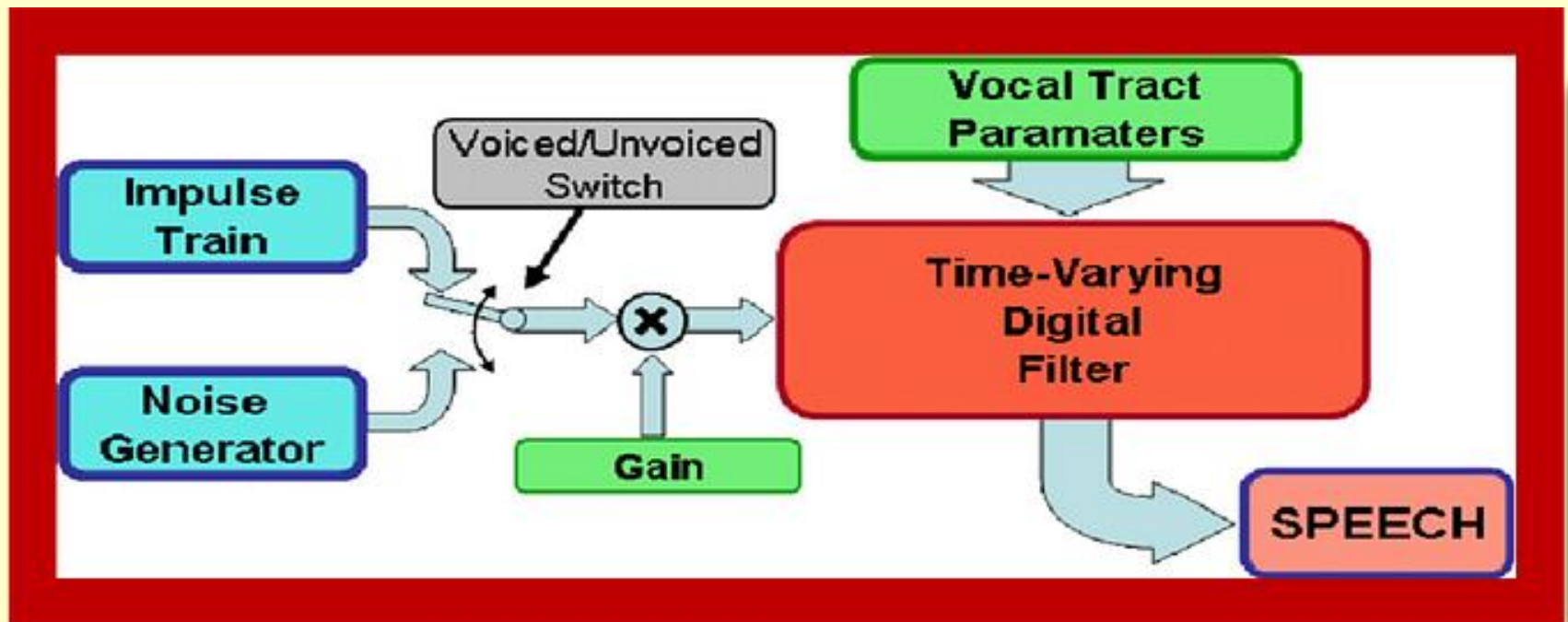
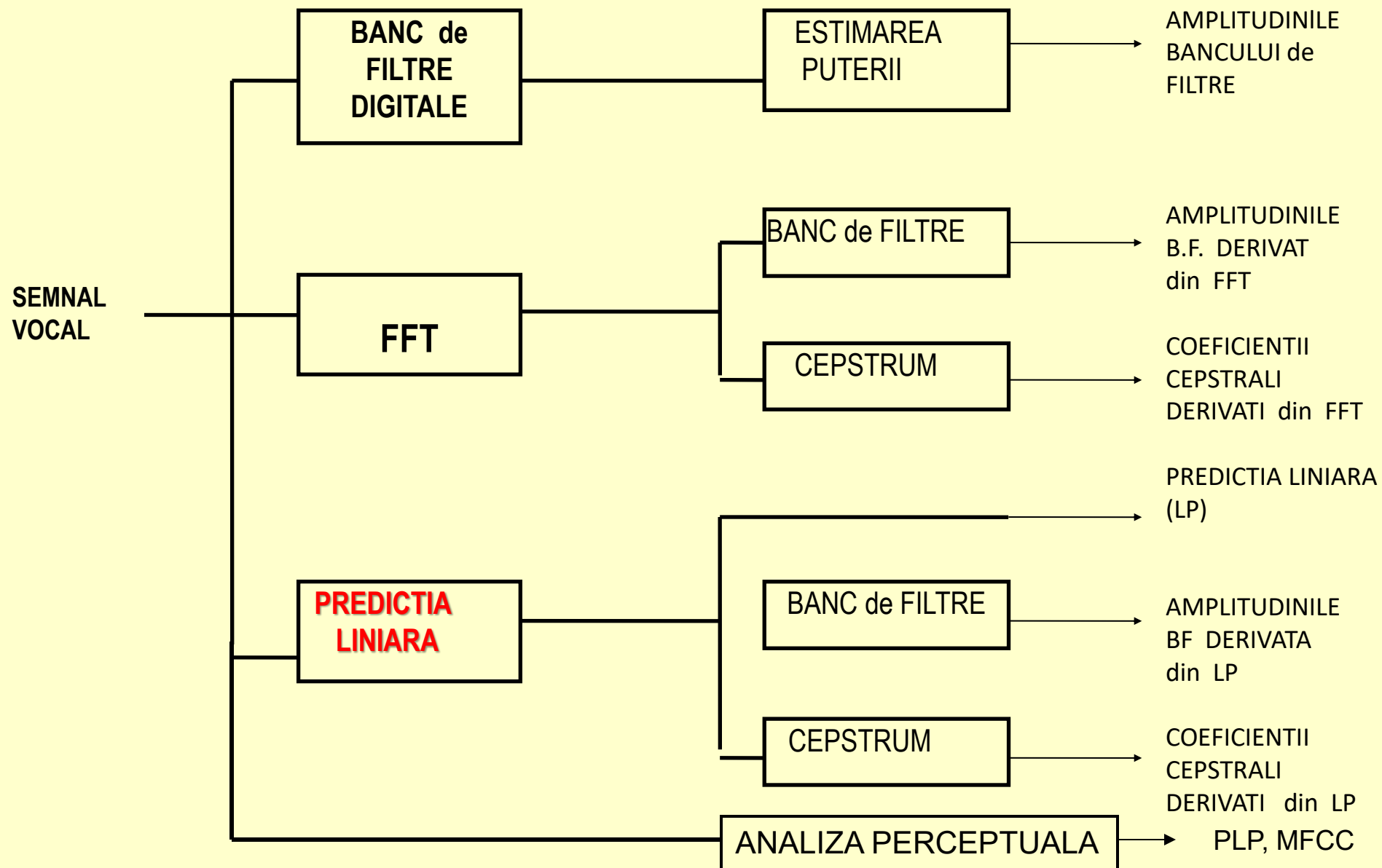


ASRSV – curs 5

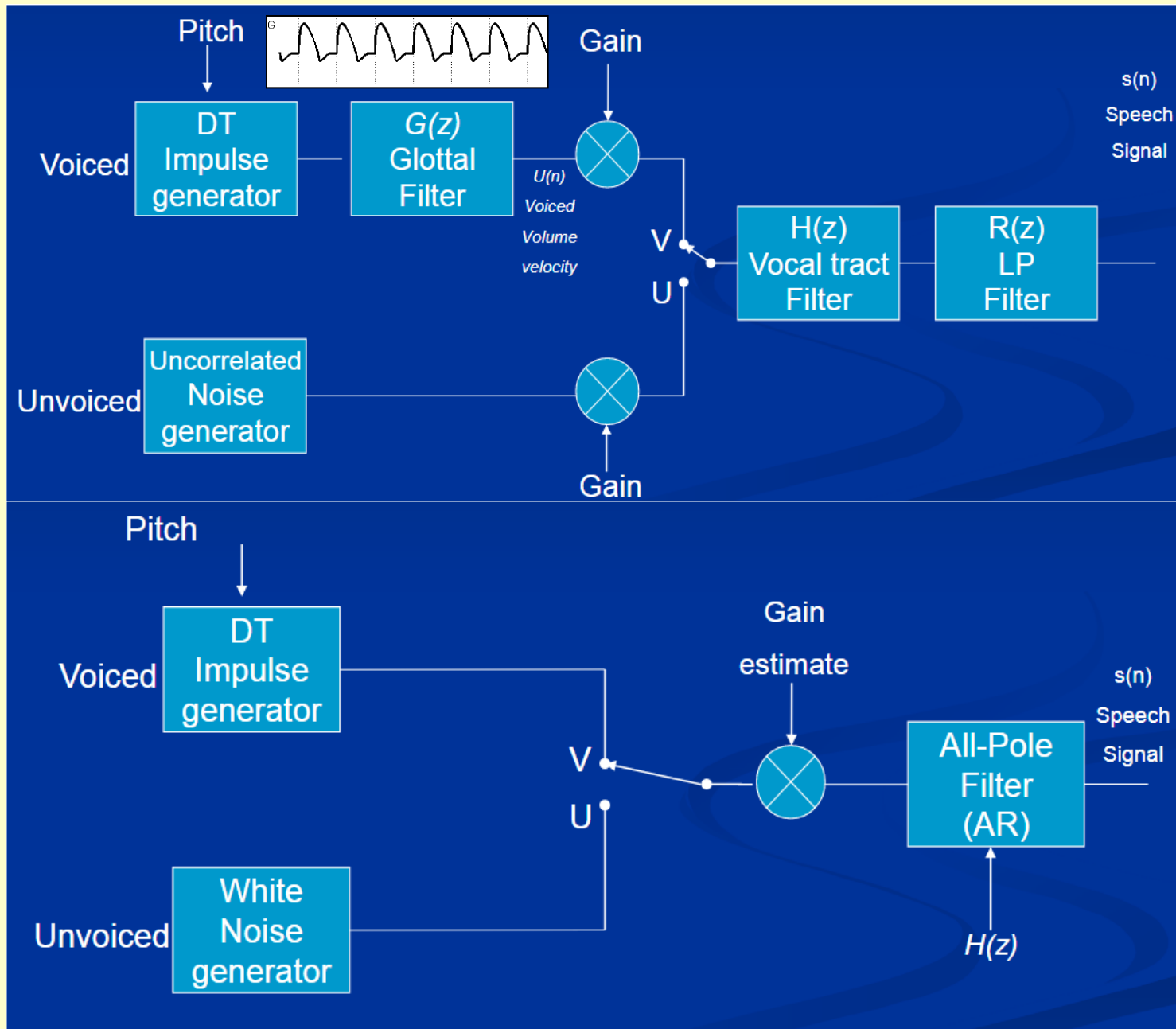
Analiza Semnalului Vocal (3)

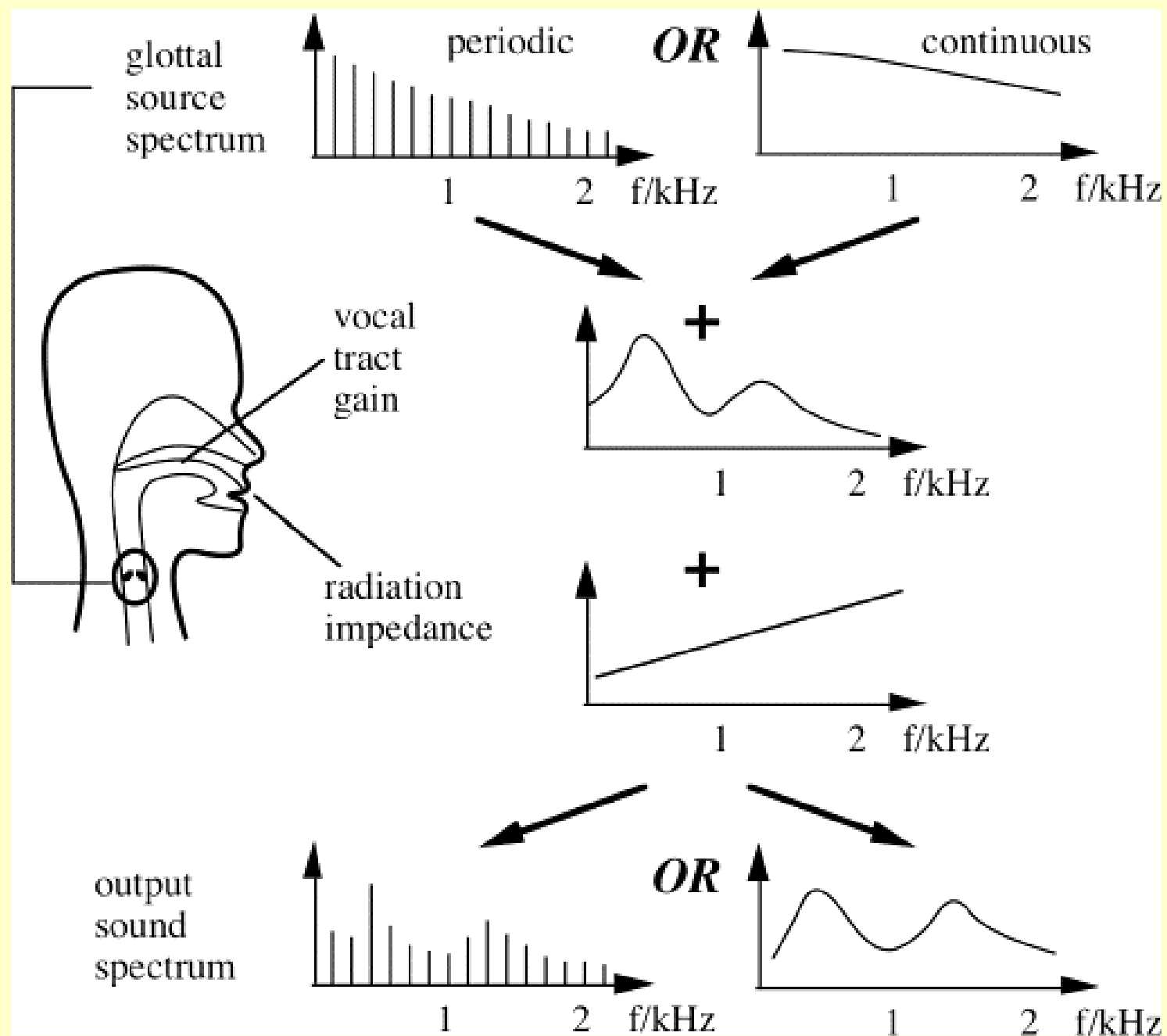




Algoritmi folosiți în analiza spectrală [adaptare dupa Picone]

1. ANALIZA PRIN PREDICTIE LINIARA - LPC





$$s(n) \approx a_1 s(n-1) + a_2 s(n-2) + \dots + a_p s(n-p),$$

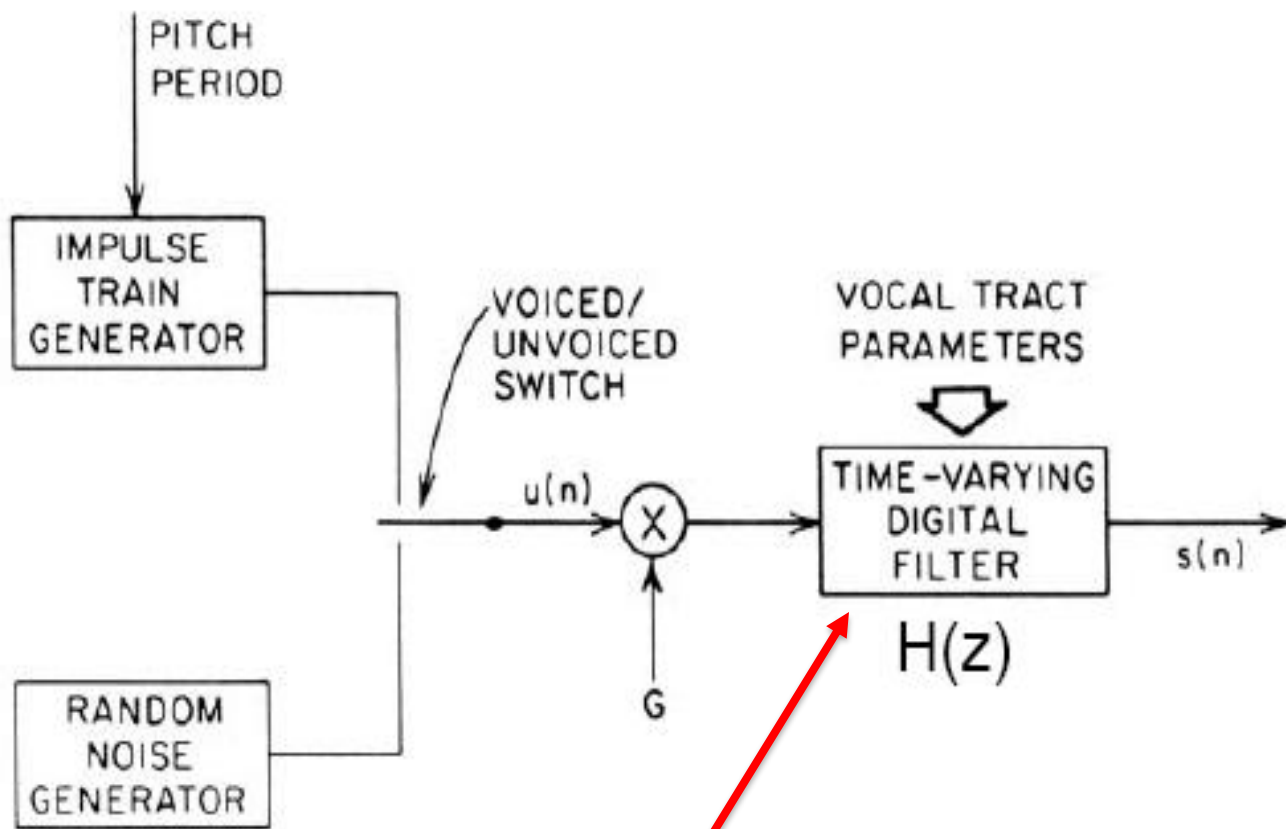
Motivele folosirii modelului LPC:

1. Are o interpretare fizica adecvata (rezonante=formanti)
2. Face o bună ~ a semnalului vocal mai ales în zonele sonore
3. Analiza LPC duce la separarea sursă/tract
4. Volum scăzut de calcule
5. Adecvat aplicațiilor de recunoaștere/sinteză
6. Se pretează la implementări hard

Interpretările modelului predicției liniare:

1. Identificarea sistemului
2. Filtrare inversă
3. Predicție liniară
4. Netezire spectrală

PARAMETRII MODELULUI

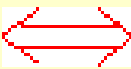
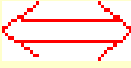
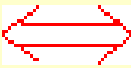
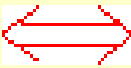


Relatia intre modelul matematic si reprezentarea fizica :

- Coeficienții filtrului (a_i)
- Câștigul (G)
- Decizia Sonor/Nesonor (voiced/unvoiced)
- Frecvența fundamentală (F_0)

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

Relatia intre modelul matematic si reprezentarea fizica :

Tract Vocal	 $H(z)$ (Filtrul LPC)
Aer	 $u(n)$ (esantioane)
Vibratiile corzilor vocale	 V (sonor-voiced)
Perioada vibratiilor	 T (perioada pitch)
Fricative/Plozive	 UV (nesonor-unvoiced)
Volum Aer	 G (Castig)

- PL se bazeaza pe modele de productie si sinteza a SV - vocea poate fi modelata ca iesirea unui sistem variabil in timp, excitat de impulsuri quasi-periodice sau zgomot
- PL furnizeaza o metoda *robusta, precisa si sigura* pentru estimarea parametrilor sistemului liniar (combina tractul vocal, impulsurile glotale si radiatia caracteristica sunetelor sonore)

Obs.

- Modelul LPC poate fi reprezentat ca un vector de forma ($p=10$):

$$\mathbf{A} = (a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}, G, V/UV, T)$$

- $\mathbf{A}$ se modifica la fiecare 20 msec (daca avem o frecventa de esantionare de 8kHz), 20 msec este echivalent cu 160 de esantioane.
- SV esantionat se imparte in cadre de $T = 20$ msec, deci avem 50 cadre/sec.
- Modelul sustine ca:

$$\mathbf{A} = (a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}, G, V/UV, T)$$

|||

Este echivalent cu

- DECI :

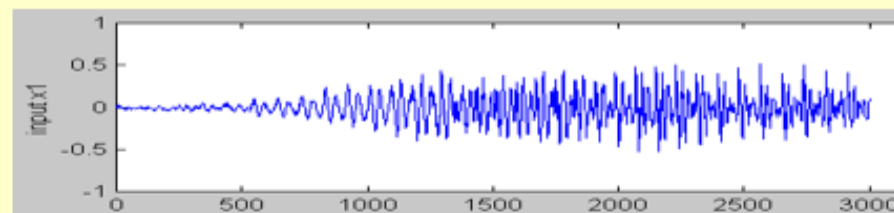
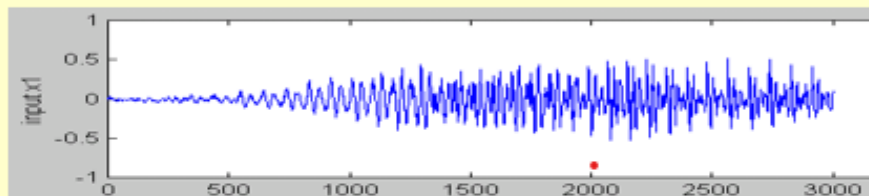
$$\mathbf{S} = (s(0), s(1), \dots, s(159))$$

160 de valori din $S(n)$ (esantioane) sunt reprezentate de 13 valori din $\mathbf{A}$ (model LPC)

compresie > 12

EXEMPLU

Semnal intrare



Se poate reconstrui semnalul din coef.
LPC+...



cadru y

20ms

cadru y+1

20ms

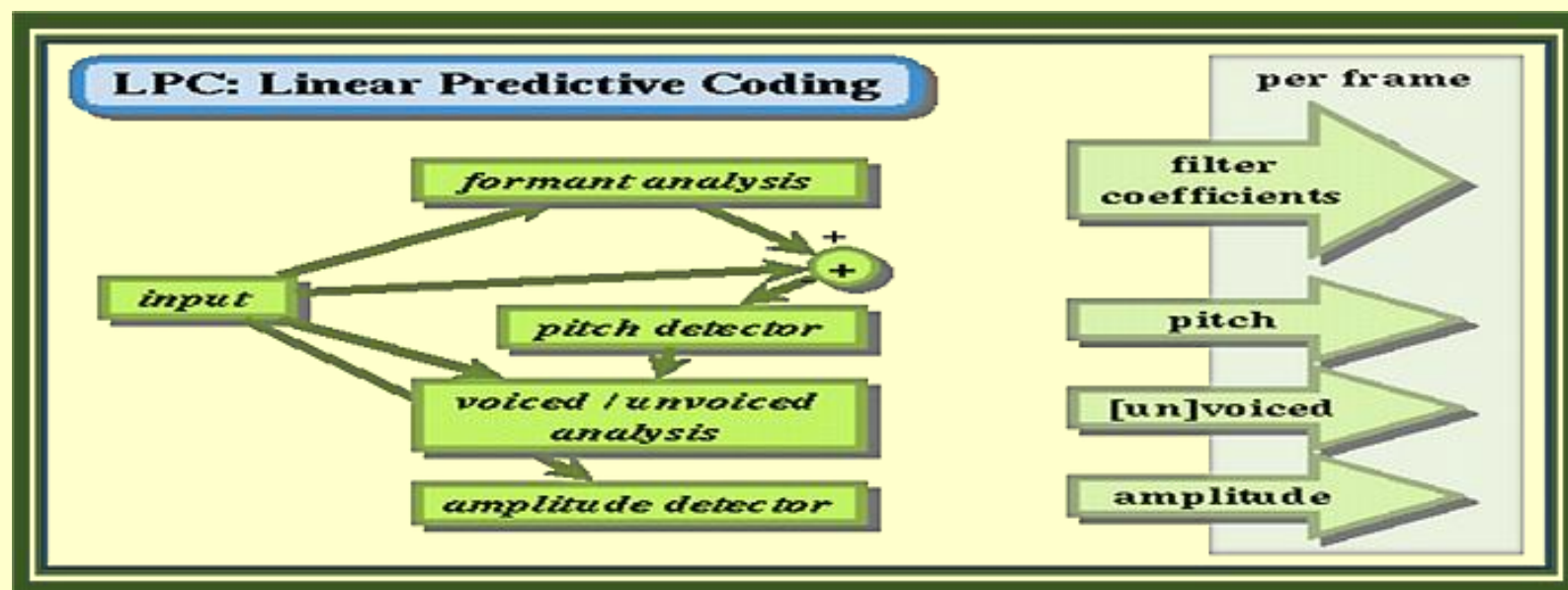
cadru y+2

20ms

$\alpha_1, \alpha_2, \alpha_3, \alpha_4, \dots, \alpha_8$

$\alpha'_1, \alpha'_2, \alpha'_3, \alpha'_4, \dots, \alpha'_8$

$\alpha''_1, \alpha''_2, \alpha''_3, \alpha''_4, \dots, \alpha''_8$



Exemple de codare LPC

Un SV eșantionat la 8kHz și fiecare eșantion cuantizat pe 8 biți necesită un debit binar de transmisie de 64 kbps (PCM)

1. O reprezentare **LPC** cu ajutorul unui cadru (m) de 25ms, necesită: 1bit pentru V_m (S/N), 3 biți pentru G_m (castig), 4 biți pentru T_m (F_0) și 12 coeficienți*8 biți de fiecare dată pentru H_m , adică un debit binar de 4.16 kbps.
 2. În celula **GSM-standard de telefonie**, se folosește o codare LPC modificată (RPE-LTP - Regular Pulse Excitation - Long Term prediction), permite o reducere a debitului vocii de calitate transmisă prin telefon de la 104 kbps (8 kHz, 13 biți) la 13 kbps (20 ms/ cadru), deci un factor de compresie de 8.
 3. **LPC10[FS1015]** – standardul de codare pentru comunicare vocală la 2.4kbps. LPC10 utilizează $f_s = 8\text{kHz}$, cadre de 22.5ms și 10 coeficienți LPC.
- Raporturi de compresie mai mari sunt posibile cu ajutorul unor algoritmi mai complecși.

2. Ecuatiile de baza ale PL

-Un predictor de ordin p este un sistem de forma:

$$\tilde{s}(n) = \sum_{k=1}^p \alpha_k s(n-k) \Leftrightarrow P(z) = \sum_{k=1}^p \alpha_k z^{-k} = \frac{\tilde{S}(z)}{S(z)}$$

- eroarea de predictie este de forma:

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

- Eroarea e(n) reprezinta iesirea unui sistem cu FT de forma:

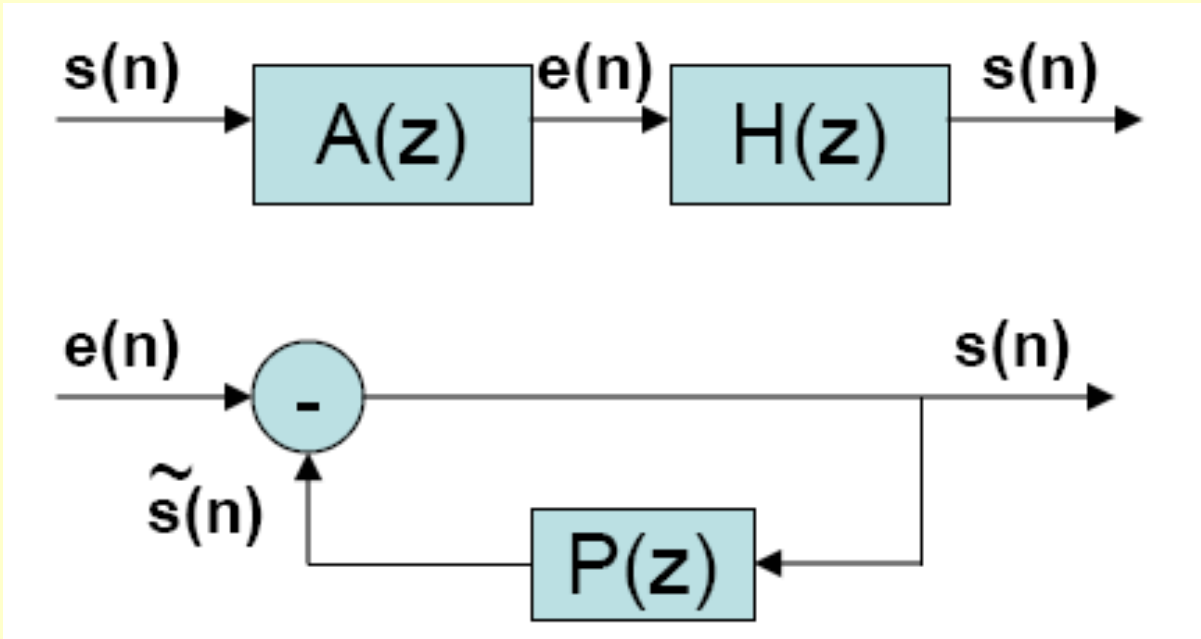
$$A(z) = \frac{E(z)}{S(z)} = 1 - \sum_{k=1}^p \alpha_k z^{-k}$$

- Daca SV respecta exact modelul de productie si

$$\alpha_k = a_k, 1 \leq k \leq p \Rightarrow e(n) = Gu(n)$$

si $A(z)$ este filtrul invers al lui $H(z)$.

$$H(z) = \frac{G}{A(z)}$$



ESTIMAREA MODELULUI

- trebuie determinati coeficientii de predictie α_k , astfel incat modelul sa estimeze cat mai bine variatia in timp a spectrului SV;
- *estimarea se va face pe cadre scurte ale SV si se va minimiza eroarea medie patratica de predictie pe aceste segmente;*
- α_k rezultati se presupun a fi coeficientii ($\mathbf{a}_k$) ai modelului de productie a SV

<https://www.youtube.com/watch?v=IWH-Oh5KnNY>

SOLUTIE ptr. α_k

- se selecteaza un segment de SV, in jurul esantionului n , $s_n(m)=s(m+n)$.
- eroarea medie patratica de predictie pe timp scurt E_n :

$$E = \sum_n e^2(n)$$

$$E = \sum_n \langle s(n) - \tilde{s}(n) \rangle^2$$

$$E = \sum_n \left[s(n) - \sum_{k=1}^p \alpha_k \cdot s(n-k) \right]^2$$

- Valorile α_i se determina minimizand eroarea de predictie E_n :

$$\frac{\partial E}{\partial \alpha_i} = 0 \quad 1 \leq i \leq p$$

$$\frac{\partial E_n}{\partial \alpha_i} = 0, \quad i = 1, 2, \dots, p$$

- din care rezulta setul de ecuatii:

$$\sum_n s(n)s(n-i) = \sum_n \left[\sum_{k=1}^p \alpha_k s(n-k) \right] \cdot s(n-i)$$

- Daca definim termenii de scurta covarianta :

