Numérisation et codage de l'information
Codage de la parole

Licence ISVD

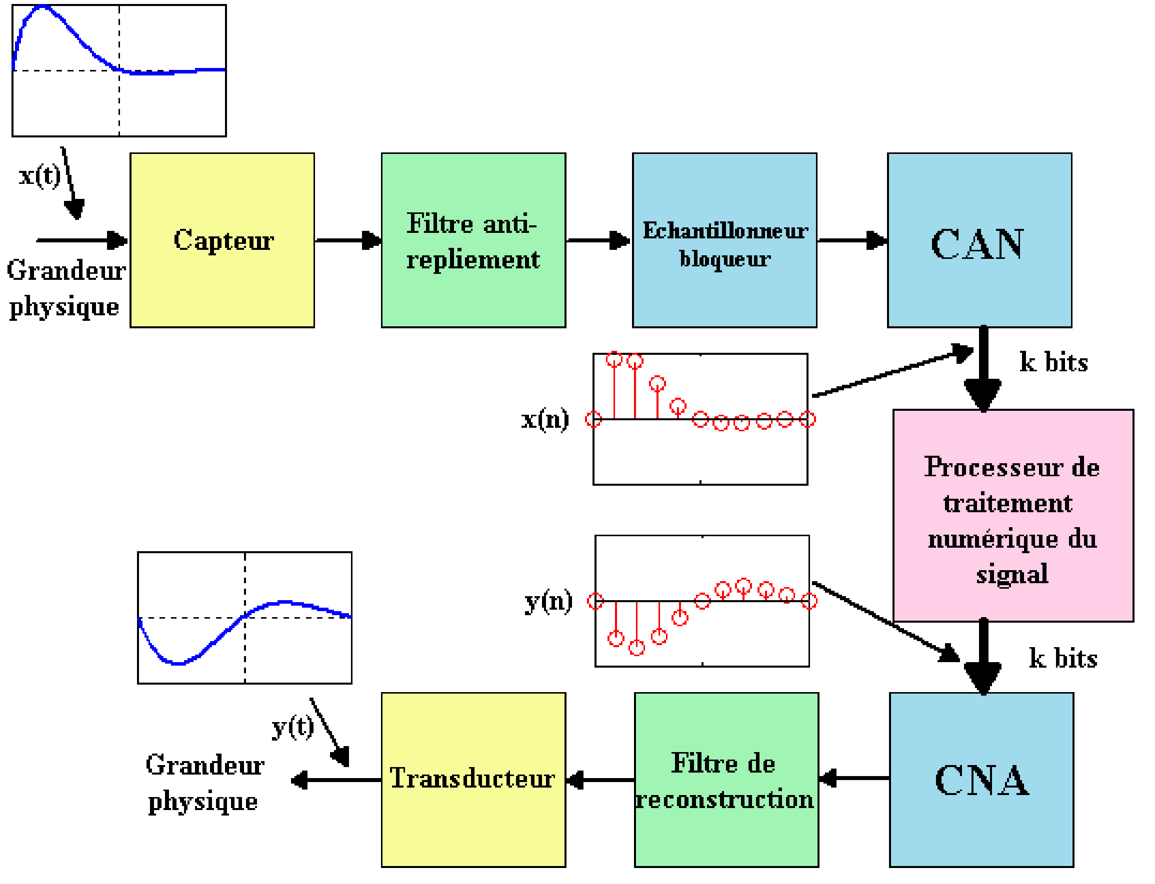
Hervé BOEGLEN

Plan

- 1. Notions de base
- 2. Théorie de l'information
- 3. Quantification
- 4. Codage de la parole
  - Qu'est-ce que la parole ?
  - Modélisation LPC
  - Codeurs en forme d'onde et par analyse et synthèse (ABS)
  - Standards

1. Notions de base

- Le traitement numérique du signal (TNS) :



1. Notions de base

- Description d'un système numérique :
  - Dans le domaine temporel par une équation aux différences, par exemple :

$$y[n] = b_0x[n]+b_1x[n-1]+a_1y[n-1]$$

  - Dans le domaine de la transformée en z, par une fonction de transfert, par exemple :

$$H(z)=\frac{b_0+b_1z^{-1}+b_2z^{-2}}{1+a_1z^{-1}+a_2z^{-2}}$$

1. Notions de base

- La transformée en z est la transformée de Fourier d'un signal échantillonné :

$$X(z) = \sum_{n=0}^{+\infty} x[n] \cdot z^{-n}$$

- C'est un outil incontournable pour l'étude des signaux numérisés. En pratique, on utilise une table de transformées.

1. Notions de base

Les opérations de base du TNS :

- Le filtrage : il existe deux types de filtres numériques :
  - Les filtres FIR (Finite Impulse Response) qui sont obtenus en échantillonnant la réponse impulsionnelle du filtre que l'on souhaite obtenir. On a la forme suivante :

$$H(z) = \sum_{k=0}^{M-1} b_k \cdot z^{-k}$$

Ils sont simples à mettre en œuvre et possèdent une phase linéaire

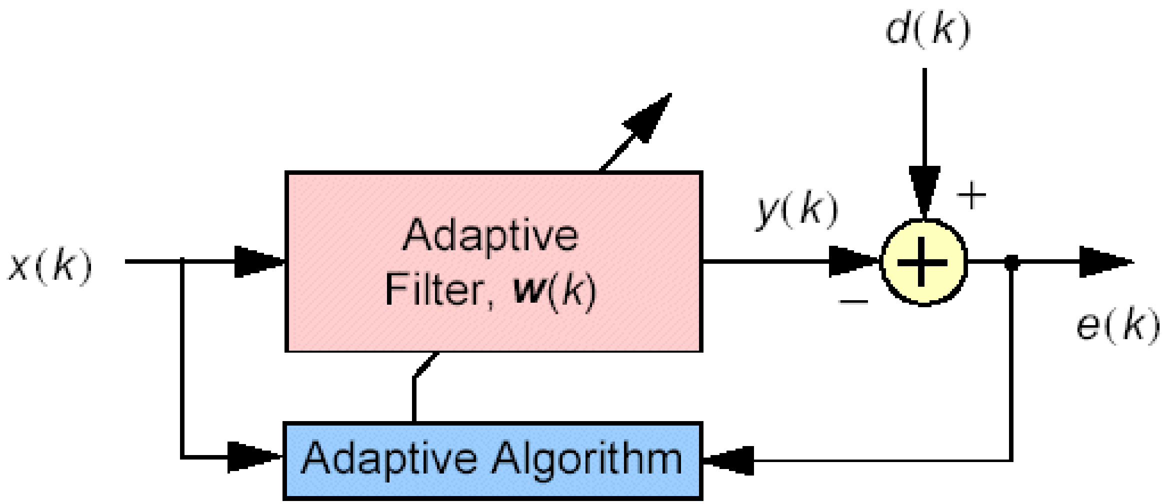
1. Notions de base

- Les filtres IIR (Infinite Impulse Response) qui sont obtenus à partir des fonctions analogiques (passage de s à z). La fonction de transfert possède la forme suivante :

$$H(z) = \frac{\sum_{n=0}^{M-1} b_0 z^{-n}}{1 + \sum_{k=1}^{N-1} a_1 z^{-k}}$$

- Les filtres adaptatifs dont les coefficients sont mis à jour régulièrement en minimisant un critère (moindres carrés). Ils peuvent être de type FIR ou IIR.

1. Notions de base



$$y(k) = \text{Filter}\{x(k), \mathbf{w}(k)\}$$

$$\mathbf{w}(k+1) = \mathbf{w}(k) + e(k) f\{d((k), x(k))\}$$

1. Notions de base

✦ La transformée de Fourier discrète :

$$X(k) = \sum_{n=0}^{N-1} x[n] e^{-j2\pi kn/N} \qquad 0 \le k \le N-1$$

- En pratique cette somme est calculée en utilisant des algorithmes rapides et on parle de FFT (Fast Fourier Transform) (cf. fonction fft dans Matlab).

✦ La corrélation : exprime la notion de similitude entre deux signaux. On distingue :

1. Notions de base

- L'intercorrélation :

$$r_{xy}[l] = \sum_{n=-\infty}^{+\infty} y[n] \cdot x[n-l]$$

- L'autocorrélation :

$$r_{xx}[l] = \sum_{n=-\infty}^{+\infty} x[n] \cdot x[n-l]$$

<u>Remarque</u> : $r_{xx}[0] = E_x$

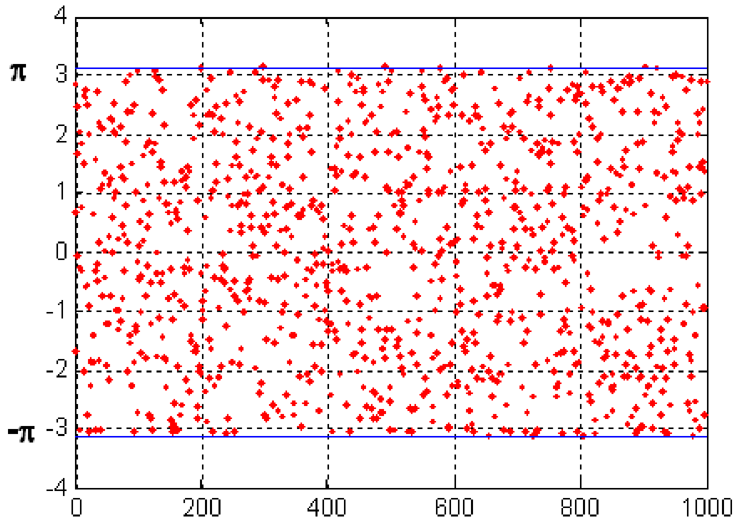
1. Notions de base

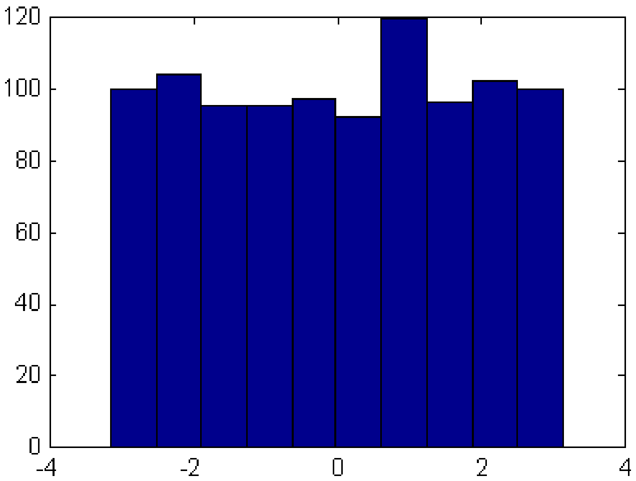
- Notions de signaux aléatoires :
  - Les signaux seront considérés comme stationnaires et ergodiques :
    - Stationnarité ➔ les moments statistiques (moyenne, variance) ne dépendent pas du temps.
    - Ergodicité ➔ on peut confondre les moments statistiques et les moments temporels (moyenne temporelle, puissance).

