



UNIUNEA EUROPEANĂ



GUVERNUL ROMÂNIEI



Instrumente Structurale  
2007-2013



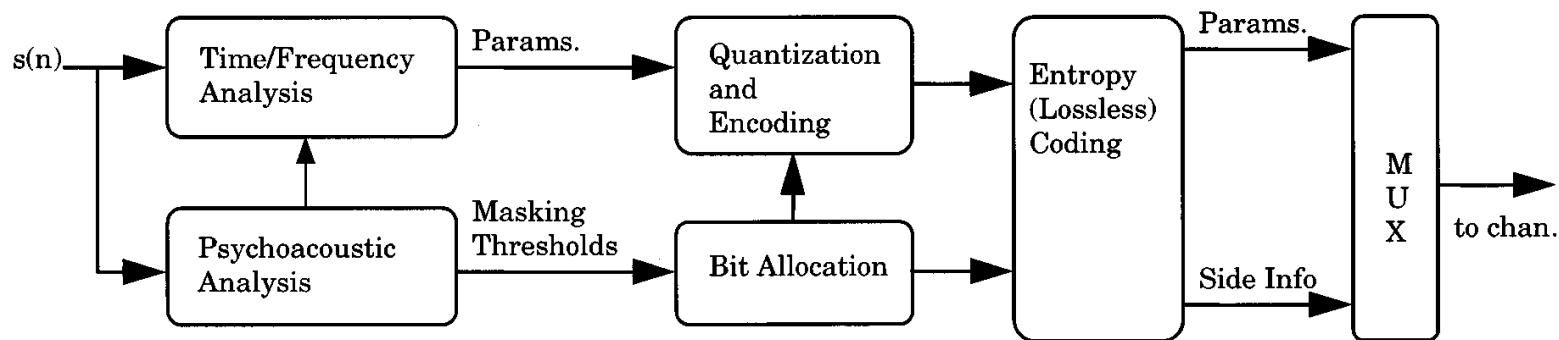
# Platformă de e-learning și curriculum e-content pentru învățământul superior tehnic

## Transmisia datelor multimedia in rețele de calculatoare

### 26. Compresia audio

# Compresia audio bazata pe modelul perceptiei audio (*Perceptual coding*)

- Schema de baza a unui codor-decodor audio, bazat pe model, este prezentata mai jos
- Se remarca existenta a doua piese de baza la partea de codare:
  - Un bloc pentru transformarea semnalului intr-o reprezentare care modeleaza precis sursa
  - Un model care modeleaza acusto-fiziologic (*psihacustic*, in engleza) aparatul auditiv
    - Acesta furnizeaza o aproximare a mecanismelor de perceptie ale urechii umane
    - Cele doua blocuri (transformari) determina marimea reducerii ce poate fi obtinuta prin procesul de codarea



# Compresia audio bazata pe modelul perceptiei audio (*Perceptual coding*)

- In etajul de reprezentare al semnalului are loc reducerea redundantei semnalului
  - In *codarea vorbirii*, un model fizic al tractului vocal este folosit pentru definirea parametrilor vorbirii
  - Acesti parametri, impreuna cu informatia reziduala, sunt codati. In timp ce aceasta tehnica permite obtinerea unor rapoarte de compresie foarte mari, nu este foarte eficienta cu semnale muzicale pentru ca este foarte dificil sa se modeleze bine toate sursele posibile de muzica
- In *codarea audio*, reprezentarea din domeniul timp a semnalului este tipic transformata intr-o reprezentare timp-frecventa cu ajutorul unui banc de filtre (cascada de filtre)
  - In acest caz, iesirile din domeniul frecventa ale bancului de filtre furnizeaza reprezentarea primara a semnalului

# Compresia audio bazata pe modelul perceptiei audio (*Perceptual coding*)

- Urechea umana realizeaza o analiza Fourier echivalenta cu un bank (lant) de filtre, unde transformarea frecventelor de lucru in „lucru mecanic” are loc in membrana basilara.
  - Spectrele de putere nu sunt reprezentate pe o scara liniara a frecventelor, ci pe benzi de frecventa limitate, denumite benzi critice
- *Sistemul auditiv* poate fi descris ca un banc de filtre trece banda, a caror caracteristica amplitudine – frecventa se suprapun
  - Filtrele au banda de ordinul 50-100 Hz pentru frecvente mai mici de 500 Hz si de pana la 5000 Hz pentru semnale de frecventa mai mare
  - In gama audio, se pot defini astfel 26 benzi critice acoperind intervalul de frecventa de pana la 24 KHz

