

ASRSV – curs 11

Metoda codării TESPAP

- **Sisteme și metode utilizate în recunoaștere**

- Abordările de recunoaștere a vorbirii/vorbitorului în raport cu metodele și tehnicile folosite se pot grupa în mai multe familii :

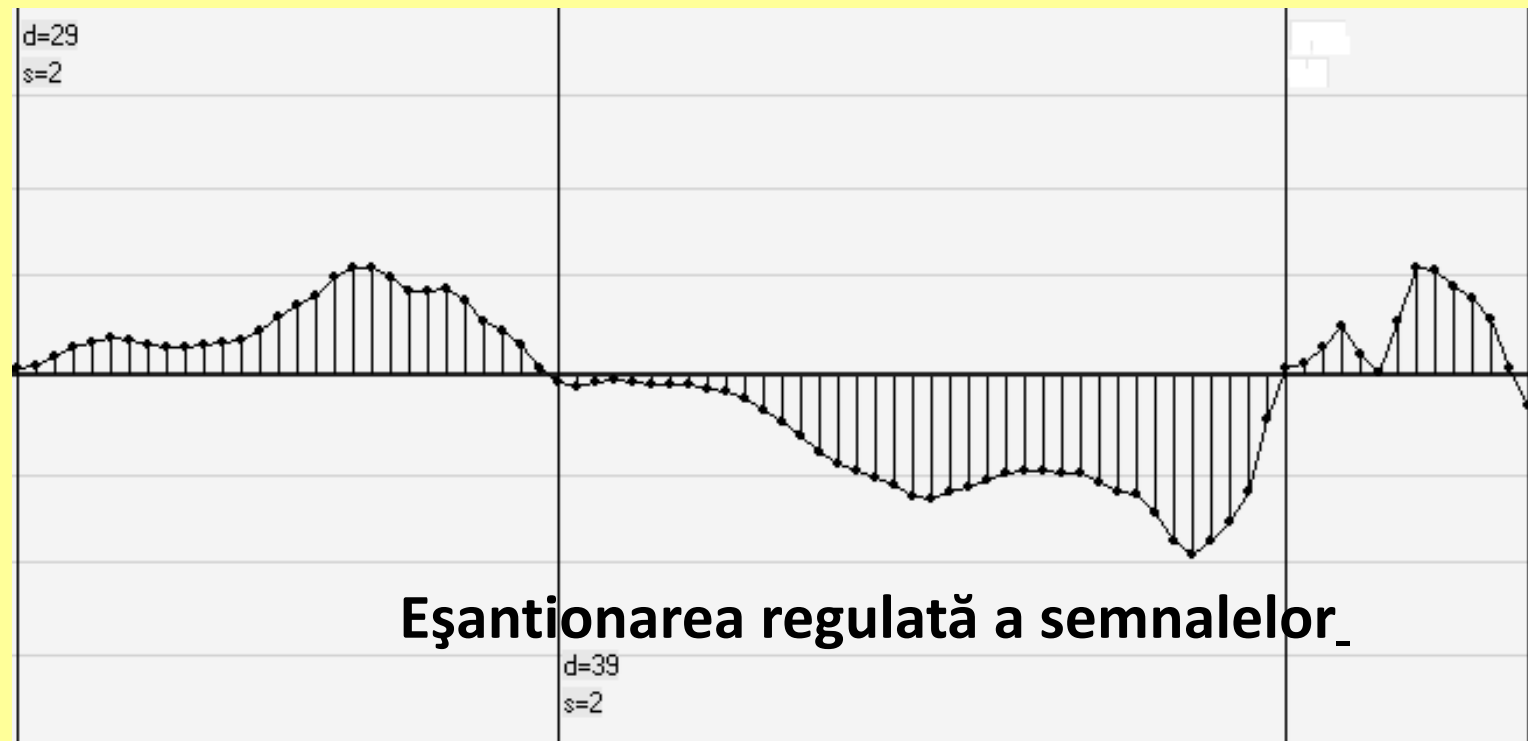
- sisteme bazate pe spectrograme
- sisteme bazate pe metodele programării dinamice (DTW)
- sisteme care folosesc cuantizarea vectorială (VQ)
- sisteme bazate pe modele Markov ascunse (MMA) - HMM
- sisteme bazate pe mixturi Gaussiene GMM
- sisteme utilizând rețelele neuronale NN
- **sisteme bazate pe metoda TESPAN FANN**
- sisteme folosind metode algebrice/statistice
- combinații ale lor.

C. E. Shannon

„A Mathematical Theory Of Communication”, 1948

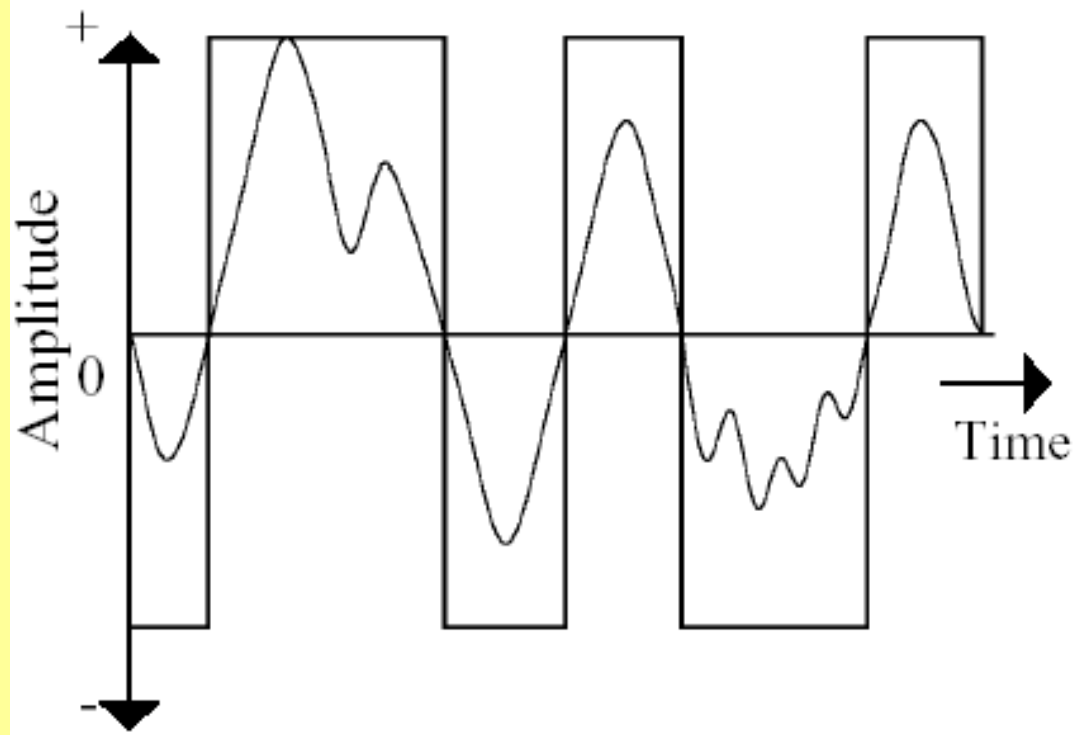
„Communication In The Presence Of Noise” 1949

- *Dacă o funcție $f(t)$ nu conține frecvențe mai mari de W cicli/sec (banda), atunci ea este complet determinată de valorile sale aflate la intervale de $1/2W$ secunde una față de cealaltă.*



- Datele obținute prin eșantionare sunt mai departe folosite ca bază pentru diferite transformări (în general liniare) ca:
 - Fourier,
 - Wavelet
 - codoare Walsh
 - pentru a extrage parametrii cheie în analiza semnalelor.
- Metodele uzuale de codare implică 3 cerințe:
 - *utilizarea eșantionării regulate*
 - *utilizarea unor descriptori de amplitudine*
 - *un domeniu de aproximare pt. amplitudine (cuantizare)*

- C. E. Shannon
- Liclidder și Pollack
- Preston C. Hammer
- Bond & Cahn



Infinite clipping

- Liclidder și Pollack - efectele lipsei informației de amplitudine în ineligibilitatea vorbirii;
- excluderea informației legate de amplitudine a păstrat inteligibilitatea în proporție de peste 97%;
- Ceea ce demonstrează că o parte esențială a informației transmisă vocal (inteligibilitatea) este conținută în **trecherile prin zero** ale formei de undă sonoră

- Aceste descoperiri pun sub semnul întrebării necesitatea celor trei cerințe exprimate mai sus
- Păstrând doar informația legată de trecerile prin zero și neglijând informațiile legate de amplitudine, obținem valori ce nu sunt egal depărtate în domeniul timp, dar care derivă din semnal – *este clar că în acest caz eșantionarea nu mai este regulată*
- Contradicțiile aparente cu teorema eșantionării a lui Shannon care par atât de evidente dispar la o privire mai atentă asupra lucrării care face prima dată publică această teoremă [Shannon]:
 - „*Cele $2TW$ numere folosite pentru descrierea funcției, nu trebuie să fie neapărat cele $2TW$ numere egal distanțate în timp*” .
 - De ex. eșantioanele pot fi inegal distanțate”, unde T este durata semnalului și W banda lui
 - „*Generalizând, orice set de $2TW$ numere independente asociate cu o funcție pot fi utilizate pentru a o descrie*”
- încurajează studiile în aproximarea semnalelor de bandă limitată utilizând zerourile

- Ultima cerință legată de domeniul de aproximare ridică totuși unele întrebări importante cum ar fi:

1. Este informația conținută în setul de date determinat de trecerile prin zero, inegal spațiate în timp, necesară dar nu și suficientă pentru o aproximare suficient de bună a semnalului vocal?

2. Mai general, poartă zerourile informații suficiente pentru analiza formelor de undă în general?

- „*Approximation spaces*”, Preston C. Hammer oferă destule răspunsuri și întrebări despre aproximări, care împreună cu ideile lui Shannon, Pollack și Licklider să încurajeze continuarea studiilor în aproximarea semnalelor utilizând zerourile.