1. Notions de base

- Lois usuelles :
  - Loi uniforme sur (a,b) :

$$p_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{pour} \quad x \in (a,b) \\ 0 & \text{sinon} \end{cases}$$

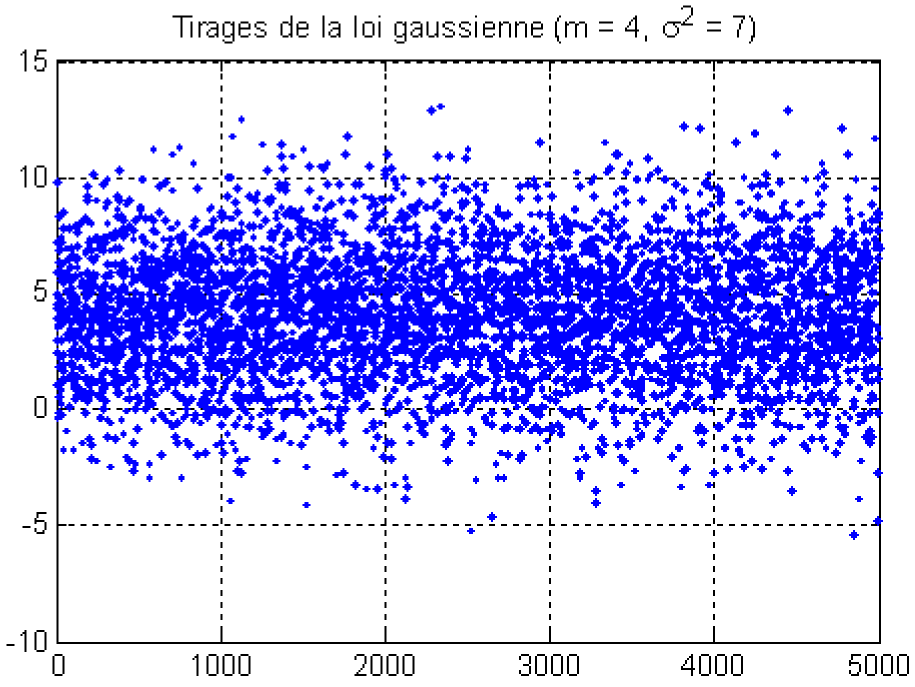


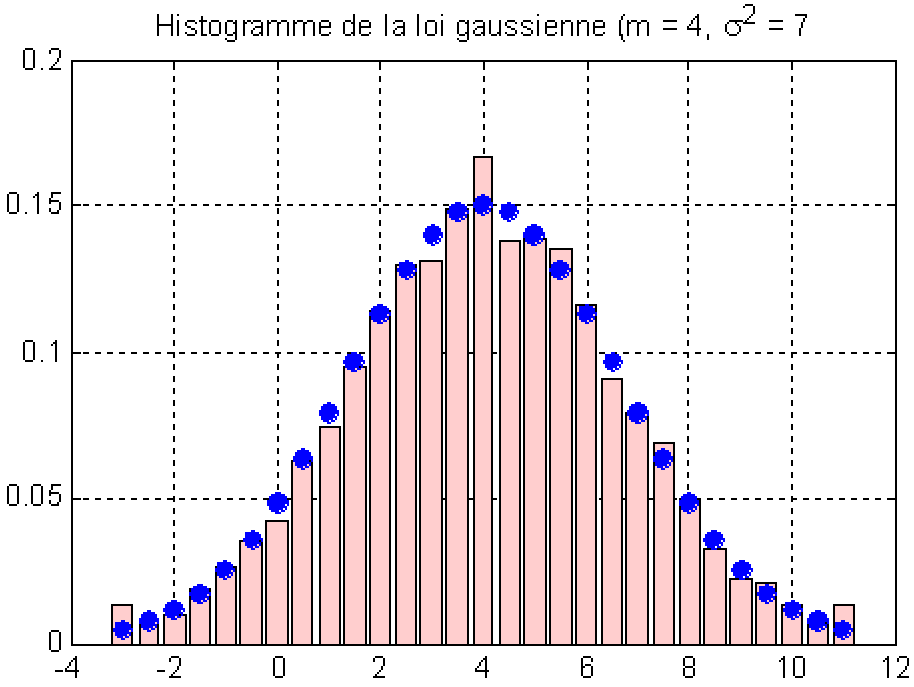


1. Notions de base

- Loi Gaussienne :

$$p_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2} \right)$$

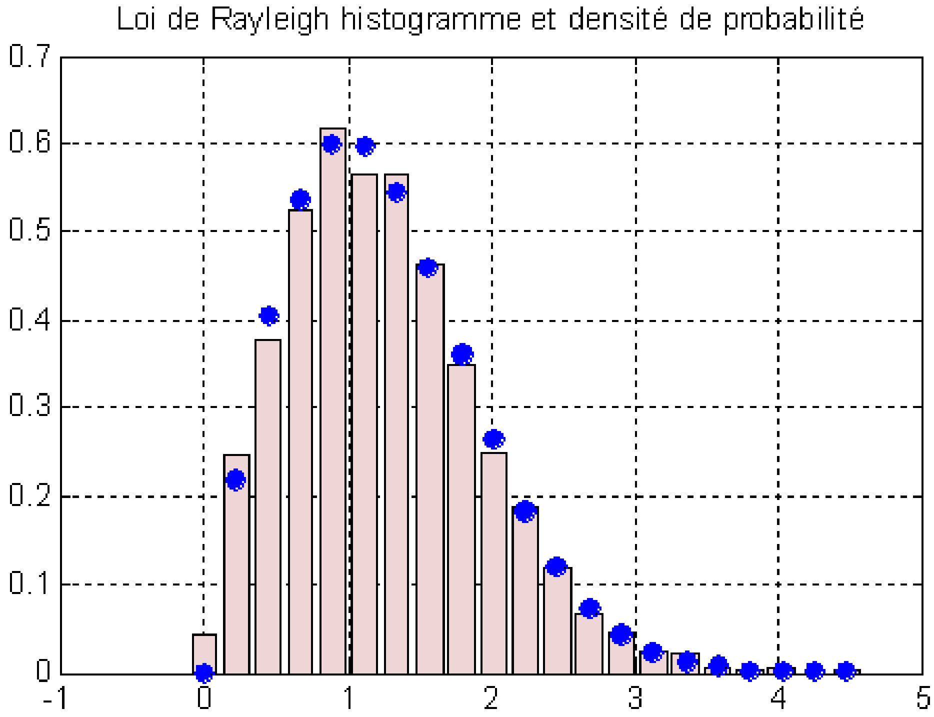




1. Notions de base

- Loi de Rayleigh :

$$p_X(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2} \right) \quad x \geq 0$$



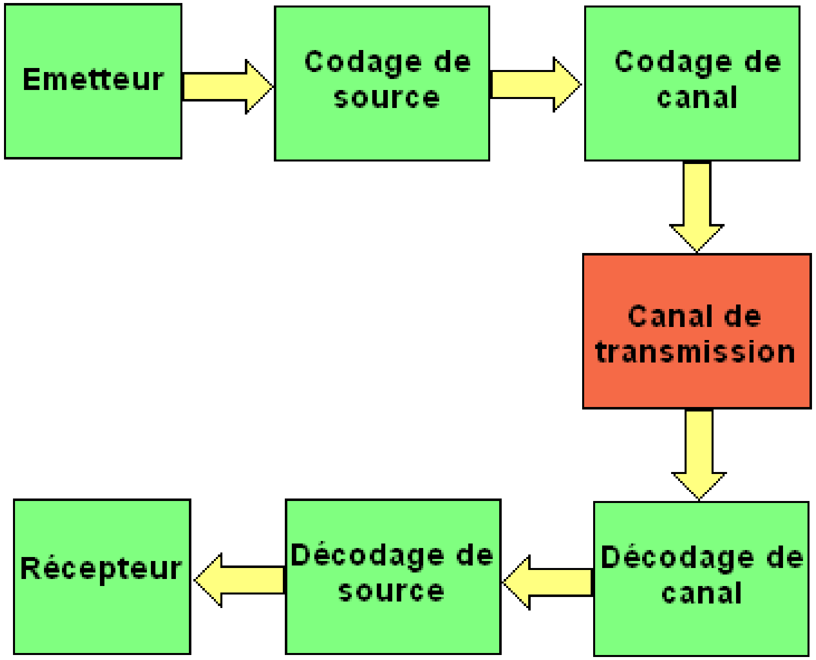
1. Notions de base

- Rapport signal sur bruit (SNR) :

$$SNR_{dB} = 10 \cdot \log_{10}\left(\frac{P_{signal}}{P_{bruit}}\right)$$

2. Théorie de l'information

- Le système de transmission numérique tel que défini par Shannon en 1948 :



2. Théorie de l'information

➢ Définition de l'information :

- Soit une variable aléatoire discrète S qui peut prendre les valeurs $S = \{s_0, s_1, \dots, s_{K-1}\}$ avec les probabilités $P(S = s_k) = p_k$ avec $k = 0, 1, \dots, K-1$, on définit la quantité d'information d'un symbole par :

$$I(s_k) = -\log_2(p_k) \text{ (bit)}$$

Il est intéressant de connaître l'information moyenne produite pas une source = entropie.

2. Théorie de l'information

$$H(X) = E[I(s_k)] = \sum_{k=0}^{K-1} p_k I(s_k) = -\sum_{k=0}^{K-1} p_k \log_2(p_k)$$

✦ Propriétés :

- $0 \leq H(X) \leq \log_2(K)$
- $H(X) = 0$ si $p_k = 1$ et les autres probabilités sont nulles ➔ aucune incertitude
- $H(X) = \log_2(K)$ si $p_k = 1/K \ \forall k$ ➔ incertitude maximale

2. Théorie de l'information

➢ Le théorème du codage de source :

Si l'on associe un code à la source S de longueur moyenne :

$$\overline{L} = \sum_{k=0}^{K-1} p_k I_k$$

On aura toujours :

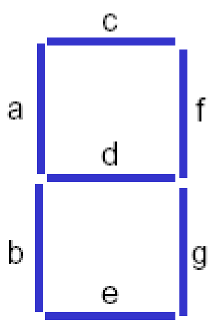
$$\overline{L} \geq H(X)$$

On peut ainsi définir l'efficacité d'un code par :

2. Théorie de l'information

$$\eta = \frac{H(X)}{\overline{L}}$$

➢ Exemple : l'afficheur 7 segments :



digit	0	1	2	3	4	5	6	7	8	9
a	1	0	0	0	1	1	1	0	1	1
b	1	0	1	0	0	0	1	0	1	0
c	1	0	1	1	0	1	1	1	1	1
d	0	0	1	1	1	1	1	0	1	1
e	1	0	1	1	0	1	1	0	1	1
f	1	1	1	1	1	0	0	1	1	1
g	1	1	0	1	1	1	1	1	1	1

- Les symboles sont équiprobables ➔ $p_k = 0.1$
- Entropie $H(X) = 3.32$ bits
- Rendement $\eta = 52.54\%$

2. Théorie de l'information

- Compression de l'information :
  - Les signaux physiques comprennent de l'information redondante → utilisation inutile du canal de transmission
  - On utilise un codage à longueur variable dont le principe consiste à assigner des descriptions longues aux symboles peu fréquents et des descriptions plus courtes aux symboles plus fréquents.

2. Théorie de l'information

- Les codes préfixes : aucun mot du code n'est le préfixe d'un autre mot du code. Ils satisfont l'inégalité de Kraft-McMillan :

$$\sum_{k=0}^{K-1} 2^{l_k} \leq 1$$

Leur longueur moyenne est bornée par :

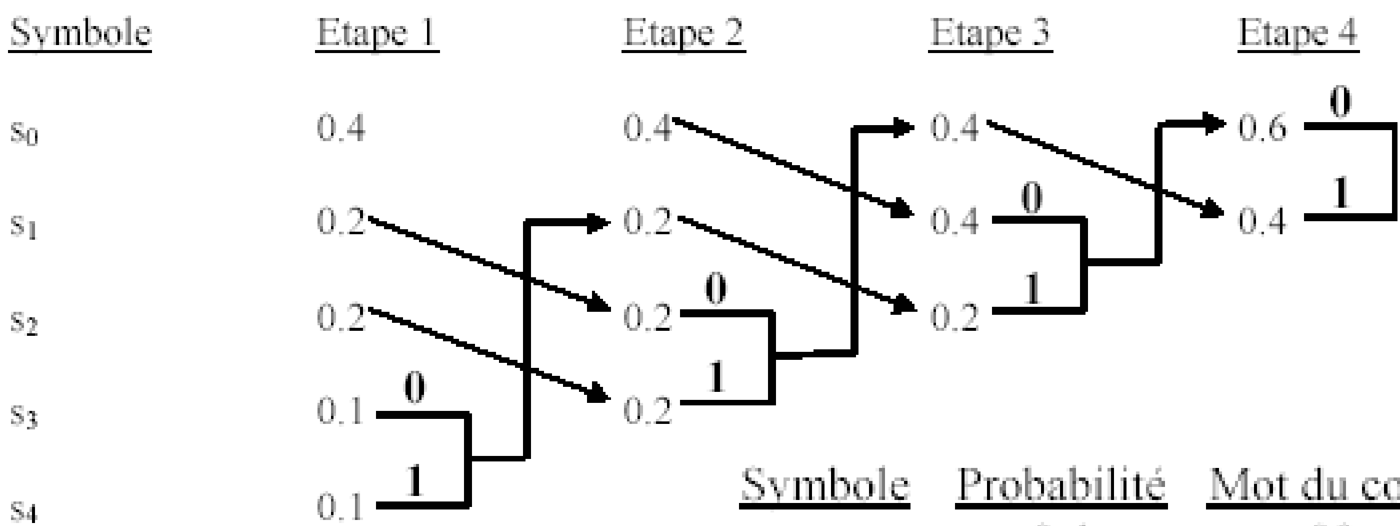
$$H(X) \leq \overline{L} < H(X)+1$$

<u>Exemple</u> :

Symbole	Probabilité	Code I	Code II	Code III
s_0	0.5	0	0	0
s_1	0.25	1	10	01
s_2	0.125	00	110	011
s_3	0.125	11	111	0111

2. Théorie de l'information

✦ Le codage d'Huffmann



Symbole	Probabilité	Mot du code
s_0	0.4	00
s_1	0.2	10
s_2	0.2	11
s_3	0.1	010
s_4	0.1	011

2. Théorie de l'information

✦Le codage LZW (WINZIP et Cie) dictionnaire construit dynamiquement :

1 0 0 1 0 1 1 1 0 0 0 1 1 1 0 0 1 0 1 . . .

