

2/191

# THESE

Présentée à

L'ECOLE NATIONALE POLYTECHNIQUE D'ALGER  
(DEPARTEMENT ELECTRONIQUE)

Pour obtenir

LE GRADE DE DOCTEUR ES-SCIENCES

Par

Daoud BERKANI

“Design” d'un Quantificateur Vectoriel  
à deux dimensions:

Le “SPIRAL QUANTIZER”

Soutenue devant la commission d'examen composée de:

M. A. CHEKIMA, Professeur à l'E.N.P.

Président

M. J-P. ADOUL, Professeur à l'Université de Sherbrooke, CANADA.

Rapporteur

M. M. BOULEMDEN, Professeur, Recteur de l'Université de BATNA.

Examineur

M. N. BELKHEMZA, Professeur à l'Université de BLIDA.

Examineur

M. A. FARAH, Maître de Conférences à l'E.N.P.

Examineur

## *Remerciements*

*Je suis très heureux de pouvoir remercier Monsieur A. Chekima , Professeur à l'Ecole Nationale Polytechnique, Responsable de la post-graduation d'Electronique, pour son aide soutenue, ses conseils avisés et pour avoir bien voulu accepter la présidence de ce Jury.*

*Je tiens particulièrement exprimer ma profonde gratitude à Monsieur J.P. Adoul, Professeur à l'Université de Sherbrooke, pour les conseils judicieux qu'il m'a prodigués, pour tous les efforts qu'il a consentis pour me guider et m'encourager pendant toute la durée de mes recherches. Je reste très sensible à l'accueil qu'il m'a réservé au sein de son laboratoire et toute l'attention qu'il m'a accordée pendant mes séjours au Canada, à Sherbrooke.*

*Mes sincères remerciements vont à Monsieur M. Boulemden, Professeur, Recteur de l'Université de Batna pour avoir bien voulu accepter de participer au jury d'examen.*

*Mes sincères remerciements vont également à Monsieur N. Belkhemza, Professeur à l'Université de Blida pour l'intérêt qu'il a bien voulu porter à ce travail.*

*Mes vifs remerciements vont à Monsieur A. Farah, Maître de Conférences à l'Ecole Nationale Polytechnique pour ses encouragements et ses remarques.*

*J'aimerais exprimer toute ma gratitude à Messieurs L. Abida, Professeur, Directeur de l'Institut d'Electronique de l'Université de Constantine et B. Derras, PhD à l'Ecole Nationale Polytechnique pour leurs encouragements.*

*Je saisi l'occasion qui m'est offerte pour exprimer ma profonde reconnaissance à toutes les personnes qui m'ont soutenu pendant la durée de mes travaux.*

*J'aimerais citer en particuliers toute l'équipe du Centre de Recherche sur les Communications et celle du Département Génie Electrique de l'Université de Sherbrooke qui m'a beaucoup faciliter ma tache et pour son amabilité.*

*Ce travail de recherche a été fait dans le cadre d'un projet de coopération ENP(Algérie) / ACDI(Canada). Aussi, je tiens à souligner les conditions de travail agréables qui m'ont été offertes par les structures tant algériennes que canadiennes et à exprimer ainsi mes remerciements les plus sincères aux responsables de ce projet.*

# Table des matières

INTRODUCTION.....	1
CHAPITRE 1: NOTIONS FONDAMENTALES DE LA THEORIE DE LA DISTORSION....	4
1.1- Entropie d'une Source.....	4
1.2- La fonction de distorsion.....	6
1.3- Cas d'une source bidimensionnelle.....	12
CHAPITRE 2 : LA QUANTIFICATION .....	14
Introduction .....	14
2.1 principe.....	14
2.2 La quantification vectorielle.....	17
2.2.1 Performances asymptotiques des quantificateurs vectoriels.....	20
a- Bornes asymptotiques des performances .....	21
b- Application à une source gaussienne sans mémoire.....	24
2.2.2 - Codage, modulation et quantification .....	26
2.3 La quantification par treillis (TCQ).....	31
2.3.1 - Description du TCQ.....	32
2.3.2 - Interprétation .....	37
2.4 Les quantificateurs polaires .....	39
Conclusion .....	43
CHAPITRE 3 : LE "SPIRAL QUANTIZER".....	44
Introduction .....	44
3.1- Description du "Spiral Quantizer".....	45
3.2- La source d'excitation .....	50
3.3- Optimisation des paramètres du quantificateur .....	54
3.4- Etude de la variation du RSB en fonction du nombre de vecteurs.....	59
3.5- Performances du quantificateur en spirale.....	61
3.6- Résultats obtenus avec Le "set-partitioning".....	67

3.6.1- Présentation des constellations types.....	67
3.6.2- Performances obtenues en "Set Partitioning" .....	76
Conclusion .....	78
<b>CHAPITRE 4 : INTERPRETATION DES RESULTATS .....</b>	<b>80</b>
Introduction .....	80
4.1 - Quantification par réseaux réguliers.....	80
4.2 - Etude comparative du "Spiral Quantizer" .....	83
4.3 - Commentaires sur les résultats obtenus .....	88
4.3.1 - La discrétisation de la spirale.....	88
4.3.2 - Stabilité du "Spiral Quantizer".....	89
4.3.3 - Les performances en terme de RSB du "Spiral Quantizer".....	89
4.3.4 - Le "Set Partitioning" .....	92
Conclusion .....	93
<b>CONCLUSION.....</b>	<b>94</b>
<b>BIBLIOGRAPHIE.....</b>	<b>96</b>

# Introduction

Les systèmes numériques de communication ne sont plus de nos jours une nouveauté, ils font déjà partie de notre environnement de tous les jours. Ces systèmes de traitement de l'information exigent une représentation numérique des formes d'ondes.

L'opération de quantification, réalisée par un quantificateur est une des procédures nécessaires à cette représentation.

Ce travail apporte une contribution originale à ce domaine en proposant l'emploi d'un nouveau quantificateur qui met en œuvre un ensemble indexé de vecteurs générés par une discrétisation de la spirale d'Archimède. C'est la première fois que la spirale est exploitée à cette fin, en codage de canal ou en codage de source. Nous désignerons par le vocable de "Spiral Quantizer" ce quantificateur.

La théorie de l'information nous apprend que toute transmission de l'information reste discrète en nature. Cependant les sources réelles des signaux telles que la parole et l'image, génèrent des fonctions continues. Ainsi, dans une première étape le traitement de l'information consiste à convertir ces fonctions ou signaux analogiques émis par les sources en signaux numériques.

La quantification réalise la discrétisation des valeurs échantillonnées ou prélevées du signal analogique, elle convertit ainsi des variables réelles considérées comme un alphabet infini en variables discrètes ou bien un alphabet fini.

Les chercheurs se sont attelés à répondre à la question fondamentale qui est :

*Quelle est la meilleure façon de représenter et de transmettre une variable aléatoire continue en utilisant le moins de bits possibles?*

Le caractère aléatoire de la variable découle de la nature même de l'information; en effet, il n'y a "information" que s'il y a un fait nouveau ou inconnu ou encore un événement.

Par ailleurs, les recherches ont été beaucoup plus axées sur les phénomènes gaussiens, ce qui est en partie une conséquence du théorème de la limite centrale sur les répartition des processus appartenant au domaine de l'aléas.

Le problème de "design" du meilleur quantificateur a canalisé les recherches sur la quantification scalaire puis vectorielle. Dans les deux cas de quantification les résultats théoriques ont montré que les signaux à distribution uniforme sont les mieux représentés, soit avec un minimum possible de distorsion entre l'information originale et le signal codé.

La théorie de la distorsion a même montré que la distorsion peut être réduite par un codage de blocs toujours plus longs.

De nombreux procédés et algorithmes de design de quantificateurs ont été mis au point. Ainsi pour une quantification dite statistique, des dictionnaires ont été mis en œuvre; ceux-ci occupent souvent par nécessité, des espaces mémoires importants. Leur complexité et leur lourdeur dans la recherche du meilleur représentant d'une valeur d'entrée considérée demeure leur principal inconvénient.

L'approche algébrique dispose de dictionnaires très bien structurés et bénéficie d'algorithmes de recherche très efficaces et rapides. Ces dictionnaires sont formés à partir de réseaux réguliers de points et de codes correcteurs d'erreurs.

Les réseaux réguliers ont été largement exploités en modulation ou codage de canal grâce aux travaux de Ungerboeck. Forney a formalisé la construction des réseaux réguliers par le concept de décomposition en "cosets" et dont le prolongement a abouti à la représentation des réseaux réguliers par des treillis fortement structurés.

La quantification par treillis (TCQ) a été introduite par Marcellin et Fisher qui ont mis en évidence encore la dualité entre la modulation et la quantification.

Cependant, dans le domaine du codage de source, les réseaux réguliers s'adaptent mal aux sources réelles qui sont loin d'avoir des distributions uniformes.

Diverses solutions et méthodes ont été avancées notamment pour un codage à bas débit dont le LPQ (Lattice Phase Quantization) proposé par Adoul.

Le passage d'un système cartésien de coordonnées à un système de coordonnées polaires a la vertu de présenter une distribution uniforme de la phase de la variable à quantifier. Cet aspect a été exploité par plusieurs chercheurs.

En tenant compte des diverses réflexions qui ont été émises, nous nous sommes attachés à répondre à la question de design posée plus haut en présentant le "Spiral Quantizer" : un quantificateur possédant une structure régulière capable d'approcher les

performances optimales en terme d'erreur quadratique moyenne pour une source gaussienne à moyenne nulle. Ce quantificateur présente d'autres avantages que nous développerons dans cette thèse.

Le premier chapitre présente les notions fondamentales de la théorie de la distorsion.

Le second chapitre est consacré à la quantification. La rétrospective des travaux de recherche présentée fait ressortir les solutions préconisées par les chercheurs et situe le problème de design des quantificateurs. Elle situe également le niveau actuel de la recherche dans le domaine .

La troisième partie de ce travail décrit le "Spiral Quantizer" et la procédure de discrétisation de la spirale puis présente les performances et résultats obtenus.

Les commentaires sur les propriétés de notre quantificateur et sur les résultats et performances obtenus sont fait dans la dernière partie de ce travail.

# Chapitre 1

## Notions fondamentales sur la théorie de la distorsion

La théorie de la distorsion est une discipline de la théorie de l'information qui étudie le codage de canal et de source de séquences stochastiques. Parmi les différents types de sources, seules seront considérées les séquences gaussiennes sans mémoire et à moyenne nulle.

La théorie de la distorsion occupe une place importante dans la théorie de l'information. C'est l'étude théorique sur la meilleure façon de représenter la séquence aléatoire de la source qui a pris la dénomination de la "théorie de la distorsion" ou de "rate distortion theory" [1]. On peut dire que c'est Shannon [2], qui en 1948, a lancé les principes de base de cette théorie.

Le but essentiel de cette partie est de montrer que la fonction de la distorsion est l'expression mathématique de la meilleure performance que peut atteindre un quantificateur.

Cette partie introduit également les différents développements des idées et questions afférentes à la quantification et que nous exposerons tout le long de notre travail.

### 1.1- Entropie d'une Source

considérons une source stationnaire, ergodique, sans mémoire  $X$ . Chaque échantillon temporel  $x(i)$  est indépendant des échantillons  $x(i-1), x(i-2), \dots, x(N-1)$  et qu'il ne peut prendre que  $N-1$  valeurs différentes dans l'ensemble  $\{a_0, a_1, \dots, a_{N-1}\}$ .

Chaque échantillon est associé à une probabilité  $p(x(i) = a(i)) = P_x(a_i)$ . On définit ainsi les probabilités :  $P(a_0), P(a_1), \dots, P(a_{N-1})$ .

L'entropie de  $X$  est définie comme suit :

$$H(X) = -E [\log_2 P(X)] = - \sum_{i=0}^{N-1} P_x(a_i) \log_2 P_x(a_i) \quad (1.1.1)$$

la grandeur  $I(X) = -\log_2 P(X)$  est appelée l'auto-information ou la surprise, plus un événement est rare ( $P(X)$  est petit), plus la surprise qui lui est associée est grande, chaque lettre de la source a donc une surprise qui lui est propre. La surprise moyenne de la source n'est autre que l'entropie de la source :

$$H(X) = E [I(X)] \quad (1.1.2)$$

$H(x)$  ainsi définie s'exprime en bits, elle mesure l'information moyenne d'une source. Dans notre cas on peut écrire :

$$0 < H(X) < \log_2 N \quad (1.1.3)$$

De cette double inégalité, on peut souligner les remarques suivantes :

- Lorsque  $H(x)$  est nulle, cela se traduit par une seule probabilité non nulle d'un élément de l'alphabet. Dans ce cas la source est totalement prédictible.
- On obtient une probabilité uniforme  $P_x = \frac{1}{N}$  dans le cas où  $H(X) = \log_2 N$ , la source serait alors totalement non prédictible.
- L'entropie d'une source discrète est toujours positive.

Si nous considérons deux variables aléatoires  $x$  et  $y$  associées aux ensembles de valeurs  $\{a_0, a_1, \dots, a_{N-1}\}$  et  $\{b_0, b_1, \dots, b_{M-1}\}$ , avec une distribution de probabilité conjointe donnée, on peut définir l'entropie conditionnelle déduite de cette notion de densité de probabilité conditionnelle, on obtient ce qui suit:

$$H(X|Y) = -E [\log_2 P_{XY}(X|Y)] = - \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} P_{XY}(a_n, b_m) \log_2 P_{XY}(a_n | b_m) \quad (1.1.4)$$

Cette relation exprime l'incertitude moyenne qui reste à propos de  $X$  après la connaissance de la valeur de  $Y$ , ou de fait, l'information d'une source connaissant une autre source.

