

Département d'informatique et de génie logiciel
Compression de données
IFT-4003/IFT-7023

Notes de cours
Préliminaires mathématiques
pour le codage avec perte

Édition Hiver 2012

Mohamed Haj Taieb

Local: PLT 2113

Courriel: mohamed.haj-taieb.1@ulaval.ca

Faculté des sciences et de génie
Département de génie électrique et de
génie informatique

Plan de la présentation

□ Compression avec perte:

- Introduction
- Critère de distorsion
- Système visuel humain
- Perception auditive
- Rappel sur la théorie d'information
- Théorie de débit distorsion
- Modèles pour la conception et l'analyse des systèmes de compression avec perte

Introduction

❑ Schéma de compression:



❑ Compression sans perte: Réversible $\rightarrow X=X'$

- Codage entropique: l'entropie, tout comme la vitesse de la lumière, est une limite fondamentale.
- Le débit R est la préoccupation principale.

❑ Compression avec perte: Irréversible $\rightarrow X \neq X'$

- Réduire le débit pour une distorsion minimale D entre X et X' \rightarrow compromis de débit distorsion.

Critère de distorsion (1)

□ Notation:

- $x_{n=1:N}$: Séquence originale de données.
- $y_{n=1:N}$: Séquence reconstruite.

□ Erreur quadratique moyenne (MSE):

$$\sigma^2 = \sigma_d^2 = \frac{1}{N} \sum_{n=1}^N (x_n - y_n)^2$$

□ Rapport signal à bruit (SNR):

$$SNR = \frac{\sigma_x^2}{\sigma_d^2} = \frac{\frac{1}{N} \sum_{n=1}^N x_n^2}{\frac{1}{N} \sum_{n=1}^N (x_n - y_n)^2}$$

□ Rapport signal à bruit en décibel (dBel):

$$SNR(\text{dB}) = 10 \log_{10} \frac{\sigma_x^2}{\sigma_d^2}$$

Critère de distorsion (2)

- ❑ Peak-signal-to-noise-ratio (PSNR):

$$PSNR(\text{dB}) = 10 \log_{10} \frac{x_{peak}^2}{\sigma_d^2}$$

x_{peak} : valeur crête du signal

- ❑ Moyenne des différences absolues (norme 1):

$$d_1 = \frac{1}{N} \sum_{n=1}^N |x_n - y_n|$$

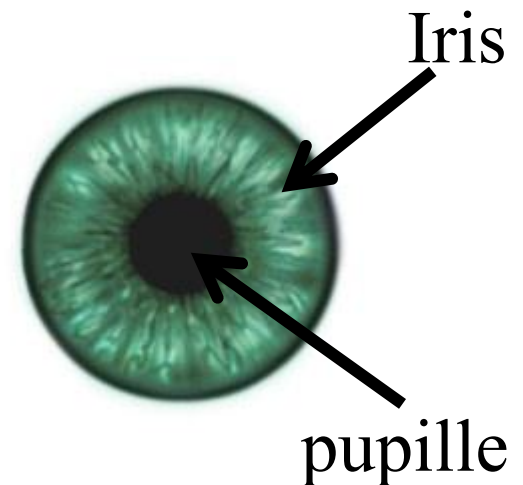
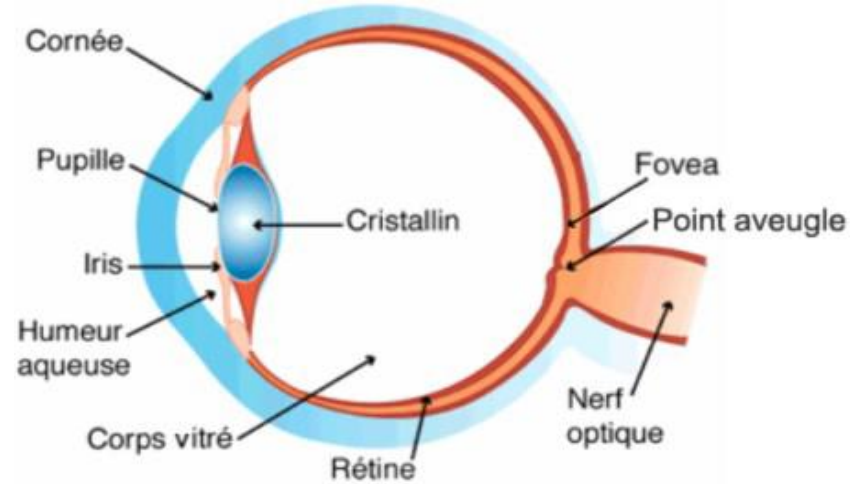
- ❑ Valeur maximale de l'erreur (norme infinie):

$$d_\infty = \max_n |x_n - y_n|$$

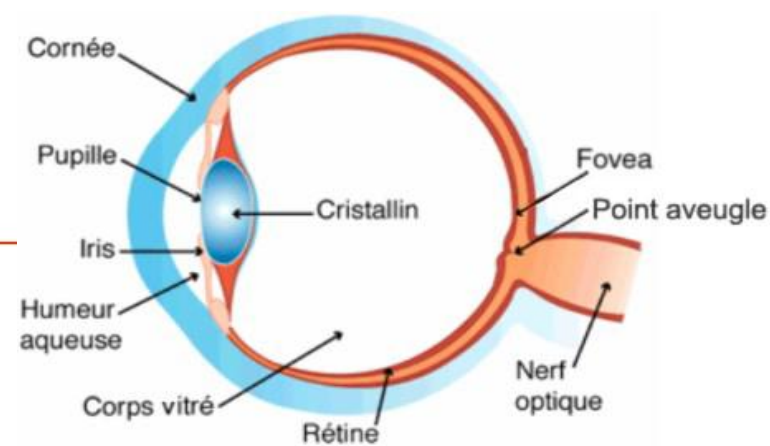
Systeme visuel humain (1)

□ Anatomie de l'œil:

- **Cornée:** protection du globe oculaire.
- **Iris:** diaphragme de contrôle de la quantité de lumière.
- **Pupille:** sa taille varie en fonction de la luminosité (diamètre=2 mm en pleine lumière, 8 mm dans l'obscurité).
- **Point aveugle:** endroit où le nerf optique vient se raccorder à la rétine.

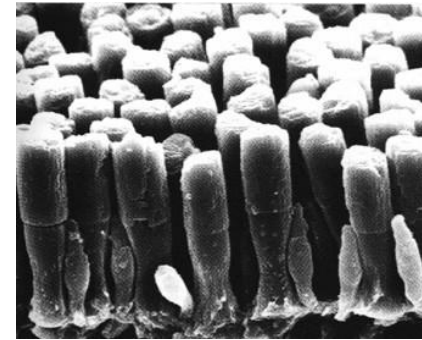


Systeme visuel humain (2)

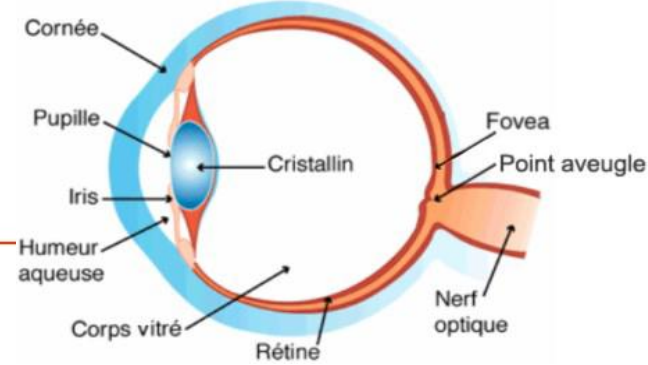


□ Anatomie de l'œil:

- **Rétine:** membrane où l'image vient se projeter. Elle est tapissée de photorécepteurs:
 - les cônes et
 - les bâtonnets.
- **Cristallin:** lentille convergente assurant la netteté de la vision grâce à la variation de sa courbure → mise au point.

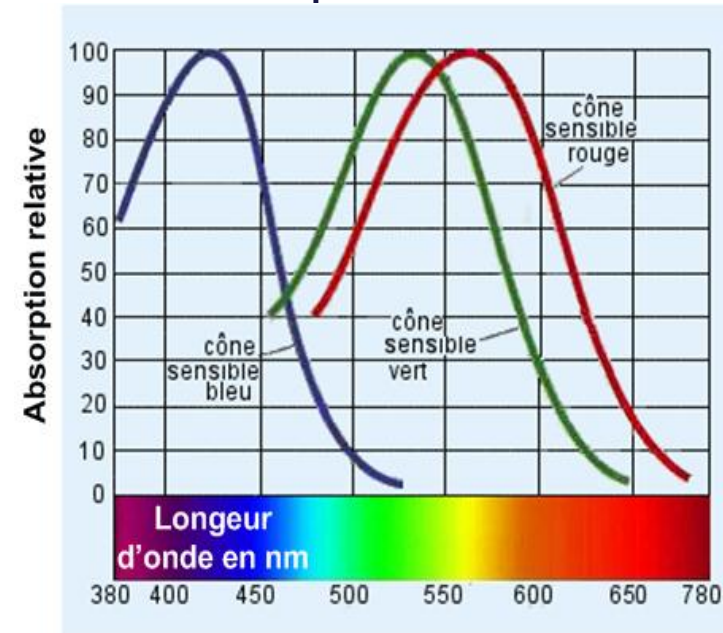


Systeme visuel humain (3)

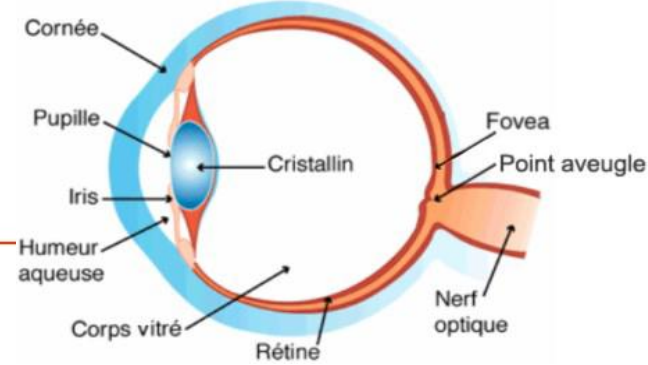


□ Anatomie de l'œil:

- **Cônes:** Il y a trois types de cônes. Chaque type est sensible à une longueur d'onde donnée du spectre visible (*short, medium, long wavelength*):
 - Spectre bleu (S): 482-465.
 - Spectre vert (M): 560-530.
 - Spectre rouge (L) : 730-622.
- Sont au nombre de 6.5 millions.
- Concentrés dans la fovéa.
- Moins nombreux que les bâtonnets.
- Assurent une meilleure résolution que les bâtonnets.
- Assurent une vision photopique (vision nette et colorée).



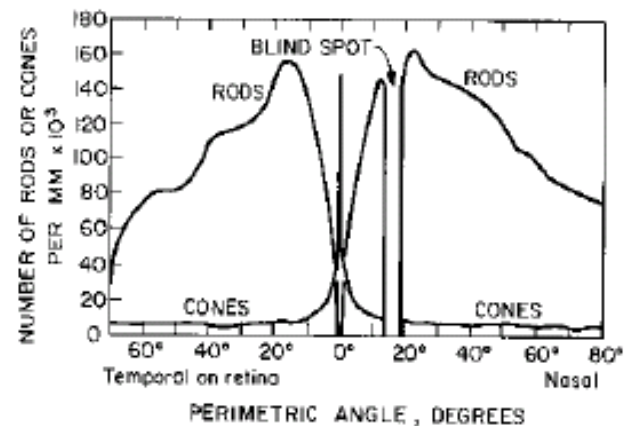
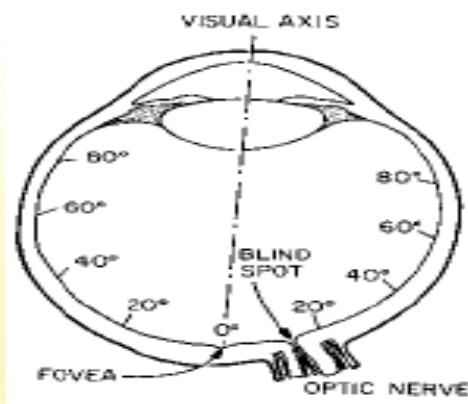
Systeme visuel humain (4)



□ Anatomie de l'œil:

- **Bâtonnets:** sont absents de la zone centrale (la fovéa) mais ils sont répartis sur tout le reste de la rétine.
- Ils sont plus sensibles à la lumière que les cônes.
- Assurent une vision scotopique (vision achromatique et moins nette) → Assurent la vision dans l'obscurité.