Dictionnaire		Mot lu		
indices	mots	valeur	indice	mot de code
0	0			
1	1			
2	10	1	1	1
3	00	0	0	00
4	01	0	0	00
5	101	10	2	010
6	11	1	1	001
7	110	11	6	110
8	000	00	3	011
9	011	01	4	0100
10	1100	110	7	0111
11	010	01	5	0101
. . .				

2. Théorie de l'information

➢ Canal discret sans mémoire :

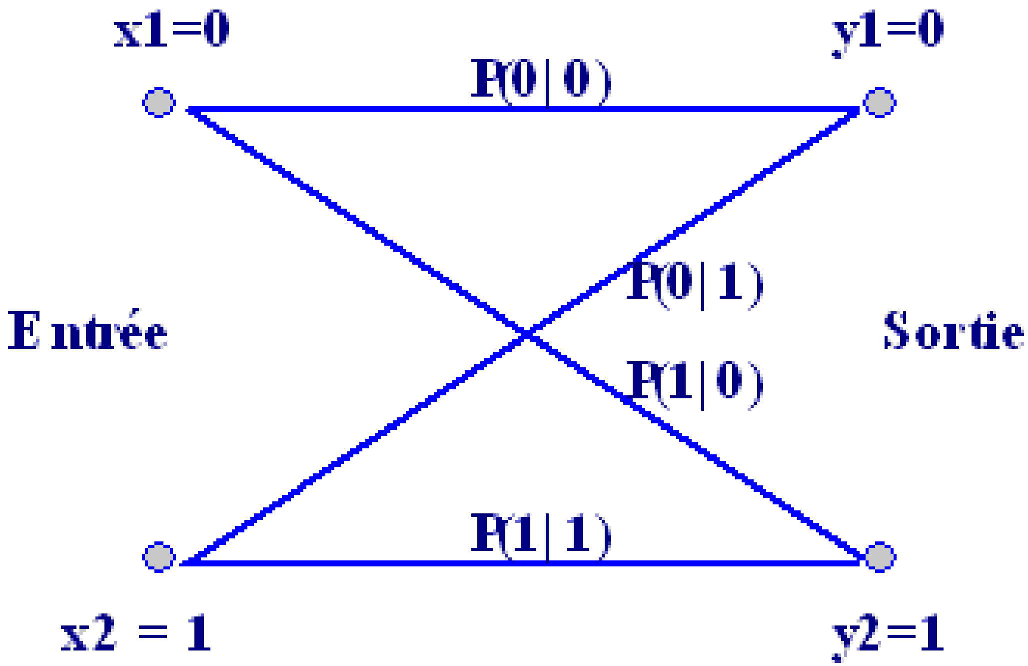
C'est un modèle statistique comportant une entrée X et une sortie Y qui est une version bruitée de X. Il est décrit par ses probabilités de transition :

$$p(y_k|x_j) = P(Y = y_k|X = x_j) \ \forall \ j, k$$

$$P = \begin{bmatrix} p(y_0|x_0) & p(y_1|x_0) & \cdots & p(y_{K-1}|x_0) \\ p(y_0|x_1) & p(y_1|x_1) & \cdots & p(y_{K-1}|x_1) \\ \vdots & \vdots & & \vdots \\ p(y_0|x_{J-1}) & p(y_1|x_{J-1}) & \ldots & p(y_{K-1}|x_{J-1}) \end{bmatrix}$$

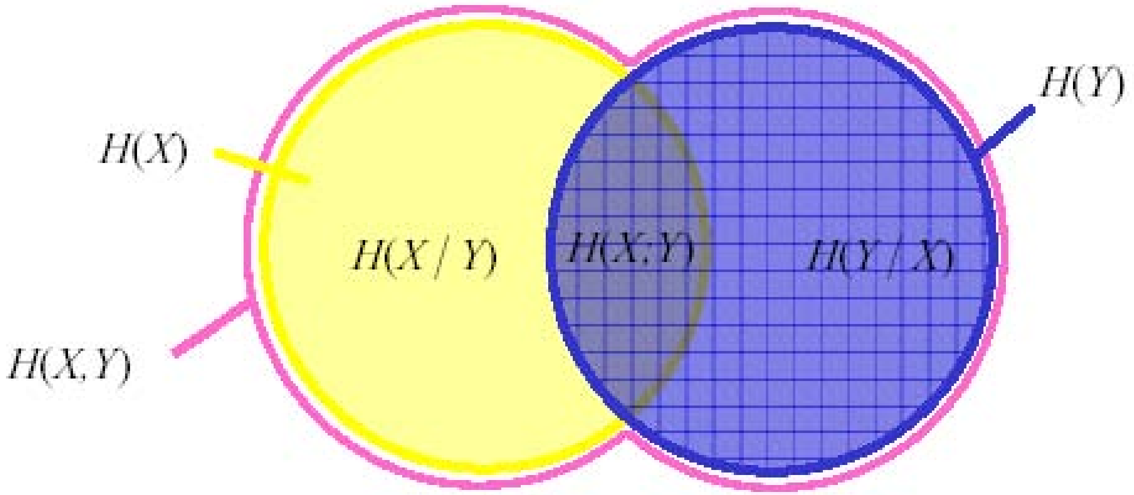
2. Théorie de l’information

- Exemple : le canal binaire symétrique :



2. Théorie de l'information

- Différentes sortes d'entropies :
  - Si l'on utilise les différentes probabilités à notre disposition on peut définir les entropies suivantes :



2. Théorie de l'information

$$H(X) = -\sum_{i=1}^{n} p(x_i)\log_2\left(p(x_i)\right)$$

$$H(Y) = -\sum_{j=1}^{m} p(y_j)\log_2\left(p(y_j)\right)$$

$$H\left(X\middle|Y\right) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\log_2\left(p\left(x_i\middle|y_j\right)\right)$$

$$H(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\log_2\left(p(x_i, y_j)\right)$$

2. Théorie de l'information

- H(X|Y) représente l'information perdue au cours de la transmission. On en déduit une nouvelle quantité appelée information mutuelle du canal :

$$I(X;Y) = H(X) - H\left(X \middle| Y\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\log_2\left(\frac{p\left(y_j \middle| x_i\right)}{p(y_j)}\right)$$

Cette quantité représente l'information correctement transmise de la source vers le destinataire.

Propriétés :

- Symétrie : I(X ;Y) = I(Y ;X)
- I(X ; Y) $\geq$ 0
- I(X ; Y) = H(X) + H(Y) – H(X,Y)

2. Théorie de l'information

- Capacité :

$$C = \max_{\{p_{x_i}\}} I(X;Y)$$

<u>Exemple</u> : capacité du canal binaire symétrique

$C = 1 + p.\log_2(p) + (1-p).\log_2(1-p)$

2. Théorie de l'information

- Théorème du codage de canal :

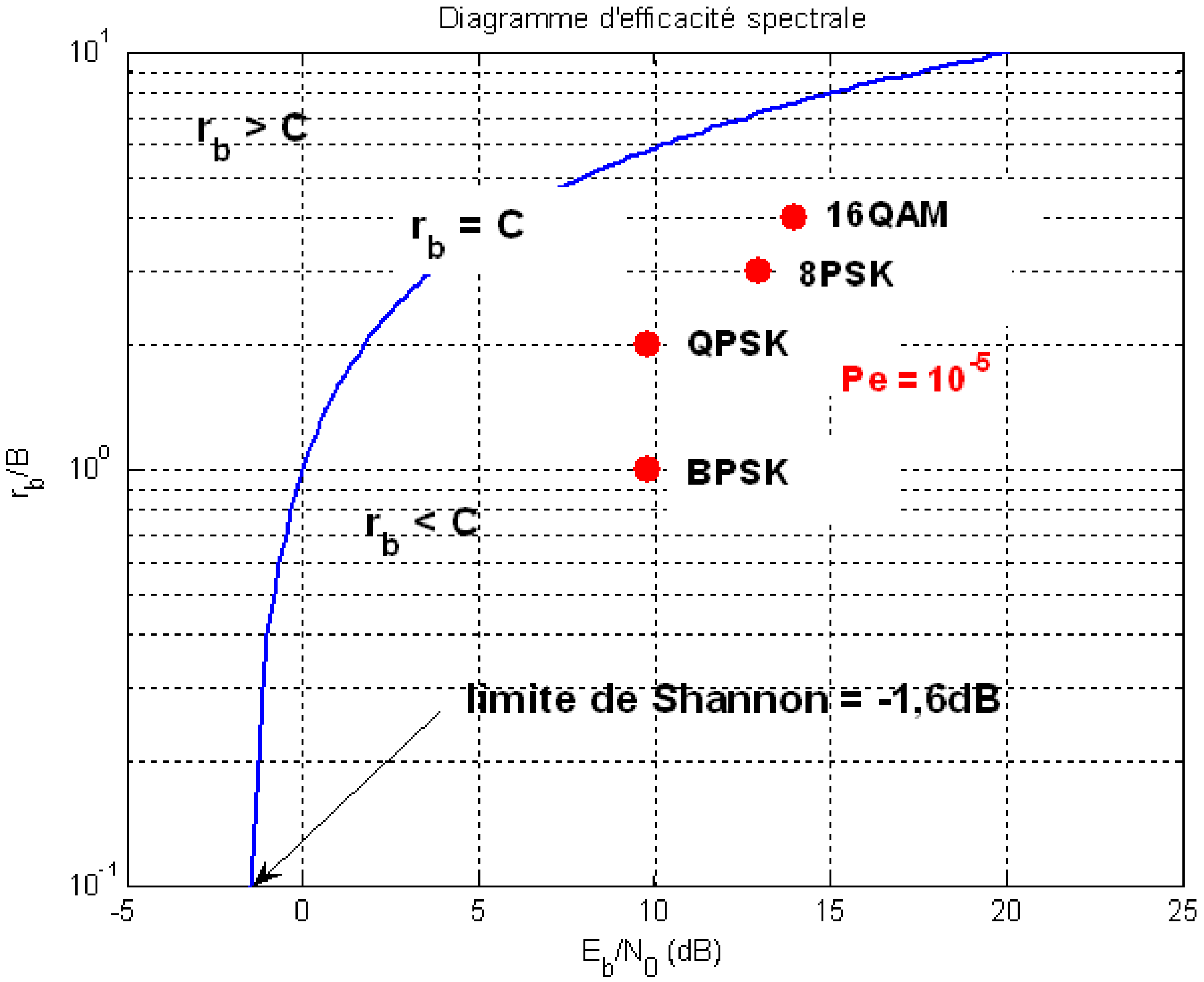
$$\frac{H(X)}{T_s} \leq \frac{C}{T_c}$$

<u>Remarque</u> : ce théorème ne donne aucune indication sur comment construire de bon codes.