L'information mutuelle s'écrit :

$$I(X;Y) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} P_{XY}(a_n, b_m) \log_2 \frac{P_{XY}(a_n, b_m)}{P_X(a_n)P_Y(b_n)} \quad (1.1.5)$$

ou bien  $I(X;Y) = H(X) - H(X|Y) \geq 0 \quad (1.1.6)$

Ainsi, l'information mutuelle peut être considérée comme la quantité d'incertitudes soustraites de  $X$  ou la quantité d'information apprises sur  $X$  en observant les valeurs de  $Y$ .

La notion d'entropie d'une source discrète telle que définie par les relations précédentes peut être étendue à des variables vectorielles en considérant les probabilités d'apparition des lettres de l'alphabet comme des probabilités conjointes des composantes des vecteurs.

### 1.2- La fonction de distorsion.

Considérons le schéma simplifié d'un quantificateur suivant (figure 1.1) :



Figure 1.1: Quantificateur type,  $x$  est la grandeur d'entrée à quantifier et  $y_i$  est la valeur de sortie.

En plus de la source déjà considérée plus haut, soient :

- une séquence de réalisations indépendantes de X composée de k échantillons,

$$X = ( x(0), x(1) , \dots , x(k-1) ) ,$$

- une autre séquence ayant ses réalisations dans l'ensemble  $\{ b_0, b_1, \dots, b_{M-1} \}$  ,

$$Y = ( y(0), y(1) , \dots, y(k-1) ) ,$$

-  $\rho_k(x, y)$  la mesure de la distorsion , le critère de fidélité ou encore le coût dû à la reproduction de la séquence X par le vecteur Y.

Notre quantificateur Q peut être considéré comme un canal de transmission. Il va associer à une lettre de l'alphabet de X une lettre de Y. Il constitue donc une matrice de taille  $M \times N$  avec en pratique  $M \leq N$  .

C'est pour cette raison que l'on représente le modèle du quantificateur comme un modèle de probabilité conditionnelle de Y pour X donné avec la contrainte du critère de fidélité  $\rho$ .

On pose l'hypothèse que l'expression de ce critère peut se mettre sous la forme :

$$\rho_k(x, y) = \frac{1}{k} \sum_{i=0}^{k-1} \rho_k(x(i) , y(i)) \quad (1.2.1)$$

où  $\rho$  est une mesure de la distorsion non négative entre les composantes correspondantes des séquences X et Y que l'on appelle souvent distorsion à "lettre unique" ou "sans mémoire" [1,3,4,5].

Les fonctions de probabilité et de probabilité conjointe de transition  $P_X$  et  $P_{Y|X}$  supposées connues, on peut obtenir la distorsion moyenne avec l'expression :

$$d(P_{Y|X}) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} P_X(a_n) P_{Y|X}(b_m | a_n) \rho(a_n, b_m) \quad (1.2.2)$$

L'expression de l'information mutuelle s'écrit:

$$I(P_{Y|X}) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} P_X(a_n) P_{Y|X}(b_m|a_n) \log_2 \frac{P_{Y|X}(b_m|a_n)}{P_Y(b_m)} \quad (1.2.3)$$

$$\text{où } P_Y(b_m) = \sum_{n=0}^{N-1} P_X(a_n) P_{Y|X}(b_m|a_n) \quad (1.2.4)$$

Pour une distorsion donnée  $D$ , on définit la fonction de taux de distorsion  $R(D)$ , pour un critère de fidélité  $\rho$  par la relation :

$$R(D) = \min_{P_{Y|X} \in \mathcal{P}_D} I(P_{Y|X}) \quad (1.2.5)$$

$$\text{avec : } \mathcal{P}_D = \{P_{Y|X} : d(P_{Y|X}) \leq D\} \quad (1.2.6)$$

Cette équation définit une fonction  $R(D)$  qui donne pour une distorsion donnée, un débit minimal  $R$  ou réciproquement pour un débit donné, la plus faible distorsion possible  $D$ .

Ce qui nous ramène à parler du codage d'une source ou d'un vecteur :

On appelle un code  $C$  de taille  $N$  et de longueur de bloc  $k$  un ensemble de  $N$  séquences  $\{y_0, y_1, \dots, y_{N-1}\}$  de dimension  $k$  chacune.

Chaque séquence de  $C$  est appelée mot de code qui est une séquence binaire de longueur  $\log_2 N$  au moins. Pour coder un vecteur  $X$  dans  $C$ , il faudra choisir le mot de code qui minimise la distorsion  $\rho(x, y)$ , notée:

$$\rho_k(x|C) = \min_{y \in C} \rho_k(x, y) \quad (1.2.7)$$

On définit ainsi la distorsion moyenne:

$$\rho(C) = E[\rho_k(x|C)] = \sum_x P(x) \rho_k(x|C) \quad (1.2.8)$$

$$\text{Avec : } P(x) = \prod_{i=0}^{k-1} P_x(x_i) \quad (1.2.9)$$

Pour un vecteur  $X$  de  $k$  échantillons, on définit le débit par échantillon du code  $C$  par la relation:

$$R = \frac{\log_2 N}{k} \text{ bits} \quad (1.2.10)$$

La théorie sur le codage nous apprend que  $R(D)$  est une limite théorique des performances des systèmes de codage, et qu'il est possible d'atteindre une performance assez proche de cette limite.

ce qui est dicté encore par l'équation ci-dessous :

$$\lim_{k \rightarrow \infty} \frac{1}{k} \log_2 N = R(D) \quad (1.2.11)$$

Il est à souligner qu'il suffit de considérer, à la place des distributions  $P_x$  les fonctions de densité de probabilité avec quelques arrangements mathématiques des relations ci-dessus pour étendre leur validité au cas des sources continues.

Cependant il nous semble utile de noter certaines remarques :

- L'alphabet de reproduction des sources continues est infini.
- L'entropie différentielle des sources continues peut être positive ou négative; elle est une mesure relative au repère alors que dans le cas discret elle est absolue. Ainsi on définit l'entropie différentielle par la relation :

$$h(X) = - E [ \log_2 f_X(X) ] = - \int f_X(x) \log_2 f_X(x) dx \quad (1.2.12)$$

ou  $f_X(X)$  est la densité de probabilité.

L'entropie conditionnelle est définie comme suit :

$$h(X|Y) = - \iint f_{XY}(x,y) \log_2 f_{XY}(x|y) dx dy \quad (1.2.13)$$

et l'information mutuelle par :

$$I(X; Y) = \iint f_{XY}(x, y) \log_2 \frac{f_{XY}(x, y)}{f_X(x)f_Y(y)} dx dy \quad (1.2.14)$$

A partir des relations précédantes, si l'on considère que  $\rho(x, y)$  est une mesure de distorsion, et  $f_{Y|X}(y|x)$  la densité de probabilité conditionnelle, on définit la distorsion moyenne suivante :

$$d(f_{Y|X}) = \iint f_X(x) f_{Y|X}(y|x) \rho(x, y) dx dy \quad (1.2.15)$$

et l'information mutuelle moyenne s'écrit:

$$I(f_{Y|X}) = \iint f_X(x) f_{Y|X}(y|x) \log_2 \frac{f_{Y|X}(y|x)}{f_Y(y)} dx dy \quad (1.2.16)$$

$$\text{où } f_Y(y) = \int f_X(x) f_{Y|X}(y|x) dx \quad (1.2.17)$$

On obtient finalement comme pour le cas discret, la limite théorique des performances ou la fonction  $R(D)$  :

$$R(D) = \min_{f_{Y|X} \in f_D} I(f_{Y|X}) \quad (1.2.18)$$

$$\text{où } f_D = \{f_{Y|X} : d(f_{Y|X}) \leq D\} \quad (1.2.19)$$

Les particularités de chaque type de cas (discret ou continu) créent une légère différence à l'origine de la fonction  $R(D)$  entre les représentations graphiques de cette fonction pour les cas des sources continues et discrètes (figure 1.2).

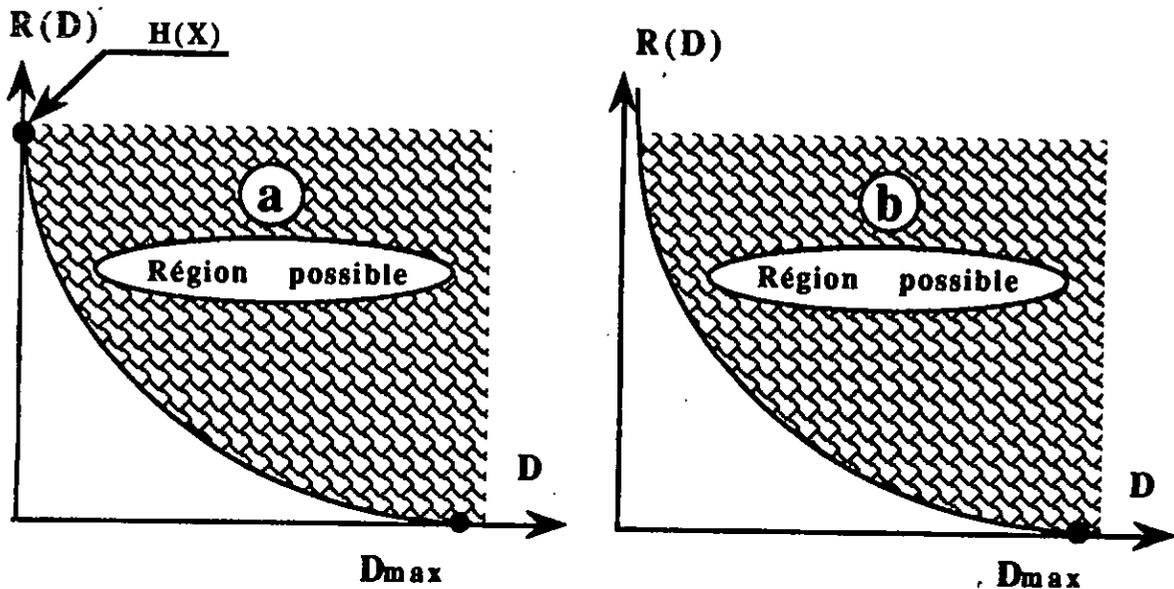


Figure 1.2: Comparaison de la fonction  $R(D)$  des sources discrète (a) et continue (b).

On remarquera certaines propriétés de la fonction  $R(D)$  [1,3,4] :

- $R(D)$  est une fonction convexe sur l'intervalle  $]0, D_{\max}[$ .
- $D_{\max}$  est la distorsion maximale, elle existe toujours.
- La valeur extrême  $D_{\max}$  est obtenue comme suit :

pour le cas discret,

$$D_{\max} = \min_{y \in C} \sum_x P(x) \rho_k(x, y) \quad (1.2.20)$$

et le cas de source continue,

$$D_{\max} = \min_y \int f_k(x) \rho(x, y) dx \quad (1.2.21)$$

- $R(D)$  est continue, monotone, décroissante.

La méthode utilisée pour la mesure de la distorsion ou le choix du critère de fidélité reste primordiale dans la conception des quantificateurs. Celle de l'erreur quadratique moyenne demeure incontestablement la plus utilisée malgré quelques cas où elle s'avère inadéquate. Appelée encore MSE ou "mean squared error", elle permet d'exprimer pour une variable aléatoire gaussienne de variance  $\sigma_X^2$ , la fonction de distorsion  $R(D)$  [4] :

$$R(D) = \frac{1}{2} \max \left( 0, \log_2 \frac{\sigma_X^2}{D} \right) = \left\{ \begin{array}{ll} \frac{1}{2} \log_2 \frac{\sigma_X^2}{D} & 0 \leq D \leq \sigma_X^2 \\ 0 & D \geq \sigma_X^2 \end{array} \right\} \quad (1.2.22)$$

La limite inférieure de la fonction  $R(D)$  est appelée la borne inférieure de Shannon (Shannon Lower Bound ou SLB). Cette caractéristique (SLB) peut être obtenue d'une façon explicite dans bien des cas ce qui ne se réalise pas pour la fonction  $R(D)$  [1, 5].

La borne inférieure de Shannon s'identifie à cette dernière pour des sources gaussiennes telles que définies au début de ce chapitre lorsque le MSE est utilisé comme estimateur.

Il serait intéressant de se pencher sur le cas d'une source gaussienne complexe bidimensionnelle.

### 1.3- Cas d'une source bidimensionnelle.

Dans la perspective d'une indépendance statistique entre les parties réelle et imaginaire, l'extension des résultats de la théorie de la distorsion exposée devient simple.

Il suffit de considérer que la source gaussienne complexe est constituée en fait de deux source réelles gaussiennes indépendantes.

La distorsion totale obtenue par suite à une opération de quantification de la source complexe n'est autre que la moyenne des distorsions de chaque source:

$$D(R) = \frac{1}{2} \{D_1(R_1) + D_2(R_2)\} \quad (1.2.23)$$

où  $R_1$  et  $R_2$  sont les débits des sources.

et  $D_1(R_1)$  et  $D_2(R_2)$  sont les distorsions.

On pourrait alors se poser une autre question comment ou bien quelle est la meilleure façon de réaliser les meilleures performances possibles ?

Il est évident qu'au moins deux solutions existent :

- 1- soit coder les parties réelles et imaginaires chacune à part.
- 2- soit coder les amplitudes et les phases.

Les deux approches ont été l'objet d'un débat des chercheurs et constituent ce que l'on appelle respectivement "quantification scalaire" et "quantification vectorielle". Cette question est développée dans le chapitre suivant.

# Chapitre 2

## La Quantification

### Introduction.

Ce chapitre a pour objet de présenter les fondements de la quantification et de faire une rétrospective succincte des travaux de recherche déjà publiés.

On verra successivement l'essence de la quantification vectorielle, la quantification par treillis, la technique dite de " Set Partitioning" et nous terminerons par la quantification sous optimale.

Les quantificateurs polaires très proches du quantificateur à base de la spirale d'Archimède que nous proposons seront présentés dans cette dernière partie.