□ Répartition des cônes et des bâtonnets dans la rétine:



Systeme visuel humain (5)

□ Sensibilité au contraste:

- L'œil est sensible à une plage énorme d'intensité: 1 à 10^{10} .
- Cependant à un instant donné, on ne peut pas voir toute la plage de luminosité.
- L'œil s'adapte à un niveau de luminosité moyen.
- L'œil est sensible au contraste de la luminosité plutôt qu'à la valeur effective de la luminosité.

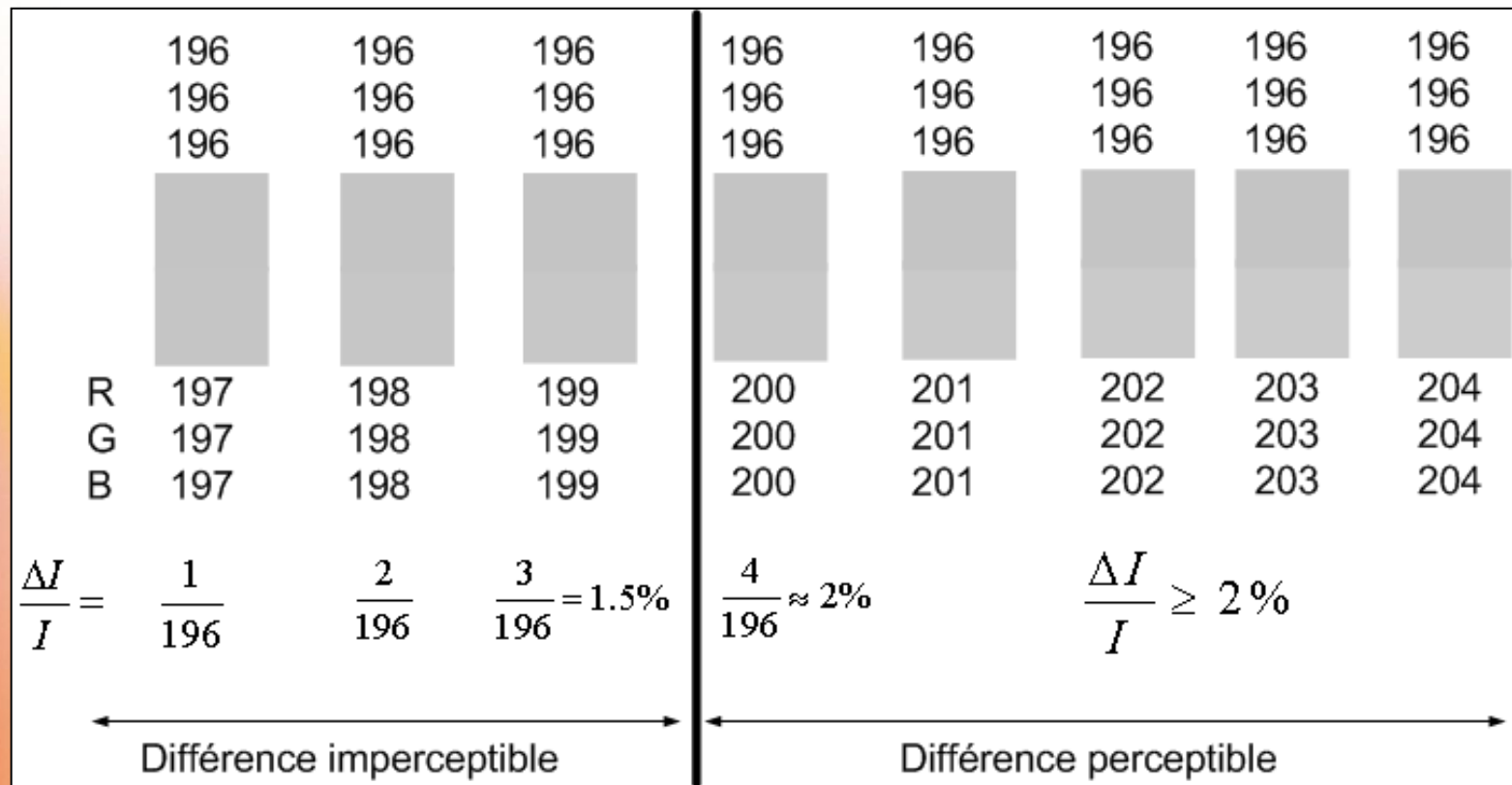
□ Rapport de Weber:

- La plus petite différence perçue (*just noticeable difference* [jnd]): ΔI
- Rapport de Weber: $\frac{\Delta I}{I} = \text{constante} = K$
- Pour la perception de la variation de luminosité: $\frac{\Delta I}{I} = K = 2\%$

Systeme visuel humain (6)

□ Rapport de Weber:

- Rapport de Weber: $\frac{\Delta I}{I} = K = 2\%$

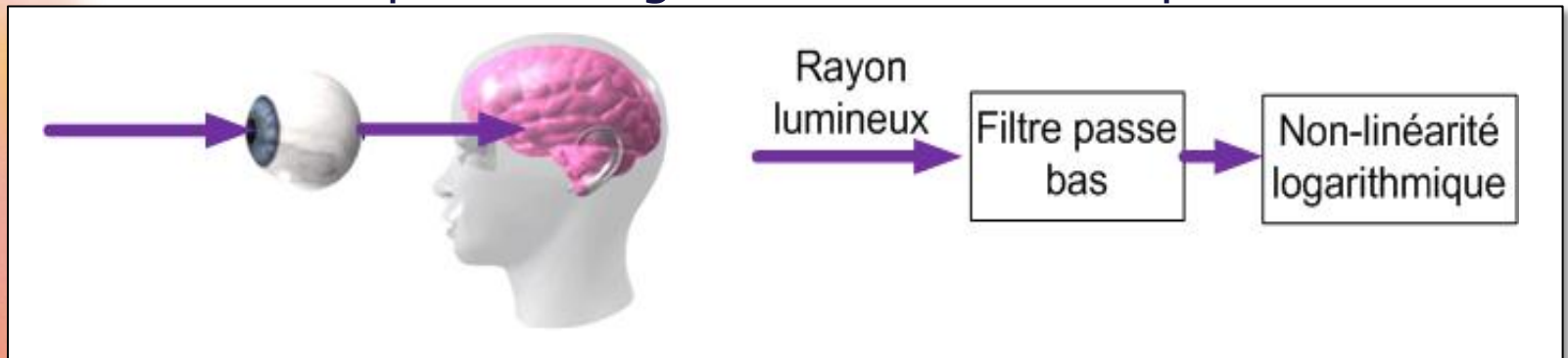


Systeme visuel humain (7)

□ Rapport de Weber: On a: $\frac{d \log I}{dI} = \frac{1}{I} \Rightarrow d \log I = \frac{dI}{I} \Rightarrow \Delta \log I = \frac{\Delta I}{I}$

Comme: $\frac{\Delta I}{I} = \text{constante} = K \Rightarrow \Delta \log I = K$

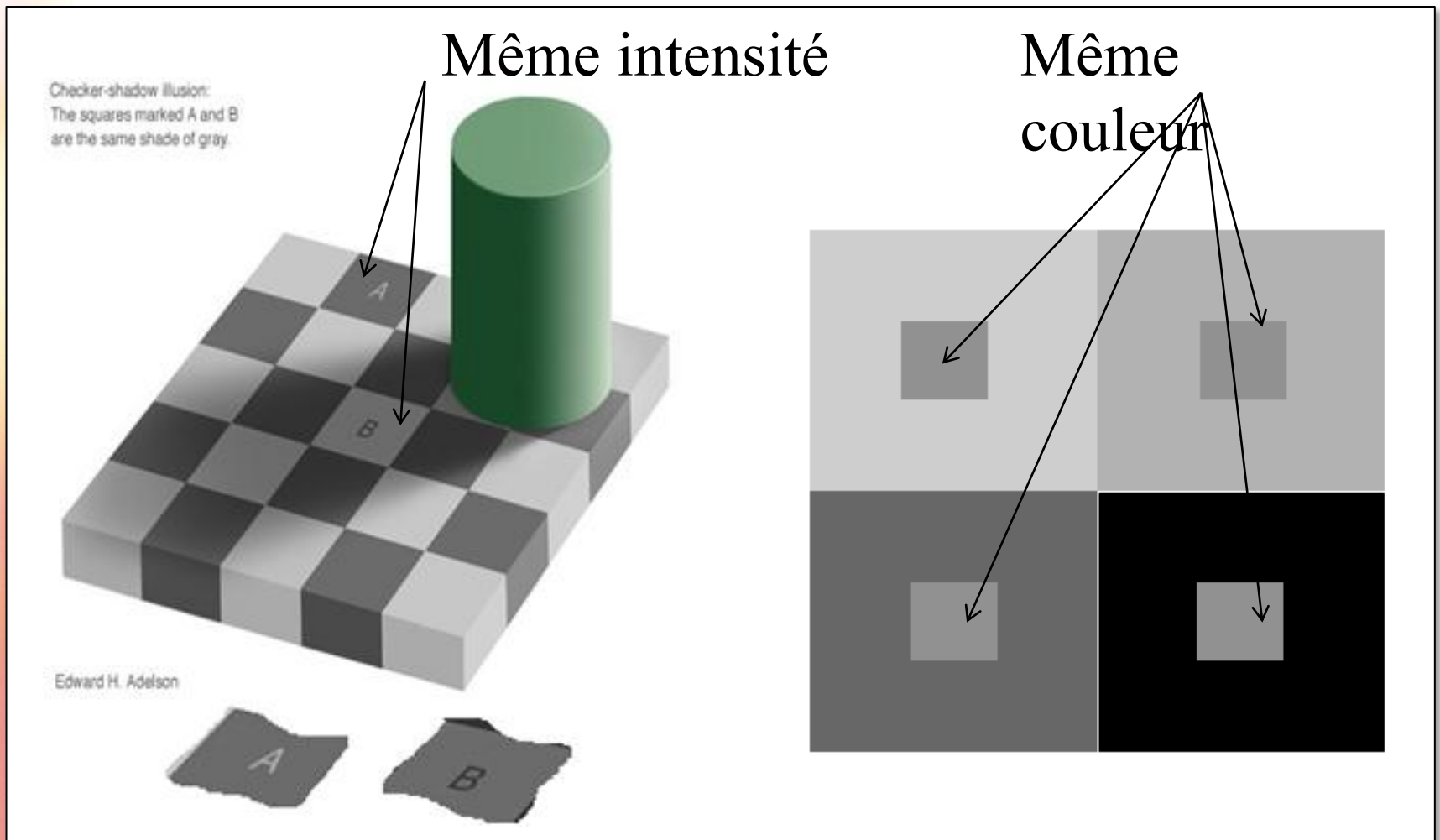
- La sensibilité de l'œil est une fonction logarithmique.
- L'œil → récepteur non linéaire mais logarithmique.
- On sait aussi que l'œil agit comme un filtre passe bas.



- → Compression introduisant une distorsion non perceptible.
- Exemple: distorsion dans les hautes fréquences.