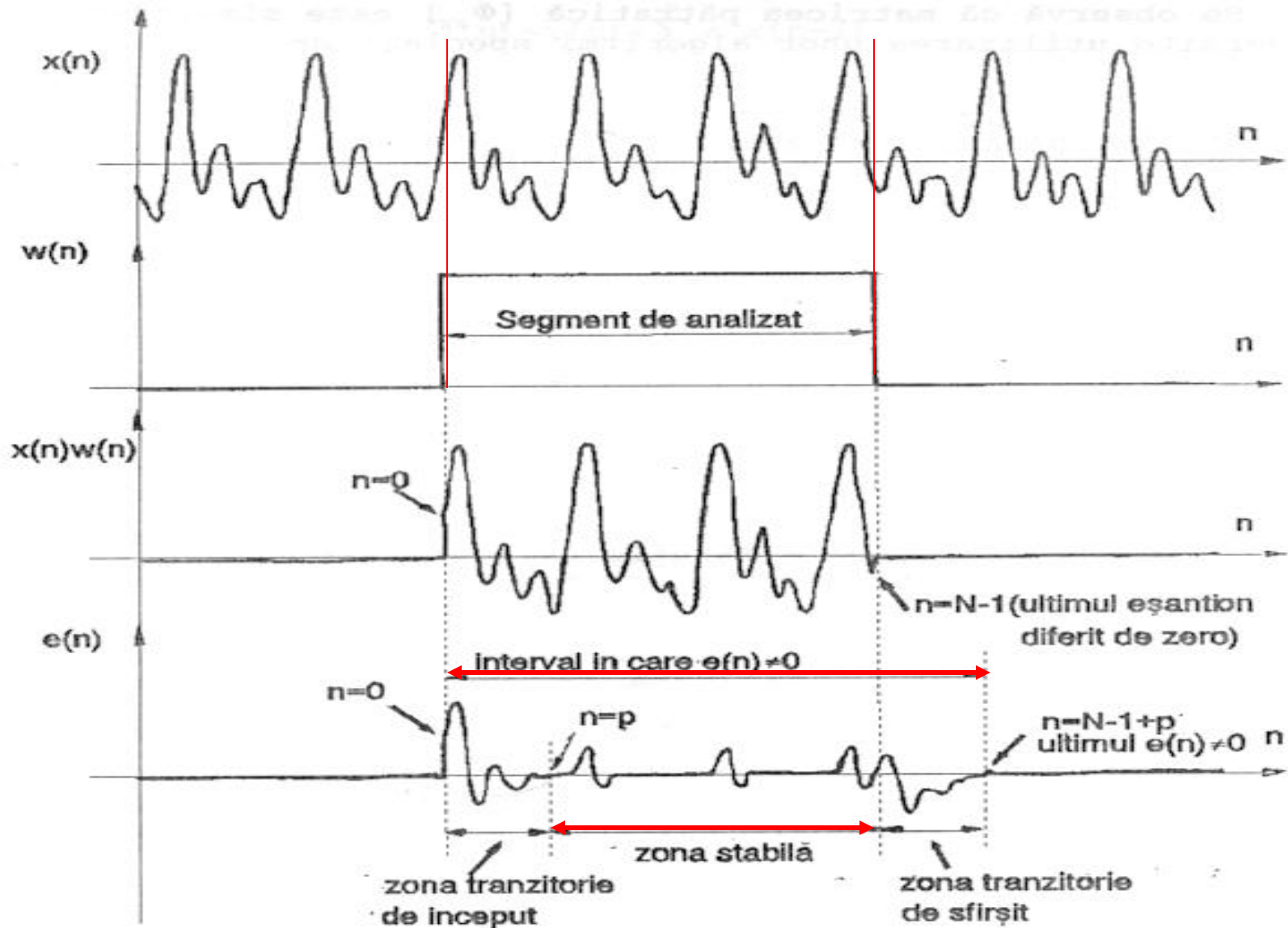
$$\phi_n(i, k) = \sum_n s(n-i)s(n-k)$$

- rezulta sistemul de ecuatii (p/p), care permite solutionare eficienta:

$$\sum_{k=1}^p \alpha_k \phi_n(i, k) = \phi_n(i, 0)$$

- eroarea medie patratica de predictie (MSE) pe timp scurt E_n :

$$E_n = \phi_n(0, 0) - \sum_{k=1}^p \alpha_k \phi_n(0, k)$$



Sunt 2 metode de calcul pentru eroare:

- Autocorelatia** - care constrangem semnalul $s_n(m) \neq 0$ in fereastra, $0 \leq m \leq N-1$
- Covarianta** - se aplica pe zona in care eroarea este stabila

1) Autocorelatia - care presupune semnalul $s_n(m) \neq 0$, $0 \leq m \leq N-1$

$$s_n(m) = \begin{cases} s(m+n).w(m), & 0 \leq m \leq N-1 \\ 0, & \text{inrest.} \end{cases}$$

, $w(m)$ fereastra de N esantioane.

- Semnalul $s_n(m) \neq 0$ pentru $0 \leq m \leq N-1$, iar eroarea este diferita de 0, pentru $0 \leq m \leq N-1+p$:

$$E = \sum_{n=0}^{N+p-1} e^2(n)$$

$$\phi_n(i, k) = \sum_{n=0}^{N+p-1} s(n-i)s(n-k)$$



$$\phi_n(i, k) = \sum_{n=0}^l s(n-i)s(n+i-k)$$

unde $l = N-1-(i-k)$

Datorită constrângerii de lungime finită, putem modifica ecuația ca:

$$\phi_n(i, k) = R_n(i-k)$$

, unde $R_n(i-k)$ este funcția de autocorelație ptr. $i-k$ care este funcție pară, deci:

$$\sum_{k=1}^p \alpha_k R_n(|i-k|) = R_n(i)$$

$$E = R_n(0) - \sum_{k=1}^p \alpha_k R_n(k)$$

Exprimata matricial, astfel:

$$\begin{bmatrix} R_n(0) & R_n(1) & \cdot & \cdot & R_n(p-1) \\ R_n(1) & R_n(0) & \cdot & \cdot & R_n(p-2) \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ R_n(p-1) & R_n(p-2) & \cdot & \cdot & R_n(0) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \cdot \\ \cdot \\ \alpha_p \end{bmatrix} = \begin{bmatrix} R_n(1) \\ R_n(2) \\ \cdot \\ \cdot \\ R_n(p) \end{bmatrix}$$

$$\Re \alpha = r$$

Cu solutia :

$$\alpha = \Re^{-1} r$$

Matricea $\mathbf{R}$ (pxp) este matrice Toeplitz si permite o metoda de rezolvare eficienta (ex. Algoritmul Levinson-Durbin).

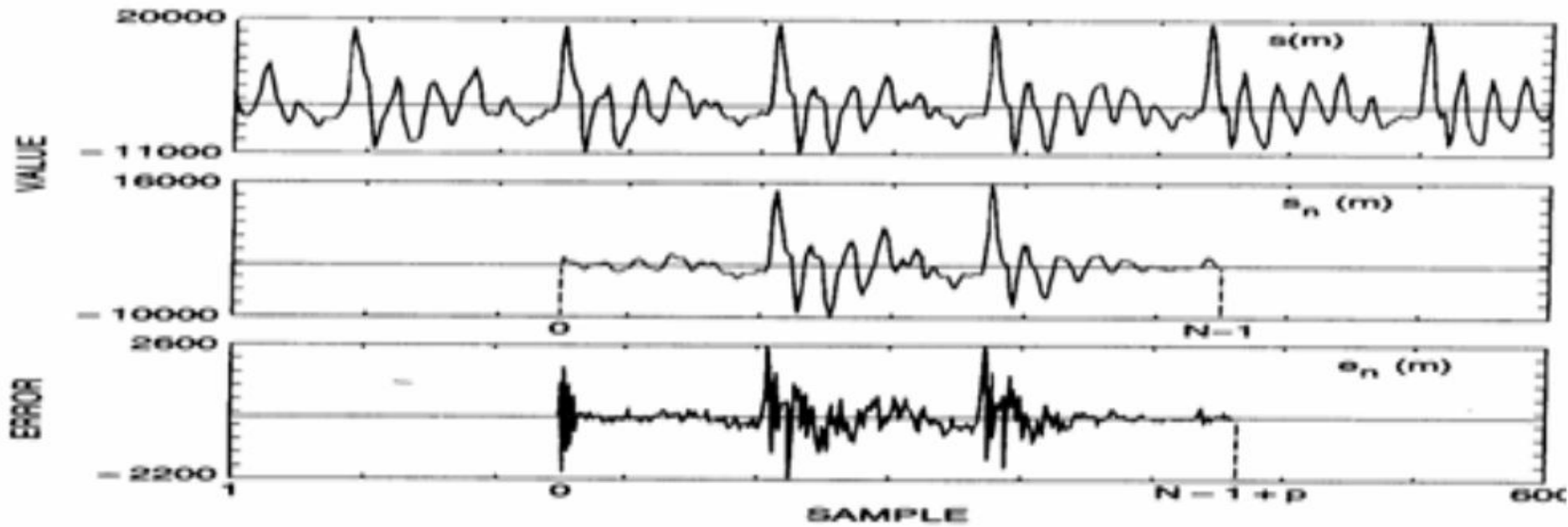


Figure 3.29 Illustration of speech sample, weighted speech section, and prediction error for voiced speech where the prediction error is large at the beginning of the section.

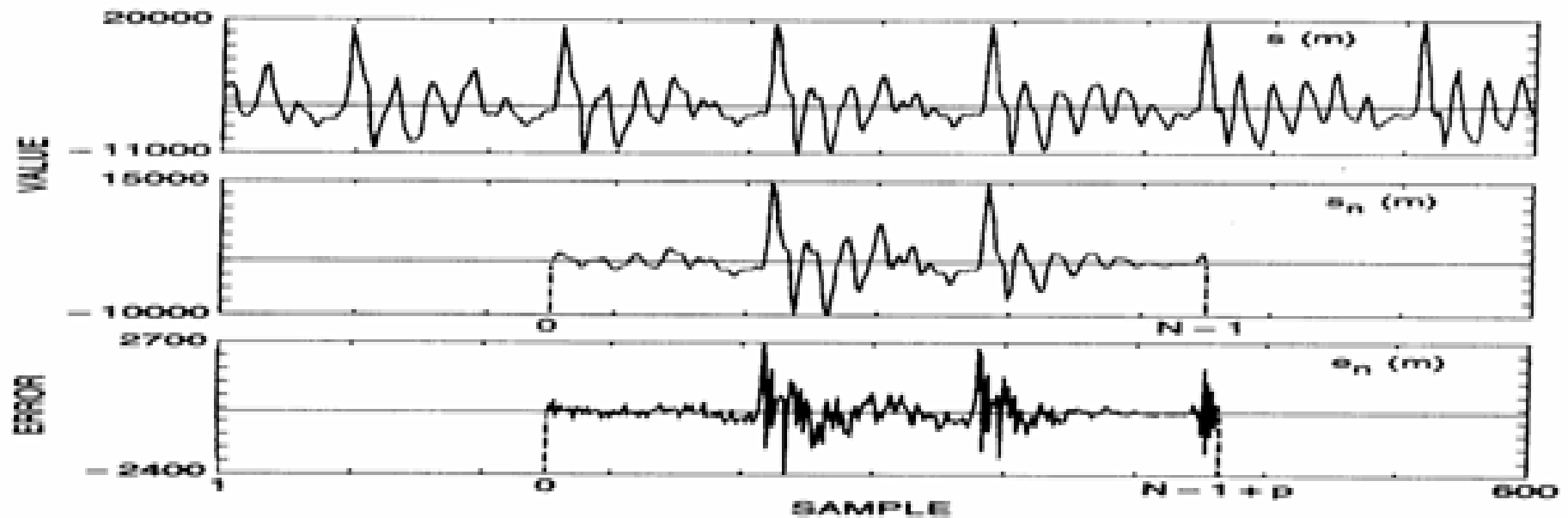


Figure 3.30 Illustration of speech sample, weighted speech section, and prediction error for voiced speech where the prediction error is large at the end of the section.

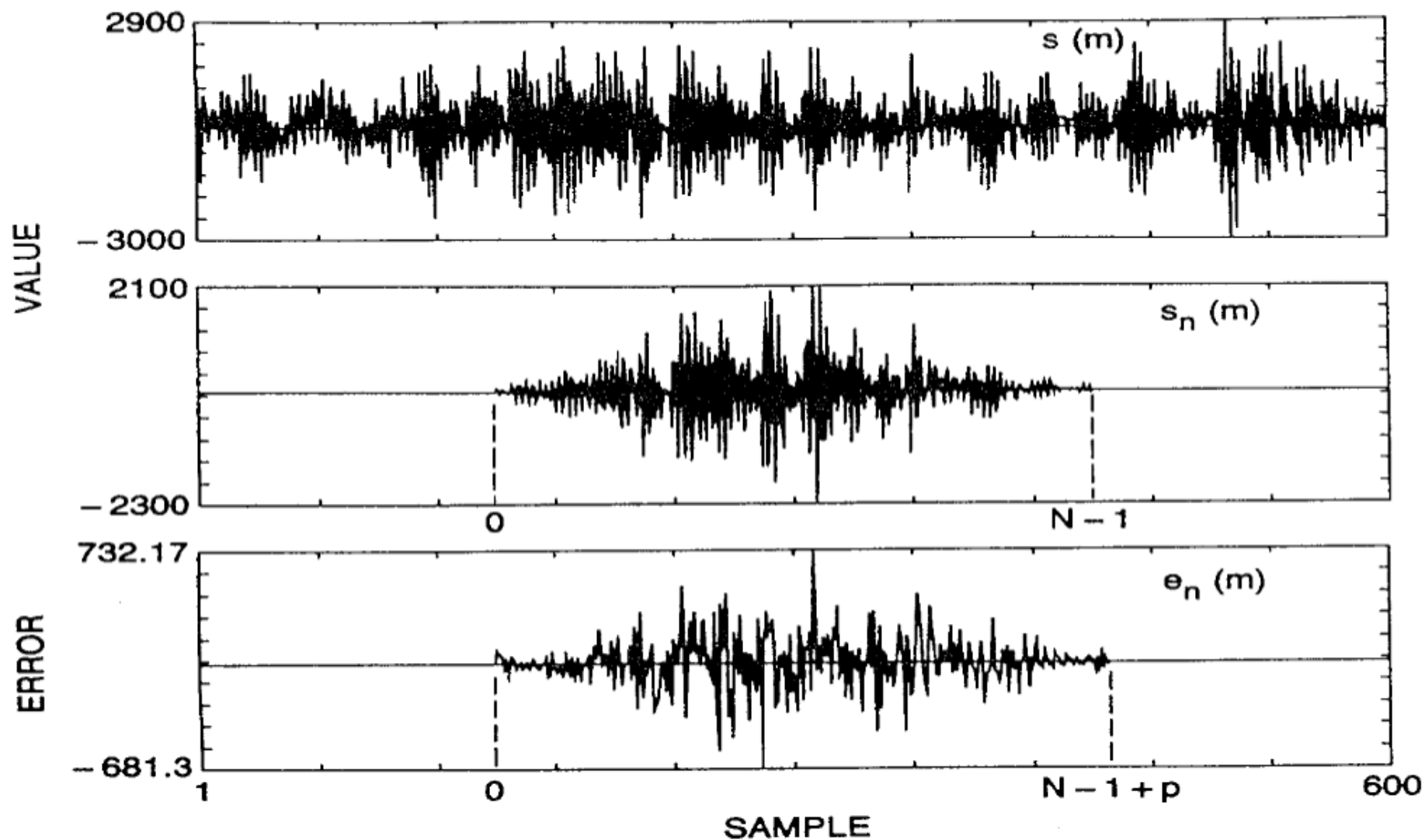
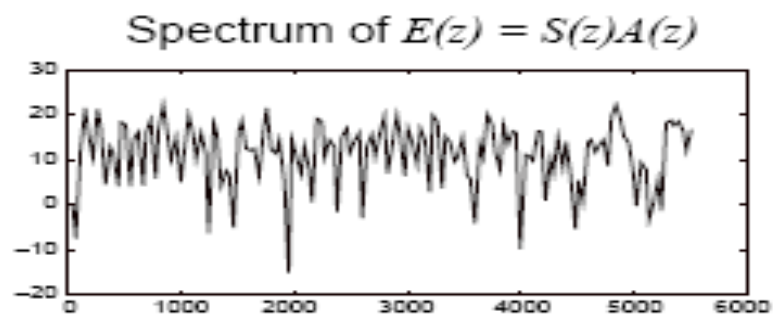
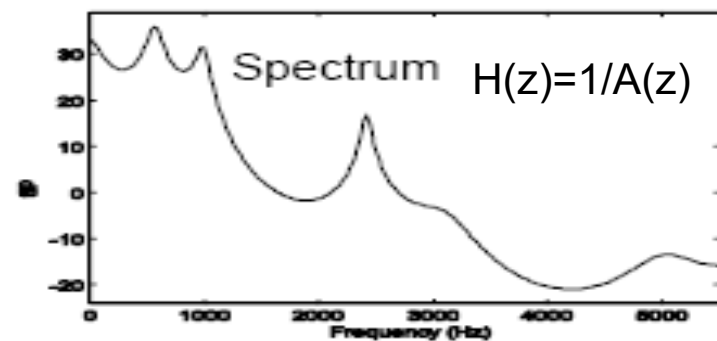
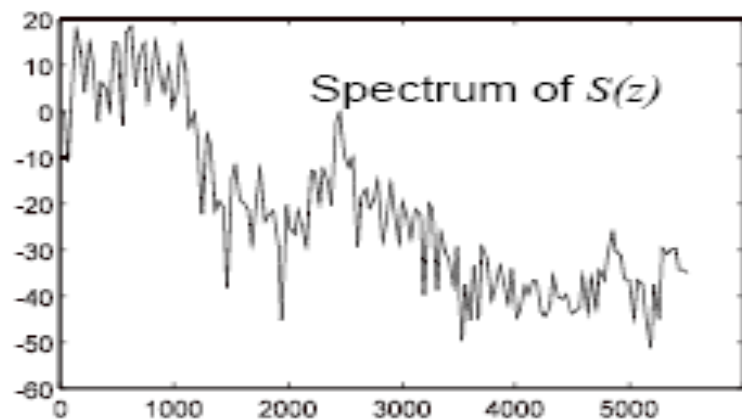
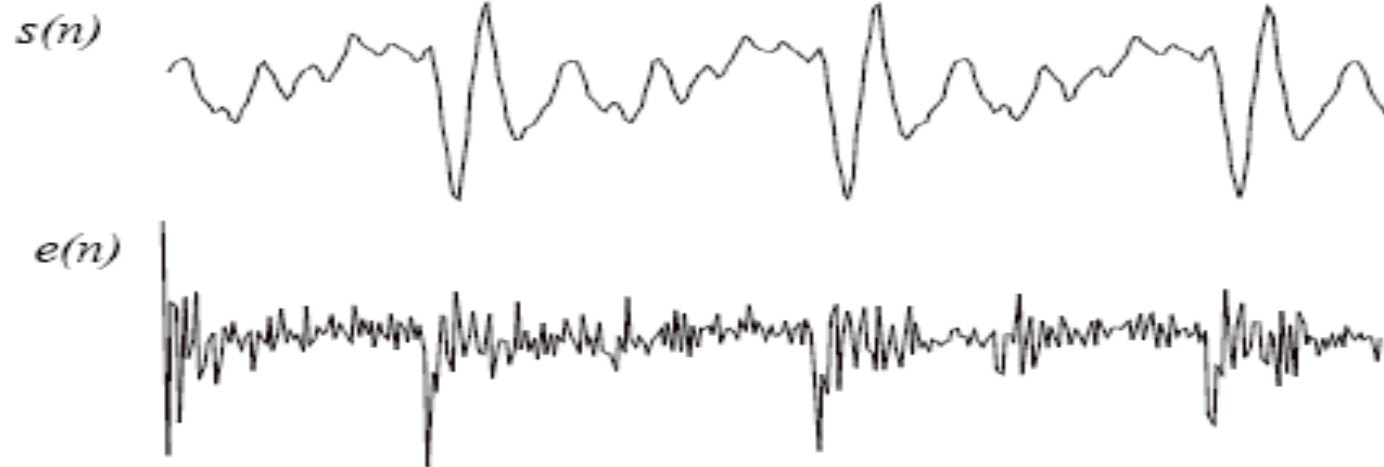
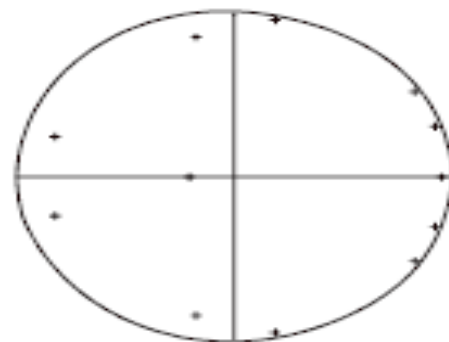