### 2.1- Principe.

La quantification est une opération qui consiste à arrondir une grandeur d'entrée en lui associant une valeur choisie parmi un nombre fini de valeurs possibles. Cette opération est indispensable dans la numérisation des signaux. Dans le cas des signaux unidimensionnels le processus est bien connu [4,6]. La fonction de transfert est une caractéristique en marche d'escalier comme le montre la figure 2.1 ; un tel quantificateur est qualifié de scalaire.

Le quantificateur établit une règle de correspondance entre la variable d'entrée  $X$  une valeur de sortie approchée ou arrondie  $Y$  choisie dans un ensemble fini de  $N$  valeurs prédéterminées.

La formulation du problème peut être faite en considérant un ensemble de  $N$  valeurs arrondies :

$$B = \{ Y_i \mid i = 0, 1, \dots, N-1 \} \quad (2.1)$$

La fonction de quantification s'écrirait alors :

$$Y_i = Q(x) \quad (2.2)$$

Un quantificateur fournit une seule information à la sortie mais celle-ci est disponible sous deux formes : la valeur arrondie  $Y$  et l'indice  $i$ .

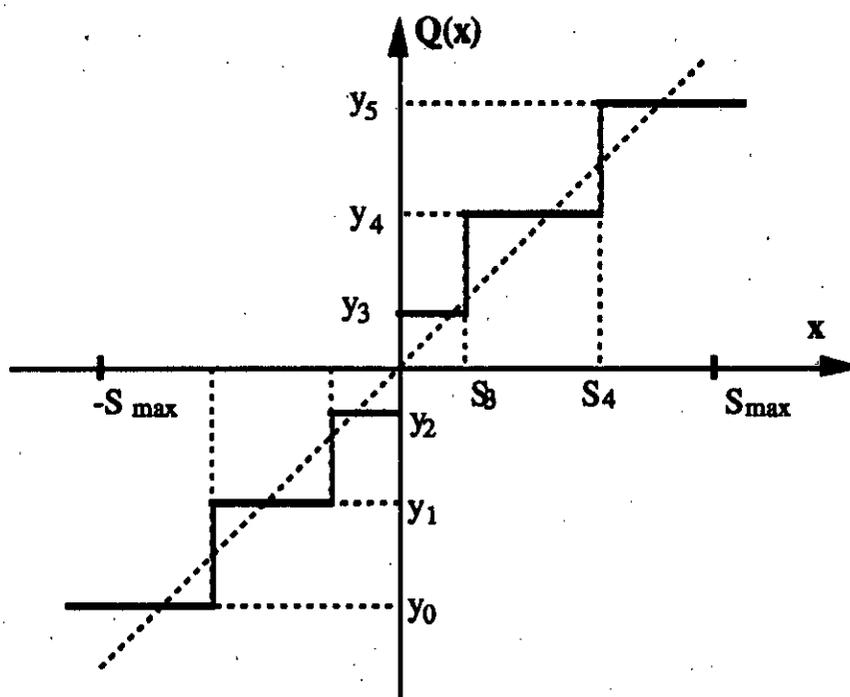


Figure 2.1: Caractéristique type d'un quantificateur scalaire à 6 niveaux de quantification ( $N=6$ ).

On peut admettre alors des variantes de conception des quantificateurs selon l'obtention ou la présentation de ces paramètres à la sortie [ 7]. Nous présentons ici ces variantes de quantificateurs (figure 2.2).

Dans le premier cas (figure 2.2-1), le quantificateur fournit , sous une forme intégrée, une même information, dans le second cas (figure 2.2-2) celle-ci est présentée sous deux formes distinctes. Le troisième schéma décrit par la figure 2.2-3 illustre le point de vue adopté en DPCM ("Differential Pulse Code Modulation") où la valeur arrondie  $Y$  est

réinjectée dans la boucle de rétroaction alors que l'indice  $i$  est directement transmis, enfin la quatrième alternative fournit tout d'abord l'indice  $i$ , tandis que la valeur arrondie s'obtient par une table de décodage.

Les chercheurs préfèrent inclure dans la quantification ainsi présentée le codage qui associe un indice à chacune des valeurs arrondies ; de là ils contournent le concept de seuil et font correspondre à l'indice  $i$  une table de décodage où sont contenues les valeurs arrondies de sortie  $Y_i$  ; ils arrivent ainsi à généraliser la notion de quantification à plus d'une variable et réduisent celle-ci à une recherche de l'indice entre 0 et  $N-1$  pour lequel la meilleure approximation est obtenue.

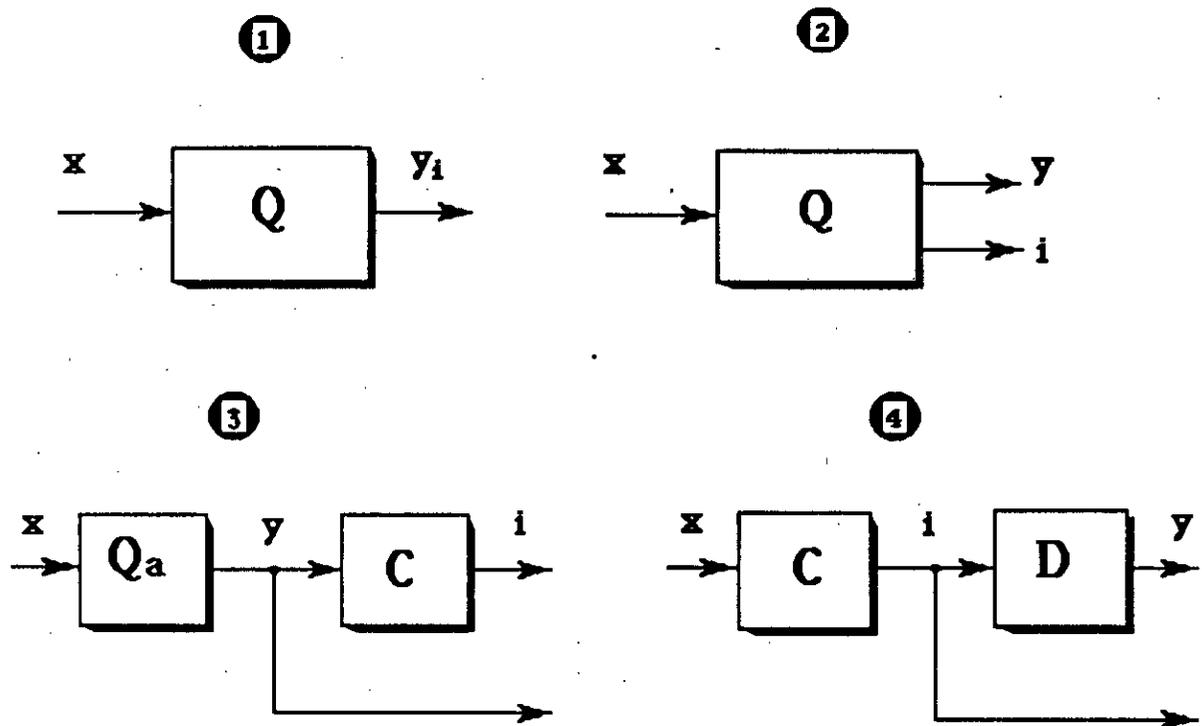


Figure 2.2: Schéma type de quantificateurs.

Cette manière de faire permet de considérer toutes les grandeurs, en particuliers l'entrée  $x$  et les valeurs arrondies  $y$ , sous forme vectorielle. La figure 2.3 illustre un exemple de quantification d'un vecteur  $x$  à deux dimensions.

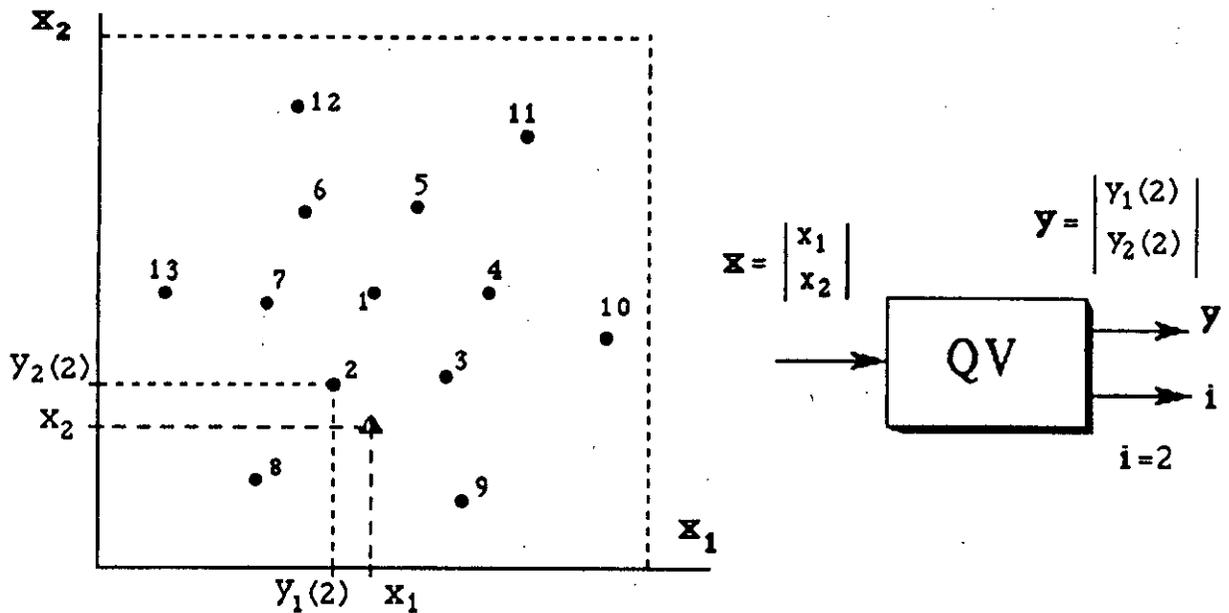


Figure 2.3 : Principe de la quantification vectorielle. L'entrée  $x$  est représentée dans l'espace par un triangle. Le vecteur arrondi  $y$  le plus proche parmi les treize possibles est le vecteur d'indice  $i=2$ .

C'est le principe de la quantification vectorielle [7,8] que nous allons développer ci-dessous .

En général, les quantificateurs vectoriels surpassent en performances les quantificateurs scalaires ; des études ont été faites dans ce sens : J.P. Adoul a comparé les quantificateurs scalaires et vectoriels et montré la supériorité de ces derniers sauf "dans quelques rares cas pathologiques où les performances sont égales " [ 7 ].

## 2.2 - Quantification Vectorielle

Un quantificateur vectoriel ( QV ) fait correspondre à tout vecteur  $x$  décrit comme suit :  $x = (x(0),x(1),...,x(k-1))$  un vecteur  $y_i = (y_i(0),y_i(1),...,y_i(k-1))$  choisi

parmi un ensemble fini  $B$  de  $N$  vecteurs de reproduction [7]. Dans la littérature, on attribue à cet ensemble  $B$  l'appellation de dictionnaire.

En pratique, dans plusieurs systèmes de communication, seul l'indice  $i$  de la séquence choisie  $y_i$  est transmis au décodeur qui choisit alors la séquence correspondante à cette valeur de l'indice. Le dictionnaire qui influe considérablement sur les performances techniques d'un QV est caractérisé par sa taille  $N$  et sa dimension  $k$ ; ces paramètres permettent de déterminer le débit binaire  $R = \log_2 N/k$  par échantillon comme cela a été souligné plus haut (chapitre 1).

Cette formulation du débit permet de faire des études comparatives de quantificateurs opérant sur des vecteurs de dimensions différentes.

Par ailleurs, il apparaît évident qu'une distorsion est introduite à chaque approximation d'un vecteur d'entrée  $x$  par un vecteur arrondi  $y$ .

La figure 2.4 schématise un exemple d'organisation structurale d'un quantificateur vectoriel souvent utilisé dans les systèmes de communication numérique.

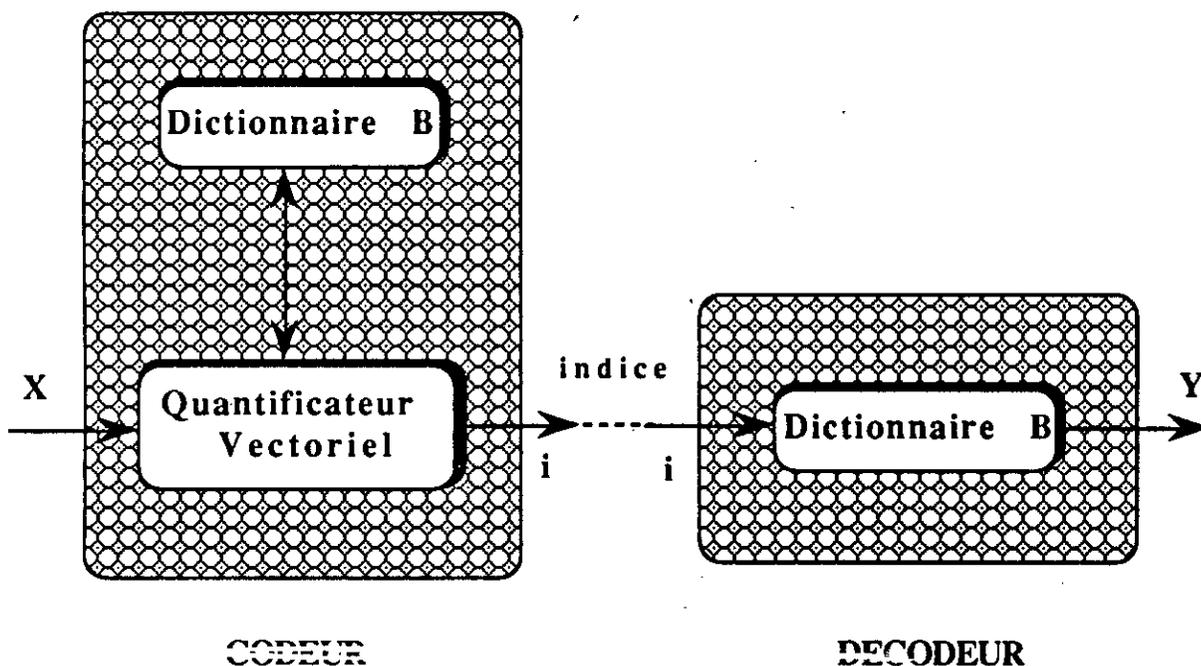


Figure 2.4 : Schéma d'un quantificateur vectoriel,  $x$  est la séquence à coder,  $i$  l'indice à transmettre et  $y$  est la séquence reproduite.

