ ,  $\beta$  și  $L$ .

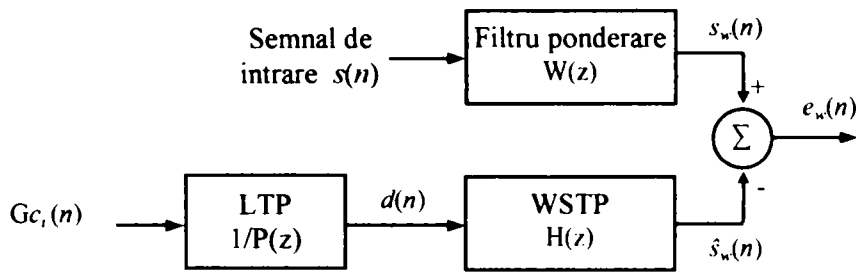


Fig. 6.4. Determinarea parametrilor LTP

În această schemă se face o căutare exhaustivă pentru toate combinațiile posibile între valorile celor doi parametri care reprezintă excitația ( $i$  – indexul din dicționar și  $G$  – câștigul) și valorile parametrilor de predicție pe termen lung  $\beta$  și  $L$ . Volumul de calcul este însă în acest caz relativ mare și se preferă o soluție suboptimală pentru a reduce volumul de calcule.

Soluția constă în a optimiza ordinul  $L$  al filtrului LTP pentru semnal de intrare zero, adică pentru câștig  $G = 0$  și apoi, odată determinat  $L$ , se menține această valoare fixă, iar ceilalți parametri se pot determina la rândul lor în două moduri:

- se calculează valorile optime pentru  $G$  și  $\beta$  pentru fiecare valoare a indexului  $i$ ;
- se determină și coeficientul  $\beta$  pentru excitație zero și apoi, cu parametrii LTP fixați, se determină indexul  $i$ .

Filtrul de sinteză LTP cu evidențierea predictorului LTP este reprezentat în figura 6.5. Așa cum s-a precizat, predictorul realizează o întârziere, pentru a aduna la semnalul de excitație curent o fracțiune din semnalul de excitație din cadrul precedent.

Această operațiune poate fi însă modelată și în felul ilustrat în figura 6.6, în care întârzierea este văzută ca selecția unui anumit vector dintr-un dicționar denumit *dicționar adaptiv*, datorită faptului că acest dicționar se reactualizează după fiecare subcadru, spre deosebire de dicționarul fix, care rămâne nemodificat.

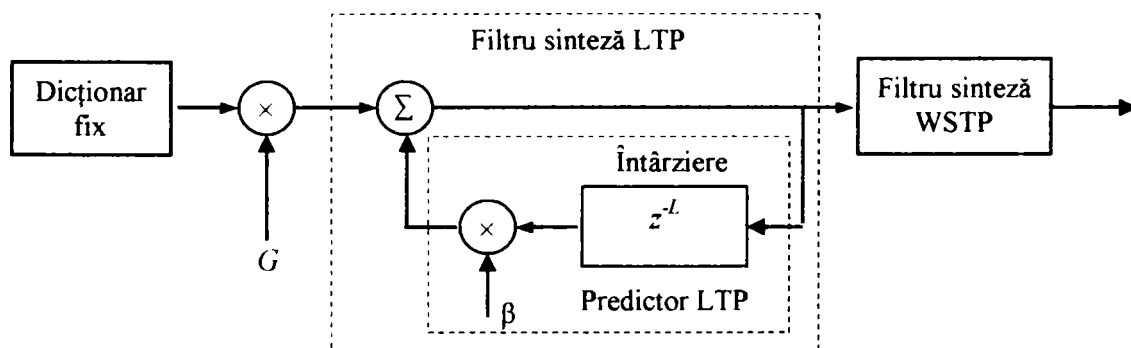


Fig. 6.5 Filtrul LTP

Astfel, în această interpretare, semnalul de excitație pentru filtrul de sinteză WSTP va fi o combinație liniară între un vector de excitație extras din dicționarul fix (stochastic) și un vector extras din dicționarul adaptiv.

În figura 6.6 este reprezentat dicționarul adaptiv sub forma unei memorii cu dimensiunea  $L_{max} + N$ , care păstrează un număr de  $L_{max}$  eşantioane precedente ale semnalului de excitație, din care se obține vectorul de excitație curent de dimensiunea  $N$ .

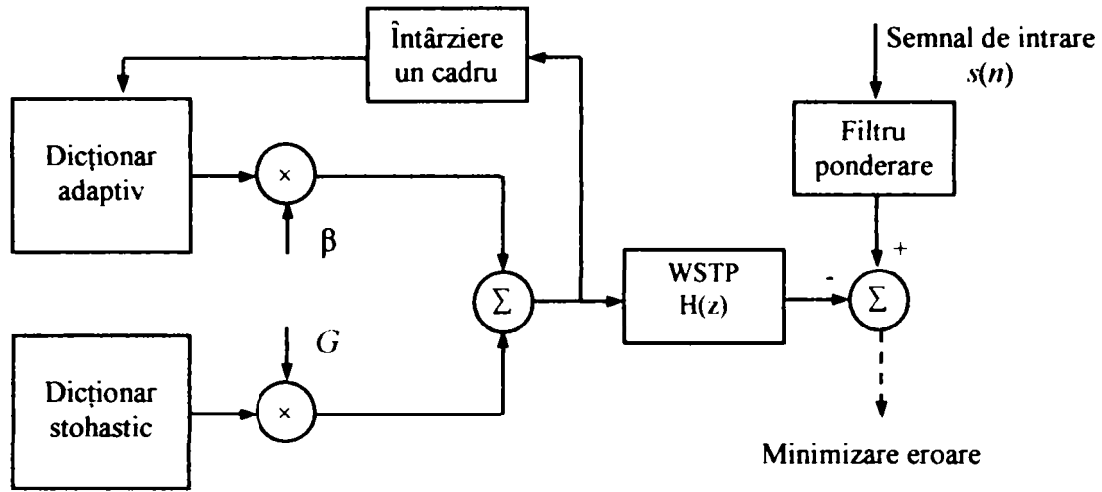


Fig. 6.6. Modelarea LTP cu un dicționar adaptiv

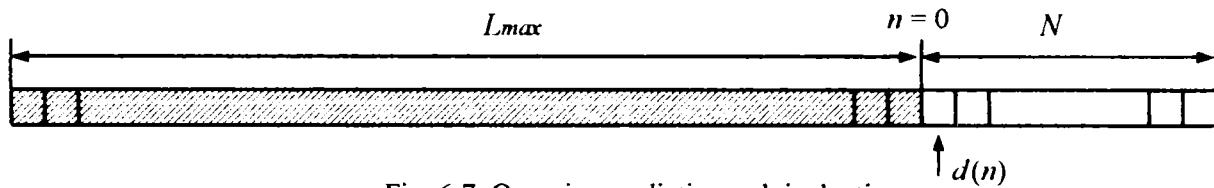


Fig. 6.7. Organizarea dicționarului adaptiv

Presupunând cadrul curent format din  $N$  eşantioane, începând cu  $n = 0$ , vectorul de excitație pentru cadrul curent  $v(n)$ , cu  $n = 0, 1, 2, \dots, N - 1$ , din dicționarul adaptiv, este construit cu ajutorul semnalului de excitație  $d(n)$  conform figurii 6.8, după cum urmează:

Se memorează excitația precedentă:

$$v(n) = d(n) \quad \text{pentru} \quad -L_{max} \leq n < 0; \tag{6.10}$$

unde  $L_{max}$  este ordinul LTP maxim care se poate asigura, apoi se construiește excitația curentă, pentru  $n \geq 0$ :

$$v(n) = v(n - L_p) \quad \text{pentru} \quad n = 0, 1, 2, \dots, N - 1, \tag{6.11}$$

unde  $L_p$  este ordinul LTP pentru intrarea cu indicele  $p$  din dicționarul adaptiv. Secvența  $v(n)$ , cu  $n = 0, 1, \dots, N-1$ , constituie semnalul de excitație pentru cadrul curent.

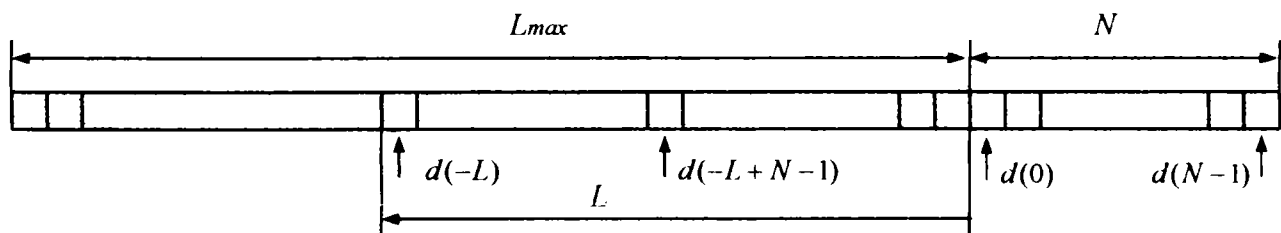


Fig. 6.8 Formarea excitației pentru cadrul curent

Construirea semnalului de excitație pentru cazul în care ordinul LTP  $L$  este mai mic decât durata  $N$  a unui cadru este ilustrată în figura 6.9. Din cele  $N$  eşantioane care formează vectorul de excitație, numai  $L$  sunt cunoscute la începutul cadrului (notate 1, 2, ...  $L$  în figură), deoarece aparțin cadrului precedent și au indicele  $n < 0$ , iar celelalte  $N - L$  (două în figură) nu sunt încă disponibile înainte de începerea cadrului curent, urmând să fie calculate abia pe parcursul acestui cadru.

În acest caz, excitația curentă se formează din cele  $L$  eşantioane disponibile, care se vor repeta periodic până la completarea dimensiunii necesare  $N$ , după cum se observă în partea dreaptă a figurii 6.9.

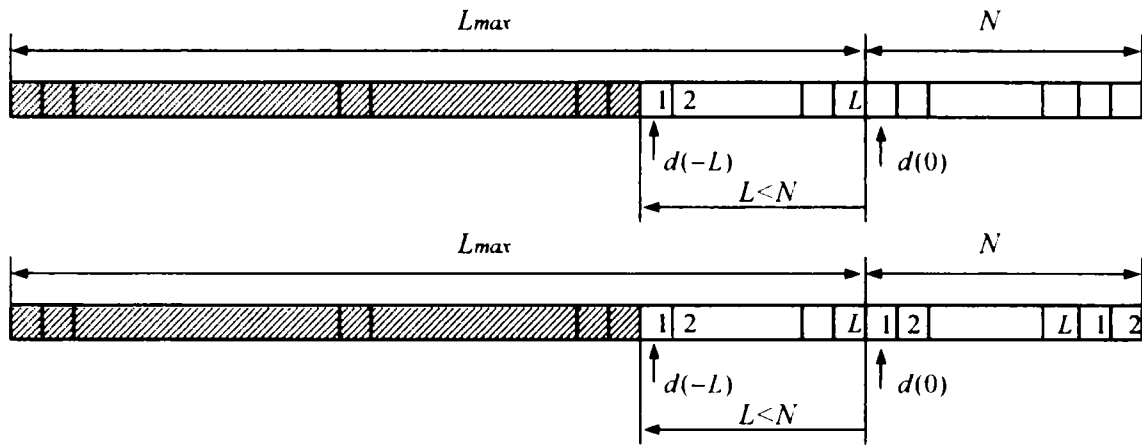


Fig. 6.9. Formarea excitației pentru cazul  $L < N$

Actualizarea dicționarului se face după fiecare cadru, prin deplasarea spre stânga a conținutului dicționarului și completarea adreselor  $(-1, -2, \dots, -N)$  cu valorile din cadrul curent. Ultimele  $N$  eșantioane din dicționar se pierd.

În figura 6.10 este reprezentat conținutul real al dicționarului adaptiv la patru momente de timp diferite, corespunzătoare primelor patru cadre de semnal.



Fig. 6.10. Actualizarea dicționarului adaptiv

Dimensiunea subcadrului este de 60 de eșantioane, iar dimensiunea dicționarului este de 240 de eșantioane și corespunde la 4 subcadre. Se observă deplasarea spre stânga a conținutului dicționarului la fiecare actualizare, astfel că după cele patru subcadre, dicționarul adaptiv este complet populat, iar în continuare, după fiecare subcadru, se face actualizarea în mod similar.

### 6.2 Deducerea relațiilor de calcul pentru parametrii LTP

Pentru cazul general al unui predictor LTP cu  $2q + 1$  coeficienți, de forma:

$$P(z) = \sum_{i=-q}^q \beta_i z^{-L+i}, \tag{6.12}$$

am dedus relațiile de calcul pentru optimizarea ordinului predictorului  $L$  și a celor  $2q + 1$  coeficienți. Am folosit apoi aceste relații pentru a implementa un predictor cu 3 coeficienți.

În acest scop am exprimat eroarea medie pătratică ponderată în domeniul  $0 \leq n \leq N - 1$  și apoi am impus condiția de a minimiza această cantitate:

$$\varepsilon = \sum_{k=0}^n (e_w(n))^2, \quad (6.13)$$

unde eroarea ponderată este:

$$e_w(n) = s_w(n) - \hat{s}_w(n) = s_w(n) - \sum_{k=-\infty}^{\infty} d(k)h(n-k), \quad (6.14)$$

iar  $s_w(n)$  reprezintă semnalul vocal de intrare după filtrarea perceptuală,  $h(n)$  este răspunsul la impuls al filtrului de sinteză WSPT și însumarea s-a făcut, pentru generalizare, între limitele  $-\infty$  și  $+\infty$ .

Ieșirea filtrului STP se poate exprima astfel:

$$d(n) = Gc_i(n) + \sum_{j=-q}^q \beta_j d(n-L+j). \quad (6.15)$$

Conform celor menționate în cap. 3, răspunsul acestui filtru poate fi exprimat ca o sumă între răspunsul unui filtru cu intrarea zero și răspunsul unui filtru cu starea inițială zero. Astfel, descompunând  $h(n)$  și având în vedere indicii de însumare, expresia erorii devine:

$$e_w(n) = s_w(n) - \sum_{k=-\infty}^{-1} d(k)h(n-k) - \sum_{k=0}^{\infty} d(k)h(n-k). \quad (6.16)$$

unde al doilea termen reprezintă răspunsul la intrarea zero, iar al treilea termen reprezintă răspunsul filtrului cu starea zero. În rel. (6.16) primii doi termeni sunt constanți pentru procesul de optimizare ai parametrilor LTP (termenul al doilea, care reprezintă răspunsul la intrarea zero a filtrului, conține coeficienții determinați în cadrul precedent de semnal) și am considerat că se pot grupa într-un singur termen notat  $\tilde{s}_w(n)$ :

$$e_w(n) = \tilde{s}_w(n) - \sum_{k=0}^{\infty} d(k)h(n-k). \quad (6.17)$$

Înlocuind  $d(n)$  exprimat prin rel. (6.15) în relația de mai sus am obținut:

$$e_w(n) = \tilde{s}_w(n) - G\tilde{c}_i(n) - \sum_{j=-q}^q \beta_j \tilde{d}(n, L+j) \quad (6.18)$$

în care am notat cu simbolul  $\sim$  excitația  $c_i(n)$  și semnalul  $d(n)$  filtrate prin filtrul WSTP:

$$\tilde{c}_i(n) = \sum_{k=0}^{N-1} c_i(k)h(n-k) \quad (6.19)$$

$$\tilde{d}(n, m) = \sum_{k=0}^{N-1} d(k-m)h(n-k). \quad (6.20)$$

Eroarea pătratică este în acest caz:

$$\varepsilon = \sum_{k=0}^{N-1} (e_w(k))^2 = \sum_{k=0}^{N-1} (\tilde{s}_w(k) - G\tilde{c}_i(k) - \sum_{j=-q}^q \beta_j \tilde{d}(k, L+j))^2. \quad (6.21)$$

Am calculat valorile câștigului  $G$  și ale coeficienților  $\beta_j$  care minimizează eroarea anulând derivatele erorii în raport cu acești parametri, și am obținut un sistem de  $(q+2)$  ecuații cu tot atâtea necunoscute, pe care l-am scris matricial în forma:

$$\Phi \mathbf{a} = \mathbf{b} \quad (6.22)$$

unde, prin efectuarea unor calcule, am notat cu  $\Phi$  o matrice de forma:



$$\Phi = \sum_{n=0}^{N-1} \mathbf{v}^{(n)} \mathbf{v}^{(n)T} \quad (6.23)$$

iar pentru vectorii care intervin în (6.22) și (6.23) am obținut expresiile:

$$\mathbf{v}^{(n)} = \begin{bmatrix} \tilde{c}_i(n) \\ \tilde{d}(n, L-q) \\ \vdots \\ \tilde{d}(n, L) \\ \vdots \\ \tilde{d}(n, L+q) \end{bmatrix}, \quad (6.24)$$

$$\mathbf{a} = \begin{bmatrix} G \\ \beta_{-q} \\ \vdots \\ \beta_0 \\ \vdots \\ \beta_q \end{bmatrix}, \quad (6.25)$$

$$\mathbf{b} = \begin{bmatrix} \sum_{n=0}^{N-1} \tilde{s}_w(n) \tilde{c}_i(n) \\ \sum_{n=0}^{N-1} \tilde{s}_w(n) \tilde{d}(n, L-q) \\ \vdots \\ \sum_{n=0}^{N-1} \tilde{s}_w(n) \tilde{d}(n, L) \\ \vdots \\ \sum_{n=0}^{N-1} \tilde{s}_w(n) \tilde{d}(n, L+q) \end{bmatrix}. \quad (6.26)$$

Cu aceasta, sistemul (6.22) este complet determinat și se poate rezolva cu una din metodele cunoscute din algebră. Cazul unui predictor simplu, cu un singur coeficient, este un caz particular în care  $q = 0$  și în acest caz sistemul va avea două ecuații cu două necunoscute.

După determinarea valorilor optime pentru parametrii de câștig, am înlocuit aceste valori în rel. (6.21) și am obținut ordinul  $L$  care minimizează expresia erorii astfel obținută.

Sistemul (6.22) se poate rezolva numai dacă se presupune că valoarea maximă admisă pentru ordinul  $L$  și durata  $N$  a cadrului, avută în vedere pentru optimizare, îndeplinesc inegalitatea  $L+q < N$ .

În acest caz, toate eșantioanele de forma  $\tilde{d}(n, M)$  care apar în sistemul (6.22) depind numai de eșantioane trecute ale semnalului de excitație, adică de  $d(n)$  cu indicii  $n < 0$ , care sunt deja cunoscute și astfel, la începutul cadrului curent, matricea  $\Phi$  este disponibilă.

În practică, această condiție nu este întotdeauna îndeplinită, mai ales pentru semnalele vocale cu frecvența fundamentală mai ridicată (voci feminine sau de copii), când ordinul LTP poate fi de până la trei ori mai mic decât durata subcadrului.

Astfel, dacă se consideră că ordinul LTP este egal cu un subcadru de analiză tipic, de 60 eșantioane, la frecvența de eșantionare 8.000 Hz, acestui ordin îi corespunde o frecvență

fundamentală de 133 Hz (200 Hz pentru subcadru de 40 eșantioane), care poate fi depășită frecvent.

Pentru a rezolva cazul în care ordinul  $L$  este mai mic decât durata cadrului există două abordări:

- recircularea excitației în filtrul LTP,
- completarea excitației cu o extensie periodică.

Recircularea excitației are dezavantajul că va conduce la un volum mai mare de calcule, pe când repetarea periodică a excitației este o soluție suboptimală.

### 6.3 Optimizarea filtrului LTP cu un coeficient

Particularizând ecuația (6.22) pentru cazul filtrului LTP cu un singur coeficient, adică pentru  $q = 0$ , se obține câștigul optim pentru filtrul LTP, care reprezintă totodată și singurul coeficient al filtrului:

$$\beta_0 = \frac{\sum_{n=0}^{N-1} \tilde{s}_w(n) \tilde{d}(n, L)}{\sum_{n=0}^{N-1} (\tilde{d}(n, L))^2}. \quad (6.27)$$

Înlocuind valoarea astfel determinată în expresia erorii, se obține:

$$\varepsilon = \sum_{n=0}^{N-1} (\tilde{s}_w(n))^2 - \frac{[\sum_{n=0}^{N-1} \tilde{s}_w(n) \tilde{d}(n, L)]^2}{\sum_{n=0}^{N-1} [\tilde{d}(n, L)]^2} \quad (6.28)$$

care se minimizează în raport cu ordinul LTP. Primul termen fiind constant, minimizarea erorii (6.28) este echivalentă cu maximizarea termenului al doilea, rezultând astfel ordinul de predicție  $L$  corespunzător.

#### 6.3.1 Recircularea excitației prin filtrul LTP

Problema care apare în cazul în care ordinul de predicție LTP este mai mic decât durata cadrului se poate rezolva, după cum s-a precizat, în două moduri. Modul cu completarea prin extensie periodică a excitației nu influențează calculul parametrilor de predicție LTP și relația (6.15) este valabilă pentru toate eșantioanele din semnalul de excitație pentru cadrul curent. Varianta de repetare prin extensie periodică se înscrie, din punct de vedere al calculelor, în cazul  $L > N$ .

În cazul în care se folosește varianta de recirculare a excitației, calculul se face în mod diferit pe diferite subintervale ale cadrului. Deoarece nu am întâlnit în literatură o analiză completă a acestei probleme, am analizat în detaliu varianta cu recircularea excitației și am calculat expresiile care conduc la obținerea câștigului de predicție. Rezultatele pe care le-am obținut sunt prezentate în continuare, pentru trei cazuri posibile, determinate de relația existentă între ordinul  $L$  și dimensiunea cadrului  $N$ . Am efectuat această analiză numai pentru cazul predictorului cu un singur coeficient.