Analiza semnalelor pe baza zerourilor

- Teoria de analiză a semnalelor pe baza zerourilor este începută de Bond și Cahn (1958)
- introducerea conceptului de zerouri complexe

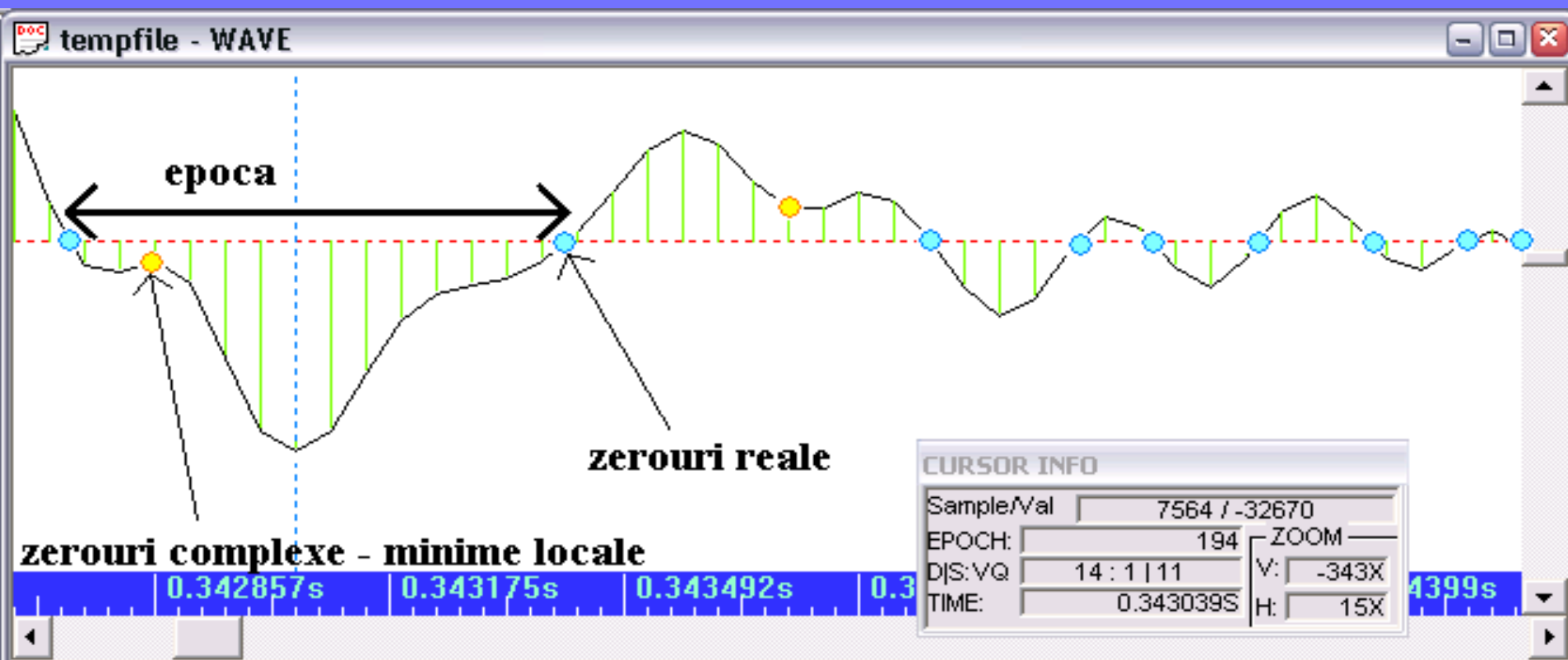
Se ajung la două concluzii importante:

- grupurile de zerouri (*reale și complexe*) se situează la limita Niquist (de eșantionare)
- funcțiile generate de sursele naturale de semnal, de bandă limitată, *conțin zerouri complexe ce nu pot fi toate detectate fizic*
- *nu este deloc ușor de determinat setul complet de zerouri complexe și reale care caracterizează o anumită formă de undă, singura metodă cunoscută care garantează găsirea tuturor zerourilor complexe și reale implică factorizarea unui polinom de gradul $2TW$ – nu e eficient*

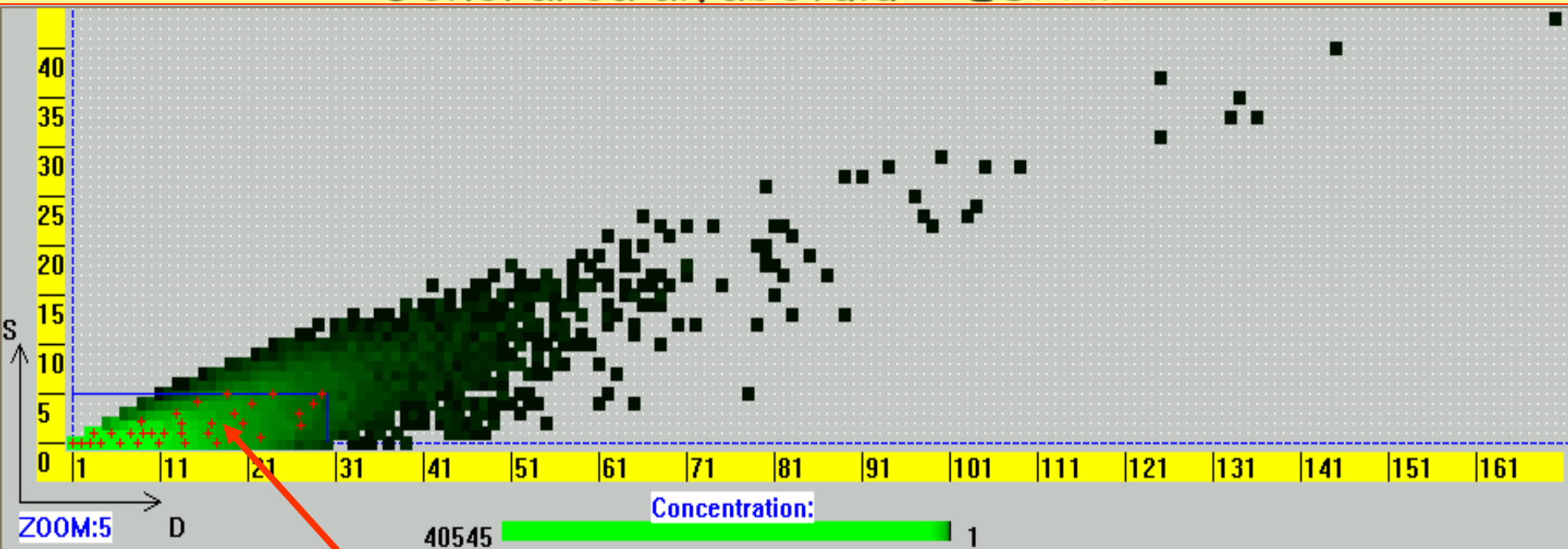
- Putem considera această abordare ca o transformată din domeniul timp >> domeniul zerourilor
- Zerourile reale din *domeniul timp* sunt identice cu locația zerourilor reale din *domeniul zerourilor*, iar zerourile complexe apar conjugate și sunt asociate cu perturbații: minime, maxime sau puncte de inflexiune, care apar în forma de undă între două treceri succesive prin zero;
- Analizând forma de undă între două treceri succesive prin zero se poate obține un număr suficient de subseturi de zerouri pentru a putea caracteriza forma de undă studiată.
- o foarte simplă aproximare a formei de undă bazată pe teoria zerourilor >> metoda codării TESPAP
- *Epoci, descriptori*
 - *durata* dintre două treceri succesive prin zero – **D**
 - *forma* dintre două zerouri reale succesive – **S**

Metoda TESPAP

- Fiecărei epoci i se asociază un simbol pe baza celor 2 descriptori S și D – se obține astfel o puternică compresie
- Totalitatea simbolurilor formează *alfabetul TESPAP*



Generarea alfabetului TESPAP



Distribuția epocilor în planul SD pentru un fișier ce conține o convorbire cu o durată de aproximativ 2 minute, eșantionată la o frecvență de 8 kHz

8 kHz	D=30
11 kHz	D=40
22 kHz	D=70

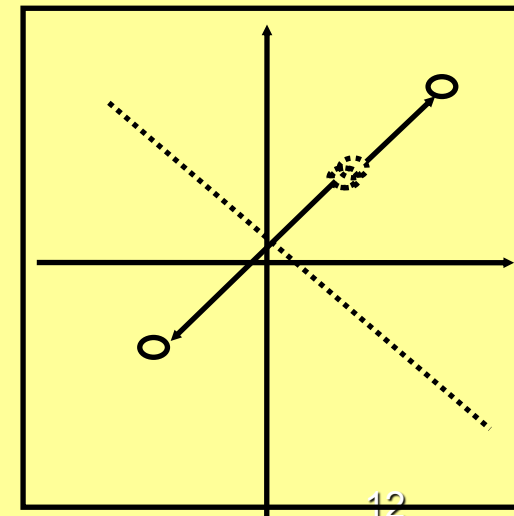
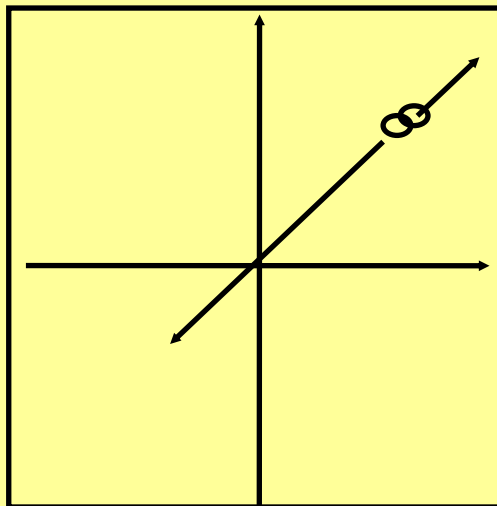
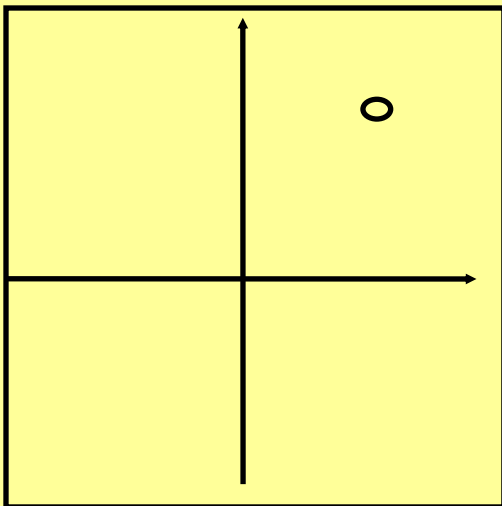
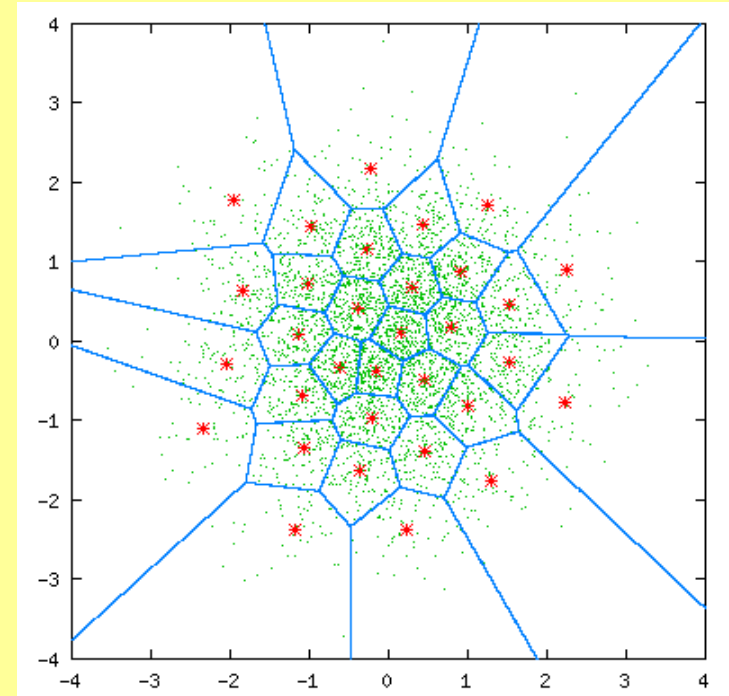
**concentrație
>96%**