Systeme visuel humain (8)

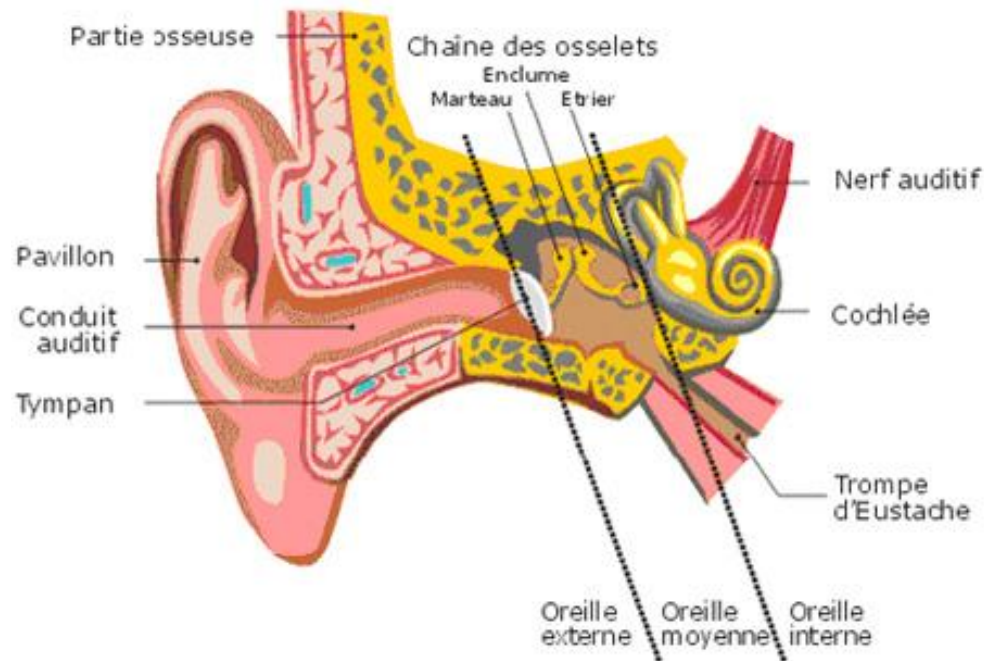
❑ Sensibilité au contraste: [illusion optique]



Perception auditive (1)

□ Système auditif humain:

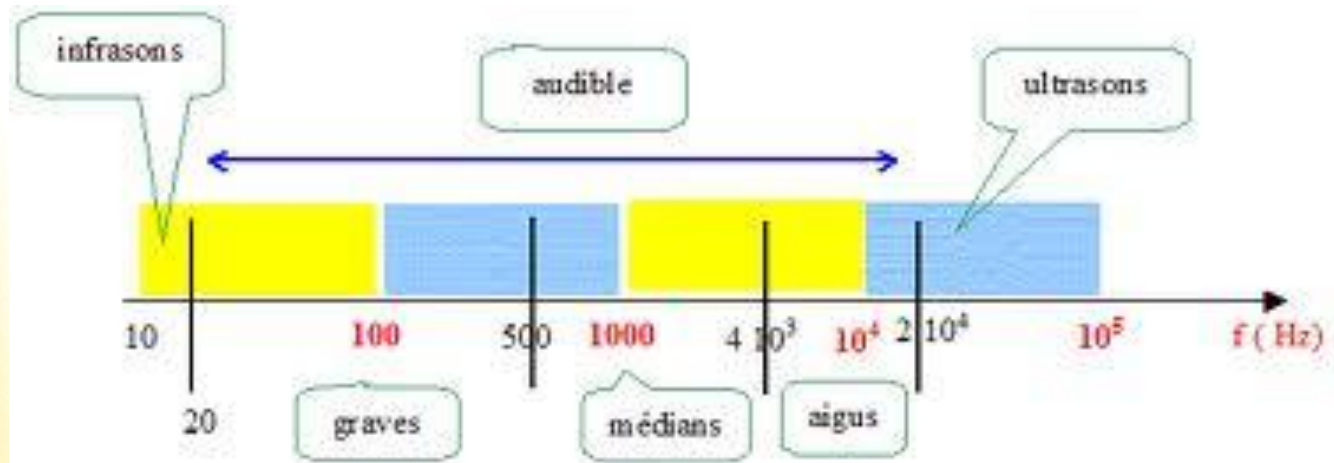
- Oreille externe: dirige l'onde sonore vers l'oreille moyenne.
- Oreille moyenne: cavité remplie d'air. Elle assure la conversion des ondes de pression en vibrations acoustiques.
- Oreille interne: conversion des vibrations en impulsions nerveuses.



Perception auditive (2)

□ Système auditif humain:

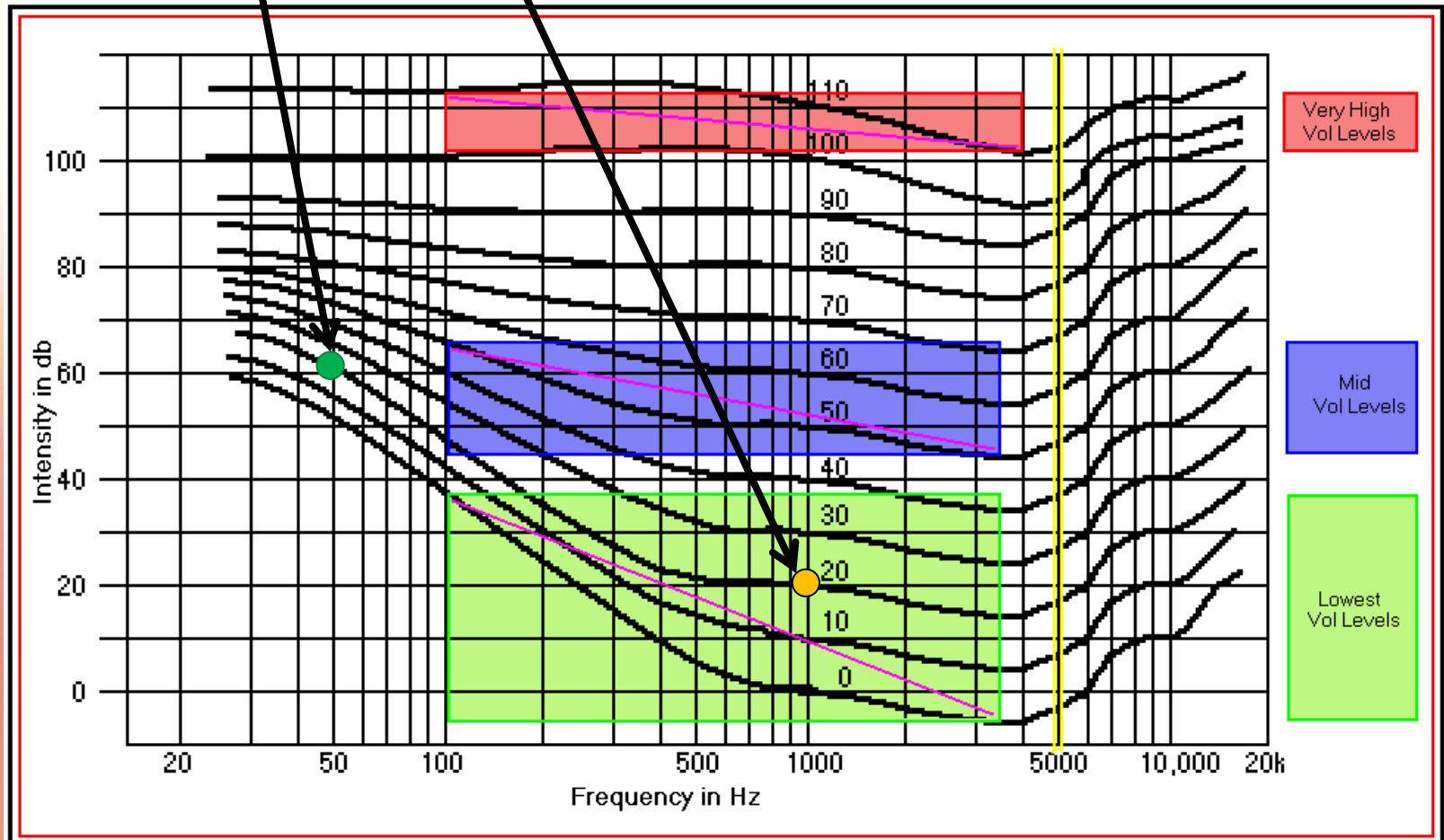
- Fréquence audible: de 20 Hz à 20 kHz.
- La perception auditive a plusieurs composantes non-linéaires.
- La réponse auditive dépend aussi de la fréquence:
- Une onde sonore à 1kHz et d'intensité 20 dB engendre la même pression acoustique qu'une onde à 50 Hz et d'intensité 60 dB.



Perception auditive (3)

□ Contours de niveau de pression acoustique égale :

- A(60dB à 50Hz) et B(20 dB à 1kHz) sont sur le même contour.



Perception auditive (4)

□ Bandes critiques:

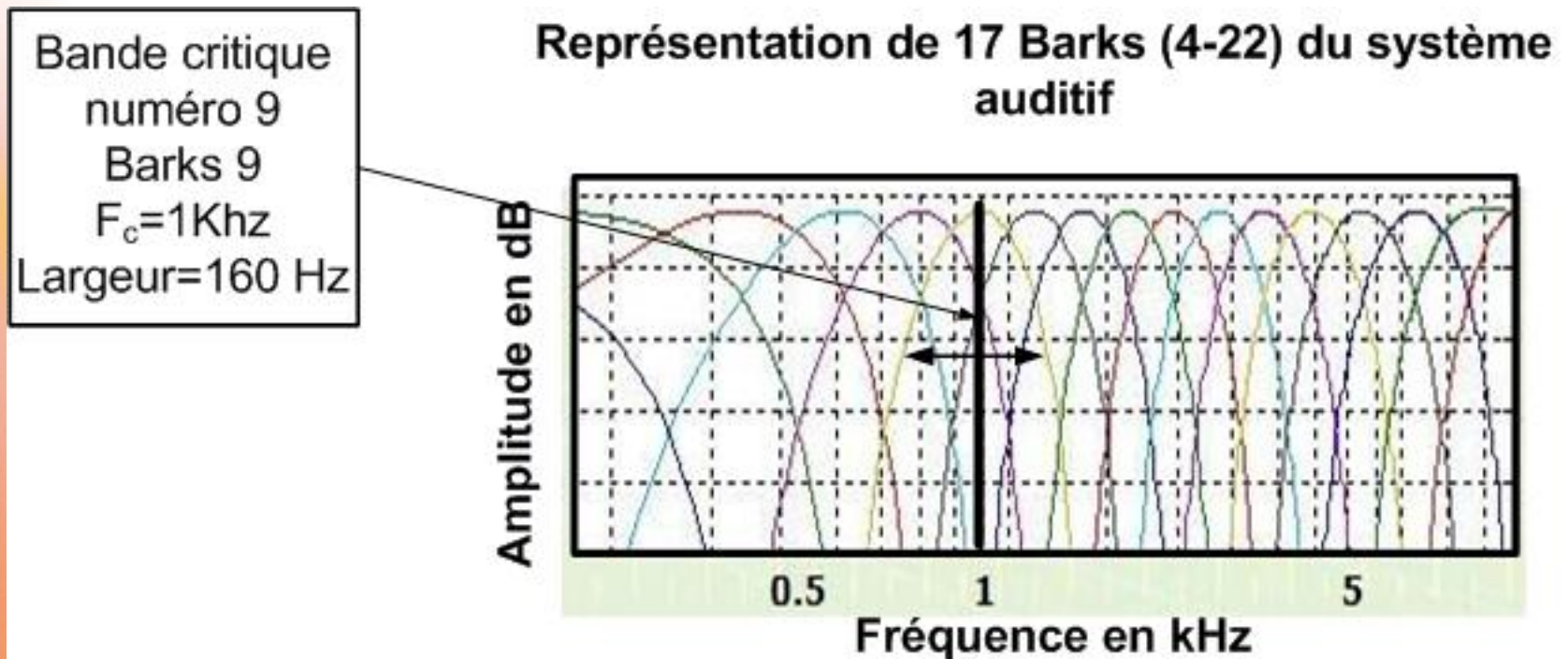
- La plage audible 20 Hz à 20 kHz comporte 24 bandes fréquentielles appelées les **bandes critiques**.

Numéro (Barks)	Fréquence Centrale (Hz)	Largeur (Hz)	Fréquences de coupure	
			Basse (Hz)	Haute (Hz)
1	50	—	—	100
2	150	100	100	200
3	250	100	200	300
4	350	100	300	400
5	450	110	400	510
6	570	120	510	630
7	700	140	630	770
8	840	150	770	920
9	1000	160	920	1080
10	1170	190	1080	1270
11	1370	210	1270	1480
12	1600	240	1480	1720
13	1850	280	1720	2000
14	2150	320	2000	2320
15	2500	380	2320	2700
16	2900	450	2700	3150
17	3400	550	3150	3700
18	4000	700	3700	4400
19	4800	900	4400	5300
20	5800	1100	5300	6400
21	7000	1300	6400	7700
22	8500	1800	7700	9500
23	10500	2500	9500	12000
24	13500	3500	12000	15500

Perception auditive (5)

❑ Bandes critiques:

- Oreille= Ensemble de filtres passe-bas qui se chevauchent.
- Si deux tonalités se trouvent dans la même bande critique, c'est la plus forte qui l'emporte.