- ordinul  $L$  este mai mare decât durata cadrului  $N$

Semnalul de excitație va fi în acest caz:

$$d(n) = \beta_0 d(n - L), \quad 0 \leq n \leq N - 1 \quad (6.29)$$

iar eroarea pătratică:

$$\varepsilon = \sum_{n=0}^{N-1} (e_w(n))^2 \quad (6.30)$$

unde

$$e_w(n) = \tilde{s}_w(n) - \beta_0 \tilde{d}(n, L). \quad (6.31)$$

Înlocuind în (6.30), am obținut:

$$\varepsilon = \sum_{n=0}^{N-1} (\tilde{s}_w(n))^2 - 2\beta_0 \sum_{n=0}^{N-1} \tilde{s}_w(n) \tilde{d}(n, L) + \beta_0^2 \sum_{n=0}^{N-1} (\tilde{d}(n, L))^2 \quad (6.32)$$

și egalând cu zero prima derivată a erorii, am obținut o ecuație liniară de gradul întâi cu soluția:

$$\beta_0 = \frac{\sum_{n=0}^{N-1} \tilde{s}_w(n) \tilde{d}(n, L)}{\sum_{n=0}^{N-1} (\tilde{d}(n, L))^2}, \quad (6.33)$$

expresie identică cu (6.27)

- ordinul LTP cuprins între  $N/2$  și  $N$

Semnalul de excitație este, în acest caz, dependent de valoarea  $n$  astfel:

$$d(n) = \begin{cases} d_1(n) = \beta_0 d(n-L), & 0 \leq n < L-1 \\ d_2(n) = \beta_0^2 d(n-2L), & L \leq n \leq N-1 \end{cases} \quad (6.34)$$

adică a doua jumătate a vectorului de excitație se obține trecând încă o dată prin filtrul LTP prima jumătate și, din această cauză, eroarea se poate descompune în doi termeni:

pentru  $0 \leq n < L-1$ :

$$e_{w1}(n) = \tilde{s}_w(n) - \beta_0 \tilde{d}_1(n, L) \quad (6.35)$$

pentru  $L \leq n \leq N-1$ :

$$e_{w2}(n) = \tilde{s}_w(n) - \beta_0 \tilde{d}_1(n, L) - \beta_0^2 \tilde{d}_2(n, 2L)$$

iar eroarea pătratică totală, obținută prin însumarea erorilor pe cele două porțiuni, devine:

$$\begin{aligned} \varepsilon = & \sum_{n=0}^{N-1} (\tilde{s}_w(n))^2 - 2\beta_0 \sum_{n=0}^{N-1} \tilde{s}_w(n) \tilde{d}_1(n, L) + \beta_0^2 \sum_{n=0}^{N-1} (\tilde{d}_1(n, L))^2 - \\ & - 2\beta_0^2 \sum_{n=L}^{N-1} \tilde{s}_w(n) \tilde{d}_2(n, 2L) + 2\beta_0^3 \sum_{n=L}^{N-1} \tilde{d}_1(n, L) \tilde{d}_2(n, 2L) + \beta_0^4 \sum_{n=L}^{N-1} (\tilde{d}_2(n, 2L))^2 \end{aligned} \quad (6.36)$$

- ordinul LTP cuprins între  $N/2$  și  $N/3$

În acest caz, semnalul de excitație se exprimă diferit în funcție de cele trei cazuri posibile pentru  $n$ , observându-se că pentru  $L \leq n < 2L-1$  situația este similară cu cea precedentă, semnalul fiind trecut tot de două ori prin filtrul LTP, iar pentru  $2L \leq n < N-1$  semnalul se trece de trei ori consecutiv prin filtru:

$$d(n) = \begin{cases} d_1(n) = \beta_0 d(n-L), & 0 \leq n < L-1 \\ d_2(n) = \beta_0^2 d(n-2L), & L \leq n < 2L-1 \\ d_3(n) = \beta_0^3 d(n-3L), & 2L \leq n \leq N-1 \end{cases} \quad (6.37)$$

Exprimând eroarea pentru cele trei cazuri:

pentru  $0 \leq n < L-1$ :

$$e_{w1}(n) = \tilde{s}_w(n) - \beta_0 \tilde{d}_1(n, L) \quad (6.38)$$

pentru  $L \leq n \leq 2L-1$ :

$$e_{w2}(n) = \tilde{s}_w(n) - \beta_0 \tilde{d}_1(n, L) - \beta_0^2 \tilde{d}_2(n, 2L) \quad (6.39)$$

pentru  $2L \leq n \leq N-1$ :

$$e_{w3}(n) = \tilde{s}_w(n) - \beta_0 \tilde{d}_1(n, L) - \beta_0^2 \tilde{d}_2(n, 2L) - \beta_0^3 \tilde{d}_3(n, 3L) \quad (6.40)$$

eroarea totală devine:

$$\varepsilon = \sum_{n=0}^{L-1} (e_{w1}(n))^2 + \sum_{n=L}^{2L-1} (e_{w2}(n))^2 + \sum_{n=2L}^{N-1} (e_{w3}(n))^2 \quad (6.41)$$

și având în vedere expresiile erorilor pe cele trei domenii, am obținut pentru eroarea totală forma:

$$\begin{aligned} \varepsilon = & \sum_{n=0}^{N-1} (\tilde{s}_w(n))^2 - 2\beta_0 \sum_{n=0}^{N-1} \tilde{s}_w(n) \tilde{d}_1(n, L) + \beta_0^2 \sum_{n=0}^{N-1} (\tilde{d}_1(n, L))^2 - \\ & - 2\beta_0^2 \sum_{n=L}^{N-1} \tilde{s}_w(n) \tilde{d}_2(n, 2L) + 4\beta_0^3 \sum_{n=L}^{N-1} \tilde{d}_1(n, L) \tilde{d}_2(n, 2L) + \\ & - 2\beta_0^2 \sum_{n=2L}^{N-1} \tilde{s}_w(n) \tilde{d}_3(n, 3L) + \beta_0^4 \sum_{n=L}^{N-1} (\tilde{d}_2(n, 2L))^2 + \\ & + 2\beta_0^4 \sum_{n=L}^{N-1} \tilde{d}_1(n, L) \tilde{d}_3(n, 3L) + 2\beta_0^5 \sum_{n=2L}^{N-1} \tilde{d}_2(n, 2L) \tilde{d}_3(n, 3L) + \\ & + \beta_0^6 \sum_{n=2L}^{N-1} (\tilde{d}_3(n, 3L))^3 \end{aligned} \quad (6.42)$$

Expresia pe care am obținut-o pentru eroarea (6.42) este un polinom de gradul 6 în raport cu  $\beta$ , deci prin calculul derivatei de ordinul I și egalarea cu zero a rezultatului, rezultă o ecuație de gradul 5, a cărei rezolvare ridică probleme într-o implementare în timp real.

Pentru a evita rezolvarea unor ecuații de grad mai mare ca unu, se poate folosi o aproximare, introdusă deja la implementarea LTP prin dicționar adaptiv, prin care semnalul de excitație pentru cadrul curent se obține prin repetarea periodică a celor  $L$  valori disponibile din trecut. Astfel, semnalul de excitație se poate scrie, pentru cazul general:

$$d(n) = \beta_0 d((n \bmod L) - L), \text{ pentru } 0 \leq n \leq N-1. \quad (6.43)$$

Introducând notația:

$$\bar{d}(n, m) = \sum_{k=0}^{N-1} d((k \bmod L) - m) h(n-k) \quad (6.44)$$

eroarea se poate exprima prin:

$$e_w(n) = \tilde{s}_w(n) - \beta_0 \bar{d}(n, L). \quad (6.45)$$

## 6.4 Creșterea rezoluției filtrului LTP

Filtrul LTP cu un singur coeficient are un dezavantaj major în ceea ce privește rezoluția în timp. Acest dezavantaj este datorat faptului că ordinul filtrului, fiind un număr natural, determină ca rezoluția sa în timp, în determinarea exactă a perioadei corespunzătoare frecvenței fundamentale a semnalului vocal, să fie limitată la multiplii întregi ai perioadei de eșantionare. Acest dezavantaj se manifestă printr-o degradare a calității semnalului, deoarece auzul uman este sensibil la variații ale frecvenței fundamentale.

Ordinul LTP în semnalul vocal original este în fapt o mărime cu variație continuă în timp, iar valorile întregi sub care se reprezintă sunt o consecință a operației de eșantionare. Pe lângă alte metode de a rezolva această problemă [KRK94], [HRL01], [YCK02], [RAL03], prelucrarea multirată a semnalelor [MAT98] poate fi folosită pentru a se depăși limitele impuse de eșantionarea cu frecvență fixă a semnalului vocal înainte de prelucrare. În acest mod se poate obține pentru ordinul LTP o valoare care este echivalentă cu un ordin fracționar și pentru care se va folosi denumirea de ordin LTP fracționar, întâlnită în literatură.

### 6.4.1 Filtrul LTP cu ordin fracționar

Un ordin fracționar  $T$  poate fi exprimat ca un ordin întreg  $L$ , la frecvența de eșantionare  $F_e$ , căreia îi corespunde perioada  $T_s$ , la care se adaugă o componentă fracționară  $\eta$  [KRK94]:

$$T = (L + \eta)T_s \quad (6.46)$$

unde  $0 \leq \eta < 1$ , astfel încât:

$$\frac{T}{T_s} - 1 < L \leq \frac{T}{T_s} \quad (6.47)$$

Dacă se presupune că  $\eta$  este un exprimat printr-o fracție subunitară, ca raport de două numere întregi de forma  $t/I$ ,  $t = 0, 1, \dots, I - 1$ , se poate considera că  $L$  este rezoluția fracționară corespunzătoare unei frecvențe de eșantionare multiplu al frecvenței de eșantionare originale. Ordinul fracționar  $t/I$  la frecvența de eșantionare  $F_e$ , corespunde ordinului întreg  $t$  la frecvența de eșantionare  $IF_e$ . Astfel, un ordin fracționar poate fi realizat prin creșterea frecvenței de eșantionare a semnalului de excitație extras din dicționarul adaptiv

Creșterea frecvenței de eșantionare se poate face prin interpolare, folosind o funcție de interpolare adecvată, ca de exemplu o funcție de forma [CHU03-2]:

$$w(n, t) = w_0(n + t) \frac{\sin((n + t)\pi)}{(n + t)\pi} \quad (6.48)$$

unde  $w_0$  este fereastra Hamming centrată în origine:

$$w_0(n) = 0,54 + 0,46 \cos\left(\frac{2\pi n}{N}\right). \quad (6.49)$$

Interpolarea cu factorul întreg  $I$  este echivalentă cu inserarea a  $I - 1$  zerouri între fiecare două eșantioane ale semnalului original și filtrarea trece jos pentru a obține o variantă interpolată. Acest semnal este întârziat cu  $t$  eșantioane, după care este subeșantionat, prin selectarea eșantioanelor cu pasul  $I$ , revenind astfel la frecvența de eșantionare  $F_e$ . Semnalul astfel obținut reprezintă o variantă a semnalului original, întârziată cu valoarea fracționară  $t/I$  [MAT98].

O implementare a unui filtru LTP cu ordin fracționar este realizată în codorul CELP G.729 [ITU96-1], conform tabelului 6.1.

Tabelul 6.1 Domeniul și rezoluția ordinului LTP în codorul CELP G.729

Subcadrul	Ordin	Domeniu	Rezoluție
1	$T_1$	$19\frac{1}{3} - 84\frac{2}{3}$ $85 - 143$	$1/3$ 1
2	$T_2$	$\lfloor T_1 \rfloor - 5\frac{2}{3} \div \lfloor T_1 \rfloor + 4\frac{2}{3}$	$1/3$

O variantă simplă de calcul prin interpolare a ordinului de predicție LTP fracționară presupune cunoscut ordinul întreg  $L$  și folosind o interpolare liniară pentru un eșantion intermediar, conform relației:

$$s(n + \eta) = (1 - \eta)s(n) + \eta s(n + 1) \quad (6.50)$$

utilizează această valoare pentru a calcula componenta fracționară  $\eta$ . Dacă valoarea rezultată nu se încadrează în intervalul  $0 \div 1$ , înseamnă că partea întreagă  $L$  nu este bine aleasă, și se reia calculul folosind pentru  $L$  o valoare mai mare sau mai mică cu o unitate, după caz [MYC91].

### 6.5 Filtrul LTP cu trei coeficienți modificat

Un filtru cu trei coeficienți poate să modeleze periodicitatea semnalului de excitație mai bine decât un filtru cu un coeficient, dar ca efect secundar crește ușor debitul binar, deoarece pentru fiecare subcadru este necesar să se codeze și transmită trei coeficienți în loc de unul singur și, în plus, crește volumul de calcul necesar pentru determinarea celor trei coeficienți.

Analizând răspunsul în frecvență al unui filtru de tipul (6.3), reprezentat în figura 6.10 pentru valorile  $L = 80$  și  $\beta = 0,5$ , se observă caracterul periodic, care se repetă uniform până la frecvența maximă considerată, în timp ce în spectrul semnalului vocal real, caracterul periodic datorat frecvenței fundamentale este atenuat tot mai mult spre frecvențele înalte.

Această situație este modelată mai exact de răspunsul unui filtru cu trei coeficienți:

$$\frac{1}{P(z)} = 1 - \sum_{i=1}^L \beta_i z^{-L+i} \quad (6.51)$$

al cărui răspuns este reprezentat, pentru aceeași valoare a ordinului LTP ( $L = 80$ ) și valorile coeficienților  $\beta_{-1} = 0,25$ ,  $\beta_0 = 0,5$  și respectiv  $\beta_1 = -0,05$  în figura 6.11, observându-se o atenuare a structurii periodice spre frecvențele înalte.

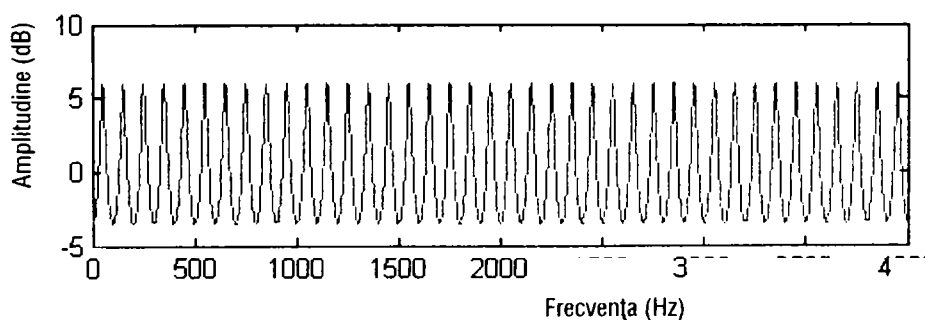


Fig. 6.10. Caracteristica filtrului LTP cu un coeficient

Această caracteristică de atenuare este favorabilă și pentru cazul unor valori fracționare ale perioadei fundamentale. În acest caz, la frecvențe joase, eroarea datorată reprezentării aproximative a unui ordin fracționară cu un ordin întreg este mai mică, iar efectul cumulativ spre frecvențe înalte este atenuat, reducând efectul subiectiv asupra semnalului vocal.



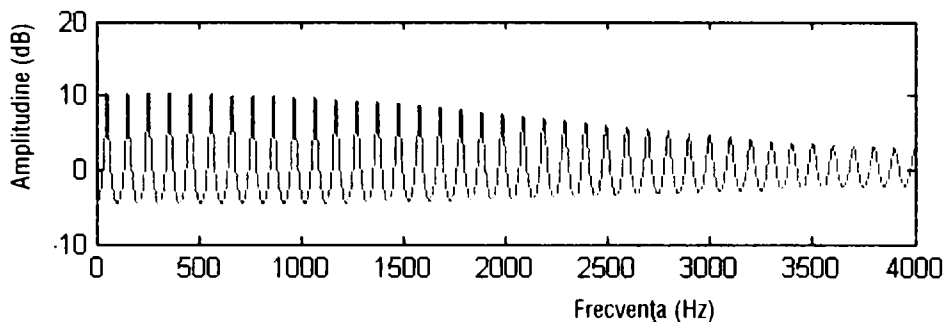


Fig. 6.11. Caracteristica filtrului LTP cu trei coeficienți

### 6.5.1 Calculul parametrilor LTP pentru un filtru cu trei coeficienți

Pentru un filtru LTP cu trei coeficienți am reluat calculele făcute în §6.2 pentru a obține o formă adecvată pentru relațiile de calcul ale parametrilor acestui filtru, în vederea unei implementări în Matlab și verificării comparative a performanțelor.

Pentru un filtru LTP cu trei coeficienți, la care semnalul de excitație se obține prin completare cu o extensie periodică, semnalul de ieșire este:

$$d(n) = \sum_{j=-1}^1 \beta_j d((n \bmod L) - L + j) \quad (6.52)$$

iar eroarea ponderată devine:

$$e_w(n) = \tilde{s}_w(n) - \sum_{k=0}^{\infty} d(k)h(n-k), \quad (6.53)$$

unde  $\tilde{s}_w$  are aceeași semnificație ca și în rel. (6.17).

Folosind notația introdusă în (6.44), am exprimat eroarea în forma:

$$e_w(n) = \tilde{s}_w(n) - \sum_{j=-1}^1 \beta_j \bar{d}(n, L + j). \quad (6.54)$$

Calculând eroarea pătratică pe baza expresiei (6.54), am obținut:

$$\varepsilon = \sum_{n=0}^{N-1} (e_w(n))^2 = \sum_{n=0}^{N-1} (\tilde{s}_w(n) - \sum_{j=-1}^1 \beta_j \bar{d}(n, L + j))^2 \quad (6.55)$$

și apoi, derivând în raport cu  $\beta_j$  și anulând derivatele, rezultă un sistem de 3 ecuații cu tot atâtea necunoscute, pe care l-am scris matricial în forma:

$$\Phi \beta = \delta \quad (6.56)$$

unde

$$\delta = \begin{bmatrix} \sum_{n=0}^{N-1} \tilde{s}_w(n) \bar{d}(n, L-1) \\ \sum_{n=0}^{N-1} \tilde{s}_w(n) \bar{d}(n, L) \\ \sum_{n=0}^{N-1} \tilde{s}_w(n) \bar{d}(n, L+1) \end{bmatrix} \quad (6.57)$$

și  $\Phi$  are forma unei matrici de autocorelație:

$$\Phi = \begin{bmatrix} \sum_{n=0}^{N-1} (\bar{d}(n, L-1))^2 & \sum_{n=0}^{N-1} \bar{d}(n, L-1)\bar{d}(n, L) & \sum_{n=0}^{N-1} \bar{d}(n, L-1)\bar{d}(n, L+1) \\ \sum_{n=0}^{N-1} \bar{d}(n, L)\bar{d}(n, L-1) & \sum_{n=0}^{N-1} (\bar{d}(n, L))^2 & \sum_{n=0}^{N-1} \bar{d}(n, L)\bar{d}(n, L+1) \\ \sum_{n=0}^{N-1} \bar{d}(n, L+1)\bar{d}(n, L-1) & \sum_{n=0}^{N-1} \bar{d}(n, L+1)\bar{d}(n, L) & \sum_{n=0}^{N-1} (\bar{d}(n, L+1))^2 \end{bmatrix}. \quad (6.58)$$

iar  $\beta$  este matricea necunoscutelor:

$$\beta = \begin{bmatrix} \beta_{-1} \\ \beta_0 \\ \beta_1 \end{bmatrix} \quad (6.59)$$

Cu aceasta sistemul se poate rezolva folosind metode algebrice iar în final, eroarea minimă care rezultă pentru valoarea optimă a coeficienților LTP este:

$$\varepsilon_{\min} = \sum_{n=0}^{N-1} (\tilde{s}_w(n))^2 - \beta_{opt}^T \delta \quad (6.60)$$

Dezavantajul filtrului LTP cu trei coeficienți este creșterea debitului binar, fiind necesar să se cuantizeze o valoare pentru ordinul LTP și trei valori pentru câștigul LTP.

Pentru a reduce numărul de elemente care se cuantizează și se transmit, am propus un filtru LTP cu trei coeficienți cu constrângeri, în două variante, care sunt prezentate în continuare.

### 6.5.2 Filtru LTP cu constrângeri cu trei parametri – LTP3

Asupra unui filtru cu trei coeficienți, cu o structură descrisă de relația:

$$P(z) = 1 - \sum_{i=-1}^1 \beta_i z^{-L+i} \quad (6.61)$$

am impus o constrângere prin care numărul coeficienților independenți este redus la doi în loc de trei:

$$\beta_{-1} = \beta_1 = \alpha\beta_0 \quad (6.62)$$

$$\beta_0 = \beta \quad (6.63)$$

la care se adaugă ordinul de predicție, rezultând în total trei parametri independenți. Am denumit acest filtru LTP3.

Prin condițiile (6.62) și (6.63) coeficienții filtrului devin simetrici față de valoarea centrală, iar semnalul de excitație se poate scrie, având în vedere (6.52):

$$d(n) = \alpha\beta d((n \bmod (L-1)) - (L-1)) + \beta d((n \bmod L) - L) + \alpha\beta d((n \bmod (L+1)) - (L+1)) \quad (6.64)$$

pentru  $0 \leq n \leq N-1$ .