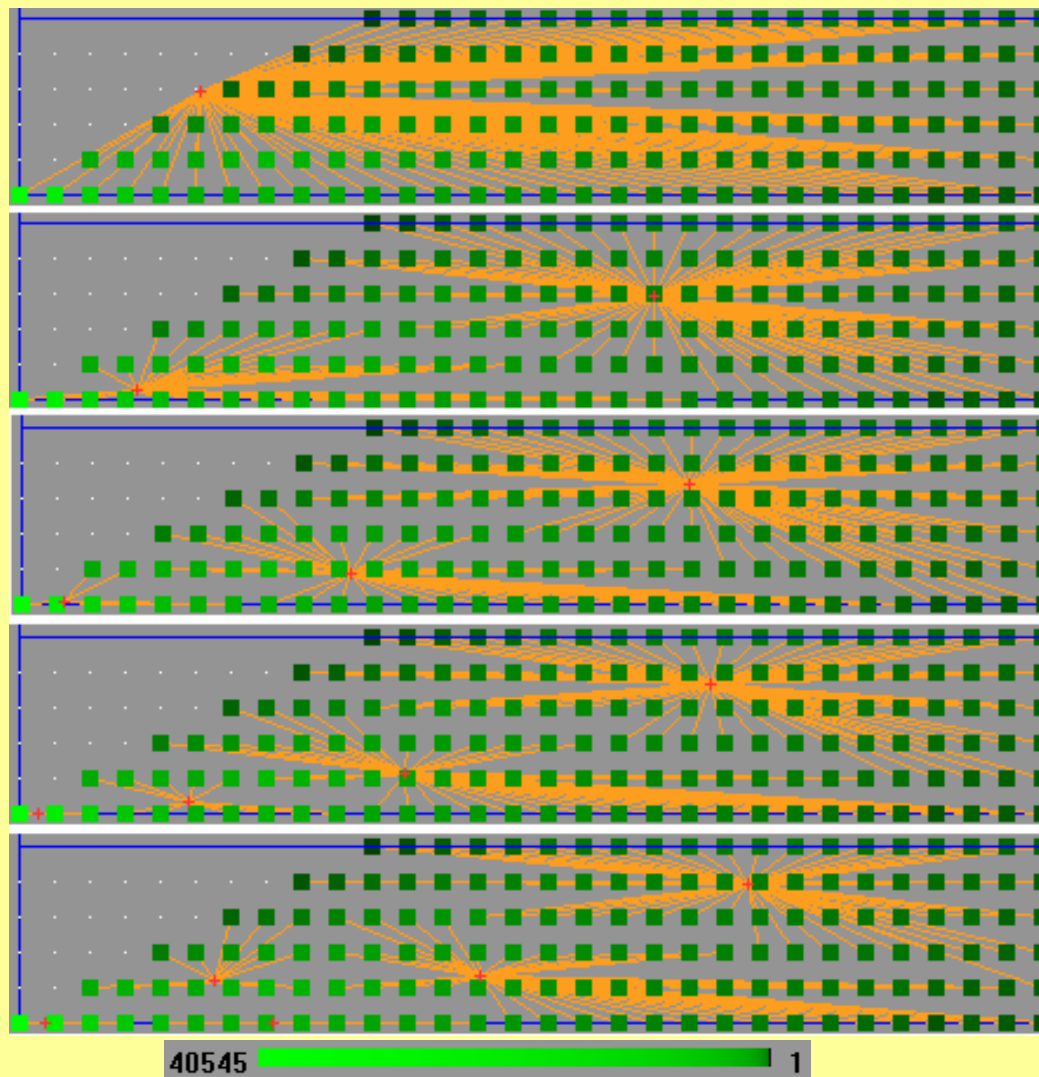
8 kHz	S=5
11 kHz	S=5
22 kHz	S=5

- Se observă o mare concentrare a epocilor în jurul originii

Alfabetul TESPAP – LBG-VQ

- Se determină pe baza proprietăților statistice ale vorbirii umane
- Permite o mai mare compresie
- Elimină o parte deranjantă a variabilității vocii umane
- Este folosit algoritmul de cuantizare vectorială LBG – se urmărește minimizarea distorsiunii globale





Valoarea concentrației epocilor

- 5 pași ai algoritmului de cuantizare vectorială

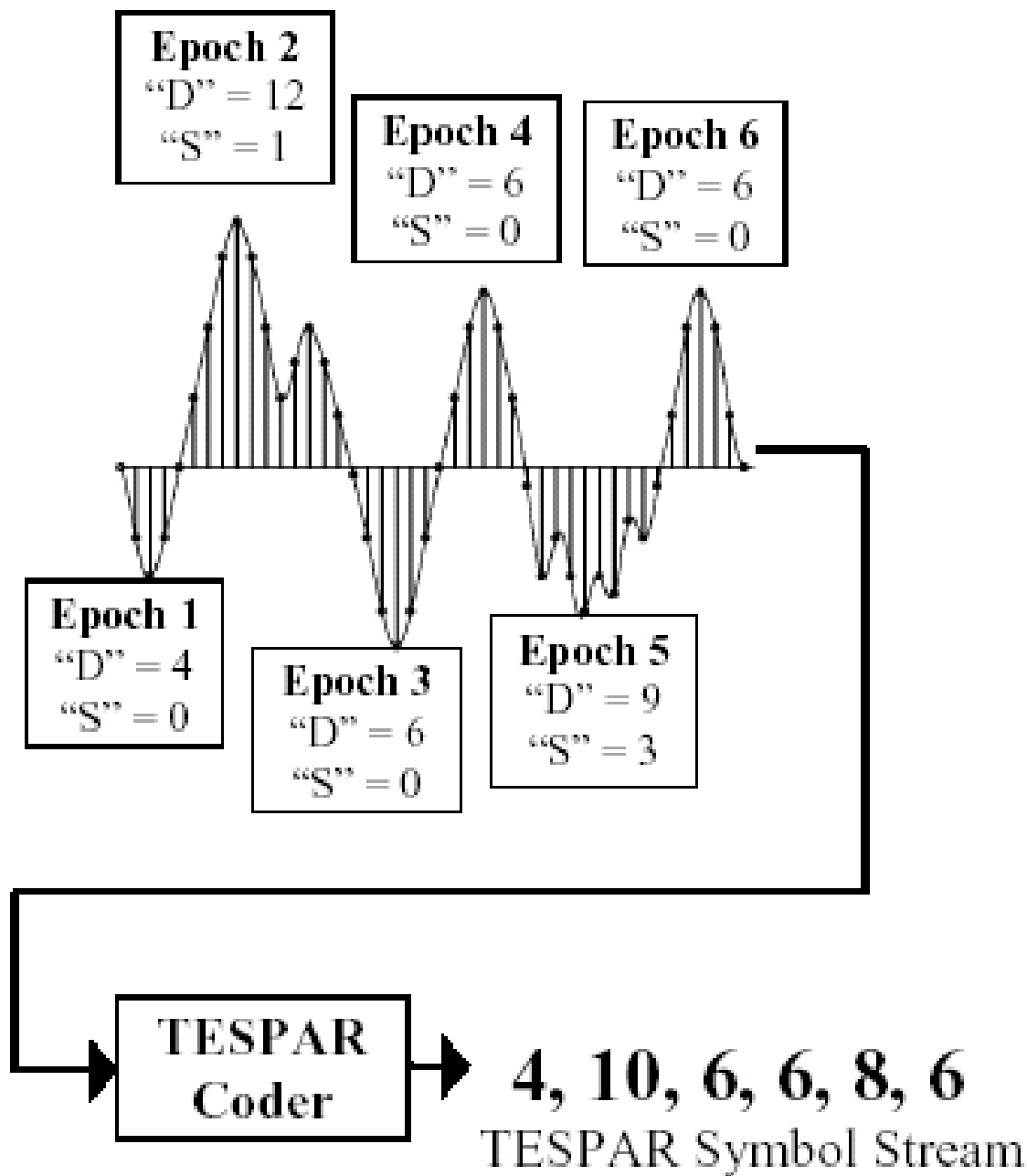
S/D	0	1	2	3	4	5
1	0	-1	-1	-1	-1	-1
2	6	-1	-1	-1	-1	-1
3	14	10	-1	-1	-1	-1
4	4	10	-1	-1	-1	-1
5	30	10	10	-1	-1	-1
6	11	10	25	-1	-1	-1
7	11	9	25	25	-1	-1
8	17	9	25	25	-1	-1
9	1	5	25	25	21	-1
10	1	5	12	21	21	-1
...						...
26	8	20	20	20	3	3
27	8	20	20	20	3	3
28	8	20	20	20	3	3
29	20	20	20	20	3	3
30	20	20	20	20	3	3