# Compresia audio bazata pe modelul perceptiei audio (*Perceptual coding*)

- Ipoteza generala este ca semnalele audio sunt *cvasi-stationare*, deci, o corespondenta in domeniul frecventa rezulta intr-o reprezentare a semnalului care este mai eficienta decat PCM directa
  - Cu cat fereastra de analiza este mai mare cu atat presupunerea este mai buna
  - In general, se face un compromis intre o codare de eficienta inalta, data de un bank de filtre cu inalta rezolutie, obtinuta in etajul de reprezentare a semnalului si costul memoriei/intarzierii necesara intregului sistem de codare
- Cercetarile din domeniul rezolutiei spectrale au arata ca o alegere buna pentru o rezolutie de frecventa este de 20 Hz ce corespunde unei rezolutii in domeniul timp de 25ms

# Compresia audio bazata pe modelul perceptiei audio (*Perceptual coding*)

- Bancuri de filtre variante in timp sunt des utilizate pentru a evita imprastierea zgomotului de cuantizare in timp in semnalul reconstituit
- Aceste metode permit bancului de filtre sa se adapteze la o rezolutie de timp marita, tipic egala cu 5ms, in prezenta unui tranzient (impuls).

# Reducerea ratei de informatie

- Interesul pentru reducerea ratei de bit (low bit rate coding) este dat de necesitatea minimizării costurilor de transmisiei și a necesarului de memorie pentru stocare, de necesitatea de a transmite pe canale cu capacitate limitată, așa cum sunt canalele radio mobile, și pentru a suporta codarea cu pas variabil
- Compresia ratei de informație pentru semnale audio și vorbire se poate face prin: tehnici predictive, codare sub-banda, codare prin transformare, diferite forme de cuantizare vectorială
  - Se obțin astfel rate de 0.5-1 bit/esanțion pentru vorbire și 1-2 bit/ esanțion pentru semnale audio

# Codarea in domeniul de frecventa

## (*Frequency-domain coding*)

- In domeniul frecventa se poate defini *redundanta* ca fiind o caracteristica spectrala ascutita (ne-plata) pe termen scurt a sursei de semnal
  - *Irelevanta* inseamna componente spectrale cu amplitudine mai mica decat pragul psihoacustic
  - Cele doua trasaturi (redundanta si irelevanta) sunt folosite pentru a reduce cantitatea de simboluri in raport cu PCM
  - Acest lucru se obtine prin impartirea spectrului sursei in benzi de frecventa pentru a genera componente spectrale necorelate si prin cuantizarea acestor componente separat
- Se folosesc doua categorii de codare:
  - Codarea transformarii (TC=transform coding)
  - Codarea sub-benzii (SBC=subband coding)



# Codarea in domeniul de frecventa

- Diferentele dintre cele doua categorii sunt de ordin istoric
  - Amandoua utilizeaza o analiza *filterbank* in codor pentru a descompune semnalul de iesire in componente spectrale sub-esantionate
  - Componentele spectrale se numesc *esantioane subbanda* (subband samples) daca bancul de filtre are o rezolutie mica; in caz contrar se numesc *linii spetrale* sau *coeficienti de transformare*
  - Aceste componente spectrale sunt recombinate in decodor prin sinteza fitrului bank

