

"Nimic nu e mai inutil decât să *faci bine* ceea ce nu ar trebui să faci,,

Peter Drucker

ASRSV – curs 2

Prelucrarea Semnalului Vocal

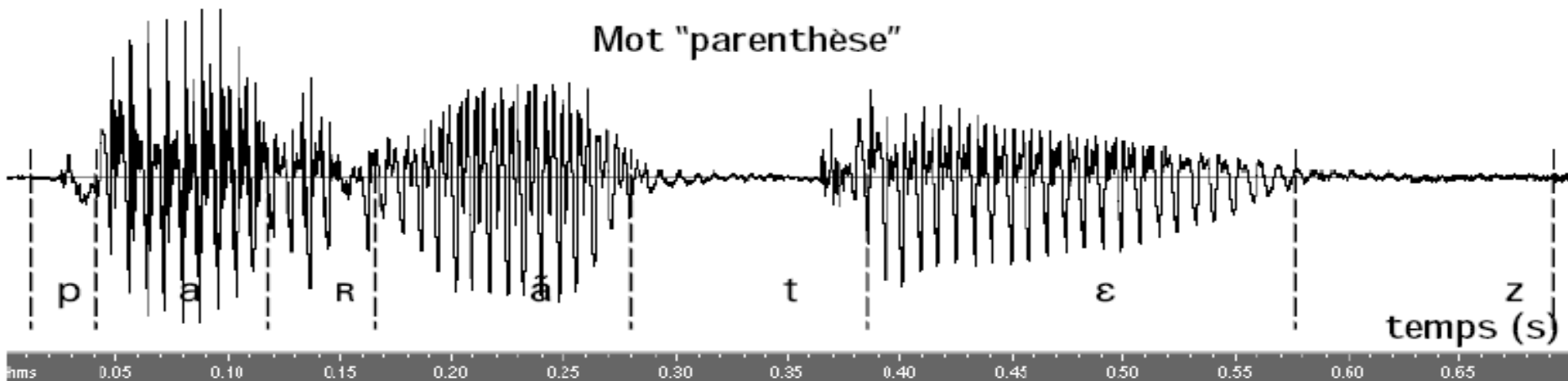
- 1. Generalitati despre SV si ap. auditiv**
- 2. Prelucrarea SV**
- 3. Tehnica ferestruirii**

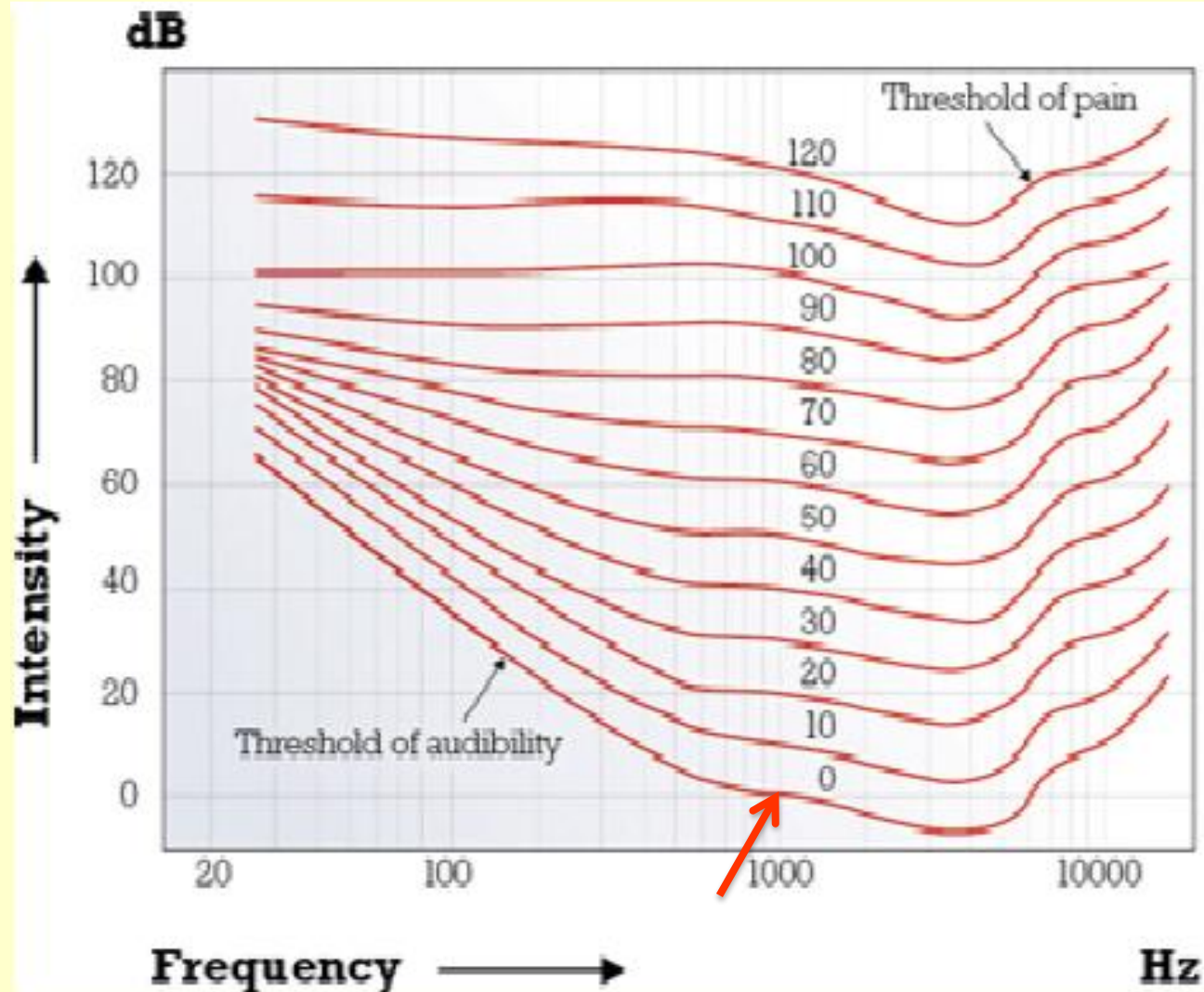
1. Generalități despre SV si ap. auditiv

- Semnalul vocal :
 - pe termen lung este semnal *nestaționar, continuu, de energie finită*;
 - pe perioade scurte de timp (10-30 ms) *cvasistaționar* →

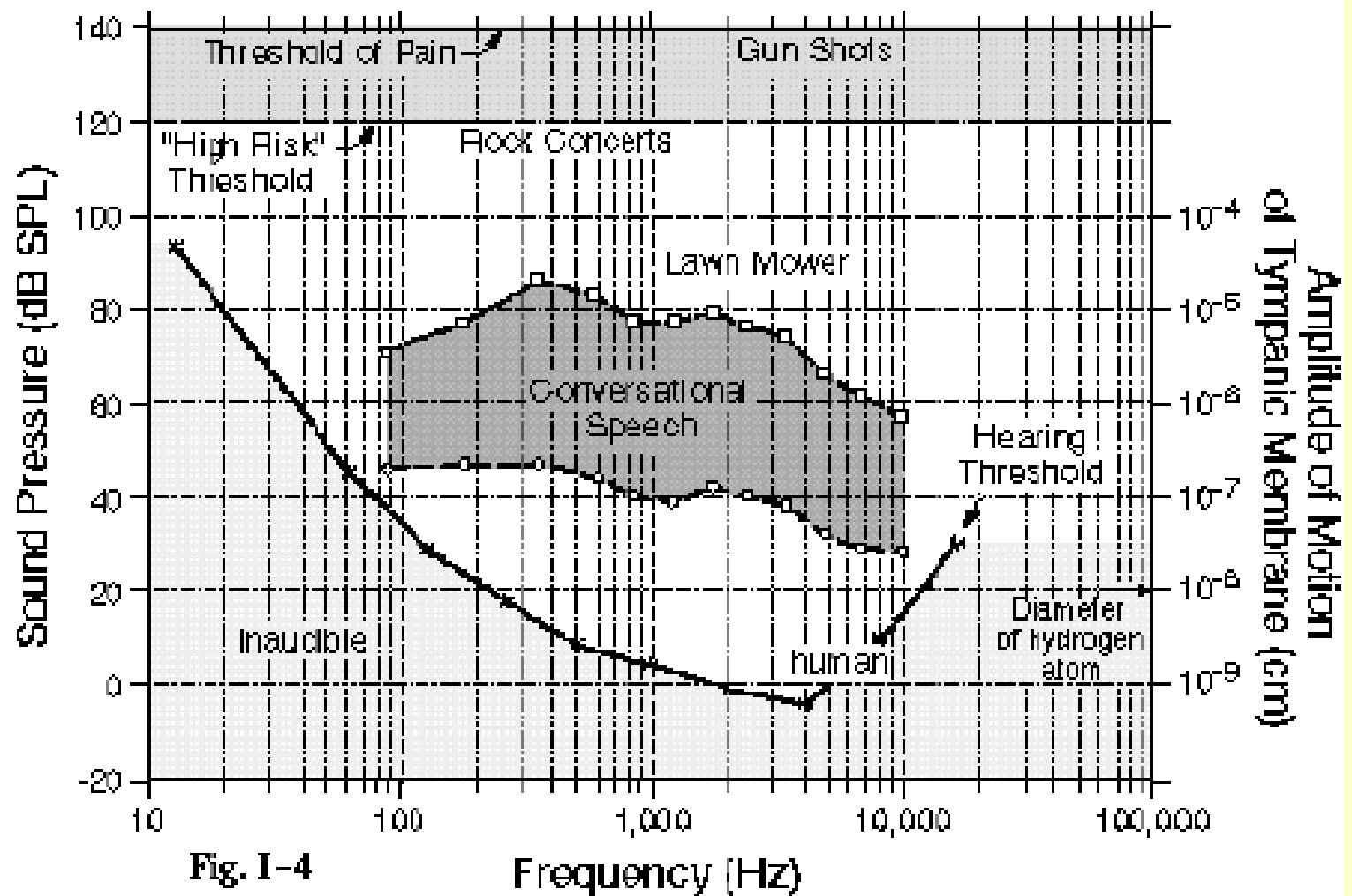
ANALIZA pe TIMP SCURT

- Semnalului vocal are *redundanța mare*, care se încearcă a se elimina prin prelucrări adecvate (transmisii eficiente, stocare optima sau la recunoasterea vorbirii)
- **Redundanța** este utilă, de exemplu la transmisii în medii zgomotoase, ajutând aparatul auditiv la recunoașterea mesajului transmis chiar în condițiile unui raport semnal zgomot (SNR) redus; (semantica) – faciliteaza intelegerea

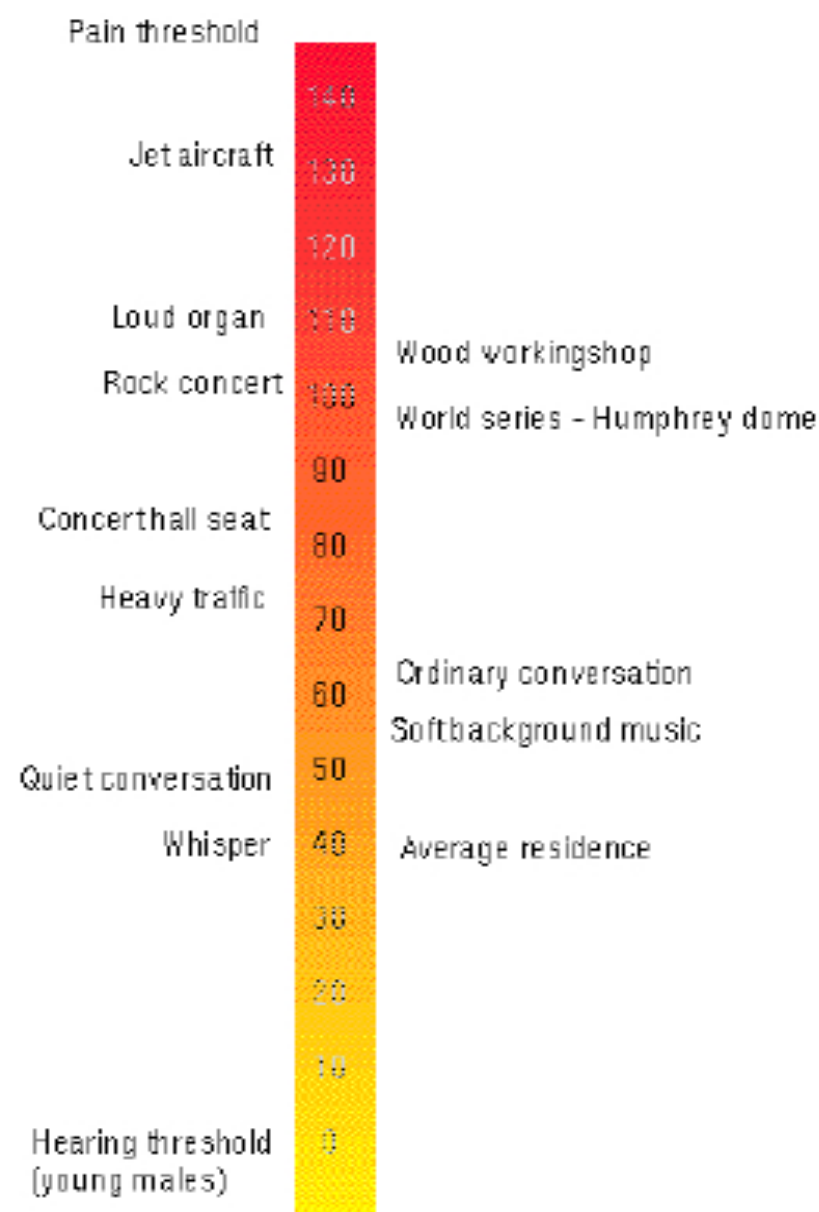




- proprietatile ap. auditiv → *caracteristica neliniara*
- urechea este capabilă să perceapă anumite sunete numai dacă intensitatea lor depășește o anumită valoare, prag de audibilitate, care depinde de ascultător și este stabilit la $10\text{-}16 \text{ W/cm}^2$ la o frecvență de 1kHz (0dB)

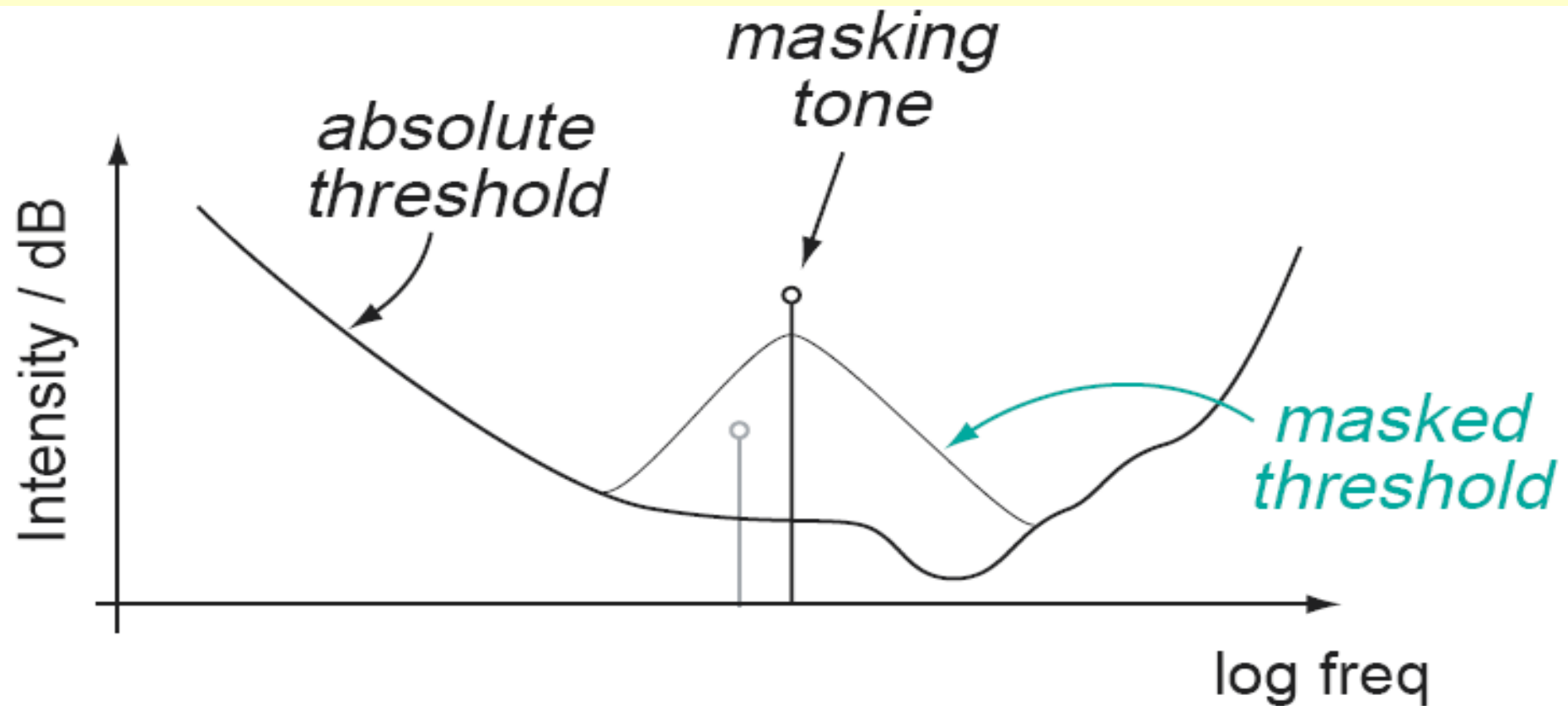


SOUND PRESSURE LEVEL OF COMMON ENVIRONMENTAL SOUNDS (dB)



Sistemul auditiv:

- pentru a putea sesiza un sunet pur acesta trebuie să aibă o durată $>60\text{ms}$;
- poate face distincția între două sunete a căror frecvență de repetiție este sub 100Hz ;
- Diferențele mici de fază nu sunt percepute (pentru întârzieri $< 50\text{ms}$), dar la valori mai mari se percepe o reverberație;
- Rezoluția de percepere a frecvențelor este de 3Hz în jurul frecvenței de 1KHz .



