Chapitre 1

Codage de source et codage de canal :
Notions fondamentales

1. Introduction

La théorie des communications s'intéresse aux moyens de transmettre une information depuis la source jusqu'à un utilisateur à travers un canal, conformément au schéma conventionnel de communication ci-dessous.

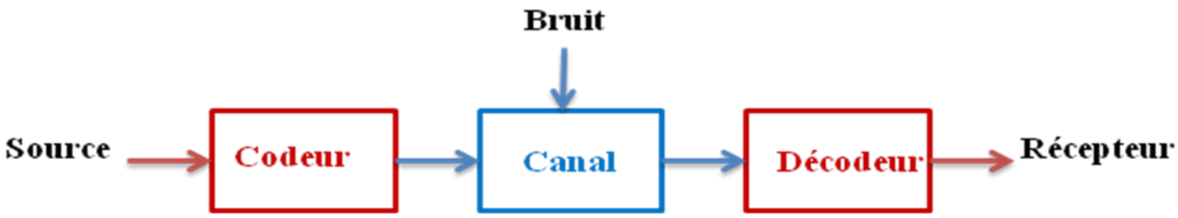

Fig.1 – Schéma de communication.

2. Définitions

2.1. Source

Tout dispositif capable de générer des messages. Il peut s'agir, par exemple, d'une voix, d'un signal électromagnétique, d'une image, d'une séquence de symboles binaires, etc.

2.2. Canal

Support de transmission de l'information. Peut être une ligne téléphonique, une liaison radio, un support optique, etc.

2.3. Bruit

C'est un perturbateur du canal. Par exemple, un décalage d'image, des rayures, etc.

2.4. Codeur

Représente l'ensemble des opérations effectuées sur la sortie de la source avant la transmission. Ces opérations peuvent être par exemple la modulation, la compression, le cryptage, l'ajout de redondance pour combattre les effets du bruit, ou encore l'adaptation à des contraintes de spectre. Elles ont pour but de rendre la sortie de la source compatible avec le canal.

2.5. Décodeur

Doit pouvoir restituer de façon acceptable l'information fournie par la source.

3. Intérêt du codage de source et du codage de canal

La séparation des modèles de sources et les modèles de canaux, par soucis de simplification, nous donne le schéma de communication représenté sur la figure 2.

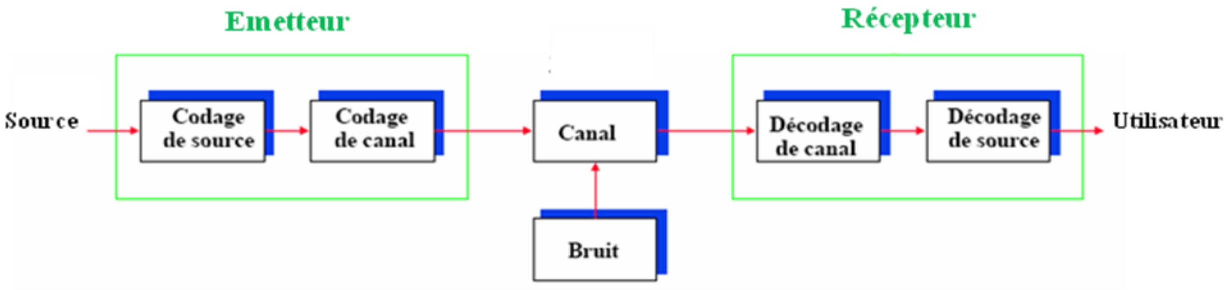

Fig.2 – Codeur de source et codeur de canal.

- Le codage de source ou compression de données sert à fournir une représentation binaire des données la plus économique possible, tout en préservant l'information essentielle qu'elles portent. Il est employé pour le stockage ou la transmission de ces données.

  On appelle données le résultat de la numérisation de signaux comme ceux de parole ou d'images ou plus généralement les données disponibles sur un fichier d'ordinateur.

- Le codage de canal consiste en une signature que l'on ajoute sur tout paquet d'information à transmettre. Cela permet de combattre à la fois le bruit de communication et l'interférence inter-symboles amenés par le canal. Le codage de canal sert principalement à détecter les erreurs et les corriger.

4. Sources et codage de source

4.1. Source discrète

C'est un dispositif qui fournit aléatoirement des séquences de symboles issus d'un ensemble discret fini. Une source d'information discrète peut être n'importe quel signal numérique. Une source est modélisée par un ensemble de variables aléatoires, à valeurs dans un alphabet de taille finie, $\Omega = \{x_1, ..., x_n\}$. Ω est appelé l'ensembles des symboles de source.

Exemple :

- $\Omega = \{12, 58, -69, 36, 47, -15, 127, -9\}$.
- Ω = {J e m a n g l p o}, pour le message: **Je mange la pomme**.