Considerând cazul în care semnalul de excitație se prelungeste prin repetare periodică, eroarea se poate exprima similar cu (6.54) astfel:

$$e_w(n) = \tilde{s}_w(n) - (\alpha\beta\bar{d}(n, L-1) + \beta\bar{d}(n, L) + \alpha\beta\bar{d}(n, L+1)). \quad (6.65)$$

Eroarea pătratică rezultată este:

$$\varepsilon = \sum_{n=0}^{N-1} e_w^2(n) \quad (6.66)$$

și minimizarea acestei erori în raport cu  $\alpha$  și  $\beta$ , prin anularea derivatelor în raport cu acești parametri, conduce la un sistem de forma:

$$\begin{bmatrix} A & B \\ B & D \end{bmatrix} \begin{bmatrix} \alpha\beta \\ \beta \end{bmatrix} = \begin{bmatrix} E \\ F \end{bmatrix} \quad (6.67)$$

unde notațiile introduse au semnificația:

$$A = \phi(L-1, L-1) + \phi(L+1, L+1) + 2\phi(L-1, L+1), \quad (6.68)$$

$$B = \phi(L-1, L) + \phi(L, L+1), \quad (6.69)$$

$$D = \phi(L, L), \quad (6.70)$$

$$E = \phi(0, L-1) + \phi(0, L+1), \quad (6.71)$$

$$F = \phi(0, L), \quad (6.72)$$

$$\phi(i, j) = \begin{cases} \sum_{n=0}^{N-1} \bar{d}(n-i)\bar{d}(n-j), & i, j \neq 0, \\ \sum_{n=0}^{N-1} \tilde{s}_w(n)\bar{d}(n-j), & i = 0, j \neq 0, \\ \sum_{n=0}^{N-1} (s_w(n))^2, & i, j = 0, \end{cases} \quad (6.73)$$

Folosind regula lui Cramer, am obținut soluțiile sistemului (6.67):

$$\beta = \frac{AF - BE}{AD - B^2} \quad (6.74)$$

$$\alpha\beta = \frac{DE - BF}{AD - B^2} \quad (6.75)$$

### 6.5.3 Filtru LTP cu constrângeri cu doi parametri – LTP2

În relația (6.62) am impus condiția suplimentară  $\alpha = const$  și astfel am obținut un filtru LTP cu trei coeficienți cu constrângeri suplimentare, care se poate descrie doar prin doi parametri, și anume ordinul  $L$  și coeficientul  $\beta_0$ . Am denumit filtrul obținut în acest caz filtru LTP2.

Singura necunoscută în ecuația (6.67) rămâne  $\beta$ , pe care am obținut-o particularizând (6.74) și ținând cont că  $\alpha$  este o constantă:

$$\beta_{opt} = \frac{\alpha\phi(0, L-1) + \phi(0, L) + \alpha\phi(0, L+1)}{\alpha^2 A + \phi(L, L) + 2\alpha B}, \quad (6.76)$$

unde  $A$  și  $B$  sunt definiți în (6.68) și (6.69).

În acest caz, este posibil ca ponderile atribuite celor doi coeficienți extremi să nu fie egale:

$$\beta_{-1} = \alpha_{-1}\beta_0, \quad \beta_1 = \alpha_1\beta_0, \quad (6.77)$$

unde, pentru simetrie, ponderile s-au notat  $\alpha_{-1}$  și  $\alpha_1$ .

Cu aceasta, matricea necunoscutelor (6.59) devine succesiv:

$$\beta = \begin{bmatrix} \beta_{-1} \\ \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \alpha_{-1}\beta_0 \\ \beta_0 \\ \alpha_1\beta_0 \end{bmatrix} = \beta_0 \begin{bmatrix} \alpha_{-1} \\ 1 \\ \alpha_1 \end{bmatrix} = \beta_0 \alpha \quad (6.78)$$

și soluția sistemului (6.56) va fi:

$$\beta_{opt} = \frac{\delta}{\phi \alpha} \quad (6.79)$$

Atât pentru filtrul LTP cu 3 coeficienți (6.61), pe care l-am denumit în continuare, în prezentarea rezultatelor experimentale, filtrul LTP original, cât și pentru cele două variante modificate LTP3 și LTP2 pe care le-am propus, cu constrângerile prezentate, se poate aplica interpolarea în vederea unei creșteri a rezoluției în timp și a reprezentării mai exacte a semnalului extras din dicționarul adaptiv.

#### 6.5.4 Răspunsul în frecvență al filtrului LTP

Răspunsul în frecvență al filtrului LTP cu trei parametri este, pentru cazul general:

$$G(e^{j\omega}) = \frac{1}{1 - \beta_{-1}e^{j\omega(L-1)} - \beta_0e^{j\omega L} - \beta_1e^{j\omega(L+1)}} \quad (6.80)$$

cu modulul:

$$|G(e^{j\omega})|^2 = \frac{1}{[\cos \omega L - \beta_0 - (\beta_{-1} + \beta_1) \cos \omega]^2 + [(\beta_1 + \beta_{-1}) \sin \omega + \sin \omega L]^2} \quad (6.81)$$

care, pentru cazul particular LTP3 devine:

$$|G(e^{j\omega})|^2 = \frac{1}{[\cos \omega L - \beta - 2\alpha\beta \cos \omega]^2 + [\sin \omega L]^2} \quad (6.82)$$

punând în evidență un caracter crescător pentru înfășurătoarea anvelopei spectrale dacă  $\alpha < 0$  și un caracter descrescător, dacă  $\alpha > 0$ .

Pentru cazul filtrului LTP2, caracteristica are întotdeauna o anvelopă de formă descrescătoare, dată de:

$$|G(e^{j\omega})|^2 = \frac{1}{[\cos \omega L - \beta(1 + 2\alpha \cos \omega)]^2 + [\sin \omega L]^2} \quad (6.83)$$

#### 6.5.5 Stabilitatea filtrului de sinteză LTP

Stabilitatea filtrului de sinteză LTP definit prin relația (6.3) este o problemă importantă în codorul CELP, datorită faptului că un filtru instabil poate produce perturbații ale semnalului de ieșire, manifestate sub forma unor „pocnituri”. În plus, principiul LTP presupune ca o porțiune din semnalul de la ieșirea filtrului LTP memorat în dicționarul adaptiv, porțiune aleasă corespunzător, să fie aplicată încă odată la intrarea filtrului LTP în unul din cadrele imediat următoare. În cazul unui filtru instabil, aceasta conduce la o propagare a perturbației datorate instabilității filtrului în anumite cadre de semnal, și în cadrele în care filtrul ar putea fi stabil.

Stabilitatea este asigurată dacă polii filtrului, respectiv zerourile numitorului, se află în interiorul cercului unitate în planul  $z$ . Polinomul care reprezintă numitorul în (6.3) deși are un grad ridicat, egal cu ordinul LTP, are foarte puțini coeficienți diferiți de zero.

Două condiții de stabilitate pentru filtrul LTP cu trei coeficienți sunt deduse în [RPK87].

Prima condiție este o condiție de suficiență și anume: filtrul este stabil dacă coeficienții au modul mai mic decât unitatea.

A doua condiție este mai restrictivă. Conform acestei condiții, filtrul este stabil în următoarele cazuri:

- dacă  $|a| \geq |b|$ , condiția de stabilitate este:

$$|\beta_{-1}| + |\beta_0| + |\beta_1| < 1. \quad (6.84)$$

- dacă  $|a| < |b|$ , condiția de stabilitate este:

$$|\beta_0| + |a| < 1, \quad (6.85)$$

sau

$$b^2 < |a| \text{ sau } b^2\beta_0^2 - (1 - b^2)(b^2 - a^2) < 0, \quad (6.86)$$

unde s-a notat  $a = \beta_{-1} + \beta_1$  și  $b = \beta_{-1} - \beta_1$

Pentru filtrele modificate pe care le-am propus LTP3 și LTP2, precum și pentru filtrul cu un coeficient, condiția de stabilitate se reduce la condiția (6.84), deoarece se poate ușor verifica că în toate cazurile coeficienții îndeplinesc condiția  $|a| \geq |b|$ .

Pentru a asigura stabilitatea în toate cazurile, se introduce un coeficient de stabilitate  $F$ , astfel definit, pe baza (6.85), (6.86) și (6.87), încât pentru un filtru stabil să fie îndeplinită relația  $F < 1$ , condiția de stabilitate fiind astfel simplu de verificat. Coeficientul de stabilitate se calculează astfel:

$$F = |\beta_{-1}| + |\beta_0| + |\beta_1| \quad \text{dacă } |a| > |b|, \quad (6.87)$$

$$F = \sqrt{\frac{b^4 + b^2\beta_0^2 - b^2a^2}{b^2 - a^2}} \quad \text{dacă } b^2 > |a|, \quad (6.88)$$

$$F = |\beta_0| + |a| < 1 \quad \text{dacă } b^2 \leq |a|. \quad (6.89)$$

Testul stabilității presupune calculul coeficientului  $F$  și verificarea condiției  $F < 1$ .

Pentru a asigura stabilitatea în cazul unui filtru instabil se pot aplica mai multe metode [RPK87]. Astfel, se poate aplica o deplasare radială a polilor astfel încât aceștia să se încadreze în cercul unitate. Această metodă presupune calcule laborioase, pentru determinarea optimă a coeficienților care asigură această cerință. O altă metodă constă în înlocuirea fiecărui pol exterior cercului unitate cu simetricul său față de acest cerc. Deși în acest caz se păstrează forma răspunsului în frecvență al filtrului, sunt necesare factorizări ale unui polinom de grad relativ mare.

Metoda pe care am folosit-o pentru stabilizare este simplă și nu necesită calcule semnificative, constând în scalarea fiecărui coeficient cu valoarea  $1/F$ , dacă condiția de stabilitate nu este îndeplinită. Ca rezultat, predictorul va fi suboptimal, având eroarea pătratică mai mare decât un predictor optimal cu cantitatea:

$$(1 - 1/F)\beta_{opt}^T \delta \quad (6.90)$$

în care  $\beta$  și  $\delta$  sunt definite în (6.59) și (6.57).

## 6.6 Rezultate experimentale

Pentru a compara performanțele diferitelor modele de filtre LTP am simulat în Matlab un codor CELP, conform schemelor bloc generale prezentate în cap. 3. Ca semnale de intrare am folosit semnale alese din baza de date descrisă în cap. 5.

Frecvența de eșantionare a semnalelor a fost 8 kHz, corespunzător unei perioade de eșantionare de 125  $\mu$ s. Această valoare constituie totodată și rezoluția în timp a filtrului LTP fără interpolare.

Dicționarul adaptiv a fost implementat în două variante, cu dimensiunea totală de 200 sau 240 eșantioane. În ambele cazuri am folosit 4 subcadre cu dimensiunea egală cu  $\frac{1}{4}$  din dimensiunea dicționarului, respectiv 50 sau 60 eșantioane.

Analiza pe termen scurt STP am făcut-o în toate cazurile odată pentru fiecare subcadru. Parametrii LPC au fost convertiți în LSP și cuantizați scalar neliniar conform standardului FS 1016. Eroarea de predicție a fost ponderată cu coeficientul  $\gamma = 0,8$ .

Dicționarul fix pe care l-am folosit are mărimea 256 vectori a 50, respectiv 60 eșantioane, și a fost generat aleator și memorat, folosind în toate experiențele același dicționar.

Ca măsură a performanțelor am folosit câștigul de predicție, definit în cap. 3.

De remarcat că am constatat diferențe semnificative în privința valorii câștigului de predicție obținut, între semnale vocale produse de vorbitori masculini și respectiv feminini, dar comparativ toate tipurile de filtre LTP studiate s-au încadrat în forma generală prezentată și performanțele comparative s-au păstrat. În general, pentru vocile masculine câștigul de predicție obținut în experimentările pe care le-am făcut a fost mai mic, cu valori cuprinse între 0.8 și 1,5 dB, dependent de semnal. Pentru măsurătorile de performanță făcute s-au ales numai cadre de semnal sonor, adică acele tipuri de cadre în care intervine modelarea excitației cu ajutorul filtrului LTP.

În figura 6.12 am reprezentat variația câștigului de predicție în funcție de ordinul de interpolare pentru filtrul LTP original, cu trei coeficienți, și pentru filtrele LTP3 și LTP2 pe care le-am propus.

Am realizat interpolarea pentru ordine fracționare de la 2 până la 8, pentru care am folosit funcțiile de interpolare de forma (6.48) – (6.49). Cu valoarea 1 am notat varianta filtrului cu ordin întreg, fără interpolare.

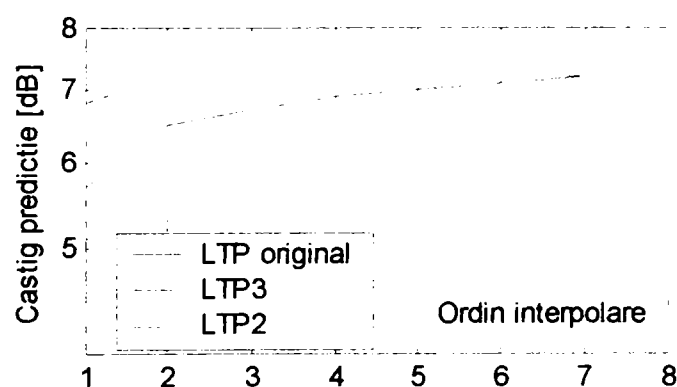


Fig. 6.12. Câștigul de predicție în funcție de ordinul de interpolare

Am constatat că filtrul LTP original, cu trei coeficienți independenți, are un câștig practic constant pentru ordine fracționare peste 2. Folosirea interpolării crește performanțele atât pentru filtrul LTP3 cât și pentru LTP2, rezultând că pentru ordin peste 6, practic filtrul LTP3, pe care l-am propus, egalează filtrul LTP original, economisind însă un coeficient.



Filtrul LTP 3 are în mod constant performanțe superioare datorită adaptării valorii coeficientului  $\alpha$  față de filtrul LTP2, la care  $\alpha$  este fix.

În figura 6.13 am reprezentat variația câștigului de predicție în funcție de valoarea coeficientului  $\alpha$ , în domeniul  $0 \div 0,5$ , pentru filtrul LTP2.

Pentru referință am reprezentat și filtrul LTP original, în varianta cu și fără interpolare.

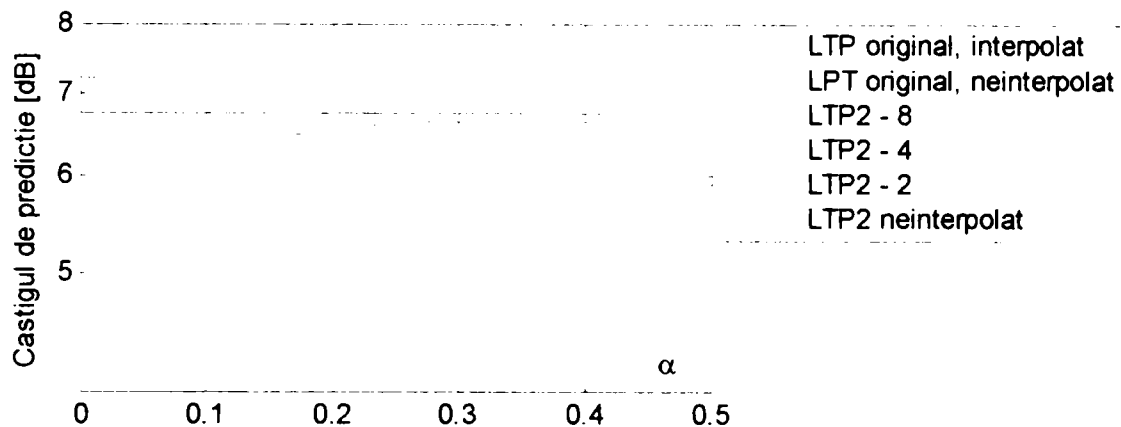


Fig. 6.13. Câștigul de predicție în funcție de coeficientul  $\alpha$

Filtrul LTP2 cu ordinul de interpolare 8 depășește ca performanță filtrul LTP original fără interpolare, iar filtrul LTP2 cu interpolare de ordin 4 se apropie ca performanță de acesta. Pentru interpolare cu 4, respectiv 2, performanțele sunt corepunzător mai scăzute.

În tabelul 6.2 am sintetizat performanțele generale ale filtrelor considerate, numai pentru ordin întreg, fără interpolare.

În tabel sunt indicate, ca măsuri comparative pentru performanță și raportul semnal – zgomot SNR precum și raportul semnal – zgomot segmental SEGSNR.

Tabelul 6.2 Performanțele generale ale filtrelor LTP3 și LTP2

Tipul predictorului	Câștig [dB]	SNR [dB]	SEGSNR [dB]
LTP un coeficient (S)	5,73	7,89	7,49
LTP un coeficient (NS)	5,80	7,74	7,68
LTP trei coeficienți (S)	6,75	7,84	7,56
LTP trei coeficienți (NS)	6,83	4,66	7,95
LTP3 (S)	5,08	6,79	7,02
LTP3 (NS)	5,87	4,97	6,86
LTP2 (S), $\alpha = 0,1$	5,12	7,19	7,12
LTP2 (S), $\alpha = 0,15$	5,48	7,91	7,54
LTP2 (S), $\alpha = 0,2$	5,23	7,16	7,10
LTP2 (NS), $\alpha = 0,1$	5,48	7,38	7,15
LTP2 (NS), $\alpha = 0,15$	5,72	7,98	7,65
LTP2 (NS), $\alpha = 0,2$	5,51	7,45	7,22

În tabelul 6.2 am considerat, pentru fiecare tip de filtru, două situații: filtrul nestabilizat (NS), cu coeficienții rezultați din calcul, și filtrul stabilizat (S), la care după calculul coeficienților am verificat indicele de stabilitate și în cazul în care filtrul a rezultat instabil, am aplicat stabilizarea prin scalarea parametrilor.

Am constatat că stabilizarea are ca efect în toate cazurile o scădere ușoară a performanțelor, datorită modificării coeficienților la valori suboptimale prin scalare.

În figura 6.14 am reprezentat un cadru de semnal (sus) și cel mai bun semnal de excitație (jos), ales din dicționarul adaptiv și filtrat prin filtrul LTP, pentru două variante de filtru, și anume filtrul LTP cu un coeficient și filtrul LTP2, corespunzător valorii  $\alpha = 0.15$ .

Pentru comparație am reprezentat, în fig. 6.14 sus, alături de semnalul original și semnalul de excitație după filtrarea STP.

În figura 6.15 am reprezentat un cadru similar, cu dimensiunea de 50 eșantioane, împreună cu întreg conținutul dicționarului adaptiv, cu dimensiunea de 200 de eșantioane, indicându-se porțiunea de semnal din dicționarul adaptiv care constituie excitația.

În figura 6.16 am reprezentat semnalul original și contribuția semnalului extras din dicționarul adaptiv, filtrat prin filtrul de sinteză STP (sus) și pentru comparație, semnalul diferență (jos) între aceste două semnale, pe o durată de timp corespunzătoare cu 10 cadre de câte 60 de eșantioane. Se constată caracterul asemănător unui zgomot al acestui semnal diferență, care va fi modelat de dicționarul fix.

De remarcat că în funcționarea normală a codorului CELP, la semnalul extras din dicționarul adaptiv și filtrat prin filtrul LTP, se adaugă contribuția dicționarului fix, înmulțită cu coeficientul de câștig corespunzător, pentru a genera, prin filtrare STP, semnalul de ieșire.