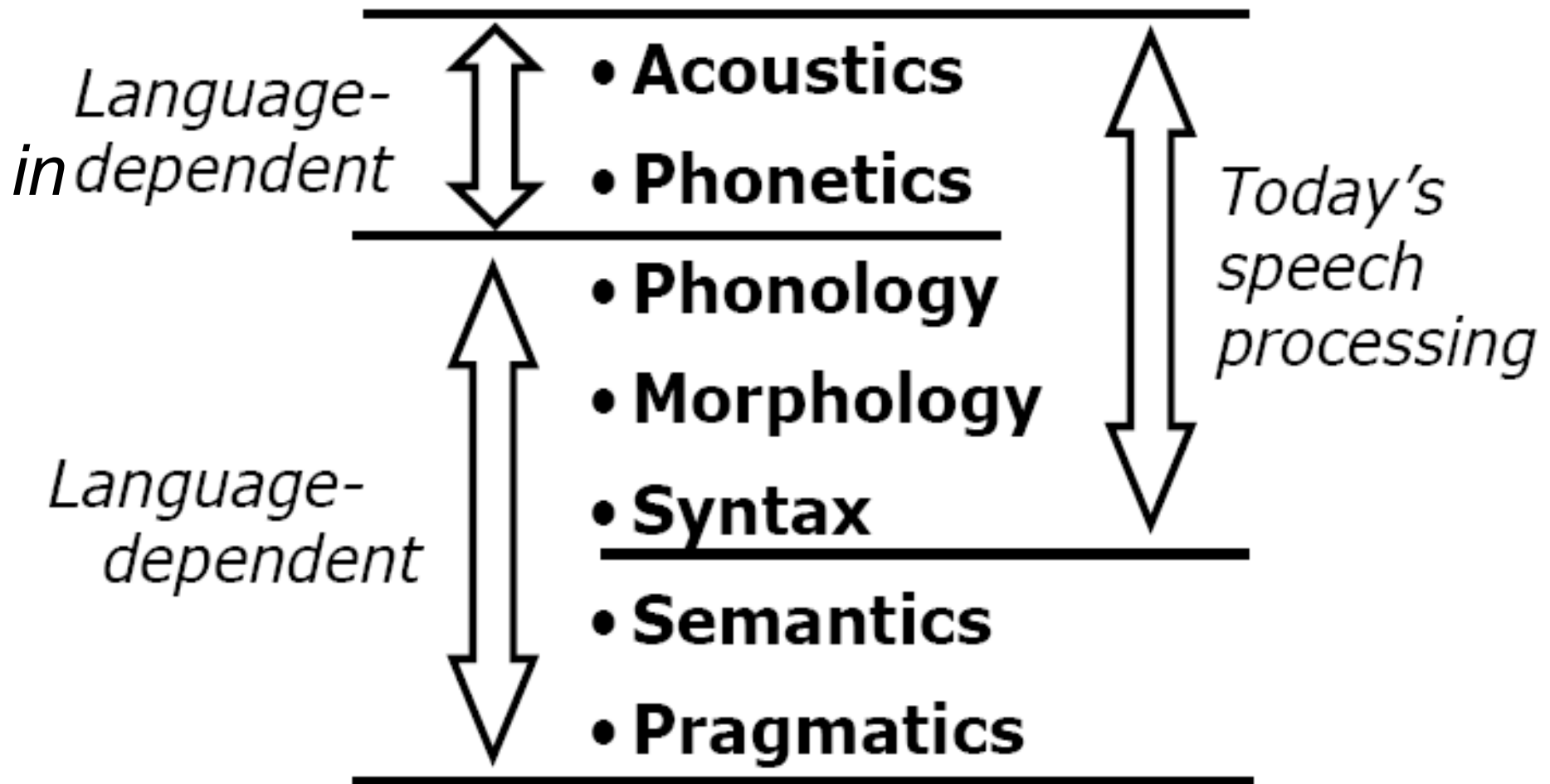
Principalele cauze ale variabilității SV

MEDIU	• Zgomot de reverberatie corelat cu vorbirea • Zgomot de reflexie necorelat • Zgomot aditiv (stationar/nestationar)
VORBITOR	• Atributele vorbitorului: dialect, sex, vârsta, sanatate • Modul de vorbire : zgomot de respiratie si buze - stres - efect LOMBARD - debit verbal - nivel - frecventa fundamentala - cooperativitate
ECHIPAMENT DE INTRARE	• Microfon (transmitator) • Distanța de microfon • Filtre • Sistemul de transmisie (distorsiuni, zgomot, ecou) • Echipament de înregistrare

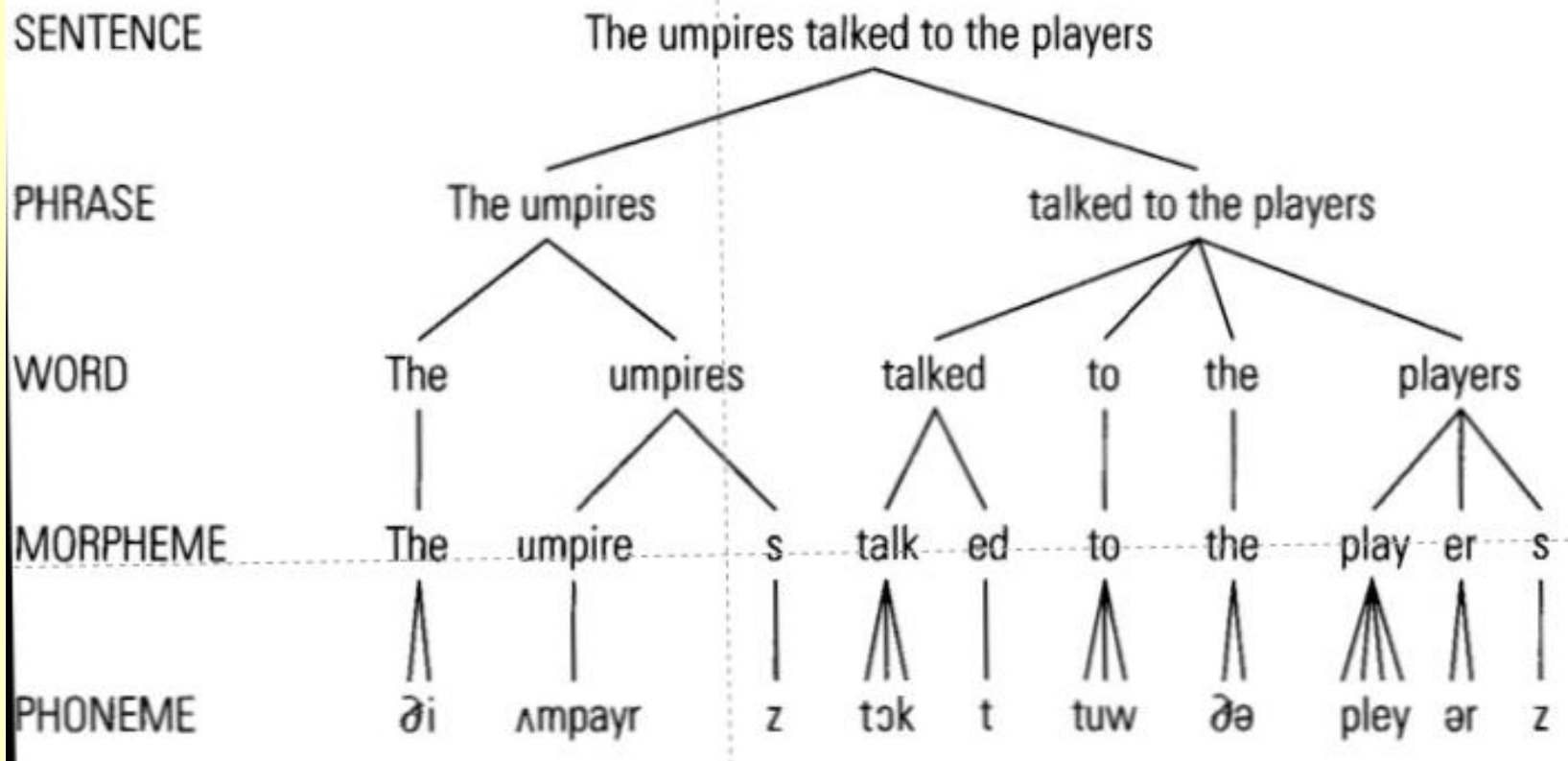
Efectul Lombard - tendința involuntară a vorbitorilor de a spori efortul lor vocal atunci când vorbesc în zgomot puternic pentru a crește audibilitatea vocii lor. Aceasta constă în modificarea parametrilor semnalului vocal (creșterea energiei și F_0 , modificarea componentelor spectrale (crește valoarea lui F_1 și scade a lui F_2), ritm și durata silabelor. Această compensare are ca efect o creștere a raportului semnal-zgomot al cuvintelor rostite de vorbitor.

2. PRELUCRAREA SEMNALULUI VOCAL

- sunt 7 nivele pentru descrierea procesului vorbirii



The Hierarchy of Linguistic Units

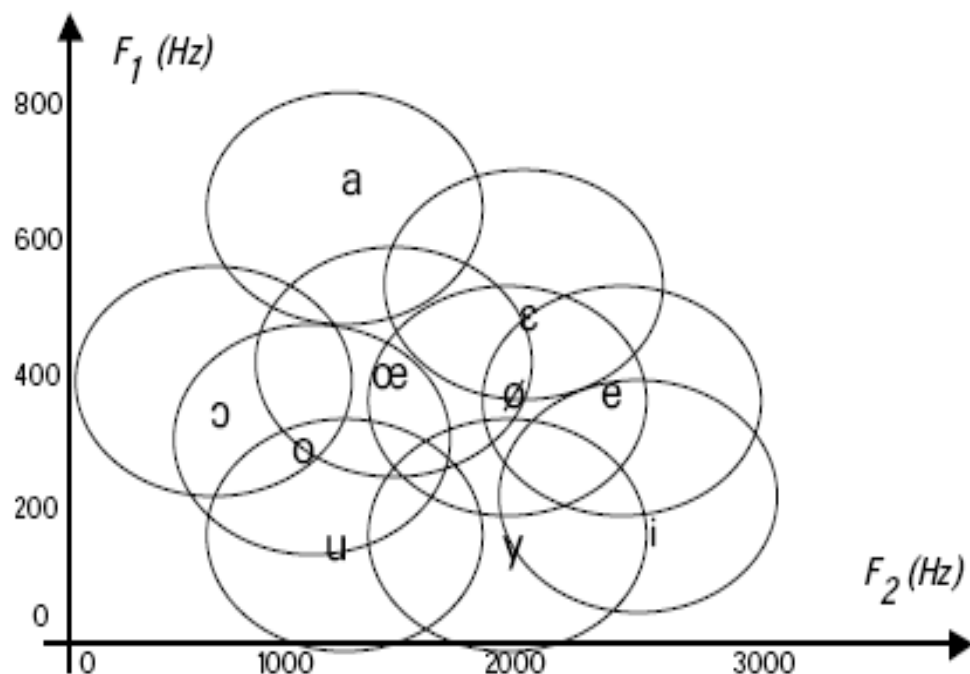
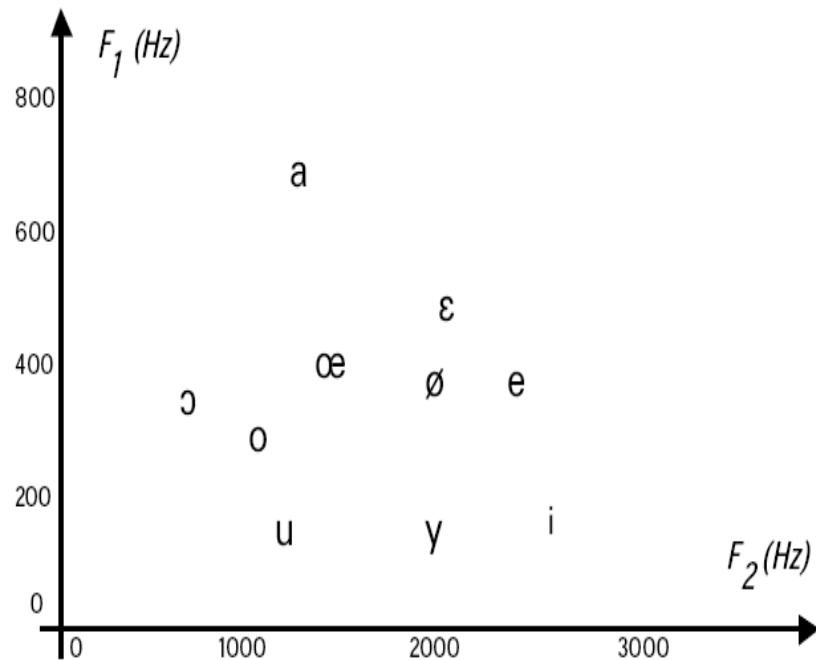


Morfemul este unitatea cea mai mică parte din structura unui cuvânt care poartă o informație. Morfemele se compun la rândul lor din foneme, dar un morfem nu este același lucru cu o silabă !!!

Un morfem poate fi un cuvânt rădăcină, un prefix sau un sufix.

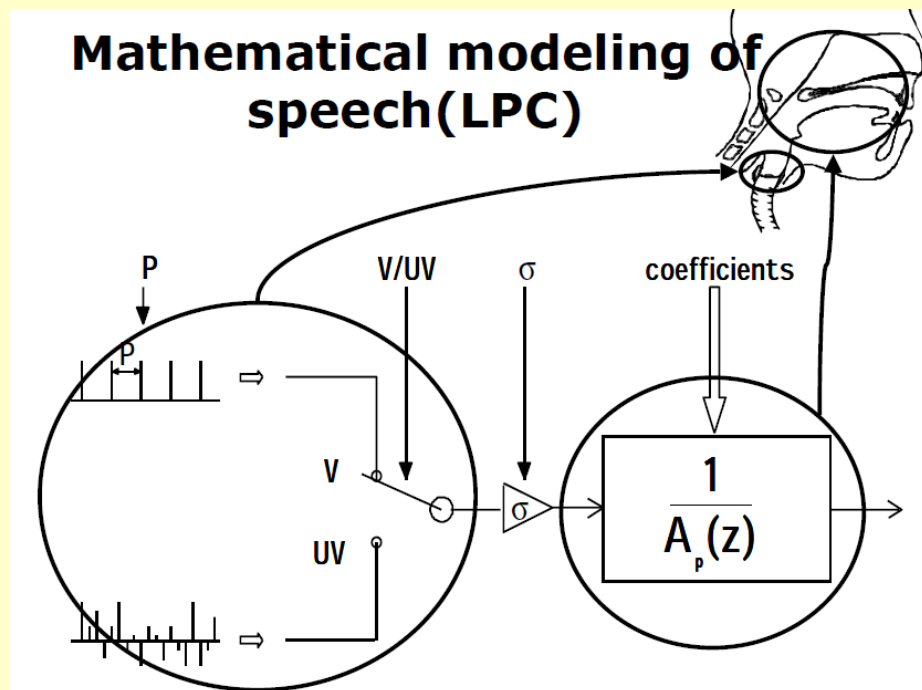
Ex. like, a+like, un+like, like+ly, like+lable

Coral = cor+al ; tribal = trib+al ; moral = mor+al...