Exemplu de alfabet pentru voce esantionata la $f_e=8\text{kHz}$

S/D	0	1	2	3	4	5
1	2	-1	-1	-1	-1	-1
2	17	-1	-1	-1	-1	-1
3	17	7	-1	-1	-1	-1
4	8	7	-1	-1	-1	-1
5	8	7	21	-1	-1	-1
6	25	7	21	-1	-1	-1
7	25	7	21	21	-1	-1
8	25	7	21	21	-1	-1
9	25	7	21	21	9	-1
10	12	11	21	21	9	-1
11	12	11	21	21	9	-1
12	12	11	21	21	9	-1
13	12	11	21	9	9	27
14	12	11	21	9	9	27
15	12	11	21	9	22	27
16	18	11	5	9	22	27
17	18	11	5	9	22	27
18	18	11	5	9	22	27
19	18	23	5	9	22	27
20	18	23	5	9	22	27
21	18	23	5	9	22	27
22	18	23	5	9	22	27
23	18	23	5	9	22	27
24	3	23	29	9	22	27
25	3	23	29	9	22	27
26	3	23	29	9	22	27
27	3	15	29	9	22	27
28	3	15	29	24	31	27
29	3	15	14	24	31	27
30	3	15	14	24	31	27

S/D	0	1	2	3	4	5
31	3	15	14	24	31	4
32	3	15	14	24	31	4
33	3	15	14	24	31	4
34	3	15	30	24	31	4
35	3	15	30	24	13	4
36	6	15	30	24	13	4
37	6	15	30	24	13	4
38	6	15	30	24	13	4
39	6	6	30	24	13	4
40	6	6	30	13	13	4
41	6	6	30	13	13	20
42	6	6	30	13	13	20
43	6	6	30	13	13	20
44	6	6	30	13	13	20
45	6	6	16	13	13	20
46	6	16	16	13	13	20
47	6	16	16	13	10	20
48	6	16	16	13	10	20
49	6	16	16	10	10	20
50	6	16	16	10	10	20
51	6	16	16	10	10	20
52	6	16	16	10	10	28
53	6	16	16	10	10	28
54	6	16	16	10	10	28
55	6	16	16	10	10	28
56	6	16	19	19	10	28
57	6	16	19	19	10	28
58	6	16	19	19	10	28
59	6	16	19	19	10	28
60	6	16	19	19	10	28
61	6	16	19	19	10	28
62	16	16	19	19	10	28
63	16	16	19	19	10	26
64	16	19	19	19	26	26
65	16	19	19	19	26	26
66	16	19	19	19	26	26
67	16	19	19	19	26	26
68	16	19	19	19	26	26
69	16	19	19	19	26	26
70	16	19	19	19	26	26

Tabel 16: Alfabetul implicit pentru frecvența de 22kHz

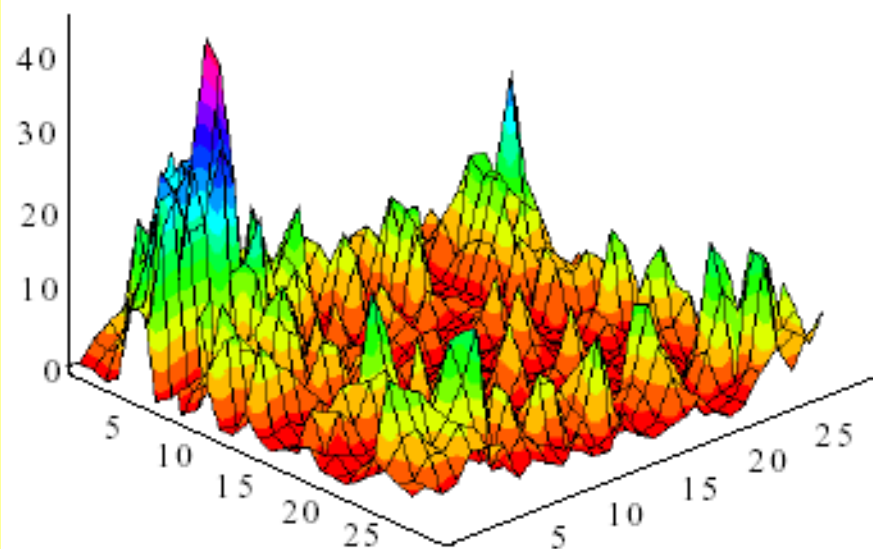
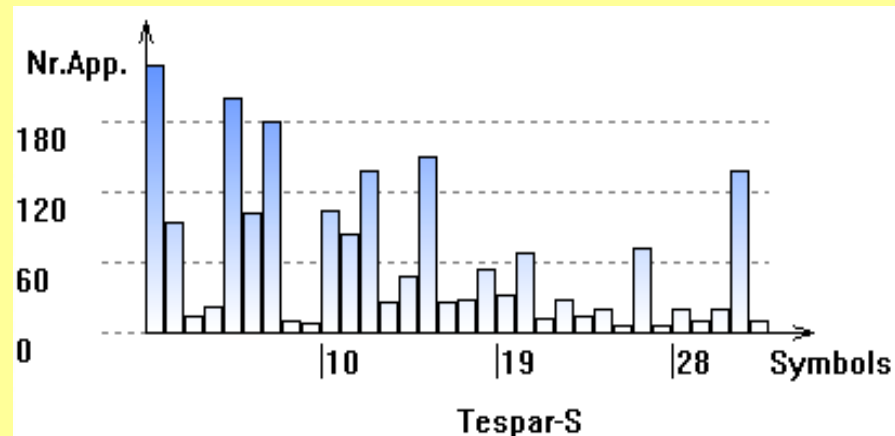


Matricile S și A

- Din sirul de simboluri TESPAP obținut la ieșirea codorului se pot crea matricile TESPAP:

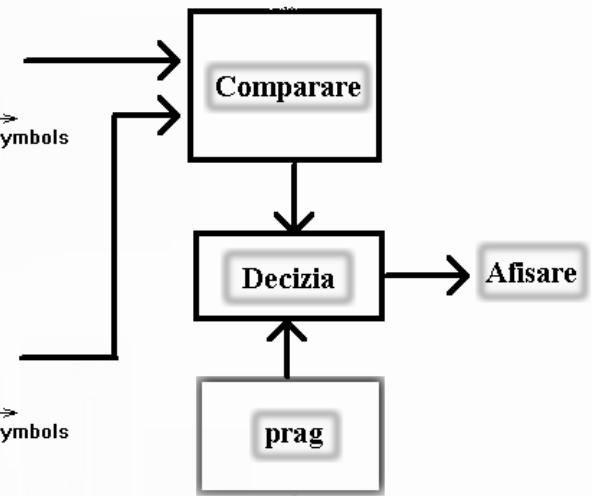
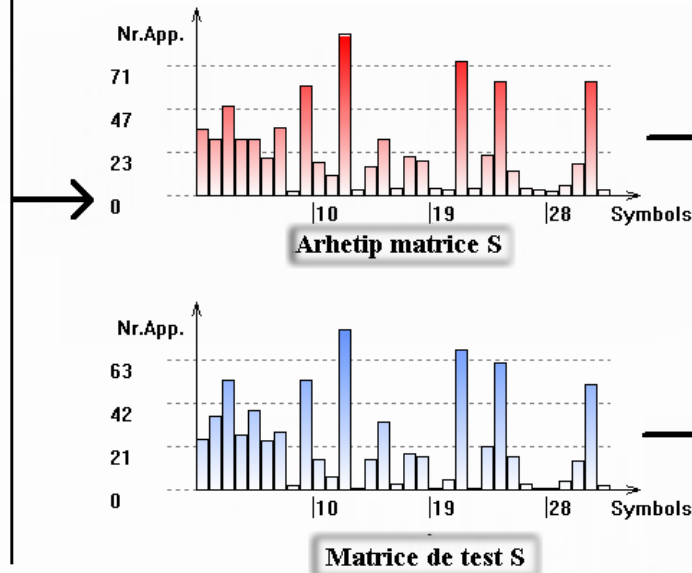
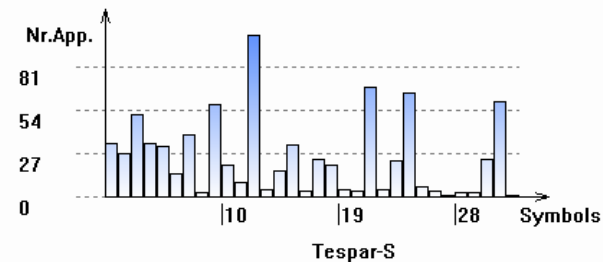
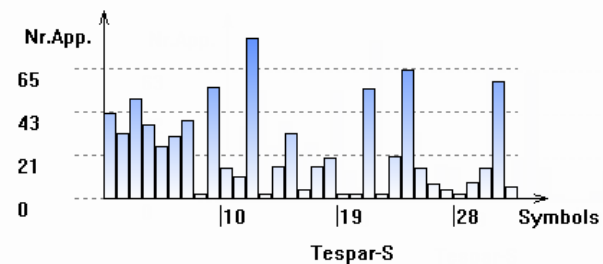
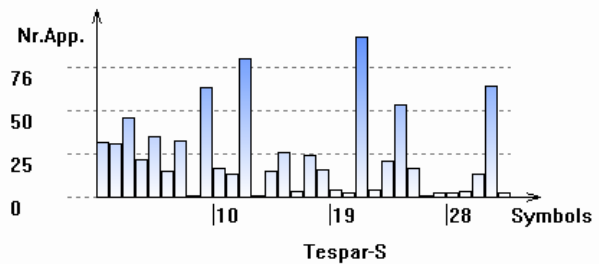
Matricea S – (unidimensională)
păstrează doar numărul de apariții
ale simbolurilor (histograma)