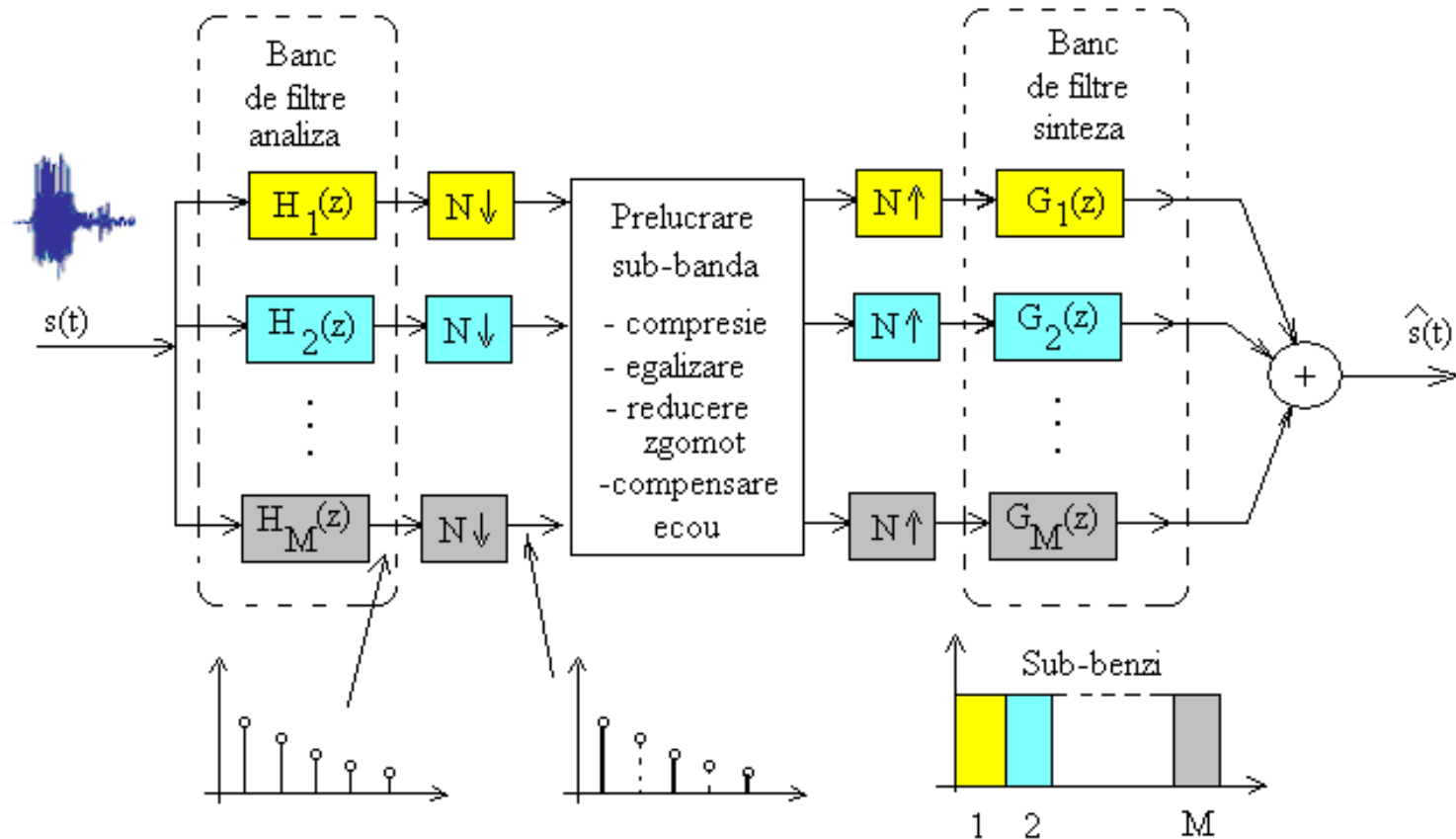
## Codarea sub-benzii (SBC=subband coding)

- In SBC, sursa de semnal este supusa unei analize filterbank constand din  $M$  filtre trece banda care sunt continue in frecvente astfel incat multimea semnalelor subbanda poate fi recombinate aditiv pentru a produce semnalul original sau o versiune apropiata a acestuia
  - Fiecare iesire a filtrului este decimata la limita (se mai spune *critic decimata*)(adica esantionata la dublul benzii nominale) cu un factor egal cu  $M$ , numarul de filtre trece banda
  - Aceasta decimare rezulta intr-un umar de esantioane subbanda ce egaleaza acelea din semnalul original
  - In receptor, rata de esantionare a fiecărei subbenzi este crescuta la cea a sursei de semnal prin completarea la un numar apropiat cu esantioane zero

# Codarea sub-benzii

- Semnalele subbanda interpolate apar la iesirile FTB ale filtrului de sinteza
- Procesul de esantionare poate introduce distorsiune de esantionare (aliasing distortion) din cauza suprapunerii subbenzilor
- Daca se folosesc filtre perfecte, (filtre dublu-banda in cuadratura sau filtre polifazice), termenii aliasing sunt eliminati si suma iesirilor filtrelor trecebanda este egala cu semnalul de intrare, in absenta cuantizarii

# Codarea sub-benzii



# Codarea transformării (TC=transform coding)

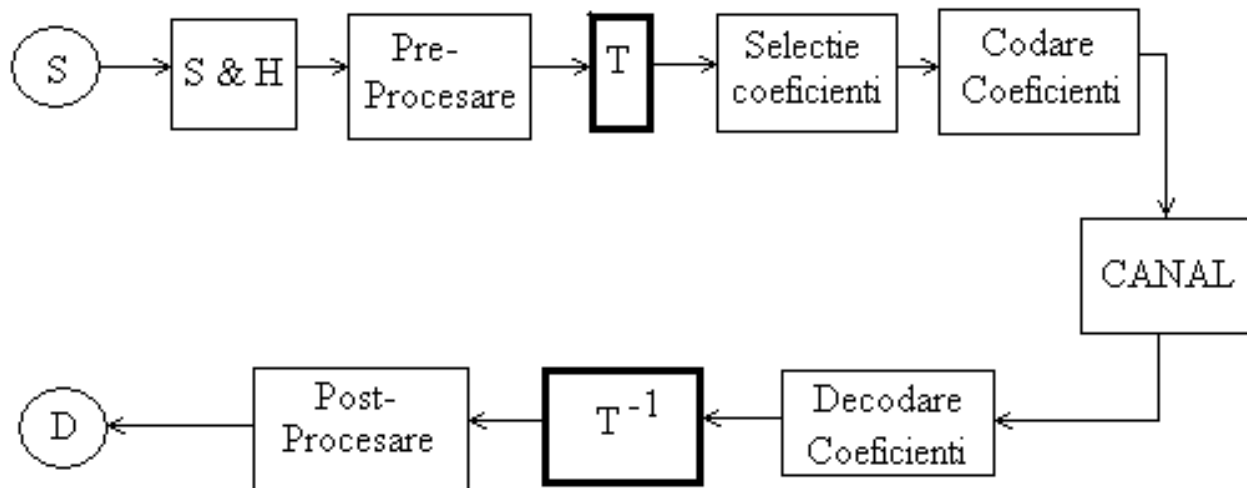
- In TC, un bloc de esantioane de intrare este transformat liniar cu o transformata discreta intr-un set de coeficienti de transformare, aproape ne-corelati
  - Acesti coeficienti sunt cuantizati si transmisi in forma numerica spre decodor
- In decodor are loc transformarea inversa, refacand semnalul in domeniul timp
  - In absenta erorilor de cuantizare sinteza furnizeaza o reconstructie exacta
- Transformari tipice sunt: *transformata Fourier discreta*, *transformata cosinus discreta* (DFT), sau *transformata cosinus discreta modificata* (MDCT)