- Capacité du canal BBAG de puissance et BP limités :

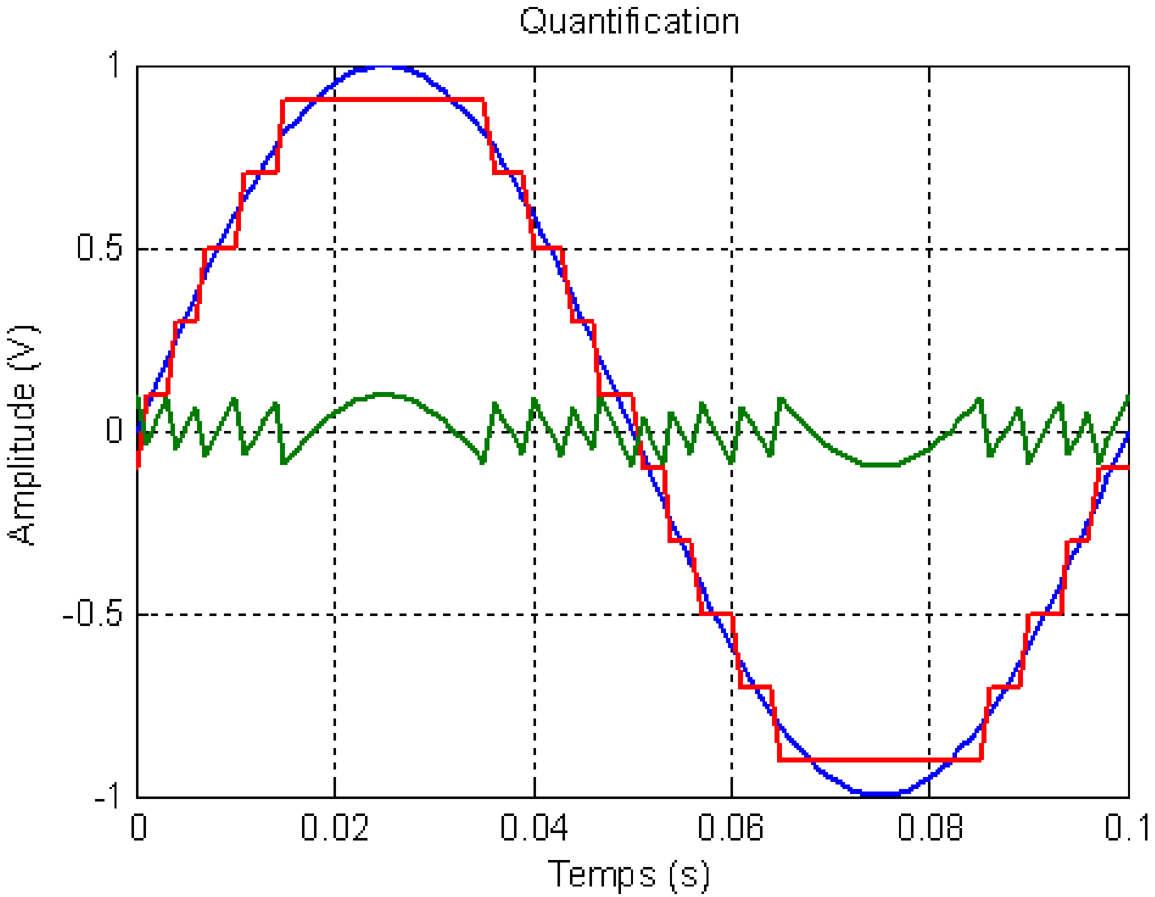
$$\frac{C}{B} = \log_2\left(1 + \frac{E_b}{N_0} \cdot \frac{C}{B}\right)$$

2. Théorie de l’information



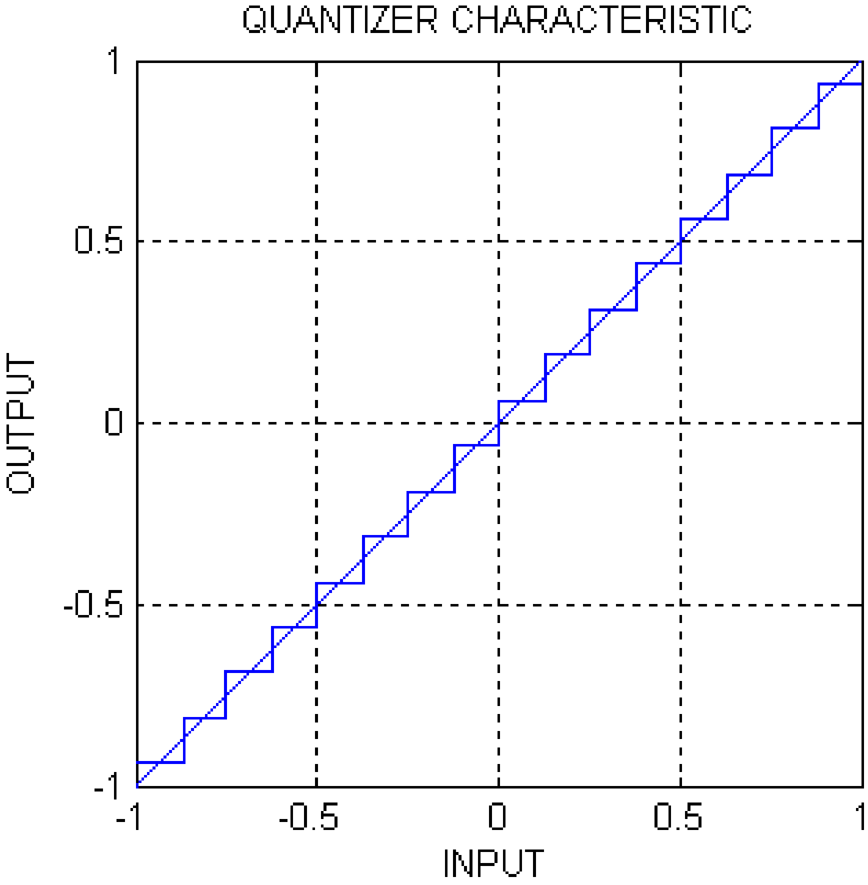
3. Quantification

➢ Illustration :



3. Quantification

➢Quantification uniforme :



3. Quantification

Bruit de quantification :

- Si le signal à quantifier est compris entre $(-m_{max}, m_{max})$, pour L niveaux on aura :

$$\Delta = \frac{2m_{\max}}{L}$$

- L'erreur de quantification Q est une variable aléatoire uniformément distribuée :

$$f_Q(q) = \begin{cases} \dfrac{1}{\Delta} & -\dfrac{\Delta}{2} < q \leq \dfrac{\Delta}{2} \\ 0 & autrement \end{cases}$$

3. Quantification

- La variance de l'erreur est égale à :

$$\sigma_Q^2 = \frac{\Delta^2}{12}$$

- Comme L = 2^R, on obtient :

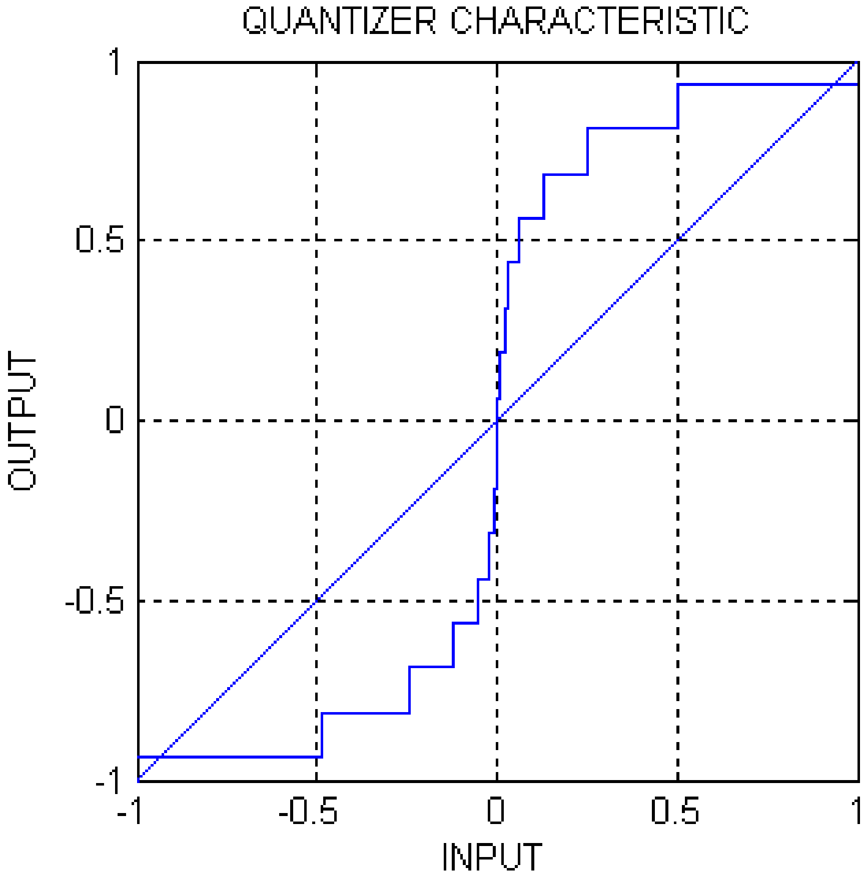
- Finalement : $$SNR = \frac{P}{\sigma_Q^2} = \left(\frac{3P}{m_{\max}^2}\right) 2^{2R}$$

$$\sigma_Q^2 = \frac{1}{3} m_{\max}^2 \, 2^{-2R}$$

- Exercice : calculer le SNR en dB si le signal à quantifier est une sinusoïde d'amplitude max = A_m

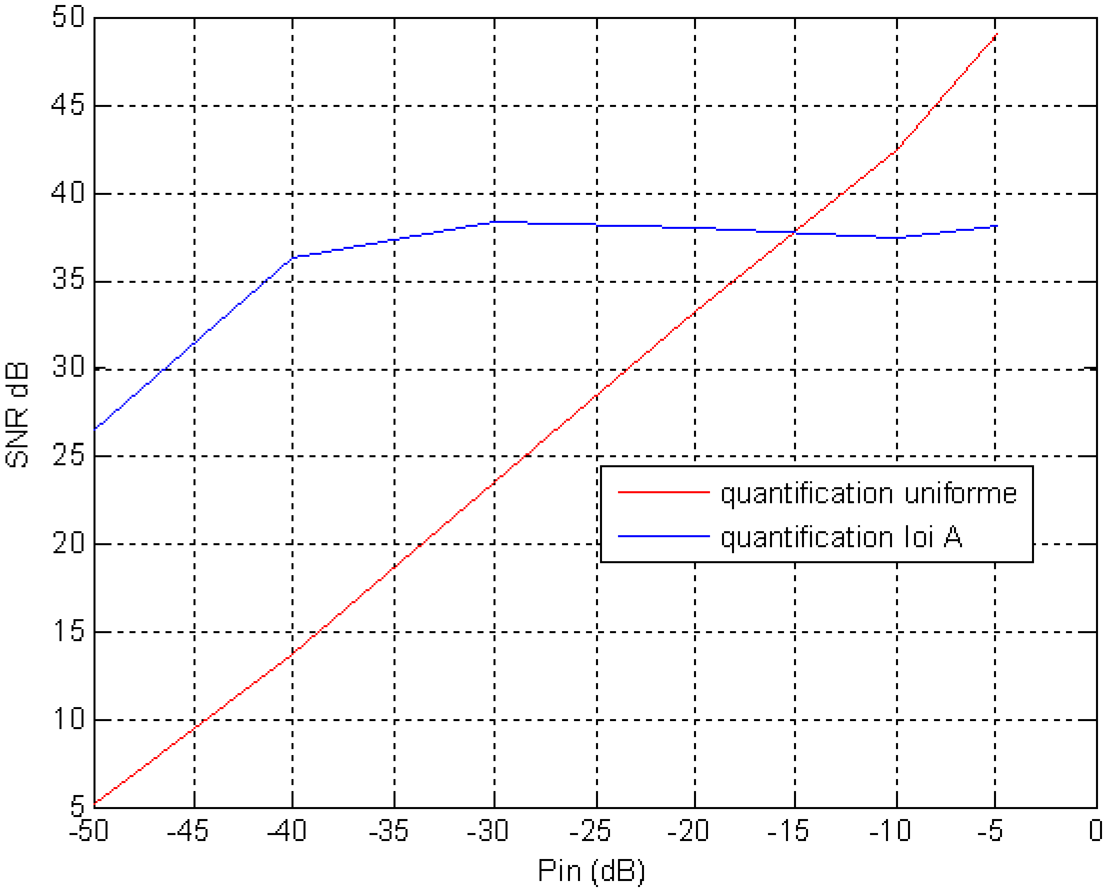
3. Quantification

➢ Quantification non linéaire :