Figure 3.31 Illustration of speech sample, weighted speech section, and prediction error for unvoiced speech where there are almost no artifacts at the boundaries of the section.



Poles of $H(z)$



2. Metoda covariantei.

Fixarea intervalului pe care se calculează eroarea medie pătratică oferă o modalitate diferită de a rezolva ecuațiile predictorului liniar, astfel $\phi_n(i, k)$, pentru $1 \leq i \leq p$ și $0 \leq k \leq p$ poate fi scris :

$$\phi_n(i, k) = \sum_{n=0}^{N-1} s(n-i)s(n-k)$$

și schimbând indicii de sumare se simplifica

$$\phi_n(i, k) = \sum_{n=0}^{N-i-1} s(n)s(n+i-k)$$

$$\phi_n(i, k) = \sum_{n=0}^{N-k-1} s(n)s(n+k-i)$$

$$\phi_n(i, k) = \phi_n(k, i)$$

Folosind ecuațiile putem scrie ecuațiile predictorului sub forma matricială

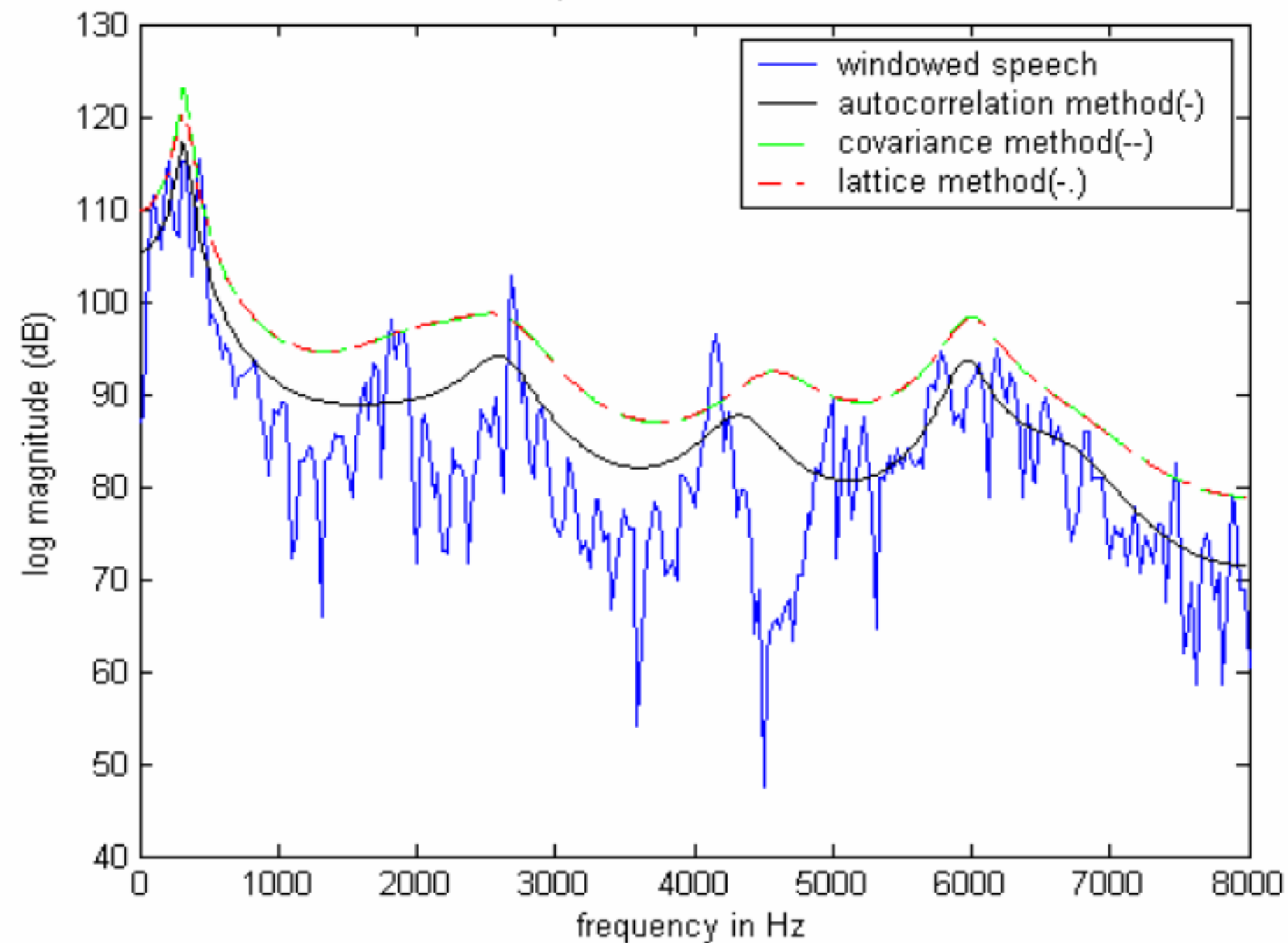
$$\sum_{k=1}^p \alpha_k \phi_n(i, k) = \phi_n(i, 0)$$

$$\bar{\phi} = \underline{\Phi} \cdot \bar{\alpha}$$

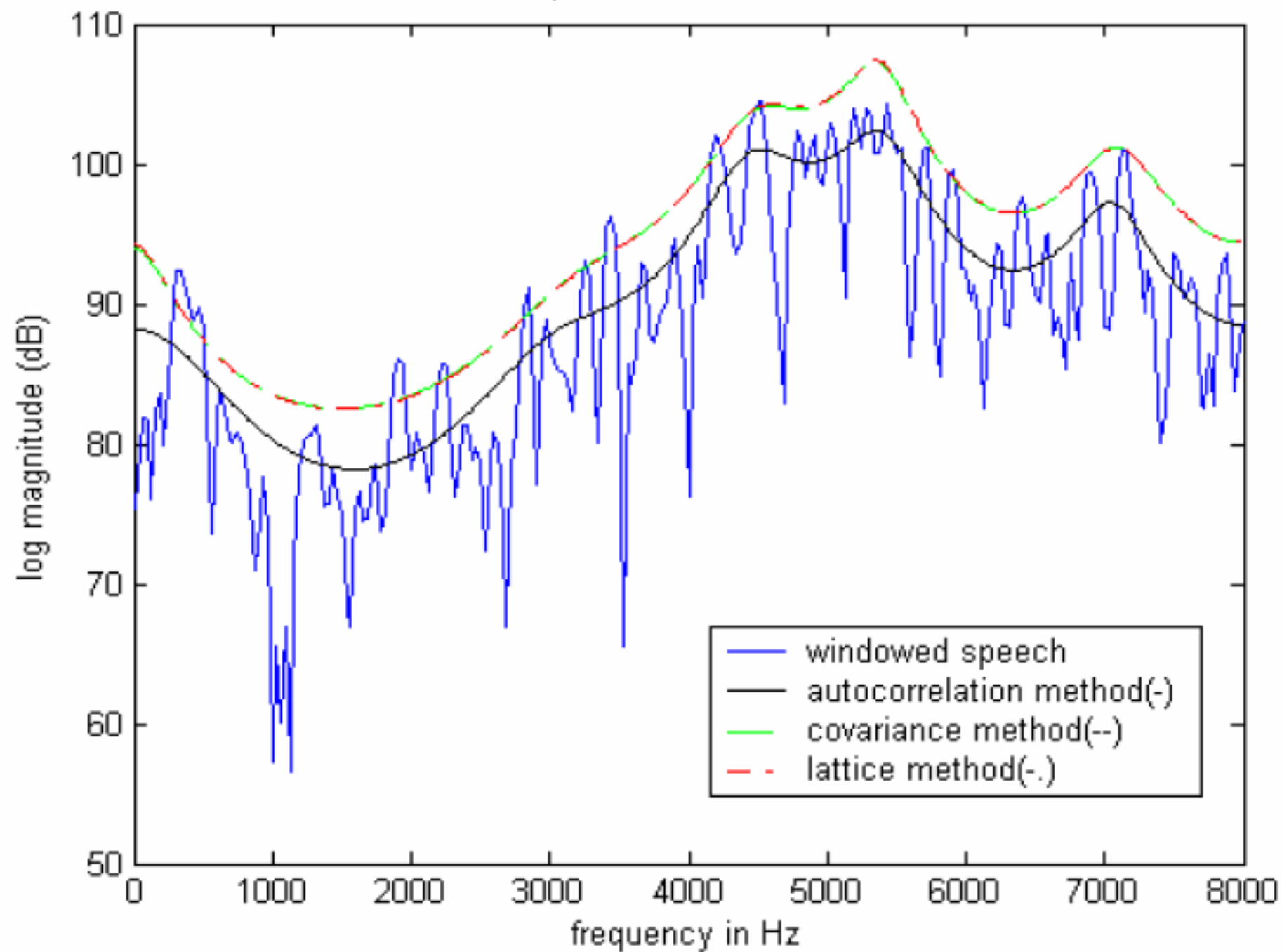
$$\begin{bmatrix} \phi_n(1,1) & \phi_n(1,2) & \phi_n(1,3) & \cdots & \phi_n(1,p) \\ \phi_n(2,1) & \phi_n(2,2) & \phi_n(2,3) & \cdots & \phi_n(2,p) \\ \phi_n(3,1) & \phi_n(3,2) & \phi_n(3,3) & \cdots & \phi_n(3,p) \\ \vdots & \vdots & \vdots & & \vdots \\ \phi_n(p,1) & \phi_n(p,2) & \phi_n(p,3) & \cdots & \phi_n(p,p) \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \vdots \\ \hat{a}_p \end{bmatrix} = \begin{bmatrix} \phi_n(1,0) \\ \phi_n(2,0) \\ \phi_n(3,0) \\ \vdots \\ \phi_n(p,0) \end{bmatrix}.$$

- Matrice simetrica, dar nu Toeplitz; => alte metode de rezolvare (descompunere Cholesky)
- Metodele covariantei si autocorelatiei utilizeaza 2 pasi ptr aflarea solutiei:
 1. Calculul matricei de valori de corelatie
 2. Gasirea eficienta a unei solutii pentru un set de ecuatii lineare
- S-a dezvoltat o alta clasa de metode de PL, numite *metode latice*, in care cei 2 pasi sunt combinati intr-un algoritm recursiv pentru determinarea parameterilor de PL (se porneste de la algoritmul Levinson-Durbin)

file: test_6k, ss: 1000 N: 400 p: 12



file: test₁6k, ss: 3000 N: 300 p: 12



Estimarea Castigului (G)

- se determina din egalitatea energiei semnalului si a esantioanelor prezise

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

Eroarea minima:

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

$$e(n) = Gu(n)$$

la potrivirea exacta

$$\alpha_k = a_k$$

- Intrucat potrivirea perfecta nu exista, se foloseste egalitatea dintre energia erorii de predictie si energia excitatiei.

-se presupune: *ptr segmente sonore*
segmente nesonore

$$u(n) = \delta(n)$$

$$u(n) = \text{zgomot alb, } (m=0, \text{ var}=1).$$

$$G^2 = R_n(0) - \sum_{k=1}^p \alpha_k R_n(k) = E_n$$

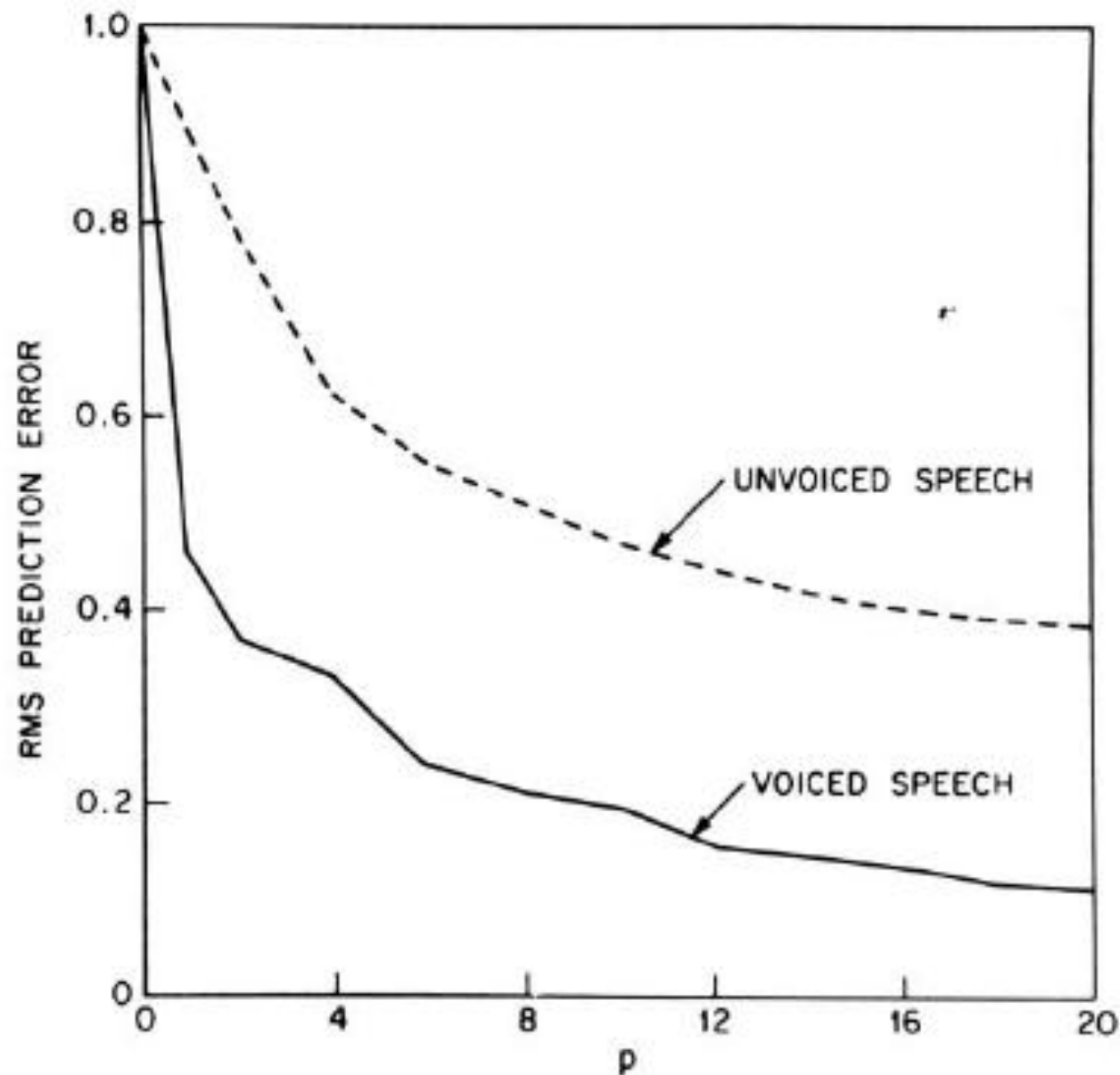


Fig. 8.4 Variation of the RMS prediction error with the number of predictor coefficients, p . (After Atal and Hanauer [3].)