# Codarea transformarii

- MDCT (Modified DCT) se bazeaza pe o suprapunere de 50% a filtrelor succesive de analiza
  - Efectul consta in obtinerea unui castig mai mare si a unui raspuns al filtrelor individuale mai bun.
- S-a mentionat ca decodorul bazat pe transformata inversa poate fi privit ca un filtru bank; raspunsul la impuls al filtrelor trece banda componente este egal cu secventa transformarilor de baza
- Raspunsul la impuls al filtrului de analiza este versiunea inversata in timp

# Codarea transformarii

- Schema a bloc a sistemului cu compresia audio folosind codarea transformarii



# Codarea transformarii

- Pre-procesarea inseamna una sau mai multe prelucrari de tipul: filtrare, compresia domeniului de variatie a amplitudine, folosirea unui model perceptual pentru indepartarea informatiei nerelevante
- Prin **T** este reprezentata o transformare de tipul Fourier, DCT sau MDCT
  - Selectia coeficientilor inseamna selectarea celor mai importanti coeficienti pentru indeplinirea unui criteriu de eroare minima
  - Codarea coeficientilor inseamna – de fapt – o codare de canal, adica adaugarea unor simboluri de control pentru protectia la perturbatii si, eventual, a unor informatii privitoare la parametrii transformarii din etapa de pre-procesare
- La receptie are loc decodarea coeficientilor si aplicarea unei transformari inverse pentru refacerea semnalului initial



# Metoda de compresie LPC

- LPC = Linear Predictive Coding
- LPC exploateaza redundanta semnalului vorbire prin mecanismul de generare a vorbirii: un filtru liniar excitat de un semnal numit semnal de excitatie
- Semnalul excitatie se mai numeste si semnal rezidual
- Codoarele pentru vorbire prelucreaza un grup de esantioane, numit cadru sau segmen
- Codorul vorbirii calculeaza coeficientii filtrului si semnalul de excitatie pentru fiecare cadru
  - Coeficientii filtrului sunt astfel calculati pentru a asigura ca energia la iesirea filtrului pentru cadrul analizat sa fie minimizata
  - Un astfel de filtru se numeste filtru LP analizor
  - Semnalul rezultat se numeste semnal *rezidual* pentru acel cadru particular
- La decodare, filtrul LP lucreaza ca un filtru de sinteza, in timp ce semnalul rezidual actioneaza ca semnal de excitatie pentru filtrul de sinteza