3. Quantification

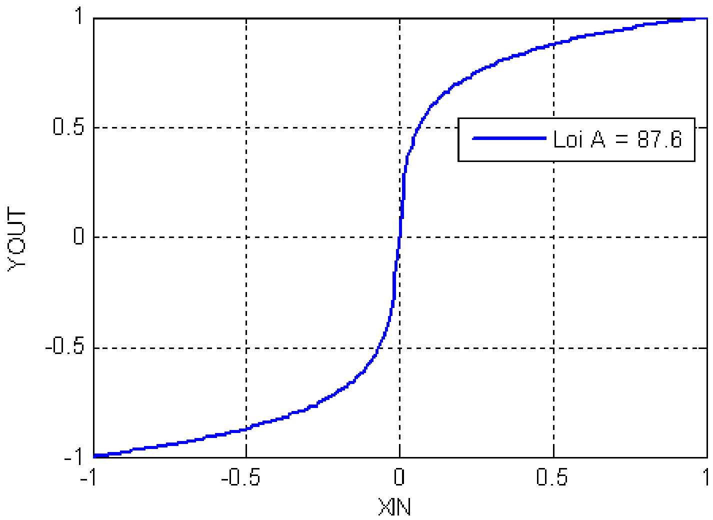
- Comparaison SNR = f(P_{in}) quantification uniforme et non linéaire loi A, A = 87.6 :



3. Quantification

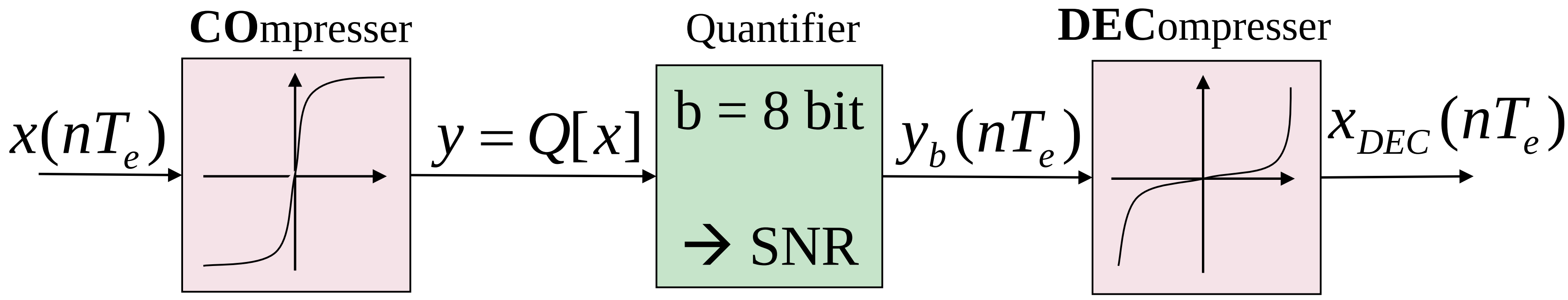
➤ Un exemple pratique : ITU-T G711 loi A :

$$c(x) = \begin{cases} \dfrac{A|x|}{1+\ln(A)}\,\mathrm{sgn}(x) & 0 \le \dfrac{|x|}{x_{\max}} \le \dfrac{1}{A} \\ x_{\max}\dfrac{1+\ln(A|x|/x_{\max})}{1+\ln(A)}\,\mathrm{sgn}(x) & \dfrac{1}{A} \le \dfrac{|x|}{x_{\max}} \le 1 \end{cases}$$

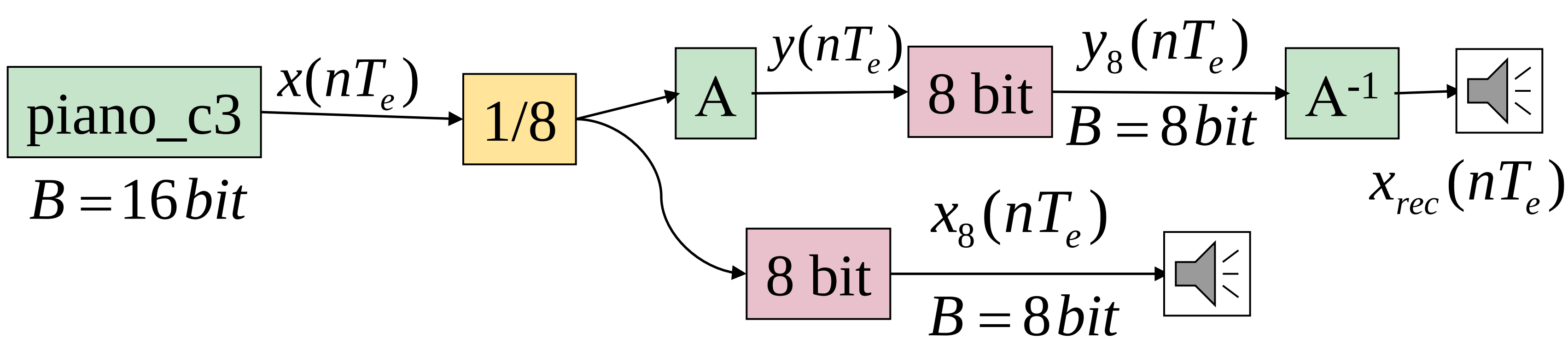


3. Quantification

Companding (Compress Expand) :

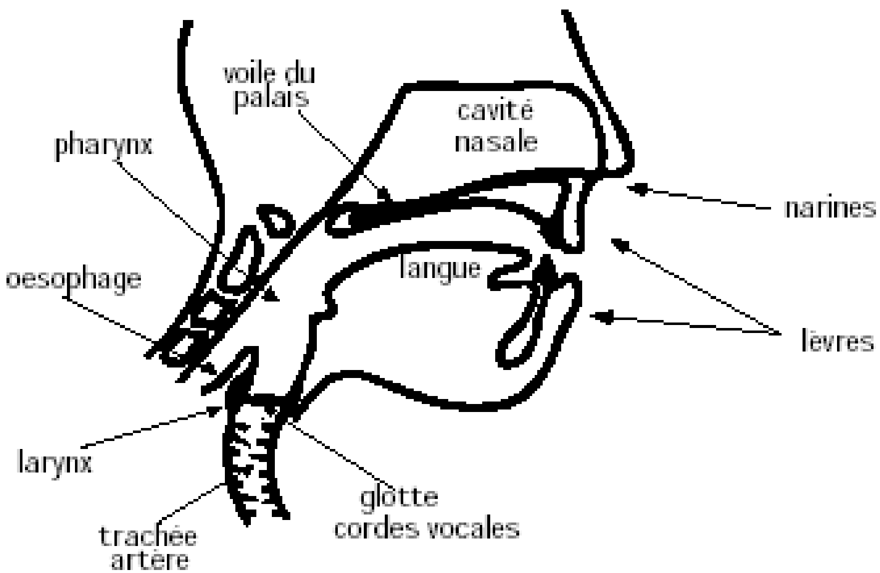


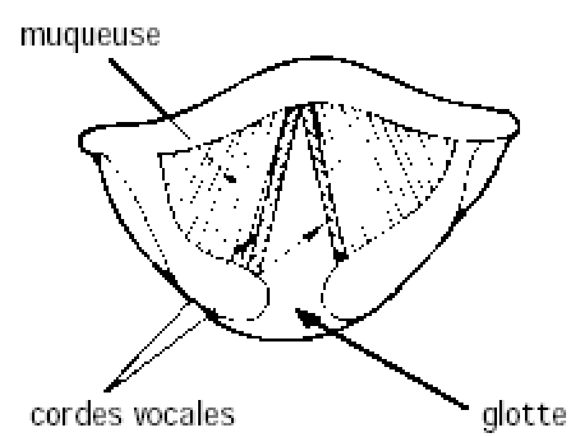
Amélioration du SNR :



4. Codage de la parole

- Le signal de parole : différents niveaux de description :
  - Le niveau phonétique : étudie la façon dont le signal est produit et perçu :
    - Appareil phonatoire





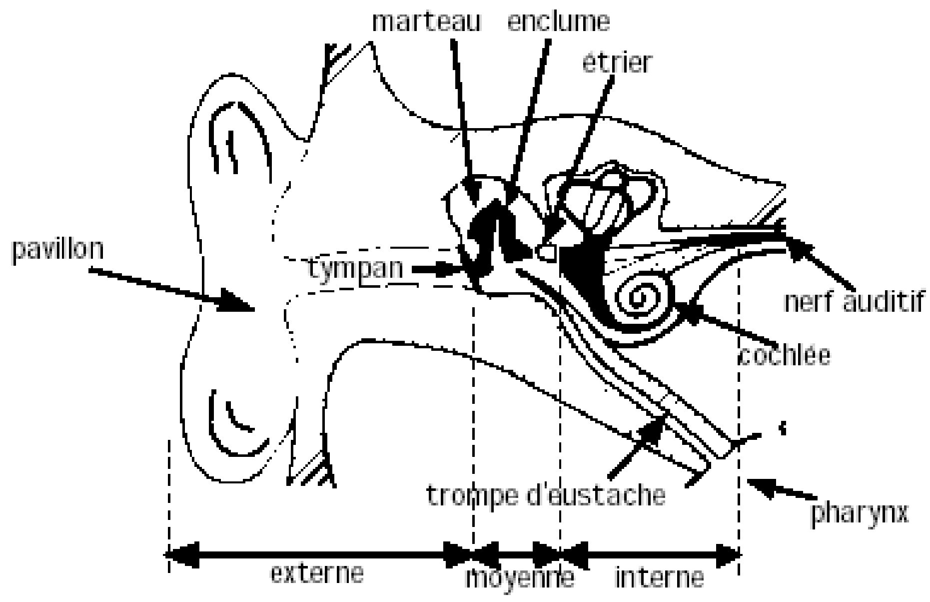
4. Codage de la parole

- L'alphabet phonétique international :

IPA	EXEMPLES	IPA	EXEMPLES
i	idée, ami	p	patte, repas, cap
e	ému, ôté	t	tête, ôter, net
ɛ	perdu, modèle	k	carte, écaille, bec
a	alarme, patte	b	bête, habile, robe
ɑ	bâton, pâte	d	dire, rondeur, chaud
ɔ	Obstacle, corps	g	gauche, égal, bague
o	auditeur, beau	f	feu, affiche, chef
u	coupable, loup	s	sœur, assez, passe
y	punir, élu	ʃ	chanter, machine, poche
ø	creuser, deux	v	vent, inventer, rêve
œ	malheureux, peur	z	zéro, raisonner, rose
ə	petite, fortement	ʒ	jardin, manger, piège
ɛ̃	peinture, matin	l	long, élire, bal
ɑ̃	vantardise, temps	ʁ	rond, chariot, sentir
ɔ̃	rondeur, bon	m	madame, aimer, pomme
œ̃	lundi, brun	n	nous, punir, bonne
j	piétiner, briller		agneau, peigner, règne
w	oui, fouine	ŋ	jumping, smoking
ɥ	huile, nuire	h	halte, hop (exclamations)