Matricea A – (bidimensională)
conține și informații temporale –
implică un alt parametru $L=LAG$
*-se contorizeaza nr. de perechi de
simboluri aparute la o distanta egala
cu lag-ul*



Există două metode principale de clasificare utilizând această metodă:

- clasificarea utilizând arhetipuri (medii ale matricilor model/antrenare)
 - clasificarea cu ajutorul rețelelor neuronale
-
- Principalul avantaj al prelucrării semnalelor variabile în domeniul timp utilizând metoda TESPAR - față de metodele convenționale care utilizează descriptori în domeniul frecvență constă în faptul că *matricile TESPAR sunt structuri de dimensiune fixă si ofera un grad inalt de compresie*



Generarea arhetipului și utilizarea lui pentru clasificare

- Arhetip = media mai multor matrici de intrare S/A, elimină variabilitatea
- Se poate urmări variația valorilor datelor pe fiecare dimensiune pentru a le da o pondere sporită celor care variază mai puțin

Distanțe

- Norma HOLDER

$$d_p(\mathbf{x}, \mathbf{y}) = \left[\sum_{k=1}^N |\mathbf{x}_k - \mathbf{y}_k|^p \right]^{1/p} = \|\mathbf{x} - \mathbf{y}\|_p$$

forme particulare ale normei Holder:

- Distanța euclidiană

$$d_2(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^N (\mathbf{x}_i - \mathbf{y}_i)^2}$$

- Distanța City-block

$$d_1(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^N |\mathbf{x}_i - \mathbf{y}_i|$$

Nu toate distanțele utilizate sunt norme:

- Distanța Mahalanobis

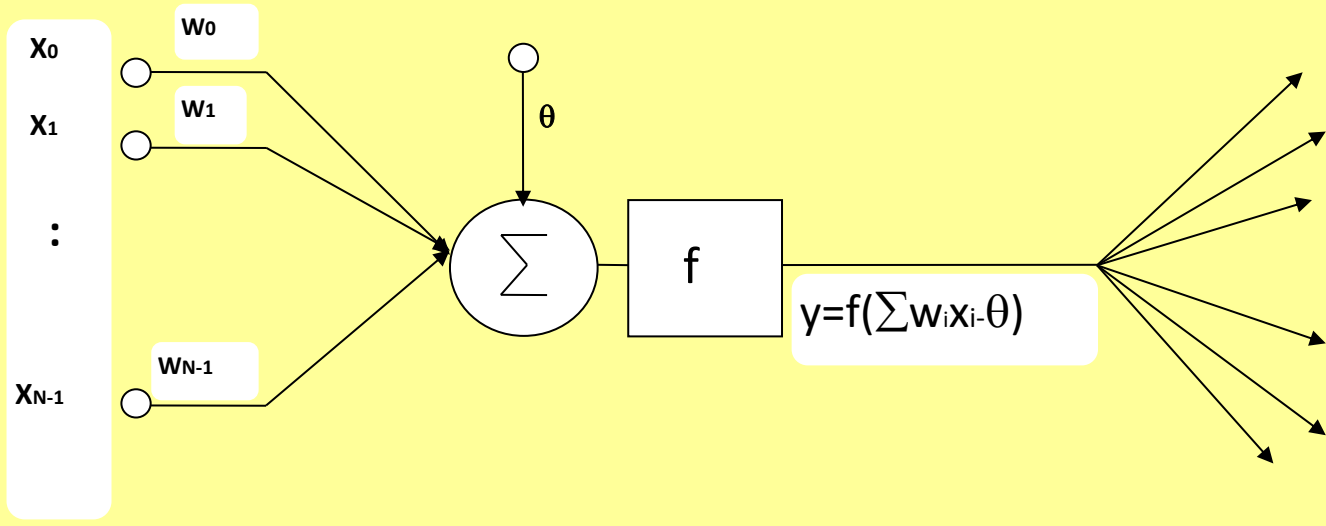
$$d(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^T \mathbf{W}^{-1} (\mathbf{x} - \mathbf{y})$$

$\mathbf{W}$ - matricea de covarianță e dificil de obținut

putem pondera diferențele pe fiecare dimensiune în funcție de variația dimensiunilor în procesul de generare al arhetipului cu factorul:

$$f = \frac{1}{1 + a * \frac{\Delta \text{simbolarhetip}}{\text{simbolarhetip}}}$$

Retele neuronale



Neuronul artificial

$$\sum_{i=1}^N W_i X_i - \theta = W \cdot X^t - \theta \quad \text{și} \quad y = f(W \cdot X^t - \theta)$$

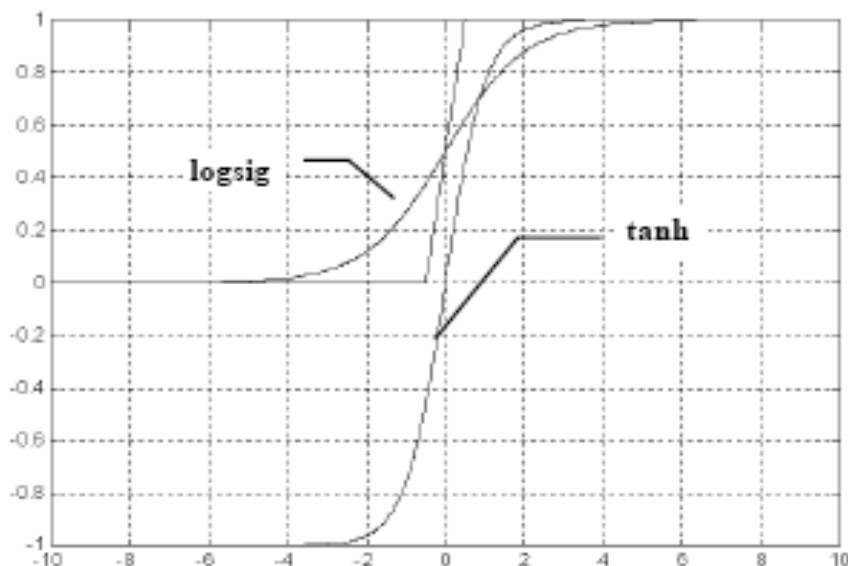


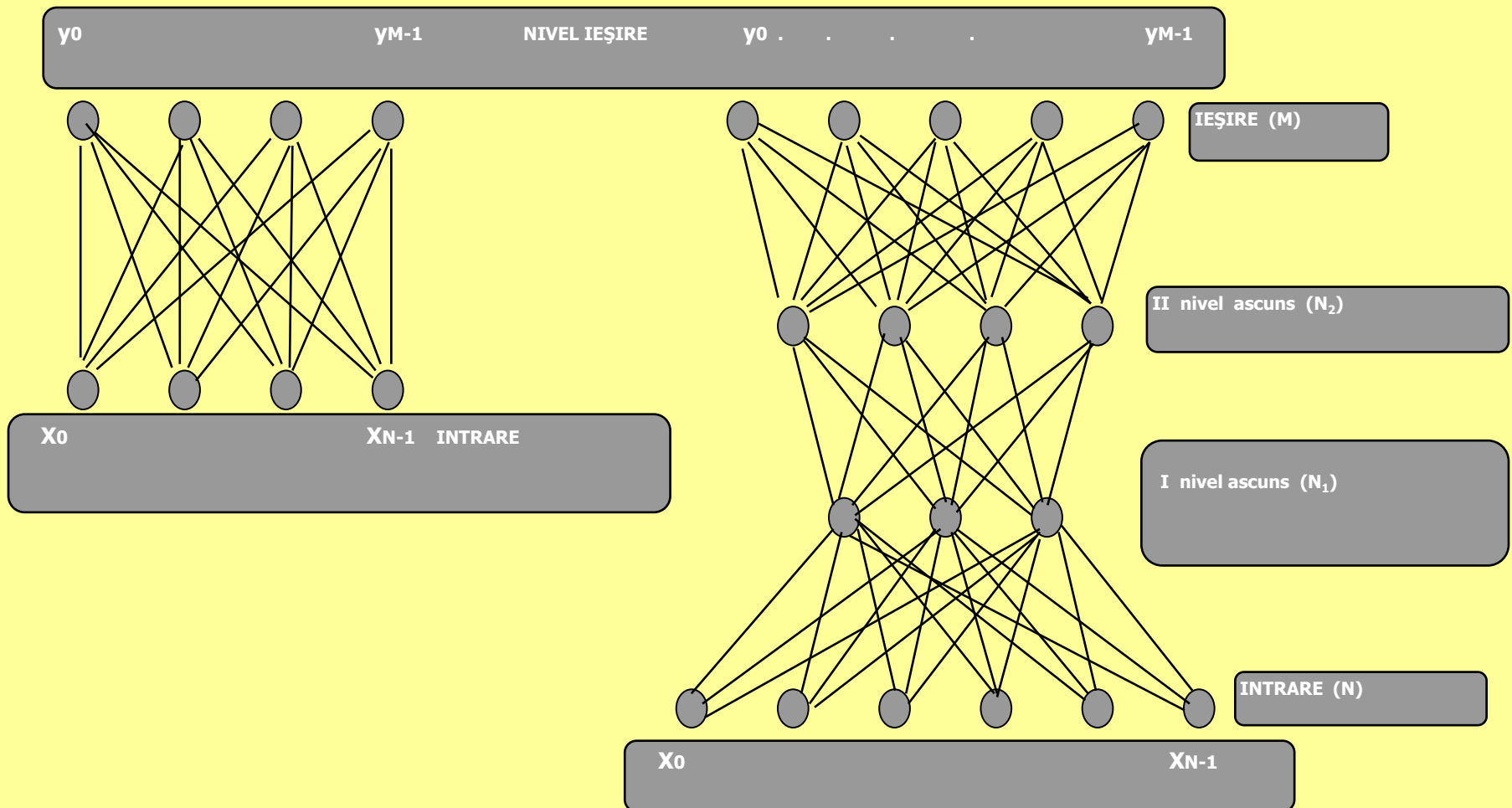
Figura 10: Neliniarități folosite pentru funcția de activare

$$f(x) = \begin{cases} +1, & x \geq 0 \\ 0, & x \leq 0 \end{cases} \quad \text{sau}$$

$$f(x) = \tanh(\beta x) = \frac{e^{\beta x} - e^{-\beta x}}{e^{\beta x} + e^{-\beta x}}, \quad \beta > 0$$