# Metoda de compresie LPC

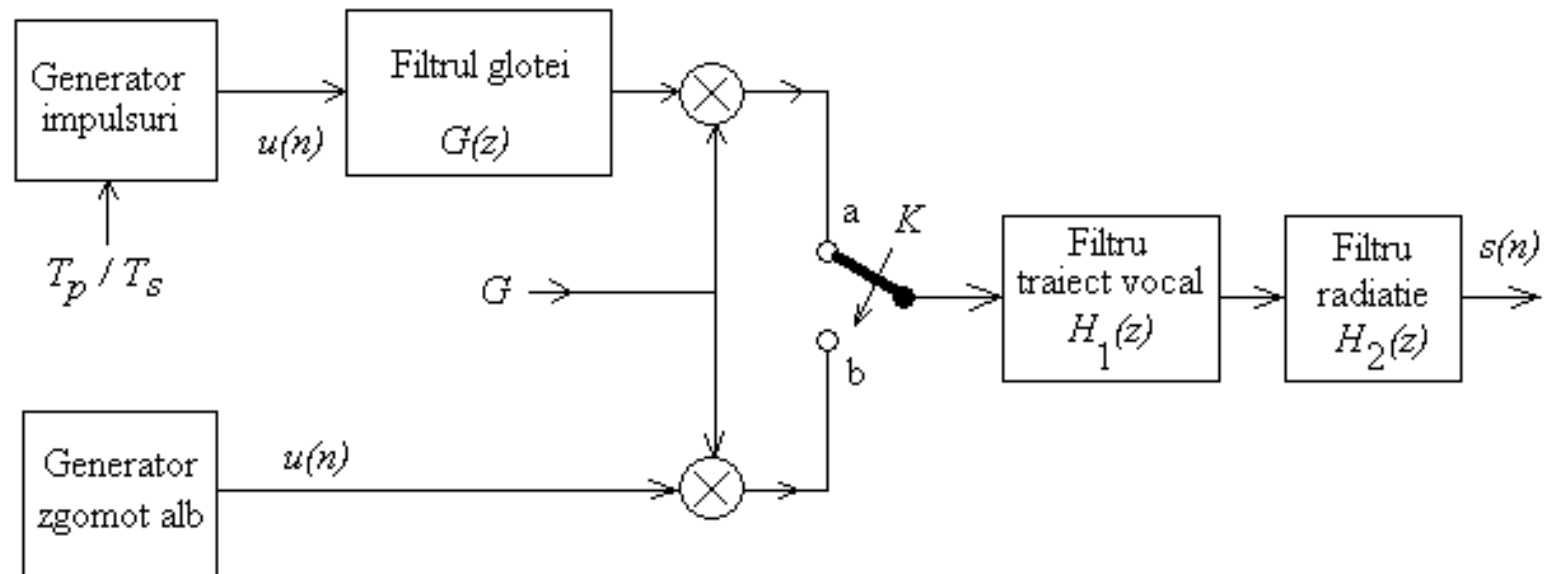
- In vederea reducerii ratei de bit totale, codoarele vorbirii cum sunt CELP (*code excited linear prediction*) nu transmit intreg semnalul rezidual, intrucat se foloseste un tabel de codare vectorial pentru a coda semnalul de excitatie
- Tehnica este numita cuantizare vectoriala (VQ), astfel incat codorul selecteaza unul dintre semnalele de excitatie dintr-un tabel predeterminat, si va transmite indexul semnalul de excitatie care se afla cel mai aproape de cel care trebuie transmis
- Tabelul cu semnalele de excitatie este cunoscut atat de codor cat si de decodor
- Semnalul excitatie este selectat astfel incat distorsiunea dintre cadrul original si cel reconstruit sa fie minima

# Metoda de compresie LPC

- Tipic, se folosesc rate de esantionare de 8 KHz iar lungimea cadrului este de 20 ms, astfel incat exista 160 esantioane in fiecare cadru
- Se gaseste ca un filtru de ordinul 10 este suficient pentru modelarea anvelopei spectrale pentru o rata de esantionare de 8 KHz, deci codorul lucreaza cu 12 parametri (10 coeficienti, castigul filtrului si indexul semnalului de excitatie) in loc de 160 de esantioane pentru un cadru

# Metoda de compresie LPC

- Modelul vorbirii

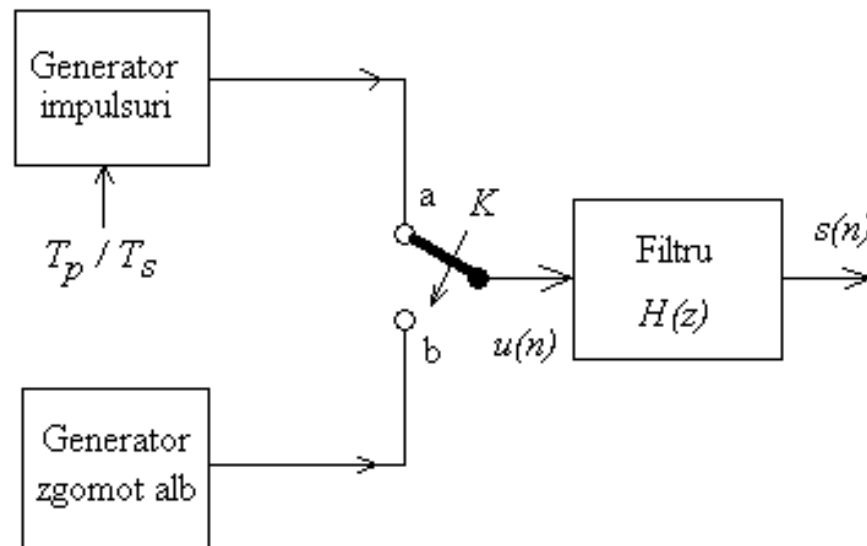


# Metoda de compresie LPC

- Modelului de vorbire are doua moduri de functionare
  - In primul mod, comutatorul  $K$  este in pozitia a si se genereaza fonemele sonore
    - Semnalul excitatie este constituit dintr-o serie de impulsuri dreptughiulare cu perioada  $T_p$ ,  $T_s$  este frecventa de esantionare
  - Al doilea mod de lucru necesita comutatorul  $K$  in pozitia b, si corespunde fonemelor insonore
    - Exista trei filtre care modeleaza efectele glotei, ale tractului vocal si al radiatiei sonore (efectul buzelor)
    - Multiplicarea cu constanta  $G$  este pentru modificarea intensitatii.