- fonemele nu ocupa o regiune exclusiva in spatiul acustic, ci zone suprapuse – procesul de coarticulare
- **Sinteza** : trebuie sa fie capabila sa imite/reproduca coarticularea
- **Recunoasterea** : trebuie sa depaseasca/elimine coarticularea

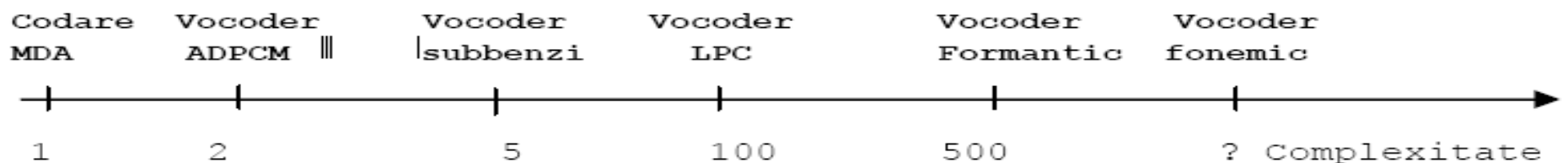
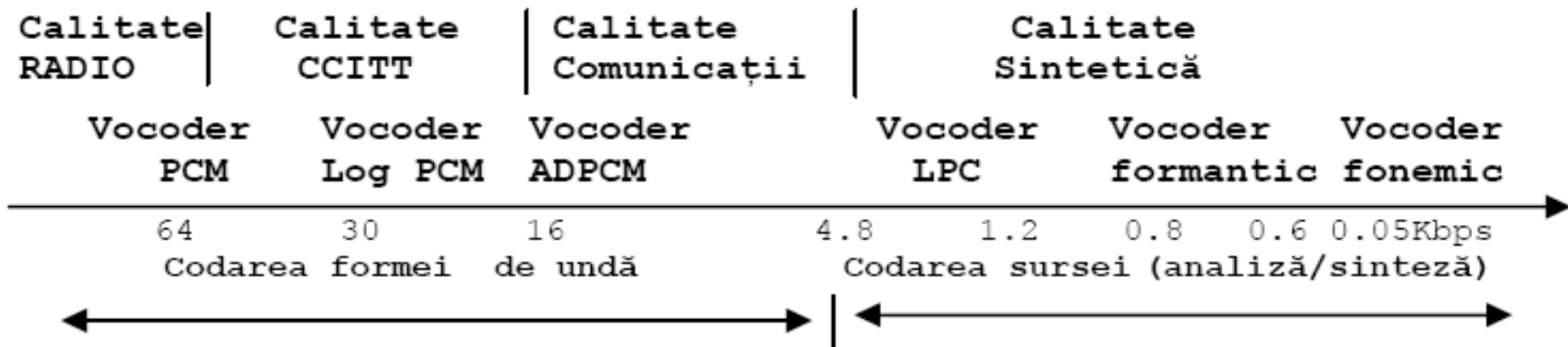
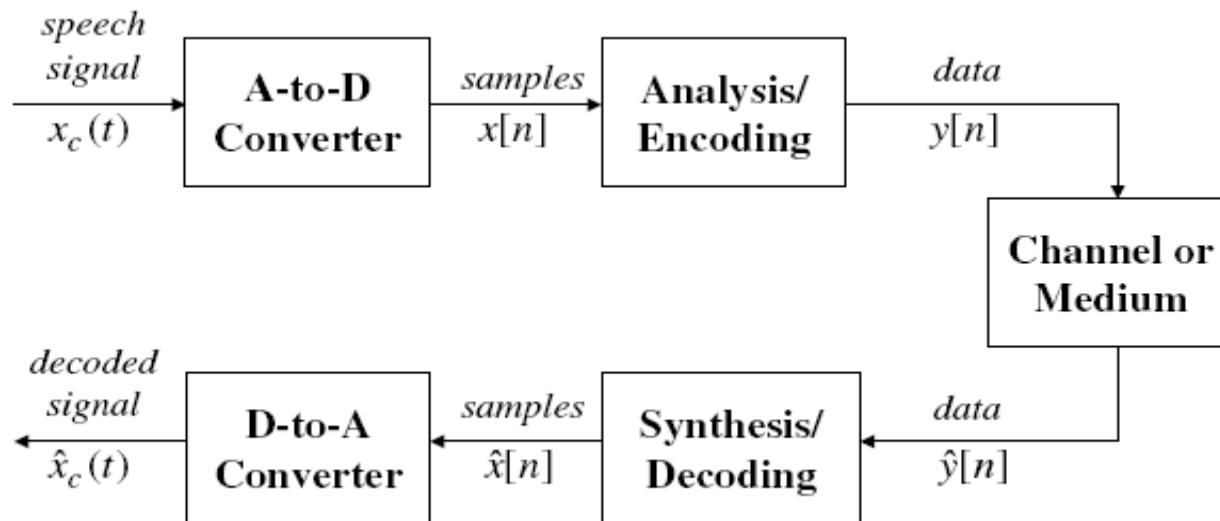
- Prelucrarea SV, foloseste uzual :
 - transformate (FFT, TZ, TWD,...) – nu tin seama de modul de productie a SV (abstracte)
 - prelucrari care separa sursa/tract (LPC, cepstru,...) – modele simplificate de productie a vorbirii
 - prelucrari pe baza modelelor de perceptie (vocoder)
 - Prin prelucrarea SV se incearca modelarea SV (asocierea unui model) prin masurare, parametrizare, observare

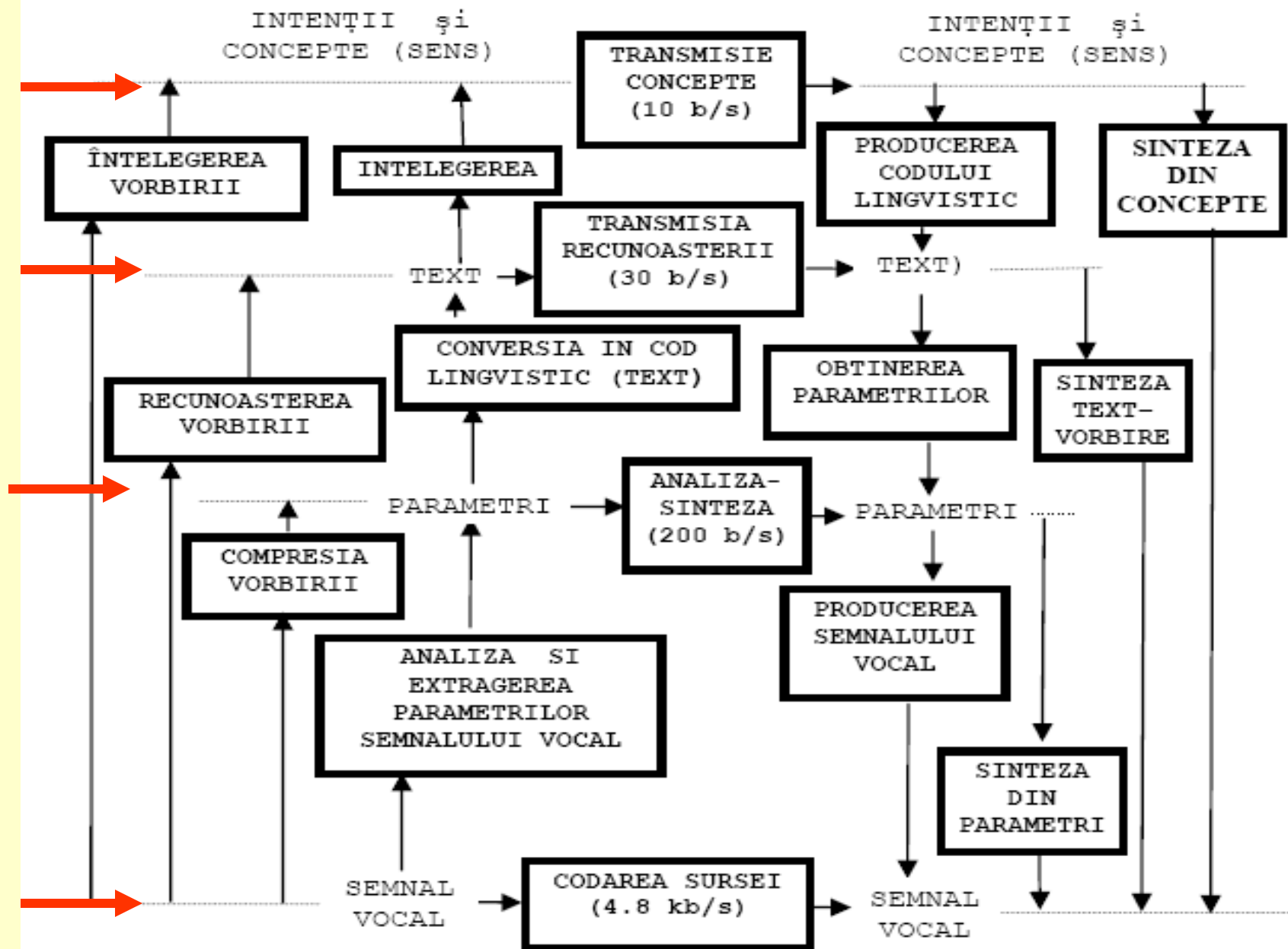


PRELUCRAREA SEMNALULUI VOCAL

1. **ANALIZA** care constă în preluarea, filtrarea, eșantionarea și prelucrarea eșantioanelor în vederea obținerii unor parametri și caracteristici acustice
2. **SINTEZA** realizează vocea artificială pe baza unor elemente specifice vorbirii (foneme, demisilabe, prozodie etc.) obținute în faza de analiză
3. **RECUNOAȘTEREA** vorbirii se face pe baza extragerii unor parametri și caracteristici ale semnalului vocal și utilizând algoritmi adecvați pentru determinarea cuvântului (fonemă, cuvânt, propoziție) rostit prin compararea cu niște modele/referințe
4. **COMPRESIA/CODAREA** semnalului vocal constă *în reducerea redundanței* semnalului vocal în vederea creșterii eficienței transmisiei sau stocării prin extragerea unor parametri și caracteristici semnificative

Codarea semnalului vocal





Principalele tehnologii de prelucrare a informațiilor vocale și relațiile între ele [după Fur01]

Digital Speech Processing Techniques

```
graph TD; A[Digital Speech Processing Techniques] --> B[Digital Transmission & Storage]; A --> C[Speech Synthesis]; A --> D[Speech Recognition]; A --> E[Speaker Verification/Identification]; A --> F[Enhancement of Speech Quality]; A --> G[Aids for the Handicapped];
```

**Digital
Transmission
& Storage**

**Speech
Synthesis**

**Speech
Recognition**

**Speaker
Verification/
Identification**

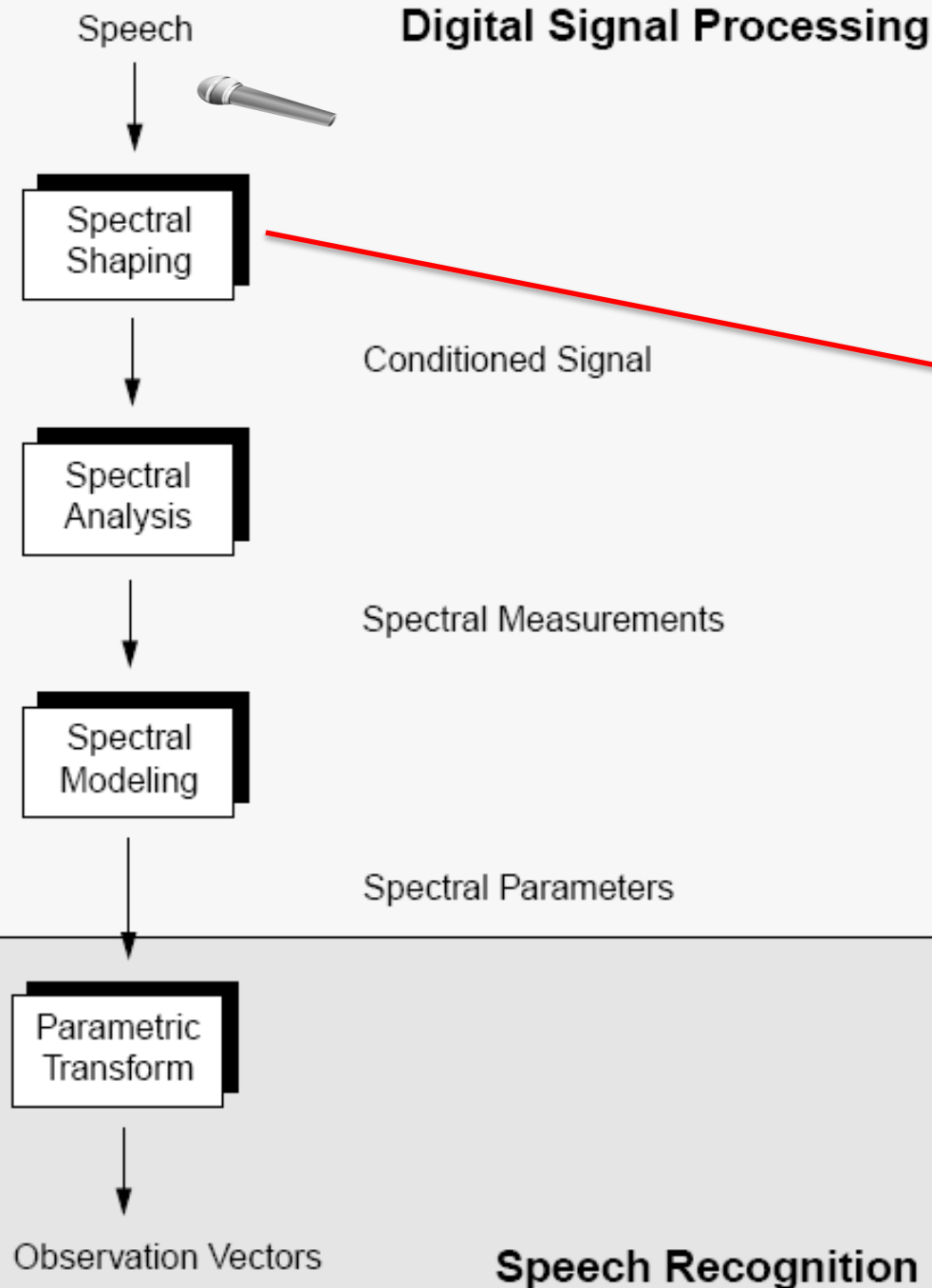
**Enhancement
of Speech
Quality**

**Aids for the
Handicapped**

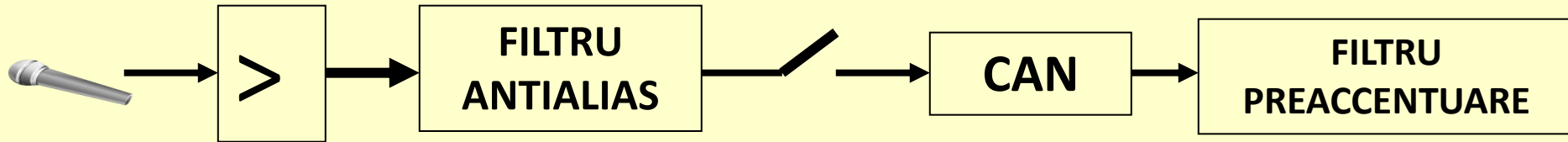
Aplicatii tipice ale prelucrarii semnalului vocal