Comparatii intre solutiile PL

- stabilitatea

- garantata ptr. Metoda autocorelatiei, lattice
- nu se poate garanta ptr. Metoda covariantei (dimensiunea ferestrei)

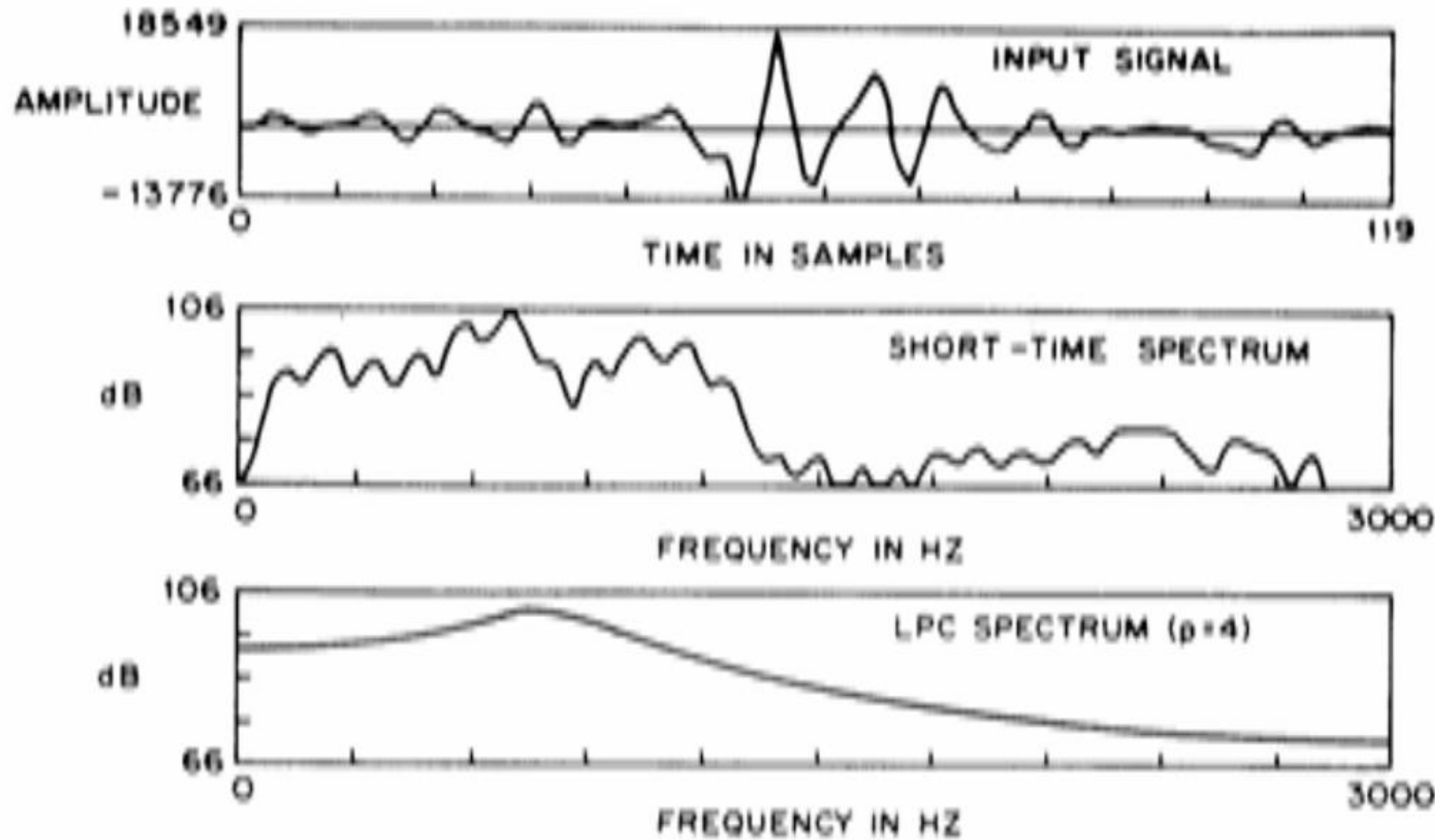
- alegerea parametrilor de analiza prin PL

- necesita 2 poli ptr. fiecare rezonanta (formant) a tractului sub $F_s/2$
- necesita 3-4 poli ptr. a reprezenta forma sursei si sarcina de radiatie
- uzual $p \approx 12-16$

Empiric: memoria predictorului (p) sa fie cel putin egala cu timpul parcurs de sunet pe dublul distantei tractului vocal

$$T_p = 2 L_{\text{tract}} / V_{\text{sunet}} = (2 * 0.17 \text{ m} / 340 \text{ m/s}) \sim 1 \text{ ms}$$

$$F_s = 8 \text{ kHz} \quad (T_s = 125 \mu\text{s}) \Rightarrow p = 8$$



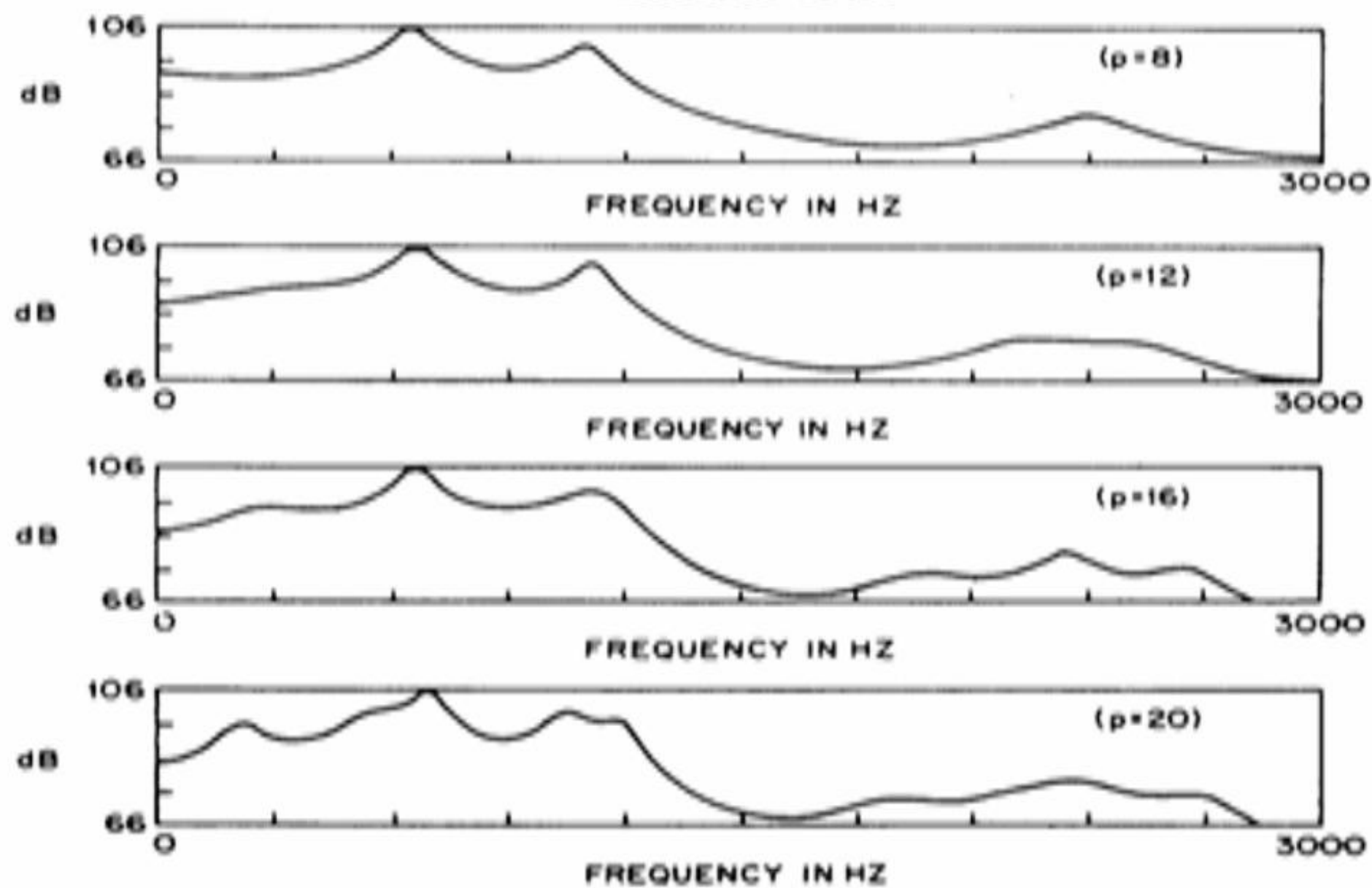
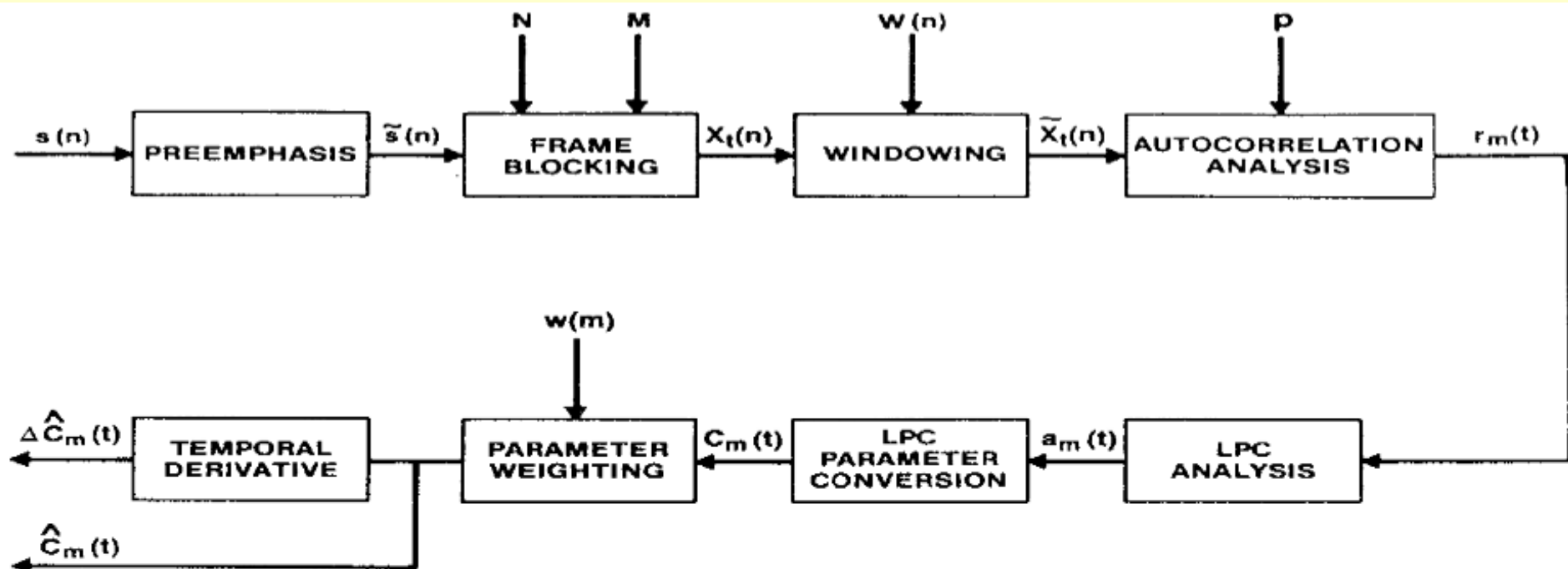
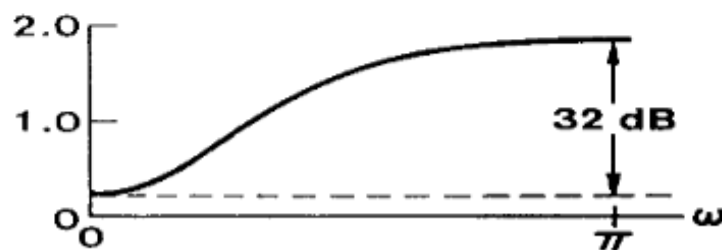


Figure 3.36 Spectra for a vowel sound for several values of predictor order, p .

3. Procesarea LPC pentru recunoașterea automata a SV



Block diagram of LPC processor for speech recognition.



Magnitude spectrum of LPC preemphasis network for $\tilde{a} = 0.95$.

Preemphasis - Preaccentuarea

- Fct. de transfer a glotei se poate modela astfel:

$$U_g(z) = \frac{1}{(1 - \mu_1 z^{-1})(1 - \mu_2 z^{-1})} \quad 0.9 \leq \mu_1, \mu_2 \leq 1.0$$

- Efectul de radiatie poate fi modelat ca:

$$R(z) = 1 - \mu_1 z^{-1}, \quad 0.9 \leq \mu \leq 1.0$$

- Prin urmare, pentru a obține funcția de transfer a tractului vocal celălalt pol trebuie să fie anulat, după cum urmează:.

$$H(z) = 1 - \mu_2 z^{-1}, \quad 0.9 \leq \mu \leq 1.0$$

- Semnalul achiziționat $s(n)$ este filtrat cu un FTS de ordinul I pentru a netezi spectrul semnalului și a ridica nivelul componentelor de frecvență mai ridicată la al celor de joasă frecvență.

$$H(z) = 1 - \tilde{a} z^{-1}, \quad 0.9 \leq a \leq 1.0.$$

$$\tilde{s}(n) = s(n) - \tilde{a} s(n-1).$$

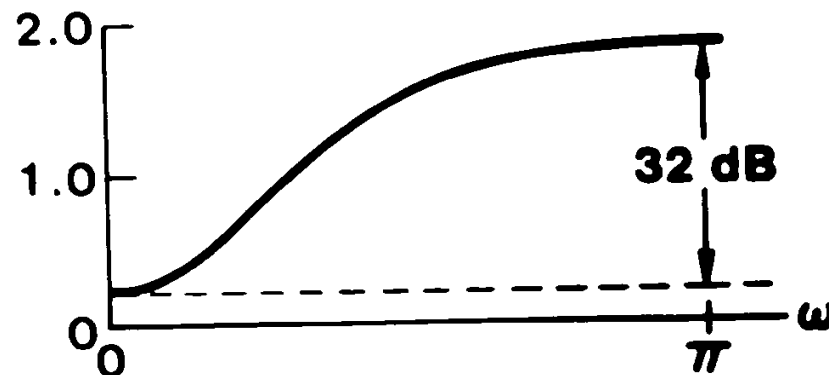


Figure 3.38 Magnitude spectrum of LPC preemphasis network for $\tilde{a} = 0.95$.