# Metoda de compresie LPC

- In aplicatiile practice se utilizeaza o schema mai compacta, unde cele trei filtre sunt inlocuite cu unul singur,  $H(z)$



# Metoda de compresie LPC

- Pentru filtru se foloseste structura ARMA (*Autoregressive Moving Average*)
  - Un semnal vorbire  $s(n)$  este considerat ca iesire a unui sistem excitat de un semnal excitatie  $u(n)$
  - Esantionul vorbirii  $s(n)$  este modelat ca o combinatie liniara a iesirilor vechi si prezente si a intrarilor vechi, dupa relatia:

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

unde  $G$  este castigul filtrului si  $\{a_k, b_k\}$  sunt parametrii modelului

- Numarul  $p$  implica folosirea a  $p$  esantioane trecute, si se numeste *ordinul* predictiei liniare
  - Functia de transfer  $H(z)$  a modelului se obtine prin aplicarea transformatei in  $z$  ecuatiei:

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

ceea ce arata un model de tip poli-zerouri. In spectrul vorbirii, nazalele sunt reprezentate de zerouri si formantii sunt reprezentati de poli.

# Metoda de compresie LPC

- Exista doua cazuri speciale ale acestui model:
  - Cazul modelului de tip auto-regresiv, cand  $H(z)$  are numai poli, deci coeficientii  $b_k$  sunt nuli
    - Acest model se foloseste din motive de simplitate si eficienta a implementarii. Ideea este ca in rezolvarea unui model poli-zeroruri este nevoie de rezolvarea unui set de ecuatii neliniare in timp ce modelul numai cu poli necesita rezolvarea unui set liniar de ecuatii
  - Cazul modelului de tip medie alunecatoare, cand  $H(z)$  are numai zerouri, deci coeficientii  $a_k$  sunt zero



# Metoda de compresie LPC

- Transformata  $z$  inversa a modelului cu toti polii este

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

- Coeficientii  $\{a_k\}$  sunt numiti coeficientii LP ai filtrului liniar
- Semnalul eroare  $e(n)$  este diferenta dintre semnalul de intrare si cel estimat, dupa relatia

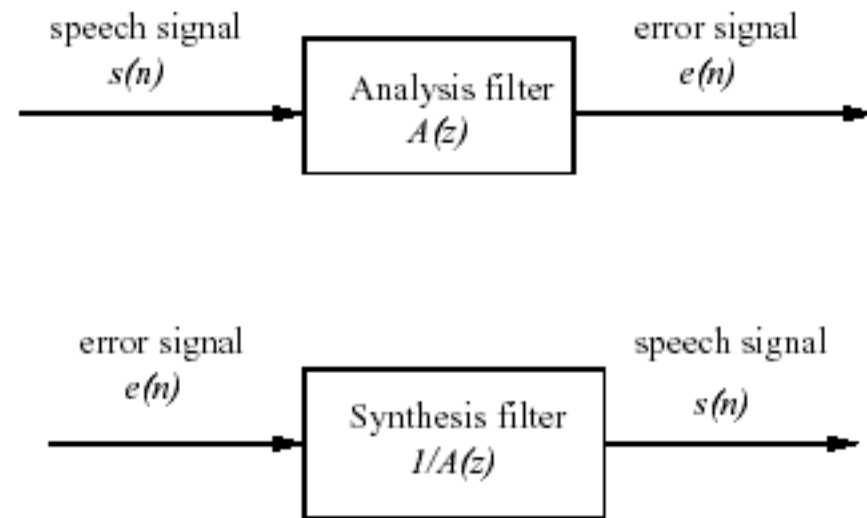
$$e(n) = s(n) - \hat{s}(n) = s(n) - \sum_{k=1}^p a_k \cdot s(n-k)$$

iar in domeniul  $z$  avem relatia de legatura

$$E(z) = S(z)A(z)$$

# Metoda de compresie LPC

- Ca urmare, modelul global poate fi descompus in doua parti
  - Partea de *analiza* analizeaza semnalul vorbire si produce semnalul eroare
  - Partea de *sinteza* preia semnalul eroare ca semnal de intrare, este filtrat de filtrul  $1/A(z)$  si se obtine semnalul vorbire
- Semnalul eroare se mai numeste semnal *rezidual* sau semnal de *excitatie*