EXEMPLU de prelucrare a SV ptr. recunoastere

1. Conditionarea semnalului
 - preluarea, esantionare, filtrare
2. Analiza spectrala (LPC, cepstrala, STFT, ...)
3. Prelucrarea parametrilor spectrali
4. Evaluare, clasificare



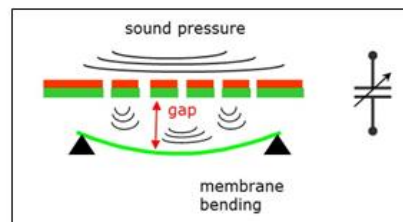
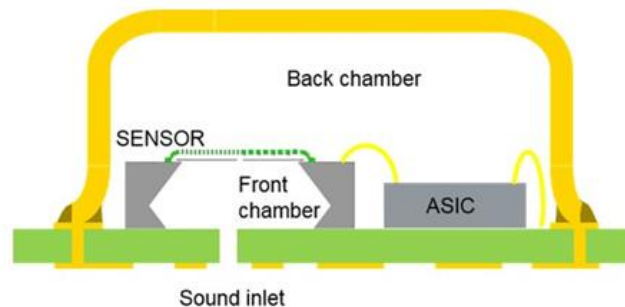
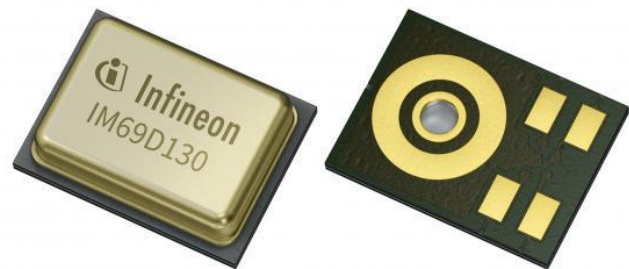
Preluarea si esantionarea SV



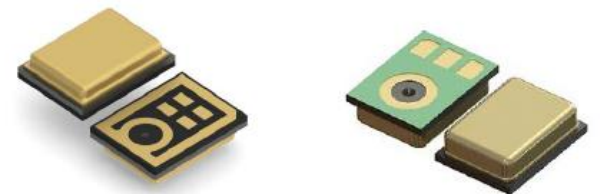
1. MICROFON - conversia sunet >> semnal electric ([Condenser](#), [Electret condenser](#), [Dynamic Ribbon](#), [Carbon](#), [Piezoelectric](#), [Fiber-optic](#), [Laser](#), [Liquid](#), [MEMS](#)) <http://en.wikipedia.org/wiki/Microphone>

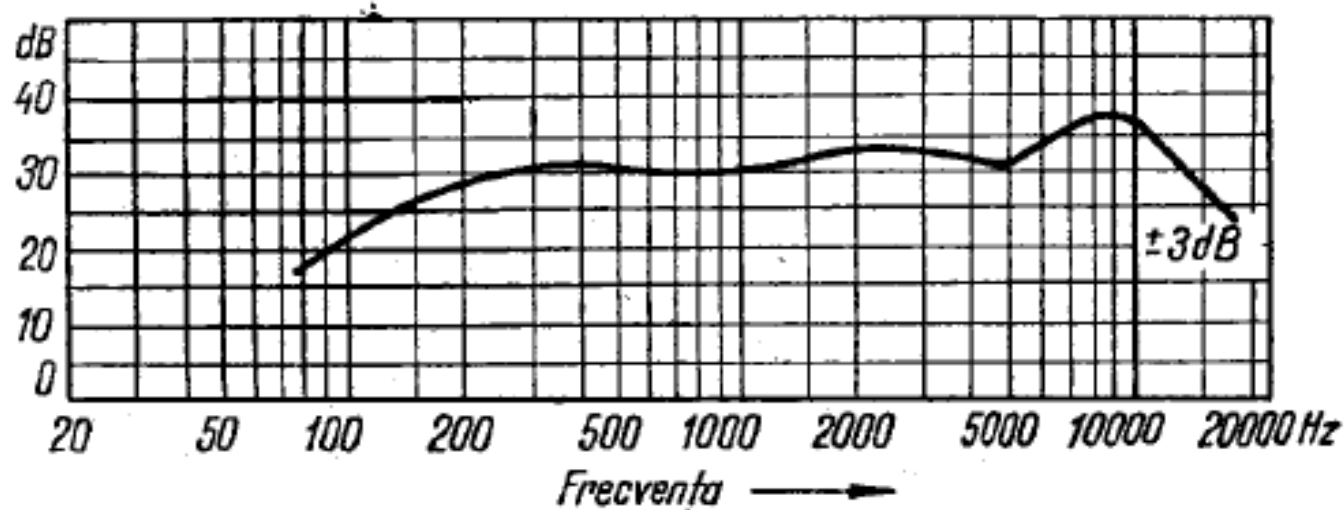
- caracteristica liniara ($\pm 4-6$ dB)
- Banda mare (50-12.000Hz) (electret/condensator/el.dinamice)
- Directivitatea (variatiia sensibilitatii cu directia) : sferica, cardioida, dublu-sferica
- Sensibilitate (0.2-1mV/ μ bar- bun)

(microelectromechanical systems)

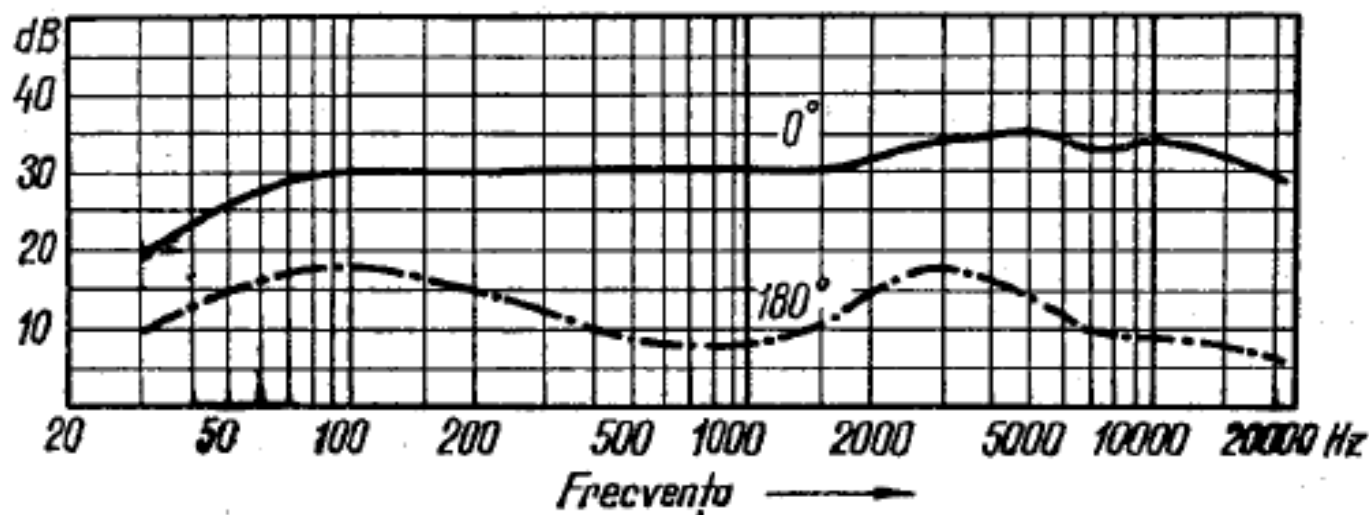


Microfon MEMS





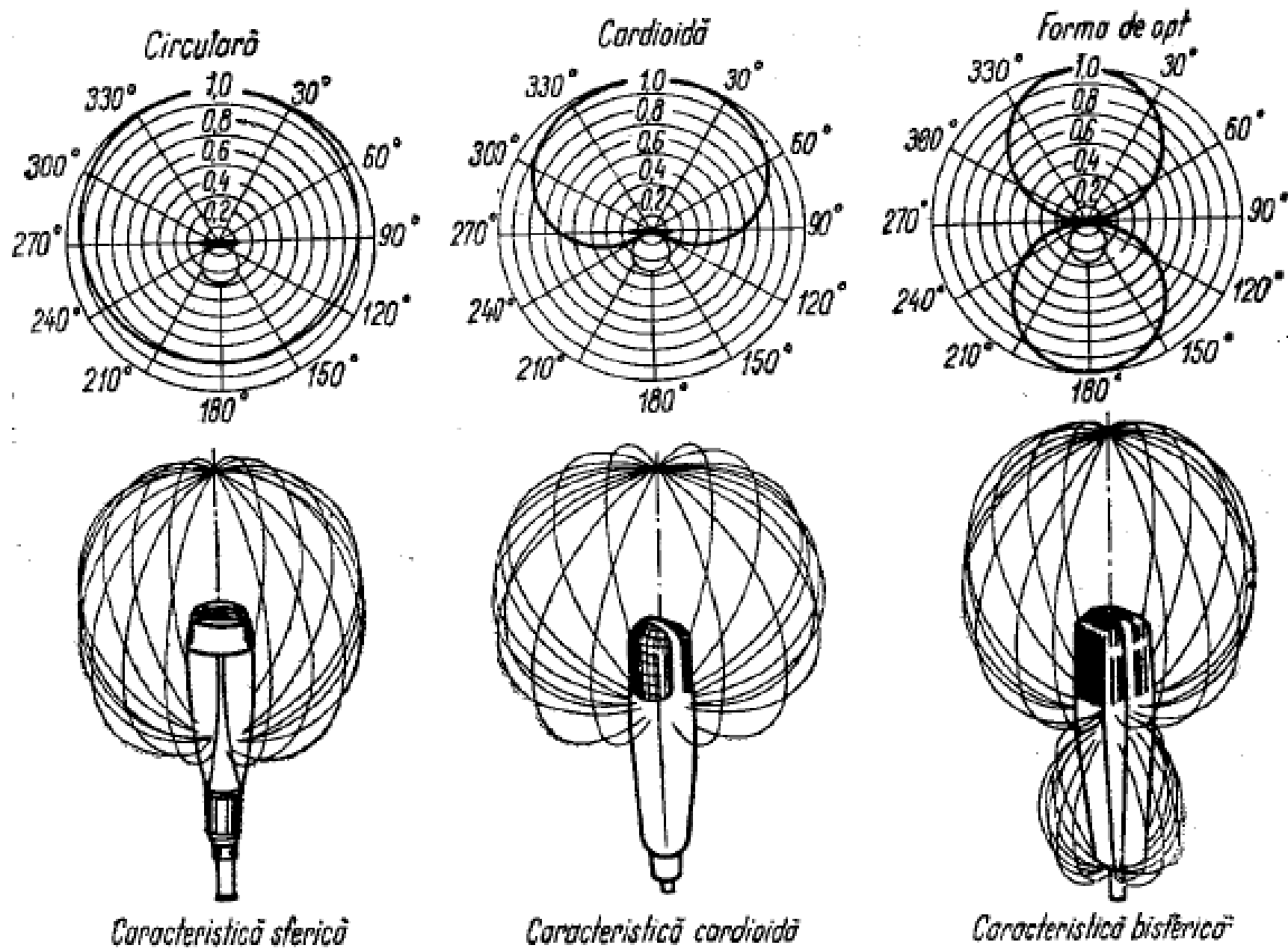
a



b

Figura 16 — Caracteristicile de frecvență ale microfoanelor cu electret.

a) microfon simplu din casetofon; b) microfon cu electret Hi-Fi.



Figură 5 — Compararea caracteristicii de directivitate a diferitelor microfoane.

2. PREAMPLIFICATOR – liniar si zgomot redus (381,387, TL082..)

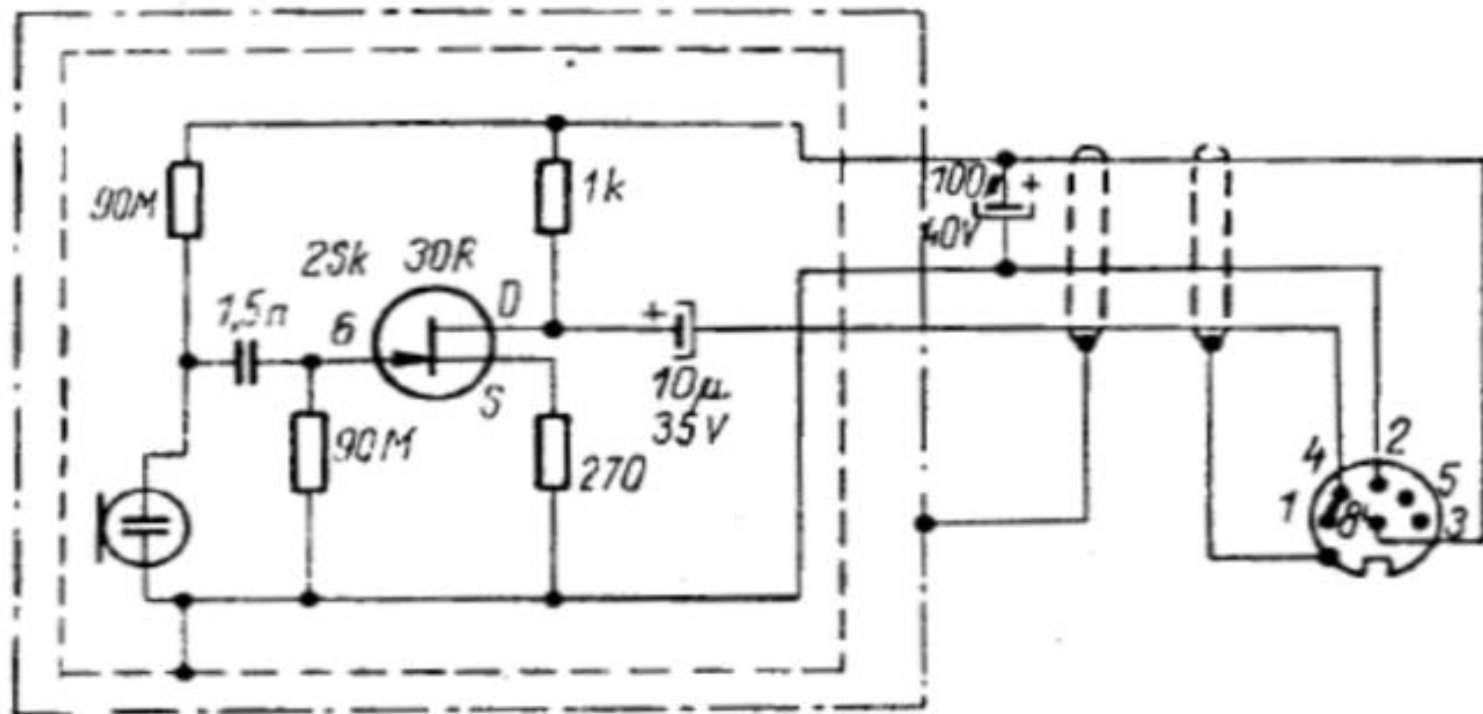
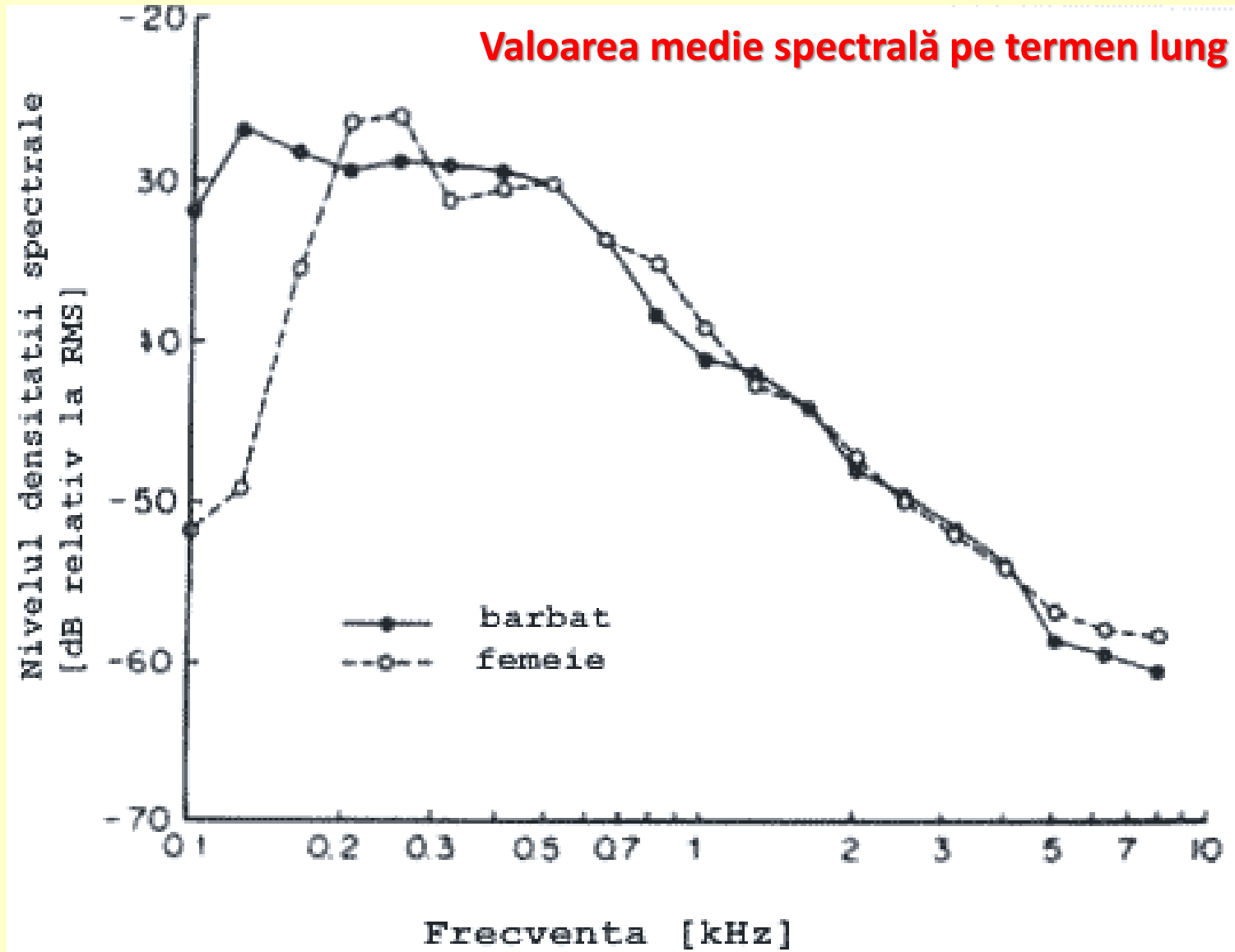