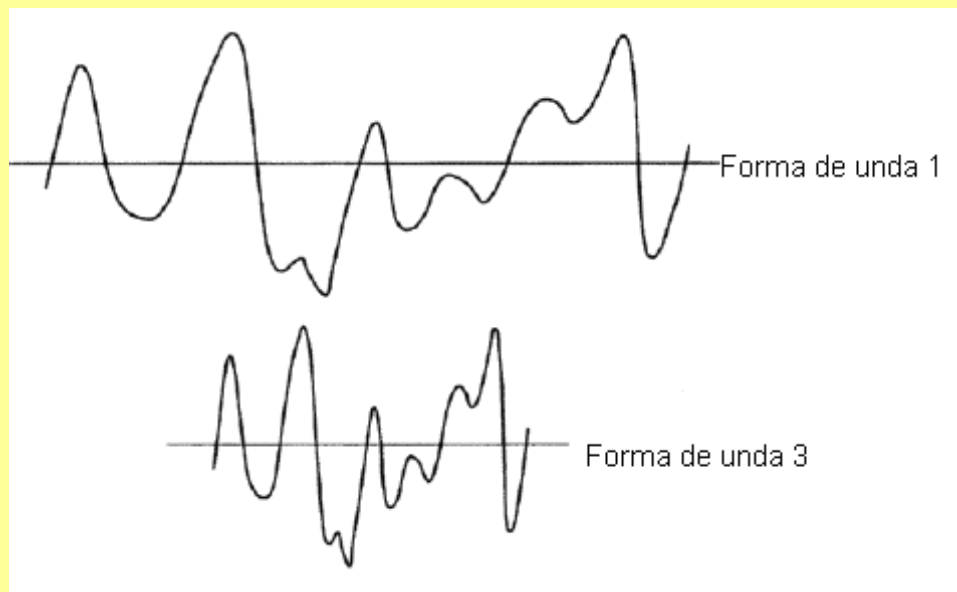
$$f(x) = \frac{1}{1 + e^{-\beta x}}, \quad \beta > 0$$

Multi Layer Perceptron - backpropagation

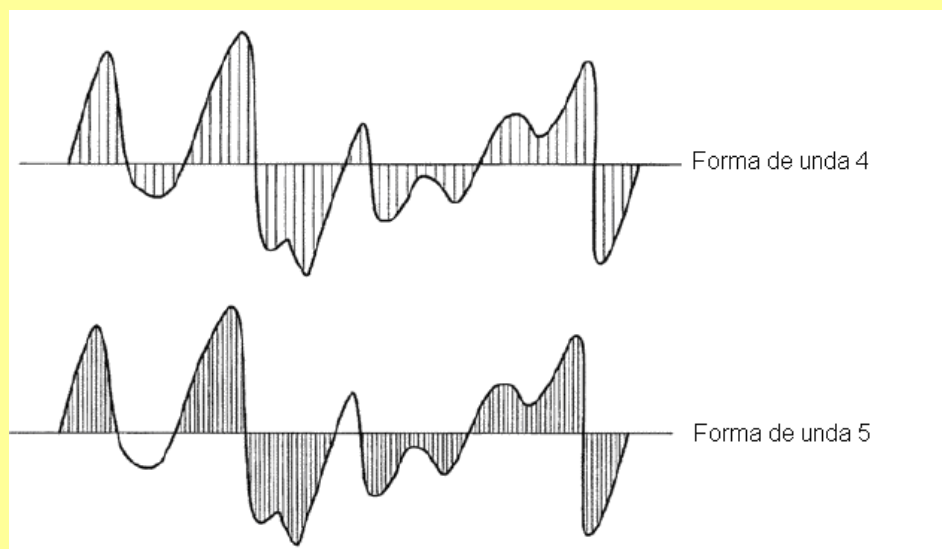


- Pentru furnizarea de rezultate bune se folosesc Masive Paralele MLP – mai multe rețele paralele a caror iesiri, pe baza votului majoritar, decid iesirea corespunzatoare intrarii

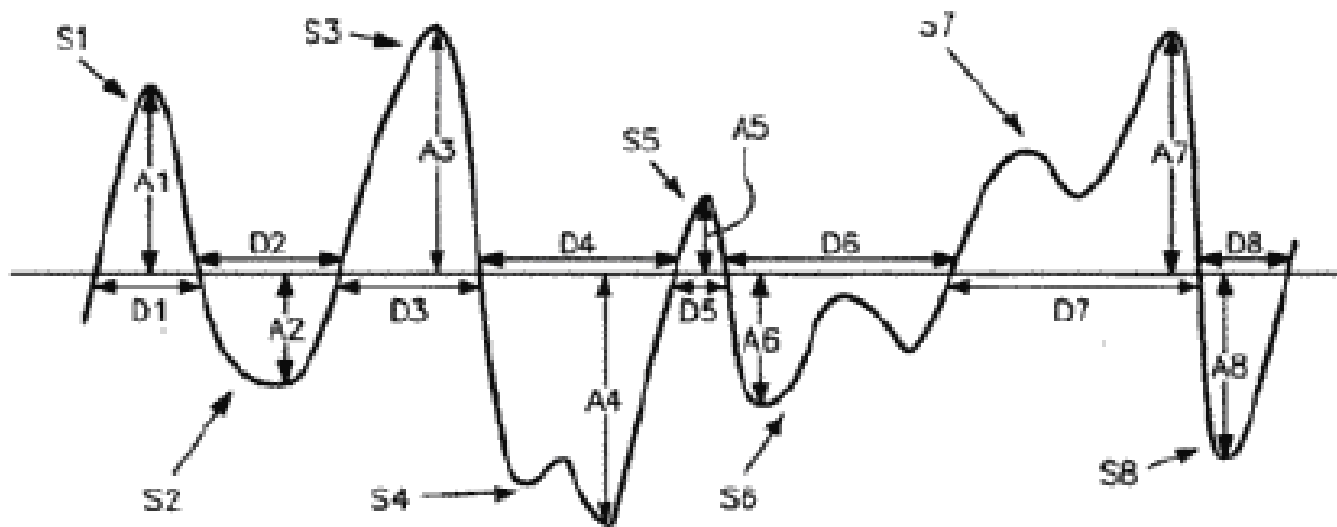
Metoda TESPAP DZ



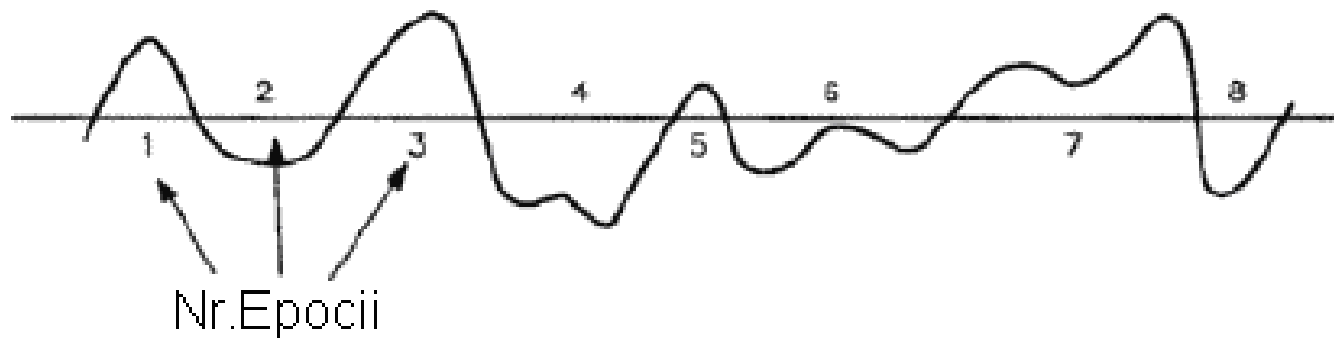
Forma de undă inițială și cea eșantionată la o frecvență dublă