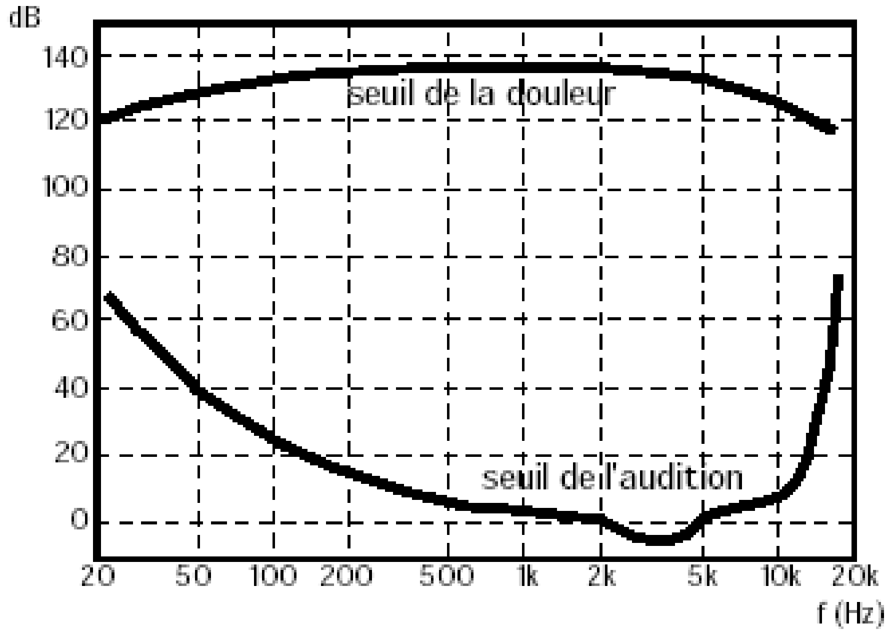
4. Codage de la parole

- Phonétique articulatoire :
  - Voyelles : degré d'ouverture du conduit vocal
  - Consonnes : passage forcé de l'air
  - Semi-voyelles (ié, oui) ou liquides (Long, Rond)
- Audition – Perception : l'appareil auditif :



4. Codage de la parole

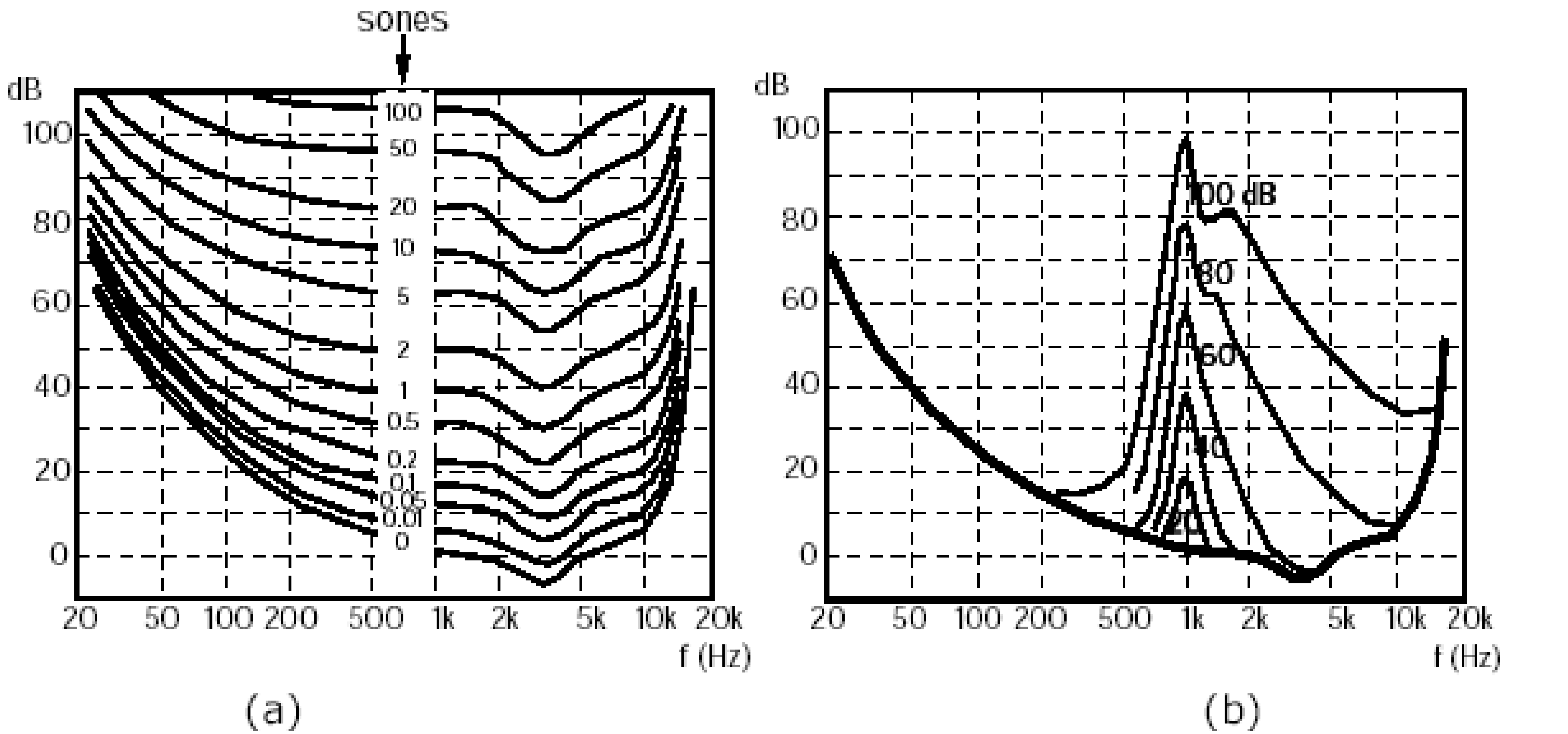
- Champ auditif [500Hz, 10KHz] :



$I(dB) = 10\log(I/I_0)$ avec $I_0 = 10^{-12} W.m^2$ (1KHz)

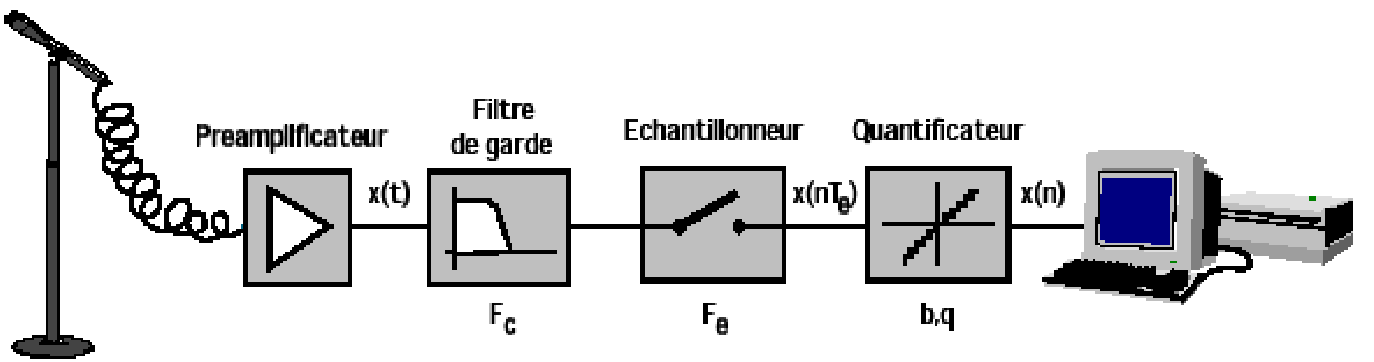
4. Codage de la parole

- Courbes isosoniques et phénomène de masquage :



4. Codage de la parole

- Le niveau acoustique : on étudie le signal électrique associé :



Traits acoustiques : fréquence fondamentale, énergie et spectre. Chaque trait acoustique est lié à une grandeur perceptuelle (pitch, intensité et timbre)

4. Codage de la parole

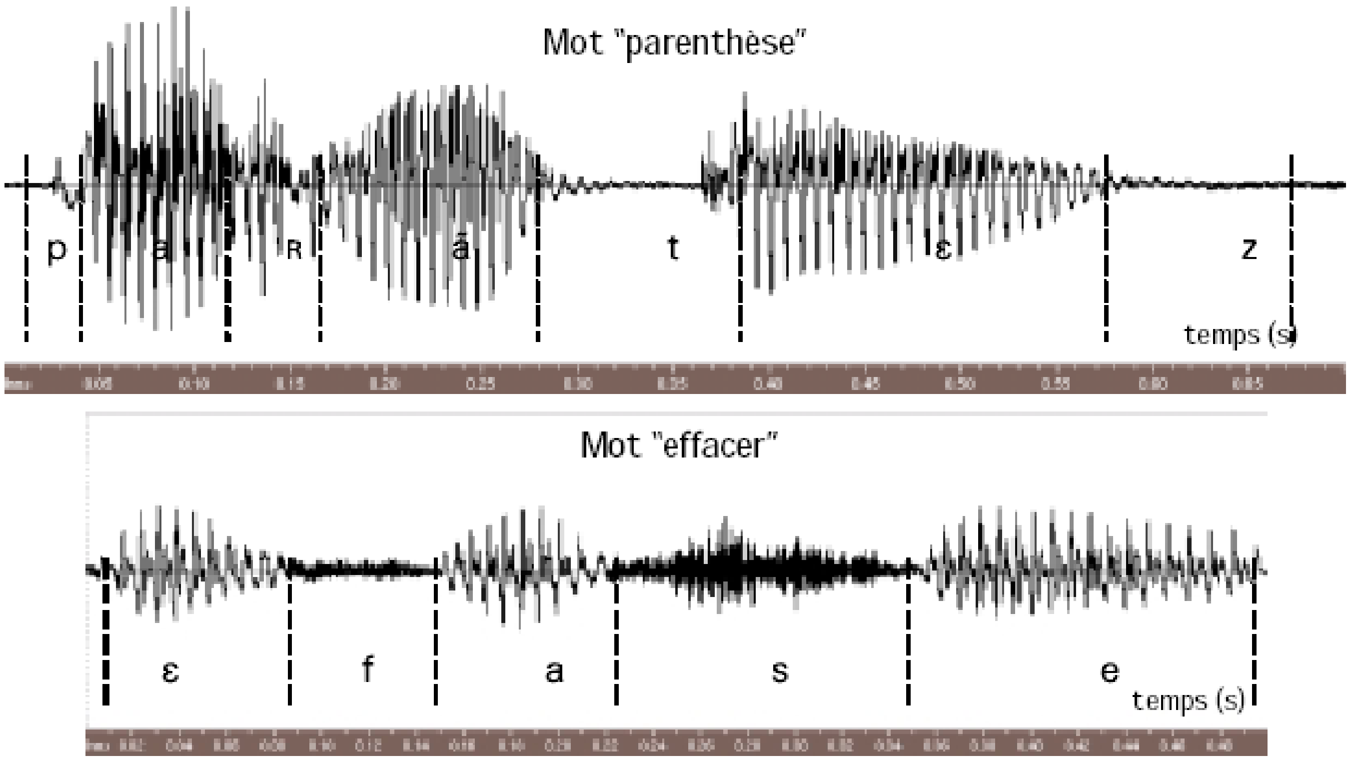
- Le signal vocal est caractérisé par une grande redondance. L'entropie H(X) = 4.73 bits pour les 37 phonèmes de la langue française. Si on prononce en moyenne 10 phonèmes/s ➔ 50bits/s

- Débits courants :

Type	Bande transmise	Fréquence d'échantil-lonnage	Nombre de bits du conver-tisseur A/D	Débit d'information (en kbit/s)	Applications principales
Voix pour le téléphone	300-3 400 Hz	8 kHz	12 ou 13	96 ou 104	Réseau téléphonique analogique et RNIS, téléphone cellulaire
Parole et audio large bande	50-7 000 Hz	16 kHz	14 ou 15	224 ou 240	Vidéo et conférence audio, radio FM
Parole et audio très haute qualité	30-15 000 Hz	32 kHz	16	512	Télévision (NICAM)
	20-20 000 Hz	44,1 kHz	16	706	CD audio
	10-22 000 Hz	48 kHz	Jusqu'à 24	1 152	Audio professionnel