Ainsi, un quantificateur est complètement défini par un ensemble  $B$  de vecteurs arrondis et une distance définissant le degré de distorsion qui s'introduit lors d'une approximation d'un vecteur  $x$  par un vecteur approché  $y$ . Cette distance est une fonction réelle, non négative et doit satisfaire au principe suivant : la distance d'un point à lui-même est toujours plus petite que la distance de ce point à tout autre point. La forme la plus utilisée est la suivante [9] :

$$d(x,y) = \|x - y\|^r \quad (2.3)$$

Cette distance devient le carré de la distance euclidienne lorsque  $r = 2$ . Elle exprimerait alors la puissance de l'erreur de quantification. Le principal inconvénient de la quantification vectorielle réside dans sa complexité qui croît de façon exponentielle avec la dimension et la taille du dictionnaire. La recherche exhaustive à travers les  $N$  vecteurs arrondis ou mots de code du dictionnaire nécessite une comparaison de la séquence d'entrée avec tous les  $N$  mots de code, ce qui requiert le calcul de  $N$  distances sachant que :  $k$  multiplications et  $2k-1$  additions sont nécessaires pour estimer une seule distance euclidienne.

Cependant, les performances en terme de rapport signal à bruit obtenues par les QV s'améliorent sensiblement lorsque la dimension  $k$  augmente. De plus, la théorie de la distorsion garantit, pour  $k$  suffisamment grand, l'existence d'un ensemble fini de  $N$  vecteurs arrondis pour un débit  $R$  comme défini par la relation (1.2.10), tel que la distorsion moyenne par échantillon est proche de la distorsion minimale idéale  $D(R)$ . C'est pour des raisons de complexité de conception que l'on fait recours aux méthodes sous-optimales de quantification, sujet que nous présenterons plus loin.

Concernant les méthodes (ou quantificateurs) optimales et sous-optimales : on considère un quantificateur, tel que déjà défini, optimal si de tous les quantificateurs, il est celui qui minimise l'espérance mathématique de la distance :

$$E [d(x,y)] = E [d(x,Q(x))] = \int d(x,Q(x)) P(x) dx \quad (2.4)$$

où  $P(x)$  est la distribution du vecteur d'entrée .

Approcher les performances d'un tel quantificateur optimal n'est pas chose aisée ; aussi, pour des cas particuliers en quantification scalaire, Lloyd [10] a proposé un algorithme itératif appelé algorithme de Lloyd - Max pour construire un quantificateur

optimal et Max a tabulé les valeurs arrondies de la distribution gaussienne [11]. Paez et Glisson ont étendu ces résultats aux densités de Laplace et Gamma [12] et Pearlman à la distribution de Rayleigh [13].

Nous donnons ci-dessous quelques aspects sur les limites des performances que peut atteindre un quantificateur sur la base des résultats théoriques obtenus par Gray et Al.[15,16] ainsi que d'autres chercheurs [14].

### 2.2.1- Performances asymptotiques des QV.

Etudions les limites des performances des quantificateurs vectoriels en fonction de la dimension  $k$  des blocs ou vecteurs [14,15]. Considérons une variable aléatoire dans l'espace euclidien de dimension  $k$  ( $\mathcal{R}^k$ ) avec  $P(x)$  comme densité de probabilité :

$$x = (x(0), x(1), \dots, x(k-1)) \in \mathcal{R}^k$$

Dans une première étape, nous allons positionner le problème :

On définit un quantificateur vectoriel QV par l'ensemble de  $N$  vecteurs arrondis  $y_0, y_1, \dots, y_{k-1}$  et par la partition  $\{S_0, S_1, \dots, S_{N-1}\}$  de  $\mathcal{R}^k$  (les  $S_i$  sont disjoints et couvrent tout l'espace  $\mathcal{R}^k$ ) tel que :

$$Q(x) = y_i \quad \text{si } x \in S_i \quad (2.5)$$

soit la région  $G \subset \mathcal{R}^k$ . On définit le volume  $V(G)$  de  $G$  par :

$$V(G) = \int_G dx \quad (2.6)$$

soit  $V_k$  le volume de la sphère unité de dimension  $k$ :

$$V_k = \text{Vol}(\{u : \|u\| \leq 1\}) = \frac{2\Gamma(1/2)^k}{k\Gamma(k/2)} \quad (2.7)$$

$$\text{où } \|u\| = \left( \sum_{j=0}^{k-1} u^2(j) \right)^{1/2} \quad (2.8)$$

et  $\Gamma$  est la fonction Gamma ( $\Gamma(1/2) = \sqrt{\pi}$ )

La performance d'un QV est mesurée par la distorsion moyenne :

$$D = E [d(x, Q(x))] = \sum_{i=0}^{N-1} \int_{S_i} p(x) \|x-y\|^2 dx \quad (2.9)$$

Et le débit du quantificateur Q est mesuré par son entropie  $H_Q$ :

$$H_Q = - \sum_{i=0}^{N-1} P_i \log_2 P_i \quad (2.10)$$

$$\text{Avec } P_i = \int_{S_i} p(x) dx = P_X(x \in S_i) \quad (2.11)$$

Voyons à présent quelle serait la distorsion moyenne minimale d'un quantificateur Q ayant une entropie donnée  $H_Q$ .

a-bornes asymptotiques des performances:

Considérons, par souci de simplicité, le cas asymptotique où la performance des quantificateurs tend vers une meilleure valeur possible. Dans ce cas, le nombre de "niveaux" ou de vecteurs arrondis N, sera assez grand ( $H_Q$  grand). Par conséquent nous pouvons considérer que la probabilité  $p(x)$  varie peu à l'intérieur de la région  $S_i$ , on peut donc assimiler cette dernière à  $p_i$ , si  $x \in S_i$ , ( $p(x) \approx p_i$ ).

A partir des relations déjà citées nous obtenons :

$$P_i = \int_{S_i} p(x) dx \approx p_i \int_{S_i} dx = p_i V(S_i) \quad (2.12)$$

L'expression de la distorsion peut être réécrite comme suit :

$$D = - \sum_{i=0}^{N-1} \frac{P_i}{V(S_i)} \int_{S_i} \|x - y_i\|^2 dx \quad (2.13)$$

On montre que [ 17] :

$$\int_S d(x, y) dx = \int_{S_i} \|x - y_i\|^2 dx \geq \frac{V(S)}{V_k} \int_{u: \|u\| \leq 1} \left\| \left( \frac{V(S)}{V_k} \right)^{1/k} \cdot u \right\|^2 du \quad (2.14)$$

Soit  $F_k(v)$  la fonction :

$$F_k(v) = \frac{1}{V_k} \int_{u: \|u\| \leq 1} \left\| \left( \frac{V(S)}{V_k} \right)^{1/k} \cdot u \right\|^2 du \quad (2.15)$$

L'inégalité précédente peut s'écrire :

$$\int_S d(x, y) dx = \int_{S_i} \|x - y_i\|^2 dx \geq V(S) \cdot F_k \left( V(S)^{1/k} \right) \quad (2.16)$$

Et nous obtenons [17] :

$$D \geq \sum_{i=0}^{N-1} P_i F_k \left( V(S_i)^{1/k} \right) = E \left[ F_k \left( V(X)^{1/k} \right) \right] = D_L \quad (2.17)$$

$D_L$  est une borne inférieure aux performances asymptotiques du QV obtenues par la distorsion moyenne  $D$ .

Sachant que  $F_k$  est une fonction convexe, l'application de l'inégalité de Jensen à l'équation précédente donne :

$$D \geq D_L \geq F_k \left( E \left[ V(X)^{1/k} \right] \right) \quad (2.18)$$

$$\text{où } F_k^{-1}(D) \geq F_k^{-1}(D_L) \geq E \left[ V(X)^{1/k} \right] \quad (2.19)$$

A l'aide des équations (2.10) et (2.12), nous obtenons une nouvelle expression de l'entropie :

$$H_Q = \sum_{i=0}^{N-1} \int_{S_i} p(x) \log_2 p(x) dx - \sum_{i=0}^{N-1} \int_{S_i} p(x) \log_2 V(S_i) dx \quad (2.20)$$

on peut en déduire :

$$H_Q = h(p) - E \left[ \log_2 V(X) \right] \quad (2.21)$$

où  $h(p)$  est l'entropie différentielle

$$h(p) = - \int p(x) \log_2 p(x) dx \quad (2.22)$$

L'inégalité de Jensen donne :

$$H_Q = h(p) - k E \left[ \log_2 V(X)^{1/k} \right] \geq h(p) - k \log_2 \left( E \left[ V(X)^{1/k} \right] \right) \quad (2.23)$$

En utilisant la relation (2.19), on obtient :

$$H_Q \geq h(p) - k \log_2 \left( F_k^{-1}(D_L) \right) \geq h(p) - k \log_2 \left( F_k^{-1}(D) \right) \quad (2.24)$$

Ce qui donne après avoir élevé au carré :

$$D \geq D_L \geq F_k \left( 2^{-\left( H_Q - h(p) \right) / k} \right) \quad (2.25)$$

La relation exprime une borne inférieure asymptotique de la distorsion moyenne obtenue par un quantificateur Q ayant une entropie fixe  $H_Q$ . En remplaçant  $F_k(v)$  dans la relation ci-dessus, on obtient :

$$D \geq D_L \geq \frac{k V_k^{-2/k}}{k+2} 2^{-2/k} (H_Q - h(p)) \quad (2.26)$$

b-Application à une source gaussienne sans mémoire :

Comparons la borne inférieure D exprimée par la relation (2.26) à celle de Shannon  $D_{LSB}(R)$ . Il est à remarquer que nous utilisons la distorsion moyenne ramenée à l'échantillon, soit :  $\bar{D} = D/k$

Le débit binaire d'un quantificateur est déterminé par la relation (1.2.10), sachant que  $H_Q/k = H_k \leq \bar{R}$ , nous obtenons à partir de la relation (2.26), pour  $\bar{R}$  fixe :

$$\bar{D} = \frac{D}{k} \geq \frac{V_k^{-2/k}}{k+2} 2^{-2(\bar{R} - h_k)} = D_Q^{(k)}(R) \quad (2.27)$$

où  $H_k = \frac{H_Q}{k}$  et  $h_k = \frac{h(p)}{k}$  sont les entropies normalisées.

Par ailleurs, la borne  $D_{LSB}(R)$  peut se mettre sous la forme suivante :

$$\bar{D}_{SLB}^{(k)}(R) = \frac{D_{SLB}}{k} = \left[ 2e V_k^{2/k} \Gamma\left(1 + \frac{k}{2}\right)^{2/k} \right]^{-1} 2^{-2(\bar{R} - h_k)} \quad (2.28)$$

En combinant les équations (2.27) et (2.28), on obtient la borne :

$$\bar{D} \geq \bar{D}_Q^{(k)}(R) = \left[ \left( \frac{e}{1+k/2} \right) \Gamma\left(1 + \frac{k}{2}\right)^{2/k} \right] \bar{D}_{SLB}^{(k)}(\bar{R}) \quad (2.29)$$

La relation (2.29) détermine une borne inférieure des performances d'un quantificateur optimal de  $N$  vecteurs arrondis de dimension  $k$  chacun. Cette borne apporte des améliorations par rapport à la fonction prédite par la fonction  $D(R)$  pour  $k$  fixe, et tend vers la limite de la fonction de distorsion lorsque  $k$  tend vers l'infini.

La réciproque de la fonction  $R(D)$ , pour une source gaussienne s'exprime comme suit :

$$D(R) = 2^{-2R} \quad (2.30)$$

Aussi la distorsion que peut atteindre un quantificateur optimal  $Q$  en fonction de la dimension des blocs et pour un débit binaire fixe donné  $R$  s'écrit :

$$\bar{D}_Q^{(k)}(R) = \left[ \left( \frac{e}{1+k/2} \right) \Gamma\left(1+\frac{k}{2}\right)^{2/k} \right] 2^{-2R} \quad (2.31)$$

Il nous semble utile de rappeler que les performances peuvent être également exprimées par le rapport signal à bruit RSB ou SNR (signal to noise rate) en décibel.

Celui-ci est le rapport de la variance  $\sigma_x^2$  de la source  $X$  à quantifier et de la distorsion moyenne  $D$  introduite par le quantificateur :

$$RSB = 10 \log_{10} \frac{\sigma_x^2}{D} \quad (2.32)$$

Pour notre source gaussienne, on obtient :

$$RSB_{\max}(R) = 6.02 \times R \quad (2.33)$$

C'est la performance maximale possible que l'on peut atteindre selon la borne inférieure de Shannon pour la distorsion moyenne. En utilisant la relation (2.31), on détermine le RSB en fonction de la dimension  $k$  des blocs ou des vecteurs de quantification :

$$RSB_Q^{(k)}(R) = RSB_{\max}(R) - 10 \log_{10} \left[ \left( \frac{e}{1+k/2} \right) \Gamma \left( 1 + \frac{k}{2} \right)^{2/k} \right] 2^{-2R} \quad (2.34)$$

où le  $RSB_{\max}(R)$  est la borne de Shannon.