# Metoda de compresie LPC

- Exista doua tipuri de codoare bazate pe predictie liniara:
  - Codor adaptiv inainte sau direct (*forward adaptive coder*)
    - Predictia liniara se bazeaza pe esantioane de intrare vechi
    - Analiza LP este efectuata in codor si se transmit coeficientii LP
  - Codor adaptiv inapoi sau invers (*Backward adaptive coder*)
    - Coeficientii LP sunt calculati din esantioanele vechi reconstruite
    - Analiza LP este facuta din nou la receptie, deci in cadrul decodorului
    - In acest fel, nu este necesar sa se transmita coeficientii LP de catre codor

# Estimarea coeficientilor predictiei liniare

- Exista doua metode larg folosite in estimarea coeficientilor predictiei liniare:
  - *autocorelatia*
  - *covarianta*
- Ambele metode folosesc coeficientii filtrului LP  $\{a_k\}$  astfel incat energia reziduala (energia semnalului rezidual) sa fie minimizata
- Cea mai folosita metoda este metoda celor mai mici patrate

# Ponderarea pe ferestre (*windowing*)

- Semnalul vorbire este un semnal variabil in timp si unele variatii sunt aleatoare
  - Uzual, in timpul vorbirii cu intensitate mica, forma tractului vocal si a excitatiei nu se modifica in 200 ms
  - Dar *fonemele* au o durata medie de 80 ms
  - Cele mai multe schimbari apar mai frecvent decat intervalul de 200 ms
- Analiza semnalului presupune ca proprietatile semnalului se schimba lent in timp, ceea ce permite o analiza in timp scurt a semnalului
  - Semnalul este divizat in segmente succesive, analiza se face pe aceste segmente si se extrag anumite proprietati dinamice

# Ponderarea pe ferestre (*windowing*)

- Semnalul  $s(n)$  este înmulțit cu o fereastră de analiză  $w(n)$  pentru a extrage un anumit segment de analiză
  - Tehnica se numește *ponderare (windowing)*
- Alegerea formei ferestrei este importantă întrucât esanționale vor fi ponderate în mod diferit
  - Folosirea unei ferestre dreptunghiulare determină apariția unor lobi secundari mari și anumite efecte nedorite în domeniul frecvență
  - Pentru a înlătura aceste oscilații mari, se utilizează ferestre de filtrare fără schimbări abrupte în domeniul timp, așa cum este fereastră Hamming, Blackman, Kaiser sau Bartlett
  - Se pot folosi și combinații de ferestre, așa cum se folosește în GSM, unde fereastră de analiză are două jumătăți de ferestre Hamming de dimensiuni diferite

# Reprezentarea parametrilor LP

- Parametrii LP au gama dinamica destul de mare, astfel incat sunt dificil de cuantizat
- Alte moduri de reprezentare au o gama dinamica mai mica si posibil de estimat
- Acest lucru permite folosirea unui numar diferit de biti pentru cuantizare, cu atat mai mic cu cat efectul coeficientilor asupra raspunsului filtrului este mai mic

# Reprezentarea parametrilor LP

- Coeficientii predictiei liniare pot avea diferite reprezentari echivalente:
  - frecventele liniilor spectrale (*line spectral frequencies – LSF*)
  - coeficienti de reflectie (*reflection coefficients – RC*);
  - autocorelatii (*autocorrelations*)
  - rapoarte logaritmici ale ariilor (*log area ratios – LAR*)
  - coeficienti de reflexie arcsine (*arcsine of reflection coefficients – ASRC*)
  - raspunsul la impulse al filtrului LP ( *impulse responses of LP synthesis filter - IR*)



# Reprezentarea parametrilor LP

- Toate aceste reprezentari au relatii de legatura exacte cu coeficientii LP si pastreaza toata informatia desprea acestia
  - Dintre acestia, o parte sunt mai usor de calculat si au proprietati particulare ce le fac mai atractive pentru diferite scopuri

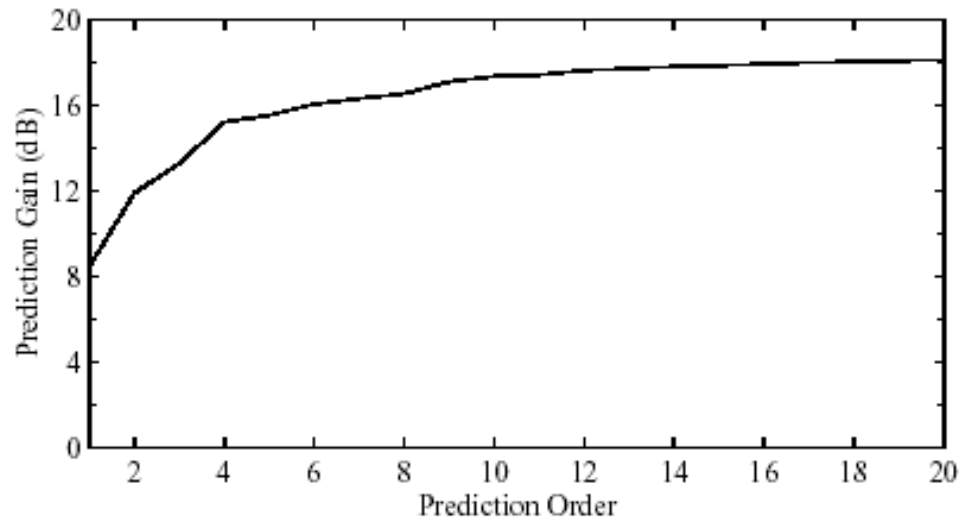


Fig. 3.3 Prediction order vs. prediction gain (male speech, 3.872 s)