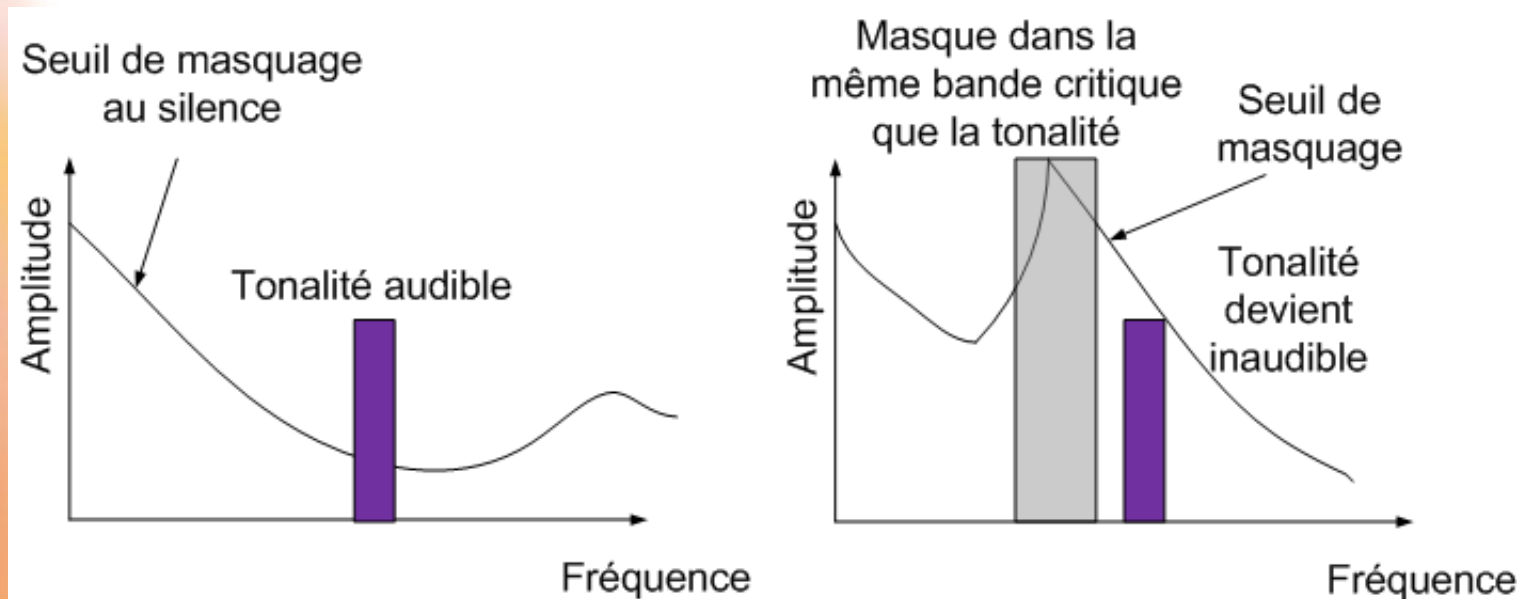


Perception auditive (6)

□ Effet du masquage:

- Seulement le bruit dans la même bande critique que la tonalité est susceptible de la masquer.
- → Même si l'on se trouve dans un grand vacarme si quelqu'un parle à une fréquence donnée on peut toujours l'entendre.

□ Seuil de masquage:



Théorie de l'information (rappel) (1)

□ Schéma de compression avec perte:



Théorie de l'information (2)

□ Information propre:

- Entropie

$$X \in \{x_0, x_1, \dots, x_{N-1}\} \Rightarrow H(X) = -\sum_{i=0}^{N-1} P(x_i) \log_2 P(x_i)$$

$$Y \in \{y_0, y_1, \dots, y_{N-1}\} \Rightarrow H(Y) = -\sum_{i=0}^{N-1} P(y_i) \log_2 P(y_i)$$

- Information propre d'un événement A

$$i(A) = \log_2 \frac{1}{P(A)} = -\log_2 P(A)$$

- Information propre conditionnelle

$$i(A|B) = \log_2 \frac{1}{P(A|B)} = -\log_2 P(A|B)$$

Théorie de l'information (3)

□ Exemple:

- Information propre conditionnelle:

$$i(A|B) = -\log P(A|B)$$

- Soit l'événement $B = \text{"Fraser n'a rien bu durant 2 jours"}$.
- Soit l'événement $A = \text{"Fraser a soif"}$.
- $P(A|B) \approx 1 \rightarrow i(A|B) \approx 0$.
- C'est-à-dire l'événement A ne fournit pas beaucoup d'information sachant qu'on connaît l'événement B .

Théorie de l'information (4)

□ Entropie conditionnelle:

$$H(X|Y) = -\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} P(x_i|y_j)P(y_j) \log_2 P(x_i|y_j) \Rightarrow H(X|Y) = -\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} P(x_i, y_j) \log_2 P(x_i|y_j)$$

$$H(Y|X) = -\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} P(x_i|y_j)P(y_j) \log_2 P(y_j|x_i) \Rightarrow H(Y|X) = -\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} P(x_i, y_j) \log_2 P(y_j|x_i)$$

□ Interprétation

- $H(X|Y)$: quantité d'incertitude restante concernant X sachant qu'on a reconstruit Y .
- La connaissance additionnelle de Y réduit l'incertitude concernant X .
- $\rightarrow H(X|Y) \leq H(X)$ [voir problème 5 pour la démonstration]

Théorie de l'information (5)

□ Exemple:

- Source équiprobable de symboles X à 4 bits.
- $X \in \{0, 1, \dots, 15\} \rightarrow P(x_i) = 1/16$.
- Compression: élimination du bit LSB puis multiplication par 2 $\rightarrow Y \in \{0, 2, \dots, 14\}$.

- Entropie de X :

$$H(X) = -\sum_{i=0}^{15} P(X = i) \log_2 P(X = i) = -\sum_{i=0}^{15} \frac{1}{16} \log_2 \frac{1}{16} = 4 \text{ bits}$$

- Entropie de Y :

$$X = b_0 b_1 b_2 b_3 \rightarrow Y = b_0 b_1 b_2 = j$$

$$Z = b_0 b_1 b_2 \bar{b}_3 \rightarrow Y = b_0 b_1 b_2 = j$$

$$P(Y = j) = P(X = j) + P(X = j + 1) = \frac{1}{16} + \frac{1}{16} = \frac{1}{8} \Rightarrow H(Y) = 3 \text{ bits}$$

Théorie de l'information (6)

- Probabilité conditionnelle $P(X|Y)$:

$$P(X = i|Y = j) = \begin{cases} 1/2, & \text{si } i = j \text{ ou } i = j + 1, \text{ pour } j = 0, 2, \dots, 14 \\ 0, & \text{sinon} \end{cases}$$

- Entropie conditionnelle $H(X|Y)$:

$$H(X|Y) = - \sum_{i=0}^{15} \sum_{\substack{j=0 \\ j \text{ pair}}}^{14} P(X = i|Y = j)P(Y = j) \log_2 P(X = i|Y = j)$$

$$H(X|Y) = - \sum_{\substack{j=0 \\ j \text{ pair}}}^{14} P(X = j|Y = j)P(Y = j) \log_2 P(X = j|Y = j)$$

$$- \sum_{\substack{j=0 \\ j \text{ pair}}}^{14} P(X = j + 1|Y = j)P(Y = j) \log_2 P(X = j + 1|Y = j)$$

$$H(X|Y) = - \sum_{\substack{j=0 \\ j \text{ pair}}}^{14} \left(\frac{1}{2} \times \frac{1}{8} \right) \log_2 \frac{1}{2} - \sum_{\substack{j=0 \\ j \text{ pair}}}^{14} \left(\frac{1}{2} \times \frac{1}{8} \right) \log_2 \frac{1}{2} = 1$$

- Intuitivement, Y informe sur 3 bits de X , il nous reste à connaître un seul bit $\rightarrow H(X|Y)=1$.

Théorie de l'information (7)

- Probabilité conditionnelle $P(Y|X)$: $P(Y = j|X = i) = \begin{cases} 1, & i = j \\ 1, & i = j + 1 \\ 0, & \text{sinon} \end{cases}$

- Entropie conditionnelle $H(Y|X)$:

$$H(Y|X) = - \sum_{i=0}^{15} \sum_{\substack{j=0 \\ j \text{ pair}}}^{14} P(X = i|Y = j)P(Y = j) \log_2 P(Y = j|X = i)$$

$$H(Y|X) = - \sum_{\substack{j=0 \\ j \text{ pair}}}^{14} P(X = j|Y = j)P(Y = j) \log_2 P(Y = j|X = i) \\ - \sum_{\substack{j=0 \\ j \text{ pair}}}^{14} P(X = j+1|Y = j)P(Y = j) \log_2 P(Y = j|X = j+1)$$

$$H(Y|X) = - \sum_{\substack{j=0 \\ j \text{ pair}}}^{14} \left(\frac{1}{2} \times \frac{1}{8} \right) \log_2 1 - \sum_{\substack{j=0 \\ j \text{ pair}}}^{14} \left(\frac{1}{2} \times \frac{1}{8} \right) \log_2 1 = 0$$

- Intuitivement, X contient 4 bits dont 3 bits forment Y .
Donc connaissant X , Y est parfaitement connu: $H(Y|X)=0$.

Théorie de l'information (8)

□ Information mutuelle moyenne

- Information mutuelle :

$$I(X;Y) = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} P(x_i, y_j) \log \frac{P(x_i|y_j)}{P(x_i)}$$

$$i(x_k; y_j) = \log \frac{P(x_k|y_j)}{P(x_k)}$$

$$I(X;Y) = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} P(x_i|y_j) P(y_j) \log \frac{P(x_i|y_j)}{P(x_i)}$$

$$I(X;Y) = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} P(x_i|y_j) P(y_j) \log P(x_i|y_j) - \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} P(x_i|y_j) P(y_j) \log P(x_i)$$

$$I(X;Y) = -H(X|Y) - \underbrace{\sum_{i=0}^{N-1} \log P(x_i) \sum_{j=0}^{M-1} P(x_i|y_j) P(y_j)}_{\text{Loi des probabilités totales : } P(x_i)}$$

$$I(X;Y) = -H(X|Y) - \sum_{i=0}^{N-1} P(x_i) \log P(x_i)$$

$$I(X;Y) = H(X) - H(X|Y)$$

Théorie de l'information (9)

□ Propriété

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = I(Y;X)$$

□ Retour à l'exemple:

- $H(X)=4$ bits.
- $H(X|Y)=1$ bit.
- $I(X;Y)=H(X)-H(X|Y)=3$ bits.
- ou encore:
- $H(Y)=3$ bits.
- $H(Y|X)=0$ bit.
- $I(X;Y)=H(Y)-H(Y|X)=3$ bits.

Théorie de l'information (10)

□ Entropie différentielle

- Soit X une variable aléatoire continue avec une fonction de densité de probabilité (pdf) $f_X(x)$.
- L'entropie différentielle est définie par:

$$h(X) = - \int_{-\infty}^{+\infty} f_X(x) \log f_X(x) dx$$

□ Exemple: X variable uniformément distribuée sur $[a,b]$

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{si } x \in [a,b) \\ 0, & \text{sinon} \end{cases}$$

$$h(X) = - \int_{-\infty}^{+\infty} f_X(x) \log f_X(x) dx = - \int_a^b \frac{1}{b-a} \log \frac{1}{b-a} dx = \log(b-a)$$

Théorie du débit distorsion (1)

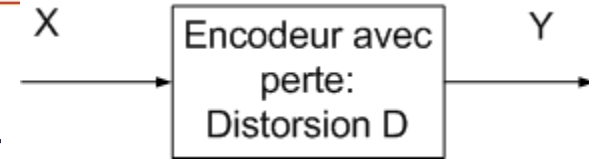
□ Définition

- Compromis entre la distorsion et le débit.
- Débit R : nombre de bits moyen par symbole.
- Fonction de débit distorsion $R(D)$: débit minimal nécessaire pour encoder une source en maintenant une distorsion ne dépassant pas D :

$$D = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} P(x_i, y_j) d(x_i, y_j)$$

□ Procédure

- Débit = $H(Y)$.
- Contrainte $D \leq D^*$.
- Recherche de tous les encodeurs assurant une distorsion $D \leq D^*$.
- Choix de l'encodeur qui génère la plus petite entropie.