### 2.2.2 - Codage, modulation et quantification.

La théorie de l'information et de la communication fait ressortir une certaine corrélation entre la théorie de la modulation qui concerne le codage de canal et le codage de la source.

L'étude des structure régulières, algébriques peut montrer à quel point les concepts "codage de source" et "modulation" sont proches l'un de l'autre.

Les deux branches se basent sur la représentation de signaux dans un espace euclidien. De nombreux travaux de recherche leurs sont communément applicables, en particuliers les applications des réseaux réguliers. Les deux concepts ont énormément bénéficié des résultats obtenues en codage par treillis et par bloc. On dit en effet, dans la terminologie propre à ce domaine qu'il y a une dualité entre le codage et la modulation.

Dans les deux domaines, les résultats théoriques sont surtout basés sur la mesure de la distance euclidienne correspondante au bruit blanc gaussien additif sur le canal des systèmes de modulation, et sur le choix du critère de l'erreur quadratique moyenne MSE comme mesure de distorsion dans les système de codage de source.

En codage de canal, on cherche le meilleur compromis entre la maximisation du nombre des signaux émis et la minimisation du taux d'erreur à la réception c'est à dire on cherche à maximiser l'information mutuelle moyenne  $I(X;Y)$ ; en codage de source on souhaite réduire le débit en limitant le nombre de représentants sans augmenter la distorsion ce qui nécessite une minimisation de l'information mutuelle  $I(X;Y)$ . De là, on voit toute la corrélation qu'il y a entre la modulation et le codage dans les systèmes de communications numériques.

Illustrons ces propos par un exemple de quantification vectorielle [17] :

Soit  $Q$  un quantificateur défini par un dictionnaire  $B = \{y_0, y_1, \dots, y_{k-1}\}$  et par une partition  $S = \{S_0, S_1, \dots, S_{N-1}\}$  de l'espace euclidien  $\mathcal{R}^k$ . Un exemple de partition dans l'espace  $\mathcal{R}^2$  est représenté par la figure ci-dessous (figure 2.5).

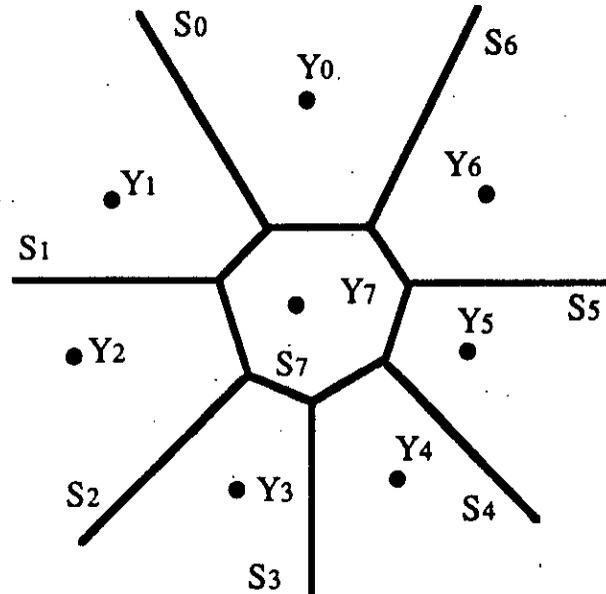


Figure 2.5: Exemple de partition dans  $\mathcal{R}^2$  pour  $N=8$

Trois opérations successives sont nécessaires pour quantifier un vecteur donné  $X$  appartenant à l'espace  $\mathcal{R}^k$  [8].

1- Retrouver la région de quantification contenant  $x$ , c'est à dire la région de Voronoï  $S_i$  telle que  $x \in S_i$  ce qui revient à déterminer le plus proche voisin de  $x$  dans le dictionnaire  $B$ .

2- Chercher l'indice  $i$  de la région de Voronoï contenant  $x$  qui est aussi l'indice du mot de code  $y$  le plus proche de  $x$ .

3- Régénérer à la réception le mot de code  $y_i$  à partir de l'indice  $i$ .

Ces trois opérations sont représentées par la figure ci-dessous (figure 2.6).

En conclusion, on remarque que le codeur de source enlève la redondance des mots ou vecteurs de la source à coder ne laissant que l'information utile telle que déterminée par le critère de fidélité. Le codeur de canal introduit une forme de redondance qui assure une immunité optimale au signal contre le bruit.

L'intérêt à présenter ici quelques notions sur le codage est que les codes peuvent être utilisés comme dictionnaire possédant des propriétés particulières pour les quantificateurs vectoriels. Cet aspect peut faciliter le décodage, organiser le dictionnaire, réduire sa taille, augmenter la vitesse des traitements [7].

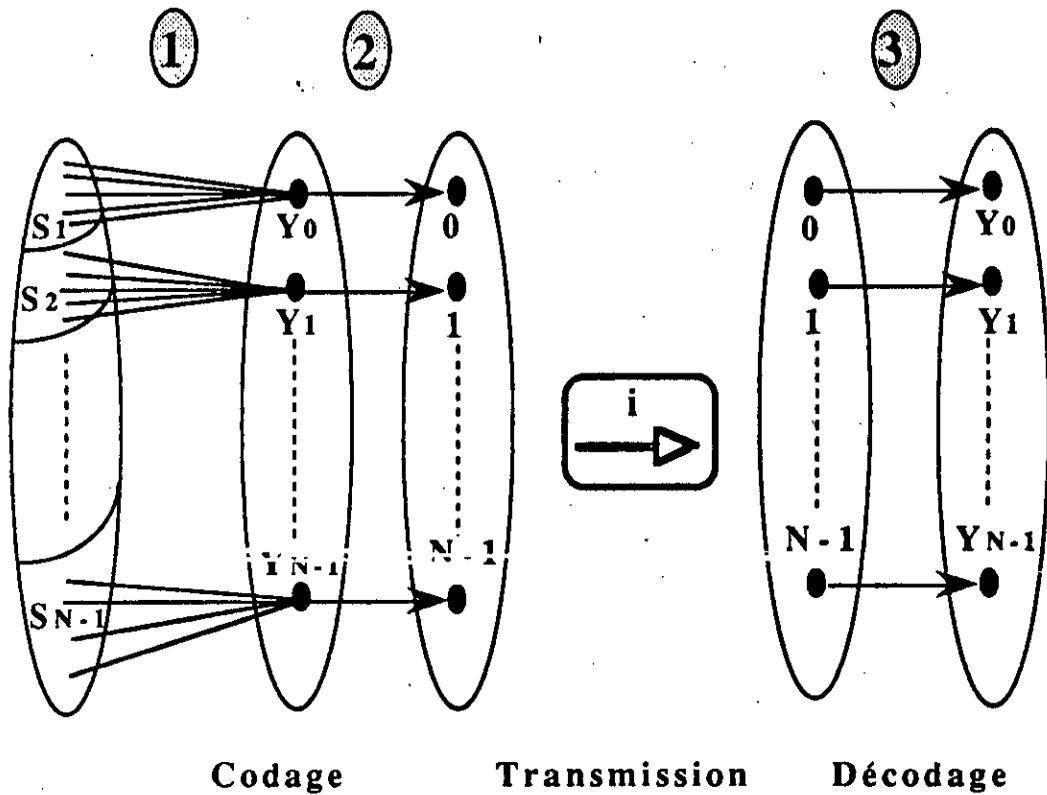


Figure 2.6: Schématisation des trois opérations d'une quantification vectorielle.

D'importantes recherches ont été faites dans le domaine des codes correcteurs d'erreurs [18, 19, 20]. Le phénomène de durcissement de sphère montre que la distribution des composantes gaussiennes du vecteur à quantifier a une symétrie sphérique et tend vers ce que l'on appelle "une coquille sphérique". Par la suite le problème de l'élaboration d'un code a été ramené à celui de l'empilement de sphère dans l'espace euclidien. Plusieurs travaux de recherche suivirent et débouchèrent sur les études des réseaux réguliers [21-27].

Il existe justement une méthode de quantification appelée sphérique qui exploite les propriétés des réseaux réguliers. Elle consiste à quantifier séparément le gain ou la norme et l'orientation ou la phase. Les avantages de cette méthode sont bien connus [7]; même la répartition des débits entre la norme et la phase a été étudiée [28].

L'utilisation des réseaux réguliers dans l'étude des quantificateurs est appelée approche algébrique; il existe une approche statistique qui consiste à appliquer un algorithme [9,29] désigné par le vocable de "K-moyenne" ou " LBG ", du nom de ses auteurs : Lynde, Buzo et Gray, à un ensemble de vecteurs normalisés d'apprentissage.

Cette approche est recommandée lorsque la distribution des orientations sur la sphère unité est non uniforme.

L'algorithme de la k-moyenne est généralement attribué à Mac Queen [30], Adoul et al. l'ont également proposé dans le contexte du traitement de la parole [31, 32] comme objet de classification et enfin, Lynde et al. ( L.B.G.) le présente comme une extension de l'algorithme de Lloyd-Max pour des variables vectorielles [29].

Nous présentons quelques détails sur l'algorithme de la k-moyenne que nous utiliserons plus loin. Cet algorithme itératif propose une fonction qui transforme les représentants ou centres d'une partition pour obtenir une nouvelle partition dont l'erreur est inférieure ou égale à celle de la partition originale . Son application successive va faire tendre l'erreur de la partition vers un minimum local [33]. Rien ne garantit la convergence de cet algorithme vers un minimum absolu. La figure ci-dessous ( figure 2.7 ) décrit cet algorithme [34].

Après une constitution des classes initiales, chaque itération correspond à une construction de centres et de classes. Au départ les centres des classes sont choisis d'une

façon arbitraire parmi les éléments de l'ensemble fini considéré B. La construction des classes consiste en fait à partitionner cet ensemble fini d'éléments B en k classes données.

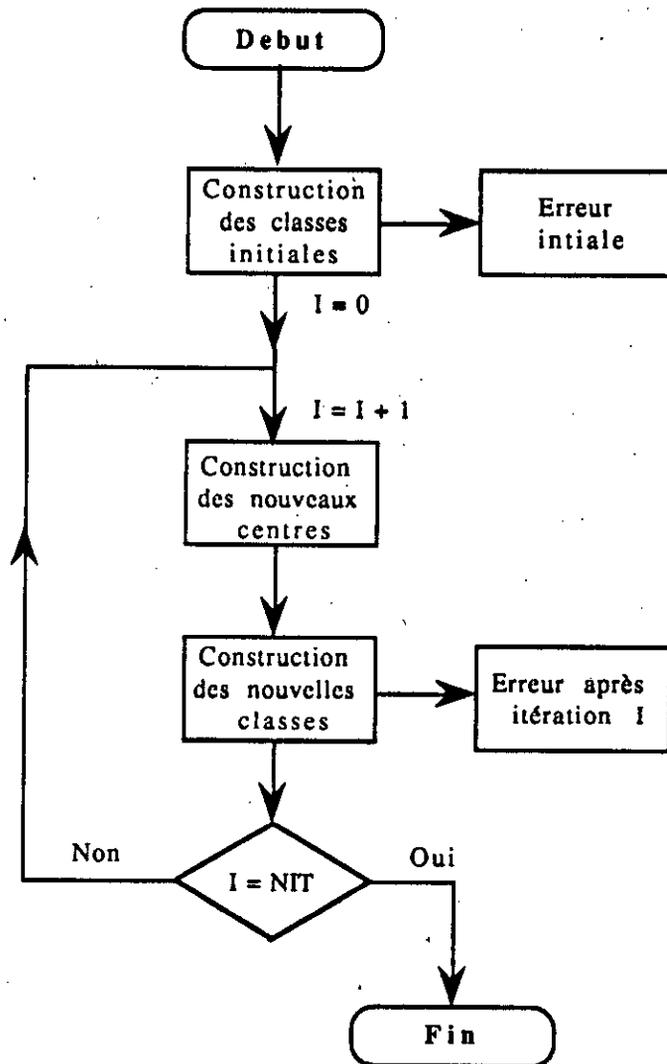


Figure 2.7: Structure générale de l'algorithme de la K-Moyenne. NIT est le nombre d'itérations.

Les éléments de l'ensemble B sont considérés un par un et sont placés dans la classe dont le centre est le plus proche. Les centres ne sont pas affectés par cette étape au cours de laquelle seule l'addition des distances entre chaque élément et le centre de sa classe ou l'erreur de la partition est déterminée.

La construction des nouveaux centres forme l'essentiel de l'algorithme. Dans cette étape les éléments des classes restent inchangés tandis que les centres seront renouvelés.

Pour chaque classe, il faudra trouver un nouveau centre tel que l'erreur due à cette classe soit minimale. Une sorte de barycentre est alors déterminé.

### 2.3 - Quantification par treillis ("Treillis Coded Quantization")

L'idée de la modulation par treillis ou "Treillis Coded Modulation" (TCM) a été introduite par Ungerboeck en 1982 [35-37]. En utilisant une technique appelée "Mapping by Set Partitioning" à des constellations de dimension 1 et 2, et des codes convolutifs, Ungerboeck a obtenu de bonnes performances dans le codage de canal à bande limitée. Ces travaux ont retenu une attention particulière des chercheurs à travers le monde. Plusieurs schémas de TCM ont été élaborés [37-39].

Forney et al. ont formalisé leur conception en montrant qu'il peuvent être décrits par les trois éléments suivants [40] (figure 2.8) :

1- Un encodeur binaire par blocs ou convolutif [41] qui permet de générer un nombre de bits plus grand à sa sortie que le nombre de bits à son entrée.

2- Ces bits codés servent à sélectionner une classe (ou sous-ensemble) d'une constellation partitionnée.

3- Des bits additionnels permettent de choisir un point dans la classe précédemment sélectionnée.

L'utilisation d'un ensemble dilaté de signaux fournit une redondance au codage tout en permettant de pouvoir maximiser la distance euclidienne minimale, appelée "free distance" entre les séquences des signaux codés.