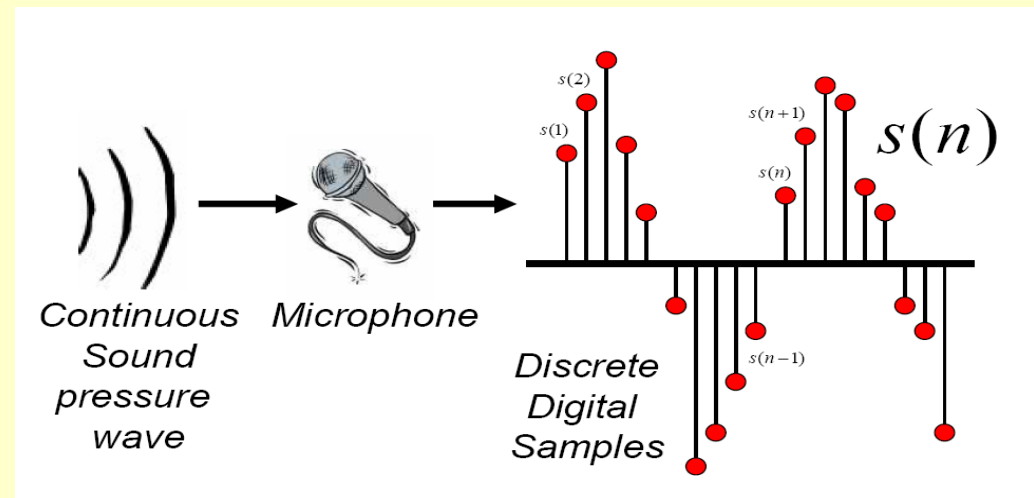
Figura 69 — Schema de principiu a unui preamplificator pentru microfon cu electret, încorporat în magnetofonul uzual cu casete.

3. Filtrarea antialias – SV 16-16000Hz => inteligibil, 0.3-3.4kHz, muzica =>20kHz

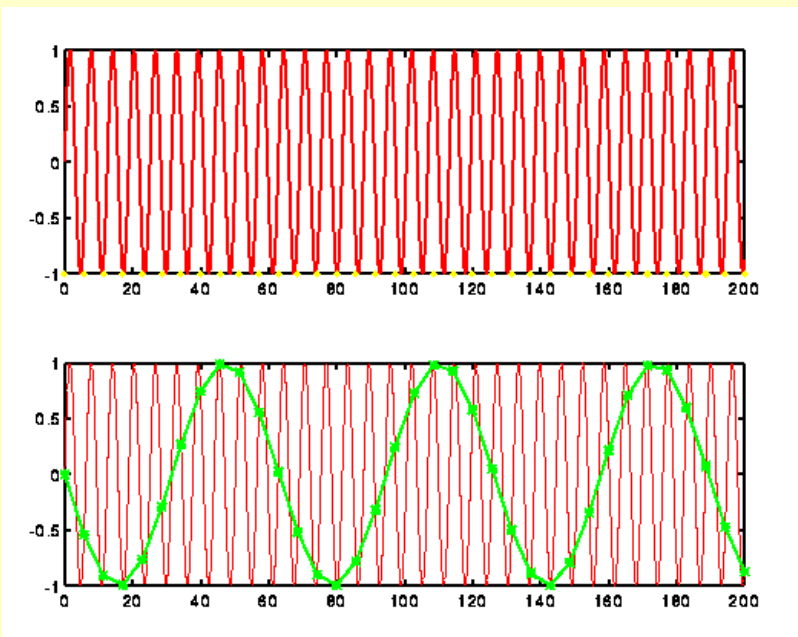
!!! Primii 3 FORMANȚI să intre in bandă



4. Eșantionarea SV — se ține cont de proprietățile SV (circuite de sample-and-hold -LF198...)

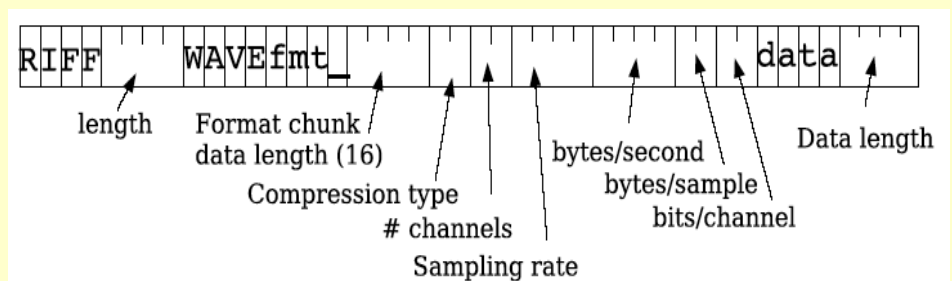


$B \leq 2 f_{\text{esantionare}}$



5. CAN - 8-16 biți/esantion, timp de conversie coresp., tip aprox. succesive, sigma-delta
- (cuantizarea se face $S/Z \geq 30\text{dB}$) – formatul :16 bits PCM, 8 bits u-law; compresie log

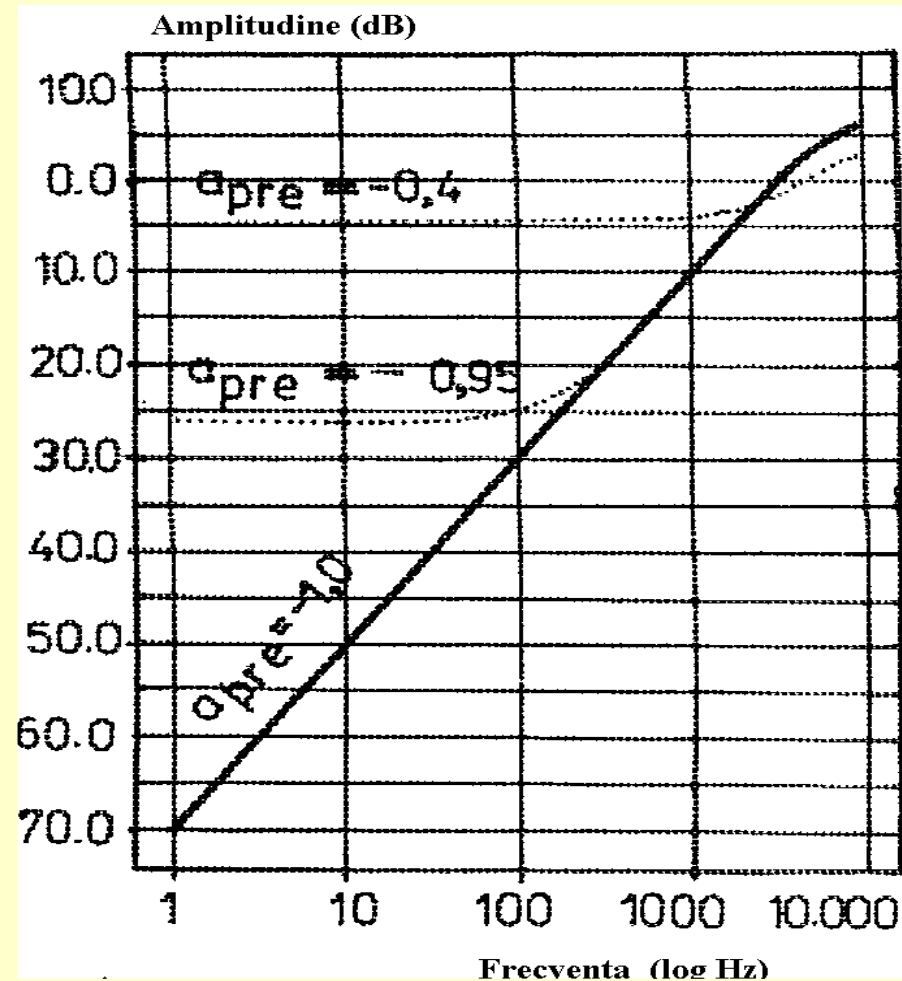
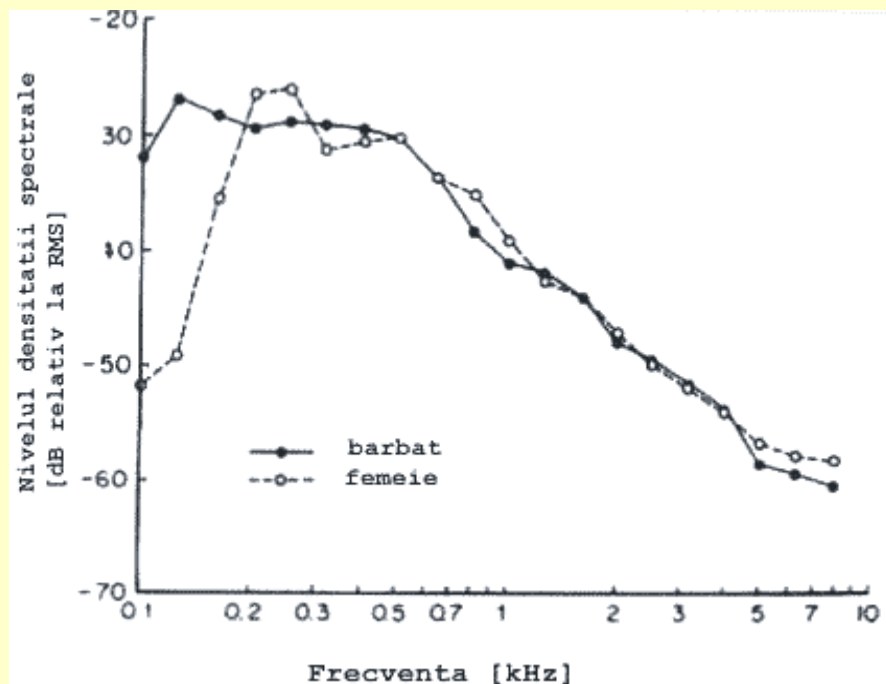
http://en.wikipedia.org/wiki/Analog-to-digital_converter



6. Preaccentuarea SV

- Elimină componenta continuă și crește ponderea frecvențelor din spectru $\geq 1\text{kHz}$ (FTS)
unde uzual $a = 0.9375$ sau 0.95

$$H(z) = 1 - a \cdot z^{-1}$$



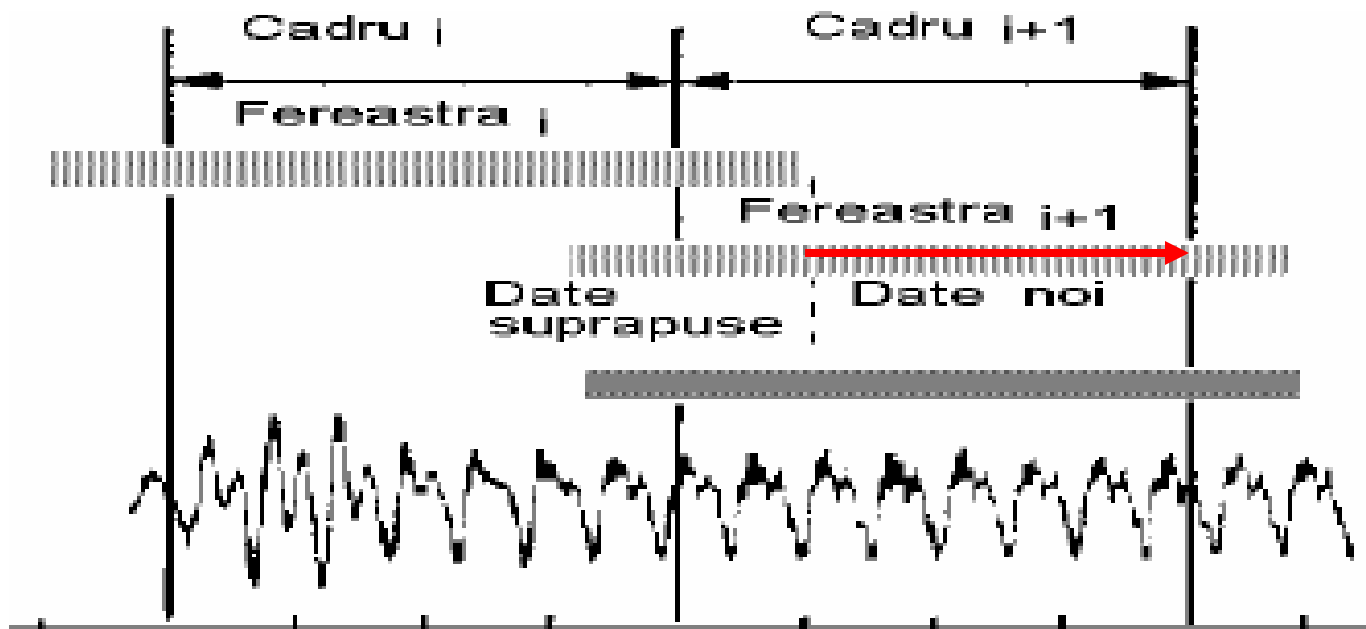
Caracteristica filtrului de preaccentuare funcție de "a"

..... Analiza spectrală etc

3. Tehnica ferestruirii

- ferestrele – functii pondere care scot in evidenta secvente finite de semnal
- SV este cvasistationar, *pe durata cadrelor parametrii ~ constanti*
- Rata modificarii parametrilor este rata cadrelor de analiza $8\text{ms} \leq T_c \leq 30\text{ms}$
- Durata cadrului este dependentă de viteza de articulare a sistemului fonator, adică de viteza de modificare a formei tractului vocal.