Théorie du débit distorsion (2)

□ Calcul de la distorsion

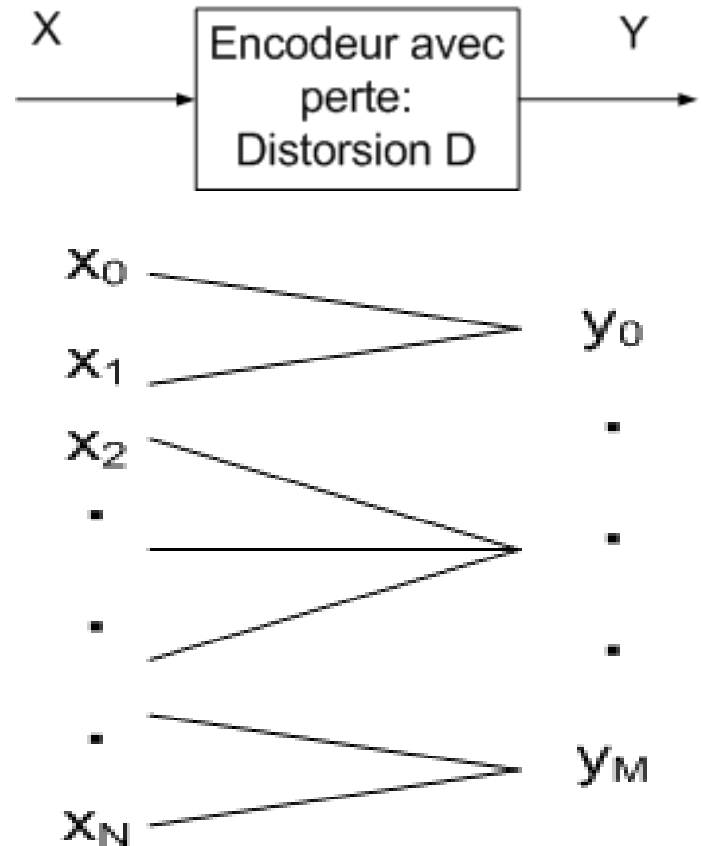
- La distorsion est déterminée par:

$$D = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} P(x_i, y_j) d(x_i, y_j)$$

$$= \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} P(y_j | x_i) P(x_i) d(x_i, y_j)$$

- Lorsque la connaissance de la valeur à l'entrée, x , spécifie parfaitement la valeur de reconstruction, y , la distorsion peut être évaluée. Dans cette situation on a:

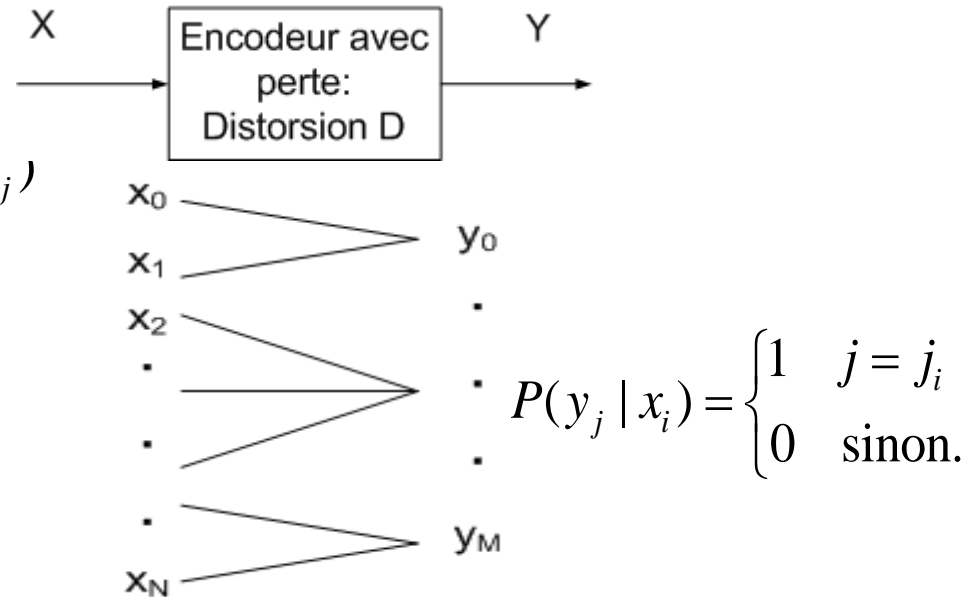
$$P(y_j | x_i) = \begin{cases} 1 & \text{pour un certain } j_i \text{ qui dépend uniquement de } i, \\ 0 & \text{sinon.} \end{cases}$$



Théorie du débit distorsion (3)

□ Calcul de la distorsion

$$\begin{aligned} D &= \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} \underbrace{P(y_j | x_i) P(x_i)}_{=0 \text{ si } j \neq j_i} d(x_i, y_j) \\ &= \sum_{i=0}^{N-1} \underbrace{P(y_{j_i} | x_i) P(x_i)}_{=1} d(x_i, y_{j_i}) \\ &= \sum_{i=0}^{N-1} P(x_i) d(x_i, y_{j_i}) \end{aligned}$$

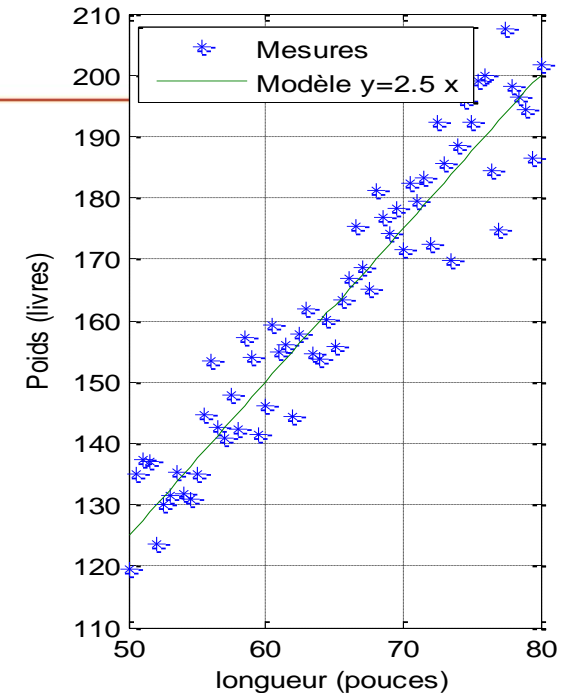


- Cependant la reconstruction, y , n'est pas toujours parfaitement déduite de la valeur à l'entrée, x .
- Il y a plusieurs techniques de compression efficaces qu'on ne traite pas dans le calcul de la distorsion.
- → Voir exemple ci-dessous.

Théorie du débit distorsion (4)

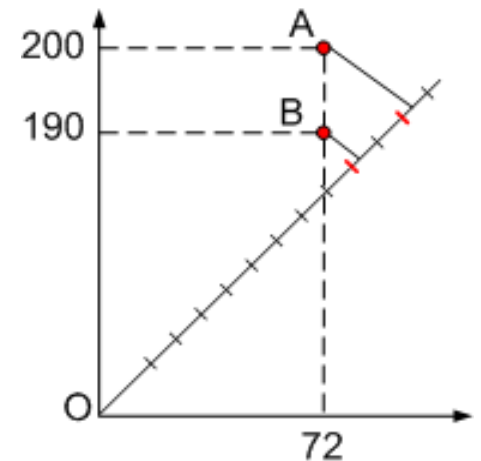
□ Exemple (7.5.1)

- Soit un ensemble de mesures de la longueur X (en pouces) et du poids Y (en livres).
- X et Y sont fortement corrélés: valeurs concentrées autour de $y=2.5 x$.



□ Schéma de compression de la paire (x,y)

- Projection des échantillons sur la droite $y=2.5 x \rightarrow$ point P
- Envoi de la distance séparant l'origine O du plus proche entier de P.



Théorie du débit distorsion (5)

□ Projection et reconstruction

- Encodage de A (72, 200)

$$P_A = \overrightarrow{OA} \cdot \vec{u} = 72 \times 0.37139 + 200 \times 0.92847 = 212.43$$

⇒ Code (A) = 212.

- Décodage de 212

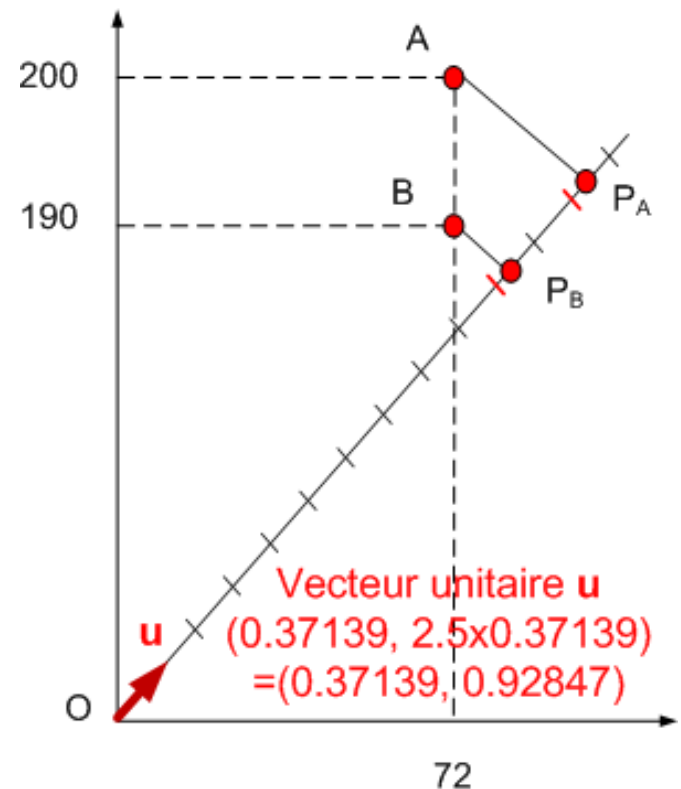
$$\sqrt{x^2 + y^2} = 212 \Rightarrow \sqrt{x^2 + (2.5x)^2} = 212$$

$$\Rightarrow \sqrt{7.25x^2} = 212$$

$$\Rightarrow x = \frac{212}{\sqrt{7.25}} = 78.735$$

$$\Rightarrow y = 2.5 \times x = 196.84$$

- Reconstruction: A^* (79, 197)

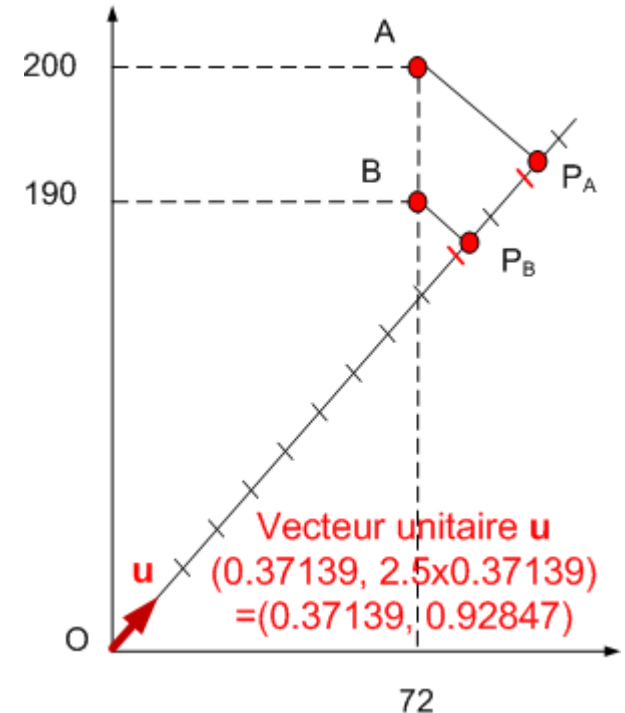


Théorie du débit distorsion (6)

□ Projection et reconstruction

- Encodage de A (72, 200) \Rightarrow Code (A)=212.
- Reconstruction: A^* (79, 197)

- Encodage de B (72, 190) \Rightarrow Code (B)=203.
- Reconstruction: B^* (75, 188)



- A et B ont la même longueur. Cependant la reconstruction A^* et B^* ont des longueurs différentes.
- Ceci est dû au fait la reconstruction de la longueur dépend aussi du poids.

Théorie du débit distorsion (7)

□ Forme générale de la distorsion
$$D = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} d(x_i, y_j) P(x_i) P(y_j | x_i)$$

- Facteur 1: mesure de la distorsion $d(x_i, y_j)$. Le choix de cette mesure dépend de l'application.
- Facteur 2: densité de probabilité de la source $P(x_i)$.
- Facteur 3: probabilité conditionnelle $P(y_j | x_i)$ décrivant le schéma de compression.
- Pour une source donnée $\{P(x_i)\}$ et une mesure $d(x_i, y_j)$ spécifiée, la distorsion est une fonction de la méthode de compression: $D = D(\{P(y_j | x_i)\})$.
- Pour une distorsion cible D^* , l'ensemble des schémas de compression est noté:

$$\Gamma = \{ \{P(y_j | x_i)\} \text{ tel que } D(\{P(y_j | x_i)\}) \leq D^* \}$$

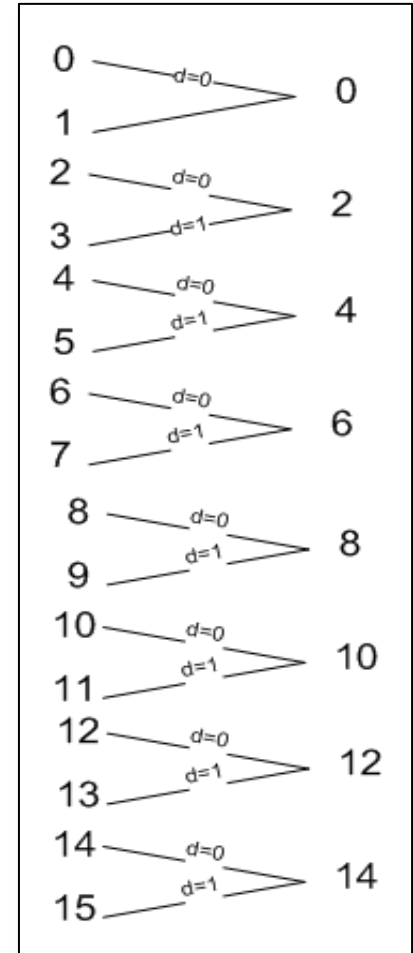
Théorie du débit distorsion (8)

❑ Exemple

- Source: $X \in \{0, 1, \dots, 15\} \rightarrow P(x_i) = 1/16$.
- Compression: $Y \in \{0, 2, \dots, 14\} \rightarrow H(Y) = 3$ bits.
- Mesure de la distorsion: $d(x_i, y_i) = (x_i - y_i)^2$

$$D = \sum_{i=0}^{N-1} P(x_i) (x_i - y_{j_i})^2 = \sum_{\substack{i=0 \\ i \text{ pair}}}^{15} \left(\frac{1}{16} \times 1 \right) + \sum_{\substack{i=0 \\ i \text{ impair}}}^{15} \left(\frac{1}{16} \times 0 \right) = 0.5$$

- ❑ Distorsion cible: Considérons $D^* = 225$,
- ❑ Même le mapping aléatoire (i.e. $P(y_j | x_i) = 1/8$) peut être considéré \rightarrow Pareillement $H(Y) = 3$ bits.
- ❑ Cependant le mapping aléatoire ne transmet aucune information. On ne peut alors rien transmettre: 0 bits.
- ❑ Ainsi, l'entropie $H(Y)$ ne peut être une mesure du débit.