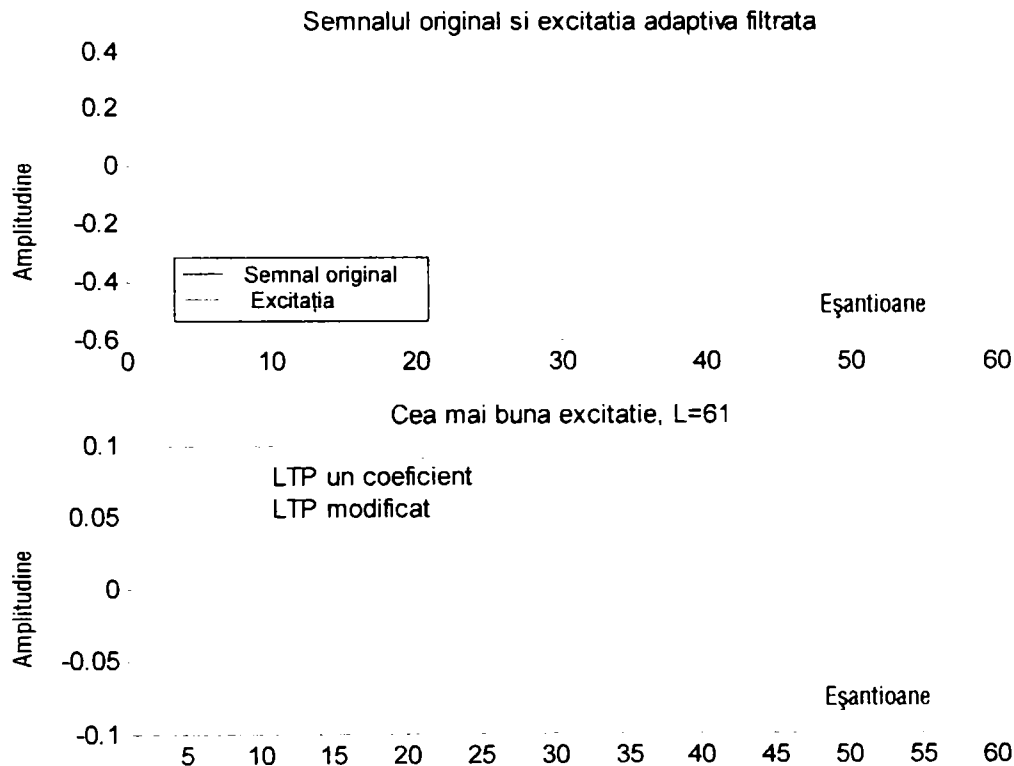


Fig. 6.14. Semnalul de excitație și semnalul original

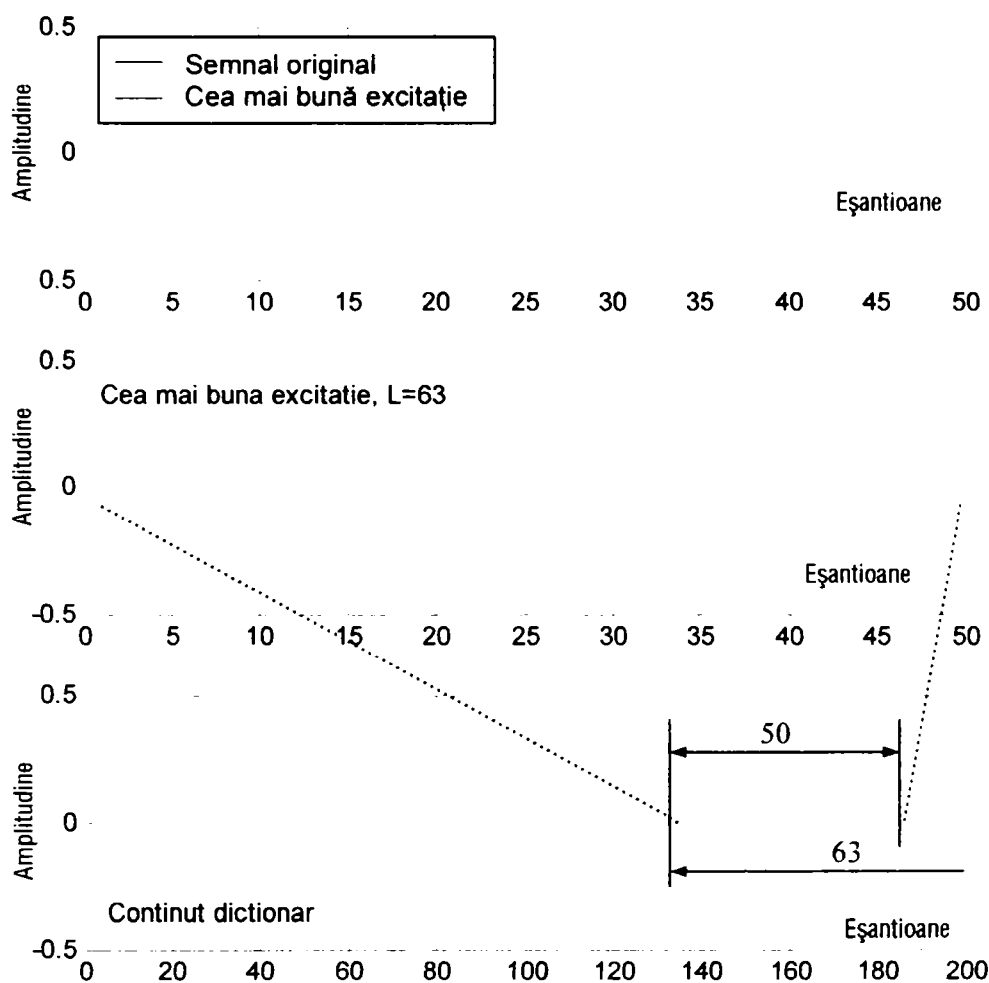


Fig. 6.15. Semnalul original, cea mai bună excitație și conținutul dicționarului

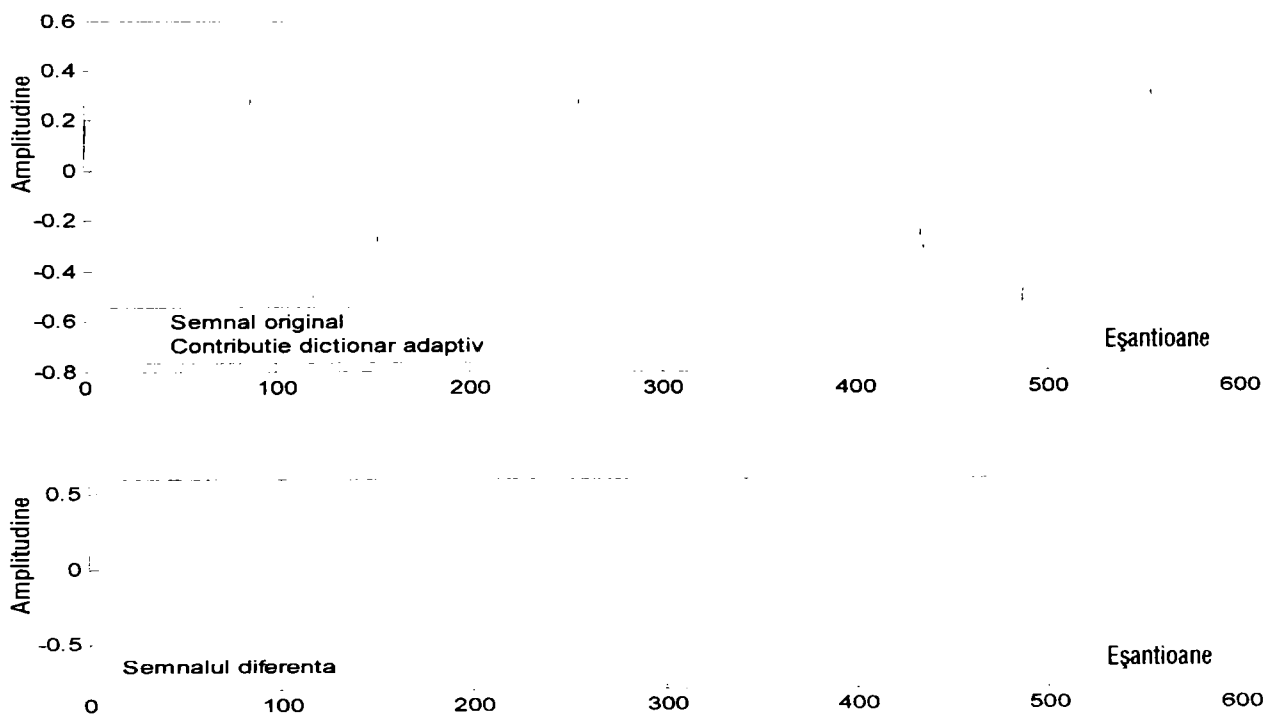


Fig. 6.16. Semnalul original, contribuția dicționarului adaptiv și semnalul diferență

# Capitolul 7

## *Cuantizarea vectorială a semnalului rezidual în codorul CELP*

### 7.1 Dicționarul fix în codorul CELP

Semnalul eroare de predicție sau semnalul rezidual care se obține după extragerea informațiilor privind spectrul pe termen scurt (STP), cu filtrul  $A(z)$  de forma (3.6) (Cap.3), și după extragerea informațiilor privind periodicitatea reziduală, prin filtrare cu un filtrul de predicție pe termen lung (LTP)  $P(z)$  de forma (6.2) (Cap. 6) este, în cazul ideal, un zgomot alb. Pentru acest semnal se folosește și termenul de inovație [STA98], sau al doilea reziduu de predicție, deoarece este obținut prin două filtrări succesive. Operațiunea inversă, de filtrarea a zgomotului alb (inovație) prin cele două filtre de sinteză, pe termen lung și pe termen scurt, va conduce la obținerea unui semnal vocal sintetic, foarte apropiat de semnalul vocal original. Această operație stă la baza modelului producerii semnalului vocal folosit în codorul CELP, model reprezentat în figura 3.10 (Cap.3).

Acest semnal rezidual nu poate fi neglijat, deși are practic caracteristicile unui zgomot alb și nu mai păstrează nimic din informațiile conținute în semnalul vocal original. Prin înlocuirea acestui reziduu de predicție cu un fragment corespunzător de zgomot alb și folosirea aceluiași fragment pentru obținerea semnalului vocal cu modelul din fig 3.10 va rezulta un semnal cu un pronunțat caracter artificial, deși modelul, caracterizat prin parametrii celor două filtre, este actualizat la fiecare cadru.

Pentru modelarea semnalului rezidual, codorul CELP folosește un dicționar, denumit dicționar stochastic sau dicționar fix. Denumirea de dicționar stochastic este justificată de caracterul de zgomot gaussian al vectorilor pe care îi conține, iar denumirea de dicționar fix sugerează caracterul stabil al acestui dicționar, spre deosebire de dicționarul adaptiv, care se actualizează după fiecare subcadru. Există și posibilitatea de a coda și transmite separat semnalul rezidual [PML95].

Dicționarul fix este disponibil atât la codare cât și la decodare și deci nu trebuie transmis. Ceea ce se transmite reprezintă doar indexul cuvântului de cod selectat. În varianta cea mai simplă, acest dicționar nu este structurat și din acest motiv căutarea în acest dicționar presupune un număr ridicat de operații de calcul [CHO95], [CBA04-1], [CBA04-4].

Având în vedere modul de prelucrare pe cadre și subcadre a semnalului în codorul CELP, vectorii de semnal gaussian conținuți în dicționarul fix vor avea dimensiunea egală cu dimensiunea subcadrului.

Dicționarul fix standard pentru codorul CELP descris în cap. 3 este reprezentat în figura 7.1 a) și constă într-o structură de tip matrice cu dimensiunea  $V \times N$ , care conține  $V$  vectori de cod cu  $N$  componente fiecare [CHU03-2]. Necesarul de memorie pentru păstrarea acestui dicționar este  $V \times N$  locații, fiecare locație având un număr corespunzător de biți pentru memorarea elementelor vectorilor. În unele aplicații dimensiunea relativ mare a acestei

memorii poate constitui o problemă. Astfel, pentru valorile  $V = 512$  și  $N = 60$ , valori care pot fi considerate tipice pentru un codor CELP, necesarul de memorie pentru dicționarul fix este de 35.720 locații.

Pentru a reduce capacitatea memoriei folosită pentru păstrarea dicționarului fix, se poate organiza acest dicționar în forma prezentată în figura 7.1 b). În această variantă dicționarul este realizat sub forma unei matrici unidimensionale, în care vectorii de cod consecutivi sunt decalajați cu  $S$  eșantioane, iar partea de vector de mărime  $S - N$  este comună cu vectorul următor. Din acest motiv, structura din fig. 7.1 b se numește dicționar cu vectori suprapuși (overlapping).

Pentru o valoare  $S$  a decalajului între doi vectori consecutivi, capacitatea totală a memoriei necesară pentru stocarea celor  $V$  vectori de mărime  $N$  este  $S(V - 1) + N$ . Varianta din fig. 7.1 a se poate considera un caz particular al structurii suprapuse, cu pasul  $S = N$ .

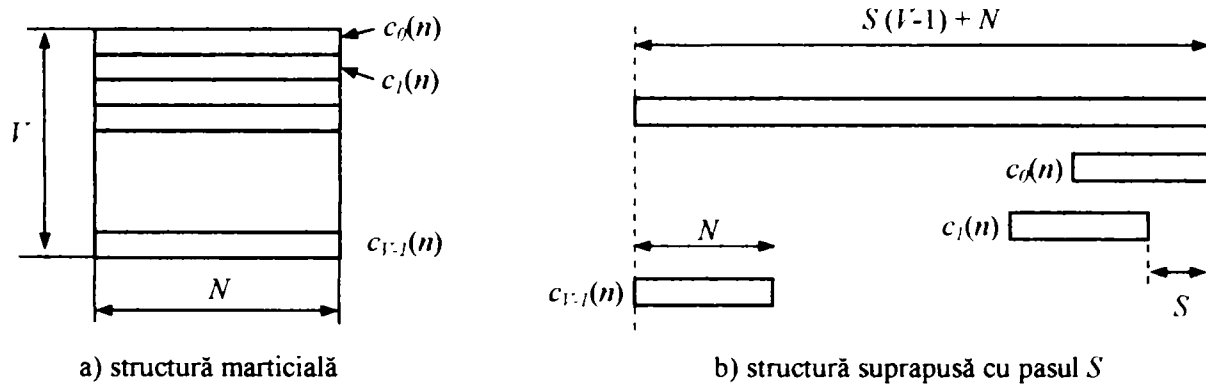


Fig. 7.1. Organizarea dicționarului fix

În afara avantajului evident privind dimensiunea memoriei, structura de dicționar fix cu vectori suprapuși permite reducerea volumului de calcul necesar găsirii celui mai bun vector de excitație, prin aplicarea recursivității asupra operației de convoluție între vectorul din dicționar și răspunsul la impuls al filtrului WSTP.

### 7.2 Operația de căutare în dicționarul fix

Căutarea celui mai bun vector de excitație din dicționarul fix este o operație similară căutării în dicționarul adaptiv. În figura 7.2 este reprezentat codorul CELP în care sunt evidențiate cele două dicționare, respectiv dicționarul adaptiv și dicționarul fix, și modul de formare a semnalului vocal sintetizat local în codor  $\hat{s}_w(n)$ .

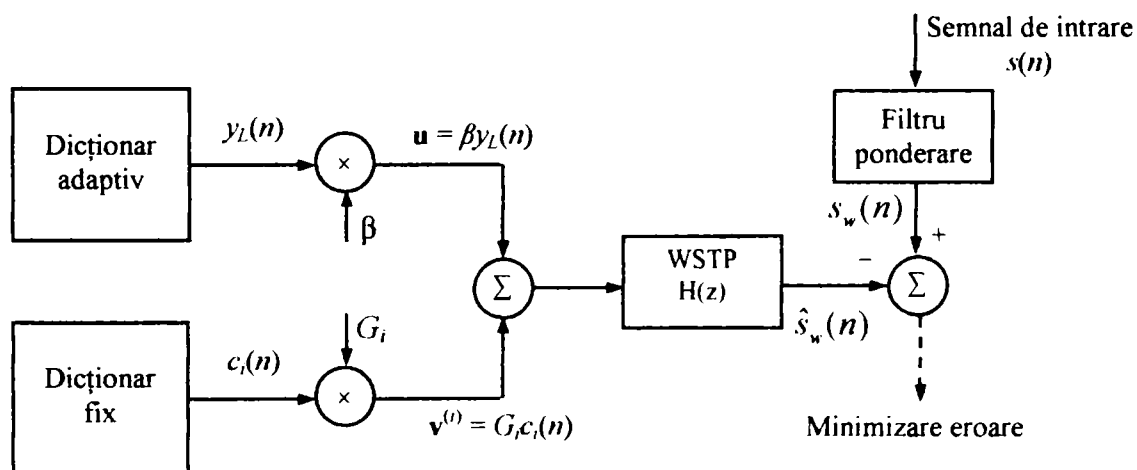


Fig. 7.2. Operația de căutare în dicționarul fix

Utilizarea celor două dicționare pentru a modela cât mai exact semnalul de excitație în codorul CELP presupune, în cazul ideal, o căutare conjugată a celor patru parametri care participă la construcția semnalului de excitație și anume cei doi vectori  $y_l(n)$  din dicționarul adaptiv și  $c_i(n)$  din dicționarul fix, precum și factorii de câștig asociați  $\beta$  și  $G_i$ . Pentru aceasta, pe durata unui subcadru, trebuie testați toți vectorii  $y_l(n)$ , din dicționarul adaptiv, și pentru fiecare valoare posibilă  $l = l_{min}, \dots, l_{max}$ , se testează fiecare din cei  $V$  vectori din dicționarul fix. Această procedură conduce la un volum inacceptabil de mare de calcule, și din acest motiv se practică o căutare suboptimală, așa cum s-a prezentat în cap. 6. Căutarea suboptimală separă căutarea în cele două dicționare și execută mai întâi căutarea în dicționarul adaptiv, considerând că semnalul provenit de la dicționarul fix este zero, după care, folosind valorile parametrilor astfel determinate pentru dicționarul adaptiv, execută căutarea în dicționarul fix.

Operația de căutare în dicționarul fix se face în sensul minimizării erorii dintre semnalul vocal de intrare, ponderat cu filtrul de ponderare perceptuală, și semnalul vocal sintetizat local în codor, prin filtrarea excitației prin filtrul de sinteză pe termen scurt ponderat WSTP [CHU03-2].

Se notează cu  $\mathbf{v}$  vectorul de excitație corespunzător dicționarului fix, cu indicele superior reprezentând indexul din dicționar, și cu  $\mathbf{u}$  vectorul de excitație extras din dicționarul adaptiv. Vectorul  $\mathbf{v}$  se poate scrie:

$$\mathbf{v}^{(i)} = G_i \mathbf{c}^{(i)}, \text{ pentru } i = 1, 2, \dots, V. \quad (7.1)$$

Fie  $\mathbf{H}$  matricea cu dimensiunea  $N \times N$ , de formă triunghiulară inferioară, ale cărei coloane conțin răspunsul la impuls al filtrului de predicție STP

$$\mathbf{H} = \begin{bmatrix} h_0 & 0 & 0 & \dots & 0 \\ h_1 & h_0 & 0 & & 0 \\ \vdots & \ddots & \ddots & \dots & 0 \\ h_{N-1} & h_{N-2} & \dots & h_0 & 0 \\ h_N & h_{N-1} & \dots & h_1 & h_0 \end{bmatrix} \quad (7.2)$$

și  $\mathbf{W}$  matricea cu aceleași dimensiuni și formă, care conține răspunsul la impuls al filtrului de ponderare perceptuală a erorii:

$$\mathbf{W} = \begin{bmatrix} w_0 & 0 & 0 & \dots & 0 \\ w_1 & w_0 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \dots & 0 \\ w_{N-1} & w_{N-2} & \dots & w_0 & 0 \\ w_N & w_{N-1} & \dots & w_1 & w_0 \end{bmatrix}. \quad (7.3)$$

Folosind reprezentarea filtrelor prin răspunsul la impuls în formă matricială, semnalul de la ieșirea filtrului se poate calcula ca produs de convoluție între semnalul de intrare și răspunsul la impuls, iar produsul de convoluție se poate exprima simplu prin produs de matrici. Astfel, semnalul obținut la ieșirea filtrului STP este:

$$\hat{\mathbf{s}}^{(i)} = \mathbf{H}(\mathbf{u} + \mathbf{v}^{(i)}) + \hat{\mathbf{s}}^{(0)}, \quad 1 \leq i \leq V \quad (7.4)$$

unde cu  $\hat{\mathbf{s}}^{(0)}$  s-a notat răspunsul filtrului STP la intrarea zero și s-a avut în vedere că vectorul  $\mathbf{u}$  este constant pentru operația de căutare în dicționarul fix.

Cu aceasta eroarea ponderată se poate scrie în forma:

$$\mathbf{e}^{(i)} = \mathbf{W}(\mathbf{s} - \hat{\mathbf{s}}^{(i)}) = \mathbf{e}^{(0)} - \mathbf{W}\mathbf{H}\mathbf{v}^{(i)} \quad (7.5)$$



unde cu  $\mathbf{e}^{(0)}$  s-a notat semnalul „țintă” în domeniul timp, exprimat prin:

$$\mathbf{e}^{(0)} = \mathbf{W}(\mathbf{s} - \hat{\mathbf{s}}^{(0)}) - \mathbf{W}\mathbf{H}\mathbf{u} \quad (7.6)$$

astfel că eroarea  $\mathbf{e}^{(i)}$  pentru vectorul cu indicele  $i$  se poate exprima ca diferență între semnalul țintă și semnalul extras din dicționar:

$$\mathbf{e}^{(i)} = \mathbf{e}^{(0)} - G_i \mathbf{y}^{(i)}, \quad (7.7)$$

unde

$$\mathbf{y}^{(i)} = \mathbf{W}\mathbf{H}\mathbf{x}^{(i)} \quad (7.8)$$

reprezintă vectorul din dicționarul fix după filtrare.

Eroarea pătratică sau norma, exprimată ca distanța între vectorul țintă și vectorul din dicționar este:

$$\begin{aligned} E_i = \|\mathbf{e}^{(i)}\|^2 &= \langle \mathbf{e}^{(i)}, \mathbf{e}^{(i)} \rangle = \mathbf{e}^{(i)T} \mathbf{e}^{(i)} = \\ &= \mathbf{e}^{(0)T} \mathbf{e}^{(0)} - 2G_i \mathbf{y}^{(i)T} \mathbf{e}^{(0)} + G_i^2 \mathbf{y}^{(i)T} \mathbf{y}^{(i)}, \end{aligned} \quad (7.9)$$

unde indicele superior  $T$  indică operația de transpunere a matricii.

Eroarea (7.9) depinde atât de indexul  $i$  cât și de câștigul  $G_i$  asociat acestuia. Pentru a minimiza eroarea, pentru un index dat, se calculează derivata în raport cu câștigul:

$$\frac{\partial E_i}{\partial G_i} = -2\mathbf{y}^{(i)T} \mathbf{e}^{(0)} + 2G_i \mathbf{y}^{(i)T} \mathbf{y}^{(i)}, \quad (7.10)$$

și se impune condiția ca această derivată să se anuleze:

$$\frac{\partial E_i}{\partial G_i} = 0, \quad (7.11)$$

Se obține câștigul asociat erorii minime, în forma:

$$G_i = \frac{\mathbf{y}^{(i)T} \mathbf{e}^{(0)}}{\mathbf{y}^{(i)T} \mathbf{y}^{(i)}}, \quad (7.12)$$

unde la numărător se recunoaște funcția de corelație între semnalul țintă și vectorul de cod filtrat, iar la numitor energia vectorului de cod filtrat.

Pentru a minimiza eroarea (7.9) aceasta se scrie în forma:

$$E_i = \mathbf{e}^{(0)T} \mathbf{e}^{(0)} - G_i (2\mathbf{y}^{(i)T} \mathbf{e}^{(0)} + G_i \mathbf{y}^{(i)T} \mathbf{y}^{(i)}), \quad (7.13)$$

și se observă că minimizarea erorii (7.13) este echivalentă cu maximizarea celui de-al doilea termen al expresiei (7.13), deoarece primul termen este o constantă. Termenul care urmează a fi maximizat, denumit și scor, este :

$$m_i = G_i (2\mathbf{y}^{(i)T} \mathbf{e}^{(0)} - G_i \mathbf{y}^{(i)T} \mathbf{y}^{(i)}), \quad (7.14)$$

și înlocuind expresia câștigului (7.12) în expresia (7.14), scorul care urmează a fi maximizat devine:

$$m_i = \frac{(\mathbf{y}^{(i)T} \mathbf{e}^{(0)})^2}{\mathbf{y}^{(i)T} \mathbf{y}^{(i)}}, \quad (7.15)$$

în care se recunosc expresiile corelației și energiei vectorului de cod filtrat, care intervin și în expresia câștigului optim.

În concluzie, căutarea în dicționarul stochastic presupune 3 operații:

- convoluția între cuvântul de cod și răspunsul filtrului de ponderare, conform relației (7.8),
- calculul numărătorului din relația (7.15), operație echivalentă cu un produs de corelație între doi vectori,
- calculul numitorului din relația (7.15), care reprezintă energia unui semnal.