- **Cadrarea semnalului în blocuri** - Semnalul filtrat este împărțit în cadre de N eșantioane. Cadrele pot fi adiacente sau se pot suprapune parțial ($M \leq N$), cadrul L fiind notat cu :

$$x_{\ell}(n) = \tilde{s}(M\ell + n), \quad \begin{array}{l} n = 0, 1, \dots, N-1 \\ \ell = 0, 1, \dots, L-1. \end{array}$$

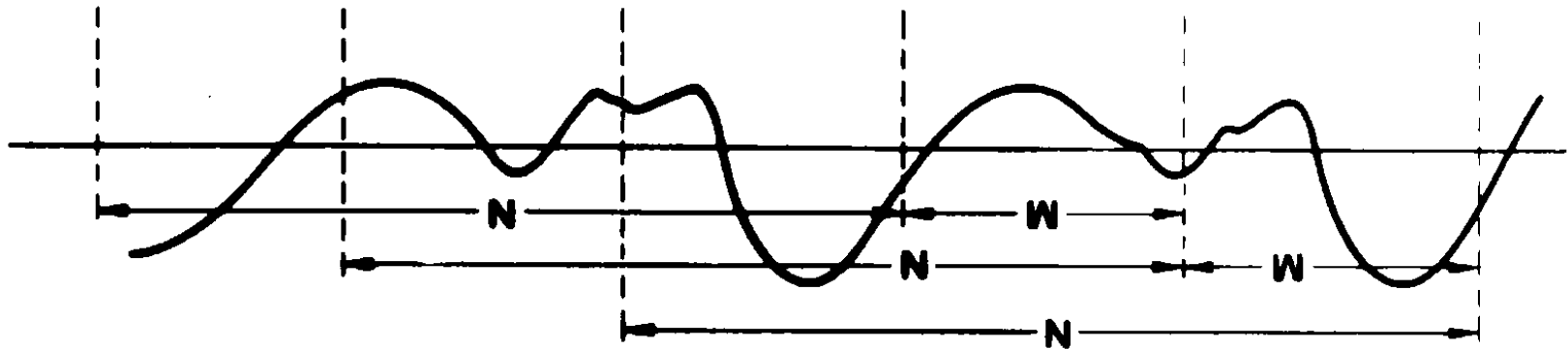
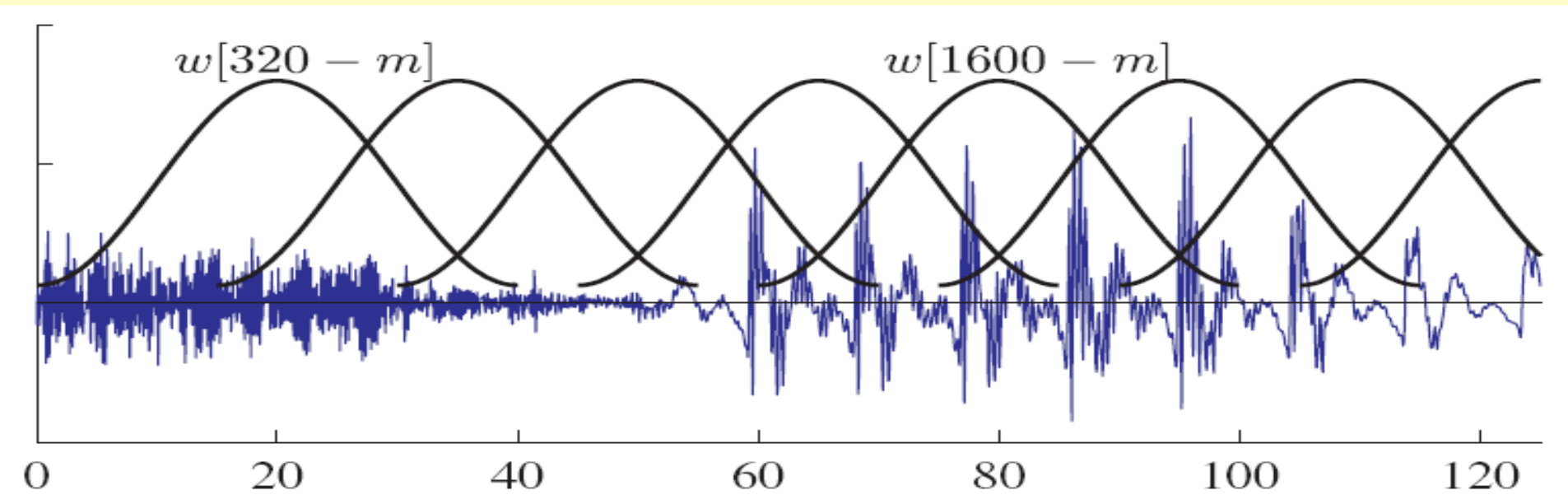


Figure 3.39 Blocking of speech into overlapping frames.

- **Ferestruirea** - Pentru reducerea discontinuităților la capetele cadrelor, acestea se ponderează folosind o funcție pondere/fereastră $w(n)$:

$$\tilde{x}_\ell(n) = x_\ell(n)w(n), \quad 0 \leq n \leq N-1.$$

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right), \quad 0 \leq n \leq N-1., \text{Hamming}$$



- **Analiza prin autocorelație** - Fiecare fereastră ponderată este apoi autocorelată pentru a obține :

$$r_{\ell}(m) = \sum_{n=0}^{N-1-m} \tilde{x}_{\ell}(n) \tilde{x}_{\ell}(n+m), \quad m = 0, 1, \dots, p,$$

- valoarea cea mai mare de autocorelație, p , este ordinul analizei LPC
- Tipic $p = 8 \dots 16$
- De remarcat faptul că funcția $r_{\ell}(0)$ reprezintă energia cadrului L .
- Energia cadrului este un parametru important în sistemele de detecție a vorbirii.

- **Analiza LPC** - care convertește fiecare cadru de $p+1$ autocorelații într-un set de parametri care pot fi :
 - coeficienți LPC
 - coeficienți de reflexie (PARCOR)
 - coeficienți log area ratio,
 - cepstrali
 - coeficienți derivați
- O metodă de conversie a coeficienților de autocorelație => parametri LPC (pentru metoda autocorelației) metoda Levinson - Durbin

$$E^{(0)} = R(0)$$

For $1 \leq i \leq p$ and $1 \leq j \leq i-1$

$$c_i = \left[R(i) - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} R(i-j) \right]$$

$$k_i = \frac{c_i}{E^{(i-1)}}$$

$$\alpha_i^i = k_i$$

$$\alpha_j^i = \alpha_j^{(i-1)} - k_i \cdot \alpha_{j-1}^{(i-1)}$$

$$E^{(i)} = (1 - k_i^2) \cdot E^{(i-1)}$$

$$a_m = \alpha_m^{(p)}, 1 \leq m \leq p, k_m, \text{ si } g_m = \log \frac{1 - k_m}{1 + k_m}$$

coeficienții LPC

coeficienții PARCOR

coeficienții log area ratio (LAR)

SPECTRUL netezit LPC

- Din coeficienții de predicție obținuți se poate determina spectrul LPC (netezit)

Notam:

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

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

• Acest polinom în z are răspunsul în domeniul timp $f(n)$

FdT a tractului vocal

$$f(n) = \begin{cases} 1, & n = 0 \\ a_n, & 1 \leq n \leq p \\ 0, & n > p \end{cases}$$

- Pentru a determina $A(j\omega)$ se poate folosi TFR care se aplică pe un număr de puncte ($N=256$ sau 512):

$$\text{FFT} (1, a_1, a_2, \dots, a_p, 0, \dots, 0)$$

Pentru a obține spectrul $H(j\omega)$ pentru $\omega=2\pi k/N$, $k=0, \dots, N-1$:

$$H(j\omega) = \frac{G}{A(e^{j\omega})}$$

$$\text{unde } G = \sqrt{\sum_{i=0}^p a_i r_1(i)} = \sqrt{r(0) \prod_{i=1}^p (1 - k_i^2)}$$

Conversia coeficienților de predicție în coeficienți cepstrali

- Cepstrul >>IFFT a logaritmului funcției de transfer, iar pentru un filtru numai poli avem :

$$\ln H(z) = \ln \frac{G}{1 + \sum_{i=1}^p a_i z^{-i}} = \sum_{n=1}^q c_n z^{-n}$$

- coeficienții LPCC, $c(m)$ se obțin diferențiind cele două părți ale expresiei în raport cu z^{-1} și rezultă formulele de recurență [Fur01]:

$$c_0 = \ln G^2 \quad G^2 \text{ este castigul modelului LPC}$$

$$c_m = a_m + \sum_{k=1}^{m-1} \left(\frac{k}{m} \right) c_k a_{m-k}, \quad 1 \leq m \leq p$$

$$c_m = \sum_{k=1}^{m-1} \left(\frac{k}{m} \right) c_k a_{m-k}, \quad m > p,$$

Ponderarea parametrilor LPCC

- Coeficientii LPCC de ordin inferior sunt sensibili la panta spectrală generală
- Coeficientii LPCC de ordin superior sunt sensibili la zgomot
- Ponderarea se face pentru a minimiza aceste sensibilitati

- valori tipice ale parametrilor folosiți în analiza LPC pentru recunoașterea vorbirii [Rab05]:

parametru	$f_e=6.67\text{KHz}$	$f_e=8\text{KHz}$	$f_e=10\text{kHz}$
N	300	240	300
M	100	80	100
p	8	10	12
q	12	12	12

N - numărul de eșantioane din cadrul de analiză

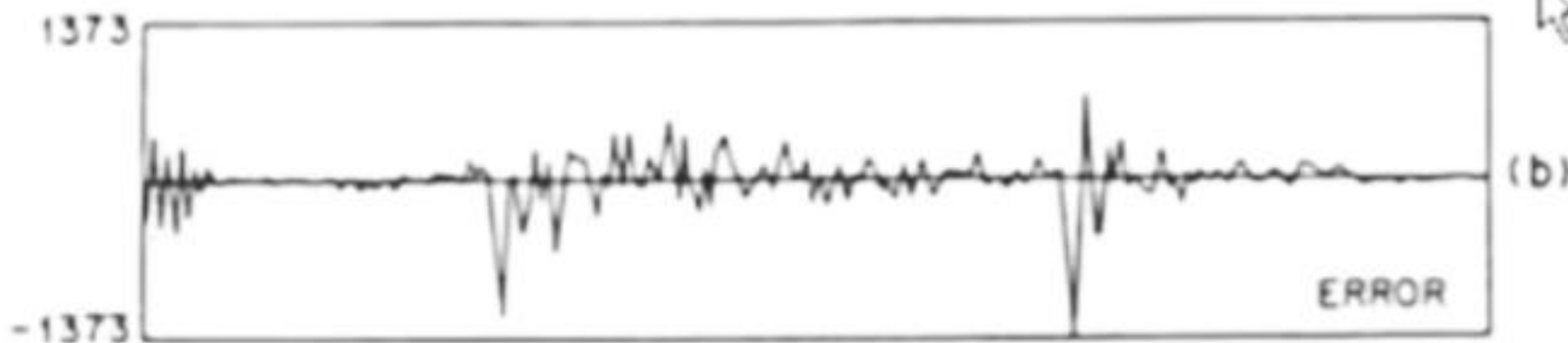
M - numărul de eșantioane de deplasare a cadrelor

p - ordinul analizei LPC

q- dimensiunea vectorului coeficienților cepstrali derivați din LPC (LPCC)

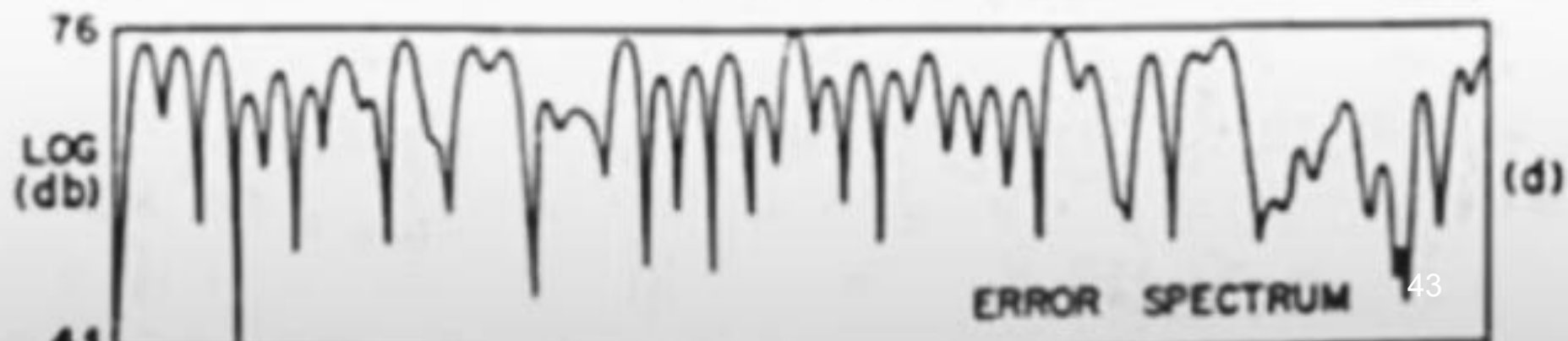
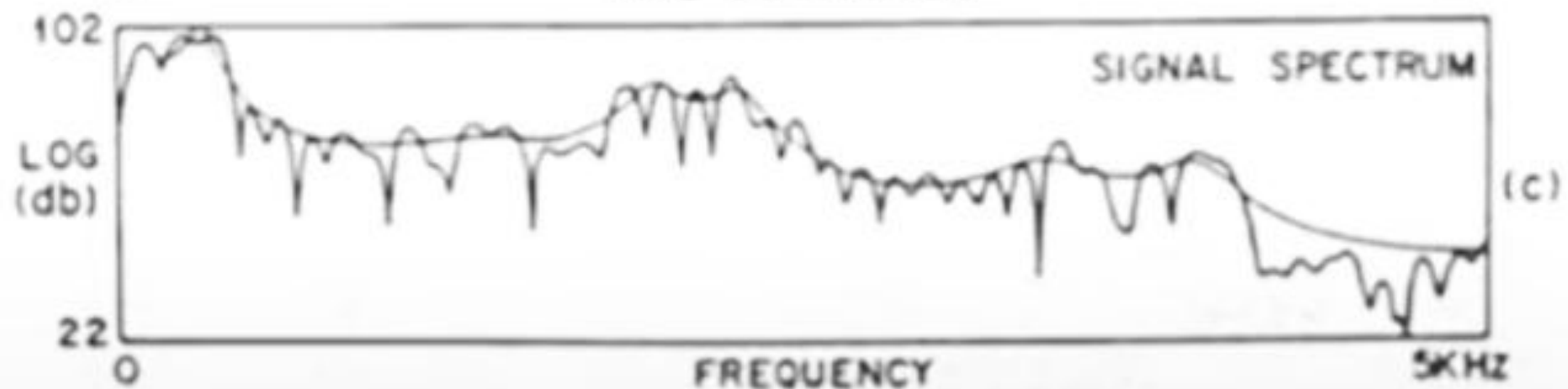
Obs. In general se utilizeaza o reprezentare cu $q > p$, unde $q \sim 1.5p$

TIME (SAMPLES)

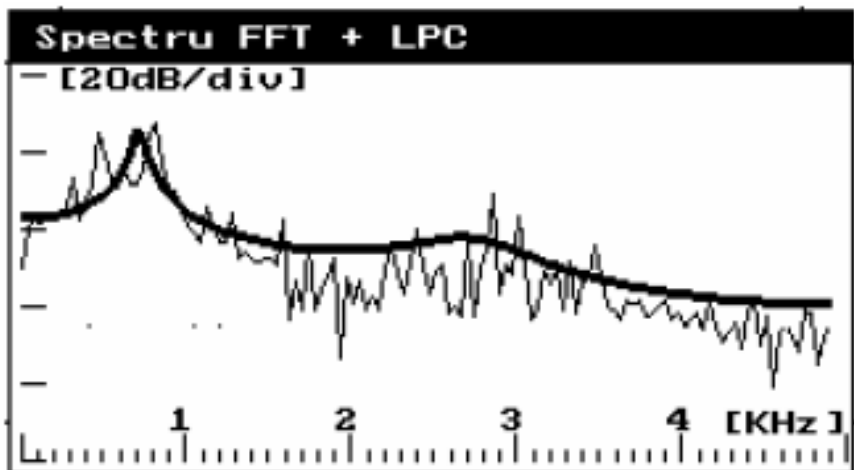


TIME (SAMPLES)