Théorie du débit distorsion (9)

❑ Débit minimum défini par Shannon

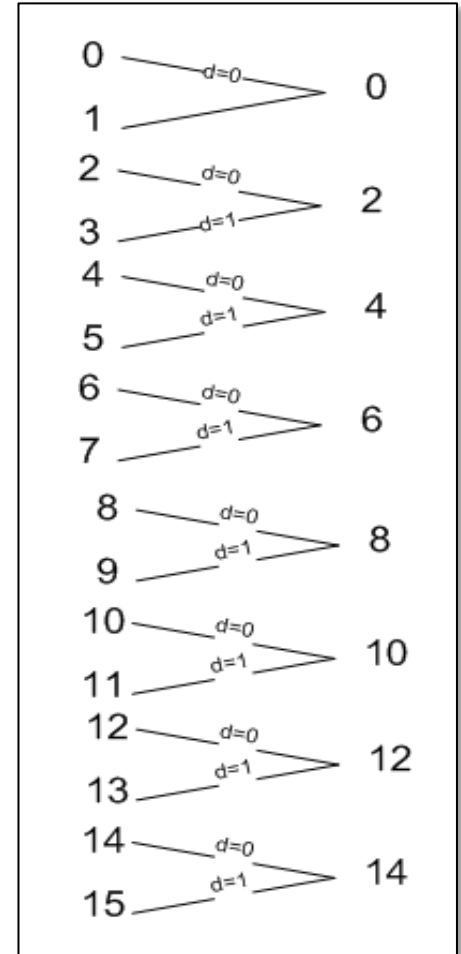
$$R(D) = \min_{P(y_j|x_j) \in \Gamma} I(X;Y)$$

❑ Reconsidérons l'exemple précédent

- $H(Y|X)=0$
- $H(Y)= 3$ bits
- $I(X;Y)=H(Y)-H(Y|X)=3$ bits

❑ Compression aléatoire

- Intuitivement, Y ne donne aucune information sur X donc :
- $I(X;Y)=0 \rightarrow$ débit = 0 au lieu de 3 bits.
- Dans ce genre de situation, où aucun transfert de données n'est effectué, il vaut mieux utiliser un débit égal à 0.



Exemple 3 : Fonction de débit-distorsion pour une source binaire (1)

- Source d'information binaire : $X \in \{0,1\}$, avec $P(0) = p$ et $P(1) = 1 - p$
- Mesure de distorsion : $d(x_i, y_j) = x_i \oplus y_j$

- Fonction de débit-distorsion :

$$I(X;Y) = H(X) - H(X|Y)$$

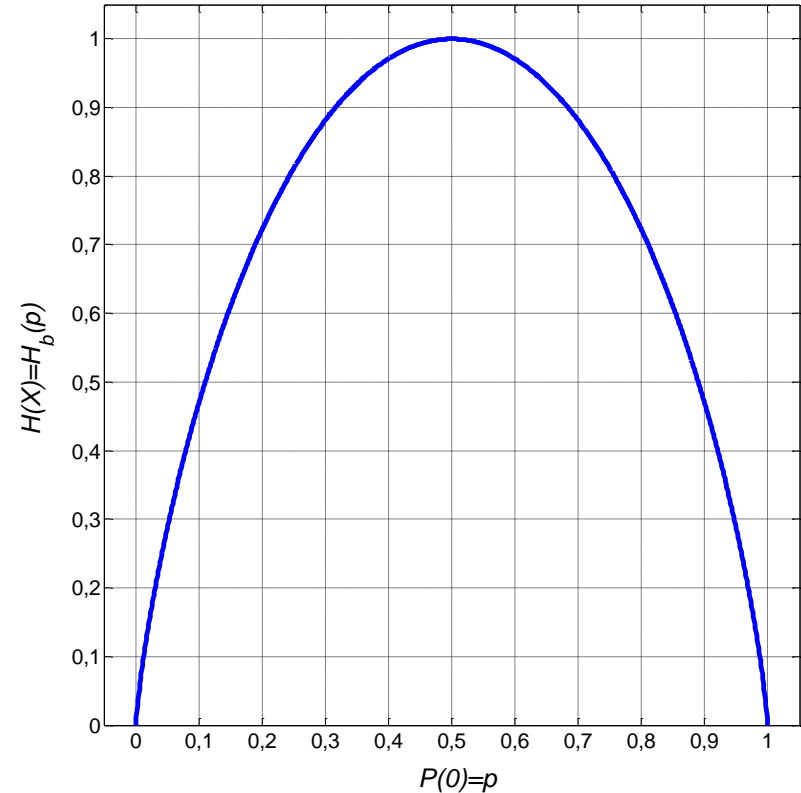
$$I(X;Y) = H(X) - H(X \oplus Y|Y)$$

$$I(X;Y) \geq H(X) - H(X \oplus Y)$$

- Entropie de X:

$$H(X) = -p \log_2 p - (1-p) \log_2 (1-p)$$

$$H(X) = H_b(p) = H_b(1-p)$$



Exemple 3 : Fonction de débit-distorsion pour une source binaire (2)

- Fonction de débit-distorsion : minimiser l'information mutuelle tout en respectant le critère de distorsion:

$I(X;Y) \geq H(X) - H(X \oplus Y)$ avec :

$\Rightarrow H(X) = H_b(p) = -p \log_2 p - (1-p) \log_2 (1-p)$: déterminé par la source

$\Rightarrow H(X \oplus Y) = H_b(P(X \oplus Y = 1))$: à maximiser sur l'ensemble $\{P(x_i | y_j)\}$

\Rightarrow sujet à la contrainte de distorsion : $E[d(x_i, y_j)] \leq D$

Avec la mesure de distorsion $d(x_i, y_j) = x_i \oplus y_j$, on a $E[d(x_i, y_j)] = E[(x_i \oplus y_j)]$:

$$E[d(x_i, y_j)] = 0 \times P(X=0, Y=0) + 1 \times P(X=0, Y=1) + 1 \times P(X=1, Y=0) + 0 \times P(X=1, Y=1)$$

$$E[d(x_i, y_j)] = P(X=0, Y=1) + P(X=1, Y=0)$$

$$E[d(x_i, y_j)] = pP(Y=1|X=0) + (1-p)P(Y=0|X=1)$$

Donc, il faut que $E[d(x_i, y_j)] = P(X \oplus Y = 1) \leq D$

Exemple 3 : Fonction de débit-distorsion pour une source binaire (3)

En tenant compte de la contrainte de $E[d(x_i, y_j)] = P(X \oplus Y = 1) \leq D$,

on peut minimiser le débit requis $I(X; Y)$ en maximisant le terme $H(X \oplus Y)$:

$$I(X; Y) \geq H(X) - H(X \oplus Y)$$

$$I(X; Y) \geq H_b(p) - H_b(P(X \oplus Y = 1))$$

$$I(X; Y) \geq H_b(p) - H_b(D)$$

Pour $0 \leq D < p \leq \frac{1}{2}$, le débit minimal avec $E[d(x_i, y_j)] \leq D$ est :

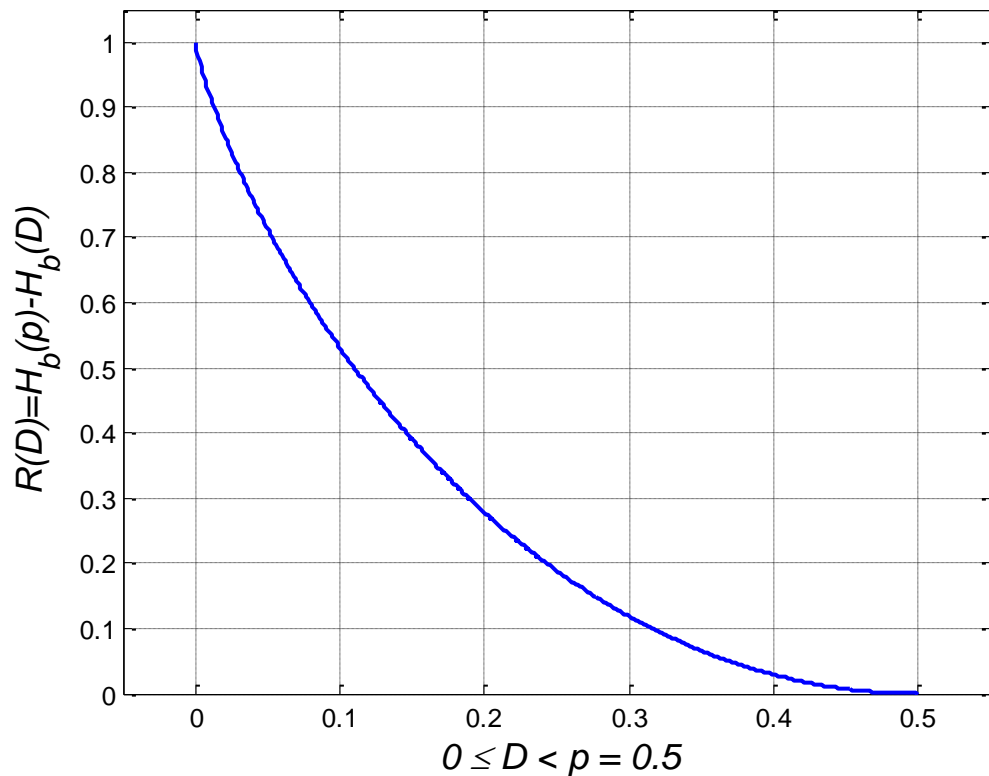
$$I(X; Y) \geq R(D) = H_b(p) - H_b(D)$$

Par symétrie de la fonction d'entropie $H_b(p)$, on obtient :

$$R(D) = \begin{cases} H_b(p) - H_b(D) & \text{pour } D < \min\{p, 1-p\} \\ 0 & \text{ailleurs} \end{cases}$$

Exemple 3 : Fonction de débit-distorsion pour une source binaire (4)

$$R(D) = \begin{cases} H_b(p) - H_b(D) & \text{pour } D < \min\{p, 1-p\} \\ 0 & \text{ailleurs} \end{cases}$$



Exemple 4 : Fonction de débit-distorsion pour une source gaussienne (1)

- Source d'information gaussienne de moyenne nulle et
- de variance σ^2
- Mesure de distorsion (erreur quadratique moyenne) :

$$d(x, y) = (x - y)^2$$

- Critère de distorsion (contrainte de distorsion) :

$$E[(x - y)^2] \leq D$$

- La fonction de débit-distorsion $R(D)$ dépend de l'information mutuelle entre X et Y :

$$I(X; Y) = h(X) - h(X|Y)$$

$$I(X; Y) = h(X) - h(X - Y|Y)$$

- Ici, $h(X)$ est l'entropie différentielle de la source gaussienne.