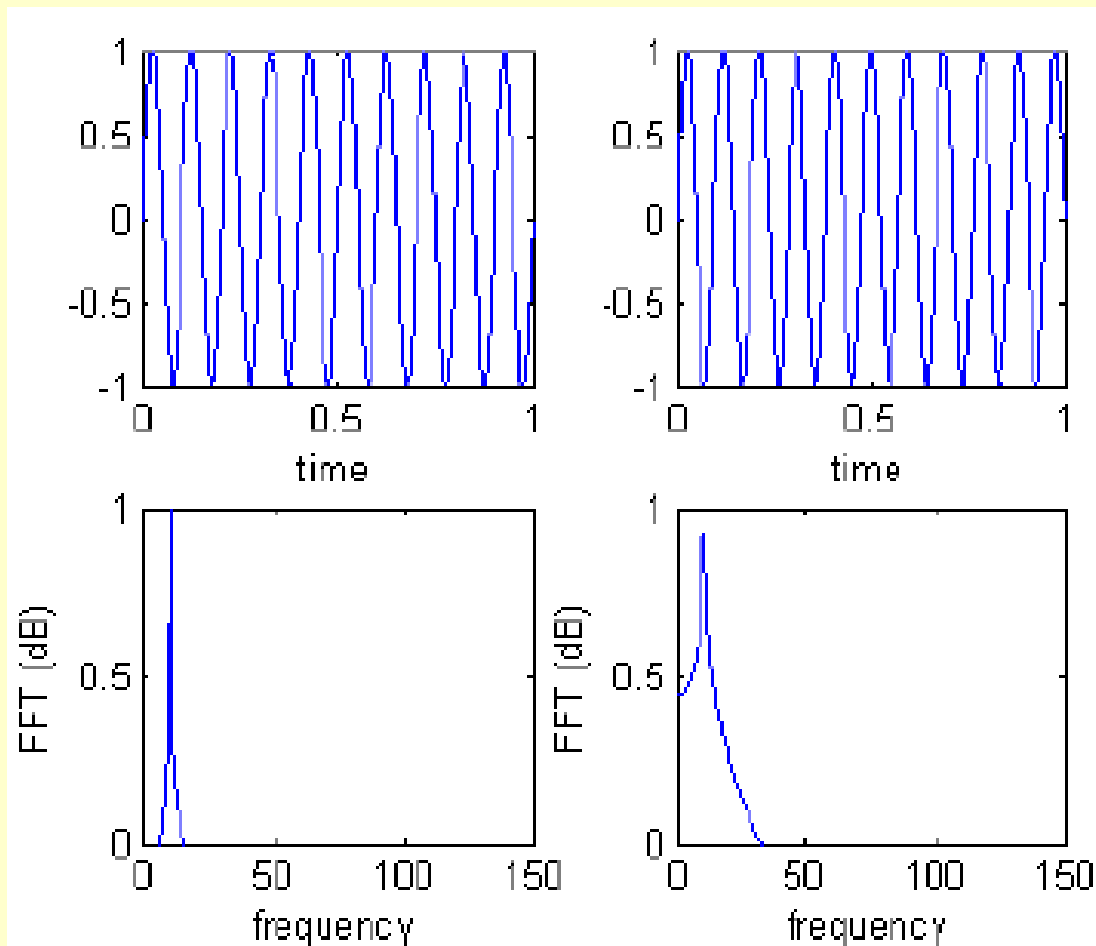
$$\% \text{ sup r} = \frac{T_f - T_c}{T_f} * 100\%$$

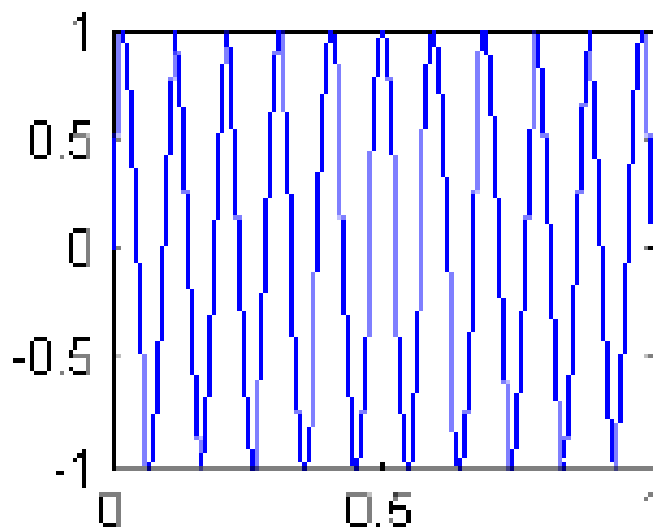


Utilizarea ferestrelor suprapuse în prelucrarea semnalului

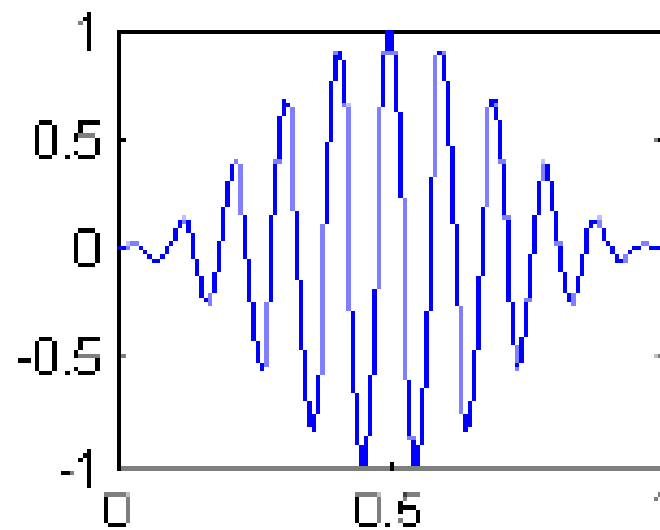
Scopul ferestruirii

- Analiza pe timp scurt se face pe blocuri de semnal (cadre)
- TFR (FFT) de ex. se face asupra acestor cadre și se bazează pe prezumția că semnalul este periodic (fig. stânga)
- Neperiodicitatea semnalelor cauzează reziduuri spectrale (leakage), adică spectrul semnalului se “dispersează” (fig. dreapta)

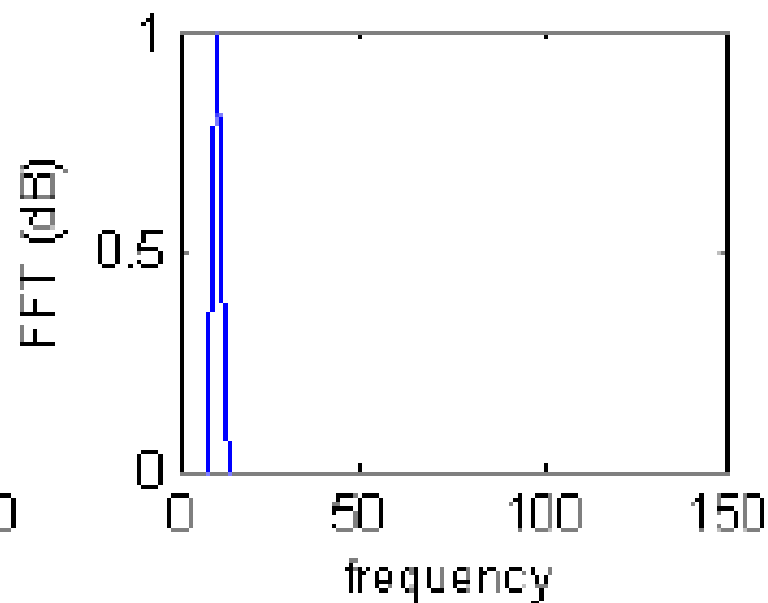
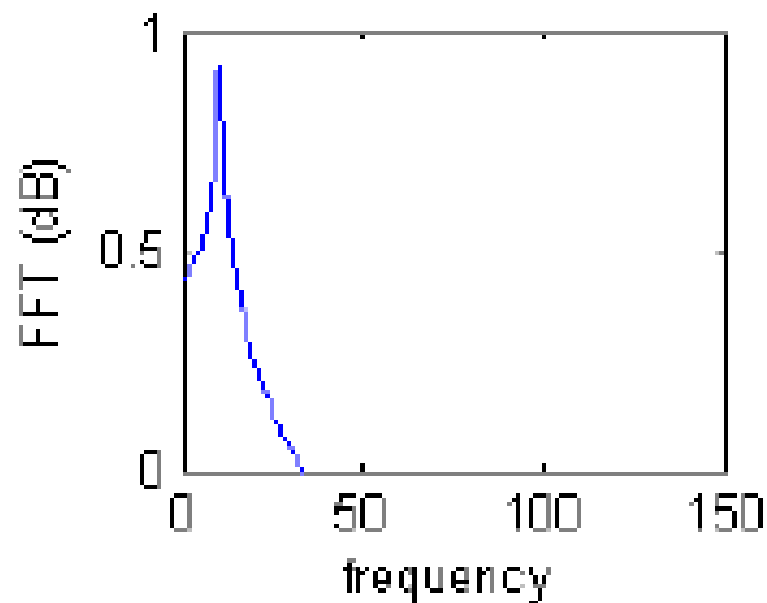




time



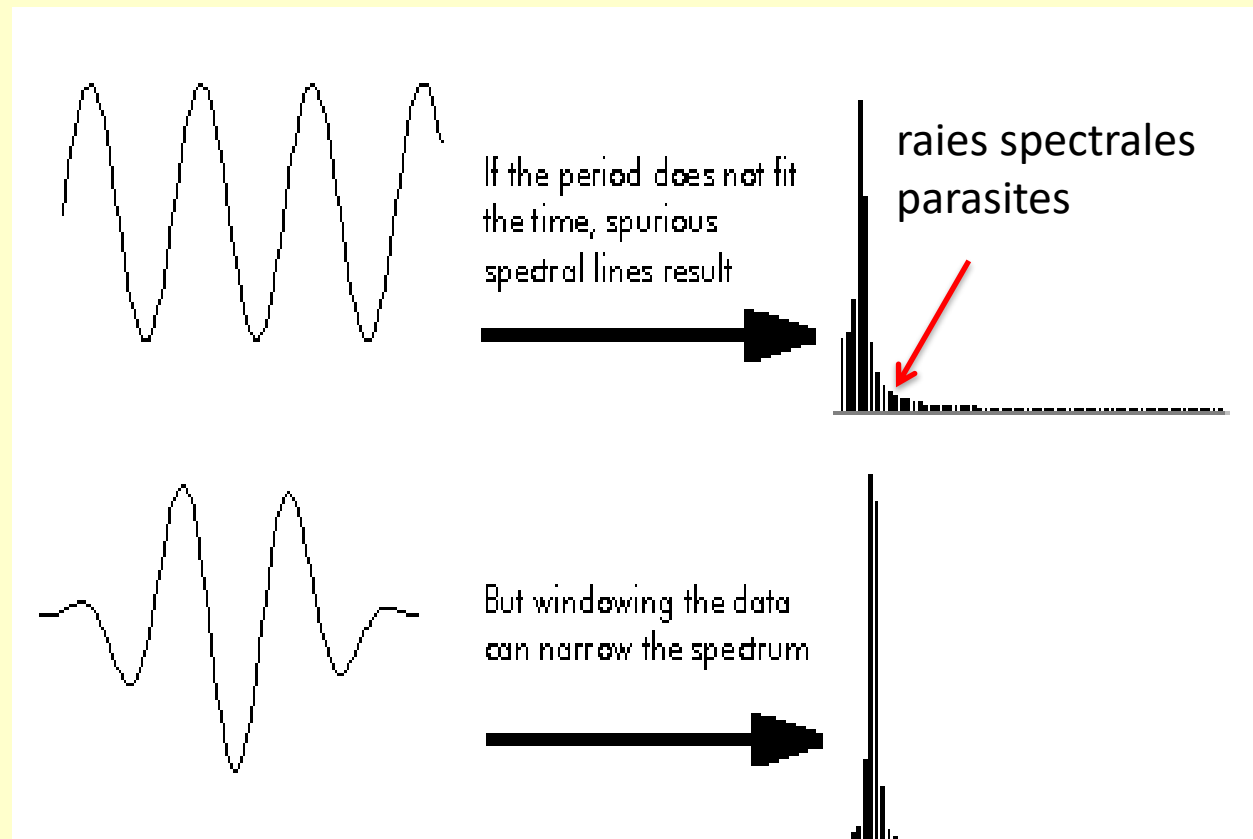
time



Comparația între un semnal sinusoidal neperiodic și spectrul său FFT și semnalul ferestruit și spectrul său FFT (dreapta) – reziduul se reduce

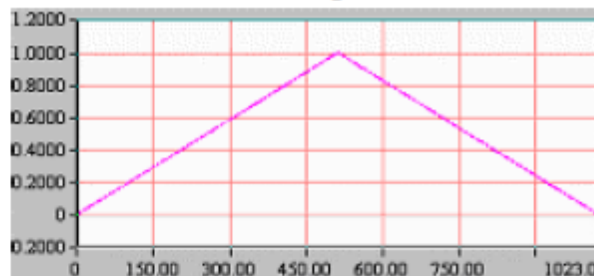
• Alegerea ferestrelor

- Ferestruirea elimină reziduul spectral, dar nu in totalitate, de fapt îi modifică forma
- fiecare tip de fereastră afectează spectrul in mod diferit
- Unele ferestre sunt mai potrivite pentru semnale aleatoare, altele pentru cele sinusoidale, iar alte ferestre au o rezoluție ridicată în frecvență



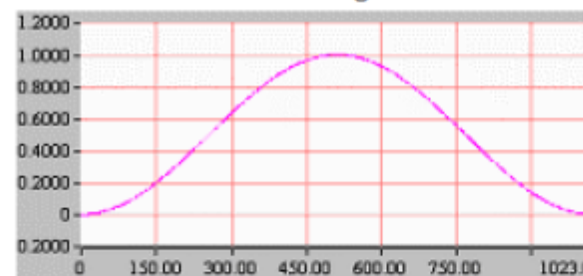
Bartlett

Highest Side lobe: -26dB



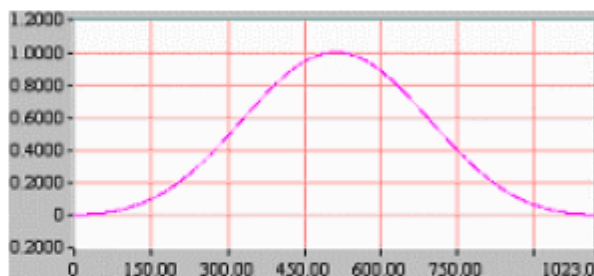
Hanning

Highest side lobe: -32 dB



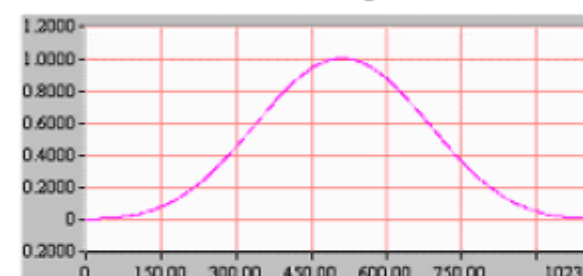
Blackman

Highest side lobe: -74 dB



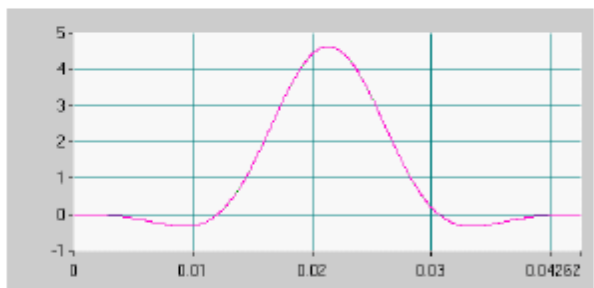
Kaiser-Bessel

Highest side lobe: -70 dB



Flat Top

Highest side lobe: -93dB



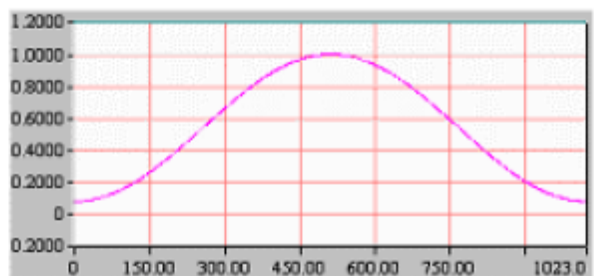
Tukey

Highest side lobe: -13 dB

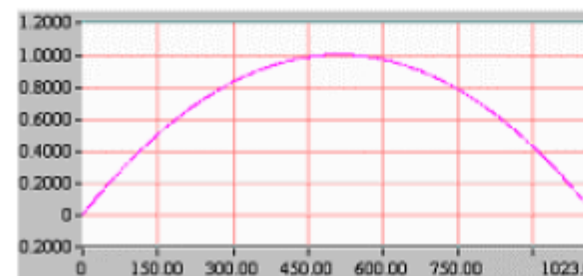


Hamming

Highest side lobe: -42dB



Welch-Highest side lobe: -21dB



Forma diferitelor ferestre folosite in domeniul timp

1. Fereastra rectangulară :

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

2. Fereastra Hamming :

$$w(n) = \begin{cases} 0.54 - 0.46 \cdot \cos(2\pi \frac{n}{N-1}) & , 0 \leq n \leq N-1 \\ 0 & , \text{în rest} \end{cases}$$