La dualité entre la modulation et la quantification a conduit Fischer et Marcellin à appliquer le concept du "Mapping by Set Partitioning" dans la quantification par treillis, appelée "Treillis Coded Quantization" (TCQ) [42].

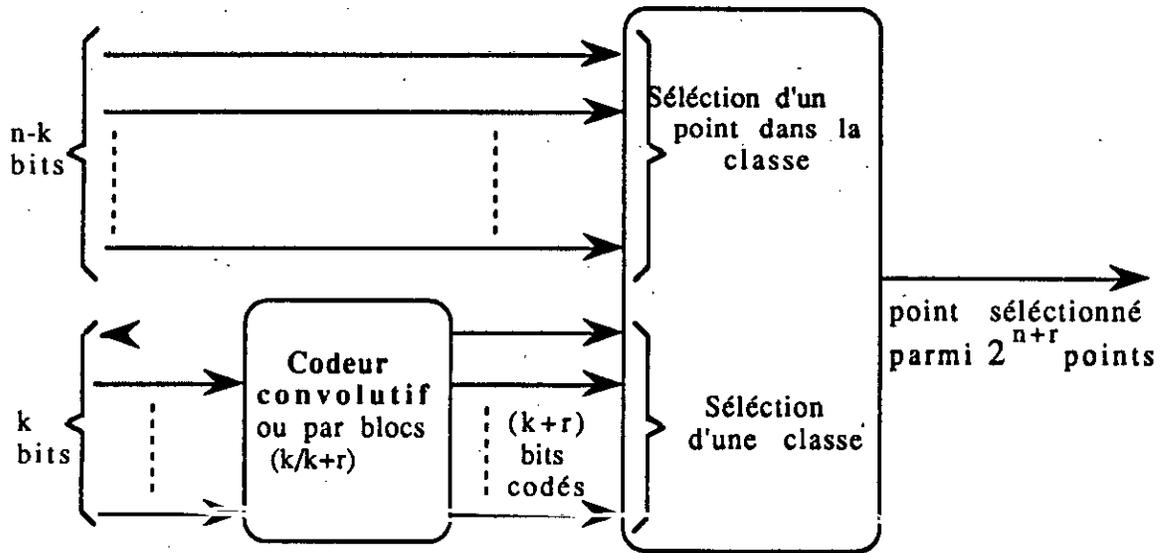


Figure 2.8: Structure d'un codeur TCM.

### 2.3.1 - Description du TCQ

Avec la quantification par treillis (TCQ), il s'agit tout d'abord de doubler la taille de l'alphabet vectoriel tout en conservant le même débit. Le TCQ impose quelques règles sur l'utilisation de l'alphabet afin de conserver le débit initial donné. Supposons que nous ayons à quantifier une source  $X$  à un débit de  $R$  bits par échantillon.

La formation de l'alphabet se fait de la manière suivante :

Soit  $Y$  l'alphabet vectoriel composé de  $2^m$  vecteurs à  $k$  dimensions. Il faudra donc doubler le nombre de vecteurs afin d'obtenir un alphabet vectoriel  $Y'$ , appelé également constellation, constitué de  $2^{m+1}$  vecteurs. On divise cette dernière en  $2^{m'+1}$  classes ou sous-alphabets  $\{ Q_0, Q_2, \dots, Q_m \}$ , contenant chacune  $2^{m-m'}$  vecteurs. Les classes seront assignées aux branches des treillis publiés par Ungerboeck, selon les règles du "Set Partitioning" [35, 36]. Par ailleurs chaque sous-alphabet doit lui-même être un bon

quantificateur [11]. Un exemple d'alphabet utilisé dans le cas d'un treillis à quatre états, avec un débit d'un demi bit par échantillon est illustré ci-dessous (figure 2.9).

Dans le cas d'une utilisation d'un quantificateur vectoriel "classique", il faudrait  $m+1$  bits et  $m-m'$  bits pour identifier respectivement la classe et le vecteur type, ce qui donnerait un débit total de  $m+1$  bits .

Cependant, en faisant appel à un treillis, il nous offrira la possibilité de ramener le débit à  $R$  bits.

La technique consiste à n'utiliser, pour chaque vecteur, que  $2^m$  des  $2^{m+1}$  vecteurs types disponibles dans la constellation mais en changeant les membres de ce groupe de  $2^m$  vecteurs selon des règles connues tant du codeur que du décodeur.

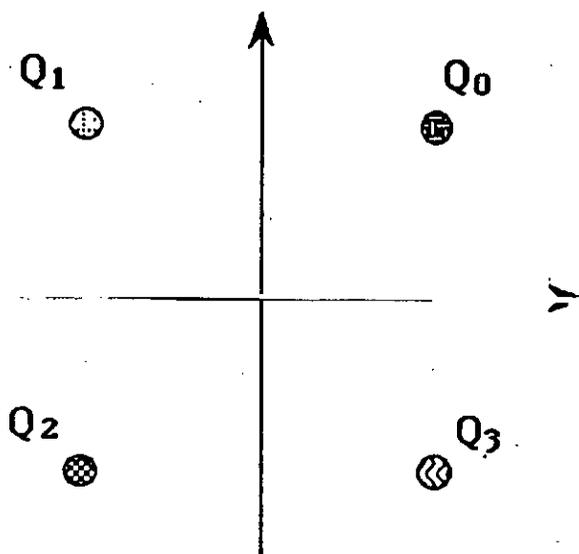


Figure 2.9: Alphabet d'un TCQ à un demi bit par échantillon.

Les treillis décrits par Ungerboeck sont déterminés à partir des transitions obtenues par le codeur convolutif de la figure 2.10.

Nous allons détailler un exemple de TCQ à quatre états avec un alphabet vectoriel de 16 vecteurs types séparés en 4 classes Q0, Q1, Q2, Q3. Notre codeur peut être représenté par un diagramme de transitions : 2 transitions arrivent à un état et 2 autres transitions permettent de quitter un état (figure 2.11 ).

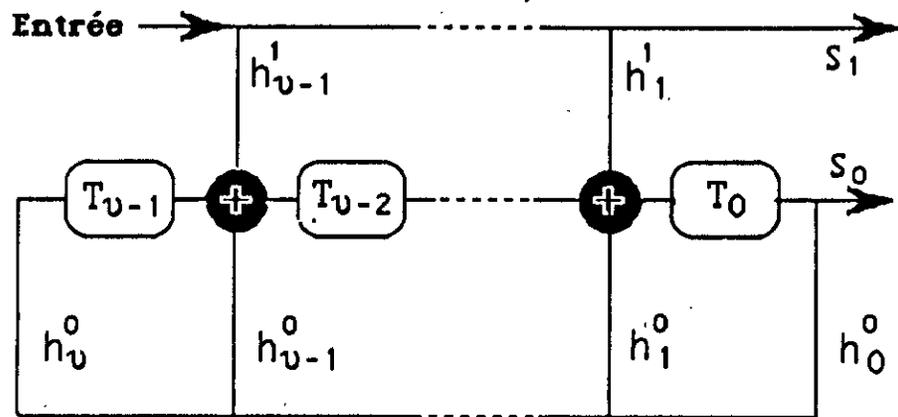


Figure 2.9 : Codeur convolutif avec un Nombre d'états égal à  $2^u$ ,  $h_0$ ,  $h_1$  sont publiés par Ungerboeck [ 37 ].

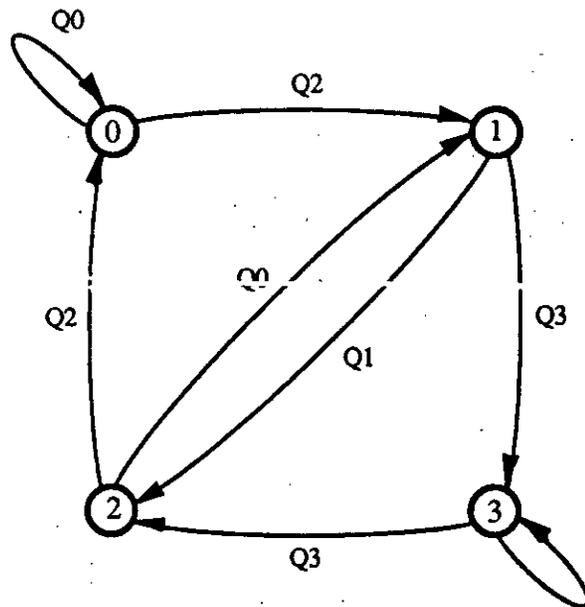


Figure 2.10 : Diagramme de transition.

Grâce au codeur convolutif ( figure 2.10 ) qui sélectionne une classe parmi les quatres, on peut donner une autre représentation du diagramme de transition, on obtient le diagramme suivant (figure 2.11 bis) :

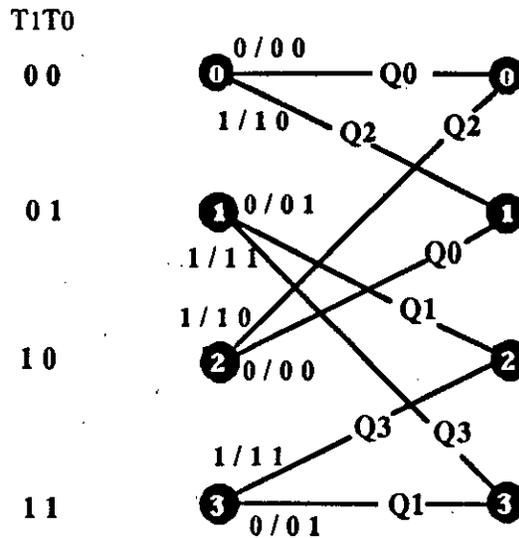


Figure 2.11 bis : Exemple de treillis à quatre états.

Dans le diagramme de transition de la figure 2.11, les états du quantificateur autorisent d'utiliser soit les classes  $Q_0, Q_2$  soit  $Q_1, Q_3$  ; de plus, le sélection d'un sous-alphabet permis occasionne une transition vers un nouvel état. De ce choix entre les quantificateurs possibles dépend la façon dont seront quantifiés les vecteurs suivants.

En effet , supposons que nous nous trouvions à l'état 0 du treillis et que nous ayons à quantifier un vecteur  $x_0$  quelconque. Dans cet état , nous avons le choix d'utiliser l'un des vecteurs types de  $Q_0$  ou de  $Q_2$  pour représenter  $x_0$ . Si nous choisissons notre vecteur type dans  $Q_0$  , nous serions tenus de quantifier le prochain vecteur  $x_1$  avec le même (  $Q_0, Q_2$  ) groupe.

Si par contre nous utilisons  $Q_2$  pour  $x_0$ ,  $x_1$  serait quantifié par l'un des quantificateurs  $Q_1, Q_3$ . Comme le choix des deux ensembles dépend des prochains vecteurs, nous attendrons de les connaître pour prendre une décision. Entre temps, nous

déterminons la distorsion ou l'erreur causée par chaque meilleur représentant de  $Q_0$  et  $Q_2$  et établissons l'arbre suivant ( figure 2.12 ).

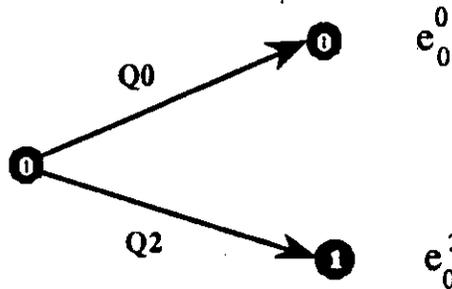


Figure 2.12: Transitions possibles pour le premier vecteur considéré (si on part de l'état 0 du diagramme). On tente de coder  $x_0$  par le meilleur représentant de  $Q_0$  et de  $Q_2$ .

Considérons maintenant le vecteur suivant  $x_1$ . Si, pour coder  $x_0$  notre choix s'est porté sur  $Q_0$  nous répétons alors la figure précédente, dans l'autre cas il faudra se servir de  $Q_1$  et  $Q_3$  pour coder  $x_1$ . En choisissant encore le meilleur représentant de chacun des sous-alphabets  $Q_n$  et en reprenant la procédure qui a permis d'obtenir la figure, un arbre plus développé peut être établi (figure 2.13 ).

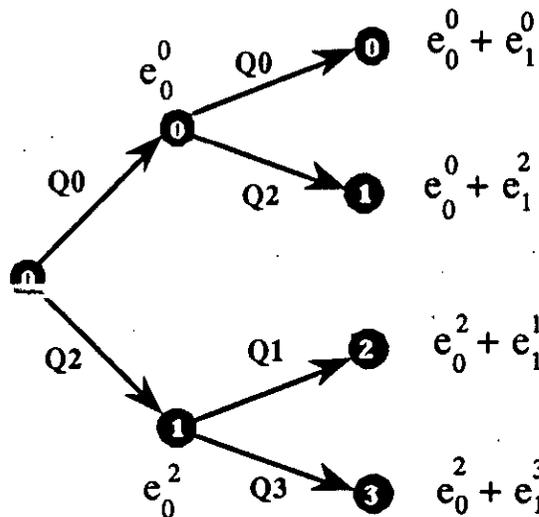


Figure 2.13: État de l'arbre binaire après la quantification du second vecteur  $x_1$ . Chacune des feuilles de l'arbre porte l'erreur accumulée sur les vecteurs  $x_0$  et  $x_1$ .

On voit déjà apparaître l'essence de la quantification par treillis. En effet l'erreur cumulée représente dans tous les cas la somme des distorsions produites sur les vecteurs  $x_0$  et  $x_1$ . La branche sur laquelle l'erreur est la plus faible indique la meilleure stratégie de quantification. Il est possible d'accepter une erreur plus grande sur  $x_0$  si ce choix s'avère payant pour les vecteurs suivants.