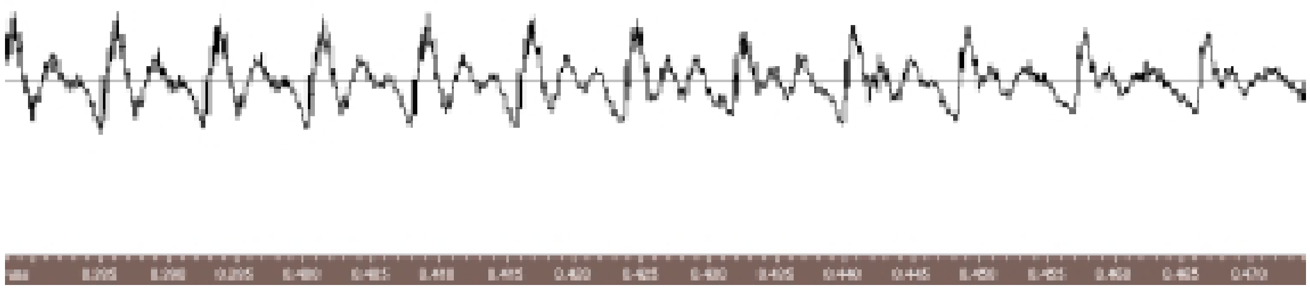
4. Codage de la parole

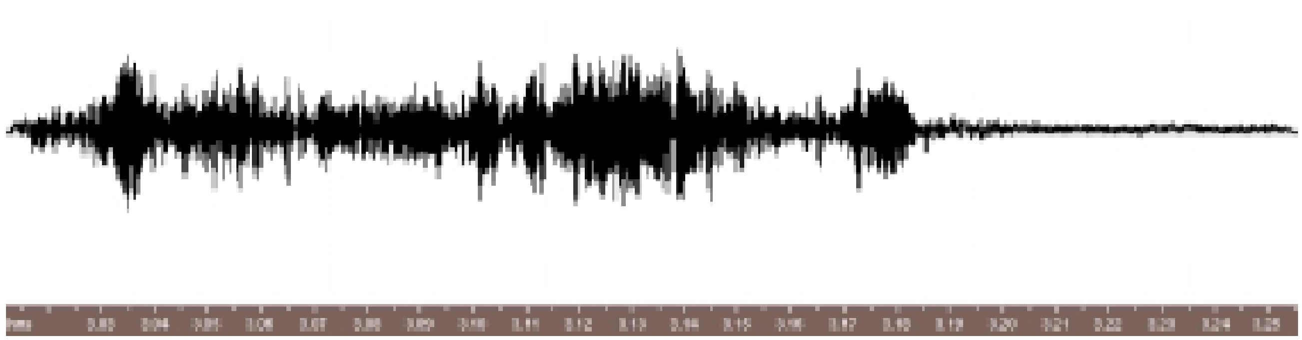
- Audiogramme : amplitude du signal vocal en fonction du temps :



4. Codage de la parole

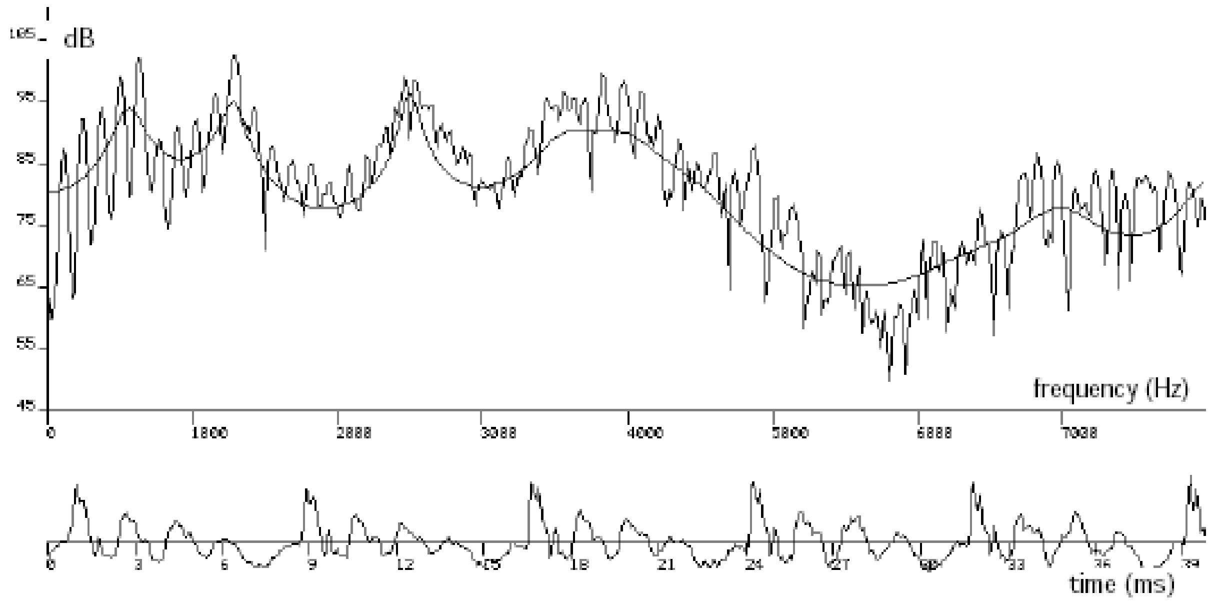
On distingue des zones voisées et non voisées :





4. Codage de la parole

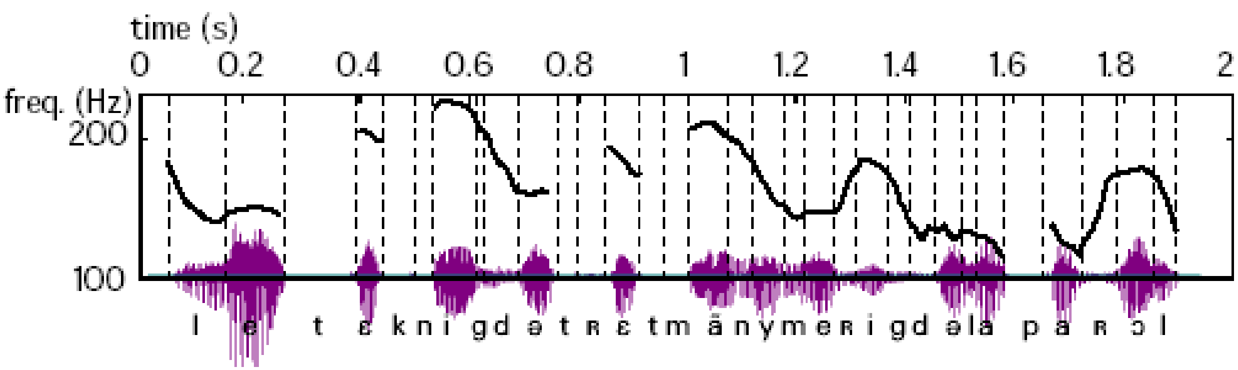
- Transformée de Fourier à court terme (fenêtres de 30ms) :



- On observe des pics de résonance que l'on appelle les formants.

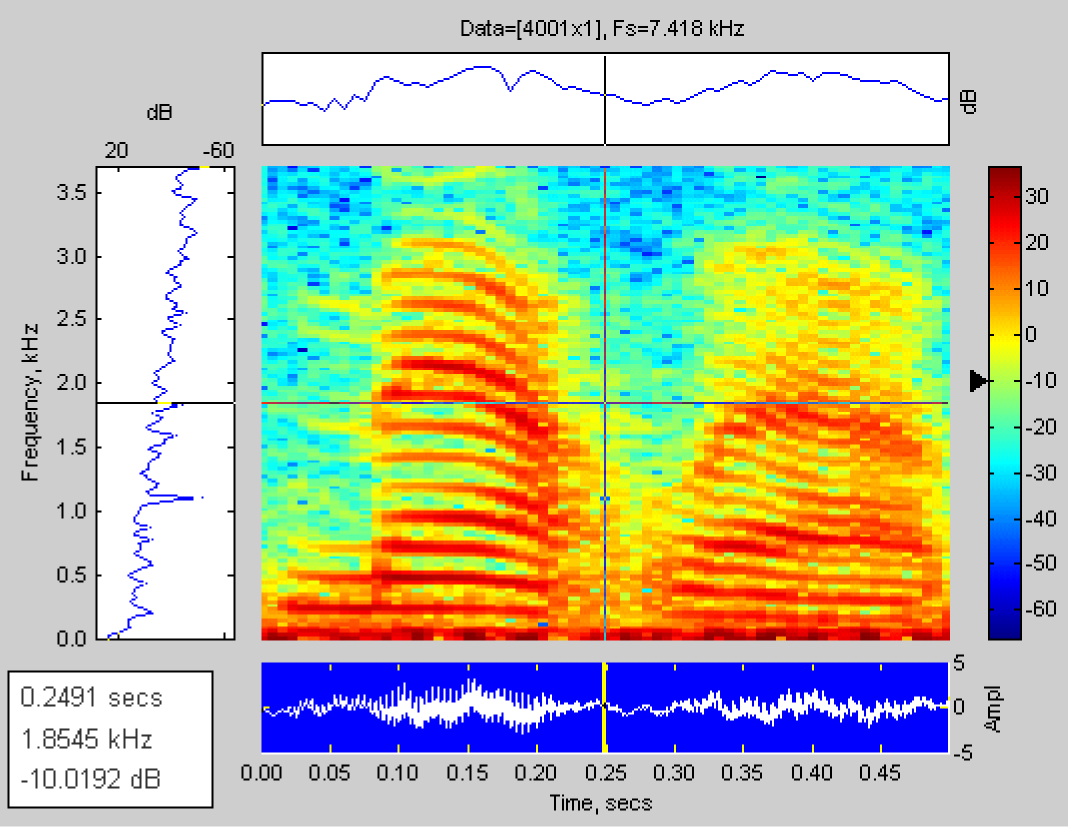
4. Codage de la parole

- Fréquence fondamentale ou pitch :
  - De 70 à 250Hz chez les hommes
  - De 150 à 400Hz chez les femmes
  - De 200 à 600Hz chez les enfants
  - Varie très peu à l'intérieur de la zone voisée



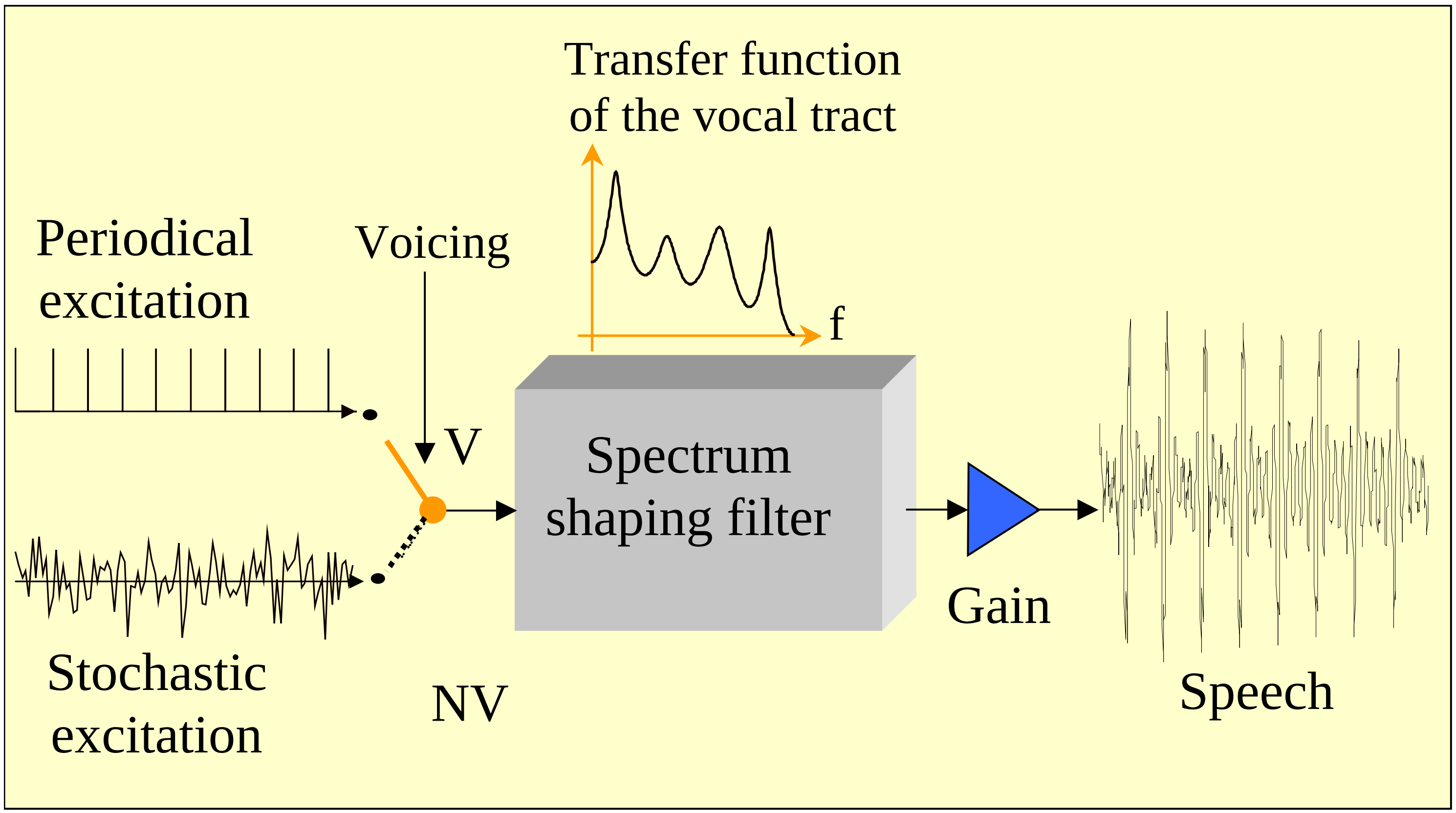
4. Codage de la parole

- Sonogramme : évolution temporelle du spectre à court terme :