### 7.3 Complexitatea căutării în dicționarul fix

Aceste trei operații se execută pentru fiecare vector din dicționarul fix, conducând astfel la un volum relativ mare de calcule.

Pentru cazul general, al unui dicționar fix care conține  $V$  vectori cu dimensiunea  $N$  și răspunsul la impuls are de asemenea  $N$  termeni, numărul de operații de înmulțire – împărțire și respectiv de adunare – scădere, pentru un singur vector din dicționar, este prezentat în tabelul 7.1.

Tabelul 7.1. Complexitatea de calcul pentru cazul general

Operație	Înmulțiri – împărțiri		Adunări – scăderi	
	Cazul general	$N = 60$	Cazul general	$N = 60$
Convoluție	$N \times (N+1)/2$	1830	$(N - 1) \times N/2$	1770
Corelație	$N$	60	$N - 1$	59
Energie	$N$	60	$N - 1$	59
Câștig	1	1	0	0
Scor	1	1	0	0
Total	$(N \times (N+1)/2) + 2N + 2$	1952	$((N - 1) \times N/2) + 2N - 2$	1888

Considerând că pentru un procesor de semnal, care dispune de instrucțiuni directe de înmulțire – împărțire, complexitatea acestor instrucțiuni este echivalentă cu a celor de adunare sau scădere, numărul total de operații pentru un vector din dicționar devine  $N^2 + 2N$ . Pentru un total de  $V$  vectori în dicționarul fix, și considerând că prelucrarea se face pe subcadre cu durata de  $C$  eșantioane, la frecvența de eșantionare  $F_e$ , rezultă în total un număr de operații pe secundă:

$$O = \frac{F_e}{C} V (N^2 + 2N) \tag{7.16}$$

Pentru cazul  $V = 512$ ,  $F_e = 8000$  Hz și  $C = 60$  eșantioane, se obține conform expresiei (7.16) un număr de operații  $O = 253.952.000$  operații/secundă.

Sunt cunoscute mai multe metode care permit reducerea volumului de calcule [TAT90], [RDS94].

O primă metodă constă în a reduce dimensiunea răspunsului la impuls al filtrului implicat în operația de convoluție. Dacă se consideră că răspunsul la impuls are doar  $W$  componente nenule, atunci pentru convoluție sunt necesare un număr de înmulțiri:

$O_1 = 1 + 2 + 3 + \dots + W + W + W + \dots + W$ , în total  $N$  termeni, sau:

$$O_1 = \frac{W(W+1)}{2} + W(N-W) = \frac{W(2N-W+1)}{2} \tag{7.17}$$

și un număr de adunări:

$O_2 = 1 + 2 + 3 + \dots + (W-1) + (W-1) + (W-1) + \dots + (W-1)$ , în total  $N-1$  termeni, respectiv:

$$O_2 = \frac{(W-1)W}{2} + (W-1)(N-W) = \frac{(W-1)(2N-W)}{2} \tag{7.18}$$

Pentru corelație, energie, câștig și scor rămân valabile și în acest caz valorile din tabelul 7.1. Această metodă este aplicată în standardul FS 1016, în care produsul celor două matrici  $\mathbf{W}$  și  $\mathbf{H}$  implicate în rel. (7.8) este precalculat [CHU03-2]. Notând matricea rezultantă cu  $\mathbf{M}$ , se observă că  $\mathbf{M}$  are aceeași formă triunghiulară ca și matricile  $\mathbf{W}$  și  $\mathbf{H}$ , dar, limitând lungimea răspunsului la impuls la 30, un număr de 450 elemente ale aceste matrici devin nule:

$$\mathbf{M} = \mathbf{HW} = \begin{bmatrix} m_0 & 0 & 0 & 0 & 0 & \dots & 0 \\ m_1 & m_0 & \ddots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & 0 & 0 & \dots & 0 \\ m_{29} & m_{28} & \dots & m_0 & 0 & \dots & 0 \\ 0 & m_{29} & \dots & m_1 & m_0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & m_{29} & m_{28} & \dots & m_0 \end{bmatrix}. \quad (7.19)$$

În acest caz, aplicând relațiile (7.17) și (7.18) se obține  $O_1 = 1830$  și  $O_2 = 1305$ , ceea ce reprezintă o reducere de 26% față de cazul numeric din tabelul 7.1, pentru aceeași valoare  $N$ .

A doua metodă de reducere a volumului de calcul constă în folosirea unei structuri de dicționar fix cu suprapunerea parțială a vectorilor de cod consecutivi, așa cum a fost ilustrat în fig. 7.1 b. Dacă pentru structura matricială a dicționarului volumul de calcul este același pentru fiecare vector din dicționar, pentru structura suprapusă calculele se fac în mod diferit pentru primul vector din dicționar față de următorii.

Astfel, pentru primul vector sunt valabile relațiile prezentate în tabelul 7.1, cu o eventuală reducere conform relațiilor (7.17) și (7.18) dacă se consideră o lungime mai mică pentru răspunsul la impuls.

Pentru următorii  $V - 1$  vectori din dicționar se poate aplica un calcul recursiv al mărimilor care intervin în rel. (7.15), având în vedere că fiecare nou vector din dicționar diferă de precedentul doar prin  $S$  eșantioane, iar restul de  $N - S$  eșantioane sunt identice cu cele din vectorul de cod precedent. În tabelul 7.2 este prezentat volumul de calcule pentru fiecare operație, pentru toți vectorii din dicționarul fix, cu excepția primului, pentru cazul răspunsului filtrului de lungime  $N$  sau trunchiat la lungimea  $W$ .

Tabelul 7.2. Complexitatea de calcul pentru toți vectorii, cu excepția primului

Operație	Înmulțiri – împărțiri		Adunări – scăderi	
	Cazul general	Răspuns trunchiat	Cazul general	Răspuns trunchiat
Convoluție	$S(2N - S + 1)/2$	$S(2W - S + 1)/2$	$(S - 1)(2N - S)/2$	$(S - 1)(2W - S)/2$
Corelație	$N$	$N$	$N - 1$	$N - 1$
Energie	$N$	$N$	$N - 1$	$N - 1$
Câștig	1	1	1	1
Scor	1	1	1	1

O a treia categorie de metode care permit reducerea volumului de calcule se bazează pe organizarea particulară a dicționarului fix.

Dicționarul se poate organiza astfel încât să se poată realiza o căutare structurată, care să permită identificarea vectorului de cod optim într-un număr de pași relativ mic față de mărimea dicționarului, fără a fi necesară testarea tuturor vectorilor din dicționar [AMS93], [MDY94].

O reducere semnificativă a volumului de calcule se poate atinge dacă se folosește un dicționar ternar sau algebric, care conține doar valorile 0, +1 sau -1. În plus, matricea care

reprezintă dicționarul poate fi o matrice rară, la care procentul de elemente egale cu zero este relativ ridicat.

Un astfel de dicționar este folosit în codorul CELP FS 1016. Dicționarul fix din codorul FS 1016 cuprinde 512 vectori în organizare suprapusă cu pasul 2, conținând în total un număr de  $2 \times 511 + 60 = 1082$  valori ternare, din care 820, respectiv 76% sunt zero.

Avantajul major al dicționarului ternar constă însă în eliminarea operațiilor de înmulțire, care, datorită faptului că valorile nenule din dicționar sunt +1 sau -1, devin simple adunări sau scăderi. Acest avantaj este important în cazul unor implementări în care operațiile de înmulțire nu sunt implementate direct și necesită un număr de cicluri procesor mult mai mare decât operațiile de adunare – scădere.

În tabelul 7.3 am calculat exact numărul de operații de tip înmulțiri – împărțiri și respectiv adunări – scăderi necesare pentru căutarea în dicționarul fix în codorul FS 1016, ținând cont de structura exactă a dicționarului.

Tabelul 7.3. Complexitatea de calcul pentru codorul FS 1016

Operație	Înmulțiri – împărțiri		Adunări – scăderi	
	Primul vector	Următorii 511	Primul vector	Următorii 511
Convoluție	1365	$2 \times 511 = 1022$	1305	$29 \times 248 = 7192$
Corelație	60	$60 \times 511 = 30660$	59	$59 \times 511 = 30149$
Energie	60	$60 \times 220 + 2 \times 291 = 13782$	59	$59 \times 220 + 2 \times 291 = 13562$
Câștig	1	$1 \times 511 = 511$	0	0
Scor	1	$1 \times 511 = 511$	0	0
Total	1487	46486	1423	50903
Total general	47973		52326	

Aceste valori se vor folosi ca referință pentru aprecierea reducerii complexității de calcul pentru metodele de căutare în dicționarul fix pe care le-am propus.

## 7.4 Reducerea complexității de căutare în dicționarul fix

Analizând rel. (7.15) care definește cantitatea ce trebuie maximizată pentru a găsi indicele celui mai bun vector de cod din dicționarul fix, se pot identifica trei căi de reducere a complexității de calcul. Am luat în considerare fiecare din aceste căi pe rând, analizând posibilitățile de reducere a complexității de calcul pentru dicționarul ternar utilizat în codorul CELP FS 1016. Efectul privind reducerea complexității de calcul este indicat prin exprimarea complexității metodei aplicate în procente față de valorile din tabelul 7.3. Efectul fiecărei metode asupra calității semnalului este apreciat prin raportul semnal – zgomot.

### 7.4.1 Trunchierea răspunsului la impuls al filtrului

Convoluția între vectorul din dicționar și răspunsul la impuls al filtrului de sinteză ponderat este o operație care trebuie efectuată pentru toți vectorii și reducerea prin trunchiere a răspunsului la impuls al filtrului va avea efect asupra volumului de calcul.

Pe baza rel. (7.17), (7.18) și a tabelului 7.2, am calculat numărul de operații necesare pentru convoluție, pentru diferite lungimi ale răspunsului la impuls.

Rezultatele obținute prin aplicarea acestei metode sunt sintetizate în tabelul 7.4, în care, în coloana a treia, am indicat valoarea pe care am obținut-o pentru raportul semnal – zgomot în fiecare caz, în procente față de valoarea de referință cu lungimea răspunsului 30.

Tabelul 7.4. Complexitatea de calcul pentru convoluție în cazul trunchierii răspunsului și SNR obținut

Lungime răspuns	Reducerea complexității	SNR
30	100 %	100 %
20	73 %	95,1 %
10	39%	88,3 %

Deoarece se poate folosi calculul recursiv pentru toți vectorii cu excepția primului, reducerea volumului de calcul în ansamblu nu este prea mare.

Coloana a doua din tabelul 7.4 arată reducerea procentuală calculată pentru primul vector din dicționar.

O reducere de aceeași valoare se obține pentru următorii vectori numai dacă se folosește un dicționar fix cu vectorii organizați sub formă de matrice; în cazul dicționarului suprapus nu se obține pentru următorii vectori nici o reducere prin această metodă.

### 7.4.2 Simplificarea calculului corelației

Pentru un dicționar suprapus, cu valori ternare, cum este cel folosit în codorul FS 1016, se poate obține o reducere a volumului de calcul necesar pentru obținerea funcției de corelație care intervine la numărătorul expresiei (7.15), prin simplificarea calculelor exploatând faptul că valorile din dicționar sunt  $-1, 0$  sau  $+1$ .

În acest scop, am rescris relația (7.8) și având în vedere regulile de calcul cu matrici, am obținut următoarea formă pentru vectorul  $\mathbf{y}$ :

$$\mathbf{y}^{(i)T} = (\mathbf{W}\mathbf{H}\mathbf{x}^{(i)})^T = (\mathbf{M}\mathbf{x}^{(i)})^T = \mathbf{x}^{(i)T}\mathbf{M}^T \quad (7.20)$$

Folosind expresia (7.20), am rescris numărătorul din (7.15) astfel:

$$cor = \mathbf{y}^{(i)T}\mathbf{e}^{(0)} = \mathbf{x}^{(i)T}\mathbf{M}^T\mathbf{e}^{(0)}, \quad (7.21)$$

și am notat în (7.21)

$$\mathbf{t} = \mathbf{M}^T\mathbf{e}^{(0)}, \quad (7.22)$$

observând că  $\mathbf{t}$  reprezintă vectorul țintă definit de rel. 7.6 și filtrat.

Elementele care intervin în (7.22) sunt constante pe un subcadru și deci vectorul  $\mathbf{t}$  poate fi calculat la începutul fiecărui subcadru și memorat.

Cu aceasta, am exprimat corelația care intervine la numărătorul expresiei (7.15) în forma:

$$cor = \mathbf{x}^{(i)T}\mathbf{t}, \quad (7.23)$$

și având în vedere modul de organizare al dicționarului fix, am scris expresia corelației pentru vectorul cu indicele  $i$  din dicționar în forma:

$$cor^{(i)} = \sum_{j=0}^{59} x_{1025-2i+j} t_j \quad (7.24)$$

Deoarece valorile  $x$  din dicționar sunt valori ternare, adică  $+1, 0$  sau  $-1$ , calculul corelației conform rel. 7.24 se reduce la testarea valorii  $x$  și efectuarea unei adunări dacă  $x = +1$ , a unei scăderi dacă  $x = -1$ , respectiv nici o operație pentru cazul  $x = 0$ .

Calculul valorii corelației conform acestui algoritm se poate face printr-o subrutină care primește indicele  $i$  și vectorul  $\mathbf{t}$  definit în (7.22) și returnează valoarea funcției de autocorelație  $cor^{(i)}$ :

```

cor = 0
for j = 1 to 60
  if x(1025 - 2i + j) = - 1
    cor = cor - t(j)
  else if x(1025 - 2i + j) = 1
    cor = cor + t(j)
  end
end
end

```

Pentru un dicționar cunoscut, numărul de operații implicat în (7.24) se poate determina exact. Astfel, pentru dicționarul folosit în codorul FS 1016, am analizat fiecare vector de cod și am calculat că pentru primul vector de cod există în dicționar 15 valori nenule, deci se efectuează 15 operații de tip adunare – scădere. Pentru următorii 511 vectori din dicționar se întâlnesc în total 7.393 valori diferite de zero, prin urmare numărul adunări – scăderi este de 7.393, reprezentând doar 24,5 % față de algoritmul folosit în codorul FS 1016. De remarcat că numărul de înmulțiri – împărțiri se reduce la una pentru fiecare vector, reprezentând ridicarea la pătrat la numărătorul rel. (7.15) și că metoda nu are nici un efect negativ asupra calității semnalului, deoarece calcul se face exact, fiind afectată doar metoda de calcul.

### 7.4.3 Simplificarea calculului energiei

Termenul de la numitorul rel. (7.15) reprezintă energia semnalului extras din dicționar și filtrat. Deoarece găsirea unui maxim al raportului (7.15) înseamnă a găsi o pereche de forma numărător maxim – numitor minim, am analizat valorile numărătorului și numitorului acestui raport și am constatat că în cele mai multe cazuri valoarea numărătorului are un rol hotărâtor în determinarea valorii maxime a raportului. Această observație a condus la posibilitatea de a simplifica calculul energiei care intervine la numitorul raportului (7.15).

Am luat în considerare, ca modalitate de simplificare, o nouă formă pentru expresia (7.15), care a devenit astfel:

$$m'_i = \frac{(\mathbf{y}^{(i)T} \mathbf{e}^{(0)})^2}{\mathbf{y}_s^{(i)T} \mathbf{y}_s^{(i)}}, \quad (7.25)$$

unde vectorul  $\mathbf{y}_s^{(i)}$  și respectiv valoarea sa transpusă reprezintă un vector obținut din vectorul original  $\mathbf{y}^{(i)}$  prin subeșantionare.

Reducerea volumului de calcul pentru energie conform (7.25) este de  $s$  ori față de cazul general, dacă în (7.25) vectorul  $\mathbf{y}_s^{(i)}$  reprezintă varianta subeșantionată cu factorul întreg  $s$  a vectorului original  $\mathbf{y}^{(i)}$ .

Pentru a verifica concordanța între rezultatul obținut folosind (7.15) și respectiv (7.25) am efectuat un experiment în care am determinat mai întâi vectorul optim din dicționarul fix folosind criteriul exact (7.15) și apoi, folosind noul criteriu (7.25) pe care l-am propus, am determinat un număr de  $C$  vectori candidați, ordonați descrescător după valoarea  $m'_i$ .

Rezultate pe care le-am obținut în urma acestui experiment sunt prezentate în tabelul 7.5.

În tabel sunt indicate, în procente, numărul de cazuri în care vectorul optim determinat cu criteriul exact (7.15) se regăsește printre cei  $C$  vectori determinați cu criteriul propus, exprimat prin rel. (7.25).

Analizând rezultatele din tabelul 7.5, am constatat posibilitatea de a folosi criteriul simplificat propus pentru a selecta din dicționarul fix cu mărimea  $V$  un număr de  $C$  vectori



candidați, după care se face căutarea numai în cadrul acestei grupe de vectori candidați, folosind criteriul exact (7.15).

Tabelul 7.5. Coincidența între criteriul exact și criteriul simplificat [%]

Factorul de subeșantionare	Numărul de vectori candidați $C$				
	1	8	16	32	64
2	98,2	99,7	100	100	100
3	94,9	98,1	99,9	100	100
4	87,5	95,3	98,2	99,0	100
5	79,5	86,7	93,5	98,8	99,8
10	63,8	77,4	84,8	96,2	98,4
Fără energie	47,5	61,2	73,6	81,7	93,5

În tabelul 7.5 a fost considerată, în ultima linie, și situația în care criteriul de maxim a ținut seama doar de autocorelația de la numărător, fără a împărți această valoare la energia semnalului. Am constatat că în acest caz probabilitatea ca vectorul de cod optim să se afle între primii 64 de vectori selectați este de 93,5 %, ceea ce permite folosirea acestui criteriu de preselecție, urmat de un calcul conform criteriului (7.15) pentru cei 64 vectori selectați.

Pentru cazul în care nu se ține cont de energie la calculul criteriului de maxim am obținut o reducere semnificativă a volumului de calcul, conform tabelului 7.6.

Tabelul 7.6. Complexitatea de calcul pentru criteriul simplificat

Operație	Înmulțiri – împărțiri		Adunări – scăderi	
	Primul vector	Următorii 511 vectori	Primul vector	Următorii 511 vectori
Convoluție	1365	0	1305	0
Corelație	0	0	15	7393
Energie	–	–	–	–
Câștig	–	–	–	–
Scor	–	–	–	–
Total	1365	0	1320	7393
Total, % față de FS 1016	2,84 %		16,65 %	

Din datele din tabelul 7.6 rezultă o reducere de aproximativ 6 ori a numărului de operații de adunare – scădere și o reducere semnificativă, de aproximativ 35 de ori, a numărului de operații de înmulțire – împărțire, deosebit de avantajoasă pentru o implementare în timp real pe un procesor de semnal.

În figura 7.3 am reprezentat un fragment de semnal vocal, împreună cu semnalul codat folosind direct vectorul cu cel mai mare scor obținut conform rel. (7.25), fără a ține cont de energie în procesul de căutare (în fig. 7.3 a), cu interpolare cu 5 (fig. 7.3 b) și cu interpolare cu 2 (fig. 7.3 c).

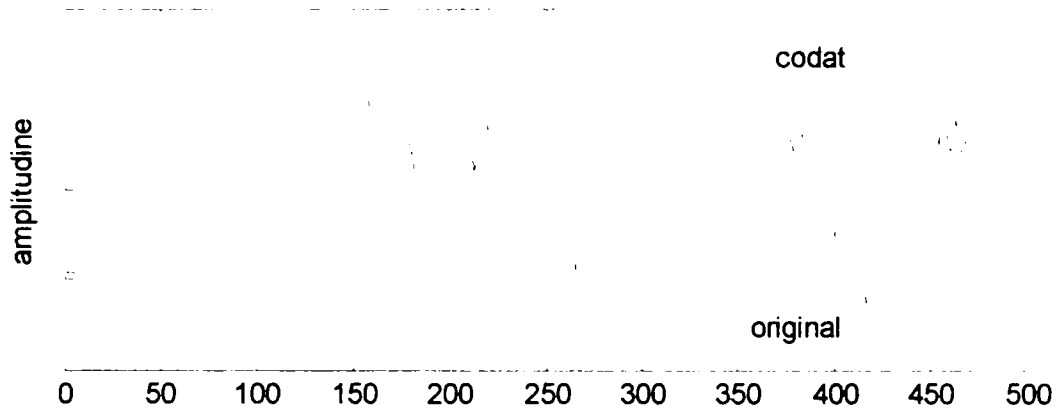


Fig. 7.3.a. Semnalul original și semnalul codat fără a ține cont de energie

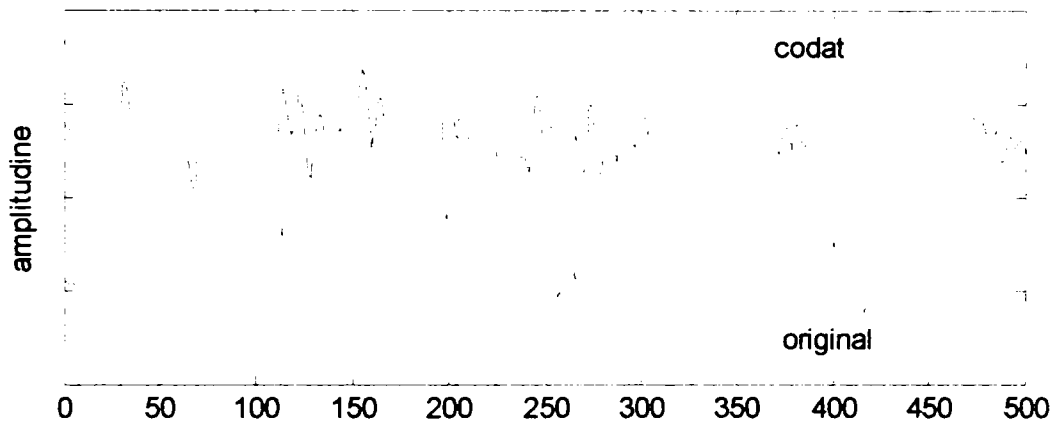


Fig. 7.3.b. Semnalul original și semnalul codat cu energia interpolată cu factorul 5

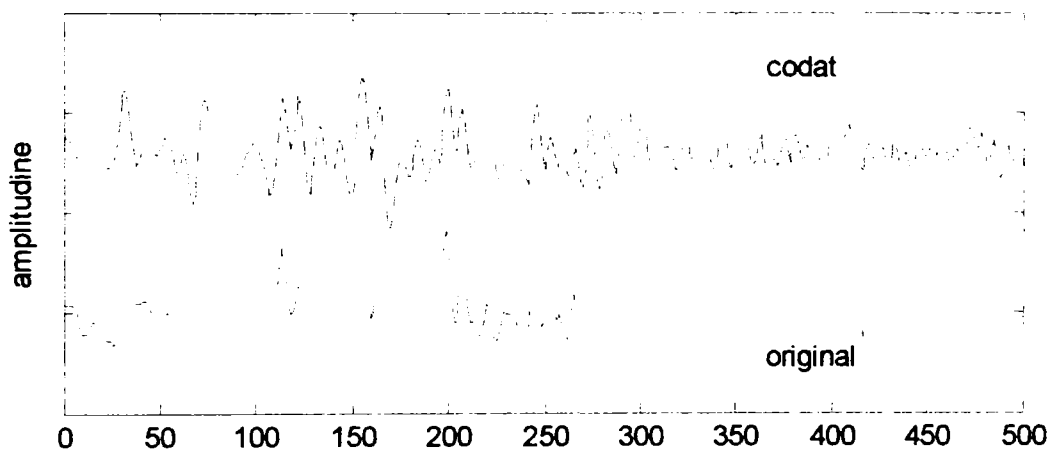


Fig. 7.3.c. Semnalul original și semnalul codat cu energia interpolată cu factorul 2

# *Capitolul 8*

## *Concluzii și contribuții*

### **8.1 Concluzii**

Subiectul abordat în lucrare este cel al cuantizării vectoriale a semnalului vocal. În acest scop s-a folosit ca structură de bază codorul CELP, identificându-se trei categorii de informații care se obțin prin analiza CELP și care urmează a fi cuantizate și transmise la decodator. Pentru fiecare dintre aceste categorii, în urma unei analize detaliate, au fost propuse diferite metode originale de calcul, de reducere a volumului de calcule, de cuantizare sau de reducere a erorilor.