Exemple 4 : Fonction de débit-distorsion pour une source gaussienne (2)

- On doit minimiser l'information mutuelle $I(X;Y)$ tout en respectant la contrainte de distorsion D . $h(X-Y)$ est alors l'entropie différentielle d'une gaussienne de variance D :

$$I(X;Y) \geq h(X) - h(X-Y)$$

$$I(X;Y) \geq \frac{1}{2} \log(2\pi e\sigma^2) - \frac{1}{2} \log(2\pi eD)$$

$$I(X;Y) \geq \frac{1}{2} \log\left(\frac{\sigma^2}{D}\right)$$

- Si Y est une gaussienne de moyenne nulle et de variance $(\sigma^2 - D)$

$$f_{X|Y}(x|y) = \frac{1}{\sqrt{2\pi D}} \exp\left(-\frac{x^2}{2D}\right) \quad \Rightarrow \quad I(X;Y) = \frac{1}{2} \log\left(\frac{\sigma^2}{D}\right)$$

- Fonction de débit-distorsion :

$$R(D) = \begin{cases} \frac{1}{2} \log\left(\frac{\sigma^2}{D}\right) & \text{pour } D < \sigma^2 \\ 0 & \text{pour } D \geq \sigma^2 \end{cases}$$

Modèles de probabilité (1)

❑ Modélisation pour la compression sans perte

- L'objectif est de reconstruire la valeur exacte.
- La modélisation est en fait une estimation empirique qui donne le même estimé à l'encodeur et au décodeur.
- Une correction d'estimation est encodée pour assurer une reconstruction exacte → codage résiduel.

❑ Modélisation pour la compression avec perte

- Modélisation de l'allure générale.
- Utilisation de fonctions de distribution de probabilité analytiquement malléables.
- La modélisation améliore à la fois la compression et la reconstruction.

Modèles de probabilité (2)

□ Distribution uniforme

- Modèle d'ignorance: toutes les valeurs sont équiprobables.

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{pour } a \leq x \leq b \\ 0 & \text{ailleurs.} \end{cases}$$

□ Distribution gaussienne

- Distribution communément utilisée pour deux raisons:
 1. Mathématiquement malléable.
 2. Constitue la limite de la somme de plusieurs variables aléatoire: loi de la limite centrale.

$$N(\mu, \sigma^2) \rightarrow f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Modèles de probabilité (3)

□ Distribution Laplacienne

- Distribution concentrée autour de zéro.
- La voix comporte énormément de silences.
- La différence entre pixels voisins d'une image est concentrée autour de 0.
- Dans ces situations la distribution gaussienne n'est pas efficace.
- Fonction de distribution d'une variable aléatoire de moyenne nulle avec une allure Laplacienne de variance σ^2 est donnée par:

$$f_x(x) = \frac{1}{\sqrt{2\sigma^2}} \exp\left(-\frac{\sqrt{2}|x|}{\sigma}\right)$$

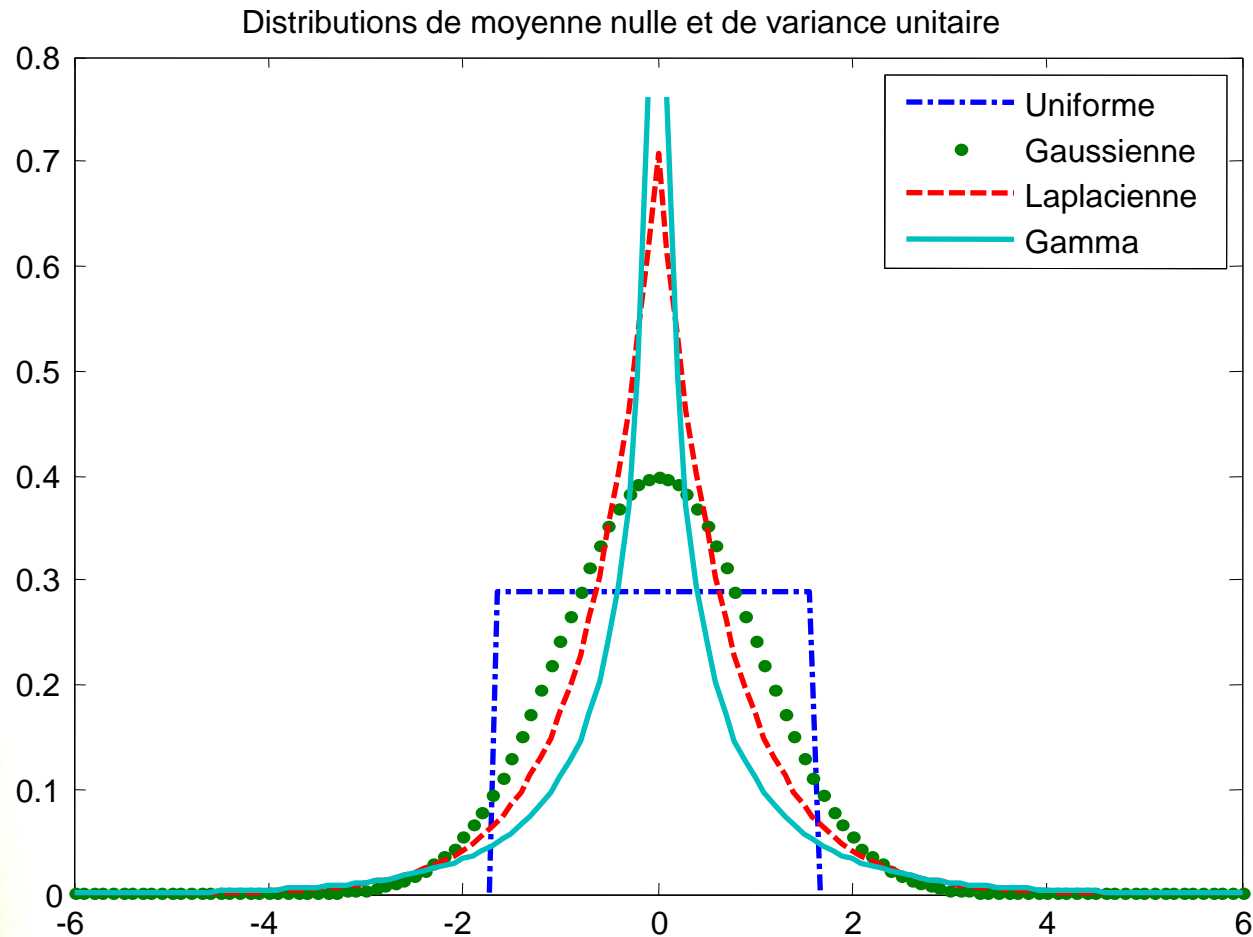
Modèles de probabilité (4)

□ Distribution Gamma

- Distribution encore plus concentrée autour de zéro.
- Mais moins malléable.
- Fonction de distribution d'une variable aléatoire de moyenne nulle avec une allure Gamma de variance σ^2 donnée par:

$$f_X(x) = \frac{\sqrt[4]{3}}{\sqrt{8\pi\sigma|x|}} \exp\left(-\frac{\sqrt{3}|x|}{2\sigma}\right)$$

Modèles de probabilité (5)



Modèles de systèmes linéaires (1)

□ ARMA (N,M) (modèles autorégressifs et moyenne mobile)

- Equation aux différences:

$$x_n = \underbrace{\sum_{i=1}^N a_i x_{n-i}}_{\text{Autorégressive}} + \underbrace{\sum_{j=1}^M b_j \varepsilon_{n-j}}_{\text{Moyenne mobile}} + \varepsilon_n$$

- $\{x_n\}$ échantillons du processus à modéliser.
- $\{\varepsilon_n\}$ séquence de bruit blanc stationnaire au sens large:

$$R_{\varepsilon\varepsilon}(k) = \begin{cases} \sigma_\varepsilon^2 & \text{pour } k = 0 \\ 0 & \text{sinon.} \end{cases}$$

Modèles de systèmes linéaires (2)

□ ARMA (N,0) = AR(N) (modèles autorégressifs)

- Équation aux différences:

$$x_n = \underbrace{\sum_{i=1}^N a_i x_{n-i}}_{\text{Autorégressive}} + \varepsilon_n$$

- Modèle utilisé dans la compression de la voix.
- Pas besoin de connaître tout l'historique de la séquence mais seulement des N échantillons passés:

$$P(x_n | x_{n-1}, x_{n-2}, \dots) = P(x_n | x_{n-1}, x_{n-2}, \dots, x_{n-N})$$

- → Il s'agit d'un modèle de Markov d'ordre N .

Modèles de systèmes linéaires (3)

□ Fonction d'autocorrélation $R_{xx}(k)$

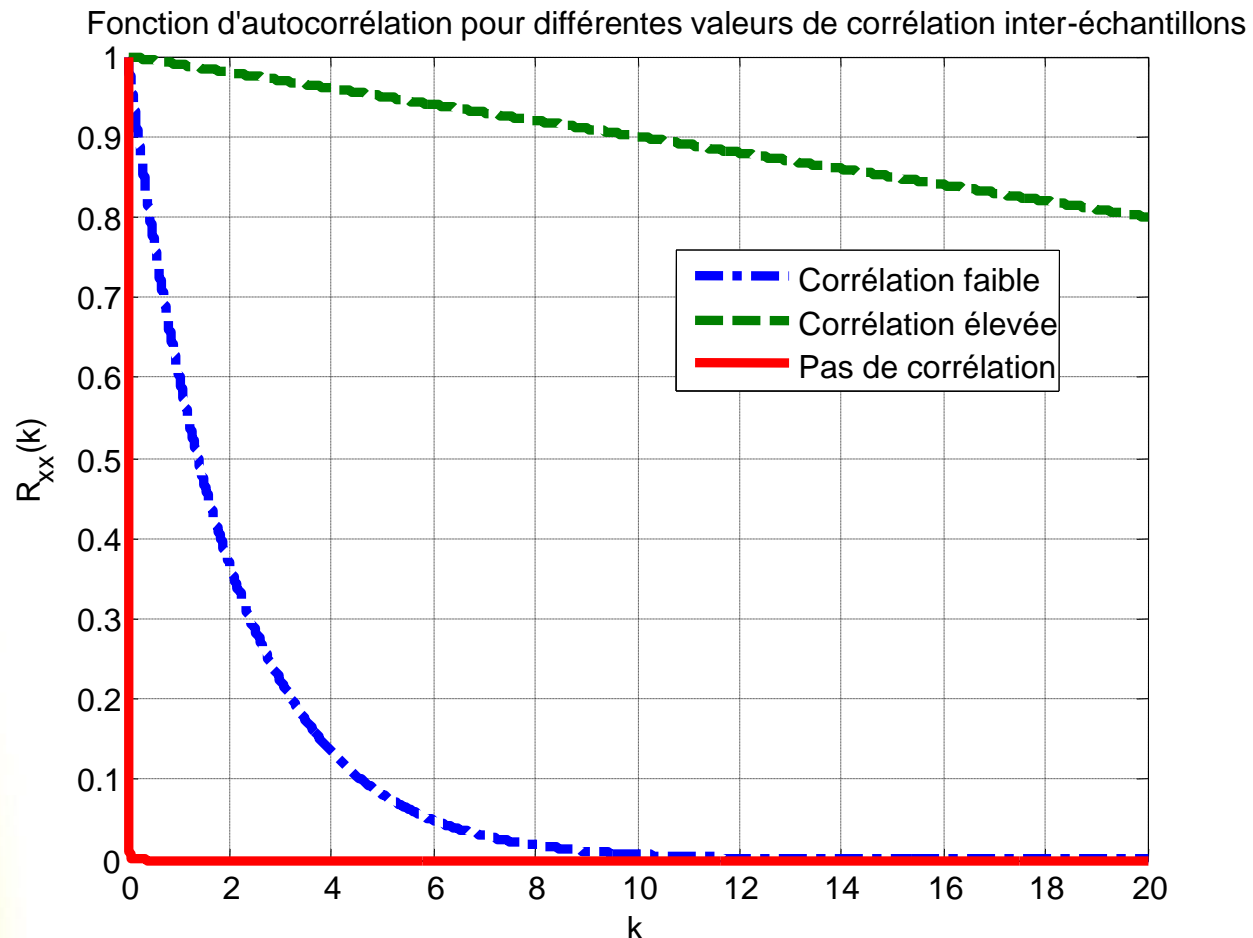
- Définition:

$$R_{xx}(k) = E[x_n x_{n-k}]$$

- Décrit le comportement inter-échantillons.
- Si $R_{xx}(k)$ décroît lentement → Corrélation inter-échantillons élevée.
- Si $R_{xx}(k)$ décroît rapidement → Corrélation inter-échantillons faible.
- Aucune corrélation inter-échantillons (cas du bruit blanc) → fonction d'autocorrélation = impulsion de Dirac.