# Metoda autocorelatiei

- Mai intai, semnalul vorbire  $s(n)$  este multiplicat cu o fereastră  $w(n)$  pentru a obtine un segment de vorbire ponderat  $s_w(n)$ , dupa relatia

$$s_w(n) = w(n)s(n)$$

- Al doilea pas este minimizarea energiei semnalului rezidual; energia reziduala se obtine cu expresia

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

# Metoda autocorelatiei

- Valorile parametrilor  $\{a_k\}$  care minimizeaza energia  $E$  se gasesc prin calcularea derivatelor partiale ale lui  $E$  in raport cu  $\{a_k\}$  si egalarea lor cu zero

$$\frac{\partial E}{\partial a_k} = 0, \quad \text{pentru } k = 1, \dots, p$$

de unde rezulta un set de  $p$  ecuatii cu  $p$  necunoscute

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

- In ultima ecuatie semnalul ponderat  $s_w(n)=0$  in afara ferestrei de filtrare

# Metoda autocorelatiei

- Ecuatiile liniare pot fi exprimate in functie de functia de autocorelatie
- Functia de autocorelatie a unui segment ponderat de vorbire este definita de relatia

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

unde  $N_w$  este lungimea ferestrei

- Prin inlocuirea valorilor autocorelatiei in penultima ecuatie se obtine

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

# Metoda autocorelatiei

- Setul de ecuatii liniare poate fi scris matricial sub forma

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

- Ultima ecuatie poate fi scrisa matricial sub forma  
 $\mathbf{R} \cdot \mathbf{a} = \mathbf{r}$

# Metoda autocorelatiei

- Matricea  $\mathbf{R}$  este o matrice Toeplitz, deci are toate elementele de pe diagonala principala egale
  - Acest lucru permite rezolvarea ecuatiei matriciale prin algoritmul Levinson-Durbin sau prin algoritmul lui Schur
- Datorita structurii matricii  $\mathbf{R}$ ,  $A(z)$  este de tip faza minima
  - Filtrul folosit la sinteza are functia de transfer  $H(z) = 1 / A(z)$ , deci zerourile lui  $A(z)$  devin polii lui  $H(z)$
  - Astfel, faza minima a lui  $A(z)$  garanteaza stabilitatea filtrului  $H(z)$

# Algoritmul Levinson-Durbin

- Algoritmul rezolva ecuatia matricala  $\mathbf{Ax}=\mathbf{b}$ , in care  $\mathbf{A}$  are o structura Toeplitz, matrice simetrica si pozitiv definita, iar  $\mathbf{b}$  este un vector arbitrar
  - Aceste cerinte sunt satisfacuate de matricea de autocorelatie  $\mathbf{R}$ , definita anterior.
- Algoritmul LD necesita o forma speciala a lui  $\mathbf{b}$ , unde  $\mathbf{b}$  consta din anumite elemente ale lui  $\mathbf{A}$ 
  - Si aceasta conditie este satisfacuta de ecuatiile autocorelatiei

# Algoritmul Levinson-Durbin

- Fie  $a_k(m)$  coeficientul  $k$  pentru un cadru oarecare  $m$  al iteratiei
- Algoritmul LD rezolva iterativ setul de ecuatii in mod iterativ pentru  $m=1,2,\dots,p$  dupa relatiile:

$$k(m) = R(m) - \sum_{k=1}^{m-1} a_k(m-1)R(m-k) \qquad a_m(m) = k(m)$$

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

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



# Algoritmul Levinson-Durbin

- Se considera initial  $E(0) = R(0)$  si  $a(0) = 0$ 
  - La fiecare iteratie, coeficientul  $a_k(m)$  pentru  $k=1,2,\dots,m$  descrie predictorul linear de ordin  $m$ ; eroarea de energie  $E(m)$  este redusa cu un factor de  $(1-k(m)^2)$ .
- Intrucat  $E(m)$  nu este negativa, rezulta ca  $|k(m)| \leq 1$
- Aceasta conditie pentru coeficientul de reflexie  $k(m)$  garanteaza ca radacinile lui  $A(z)$  vor fi in interiorul cercului unitate
  - Rezulta ca sinteza filtrului  $H(z)$  va fi stabila si – deci – metoda corelatiei determina stabilitatea filtrului cautat