Formele de undă eșantionate la frecvențe diferite

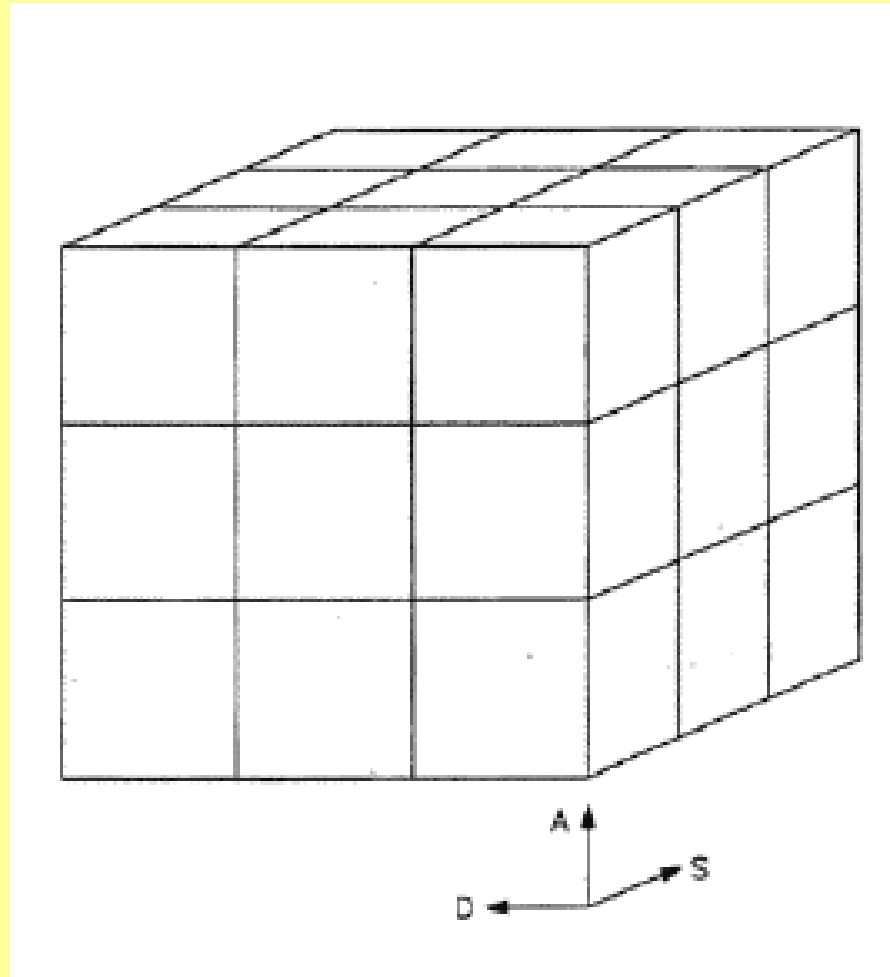


Forma de unda 1

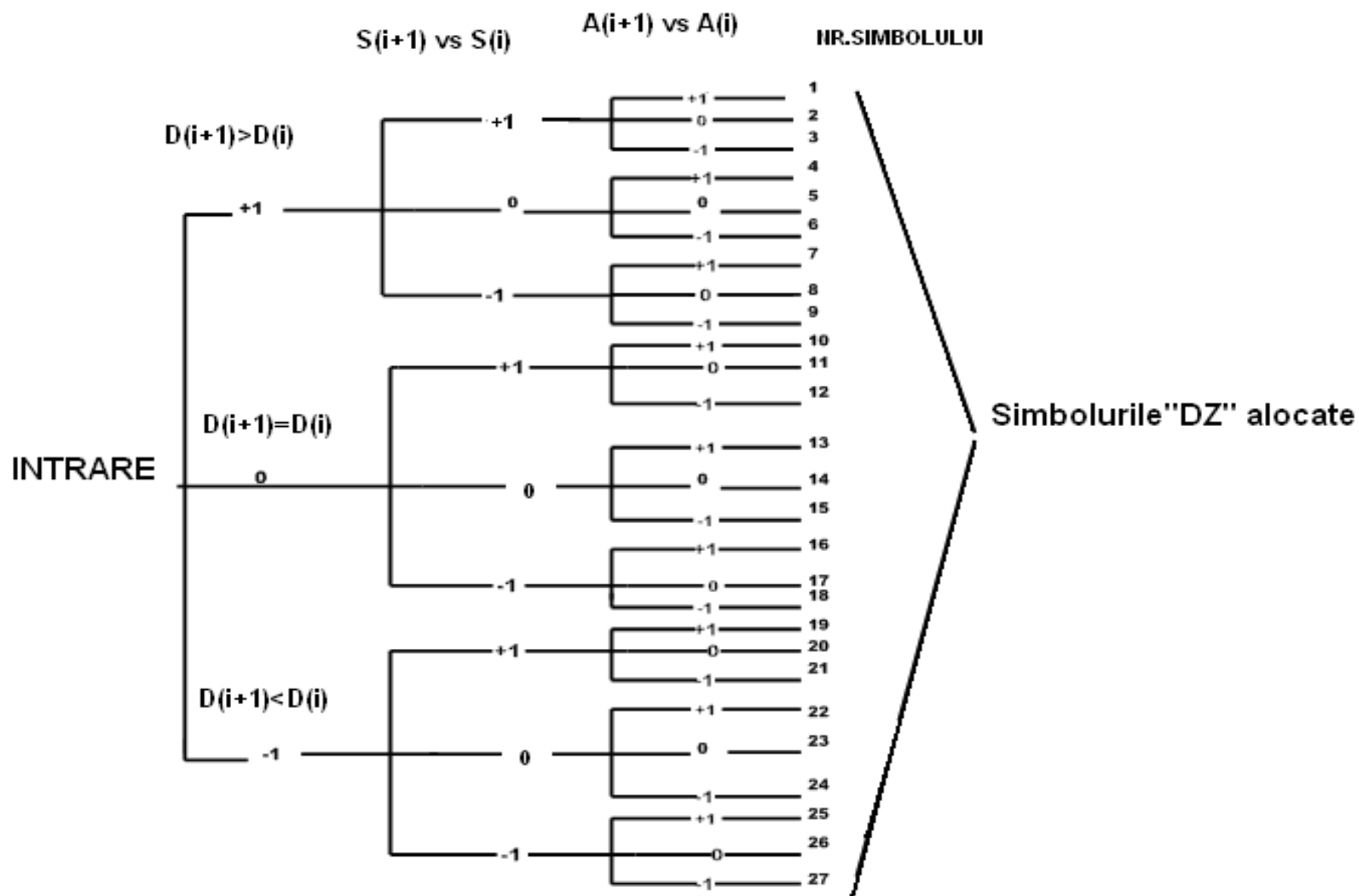


Forma de unda 2

- Sunt realizate comparații între perechi de epoci, modalitate prin care sunt comparate caracteristicile individuale D, S și A ale fiecărei perechi și este produs un vector diferențial indicând diferențele dintre D, S și A ale celor două epoci care sunt comparate.



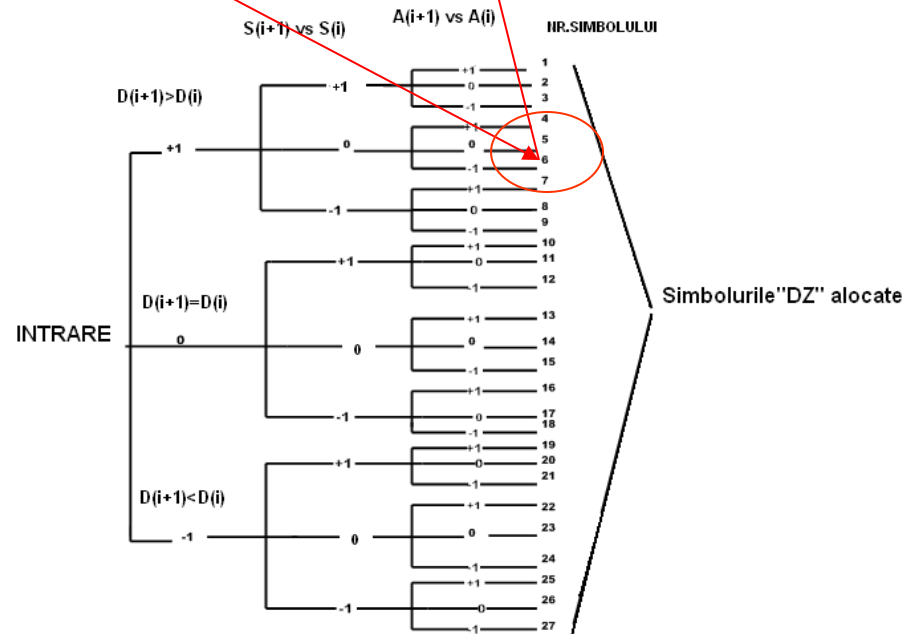
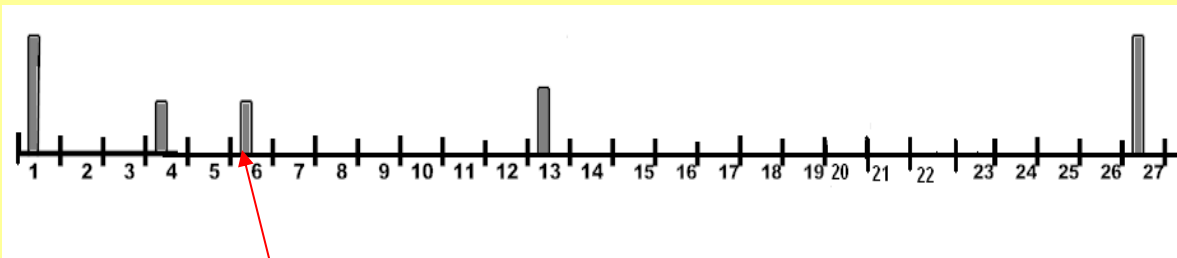
Spațiul tridimensional pentru
cei 3 descriptori