Pour le vecteur suivant  $x_2$ , on remarque que les branches de l'arbre 0 et 2 offrent les mêmes quantificateurs ( $Q_0$  et  $Q_2$ ) et ont des chemins qui aboutissent aux mêmes transitions; c'est aussi le cas pour les branches 1 et 3 avec les quantificateurs  $Q_1$  et  $Q_3$ .

En fait, à partir du troisième vecteur, l'arbre se replie sur lui-même et devient un treillis (figure 2.14).

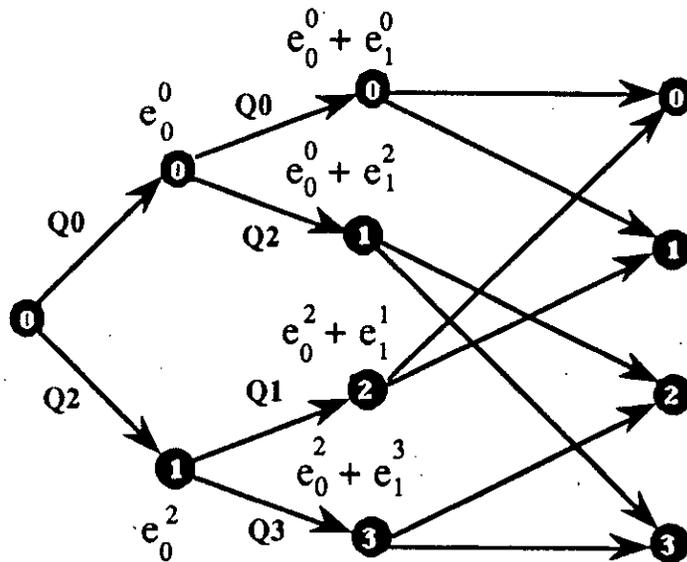


Figure 2.14 : État de l'arbre binaire au troisième vecteur. Cette configuration porte le nom de treillis.

### 2.3.2 - Interprétation

Chaque noeud dans le treillis représente un état et les branches correspondent aux transitions. Des deux chemins arrivant à chacun des états, un seul présente un certain intérêt.

En effet, chacun de ces chemins représente un passé différent, mais ils auront le même avenir car ils aboutissent au même état.

Le chemin qui aura accumulé la plus faible erreur sera le chemin restant ou le survivant. Il représente, pour les chemins arrivant à un état donné la meilleure stratégie de quantification pour les vecteurs passés.

Il y aura quatre de ces survivants (chemins arrivants à chacun des états) pour chaque nouveau vecteur  $x_i$ . Après un nombre  $n$  de vecteurs, on choisira parmi ces quatre chemins celui qui a cumulé la plus faible erreur.

La différence entre le TCQ et les autres techniques de codage par treillis est que chaque branche du treillis correspond à un quantificateur plutôt qu'à un symbole binaire.

Pour minimiser la distance euclidienne, on utilise l'algorithme de Viterbi [43,44] :

Au départ, on assigne une erreur cumulée nulle à l'un des états du treillis et une erreur infinie aux autres. Le codeur et le décodeur s'entendent préalablement sur le choix de l'état ayant une erreur nulle.

1- Dans une première étape, on calcule les distances euclidiennes pour chaque étage.

$$\left(d_j^i\right)^2 = (x_i - q_j)^2 \quad \text{avec } q_j \in Q_j, \quad j = 0, 1, \dots, 3 \quad (2.35)$$

2- Puis, on additionne les distorsions causées par chaque branche et à un noeud donné, nous gardons la branche qui cause la distorsion minimale.

3- A la fin du treillis nous comparons les quatre états et nous choisissons le noeud ayant la distorsion cumulée minimale.

Ces trois étapes nécessitent 2 multiplications, 8 additions et 2 comparaisons par échantillon pour un treillis à quatre états.

## 2.4 - Les quantificateurs polaires.

Plusieurs types de contraintes ou difficultés peuvent surgir lors de la mise en oeuvre d'un quantificateur vectoriel.

Les chercheurs préfèrent présenter le système quantificateur comme un ensemble de quantificateurs élémentaires scalaires déjà connus.

Il est parfois plus judicieux d'utiliser un repère polaire de coordonnées qu'un système cartésien. En effet un signal exprimé sous une forme complexe peut être obtenu et traité plus facilement dans un repère polaire où son amplitude et sa phase sont exprimés sous une forme plus explicite. Les travaux de recherche dans cet axe ont permis de mettre au point une nouvelle classe de quantificateurs appelés " Polar Quantizers " ( Quantificateurs polaires) [ 45].

Mais à notre sens, la majeure raison de faire appel à la quantification sous-optimale reste la difficulté de conception des quantificateurs vectoriels optimaux tels que définis par la théorie (au sens de la performance). Cette difficulté réside dans l'importance des moyens de calculs à mettre en oeuvre, ce qui exige des temps de calcul prohibitifs .

En effet, des conditions sévères sont exigées pour les centroïdes et les régions avoisinantes d'un quantificateur vectoriel optimal. Dans les algorithmes de design ou de conception, le centroïde de l'une des N régions de quantification  $S_i$  est donné par la relation :

$$y_i = \frac{\int_{S_i} x p(x) dx}{\int_{S_i} p(x) dx} \quad (2.36)$$

où  $p(x)$  est la distribution de la variable à quantifier.

Il est à remarquer que la dimension  $k$  des coordonnées d'un point implique  $k$  divisions scalaires et  $k+1$  intégrations de dimension  $k$ , soit pour les  $N$  régions à considérer, un total de  $kN$  divisions et  $kN$  intégrations.

La complexité s'accroît lors de la définition des régions avoisinantes de quantification  $S_i$  pour chaque valeur arrondie de sortie. Il s'agit de partitionner un espace de dimension  $k$  ( $\mathcal{R}^k$ ); on recherche alors les intersections des demi-espaces comme suit :

$$S_i = \bigcap_{\substack{j=1 \\ j \neq i}}^N \{ x : |x - x_i| < |x - x_j| \} \quad (2.37)$$

La relation (2.37) montre que la détermination de chaque région  $S_i$  nécessite la recherche de  $N-1$  demi-espaces de dimension  $k$  soit un total, pour  $N$  régions, de  $N^2$  intersections.

Une étude analogue peut être faite pour montrer l'ampleur des difficultés auxquelles on doit faire face lors de l'implémentation d'un quantificateur vectoriel notamment lors de l'estimation de la distorsion. Malgré l'alternative développée par Gersho [15, 46], la conception d'un QV optimal reste un art en soi.

Une étude comparative relative à la complexité des quantificateurs vectoriel et scalaire, faite par Swaszek [45], est résumée par le tableau 2.1.

La complexité des QV a incité les chercheurs à s'orienter vers la quantification sous optimale. Chacun tente par de nouvelles études, méthodes ou techniques de récolter un certain gain en performances sans exiger des conditions contraignantes ou supplémentaires; ce qui revient à dire qu'il y a réduction de la complexité de façon indirecte.

Ainsi Huang et Shultheis ont modélisé une source gaussienne à éléments corrélés par un système constitué d'une combinaison d'opérateurs linéaires et d'une banque de quantificateurs scalaires de Max-Lloyd [47].

Bucklew et Gallagher [48], Pearlman [13,28] ont développé l'étude des sources gaussiennes à variables aléatoires complexes. Cette étude bidimensionnelle, faite dans un système de coordonnées polaires, consiste, essentiellement, à minimiser la distorsion (MSE). Ils ont ainsi défini une classe de quantificateurs dits polaires. Les régions de quantification se présentent sous la forme de tronçons ou sections d'anneaux comme le montre la figure 2.15.

	Quantificateur vectoriel	Quantificateur scalaire
<b>Conception</b>		
centroïdes	$O(kN)$ k-dim intégrations	$O(kN^{1/k})$ intégrations scalaires
	$O(kN)$ divisions scalaires	$O(kN^{1/k})$ divisions scalaires
régions	$O(N^2)$ k-dim intersections	$O(kN^{1/k})$ opérations scalaires
<b>Implémentation</b>		
	$O(N^2)$ opérations scalaires	Aucune
	$O(N)$ comparaisons scalaires	$O(\log_2 N)$ comparaisons scalaires

Tableau 2.1 : Etude comparative relative la complexité des quantificateurs scalaire et vectoriel.

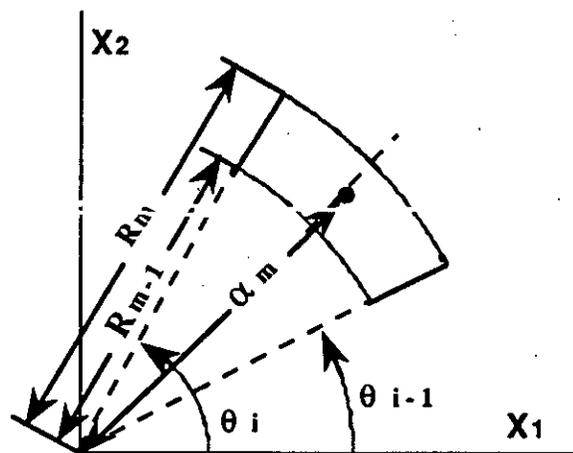


Figure 2.15: Région de quantification d'un quantificateur polaire.

Ces travaux montrent que la répartition uniforme de la phase reste une condition d'optimalité de ce type de quantificateur.

Pearlman a étudié la répartition des débits entre l'amplitude et la phase, cette dernière est sensée contenir plus d'information que l'amplitude. Ont été également déterminés et tabulés les niveaux de quantification.

Wilson [49], Swaszek et Thomas [50] ainsi que Swaszek et Ku [51] ont généralisé les études des quantificateurs polaires symétriques (polar Quantizers) aux quantificateurs symétriques dits "Unrestricted polar quantizers" (UPQ). La levée de la restriction concerne l'octroi d'un nombre variable de niveaux à la phase pour un niveau d'amplitude donné. Un exemple de quantificateur UPQ est donné par la figure 2.15 .

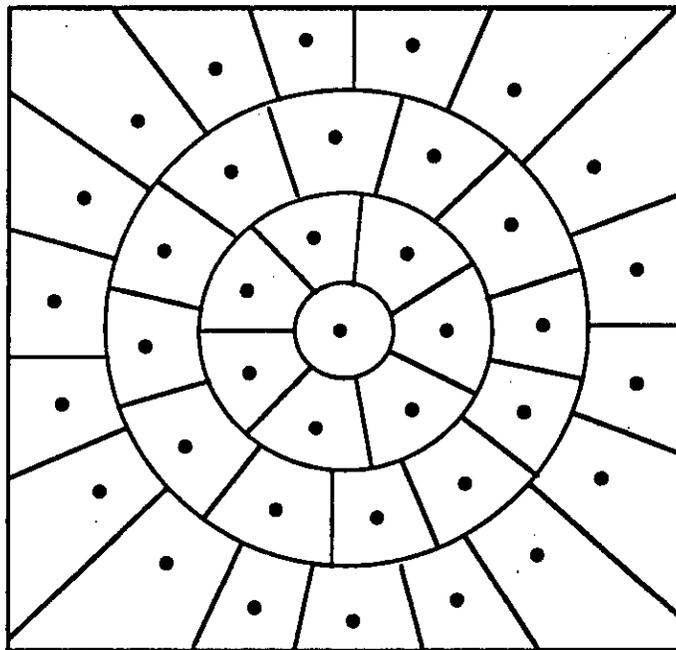


Figure 2.16: Quantificateur polaire à niveaux de phase indépendants "unrestricted polar quantizer".

Le prolongement de ces recherches a abouti à l'étude des quantificateurs polaires à plusieurs dimensions [52].

## **conclusion**

Nous avons présenté l'essentiel de la théorie de la quantification et de la modulation . Cette partie permet de situer le problème de design des quantificateurs : les limites théoriques des performances des quantificateurs restent une référence.

Diverses techniques et méthodes de quantification ont été présentées. La description des techniques et des algorithmes, sur la base desquels sont conçus et fonctionnent nos logiciels de calcul des performances, permet de mieux interpréter ces derniers.

Nous avons situé le niveau actuel des connaissances dans ce domaine grâce à la rétrospective des travaux de recherche présentée. Cette rétrospective nous apprend les solutions préconisées par les chercheurs, les difficultés rencontrées et les résultats qu'ils ont obtenus. Elle demeure, pour nous, une base de départ de toute conception d'un quantificateur nouveau.

Ainsi, dans la dernière section de ce chapitre, nous avons présenté les quantificateurs polaires de type symétrique (symétrie circulaire) qui demeurent assez proches du "Spiral Quantizer". Ces quantificateurs ont fait l'objet de travaux de recherche publiés , leurs performances sont connues et des recommandations pour leur conception ont été formulées par leurs auteurs. Pour la quantification de signaux gaussiens, dans le cas d'une indépendance statistique entre l'amplitude et la phase, ils présentent l'avantage de fournir une répartition uniforme de la phase. Cependant, on peut remarquer une dépendance géométrique entre ces deux paramètres.

Tout en conservant les avantages de ce type de quantificateurs nous avons développé l'étude du "Spiral Quantizer" dont une des particularités est que l'expression mathématique de l'amplitude et celle de la phase sont interdépendantes. Cette propriété nous permet de tirer d'autres avantages que nous exposerons plus loin. Nous entamerons cette étude par une description du "Spiral Quantizer" au chapitre suivant.

## Chapitre 3

# Le "Spiral Quantizer"

### Introduction

Dans cette partie, nous présentons un quantificateur d'un type nouveau, appartenant à la classe des quantificateurs sous-optimaux, basé sur l'utilisation de la spirale d'Archimède: le "Spiral Quantizer". Une méthode d'optimisation pour une meilleure discrétisation de la spirale sera suivie; elle nous permettra de déterminer les valeurs optimales des ses paramètres [53]. Ce sont ces derniers qu'il faudra stocker dans le dictionnaire.