4. Codage de la parole

- Modèle simplifié de production de la parole :



4. Codage de la parole

Modélisation LPC :

- On peut montrer que H(z) du filtre qui modélise l'enveloppe spectrale du signal de parole s'écrit :

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 + \sum_{i=1}^{P} a_i z^{-i}}$$

- Les coefficients a_i s'obtiennent par prédiction linéaire :

$$\tilde{x}(n) = -\sum_{i=1}^{p} a_i x(n-i)$$

$$e(n) = x(n) - \tilde{x}(n) \quad \min \sum_{n} e^2(n)$$

4. Codage de la parole

- La mise en équation donne les équations dites de Yule-Walker :

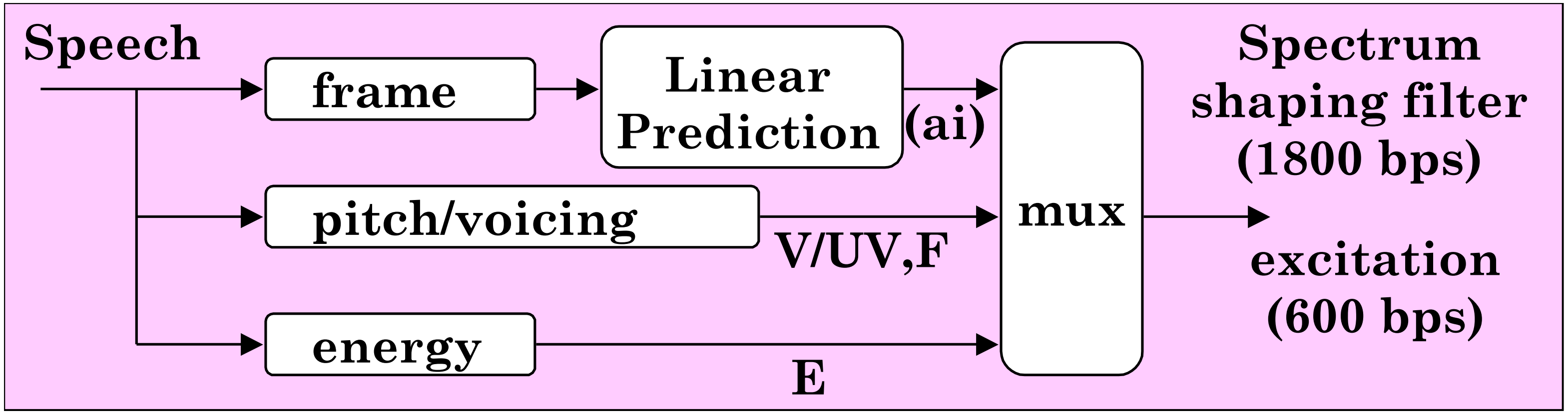
$$\begin{pmatrix} r_{1,1} & \cdots & r_{1,p} \\ \vdots & \cdots & \vdots \\ r_{p,1} & \cdots & r_{p,p} \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix} = -\begin{pmatrix} r_{0,1} \\ \vdots \\ r_{0,p} \end{pmatrix}$$

$$r_{i,j} = \sum x(n-i)x(n-j).$$

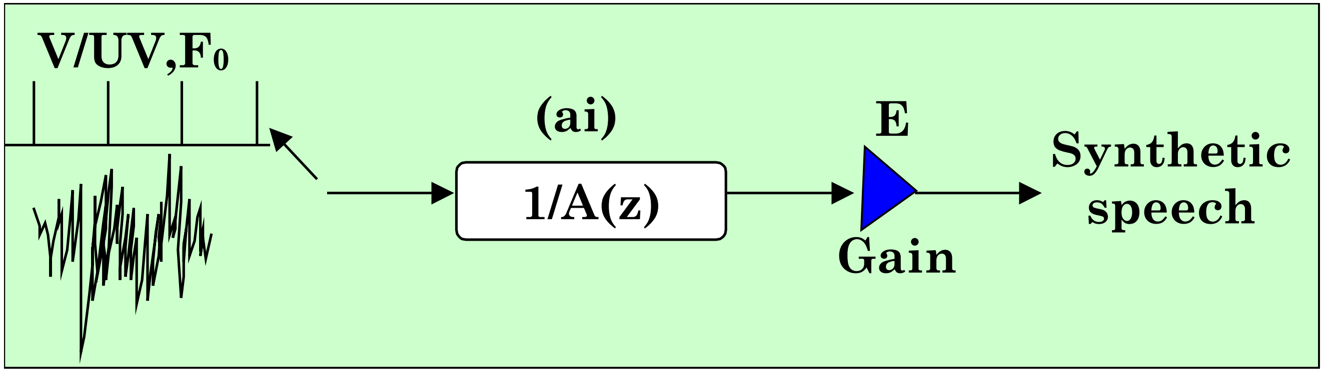
- La résolution se fait en utilisant des algorithmes récursif comme Levinson-Durbin.

4. Codage de la parole

✦ Exemple : le codeur NATO LPC10 :



CODER



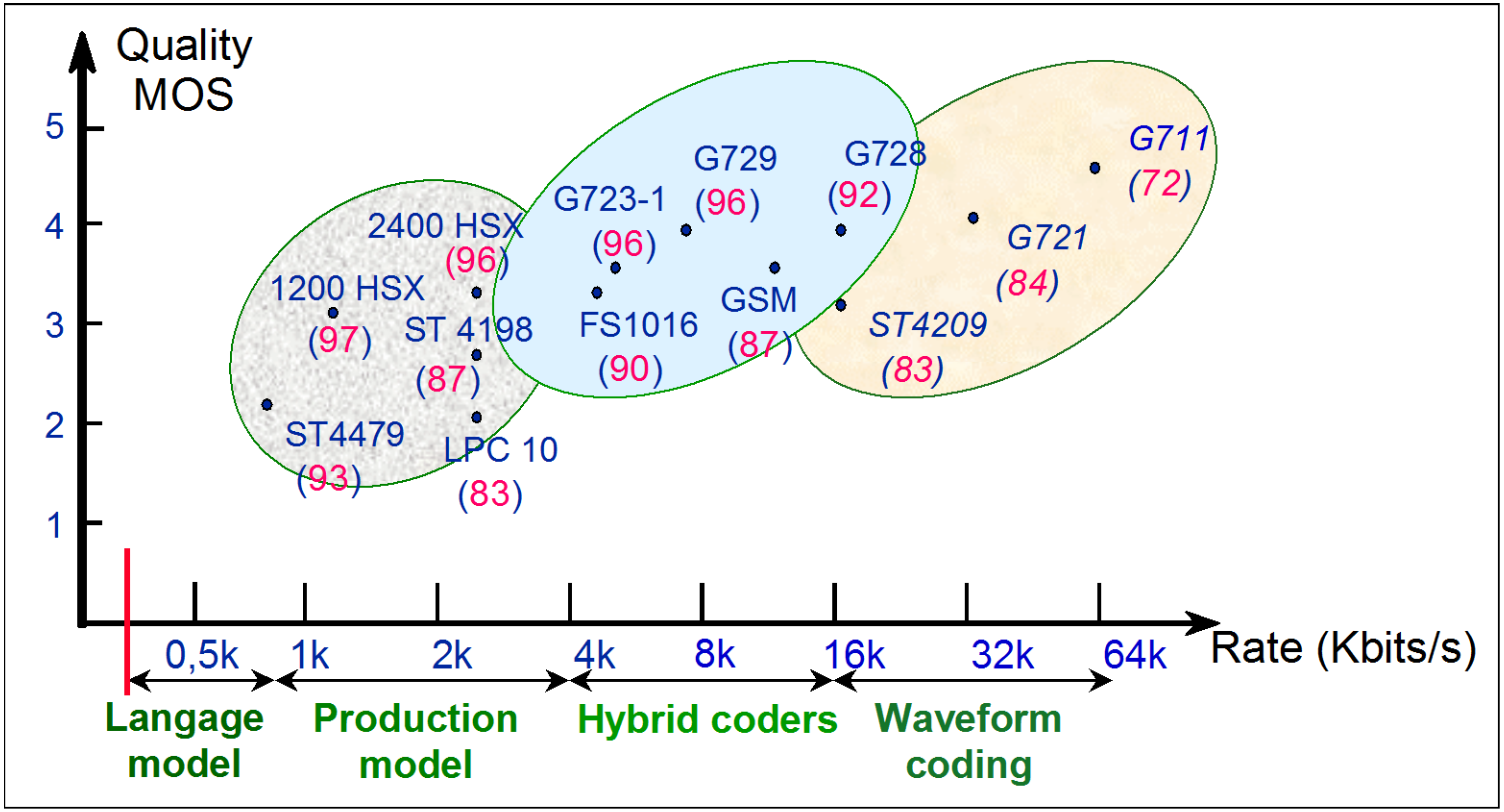
Decoder

4. Codage de la parole

- Les deux grandes familles de codeurs de la parole :
  - Codeurs en forme d’onde qui éliminent la corrélation entre échantillons en utilisant une prédiction linéaire et des quantificateurs adaptatifs ➔ 24 à 32 kbits/s
  - Codeurs ABS ou hybride (Analysis By Synthesis) combinent le meilleur des techniques LPC et forme d’onde ➔ 6 à 16 kbit/s

4. Codage de la parole

➢ Standards :



4. Codage de la parole

➢ Standards ITU :

UIT Standard	Method	Year	Bit rate in Kbps	Delay in ms	Quality MOS	Complexity in Mips
G711	PCM	1972	64	0.125	4.3	<<1
G721 G723 G726 G727	ADPCM	1984 1986 1988 1990	32 40-32-24 40-32-24-16 40-32-24-16	0,125	4.1 at 32Kbps	1.25
G728	LD-CELP	1992	16	2.5	4.0	30
G729 G729a	CS-ACELP	1994 1996	8	30	3.9	25 12
G723.1	MP-MLQ ACELP	1995	6.3 5.3	75	3.9	24

4. Codage de la parole

➢ Standards ETSI :

Standard ETSI europe	Method	Year	Bit rate in Kbps	Delay in ms	Quality MOS	Complexity in Mips
GSM	RPE-LTP	1987	13	40	3.47	6
TETRA	ACELP	1994	4.567	67.5		12
GSM HR	VSELP	1994	5.6	45		30
EFR-GSM (identical DCS)	ACELP	1995	12.8	40		15

➢ Exercice : rechercher sur Internet les normes G726 (ITU-T) et GSM 06.10 (ETSI) et retrouver les éléments fondamentaux de ces standards.

4. Codage de la parole

➢Codeur en forme d'onde : ITU-T G726