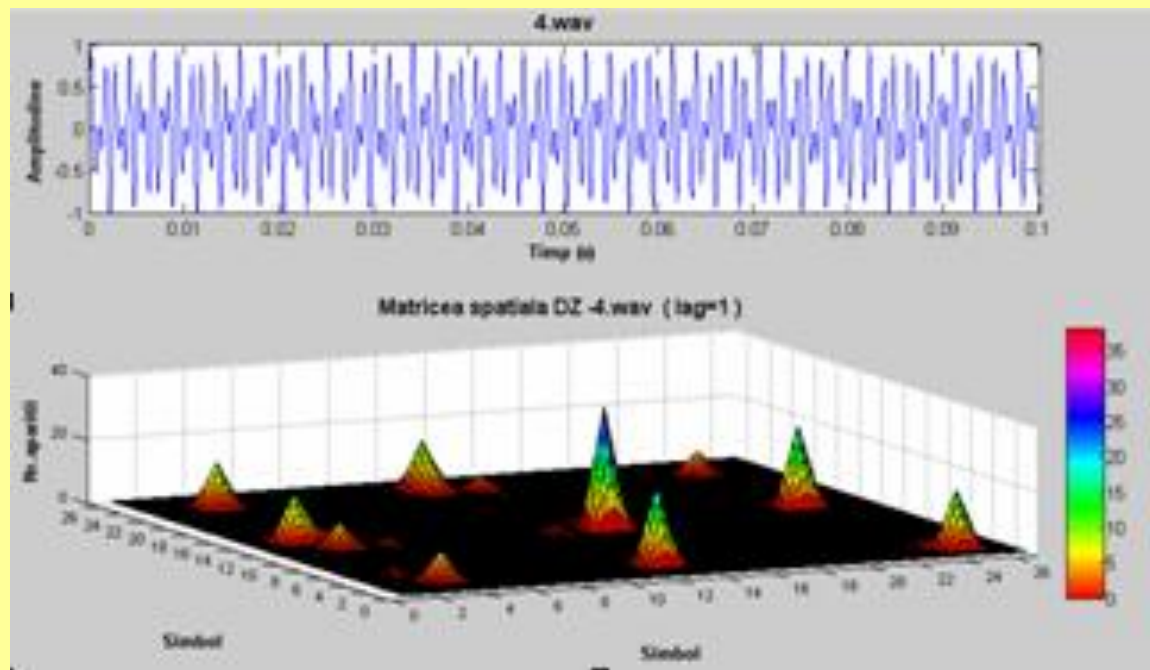
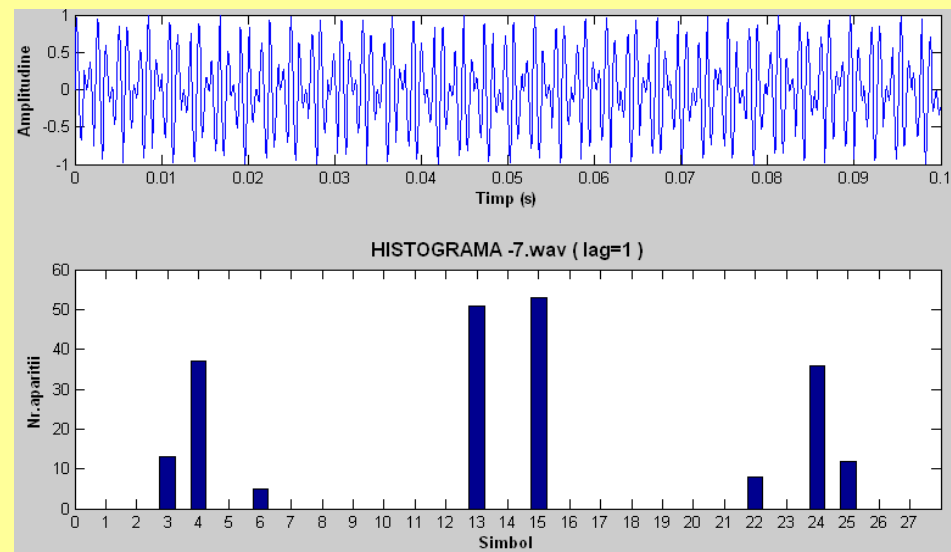
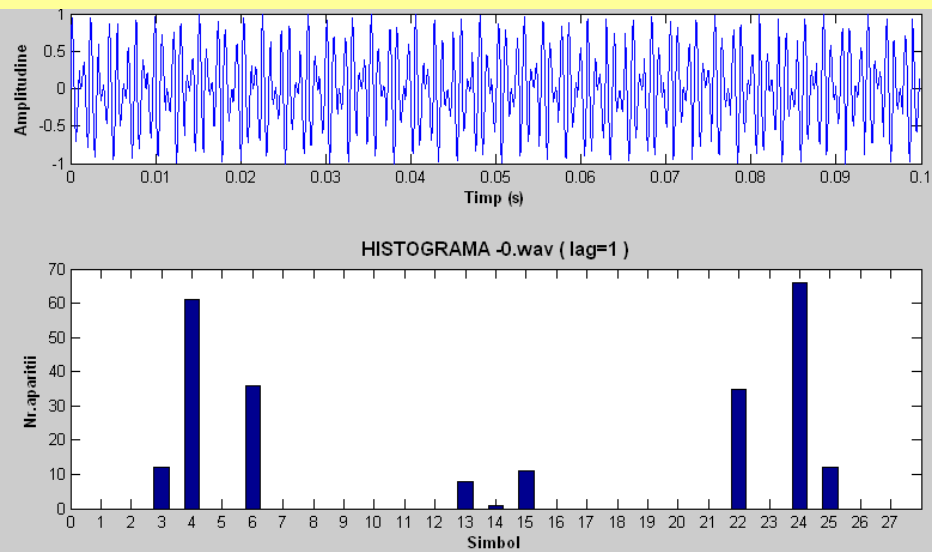


Alocarea simbolurilor în funcție de cei 3 descriptori_

- Se alocă un simbol de la 1.....27 pe baza diagramei de alocare:

E2/E1	E3/E2	E4/E3	E5/E4	E6/E5	E7/E6	E8/E7
D2>D1=>+	D3=D2=>0	D4>D3=>+	D5<D4=>-	D6>D5=>+	D7>D6=>+	D8<D7=>-
S2=S1=>0	S3=S2=>0	S4>S3=>+	S5<S4=>-	S6>S5=>+	S7=S6=>0	S8<S7=>-
A2<A1=>-	A3>A2=+	A4=A3=>+	A5<A4=>-	A6>A5=>+	A7>A6=>+	A8<A7=>-
6	13	1	27	1	4	7





- Tabelul distanțelor Hamming dintre cele 12 tonuri DTMF:

	0	1	2	3	4	5	6	7	8	9	*	#
0	0	125	133	125	215	115	93	195	80	84	133	97
1	125	0	36	112	90	26	82	98	57	83	92	60
2	133	36	0	94	88	34	82	110	67	83	112	52
3	125	112	94	0	180	104	52	190	109	61	184	118
4	215	90	88	180	0	102	164	38	139	165	98	126
5	115	26	34	104	102	0	70	102	45	71	88	42
6	93	82	82	52	164	70	0	168	65	33	140	74
7	195	98	110	190	38	102	168	0	127	169	76	118
8	80	57	67	109	139	45	65	127	0	50	79	29
9	84	83	83	61	165	71	33	169	50	0	125	61
*	133	92	112	184	98	88	140	76	79	125	0	72
#	97	60	52	118	126	42	74	118	29	61	72	0

- Observăm că obținem pe diagonală doar zerouri, iar distanța minimă este 26 și maximă 215.

AVANTAJE

- Putere și timp de calcul cu cel puțin un ordin de mărime mai reduse față de alte abordări (frecventa)
- Potrivită pentru aplicații – timp real
- Simplitate de implementare – cost redus
- Robustețe
- Poate diferenția semnale aproape identice în domeniul frecvență

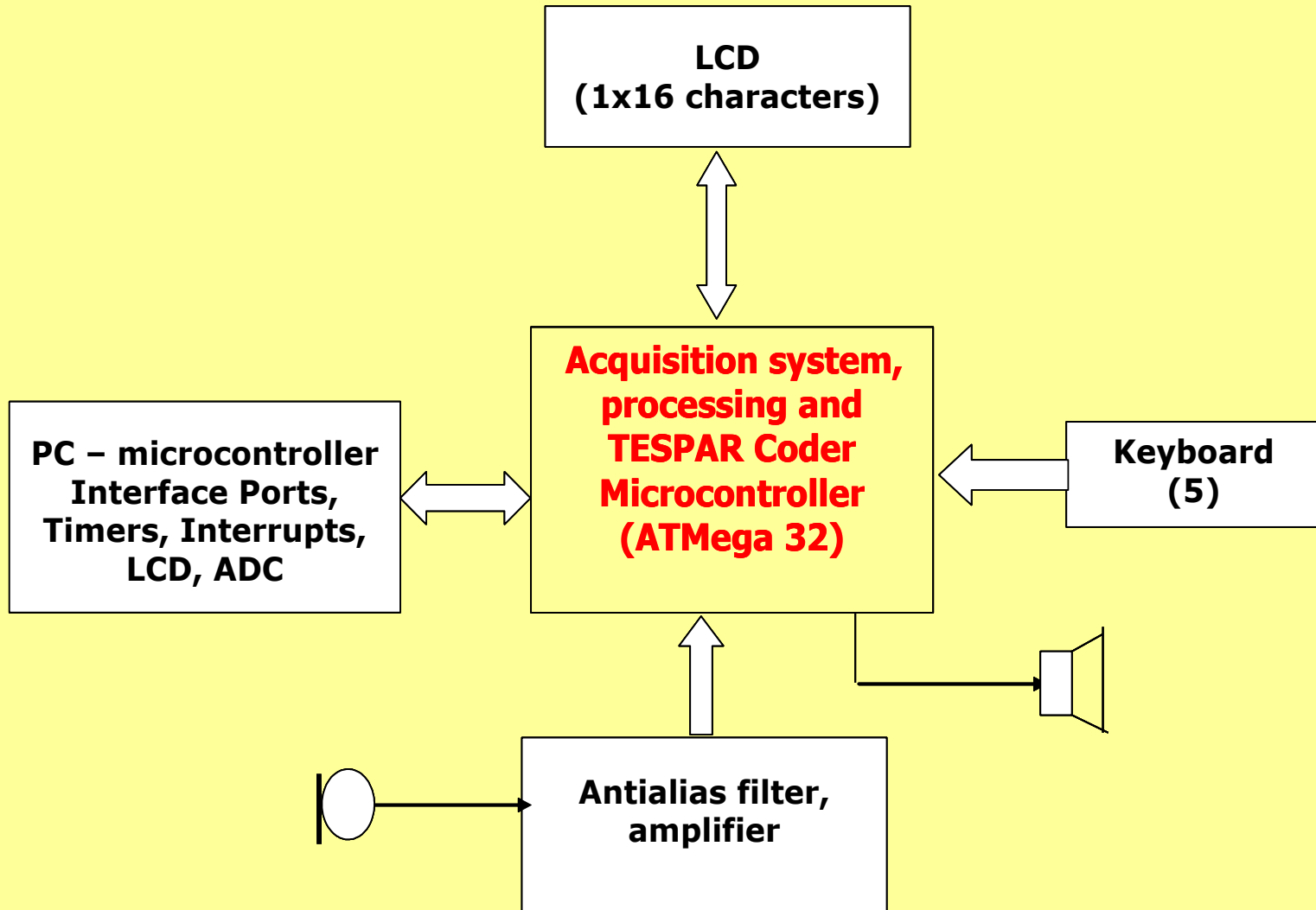
DEZAVANTAJE

- Antrenare MLP necesită un timp îndelungat
- Distanțele nu oferă un gard suficient de siguranță pentru aplicații biometrice, deci se recomandă RN

DOMENII DE APLICABILITATE

- Verificări biometrice, recunoaștere vocală
- Diagnoză pe baza vibrațiilor la poduri, turbine, motoare sau alte surse de vibrații naturale, chiar și diagnoză medicală

RECOGNITION SYSTEM overview



The system block diagram