Lucrarea cuprinde 156 de pagini și este structurată în șapte capitole, la care se adaugă un capitol de concluzii și contribuții precum și bibliografia.

Capitolul 1 este o introducere care prezintă sintetic principalele direcții în care a evoluat prelucrarea semnalului vocal. Este analizat mecanismul producerii vorbirii naturale și este introdus un model electric simplificat al producerii semnalului vocal, care stă la baza metodelor de codare a sursei de semnal.

Se prezintă pe scurt principalele caracteristici temporale, spectrale și statistice ale semnalului vocal, alături de caracteristicile sale lingvistice, având în vedere că semnalul vocal reprezintă materializarea comunicării prin vorbire.

Într-un scurt istoric se evidențiază că dacă codarea semnalului vocal în scopul transmiterii sau stocării este strâns legată de evoluția prelucrării semnalelor electrice prin metode analogice sau digitale și de evoluția circuitelor aferente, în special a microprocesoarelor și a procesoarelor de semnal, sinteza semnalului vocal a constituit o provocare pentru spiritul cercetător al omului încă din evul mediu, fiind menționate realizări de mașini vorbitoare mecanice încă din secolul XVIII.

Sunt prezentate criteriile de performanță ale sistemelor de codare a semnalului vocal și se evidențiază metodele subiective și obiective de apreciere a calității semnalului vocal codat. În finalul capitolului sunt trecute în revistă organizațiile care au elaborat standarde în domeniul codării semnalului vocal, pentru diferite aplicații, și sunt prezentate sintetic principalele standarde în domeniu.

Capitolul 2 prezintă metodele de codare a semnalului vocal. Este evidențiat faptul că majoritatea metodelor de codare a semnalului vocal sunt metode de codare cu pierderi. Este făcută o clasificare a acestor metode în metode care codează forma de undă și metode care codează sursa de semnal, la care se adaugă unele categorii de metode hibride, și se evidențiază performanțele generale ale acestor categorii de metode în funcție de debitul binar impus.

Sunt trecute în revistă metode de codare a formei de undă, între care: codarea scalară și vectorială, codarea în domeniul frecvență și codarea folosind analiza și sinteza sinusoidală. Din categoria metodelor de codare a sursei sunt prezentate vocoderele canal, formantic și homomorfic, precum și metodele de codare a semnalului vocal prin predicție liniară, în buclă deschisă și în buclă închisă.

Capitolul 3 prezintă din punct de vedere teoretic predicția liniară. Sunt introduse ecuațiile de predicție liniară și pe baza acestora este prezentat modelul de analiză și sinteză prin predicție liniară a unui semnal în general sau a semnalului vocal în particular. Sunt prezentate în detaliu relațiile de calcul a coeficienților de predicție prin metodele autocorelației și covarianței. Este prezentat de asemenea algoritmul Levinson – Durbin de rezolvare a ecuațiilor de predicție, și se introduc cu ajutorul acestui algoritm și coeficienții de reflexie, ca reprezentare alternativă a coeficienților de predicție. Se definește câștigul de predicție ca o măsură a calității predicției, măsură ce va fi folosită în capitolele 5, 6 și 7 pentru aprecierea comparativă a metodelor propuse. În acest capitol sunt prezentate în detaliu problemele legate de prelucrarea semnalului vocal pe cadre, ca modalitate practică de implementare a algoritmilor de codare, determinată de necesitatea de a considera semnalul vocal staționar pe durata unui cadru. În legătură cu prelucrarea semnalului pe cadre sunt prezentate modalitățile de asigurare a continuității prelucrării, considerând operațiile de filtrare realizate cu ajutorul a două filtre identice, unul cu starea inițială zero și unul cu intrarea zero. Ca modalitate alternativă de filtrare, este prezentată operația de filtrare ca produs de convoluție, particularizată pentru structura filtrelor folosite în codorul CELP, în scopul asigurării unei lungimi identice a cadrelor de semnal vocal de intrare și respectiv de ieșire.

Sunt prezentate în mod unitar și original problemele legate de aplicarea predicției liniare la prelucrarea semnalului vocal. Astfel, sunt descrise în detaliu operațiile de preaccentuare a frecvențelor înalte, expandarea benzii de trecere, corecția prin adăugarea unui zgomot alb și netezirea spectrală.

Sunt prezentate modalitățile alternative de reprezentare a coeficienților de predicție: coeficienții de reflexie, coeficienții LAR, coeficienții sinus invers și perechile de linii spectrale, împreună cu relațiile care le definesc și avantajele pe care le oferă privind cuantizarea.

În acest capitol este prezentat în detaliu codorul CELP, introdus în cap.2. Sunt descrise principiile de analiză prin sinteză, modelul CELP de producere a semnalului vocal și este prezentată structura de bază a codorului și respectiv a decodorului CELP, fiind definite cele trei categorii de informații care rezultă în urma analizei CELP, și care urmează a fi codate și transmise la decodor: parametrii de predicție pe termen scurt, parametrii de predicție pe termen lung și semnalul rezidual de excitație. Fiecare din aceste trei categorii de informații sunt analizate pe larg în cap. 5, cap. 6 și cap. 7, în care sunt introduse metode originale de calcul sau de simplificare a procedurilor de calcul în vederea reducerii volumului de operații necesare, precum și metode originale de cuantizare pentru acești parametri. În finalul acestui capitol se prezintă unele aspecte importante ale codării CELP care nu fac obiectul cap. 5, 6 și 7, respectiv ponderarea perceptuală a erorii de predicție și prelucrarea semnalului vocal după decodarea CELP: postfiltrarea pe termen scurt, postfiltrarea pe termen lung și controlul automat al câștigului.

Capitolul 4 prezintă cuantizarea vectorială. Sunt prezentate în detaliu principiile cuantizării vectoriale și proprietățile structurale ale cuantizorului vectorial. Sunt definite performanțele cuantizorului vectorial pe baza distorsiunii, definită ca o medie a distanțelor între vectorii de intrare și vectorul de cod, și sunt introduse câteva cazuri particulare pentru distorsiunea de cuantizare, ca distanța medie pătratică, distanța generalizată, eroarea pătratică ponderată, eroarea maximă, sau măsuri nesimetrice ale distanței. Este prezentat cuantizorul vectorial bazat pe regula celui mai apropiat vecin și un algoritm de cuantizare după această regulă.

Sunt analizate în detaliu condițiile de optimalitate pentru un cuantizor vectorial. Este definită în acest sens condiția de codor optim pentru un decodor dat, condiția de decodor optim pentru un codor dat, care introduce noțiunea de centroid, și condiția de margine cu

probabilitate zero. Sunt definite noțiunile de cuantizare vectorială local optimală și global optimală.

Se prezintă unele aspecte practice privind proiectarea unui cuantizor vectorial și este analizată în acest sens problema dicționarului inițial și este prezentat algoritmul Lloyd care conduce la obținerea unui cuantizor optim pornind de la dicționarul inițial. În acest context este analizat algoritmul Lloyd generalizat, sau algoritmul Linde – Buzo – Gray pentru date empirice, se arată convergența acestui algoritm spre un cuantizor optimal și se prezintă criteriile de oprire ale algoritmului. Este prezentat un algoritm LBG îmbunătățit, la care, ca o contribuție originală, căutarea unui optim se face pentru toate posibilitățile de deplasare a vectorului de cod, pe baza unui indice de utilitate.

Sunt trecute în revistă în acest capitol cele mai importante metode practice de cuantizare vectorială, după ce se arată care sunt problemele care pot să apară la implementarea unui cuantizor vectorial pentru o aplicație concretă. Metodele practice sunt prezentate într-o clasificare originală. Conform acestei clasificări, sunt prezentate metodele rapide de cuantizare nestructurată (metoda distanțelor parțiale, metoda proiecțiilor, metoda inegalității triunghiului și metoda arborelui  $K - d$ ) precum și metodele de cuantizare structurată (structura arborescentă, cuantizarea cu clasificare, cuantizarea prin transformare, cuantizarea de produs, cuantizarea în mai multe etaje și prin interpolare). Categoriile distincte în această clasificare o reprezintă cuantizarea predictivă și cuantizarea adaptivă și cuantizarea cu rată de bit variabilă. Metodele de cuantizare adaptivă sunt la rândul lor clasificate într-o formă originală, bazată pe noțiunea de dicționar universal, inductiv sau cunoscut apriori, fiind introdus un model al cuantizorului adaptiv bazat pe acest dicționar universale.

Capitolul 5 prezintă calculul și cuantizarea vectorială a coeficienților de predicție. Sunt definite perechile de linii spectrale LSP ca rădăcini ale unor polinoame  $P(z)$  și  $Q(z)$ , fiind deduse relațiile de recurență care permit calculul polinoamelor  $P(z)$  și  $Q(z)$  plecând de la coeficienții de predicție  $a_i$ , și sunt trecute în revistă diferite metode de rezolvare a ecuațiilor pentru obținerea LSP. Sunt prezentate câteva proprietăți utile ale LSP, care justifică alegerea acestor parametri pentru a reprezenta parametrii de predicție liniară.

Pentru cazul particular al predicției de ordinul 10 și rezolvarea ecuațiilor prin evaluarea polinoamelor în puncte succesive, este introdusă o metodă de calcul care elimină, din operația de evaluare, mai multe intervale în care am arătat că nu se pot afla rădăcini. Pentru aceasta, am demonstrat că în cazul general, pentru orice ordin  $M$  al predicției, în fiecare interval determinat de două rădăcini consecutive ale ecuației  $P(z) - Q(z) = 0$ , se află câte o singură rădăcină a polinomului  $P(z)$  și respectiv o singură rădăcină a polinomului  $Q(z)$ . Am demonstrat de asemenea că această ecuație are gradul mai mic cu o unitate față de polinoamele  $P(z)$  și  $Q(z)$  și că are întotdeauna rădăcini reale și distincte. În cazul predicției de ordinul 10, ecuația  $P(z) - Q(z) = 0$  este de gradul 4 și se poate rezolva simplu, având întotdeauna 4 rădăcini reale și distincte. Am implementat o metodă de rezolvare a acestei ecuații care realizează și ordonarea acestor rădăcini și determinarea celor 5 intervale în care sunt cuprinse valorile LSP, obținând o reducere a volumului de calcul cu aproximativ 32%.

Am analizat în detaliu modul de calcul al parametrilor LSP prin evaluarea celor două polinoame  $P(z)$  și  $Q(z)$  numai în valorile cuantizate ale LSP și erorile care apar în acest caz. Pentru reducerea acestor erori, am propus două metode de corecție. O primă metodă se referă la calculul unei singure valori LSP în mod independent, și pentru acest caz am propus o corecție bazată pe evaluarea polinomului la mijlocul intervalului de cuantizare. A doua metodă de corecție are în vedere calculul și cuantizarea a două valori LSP consecutive și pentru acest caz am propus de asemenea o metodă bazată pe evaluarea și compararea valorilor celor două polinoame la mijlocul intervalului de cuantizare.

În acest capitol este analizată și problema interpolării parametrilor de predicție, ca metodă de îmbunătățire a tranzițiilor între cadrele succesive. Am comparat experimental performanțele privind interpolarea pentru diferite reprezentări ale parametrilor de predicție, cu



concluzia că LSP prezintă cele mai bune proprietăți de interpolare, comparativ cu coeficienții de reflexie și parametrii LAR. Am enunțat și demonstrat o proprietate a LSP, conform căreia, prin interpolarea a două seturi de LSP care reprezintă filtre de sinteză stabile, se obține un nou set care asigură de asemenea stabilitatea filtrului de sinteză. Pentru îmbunătățirea interpolării am luat în considerare posibilitatea de a introduce un nou coeficient de interpolare, bazat pe energia cadrului și am introdus un exponent care să permită modificarea simplă a modului în care se ține seama de energie în operația de interpolare. Experimentările pe care le-am efectuat au demonstrat că ponderarea bazată pe energia cadrelor conduce la o scădere semnificativă a procentului cadrelor cu distorsiunea spectrală mare, îmbunătățind astfel cuantizarea în ansamblu.

Sunt trecute în revistă diferite metode de cuantizare scalară și vectorială a parametrilor LSP și sunt definite distorsiunea spectrală, ca măsură a performanței cuantizării, și cuantizarea transparentă, ca un standard de calitate a cuantizării parametrilor de predicție.

Pentru cuantizarea vectorială a LSP, am propus un model pentru studiul cuantizării parametrilor spectrali în general și am analizat două metode de cuantizare vectorială.

Prima metodă propusă are în vedere corelația semnificativă existentă între parametrii LSP din cadre succesive. Pentru a reduce propagarea erorilor și pentru a ține cont de faptul că există cadre succesive cu o corelație mai redusă, am propus un cuantizator vectorial care folosește o cuantizare SVQ (split VQ) pentru cadre pare, cuantizând astfel independent aceste cadre și o cuantizare predictivă pentru cadrele impare, folosind tot SVQ pentru a cuantiza reziduul de predicție. Am analizat două variante de descompunere a vectorului LSP, în doi subvectori cu 5 componente fiecare sau în trei subvectori, cu 3, 3, respectiv 4 componente fiecare.

A doua metodă de cuantizare vectorială propusă este cuantizarea cu clasificare (CVQ), folosind o singură variantă de descompunere a vectorului LSP în trei subvectori cu 3, 3 și 4 componente. Pentru această metodă am propus trei variante de clasificatori: un clasificator vectorial, un clasificator bazat pe media vectorilor și un clasificator bazat pe valoarea mediană.

Pentru fiecare din aceste metode au fost prezentate rezultatele obținute în ceea ce privește distorsiunea spectrală și rata de bit.

Capitolul 6 tratează problema modelării periodicității semnalului de excitație în codorul CELP. În acest capitol sunt explicate noțiunile de filtru LTP și de ordin LTP (pitch) și sunt prezentate modalitățile de calcul a ordinului LTP în buclă deschisă și în buclă închisă. Este de asemenea explicată în detaliu modelarea filtrării LTP printr-un dicționar adaptiv, modul de actualizare a acestui dicționar și problemele care apar în cazul în care ordinul LTP este mai mic decât lungimea subcadrului.

Pentru predictorul LTP cu un coeficient, am analizat posibilitățile de optimizare, calculând expresiile câștigului de predicție pentru varianta cu recircularea excitației prin filtrul LTP, pentru diferite cazuri ale ordinului LTP. Am analizat de asemenea filtrul LTP de ordin fracționar, ca modalitate de creștere a rezoluției în timp a filtrului cu un coeficient.

Pentru cazul general al unui predictor LTP cu  $2q + 1$  coeficienți, am dedus relațiile de calcul pentru parametrii de predicție LTP. Pe baza acestor relații, am dedus apoi relațiile de calcul pentru parametrii filtrului LTP cu trei coeficienți.

În vederea reducerii numărului de parametri care se cuantizează și se transmit, am propus două variante de filtru LTP cu trei coeficienți cu constrângeri, impunând restricții asupra coeficienților filtrului, care astfel nu mai sunt independenți.

Pentru cele două variante de filtru LTP cu constrângeri, am analizat stabilitatea filtrului LTP, constatând că pentru cele două variante propuse, criteriul de stabilitate se reduce la relații algebrice simple între coeficienții filtrului. Pentru a asigura stabilitatea am propus un criteriu simplu, bazat pe un coeficient de stabilitate ușor de calculat și pe scalarea coeficienților filtrului cu acest coeficient, în cazul în care filtrul nu este stabil.



Fiecare din cele două variante de filtru a fost analizat comparativ cu filtrul cu un coeficient, interpolat și neinterpolat. Pentru filtrele propuse am analizat și posibilitatea folosirii ordinului LTP fracționar, prin interpolare. Rezultatele au fost comparate din punct de vedere al câștigului de pedicție, al raportului semnal – zgomot și al raportului semnal – zgomot segmental, validând variantele de filtre propuse.

Capitolul 7 prezintă cuantizarea vectorială a semnalului rezidual în codorul CELP. Este explicată modelarea semnalului rezidual cu un dicționar fix și este prezentată în detaliu modalitatea de căutare a celui mai bun vector de excitație din acest dicționar. Având în vedere că această operație este cea mai costisitoare, din punct de vedere al volumului de calcul, în codorul CELP, este analizată în detaliu complexitatea căutării în dicționarul fix și sunt deduse relațiile care exprimă complexitatea căutării, exprimată prin numărul de operații de tip adunare – scădere, respectiv înmulțire – împărțire, pentru cazul general al unui dicționar fix organizat sub formă de matrice, cât și pentru cazul dicționarului cu vectori suprapuși și sunt trecute în revistă metodele cunoscute pentru reducerea complexității operațiilor de căutare.

Pentru codorul CELP FS 1016, considerat ca referință pentru aprecierea complexității metodelor propuse, este calculat exact numărul de operații necesare, ținând cont de structura exactă a dicționarului.

Luând în considerare un dicționar fix cu vectori suprapuși, cu structură ternară, similar celui folosit în codorul FS 1016, sunt propuse trei metode originale pentru reducerea complexității de căutare.

Prima metodă propune trunchierea lungimii răspunsului la impuls al filtrului de sinteză și analizează efectul acestei trunchieri asupra raportului semnal – zgomot.

A doua metodă propusă este bazată pe o nouă formă de exprimare a vectorului  $y$  care intervine în calculul unui produs de corelație, produs care, la rândul său, intervine în expresia scorului care trebuie maximizat în vederea alegerii celui mai bun vector de excitație. În acest sens, vectorul  $y$  este exprimat ca produs dintre un vector (care se calculează o singură dată pentru un cadru) și vectorul extras din dicționarul fix. În acest mod, produsul se reduce la testarea fiecărei valori din dicționarul fix și efectuarea unei adunări a unei scăderi, sau nici o operație, după cum valoarea este  $+1$ ,  $0$  sau  $-1$ . Această metodă asigură o reducere a volumului de calcul pentru autocorelație cu 75.5% față de metoda clasică, fără nici un efect negativ asupra calității semnalului.

A treia metodă propusă are în vedere simplificarea calcului termenului care reprezintă energia în expresia scorului care trebuie maximizat. Pentru aceasta am propus calculul unei versiuni simplificate a termenului energie, pe baza unor vectori subeșantionați cu un factor  $s$ , și am folosit această valoare pentru a calcula un număr de vectori candidați din dicționarul fix, în ordinea descrescătoare a scorului. Dintre aceștia se poate apoi alege cel mai bun vector folosind criteriul de căutare nemodificat. Pentru valori ale factorului de subeșantionare de până la 10 și a numărului de vectori candidați de până la 64, metoda propusă asigură găsirea celui mai bun vector de excitație din dicționarul fix în peste 98 % din cazuri.