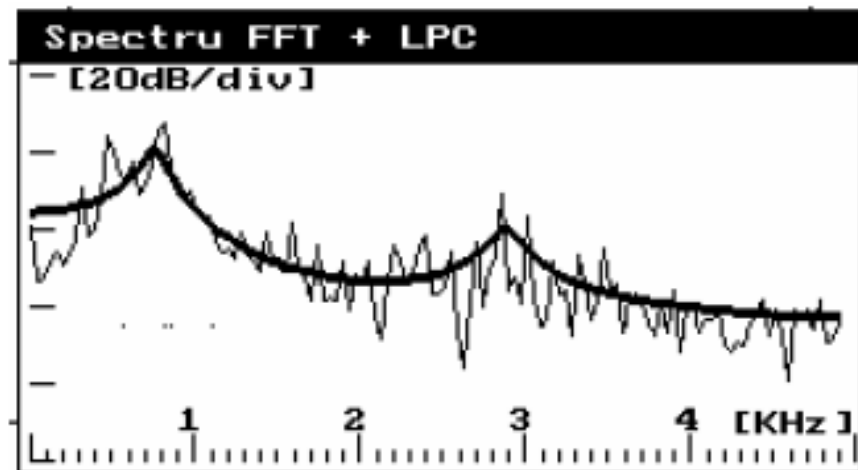
199



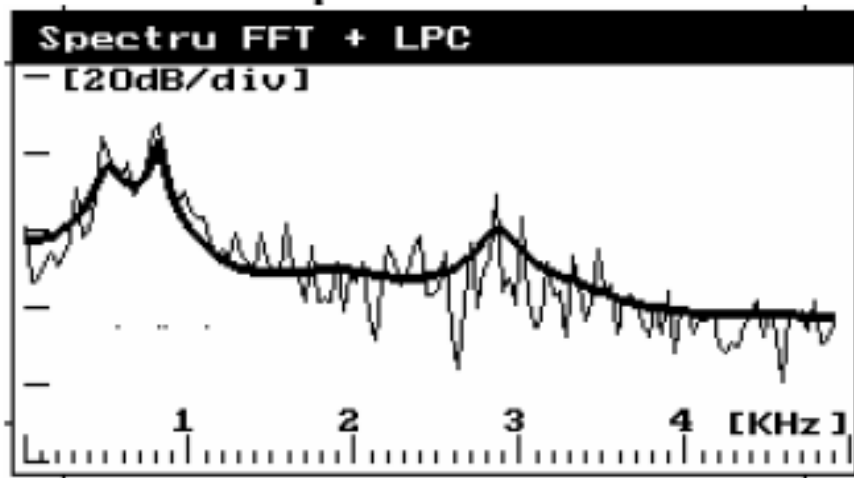
SPECTRE LPC



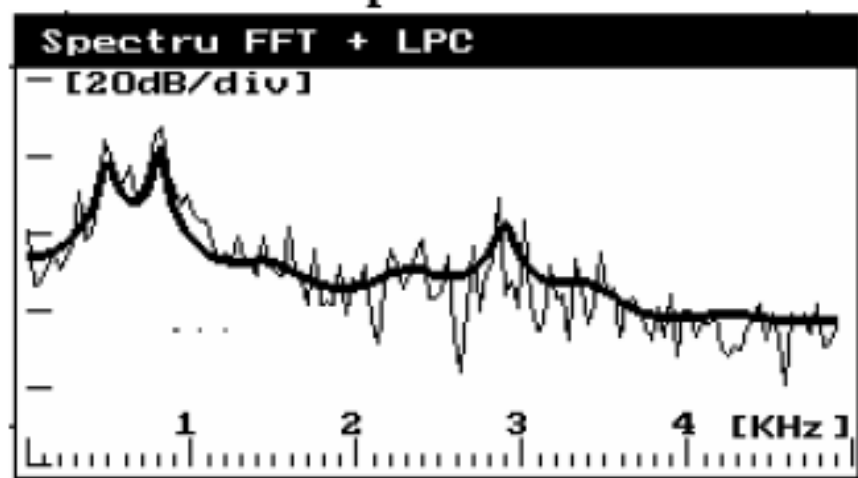
p=4



p=8



p=12

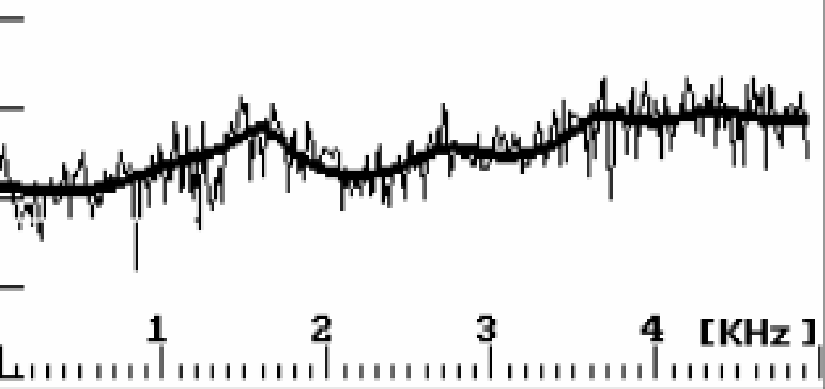


p=16

Spectrul log Fourier și LPC al vocalei "a" pentru p=4,8,12,16
folosind fereastră Hamming pe 256 de puncte

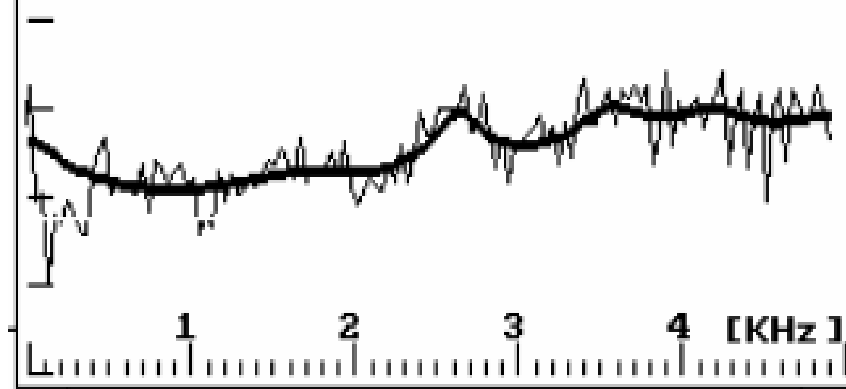
Spectru FFT + LPC

— [20dB/div]



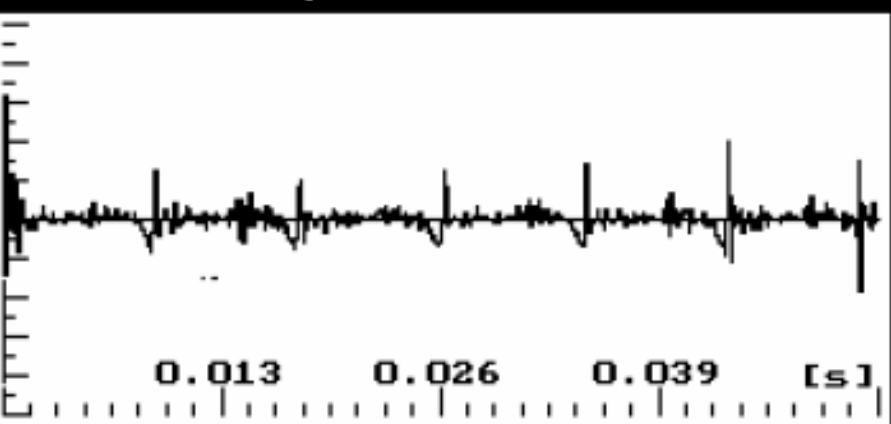
Spectru FFT + LPC

— [20dB/div]



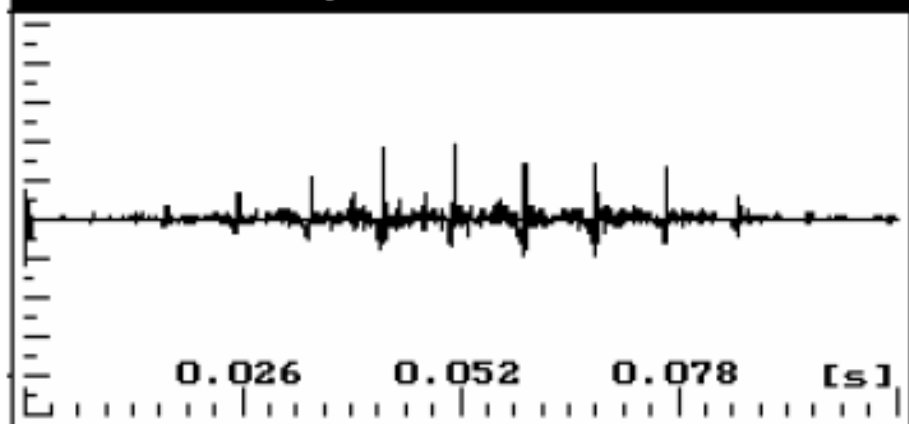
Spectrul log Fourier și LPC pentru consoanele "s" și "ș" ($p=12$)

Eroarea de predicție



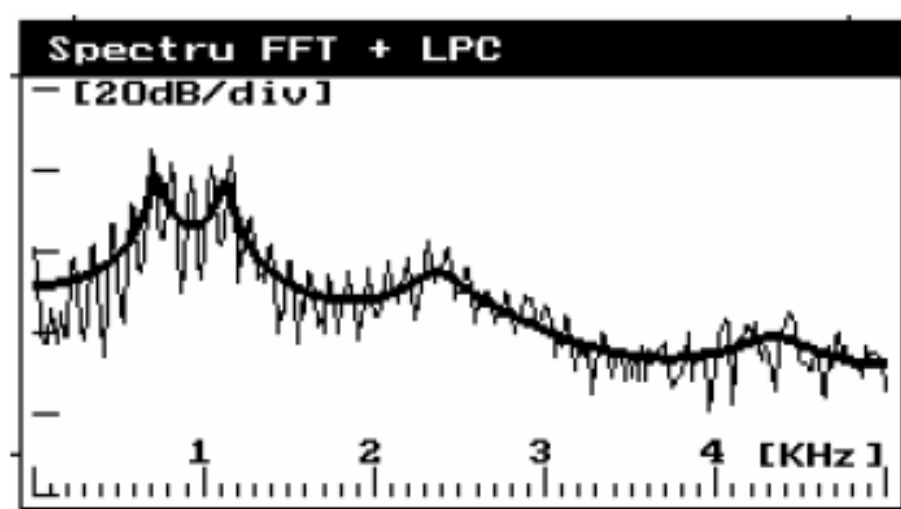
(a)

Eroarea de predicție

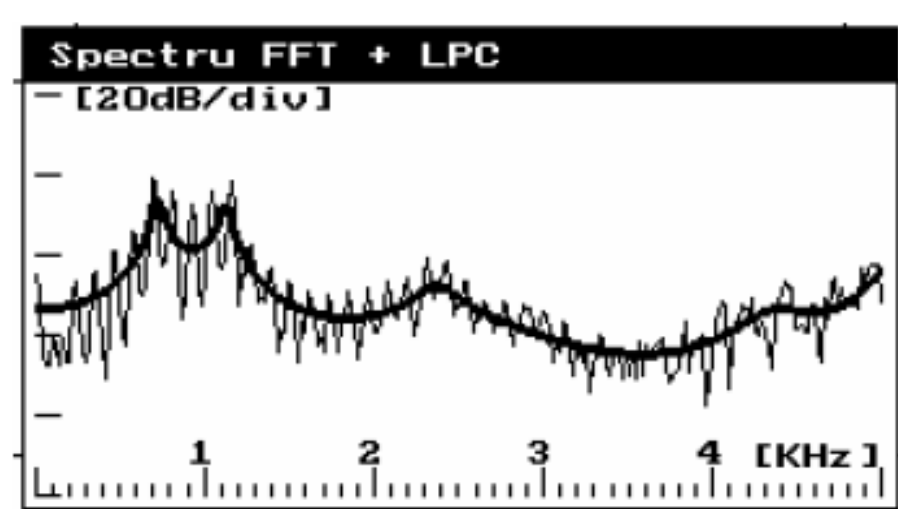


(b)

Eroarea de predicție pentru un segment sonor ("a"),
ferestruit cu fereastră dreptunghiulară (a) și Hamming (b) ($p=12$)

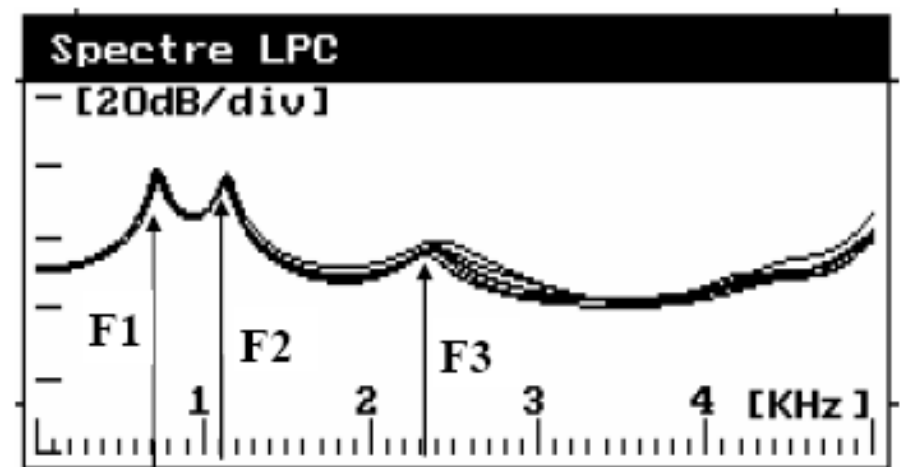


(a)

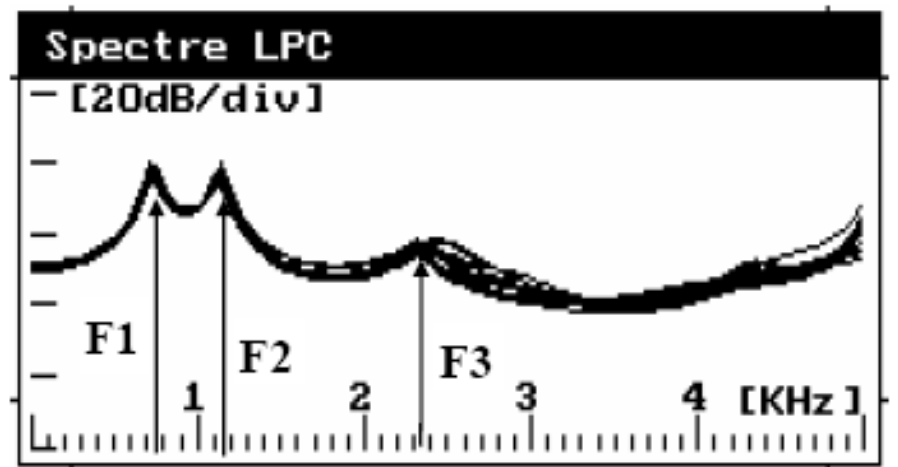


(b)

Spectrul log Fourier și LPC al vocalei "a" fără preaccentuare (a) și cu preaccentuare (b) (p=12)



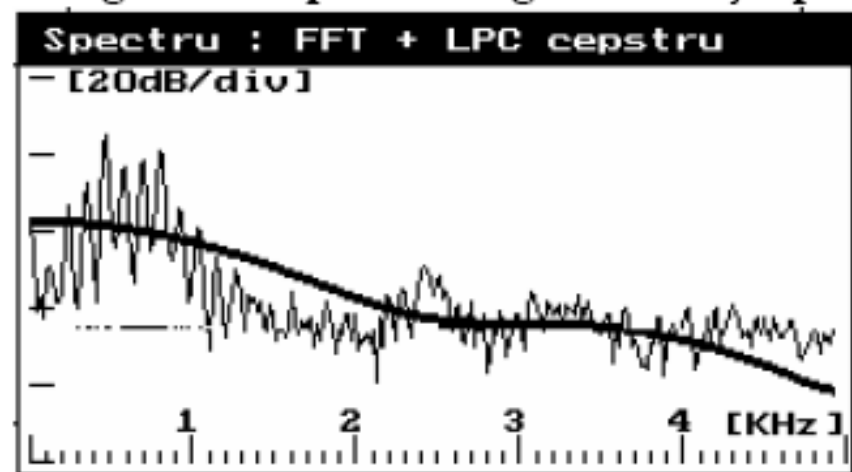
(a)



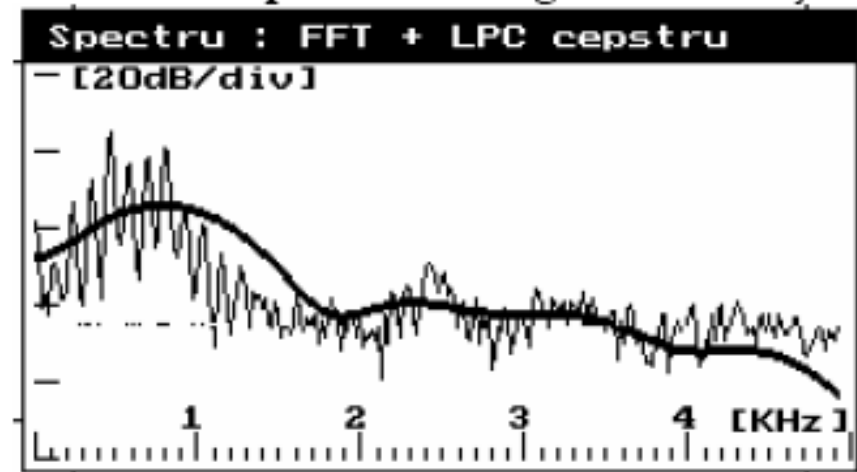
(b)