3. Fereastra Hanning :

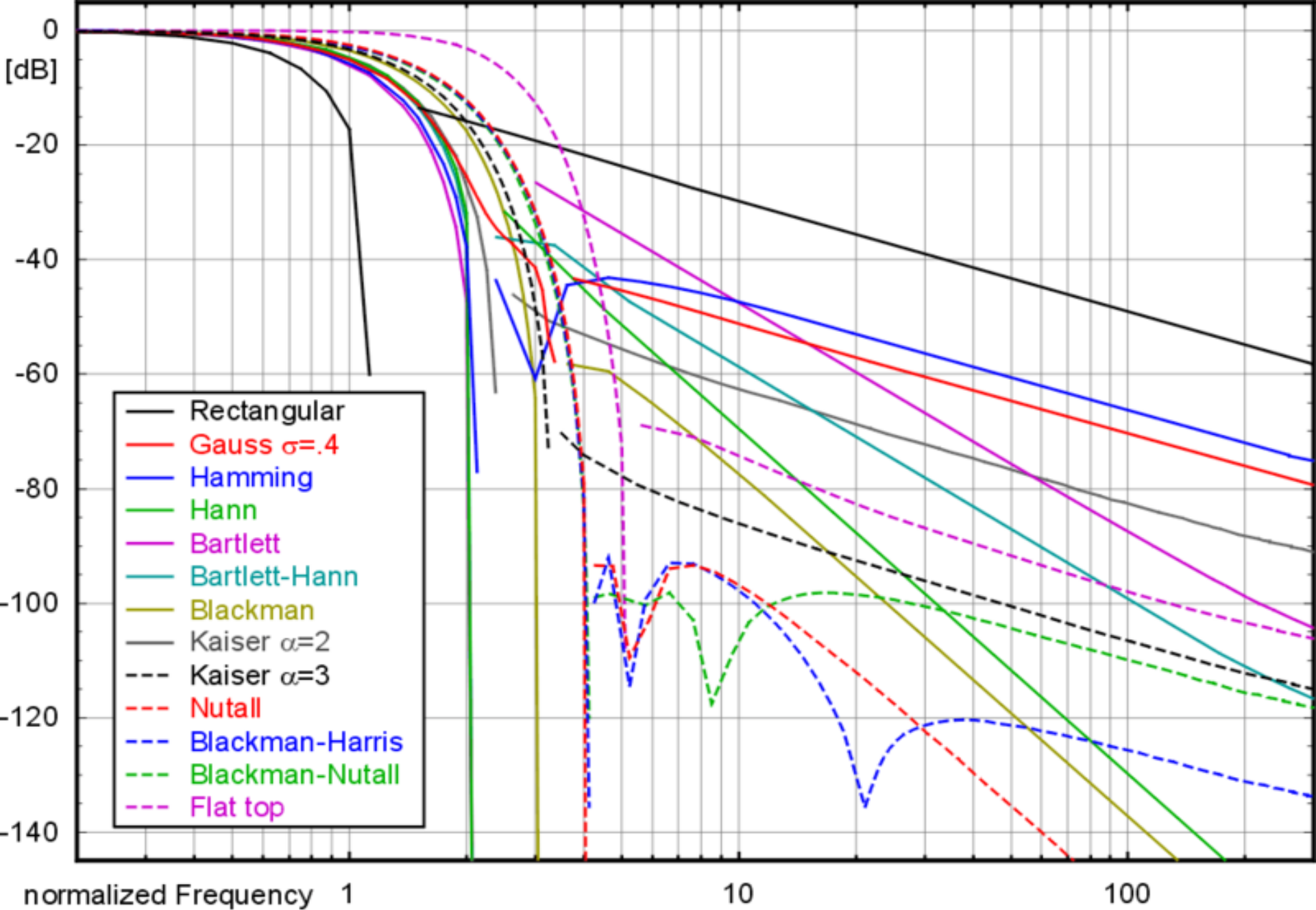
$$w(n) = \begin{cases} 0.5 - 0.5 \cos \frac{2\pi n}{N-1} & , 0 \leq n \leq N-1 \\ 0 & , \text{în rest} \end{cases}$$

4. Fereastra Barlett

$$w(n) = \begin{cases} \frac{2n}{N-1} & , 0 \leq n \leq \frac{N-1}{2} \\ 2 - \frac{2n}{N-1} & , \frac{N-1}{2} \leq n \leq N-1 \end{cases}$$

5. Fereastra Blackman :

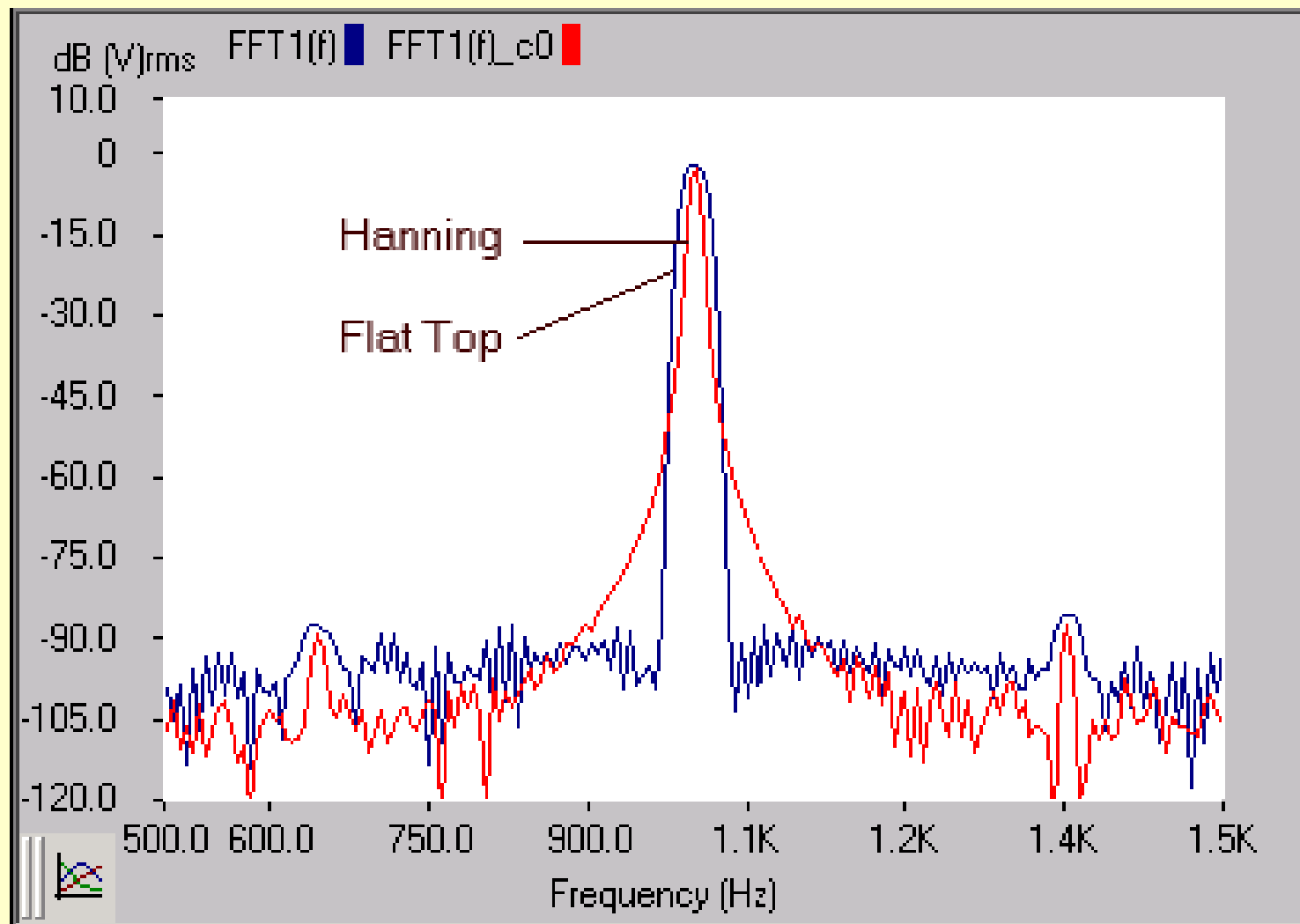
$$w(n) = \begin{cases} 0.42 - 0.5 \cos \frac{2\pi n}{N-1} + 0.08 \cos \frac{4\pi n}{N-1} & , 0 \leq n \leq N-1 \\ 0 & , \text{în rest} \end{cases}$$



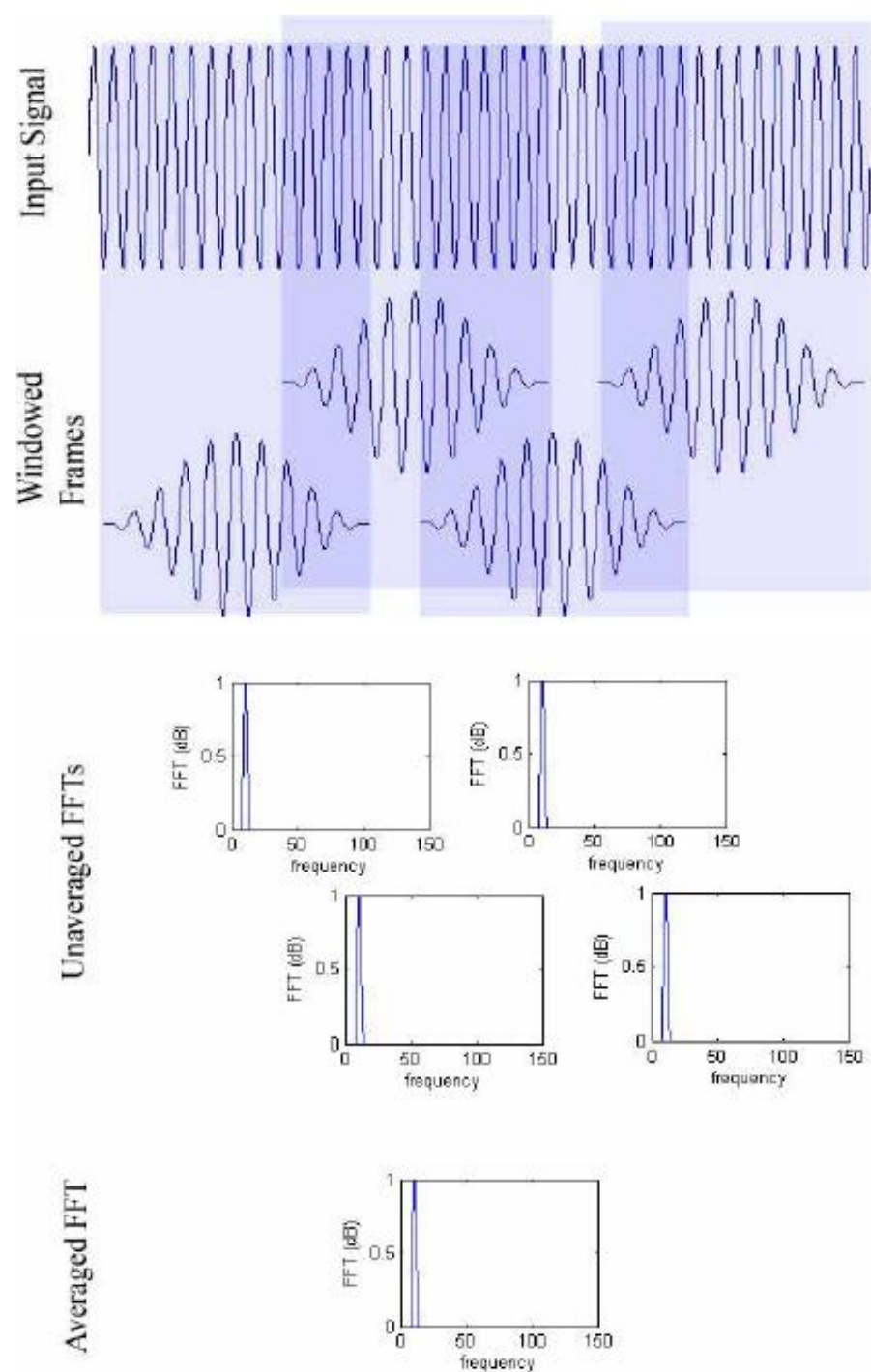
Atenuarea diferitelor ferestre in banda de oprire

Caracteristici ale unor ferestre uzuale

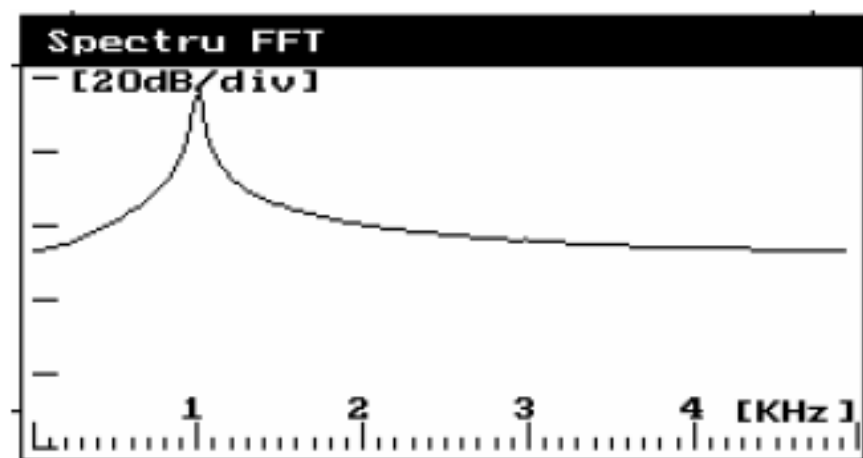
Window	Best for these Signal Types	Frequency Resolution	Spectral Leakage	Amplitude Accuracy
Barlett	Random	Good	Fair	Fair
Blackman	Random or mixed	Poor	Best	Good
Flat top	Sinusoids	Poor	Good	Best
Hanning	Random	Good	Good	Fair
Hamming	Random	Good	Fair	Fair
Kaiser-Bessel	Random	Fair	Good	Good
None (boxcar)	Transient & Synchronous Sampling	Best	Poor	Poor
Tukey	Random	Good	Poor	Poor
Welch	Random	Good	Good	Fair



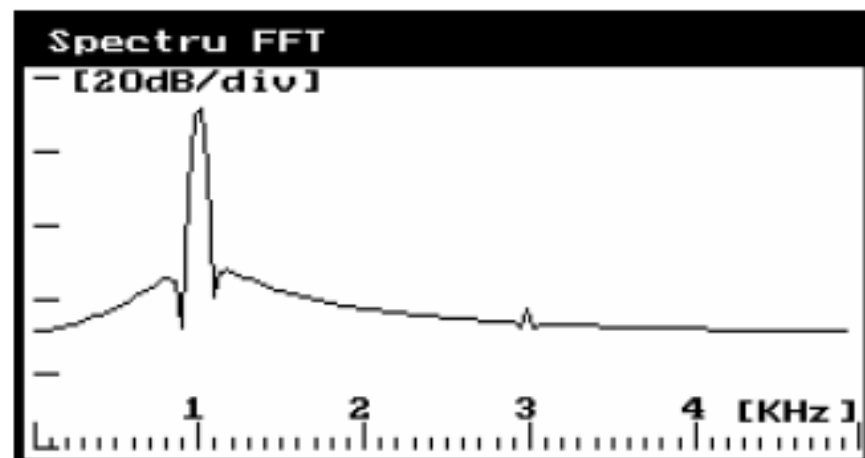
Comparatie intre functiile de ferestruire Hanning si Flat-Top pentru un semnal sinus ne-periodic