## 8.2 Contribuții originale

- prezentarea sintetică a metodelor cunoscute pentru codarea semnalului vocal (cap. 3),
- prezentarea sintetică a metodelor de cuantizare vectorială, cu o propunere de clasificare generală a acestor metode (§4.8),
- clasificarea metodelor de cuantizare adaptivă pe baza noțiunii de dicționar universal și introducerea unui model pentru cuantizorul adaptiv (§4.8.4),
- propunerea unei metode de îmbunătățire a algoritmului LBG de construcție a unui cuantizor vectorial, bazată pe un factor de utilitate al fiecărei celule și căutarea celei mai bune deplasări a vectorului de cod în toate celulele (§4.7.5),

- studiul unor aspecte privind implementarea practică a predicției liniare în general și a codorului CELP în particular asupra semnalului vocal, bazată pe prelucrare semnalului pe cadre (§3.2.5, §3.4.2, §3.4.3),
- studiul reprezentării alternative a coeficienților de predicție sub forma perechilor de linii spectrale LSP, ca rădăcini ale unor polinoame (§5.1.1, §5.1.2),
- enunțarea și demonstrarea unei proprietăți a acestor polinoame, prin care se pot încadra perechile de linii spectrale în intervale deschise, simplificând calculul LSP prin evaluare directă a celor două polinoame (§5.2.1),
- propunerea unei metode de reducere a erorilor în cazul calculului LSP prin evaluare directă, în mod individual (§5.2.1),
- propunerea unei metode de reducere a erorilor în cazul calculului LSP în perechi, metodă care asigură păstrarea proprietății de ordine a LSP și implicit stabilitatea filtrului de sinteză (§5.2.1),
- studiul posibilităților de interpolare a diferitelor reprezentări ale parametrilor de predicție (§5.3),
- enunțarea și demonstrarea proprietății de stabilitate a filtrului de sinteză realizat pe baza valorilor interpolate ale LSP (§5.3),
- propunerea unei metode de interpolare a LSP bazată pe energia cadrului de semnal, cu posibilitatea de a controla contribuția energiei în coeficientul de ponderare, cu efecte pozitive asupra procentului cadrelor cu distorsiuni spectrale mari (§5.3),
- propunerea unei metode de cuantizare vectorială predictivă a LSP, diferită pentru cadrele pare și impare, care ține seama de corelația importantă existentă între valorile LSP din cadre consecutive (§5.4.1),
- propunerea unei metode de cuantizare vectorială cu clasificare a LSP, folosind trei clasificatori diferiți, cu avantajul unei reduceri a volumului de calcul (§5.4.2),
- analiza modelării periodicității semnalului rezidual în codorul CELP (§6.1),
- deducerea relațiilor de calcul a parametrilor filtrului LTP, pentru cazul genral, și particularizarea lor pentru filtrul cu trei parametri (§6.2, §6.5.1),
- propunerea a două modele de filtre LTP cu trei parametri, cu restricții asupra parametrilor de câștig, în scopul reducerii numărului de parametri ce trebuie transmiși la decodor (§6.5.2, §6.5.3),
- studiul stabilității filtrului LTP (§6.5.4),
- propunerea unei metode simple de stabilizare a filtrului LTP (§6.5.5),
- analiza modelării excitației reziduale în codorul CELP prin dicționarul fix și analiza operației de căutare a celui mai bun semnal de excitație în dicționarul fix (§7.2),
- analiza detaliată a complexității operațiilor de căutare în dicționarul fix, pentru identificarea unor posibilități de reducere (§7.3),
- propunerea unei metode de reducere a complexității de căutare bazată pe trunchierea răspunsului la impuls al filtrului de sinteză (§7.4.1),
- propunerea unei metode eficiente de calcul a funcției de autocorelație pentru dicționarul fix cu vectori suprapuși (§7.4.2),
- propunerea unui criteriu simplificat de căutare a unei grupe de candidați cu cei mai buni vectori de excitație, bazat pe folosirea unei versiuni subeșantionate a vectorului pentru calculul energiei (§7.4.3).

Pentru verificarea practică a metodelor propuse a fost creată o bază de date de semnal vocal, descrisă în cap. 5, și a fost implementat un set de subrutine în mediul Matlab, care permit:

- a) codarea CELP a unui semnal vocal, urmată de decodare, :
  - selectarea și citirea unui fișier de intrare de tip \*.wav, care conține semnal audio mono, codat PCM pe 8 sau 16 biți, cu frecvența de eșantionare implicită 8 kHz,

- realizarea unei codări CELP, urmată de decodare.
  - scrierea semnalului rezultat prin decodare, în același format ca și semnalul de intrare, într-un fișier de ieșire de același tip, cu numele ales de utilizator,
  - stabilirea numărului de coeficienți de predicție,
  - stabilirea lungimii cadrului de analiză,
  - stabilirea tipului de fereastră de ponderare a cadrului de semnal,
  - stabilirea coeficientului  $\gamma$  de ponderare perceptuală a erorii,
  - stabilirea lungimii minime și maxime pentru dicționarul adaptiv,
  - utilizarea unui dicționar fix definit de utilizator, sau a unui dicționar aleator generat local, cu număr de vectori la alegerea utilizatorului,
  - salvarea mărimilor de ieșire ale codorului într-un fișier în format \*.mat,
  - posibilitatea de a folosi un ordin de predicție LTP fracționar, obținut prin interpolare până la ordinul 4,
  - posibilitatea de a reprezenta grafic semnalul de intrare, de ieșire sau diferite mărimi care intervin în codarea semnalului de intrare.
- b) cuantizarea vectorială a parametrilor LPC:
- citirea parametrilor dintr-un fișier în format \*.mat,
  - obținerea unui dicționar prin algoritmul LBG modificat, descris în §4.7.5, pe baza unui set de parametri de antrenare,
  - cuantizarea parametrilor LPC folosind dicționarul obținut,
  - scrierea rezultatului cuantizării într-un fișier de tip \*.mat,
  - folosirea valorilor cuantizate pentru decodorul CELP.

# Bibliografie

- [ALS84] H. Abut S. A. Luse, "Vector quantizers for subband coded waveforms", *International Conference on Acoustics, Speech, and Signal Processing*, San Diego, Calif., March 1984.
- [AMS93] M. E. Ahmet, M. I. Suwaiyel, "Fast methods for Code Search in CELP", *IEEE Transactions on Speech and Audio Processing*, vol. 1, nr. 3, July 1993.
- [ARM93] S. Arya, D. M. Mount, "Algorithms for fast vector quantization", *Proc. Data Compression Conference*, J. A. Storer, M. Cohn (editors), *IEEE Computer Society Press*, 1993.
- [ASS99] A. Ahmadi, S. Spanias, "Cepstrum-based pitch detection using a new statistical V/UV classification algorithm", *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 3, May 1999.
- [ATA82] B. S. Atal, "Predictive coding of speech at low bit rates." *IEEE Trans. Commun.*, vol. COM-30, Apr. 1982.
- [ATH71] B. S. Atal, S. L. Hanauer, "Speech analysis and synthesis by linear prediction of the speech wave." *J. Acoustical Society of America*, vol. 50, Aug. 1971.
- [BAK03] T. Bäckström, P. Alku, "All-Pole Modeling Technique Based on Weighted Sum of LSP Polynomials", *IEEE Signal Processing Letters*, vol. 10, no. 6, June 2003.
- [BBM93] W. P. LeBlanc, B. Bhattacharya, S. A. Mahmoud, V. Cuperman, "Efficient search and design procedures for robust multi-stage VQ of LPC parameters for 4 kb/s speech coding", *IEEE Transactions on Speech and Audio Processing*, vol. 1, nr. 4, oct. 1993.
- [BGG80] A. Buzo, A. H. Gray Jr., R. M. Gray, J. D. Markel, "Speech coding based upon vector quantization", *IEEE Transaction on Acoustics, Speech and Signal Processing*, ASSP-28, October 1980
- [BHT97] J. M. Buhmann, T. Hofmann, "Robust vector quantization by competitive learning", *Proc. of International Conference on Acoustics, Speech and Signal Processing*, ICASSP, Munich, 1997.
- [BME99] M. E. Borda, *Teoria transmiției informației – teoria informației și codării. Fundamente și aplicații*. Ed. Dacia, Cluj-Napoca, 1999.
- [BRG71] T. Berger, *Rate Distortion Theory*, Englewood Cliffs, N.J., Prentice Hall, 1971.
- [BTS02] B. Bessette and all., "The adaptive Multirate Wideband Speech Codec (AMR – WB)", *IEEE Transactions on Speech and Audio Processing*, vol.10, no. 8, ,nov. 2002.
- [CBA00] C. Balint, "CELP coding for wideband speech signal", *Proceedings of the Symposium of Electronics and telecommunications ETc-2000*, Timișoara, nov. 2000.
- [CBA03-1] C. Balint, "MATLAB Software for Speech Analysis and Coding", *Analele Universității Tibiscus Timișoara, seria Informatică*, vol. I, Fasc. I, 2003.
- [CBA03-2] C. Balint, "Variable Low Bit Rate CELP Speech coder", *Analele Universității Tibiscus Timișoara, seria Informatică*, vol. I, Fasc. I, 2003.
- [CBA03-3] C. Balint, M. Naforniță, "Cuantizarea vectorială. Partea I-a. Prezentarea cuantizării vectoriale", *Analele Universității Tibiscus Timișoara, seria Informatică*, vol. I, Fasc. II, 2003.
- [CBA03-4] C. Balint, M. Naforniță, "Cuantizarea vectorială. Partea a II-a. Cuantizarea vectorială adaptivă", *Analele Universității Tibiscus Timișoara, seria Informatică*, vol. I, Fasc. II, 2003.
- [CBA03-5] C. Balint, "Objective Quality Measurement for Compressed Audio Signals", *Simpozionul Internațional Competiție și integrare. Noua Europa*, Universitatea Tibiscus și Fundatia Truvist Olanda, Timișoara, oct. 2003.

- [CBA04-1] C. Balint, "Design and training of codebook for CELP coder", *Analele Universității Tibiscus Timișoara, seria Informatică*, vol. II, Fasc. I, 2004.
- [CBA04-2] C. Balint, "An improved LBG algorithm for vector quantization", *Analele Universității Tibiscus Timișoara, seria Informatică*, vol. II, Fasc. I, 2004.
- [CBA04-3] C. Balint, "Efficient vector quantization of speech spectral parameters", *Symposium of Electronics and telecommunications Etc 2004*, Timișoara, oct. 2004.
- [CBA04-4] C. Balint, "Excitation modelling in CELP speech coders", *Symposium of Electronics and telecommunications Etc 2004*, Timișoara, oct. 2004.
- [CBA04-5] C. Balint, "Fast vector Quantization of Line Spectral Frequency", *Analele Universitatii din Oradea, Fascicola Electrotehnica, Sectiunea Electronica*, 2004.
- [CBA05] C. Balint, "Efficient LSP computation and quantization", *International Symposium on Signals, Circuits & Systems*, Iași, 2005 (acceptată pentru publicare).
- [CCI88] CCITT Recommendation G.721, "32kb/s Adaptive Differential Pulse Code Modulation (ADPCM)." Blue Book, Vol. III, Fascicle III.3, Oct. 1988.
- [CCL92] J-H. Chen, R. V. Cox, Y-C. Lin, N. Jayant, M. J. Melchner, "A low-delay CELP coder for the CCITT 16 kb/s speech coding standard", *IEEE Journal on Selected Areas in Communications*, vol. 10, no. 5, june 1992.
- [CCT86] J. P. Campbell and T. E. Tremain, Voiced/unvoiced classification of speech with applications to the U.S. government LCP-10E algorithm, *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, 1986.
- [CGR86] P. C. Chang, R. M. Gray, "Gradient algorithms for designing predictive vector quantizers", *IEEE Transaction on Acoustics, Speech and Signal Processing*, ASSP-34, Aug. 1986.
- [CGT90] J. M. Combes, A. Grossmann, Ph. Tchamitchian, editors, *Wavelets*, Springer-Verlag, Berlin, 1990.
- [CHE95] J. H. Chen, "Low delay coding of speech", în *Speech Coding and Synthesis*, W. B. Klein, K. K. Paliwal, eds., Elsevier Science, 1995.
- [CHG94] W. Y. Chan, A. Gersho, "Generalized product code vector quantization: A family of efficient techniques for signal compression", *Digital Signal Process.*, vol. 4, Apr. 1994.
- [CHO95] J. Choi, "A Fast Determination of Stochastic Excitation without Codebook Search in CELP Coder", *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 6, nov. 1995.
- [CHU03-1] W. C. Chu, "Window optimization in linear prediction analysis", *IEEE Transactions on Speech and Audio Processing*, vol. 11, nr. 6, nov. 2003.
- [CHU03-2] W. C. Chu, *Speech coding algoritms. Foundation and evolution of standardized coders*, J. Wiley & Sons, N. J., 2003.
- [CLG89] P. A. Chou, T. Lookabaugh, R. M. Gray, "Entropy-constrained vector quantization", *IEEE Transaction on Acoustics, Speech and Signal Processing*, ASSP, vol 37(1), 1989.
- [CLR91] T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, *Introduction to algorithms*, MIT Press, McGraw-Hill, NY, 2001.
- [CLS01] L. Chunyani, P. Lupini, E. Shlomot, V. Cuperman, "Coding of variable dimension speech spectral vectors using weighted nonsquare transform vector quantization", *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 6, Sept. 2001.
- [COS93] Y. M. Cheng, D. O'Shaughnessy, "On 450-600 b/s natural sounding speech coding", *IEEE Transactions on Speech and Audio Processing*, vol. 1, nr. 2, apr. 1993.
- [CRL92] D. Cohn, E. Riskin, R. Ladner, "Theory and Practice of Vector Quantizers Trained on Small training Sets", *Tehcnical Report TR 92-12-08*, University of Washington, 1992
- [CSK99] W. Chung, S. Kang, "Design of a variable rate algorithm for the CS-ACELP coder" *IEICE Trans. Inf. & Syst.*, Vol. E82-D, No. 10, October 1999.
- [CSL82-1] J. H. Conway, N. J. A. Sloane, "Fast quantizing and decoding algorithms for lattice quantizers and codes", *IEEE Trans. Inform. Theory*, IT-28, March 1982.



- [CSL82-2] J. H. Conway, N. J. A. Sloane, "Voronoi regions of lattices, second moments of polytopes, and quantization", *IEEE Trans. Inform. Theory*, IT-28, March 1982.
- [CSQ95] C. Chinrungrueng, C. Sequin, "Optimal Adaptive k-means Algorithm with Dynamic Adjustment of Learning Rate", *IEEE Trans. on Neural Networks*, vol. 6 (1), 1995.
- [CTT94]. J. P. Campbell Jr., V. C. Welch, and T. E. Tremain, An expandable error-protected 4800 BPS CELP coder (U. S. federal standard 4800 BPS voice coder), *Proc. ICASSP*, 1989, 735-738.
- [CTV91] J. Campbell, T. Tremain, V. Welch, "The DoD 4.8 kbps Standard (Proposed Federal Standard 1016)", in *Advances in Speech Coding*, Kluwer Academic Publishers, Boston, 1991.
- [DPH93] J. R. Deller Jr., J. G. Proakis, J. H. L. Hansen, *Discrete-Time Processing of Speech Signal*, Macmillan, 1993.
- [DRG96] A. Das, A. V. Rao, A. Gersho, "Variable-Dimension Vector Quantization", *IEEE Signal Processing Letters*, vol. 3, no. 7, July 1996.
- [EBR97] J. S. Erkelens, P. M. T. Broersen, "Bias Propagation in the Autocorrelation Method of Linear Prediction", *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 2, March 1997.
- [ECT94] E. Erzin, A. E. Cetin, "Interframe Differential Coding of Line Spectrum Frequencies", *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 2, April 1994.
- [ESJ99] T. Eriksson, J. Linden, J. Skoglund, "Interframe LSF quantization for noisy channels", *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 5, Sept. 1999.
- [EPB95] J. S. Erkelens, P. M. T. Broersen, "On the statistical properties of line spectrum pairs", *Proc. ICASSP, Detroit*, 1995.
- [EPB98] J. Erkelens, P. Broersen, "LPC interpolation by approximation of the sample Autocorrelation function," *IEEE Transactions on Speech and Audio Processing*, vol. 6, Nov. 1998.
- [EQZ89] W. H. Equitz, "A new vector quantization clustering algorithm", *IEEE Transactions on Acoustic, Speech and Signal Processing*, October 1989.
- [ETS97-1] European Technical Standard (ETS 300 960), GSM – Digital Cellular Telecommunications (Phase 2+). Full Rate Speech: Processing Functions (GSM 06.01), 1997.
- [ETS97-2] European Technical Standard (ETS 300 961), GSM – Digital Cellular Telecommunications (Phase 2+). Full Rate Speech: Transcoding (GSM 06.10), 1997.
- [FAH97] J. E. Fowler, S. C. Ahalt, "Adaptive Vector Quantization Using Generalized Threshold Replenishment", *Proceedings of the IEEE Data Compression Conference*, IEEE Computer Society Press, March 1997.
- [FBF95] M. A. Ferrer – Ballester, A. R. Figueiras – Vidal, "Efficient Adaptive Vector Quantization of LPC Parameters", *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 4, July 1995.
- [FLA72] J. Flanagan, *Speech Analysis, Synthesis and Perception*, Springer Verlag, Berlin, 1972.
- [FMS92] S. Furui, M. M. Sondhi (editors), *Advances in Speech Signal Processing*, Marcel Dekker Inc., New York 1992.
- [FTZ97] B. Fritzke, "The LBG-U Method for vector Quantization – an Improvement over LBG inspired from Neural Network", *Neural Processing Letters*, vol. 5, no. 1, 1997.
- [GDM76] A. H. Gray, Jr., J. D. Markel, "Distance Measures for Speech processing", *IEEE Trans. On Acoustics, Speech and Signal Processing*, vol. 24, Oct. 1976.
- [GER93] A. Gersho, R. M. Gray, *Vector quantization and Signal Compression*, Kluwer Academic Publishers, London, 1993.
- [GKM99] E. Gunduzhan, K. Momtahan, "Linear prediction based packet loss concealment algorithm for PCM coded speech", *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 8, Nov. 1999.
- [GKN00] W. J. Goralski, M. C. Kolon, *IP Telephony*, McGraw – Hill, 2000.



- [GMJ91]. I. Gerson and M. Jasiuk, Vector sum excited linear prediction (VSELP), in B. S. Atal, V. S. Cuperman, and A. Gersho, eds., *Advances in Speech Coding*, Kluwer, Dordrecht, 1991.
- [GMY85] A. Gersho, M. Yano, "Adaptive Vector Quantization by Progressive Codevector Replacement". *Proceedings of the International Conference of Acoustics, Speech and Signal Processing*, May 1985.
- [GNF98] R. M. Gray, D. I. Neuhoff, "Quantization". *IEEE Transaction on Information Theory*, vol. 44(6), oct. 1988.
- [GPK91] J. Grass, P. Kabal, "Methods of improving vector-scalar quantization of LPC coefficients". *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Toronto, May 1991.
- [GRA95] W.R. Gardner, D. Rao, "Theoretical Analysis of the High-Rate Vector Quantization of LPC Parameters". *IEEE Transactions on Speech and Audio Processing*, vol. 3 .no. 5, september 1995.
- [GSM97]. Special Mobile Group (GSM), Digital Cellular Telecommunications System: Enhanced Full Rate (EFR) Speech Transcoding, Technical Report GSM 06.60, *European Telecommunication Standard Institute (ETSI)*, 1997.
- [GSM98]. Special Mobile Group (GSM), Digital Cellular Telecommunications System (Phase 2+): Adaptive Multi-rate (AMR) Speech Transcoding, Technical Report GSM 06.60, *European Telecommunication Standard Institute (ETSI)*, 1998.
- [HAG96] R. Hagen, "Robust LPC Spectrum Quantization – Vector Quantization by a Linear Mapping of a Block code". *IEEE Transactions on Speech and Audio Processing*, vol. 4, no. 4, july 1996.
- [HAN99]. M. Handley et al., SIP: Session Initiation Protocol, IETF RFC, March 1999, <http://www.cs.columbia.edu/sip/sip.html>.
- [HCH02] Y. – H. Huang, T. – D. Chiueh, "A New Audio Coding Scheme Using a Forward Masking Model and Perceptually Weighted Vector Quantization". *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, july 2002.
- [HEL89]. K. Hellwig et al., Speech codec for the European mobile radio system. *Proc. IEEE Global Telecomm. Conf.*, 1989.
- [HPG99] R. Hagen, E. Paksoy, A. Gersho, "Voicing-specific LPC quantization for variable-rate speech coding". *IEEE transactions on Speech and Audio Processing*, vol. 7, no. 5, Sept. 1999.
- [HPL98] J. H. L. Hansen, B. L. Pellom, "An effective quality evaluation protocol for speech enhancement algorithms", <http://www.ee.duke.edu/Research/Speech>.
- [HRL01] A. Harma, U. K. Laine, "A comparison of warped and conventional linear predictive coding" *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 5, July 2001.
- [HRM01] A. Harma, "Linear predictive coding with modified filter structures", *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 8, Nov. 2001.
- [HRS01] M. Hans, R. W. Schafer, "Lossless Compression of Digital Audio." *IEEE Signal Processing Magazine*, july 2001.
- [HSK00] P. Hedelin, J. Skoglund, "Vector Quantization Based on Gaussian Mixture Models". *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 4, july 2000.
- [ISO98-1] ISO/IEC, Information Technology – Coding of Audiovisual Objects, Part3: Audio, Subpart 1: Overview, Technical Report ISO/JTC 1/SC 29/N2203, ISO/IEC, 1998.
- [ISO98-2] ISO/IEC, Information Technology – Coding of Audiovisual Objects, Part3: Audio, Subpart 3: CELP, Technical Report ISO/JTC 1/SC 29/N2203CELP, ISO/IEC, 1998.
- [ISO98-3] ISO/IEC, Information Technology – Coding of Audiovisual Objects, Part3: Audio, Subpart 4: Time/Frequency Coding, Technical Report ISO/JTC 1/SC 29/N2203TF, ISO/IEC, 1998.
- [ISO98-4] ISO/IEC, Information Technology – Very Low Bitrate Audio-Visual Coding, Part 3: Audio, Subpart 2: Parametric Coding, Technical Report ISO/JTC 1/SC 29/N2203PAR, ISO/IEC, 1998.