Evoluția spectrală LPC în ferestre succesive la rostirea vocalei "a" de către un locutor masculin : (a) pentru ferestre de 512 puncte și (b) pentru ferestre de 256 puncte

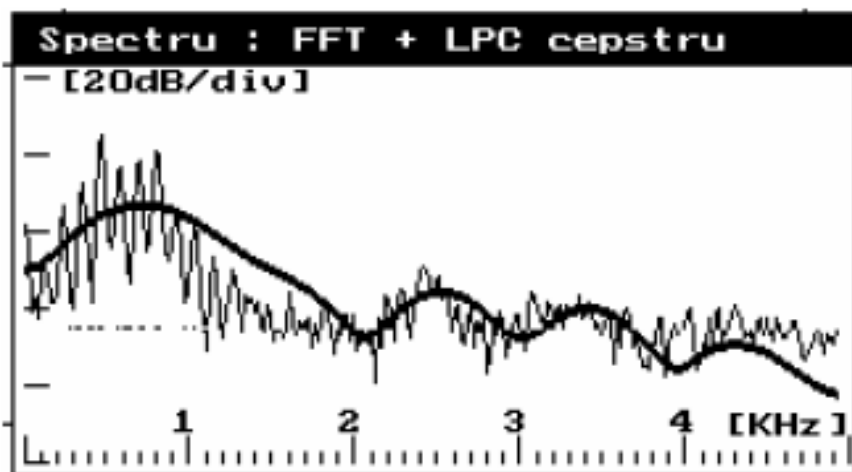
SPECTRE LPCC



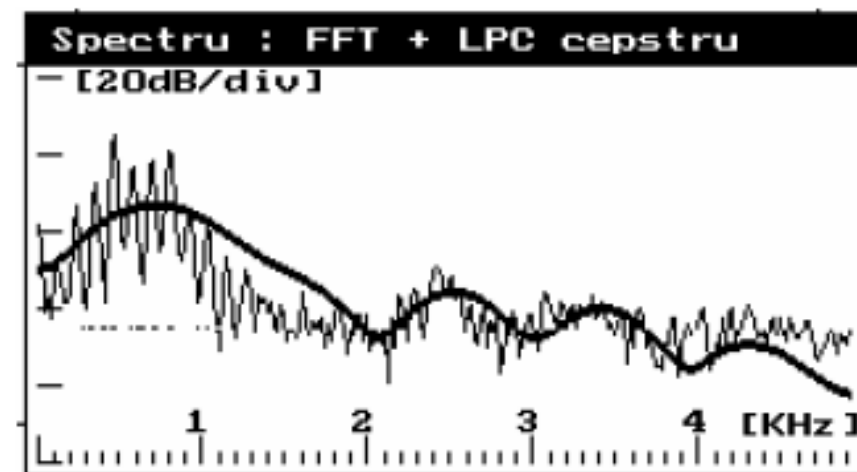
$q=4$



$q=8$



$q=12$



$q=16$

Spectrul log Fourier și spectrul cepstrului LPC al vocalei "o"
pentru $q=4,8,12,16$ ($p=12$)

Exemplu simplificat de calcul al coeficientilor LPC

Pentru simplificare, vom considera un semnal vocal foarte scurt, deja pre-procesat și ponderat (un cadru cu $N = 4$ eșantioane).

1. Semnalul de Intrare

Fie semnalul ponderat $s_w[n]$ pentru $n = 0, 1, 2, 3$:

$$s_w = \{s_w[0], s_w[1], s_w[2], s_w[3]\} = \{2, 4, 3, 1\}$$

2. Formula Autocorelației

Funcția de autocorelație pe termen scurt, $R[j]$, este definită ca:

$$R[j] = \sum_{n=0}^{N-1-j} s_w[n] s_w[n+j]$$

Pentru a calcula coeficienții de predicție de ordin p (de exemplu, $p = 2$), avem nevoie de coeficienții de autocorelație $R[0], R[1], \dots, R[p]$.

3. Calculul Coeficienților $R[j]$

a) Calculul $R[0]$ (întârziere $j = 0$)

$R[0]$ reprezintă **energia totală** a semnalului în cadrul de analiză:

$$R[0] = \sum_{n=0}^3 s_w[n]s_w[n+0] = s_w[0]^2 + s_w[1]^2 + s_w[2]^2 + s_w[3]^2$$

$$R[0] = (2)^2 + (4)^2 + (3)^2 + (1)^2 = 4 + 16 + 9 + 1 = \mathbf{30}$$

b) Calculul $R[1]$ (întârziere $j = 1$)

$R[1]$ măsoară corelația dintre eșantioanele alăturate:

$$R[1] = \sum_{n=0}^2 s_w[n]s_w[n+1] = s_w[0]s_w[1] + s_w[1]s_w[2] + s_w[2]s_w[3]$$

$$R[1] = (2 \cdot 4) + (4 \cdot 3) + (3 \cdot 1) = 8 + 12 + 3 = \mathbf{23}$$

c) Calculul $R[2]$ (întârziere $j = 2$)

$R[2]$ măsoară corelația dintre eșantioanele separate de un interval de timp de două unități:

$R[2]$ măsoară corelația dintre eșantioanele separate de un interval de timp de două unități:

$$R[2] = \sum_{n=0}^1 s_w[n]s_w[n+2] = s_w[0]s_w[2] + s_w[1]s_w[3]$$

$$R[2] = (2 \cdot 3) + (4 \cdot 1) = 6 + 4 = \mathbf{10}$$

d) Calculul $R[3]$ (întârziere $j = 3$)

$R[3]$ măsoară corelația dintre eșantioanele separate de un interval de timp de trei unități:

$$R[3] = \sum_{n=0}^0 s_w[n]s_w[n+3] = s_w[0]s_w[3]$$

$$R[3] = (2 \cdot 1) = \mathbf{2}$$

4. Sistemul de Ecuații (Ecuațiile Yule-Walker)

Coeficienții de autocorelație $\{R[0], R[1], R[2], \dots\}$ obținuți sunt folosiți pentru a construi **Ecuațiile Yule-Walker** (sau Ecuațiile Normale), care stau la baza găsirii coeficienților de predicție a_k .

Dacă alegem un **ordin de predicție** $p = 2$ (deci căutăm a_1 și a_2), sistemul devine:

$$\begin{pmatrix} R[0] & R[1] \\ R[1] & R[0] \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = - \begin{pmatrix} R[1] \\ R[2] \end{pmatrix}$$

Înlocuind valorile calculate:

$$\begin{pmatrix} 30 & 23 \\ 23 & 30 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = - \begin{pmatrix} 23 \\ 10 \end{pmatrix}$$

5. Rezolvarea pentru Coeficienții de Predicție a_k

Rezolvarea acestui sistem de ecuații liniare (de obicei prin **Algoritmul Levinson-Durbin** în aplicațiile reale) ne va da valorile numerice ale coeficienților de predicție a_1 și a_2 .

Interpretare: Acești coeficienți a_1 și a_2 (în acest exemplu simplificat) reprezintă modelul matematic al rezonanțelor tractului vocal pentru acel cadru de semnal.

TEMA 1.

Un sir de esantioane de SV are valorile $[s_0, s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8] = [1, 3, 2, 1, 4, 1, 2, 4, 3]$. Dimensiunea cadrului este de 4.

- Găsiți valorile preaccentuate ale esantioanelor dacă $\alpha = 0,98$.
- Găsiți parametrii de auto-corelație r_0, r_1, r_2 .
- Dacă vom folosi ord. 2 de predicție ($p=2$) pentru sistemul de extragere a caracteristicilor, găsiți coeficienții LPC: a_0, a_1, a_2 .
- În cazul în care numărul de esantioane care se suprapun pentru două cadre succesive este de 2, găsiți coeficienții LPC pe primul cadru.

TEMA 2.

Analizati exemplul alaturat.

Exercise 3.5

To illustrate LPC analysis via the autocorrelation method, consider a predictor of order $p = 2$. Assume an autocorrelation vector with components $R = (r(0), r(1), r(2))$. Use the Durbin method, described in the previous section, to solve for the LPC coefficients a_1 and a_2 in terms of the R s. Check your answer by solving the matrix equation

$$\begin{bmatrix} r(0) & r(1) \\ r(1) & r(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} r(1) \\ r(2) \end{bmatrix}$$

using simple matrix algebra.

Solution 3.5

Using the Durbin method, we get the following steps:

$$E^{(0)} = r(0)$$

$$k_1 = r(1)/r(0)$$

$$\alpha_1^{(1)} = r(1)/r(0)$$

$$E^{(1)} = (r^2(0) - r^2(1))/r(0)$$

$$k_2 = (r(2)r(0) - r^2(1))/(r^2(0) - r^2(1))$$

$$\alpha_2^{(2)} = (r(2)r(0) - r^2(1))/(r^2(0) - r^2(1))$$

$$\alpha_1^{(2)} = (r(1)r(0) - r(1)r(2))/(r^2(0) - r^2(1))$$

$$a_1 = \alpha_1^{(2)}$$

$$a_2 = \alpha_2^{(2)}$$

Using matrix algebra we get

$$a_1 r(0) + a_2 r(1) = r(1)$$

$$a_1 r(1) + a_2 r(0) = r(2)$$

Solving directly for a_1 and a_2 we get

$$a_1 = (r(1)r(0) - r(1)r(2))/(r^2(0) - r^2(1))$$

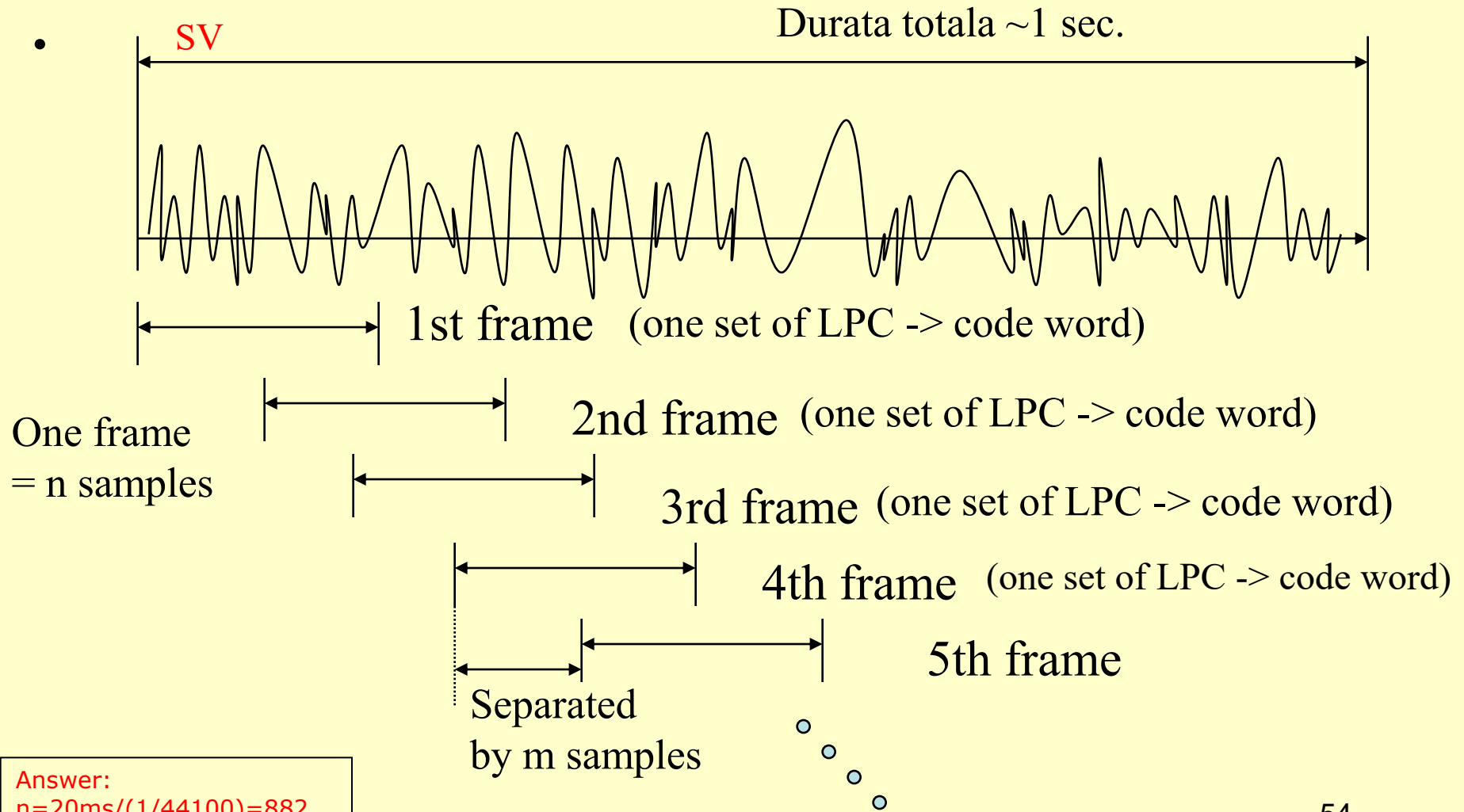
$$a_2 = (r(2)r(0) - r^2(1))/(r^2(0) - r^2(1))$$

which is the same result as obtained via the Durbin method.

Exemplu de analiza a semnalului vocal

Dimensiunea cadrului este 20ms, suprapuse cate 10 ms, SV este esantionat cu 44.1 KHz

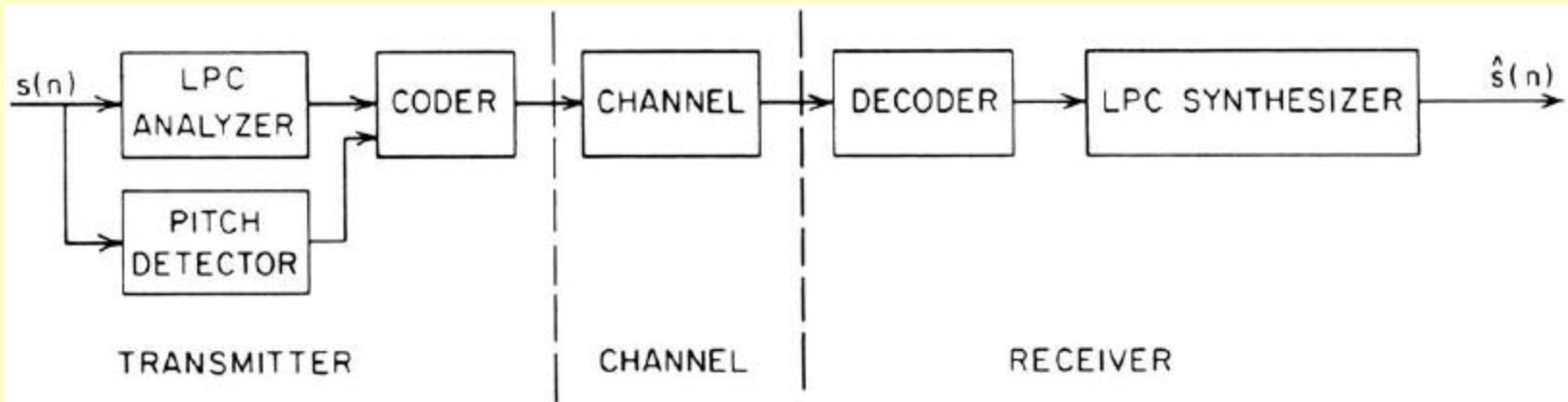
Calculati: n si m.



Answer:
 $n = 20\text{ms} / (1/44100) = 882$,
 $m = 10\text{ms} / (1/44100) = 441$

Exercises on LPC

1. Calculate LPC parameters of Speech
 - Use *LPC* function and perform pre-emphasis
2. Determination of Formants using LPC
3. Formant Tracking Using LPC
4. Inverse Filtering to determine $e(n)$
5. Determination of Pitch Period from $e(n)$
6. Pitch Period Tracking



Vocoder LPC

Referinte.

Curs: L.Rabiner **Linear Predictive Coding (LPC)**-Introduction ECE 259

<http://engineering.purdue.edu/VISE/ee438L/>

<http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html>

<http://www.mathworks.com/products/signal/demos.html?file=/products/demos/shipping/signal/lpcdemo.html>

<http://vlab.amrita.edu/?sub=3&brch=164&sim=1012&cnt=2>