4.2. Source discrète sans mémoire

Une source est dite ***sans mémoire*** si la séquence de symboles générée par cette dernière est une suite de variables indépendantes. Chaque symbole de la source est choisi aléatoirement d'après une loi de probabilité p indépendante du temps. Pour chaque symbole x_i, $p(x_i)$ est la probabilité pour que ce symbole soit choisi telle que $0 \leq p(x_i) \leq 1 \quad et \quad \sum p(x_i) = 1$. La donnée de $p(x_1), \ldots, p(x_n)$ définit la probabilité discrète p sur Ω.

4.3. Code de source

Un ***code de source*** **C** pour une variable aléatoire **X** de distribution de probabilité *p*, est une application de Ω vers l'ensemble des chaînes de symboles d'un alphabet, ***D***-aire, ***A***.

L'ensembles des chaines non nulles de symboles d'un alphabet ***D***-aire ***A*** est noté A^+. En général, cet alphabet est binaire et on a $D = 2$, $A = \{0, 1\}$. A^+ est alors l'ensemble des chaines de caractères de taille finie formées de 0 et de 1. $A^+ = \{$**0, 1, 00, 01, 10, 11, 000, 001, ...**$\}$. un code associe à un symbole de source x_i un mot de code $\mathbf{C}(x_i)$. Ce mot de code est de longueur $l\,(x_i)$, la longueur étant son nombre de bits.

Exemple :

Soit une source qui fournit comme information l'un des quatre symboles discrets x_1, x_2, x_3 et x_4. Supposons que le codage de source transforme cette information discrète en symboles binaires. Nous donnons deux exemples de codage différents.

Codage 1	
x_1	**00**
x_2	**01**
x_3	**10**
x_4	**11**

Codage 2	
x_1	**0**
x_2	**10**
x_3	**110**
x_4	**111**

i. Dans le cas où les symboles sont équiprobables ($p(x_i) = 1/4$, i = 1, 2, 3, 4), la première méthode est meilleure:

- le codage 1 nécessite en moyenne 2 bits par chiffre;
- le codage 2 nécessite en moyenne $\frac{1}{4}\times 1+\frac{1}{4}\times 2+2\times\frac{1}{4}\times 3=2{,}25$ bits par chiffre.

ii. Si l'on a une source dont la distribution de probabilités est :

$$p(x_1) = 1/2 \quad p(x_2) = 1/4; \quad p(x_3) = p(x_4) = 1/8$$

- la longueur moyenne d'un symbole codé par la première méthode est toujours 2
- le codage 2 nécessite en moyenne $\frac{1}{2}\times 1+\frac{1}{4}\times 2+2\times\frac{1}{8}\times 3=1{,}75$ bits par chiffre, soit moins de 2 bits. On dit qu'il a réalisé une compression.

Pour coder correctement une source, il est donc important de connaître son comportement statistique.

5. canal et codage de canal

Pour modéliser un canal de transmission, il est nécessaire de spécifier l'ensemble des entrées et l'ensemble des sorties possibles. Le cas le plus simple est celui du ***canal discret sans mémoire***. Par exemple, l'entrée est un symbole pris dans un alphabet fini $E = \{e_1, \ldots, e_n\}$ et la sortie est un symbole pris dans un alphabet fini $S = \{s_1, \ldots, s_m\}$. Ces symboles sont émis en séquence, et, le canal est sans mémoire si chaque symbole de la séquence reçue ne dépend statistiquement que du symbole émis de même position. Ainsi un canal discret sans mémoire est entièrement décrit par la donnée des probabilités conditionnelles p(s|e) pour tous les symboles **e** de l'alphabet d'entrée et touts les symboles **s** de l'alphabet de sortie.

Exemple :

Le canal discret sans mémoire le plus connu est ***le canal binaire symétrique*** défini par E = S = {0, 1} et dont les probabilités de transition sont représentées sur la figure 3. La probabilité pour qu'un symbole soit inchangé est $1 - p$, et la probabilité pour qu'il soit changé est p : probabilité d'erreur.

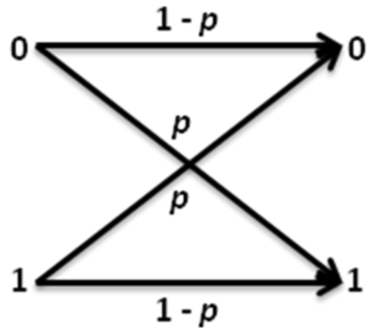

Fig.3 – Le canal binaire symétrique.

6. Codage conjoint source-canal

Historiquement la théorie de l'information, notamment les travaux de Shannon ont incité l'étude séparée des fonctions de quantification et de protection de l'information. C'est ainsi qu'ils ont préconisé le principe de séparabilité du codage de source (compression) et du codage de canal codes correcteurs d'erreurs de transmission).

Néanmoins, l'alternative proposée par le codage conjoint a prouvé, depuis plusieurs années, qu'il était possible avec des schémas de codage source canal combiné de réduire la complexité d'un système global de communication tout en perdant très peu en performances.