- [ITA75] F. Itakura, "Line Spectrum Representation of Linear Predictive Coefficients of Speech Signals", *Journal of the Acoustic Society of America*, vol. 57, 1975.
- [ITU90] ITU-T, 5-, 4-, 3- and 2-bits per Sample Embedded Adaptive Differential Pulse Code Modulation (ADPCM), Technical Report G.727, International Telecommunications Union, Geneva, 1990.
- [ITU92] ITU-T Recommendation G.728, "Coding of Speech at 16 Kbit/s using Low-Delay Code Excited Linear Prediction", Blue Book, Vol. III, Fascicle III.4, Geneva, 1992.
- [ITU93] ITU-T Recommendation G.722 7 kHz audio – coding within 64 kbit/s, Blue Book, Vol. III, Fascicle III.4, Geneva, 1993.
- [ITU96-1] ITU-T Recommendation G.729. "Coding of speech at 8 kbps using conjugate-structure algebraic code excited linear prediction (CS-ACELP)", ITU-T Recommendation, Geneva, 1996.
- [ITU96-2] ITU-T, Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 kbit/s, Technical Report G.723.1, International Telecommunication Union, Geneva, 1996.
- [ITU98-1] ITU-T, Objective Quality Measurement of Telephone-Band (300-3400 Hz) speech codecs, Technical Report P.861, International Telecommunications Union, Geneva, 1998.
- [ITU98-2] ITU-T, Packet Based Multimedia Communications Systems, Technical Report H.323, International Telecommunications Union, Geneva, 1998.
- [JAL03] M. Hasegawa-Johnson, A. Alwan, "Speech coding: Fundamentals and Applications", in *Wiley Encyclopedia of Telecommunications*, J. G. Proakis (editor), John Wiley & Sons, 2003.
- [JAR97] K. Jarvinen et al., GSM enhanced full rate speech codec, *Proc. IEEE Int. Conf. on Acoustics, Speech, Signal Processing*, 1997.
- [JJH98] J. Johnston, "Estimation of perceptual entropy using noise masking criteria", *Proc. IEEE Int. Conf. on Acoustics, Speech, Signal Processing*, (New York), 1988.
- [JLP99] Jialong He, Li Liu, G. Palm, "A discriminative training algorithm for VQ-based speaker identification", *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 3, May 1999.
- [JMK75] J. Makhoul, "Linear prediction: A tutorial review", *Proc. IEEE*, vol. 63, Apr. 1975.
- [JRW99] J. Rothweiler, "On polynomial reduction in the computation of LSP frequencies", *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 5, September 1999.
- [JSH01] Ph. J. B. Jackson, Ch. H. Shadle, "Pitch – Scaled Estimation of Simultaneous Voiced and Turbulence – Noise Components in Speech", *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 7, October 2001.
- [JWG96] F.-R. Jean, H.-Ch. Wang, "Transparent Quantization of Speech LSP Parameters Based on KLT and 2-D-Prediction", *IEEE Transactions on Speech and Audio Processing*, vol. 4, no. 1, Jan. 1996.
- [KAR86] P. Kabal, R. P. Ramachandran, "The Computation of Line Spectral Frequencies Using Chebyshev Polynomials", *IEEE Transactions on Acoustic, Speech and Signal Processing*, vol. ASSP-34, no. 6, December 1986.
- [KAT04] V. Krishnan, D. V. Anderson, K. K. Truong, "Optimal multistage vector quantization of LPC parameters over noisy channels", *IEEE Transactions on Speech and Audio Processing*, vol. 12, nr. 1, Jan. 2004.
- [KBA03] W. B. Kleijn, T. Bäckström, P. Alku, "On line spectral frequencies", *IEEE Signal Processing Letters*, vol. 10, no. 3, March 2003.
- [KDS86] P. Kroon, E. F. Deprettere, R. J. Sluyter, "Regular-pulse excitation: A novel approach to effective and efficient multi-pulse coding of speech" *IEEE Transactions on Acoustic, Speech and Signal Processing* ASSP 34, 1986.
- [KHL00] H. K. Kim, S. Ho Choi; H. S. Lee, "On approximating line spectral frequencies to LPC cepstral coefficients", *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 2, March 2000.

- [KKK90] W. B. Kleijn, D. J. Krasinski, R. H. Ketchum, "Fast methods for the CELP speech coding algorithm", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, nr. 8, aug. 1990.
- [KLE99] H. K. Kim, H. S. Lee, "Interlacing properties of line spectrum pair frequencies", *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 1, January 1999.
- [KLO00] K. Lehgwehasaitit, A. Ortega, „Rate-complexity-distortion for quadtree-based DCT coding”, *Proc. IEEE Int. Conference on Image Processing*, 2000.
- [KLO99] K. Lehgwehasaitit, A. Ortega, „Complexity-distortion tradeoffs in vector matching based of probabilistic partial distance techniques”, *Proc. IEEE International Data Compression Conference, IEEE Computer Society*, 1999.
- [KNT87] R.Boite, M.Kunt, *Traitement de la parole*, Presses Polytechniques Ronades, Lausanne, 1987.
- [KRK94] W. B. Kleijn, R. P. Ramachandran, P. Kroon, „Interpolation of the pitch-predictor parameters in analysis-by-synthesis speech coders”, *IEEE Transactions on Speech and Audio Processing*, vol. 2, nr. 1, jan. 1994.
- [KRO03] H. Khalil, K. Rose, "Predictive Vector Quantizer Design Using Deterministic Annealing", *IEEE Transactions on Signal Processing*, vol. 51, no. 1, january 2003.
- [KSL00] H. K. Kim, S. H. Choi, H. S. Lee, "On Approximating Line Spectral Frequencies to LPC Cepstral Coefficients", *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 2, march 2000.
- [KWG01] R. Kumaresan, Y. Wang, "On the duality between line-spectral frequencies and zero-crossings of signals", *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 4, May 2001.
- [LAK00] F. Lahouti, A. K. Khandani, "Quantization of line spectral parameters using a Trellis structure", *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2000*.
- [LBG80] Y. Linde, A. Buzo, R. M. Gray, "An algorithm for vector quantizer design", *IEEE Trans. Comm.*, COM-28, January 1980.
- [LCH94] K. – W. Law, Ch. – F. Chan, "Split-Dimension Vector Quantization of Parcor Coefficients for Low Bit Rate Speech Coding", *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 3, july 1994.
- [LCK96] J. H. Y. Loo, W. Y. Chan, P. Kabal, "Classified nonlinear predictive vector quantization of speech spectral parameters", *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Atlanta, May 1996.
- [LHM87] A. Lowry, S. Hossain, W. Millar, "Binary search trees for vector quantization", *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Dallas, 1987.
- [LIN00] J. Linden, "Channel Optimized Predictive Vector Quantization", *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 4, july 2000.
- [LPE99] E. Levin, R. Pieraccini, W. Eckert, "A stochastic model of human-machine interaction for learning dialog strategies", *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 1, January 1999.
- [LRT97] P. A. Laurent, "Expression of Spectral Distortion using Line Spectrum Frequencies", *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 5, september 1997.
- [LSH03] C. – C. Lee, Y. Shoham, "Trellis code excited linear prediction (TCELP) speech coding", *IEEE Transactions on Speech and Audio Processing*, vol. 11, nr. 6, nov. 2003.
- [LVC96] P. Lupini, V. Cuperman, "Nonsquare Transform Vector Quantization", *IEEE Signal Processing Letters*, vol. 3, no. 1, january 1996.
- [LWC95] J. H. Y. Loo, W. Y. Chan, "Nonlinear predictive vector quantization of speech spectral parameters", *Proc. IEEE Workshop on Speech Coding for Telecommunications*, Annapolis, Sept. 1999.
- [LWO01] I. Lee, H. C. Woo, "Encoding of Speech Spectral Parameters Using Adaptive Quantization Range Method", *ETRI Journal*, Vol. 23, No. 1, March 2001.

- [MAK75] J. Makhoul, "Linear Prediction: A Tutorial Review", *Proc. IEEE*, vol. 63, no. 4, april 1975.
- [MAR91] V. Mărâi, Gh. Mărâi, *Comanda vocală a sistemelor tehnice*, Editura Militară, București, 1991.
- [MAT98] A. Mateescu, S. Ciochină, N. Dumitriu, A. Șerbănescu, L. Stanciu, *Prelucrarea numerică a semnalelor*, Ed. Tehnică, București, 1998.
- [MCB95] A. V. McCree, T. P. Barnwell, "A Mixed Excitation LPC Vocoder Model for Low Bit Rate Speech Coding", *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 4, july 1995.
- [MCG97] S. McClellan, J. D. Gibson, "Variable-Rate CELP Based on Subband Flatness", *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 2, march 1997.
- [MDY94] N. Moreau, P. Dymarski, "Selection of excitation vectors for the CELP coders", *IEEE Transactions on Speech and Audio Processing*, vol. 2, nr. 1, jan. 1994.
- [MFG90] M. W. Marcellin, T. R. Fischer, J. D. Gibson, "Predictive trellis coded quantization of speech", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, nr. 1, ian. 1990.
- [MFS93] K. T. Malone, T. R. Fischer, "Trellis-searched adaptive predictive coding of speech", *IEEE Transactions on Speech and Audio Processing*, vol. 1, nr. 2, apr. 1993.
- [MGH90] D. S. Mazel, J. S. Geronimo, M. H. Hayes, "On the geometric sequences of reflection coefficients", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, nr. 10, oct. 1990.
- [MGR99] S. McClellan, J. D. Gibson, B. K. Rutherford, "Efficient pitch filter encoding for variable rate speech processing" *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 1, January 1999.
- [MHG76] J. T. Markel, A. H. Gray, Jr., *Linear Prediction of Speech*, Springer-Verlag, New York, 1976.
- [MNH94] N. Moayeri, D. L. Neuhoff, "Time-memory tradeoffs in vector quantizer codebook searching based on decision trees", *IEEE Transactions on Speech and Audio Processing*, vol. 2, nr. 4, oct. 1994.
- [MOR95] N. Moreau, *Tehniques de compression des signaux*, Ed. Masson, Paris, 1995
- [MTF93] K. T. Malone, T. R. Fischer, "Enumeration and trellis-searched coding schemes for speech LSP parameters", *IEEE Transactions on Speech and Audio Processing*, vol. 1, nr. 3, jul. 1993.
- [MYC91] Y. Medan, E. Yair, D. Chazan, "Super Resolution Pitch Determination of Speech Signals", *IEEE Transactions on Signal Processing*, vol. 39, no. 1, January 1991.
- [NDG94] S. H. Nam, J. D. Gibson, "Analysis of the smoothed residual-driven algorithm for speech coders", *IEEE Transactions on Speech and Audio Processing*, vol. 2, nr. 3, jul. 1994.
- [NGM01] E. Nemer, R. Goubran, S. Mahmoud, "Robust voice activity detection using higher-order statistics in the LPC residual domain", *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 3, March 2001.
- [NJY01] N. Jayant (editor), *Compresia semnalelor – Codarea vorbirii, a semnalelor audio, a textului, a imaginii și a semnalelor video*, Ed. Teora, București, 2001.
- [NMS93] C. R. Nassar, M. R. Soleymani, "Codebook design for trellis quantization using simulated annealing", *IEEE Transactions on Speech and Audio Processing*, vol. 1, nr. 4, oct. 1993.
- [NOL99] P. Noll, "MPEG digital audio coding", *IEEE Signal Processing Magazine*, 14(5), 1997.
- [NTL01] H.-W. Nein, Ch.-T. Lin, "Incorporating Error Shaping Technique into LSF Vector Quantization", *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 2, february 2001.
- [ORC91] M. Orchard, "A fast nearest neighbor search algorithm", *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Toronto, Canada, May 1991.
- [OSF89] A. V. Oppenheimer, R. W. Schafer, *Discrete – Time Signal Processing*, Prentice Hall, Englewood Cliffs, N. J., 1989.



- [PAK95] K. K. Paliwal, W. B. Kleijn, "Quantization of LPC parameters", in W. B. Kleijn and K. K. Paliwal, editors, *Speech Coding and Synthesis*, Elsevier, 1995.
- [PAL95] K. K. Paliwal, "Interpolation properties of linear prediction parametric representations", *Proc. European Conf. on Speech Commun. and Technology*, Madrid, 1995.
- [PAS00] T. Painter, A. Spanias, "Perceptual Coding of Digital Audio", *Proc. IEEE*, vol.88, no.4, april 2000.
- [PAT93] K. Paliwal, B. Atal, "Efficient Vector Quantization of LPC Parameters at 24 bit/frame", *IEEE Transaction on Speech and Audio Signal Processing*, vol. 1, no. 1, 1993.
- [PML95] A. Popescu, N. Moreau, C. Lamblin, "CELP Coding Using Trellis-Coded Vector Quantization of the Excitation", *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 6, nov. 1995.
- [PRM92] J. G. Proakis, D. G. Manolakis, *Digital Signal Processing: principles, algorithms and applications*, MacMillan Publishing Company, 1992.
- [PRP01] J. L. Pérez-Córdoba, A. J. Rubio, A. M. Peinado, A. de la Torre, "Joint source-channel coding for low bit-rate coding of LSP parameters", *Eurospeech 2001 – Scandinavia*, 2001.
- [PVT00] P. Prandoni, M. Vetterli, "R/D Optimal Linear Prediction", *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 6, nov. 2000.
- [QBC88] S. R. Quackenbush, T. P. Barnwell, M. A. Clements, *Objective Measure of Speech Quality*, Prentice-Hall, Englewood Cliffs, NJ, 1988.
- [QRH93] Y. Qi, B. R. Hunt, "Voiced-unvoiced-silence classifications of speech using hybrid features and a network classifier", *IEEE Transactions on Speech and Audio Processing*, vol.1, nr. 2, apr.1993.
- [RAL03] A. V. Rao, S. Ahmadi, J. Linden, A. Gersho, V. Cuperman, R. Heidari, "Pitch adaptive windows for improved excitation coding in low-rate CELP coders", *IEEE Transactions on Speech and Audio Processing*, vol. 11, nr. 6, nov. 2003.
- [RDS94] T. V. Ramabadran, D. Sinha, "Speech data compression through sparse coding of innovations", *IEEE Transactions on Speech and Audio Processing*, vol. 2, nr. 2, apr. 1994.
- [RGR00] R. Grigonis, *Computer Telephony Encyclopedia*, CMP Books, New York, 2000.
- [RKP99] V. Ramasubramanian, K.K. Paliwal, "Fast nearest-neighbor search based on Voronoi projections and its application to vector quantization encoding", *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 2, March 1999.
- [RMG90] R. M. Gray, *Source Coding Theory*, Kluwer Academic Press, Boston, 1990.
- [RMP03] S. A. Ramprasad, "The multimode transform predictive coding paradigm", *IEEE Transactions on Speech and Audio Processing*, vol. 11, nr. 2, mar. 2003.
- [RPK87] R. P. Ramachandran, P. Kabal, "Stability and performance of pitch filters in speech coders", *Proc. IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-35, 1987.
- [RRG98] S. Ramakrishnan, K. Rose, "Constrained-storage vector quantization with a universal codebook", *IEEE Trans. on Image Processing*, vol. 7, no. 6, june 1998.
- [RSS95] R. P. Ramachandran, M. M. Sondhi, N. Seshandri, B. S. Atal, "A Two Codebook Format for Robust Quantization of Line Spectral Frequencies", *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 3, april 1995.
- [RTB90] R. C. Rose, T. P. Barnwell, "Design and performance of an analysis-by-synthesis class of predictive speech coders", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, nr. 9, sept. 1990.
- [SAA00] L. M. da Silva, A. Alcaim, "Differential Coding of Speech LSF Parameters Using Hybrid Vector Quantization and Bidirectional Prediction", *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 2, march 2000.
- [SCA85] M. R. Schroeder, B. S. Atal, "Code-Excited Linear Prediction (CELP): High Quality Speech at Very Low Bit Rates", *IEEE ICASSP*, 1985.

- [SCC97] L. Supplee, R. Cohn, and J. Collura, "MELP: The new federal standard at 2400 bps", *Proc. ICCASP*, 1997.
- [SCH98] C. Schlegel, *Trellis Coding*, IEEE Press, New York, 1998.
- [SMF93] R. C. Snell, F. Milinazzo, "Formant location from LPC analysis data", *IEEE Transactions on Speech and Audio Processing*, vol. 1, nr. 2, apr. 1993.
- [SHD01] J. Samuelsson, P. Hedelin, "Recursive coding of spectrum parameters", *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 5, July 2001.
- [SJS87] M. Sondhi, J. Schroeter, "A Hybrid Time-Frequency Domain Articulatory Speech Synthesizer", *IEEE Transaction on Acoustic, Speech and Signal Processing*, ASSP-35, No. 7, July 1987.
- [SKB01] T. Shimamura, H. Kobayashi, "Weighted Autocorrelation for Pitch Extraction of Noisy Speech", *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 7, october 2001.
- [SOJ93] F. K. Soong, B. H. Juang, "Optimal Quantization of LSP Parameters", *IEEE Transactions on Speech and Audio Processing*, vol. 1, no. 1, january 1993.
- [SPA94] A. Spanias, "Speech coding: a tutorial revue", *Proc. IEEE*, vol 82, no. 10, oct. 1994.
- [SRO03] A. D. Subramaniam, B. D. Rao, "PDF optimized parametric vector quantization of speech line spectral frequencies", *IEEE Transactions on Speech and Audio Processing*, vol. 11, nr. 2, mar. 2003.
- [STA98] D. Stanomir, C. Negrescu, L. Jalba, *Algoritmi pentru prelucrarea semnalului vocal – Teorie și aplicații în comunicații GSM*, Editura Athena, București, 1998.
- [SWK00] J. Skoglund, W. B. Kleijn, "On Time-Frequency Masking in Voiced Speech", *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 4, july 2000.
- [SYH02] G. D. T. Schuller, B. Yu, D. Huang, B. Edler, "Perceptual Audio Coding Using Adaptive Pre – and Post Filters and Lossless Compression", *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 6, september 2002.
- [TAT90] I. M. Trancoso, B. S. Atal, "Efficient search procedures for selecting the optimum innovation in stochastic coders", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, nr. 3, mar. 1990.
- [TSA97] B. Tang, A. Shen, A. Alwan, G. Pottie, "A Perceptually Based Embedded Subband Speech Coder", *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 2, march 1997.
- [VJL91] R. A. DeVore, B. Jawerth, B. Lucier, "Data compression using wavelets: error, smoothness, and quantization", J. A. Storer and J. H. Reif, editors, *Proceedings Data Compression Conference*, April 1991.
- [VKB03] R. Veldhuis, E. Klabbers, "On the computation of the Kullback-Leibler measure for spectral distances", *IEEE Transactions on Speech and Audio Processing*, vol. 11, nr. 1, jan. 2003.
- [VLL00] H. L. Vu, L. Lois, "Efficient Distance Measure for Quantization of LSF and Its Karhunen-Loeve Transformed Parameters", *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 6, nov. 2000.
- [VOR99-1] S. Voran, "Objective Estimation of Perceived Speech Quality - Part I: Development of the Measuring Normalizing Block Technique", *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 4, july 1999.
- [VOR99-2] S. Voran, "Objective estimation of perceived speech quality – Part II: Evaluation of the measuring normalizing block technique," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no.4, july 1999.
- [VZZ97] J. Vass, Y. Zhao, X. Zhuang, "Adaptive Forward – Backward Quantizer for Low Bit Rate High – Quality Speech Coding", *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 6, november 1997.
- [WDG94] H. C. Woo, J. D. Gibson, "Low delay tree coding of speech at 8 kbit/s", *IEEE Transactions on Speech and Audio Processing*, vol. 2, nr. 3, jul.1994.
- [WEB1] <http://www.ling.su.se/staff/hartmut/kemplne.htm>



- [WEB2] <http://www.acoustics.hut.fi/~slemmet/dippa/chap2.html>
- [WGS91] L. Wang, M. Goldberg, S. Shlien, "Interleaved Image Adaptive Vector quantization", *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, Toronto, Canada, May, 1991.
- [WMG96] W. T. K. Wong, R. M. Mack, B. M. G. Cheetham, X. Q. Sun, "Low rate speech coding for telecommunications", *BT Technol J.*, Vol. 14, No. 1 January 1996.
- [WNF94] L. Wu, M. Niranjan, F. Fallside, "Fully vector-quantized neural network-based code-excited nonlinear predictive speech coding", *IEEE Transactions on Speech and Audio Processing*, vol. 2, nr. 4, oct. 1994.
- [WSS85] B. Widrow, S. Stearns, *Adaptive Signal processing*, Prentice Hall, Englewood Cliffs, New Jersey, 1985.
- [WUC97] Ch.-H. Wu, J.-H. Chen, „A Novel Two-Level Method for the Computation of the LSP Frequencies Using a Decimation –in –Degree Algorithm”, *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 2, march 1997.
- [WYG90] W.-Y. Chan, A. Gersho, "High fidelity audio transform coding with vector quantization", *Proceedings of the International Conference on Acoustic, Speech and Signal processing*, April 1990
- [XAD96] M. Xie, J.-P. Adoul, "Algebraic Vector Quantization of LSF Parameters with Low Storage and Computational Complexity", *IEEE Transactions on Speech and Audio Processing*, vol. 4, no. 3, may 1996.
- [XCP99] C. S. Xydeas, C. Papanastasiou, "Split matrix quantization of LPC parameters" *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 3, March 1999.
- [YCK02] R. Yu, C. C. Ko, "A warped linear-prediction-based subband audio coding algorithm", *IEEE Transactions on Speech and Audio Processing*, vol. 10, nr. 1, jan. 2002.
- [YHF93] Y. Hussain, N. Farvardin, "Variable-rate finite-state vector quantization and applications to speech and image coding", *IEEE Transactions on Speech and Audio Processing*, vol. 1, nr. 1, ian. 1993.
- [ZBL94] K. Zeger, A. Bist, T. Linder, "Universal Source Coding with Codebook Transmission", *IEEE Transaction on Communications*, vol. 42, Feb-March- April, 1994.
- [ZZW96] Z. Zhang, V. K. Wei, "An On-Line Universal Lossy Data Compression Algorithm via Continuous Codebook Refinement– part I: Basic Results", *IEEE Trans. on Information Theory*, vol. 42, may 1996.