- Procesul de suprapunere scurtează timpul de achiziție, prin refacerea unei porțiuni din cadrul anterior care ar fi fost pierdut datorită ferestruirii

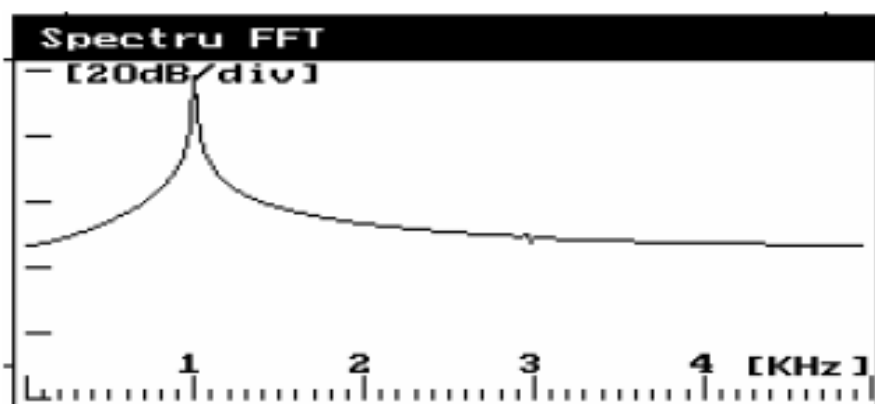


(a)

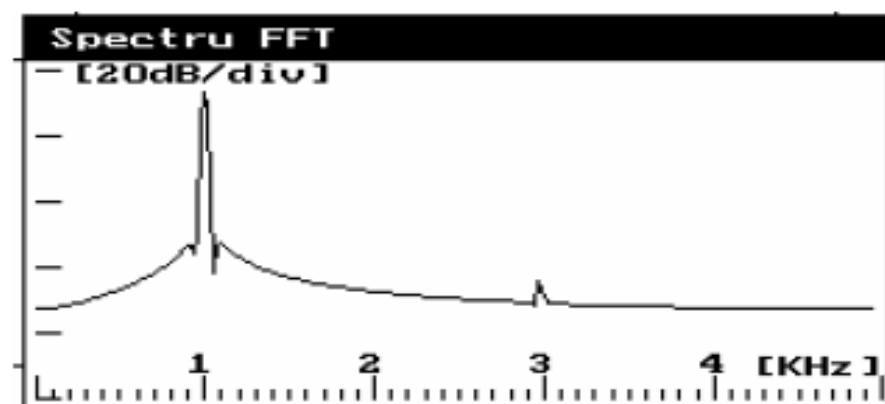


(b)

Spectrul unui semnal sinusoidal ferestruit cu fereastră dreptunghiulară (a) si Hamming (b) (N=256)



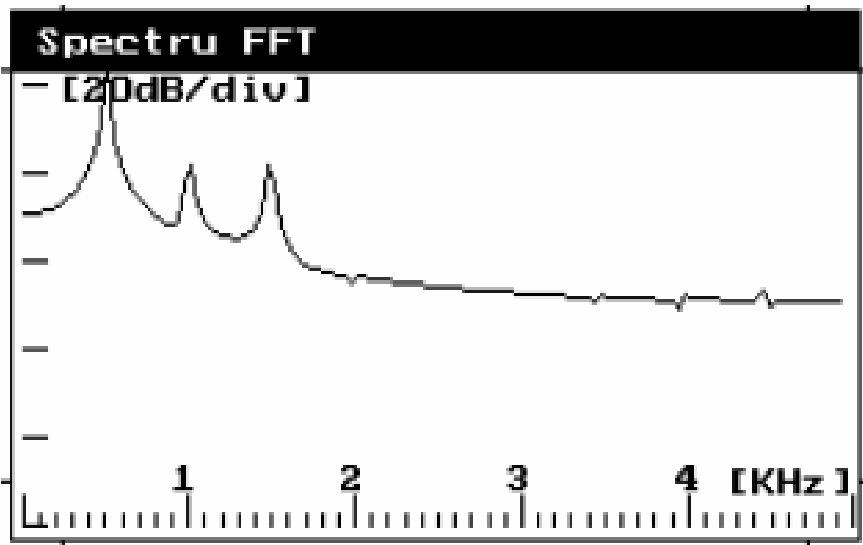
(a)



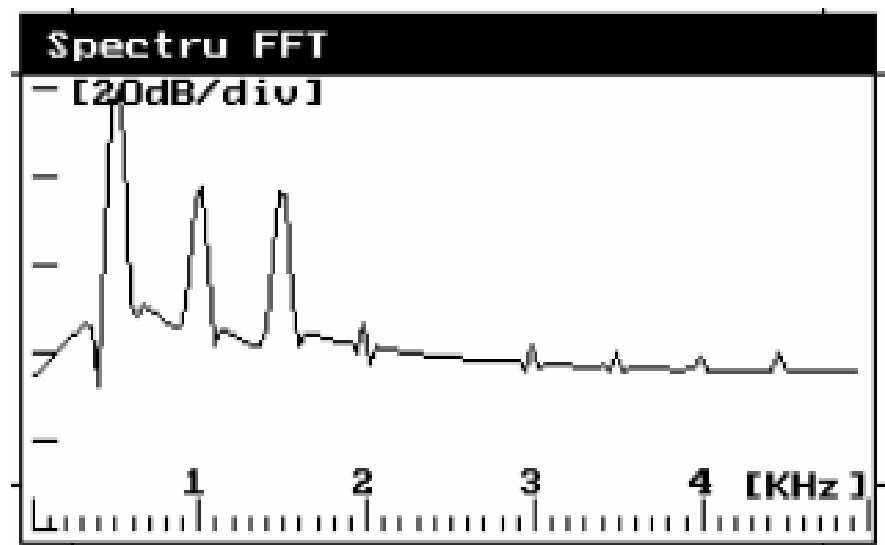
(b)

Spectrul unui semnal sinusoidal ferestruit cu fereastră dreptunghiulară (a) si Hamming (b) (N=512)

$$s(t) = \sin 2\pi ft + 0.1\sin 4\pi ft + 0.1\sin 6\pi ft \quad \text{unde } f=500 \text{ Hz}$$

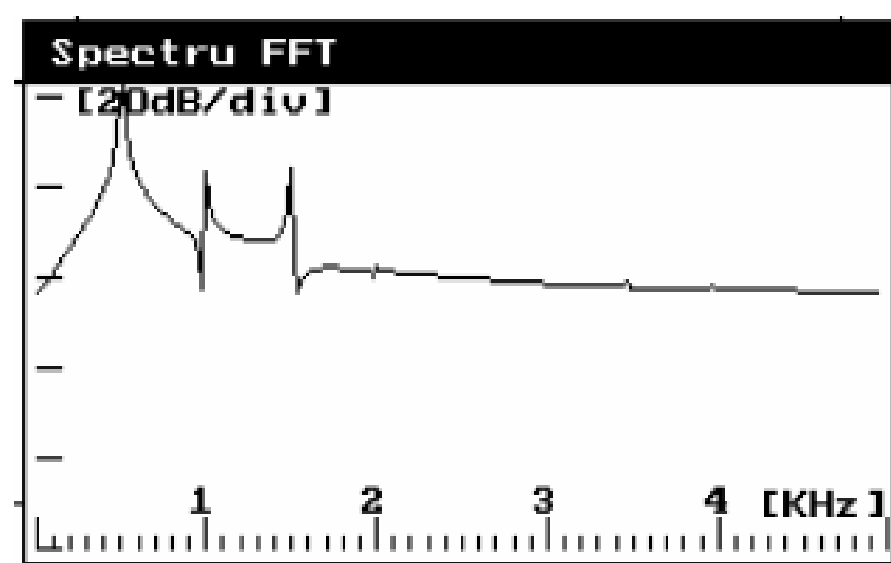


(a)

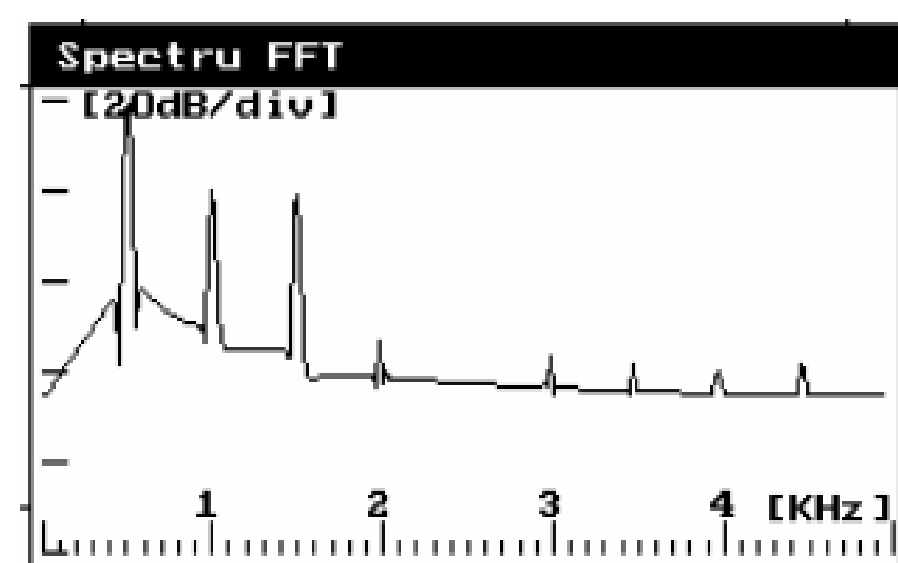


(b)

**Spectrul unui semnal sintetizat ferestruit
cu fereastră dreptunghiulară (a) si Hamming (b) (N=256)**



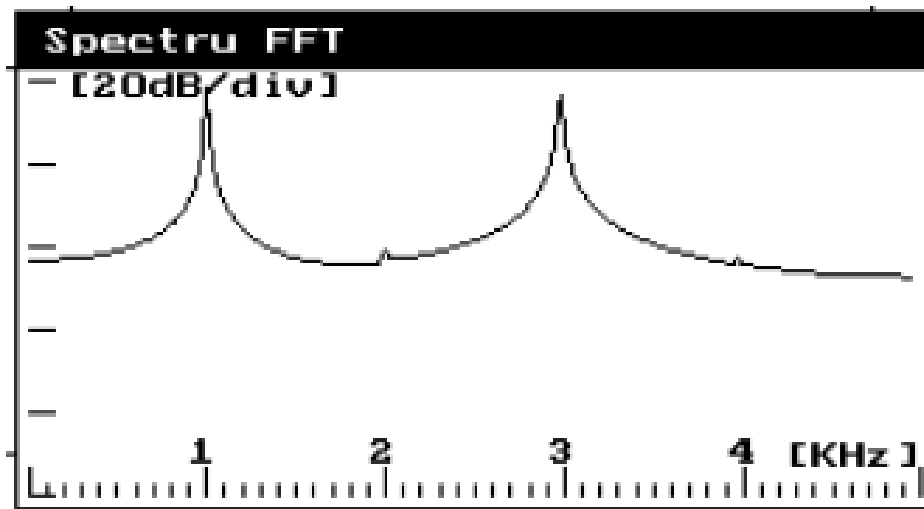
(a)



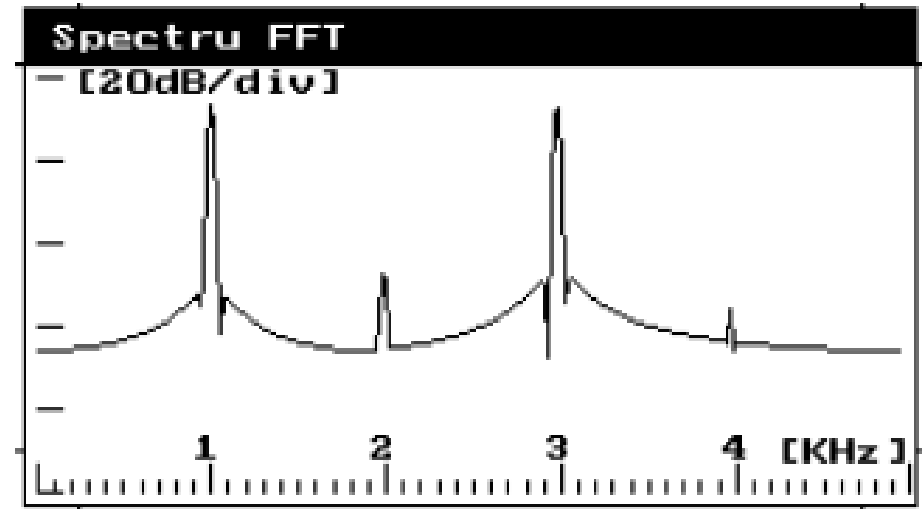
(b)

**Spectrul unui semnal sintetizat ferestruit cu
fereastră dreptunghiulară (a) și Hamming (b) (N=512)**

$$s(t) = 0.5\sin 2\pi ft + 0.005\sin 4\pi ft + 0.5\sin 6\pi ft, \text{ unde } f=1\text{kHz.}$$



(a)

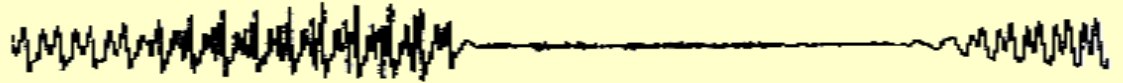


(b)

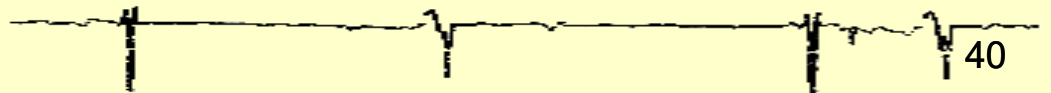
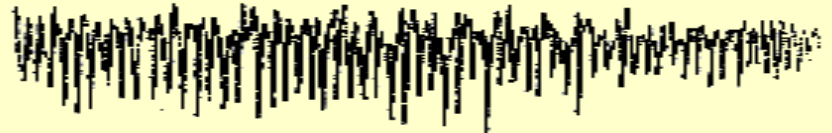
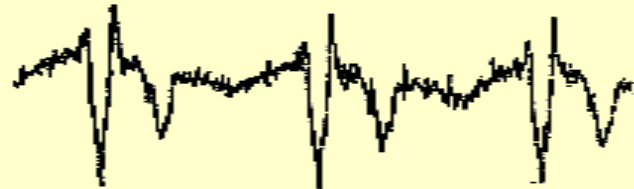
**Spectrul unui semnal sintetizat (3.3.10.) ferestruit
cu fereastră dreptunghiulară (a) si Hamming (b) (N=512)**

Concluzii

- **Vorbirea** – facultatea de a comunica gândurile printr-un sistem de sunete articulate emise de organele fonatoare sau expresii verbale ale gândurilor



- Cateva segmente de semnal



Cateva proprietati ale sV

- SV este quasi-stationar pe termen scurt, cu mare redundanta
- SV este $\sim 70\%$ din timp (quasi-)periodic (zgomot sau liniste in rest)
- SV este un semnal de banda larga (ocupa toata latimea de banda)
- SV este un semnal de banda limitata (50-8000 Hz)

Speech synthesis

- Synthesis by rule
- Text-to-speech

Speech recognition /understanding

- Speaker-independent
- Spontaneous speech

Speech coding

- Wide/narrow-band
- Very-low-bit-rate

Robustness

- Noise/distortion

Human-machine interface

- Ergonomics
- Subjective/objective evaluation

Individuality

- Speaker recognition
- Speaker adaptation/normalization
- Voice conversion

Database

Feature extraction
(dynamics)

Database

Psychology

Speech perception

Physiology

Nerve system

Speech production

Articulation

Acoustic phonetics

Memory/learning

Speech analysis

Artificial Intelligence

Signal processing

Acoustics