Modèles de systèmes linéaires (4)

□ Fonction d'autocorrélation $R_{xx}(k)$ $R_{xx}(k) = E[x_n x_{n-k}]$



Modèles de systèmes linéaires (5)

$$\begin{aligned} R_{xx}(k) &= E[x_n x_{n-k}] = E\left[\left(\sum_{i=1}^N a_i x_{n-i} + \varepsilon_n\right) x_{n-k}\right] \\ &= E\left[\sum_{i=1}^N a_i x_{n-i} x_{n-k}\right] + E[\varepsilon_n x_{n-k}] \\ &= \sum_{i=1}^N a_i E[x_{n-i} x_{n-k}] + E[\varepsilon_n x_{n-k}] \quad (\text{permutation somme-espérance}) \\ &= \begin{cases} (1) \rightarrow \sum_{i=1}^N a_i R_{xx}(k-i) & \text{pour } k > 0 \\ (2) \rightarrow \sum_{i=1}^N a_i R_{xx}(i) + \sigma_\varepsilon^2 & \text{pour } k = 0 \end{cases} \end{aligned}$$

(1) ε_n et x_{n-k} sont indépendants $\Rightarrow E[\varepsilon_n x_{n-k}] = 0$ pour $k > 0$

$$(2) E[\varepsilon_n x_n] = E\left[\left(\sum_{i=1}^N a_i x_{n-i} + \varepsilon_n\right) \varepsilon_n\right] = \sum_{i=1}^N a_i \underbrace{E[x_{n-i} \varepsilon_n]}_{=0} + \underbrace{E[\varepsilon_n \varepsilon_n]}_{\sigma_\varepsilon^2}$$

Modèles de systèmes linéaires (6)

□ Exemple: Fonction d'autocorrélation d'un processus AR(3)

$$R_{xx}(k) = \begin{cases} \sum_{i=1}^N a_i R_{xx}(k-i) & \text{pour } k > 0 \\ \sum_{i=1}^N a_i R_{xx}(i) + \sigma_\varepsilon^2 & \text{pour } k = 0 \end{cases}$$

$$R_{xx}(0) = a_0 R_{xx}(0) + a_1 R_{xx}(1) + a_2 R_{xx}(2) + \sigma_\varepsilon^2 \quad [\text{on a: } R_{xx}(k) = R_{xx}(-k)]$$

$$3 \text{ équations et 3 inconnues: } \begin{cases} R_{xx}(1) = a_0 R_{xx}(0) + a_1 R_{xx}(1) + a_2 R_{xx}(2) \\ R_{xx}(2) = a_0 R_{xx}(1) + a_1 R_{xx}(0) + a_2 R_{xx}(1) \\ R_{xx}(3) = a_0 R_{xx}(2) + a_1 R_{xx}(1) + a_2 R_{xx}(0) \end{cases}$$

Connaissance de $R_{xx}(k) \Rightarrow$ les 3 coefficients du modèle AR(3) $\{a_1, a_2, a_3\}$

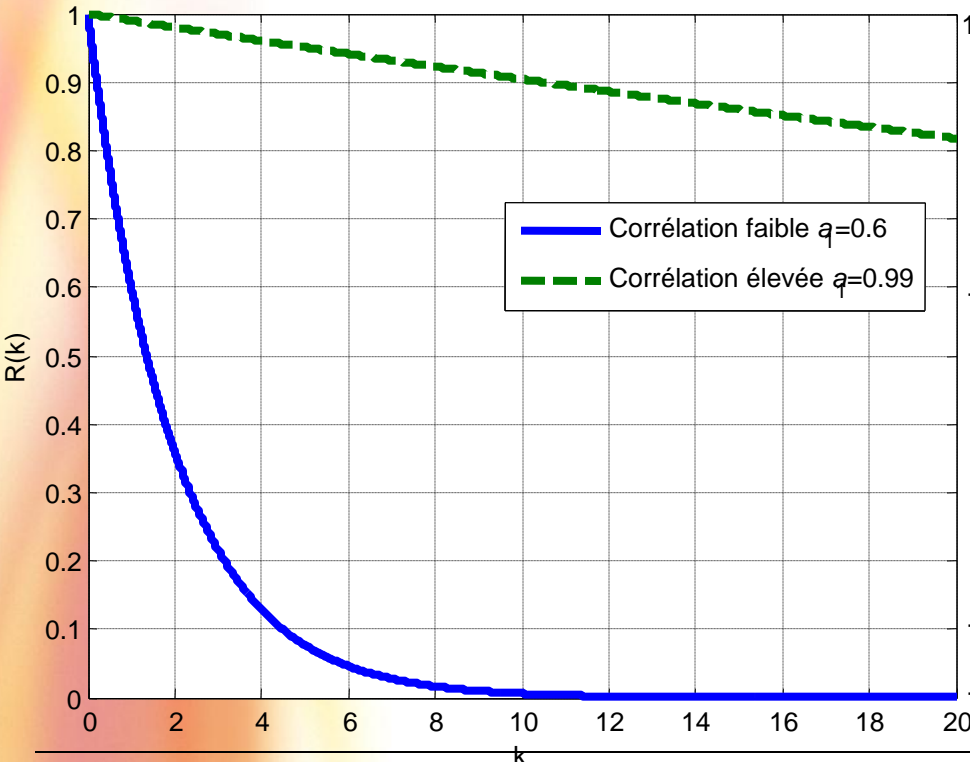
Connaissance des coefficients du modèles et de $\sigma_\varepsilon^2 \Rightarrow R_{xx}(k)$

Modèles de systèmes linéaires (7)

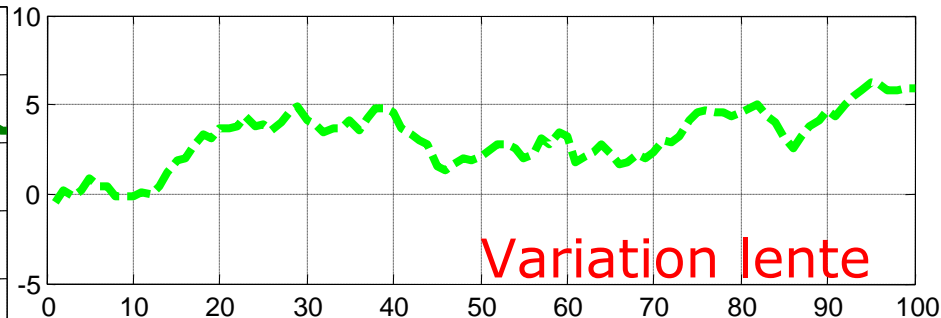
□ Exemple: Relation entre le comportement temporel et fonction d'autocorrélation:

Source AR(1): $x_n = a_1 x_{n-1} + \varepsilon_n$ Prob.8 \rightarrow $R_{xx}(k) = \frac{1}{1 - a_1^2} a_1^k \sigma_\varepsilon^2$

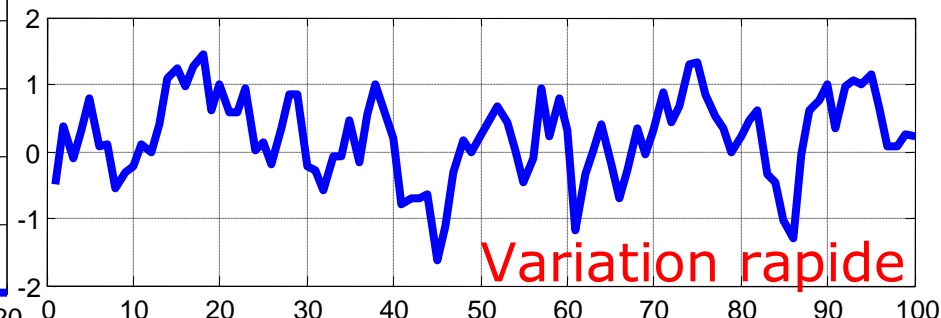
Fonction d'autocorrélation normalisée d'un processus AR(1): $\sigma_\varepsilon=0.5$



Echantillons d'un processus AR(1) avec $a_1=0.99$



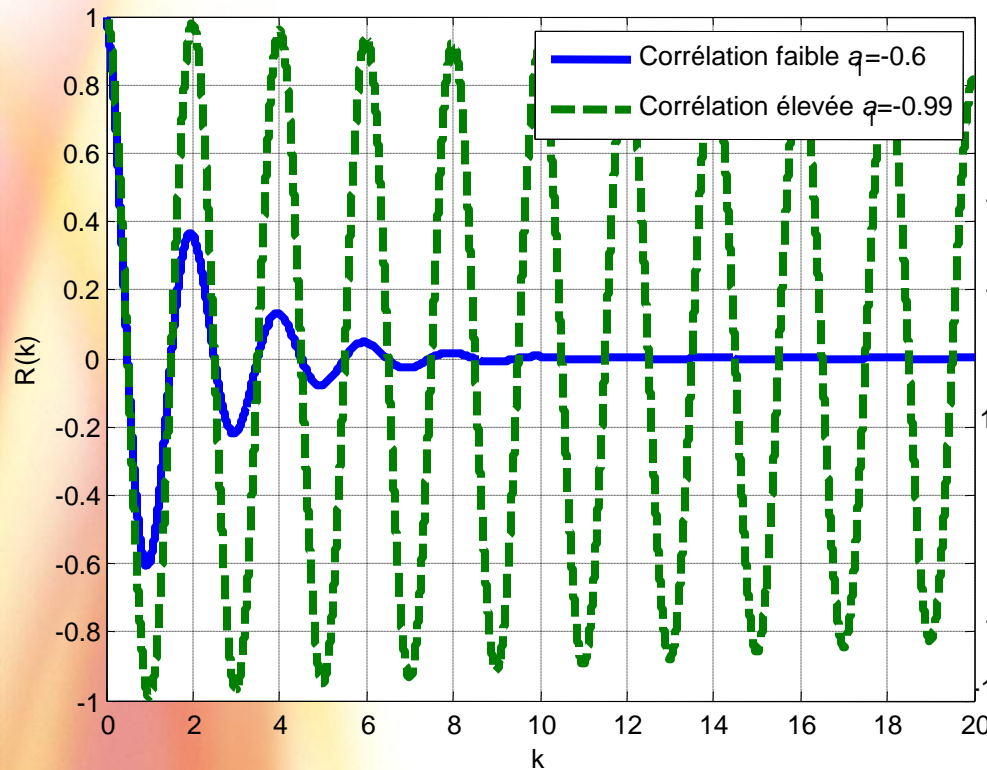
Echantillons d'un processus AR(1) avec $a_1=0.6$



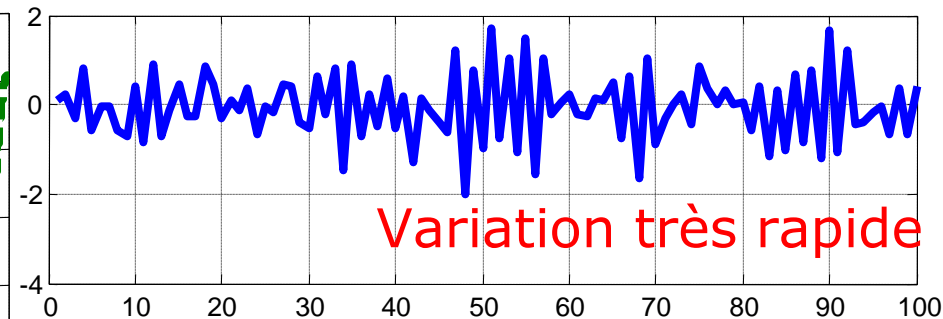
Modèles de systèmes linéaires (8)

□ Processus AR(1) avec un coefficient négatif

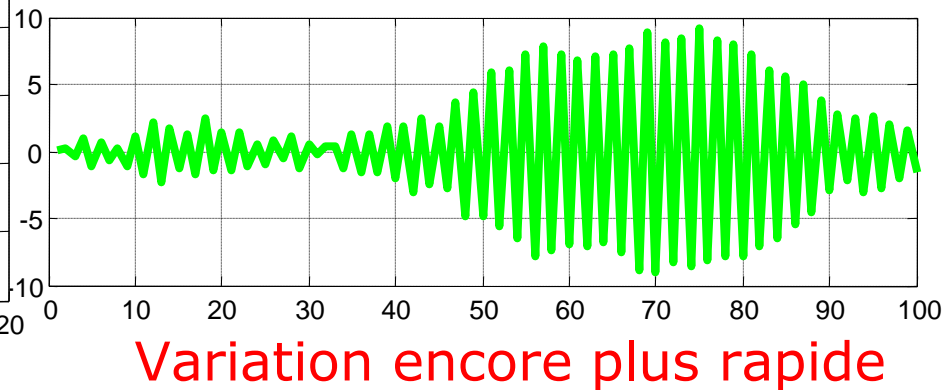
Fonction d'autocorrélation normalisée d'un processus AR(1): $\sigma_E=0.5$



Échantillons d'un processus AR(1) avec $q=-0.6$



Échantillons d'un processus AR(1) avec $q=-0.99$



Modèles de systèmes linéaires (9)

□ MA (0,M) (Processus à moyenne mobile)

- Equation aux différences:

$$x_n = \underbrace{\sum_{i=1}^M b_j \varepsilon_{n-j}}_{\text{Moyenne mobile}} + \varepsilon_n$$

- Processus à moyenne mobile d'ordre M .