Après une description du "Spiral Quantizer" et des conditions d'expérimentation qui concernent essentiellement la source des vecteurs utilisée, nous présenterons les performances atteintes par notre quantificateur en terme d'erreur quadratique moyenne et de RSB. Ces performances seront comparées aux bornes ou limites théoriques formulées par la théorie de la distorsion, de la quantification et de la modulation exposée aux chapitres précédents.

On montrera l'amélioration tant qualitative que quantitative apportée par l'algorithme de la K-Moyenne aux performances obtenues.

Nous verrons comment le RSB fourni par le "Spiral Quantizer" se stabilise après un nombre déterminé de vecteurs d'entrée, c'est à dire comment notre quantificateur atteint le régime stationnaire.

Nous verrons également comment se fait la répartition des vecteurs générés par la spirale d'Archimède en "Set Partitioning" et les performances obtenues par l'utilisation de cette nouvelle technique.

**3.1- Description du "Spiral Quantizer".**

Considérons une source aléatoire gaussienne bidimensionnelle, l'expression d'un point Z, dans l'espace E de coordonnées cartésiennes X et Y, dans un système de coordonnées polaires satisfait aux relations suivantes :

$$\begin{cases} x = r \cos \theta \\ y = r \sin \theta \end{cases} \quad \text{et} \quad \begin{cases} r = \sqrt{x^2 + y^2} \\ \theta = \arctg \frac{y}{x} \end{cases} \quad (3.1.1)$$

En notation complexe, cela se traduit par :

$$Z = x + jy = re^{j\theta} \quad \text{où} \quad j^2 = -1 \quad (3.1.2)$$

Notre quantificateur a été conçu à partir d'une discrétisation de la spirale d'Archimède que nous présentons ci-dessous (figure 3.1.1).

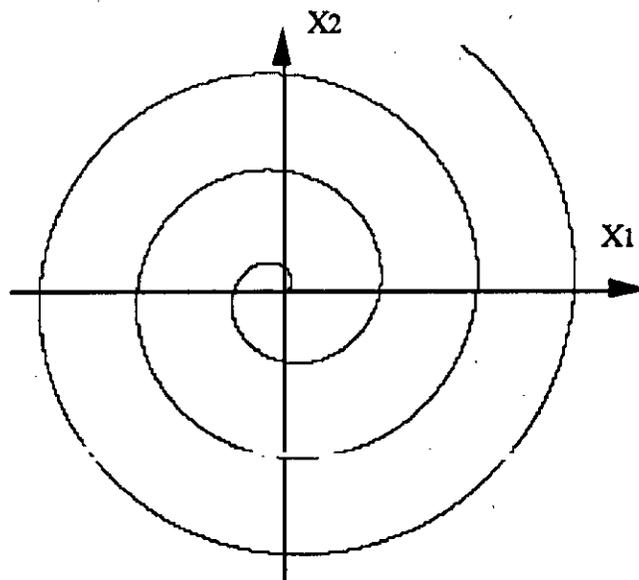


Figure 3.1.1: La spirale d'Archimède dans le système de coordonnées cartésien.

Les points décrits par cette spirale sont régis par les équations paramétriques suivantes :

$$\begin{cases} x = r \cos \theta \\ y = r \sin \theta \end{cases} \quad (3.1.3)$$

$$\text{où } r = G\varphi \quad \text{et} \quad \varphi = \theta + 2k\pi \quad (3.1.4)$$

En notation complexe, on obtient :

$$Z = x + jy = re^{j\theta} = G\varphi e^{j\theta} \quad (3.1.5)$$

Si la densité de probabilité bidimensionnelle des variables gaussiennes à quantifier s'exprime par la relation ci-dessous :

$$P_{xy}(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (3.1.6)$$

où  $\sigma^2$  est la variance.

Cette même densité dans un repère de coordonnées polaires est donnée par l'expression :

$$P_{r\theta}(r,\theta) = \frac{r}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \quad (3.1.7)$$

Exploitant l'indépendance statistique des variables malgré leur dépendance géométrique, on obtient les densités de probabilité de l'amplitude et de la phase :

$$P_r(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \quad 0 \leq r < \infty \quad (3.1.8)$$

$$P_\theta(\theta) = \frac{1}{2\pi} \quad 0 \leq \theta < 2\pi \quad (3.1.9)$$

L'étape suivante consiste à définir les centroïdes ou régions de quantification comme cela a été vu au chapitre précédent. Une des particularités de la définition de ces régions par rapport à celles d'un quantificateur polaire symétrique reste le fait que le module  $r$  n'est plus une constante comme cela a été déjà souligné par la relation 3.1.4.

Cette propriété de spirale exige une attention toute particulière. En effet, dans le cas où  $G$  est un facteur d'échelle constant ( ce qui est vrai pour une spirale d'Archimède) l'expression mathématique de l'amplitude est une fonction linéaire de la phase. Ceci va donner une autre forme à la définition des limites d'intégration et un certain avantage concernant le débit au "Spiral Quantizer" que nous commenterons à la fin de cette section.

Considérons un "Spiral Quantizer" avec  $N$  régions de quantification, on divise ainsi l'espace en  $N$  régions. Chaque région est délimitée par les spires de la spirale et les niveaux de phase considérés. On supposera que le nombre de niveaux de phase est égal au nombre de niveaux des amplitudes ; cette contrainte sera commentée un peu plus loin.

Considérons également les régions  $N'$  sous forme d'anneaux dont le nombre est identique à  $N$ , telles que décrites par la figure 3.1 et définies comme suit :

$$\theta_{m-1} \leq \theta < \theta_m \quad (3.1.10)$$

$$r_{m-1} \leq r < r_m \quad (3.1.11)$$

où  $r_m = \text{constante}$  et  $m = 1, 2, \dots, M$

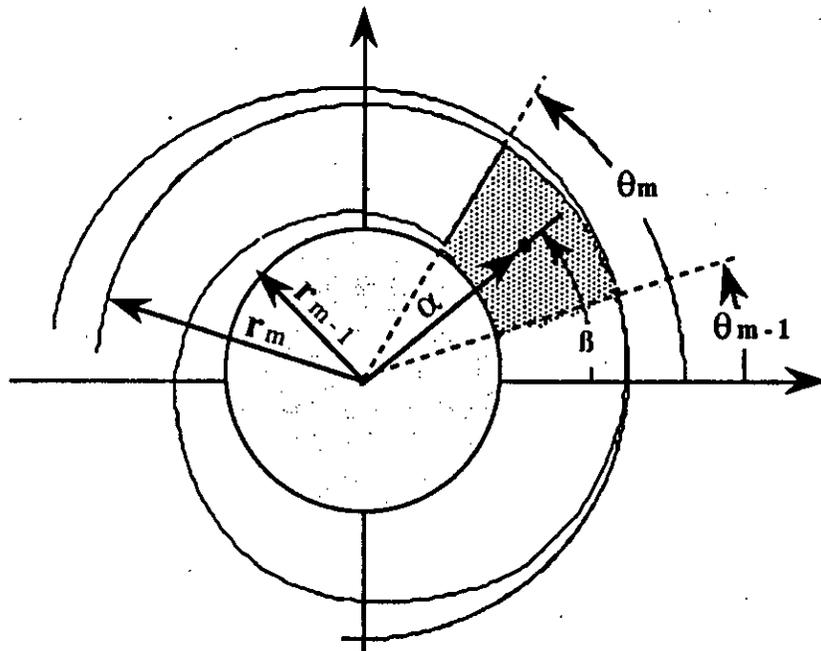


Figure 3.1.2: Discrétisation de la spirale.

Si on assimile chaque région  $N_i$  à une région correspondante  $N'_i$ , on ramène le cas de notre quantificateur à celui d'un quantificateur polaire déjà étudié par plusieurs auteurs [28,48-52]. Cette condition reste possible lorsque les écarts de phase sont suffisamment petits. En considérant le quantificateur construit à base d'une spirale d'Archimède comme un SPQ "Strict Polar Quantizer" [28], la contrainte serait :  $N = MP$  où  $M$  et  $P$  sont respectivement les nombres de niveaux d'amplitude et de phase. L'expression générale de la distorsion s'écrit:

$$D = \sum_{m=1}^M \sum_{p=1}^P \int_{r_{m-1}}^{r_m} \int_{\theta_{p-1}}^{\theta_p} \left| re^{j\theta} - \alpha e^{j\beta} \right|^2 P_{r\theta}(r, \theta) dr d\theta \quad (3.1.12)$$

Où  $\alpha$  et  $\beta$  sont les valeurs arrondies de sortie. Des solutions à cette question ont été formulées par Gallagher [54], la résolution systématique avec les niveaux de quantification appropriés ont été proposés par Pearlman et Senge [13]. De plus, dans notre cas si l'on suppose que les niveaux d'amplitude et de phase sont égaux, on peut tirer les identités suivantes :

$$r_{m-1} = G\theta_{p-1} = G\theta_{m-1} \quad (3.1.13)$$

$$r_m = G(\theta_{p-1} + 2\pi) = G(\theta_{m-1} + 2\pi) \quad (3.1.14)$$

$$p = m = 1, 2, \dots, M$$

$$\alpha_m = \text{sinc}\left(\frac{1}{M}\right) \frac{\int_{r_{m-1}}^{r_m} r P_r(r) dr}{\int_{r_{m-1}}^{r_m} P_r(r) dr} \quad (3.1.5)$$

$$r_m = \frac{\alpha_m + \alpha_{m-1}}{2 \text{sinc}\left(\frac{1}{M}\right)} \quad (3.1.16)$$

Dans le cas d'un quantificateur UPQ "Unrestricted Polar Quantizer" [49], la contrainte serait :

$$\sum_{m=1}^M P_m = N \quad \text{avec} \quad (m-1)\frac{2\pi}{P_m} \leq \theta < m\frac{2\pi}{P_m} \quad (3.1.17)$$

Où  $m = 1, 2, \dots, P_m$

Des résultats similaires à ceux déjà publiés [49,50] peuvent être obtenus.

Pearlman a établi la répartition des débits entre l'amplitude et la phase, celle-ci doit satisfaire à la relation :

$$R_\theta - R_r = 1.376 \text{ bits} \quad (3.1.18)$$

Où  $R_\theta$  et  $R_r$  sont respectivement les débits de phase et d'amplitude par variable.

Cependant la question la plus importante consiste à exploiter ces résultats, et interpréter ces méthodes en vue de donner un avantage sûr au quantificateur que nous proposons. Nous avons dégagé un avantage en ce qui concerne le débit de stockage, propriété que nous interprétons de la manière suivante.

Le module  $r$  pour notre spirale s'exprime selon la relation 3.1.4 comme suit :

$$r = G\varphi \quad \text{où} \quad \varphi = \theta + 2m\pi \quad \text{et} \quad m = 1, 2, \dots, M$$

Afin de simplifier le problème, supposons que  $|G| = 1$ , dans ce cas, on obtient:

$$r = \varphi = \theta + 2m\pi \quad (3.1.19)$$

La relation précédente montre que toute l'information de la phase, pour une spirale, est implicite dans l'amplitude, c'est une des propriétés de la spirale. On pourrait donc, si cela était nécessaire, ne stocker finalement que les  $N$  valeurs de  $\varphi$  du quantificateur. Bien entendu cette opération est en général superflue puisque les valeurs de  $\varphi$  sont directement générées par la spirale.

Puis, on calcule la phase à partir de l'amplitude selon la relation.

$$\theta = \varphi \text{ modulo } 2\pi = r \text{ modulo } 2\pi \quad (3.1.20)$$

Cette particularité peut être expliquée par le fait que les coordonnées d'un point de l'espace généré par une spirale d'Archimède est entièrement défini soit par sa distance à l'origine soit par son orientation (relations 3.1.3 à 3.1.5).

Ainsi, hormis les autres avantages que peut présenter le "Spiral Quantizer", il permet une réduction du débit de stockage, ce qui lui permet de surpasser les quantificateurs polaires déjà étudiés.

En conclusion, à partir de ce qui a été énoncé plus haut, la discrétisation d'une spirale d'Archimède [53] se résume, dans une première étape, à une discrétisation de l'angle  $\theta$ . On définira donc, dans nos logiciels de calcul des performances du "Spiral Quantizer", un pas d'incrément "delta" de cet angle et "delta" sera un des paramètres dont il faudra rechercher la valeur optimale.

La discrétisation de l'angle  $\theta$  n'est pas pour autant terminée car on estime qu'il est nécessaire d'octroyer un autre degré de liberté à la constellation de points générés par la spirale : celui de pouvoir s'éloigner de l'origine dans un mouvement de révolution. De ce fait, on définit un deuxième paramètre, une composante de l'angle  $\theta$ , l'angle initial "têta0" ou " $\theta_0$ " et qui fera également l'objet d'une optimisation par la suite.

Cependant, il faut rappeler que pour compléter ce processus de discrétisation, on tiendra compte du facteur d'échelle que l'on désignera par le "Gain" et qui permettra d'ajuster notre quantificateur vis à vis de la source gaussienne des vecteurs d'excitation.

Afin de mieux apprécier les performances et résultats obtenus, les principales caractéristiques, notamment statistiques, des vecteurs aléatoires sont l'objet de la section suivante.

### 3.2- La source d'excitation.

La source d'excitation utilisée pour tester le "Spiral Quantizer", suit une loi Normale de densité de probabilité, sa variance est égale à l'unité. Un exemple de répartition de ses points dans un espace à deux dimensions est donné par la figure 3.2.1.

La représentation bidimensionnelle de la forme  $X = [X_1, X_2]$  des éléments de la source est obtenue par couplage de variables gaussiennes deux à deux. Cependant, si cette alternative fournit la solution la plus simple et la plus directe qui permet d'obtenir un processus gaussien bidimensionnel à partir des échantillons d'un processus gaussien monodimensionnel, elle n'est pas la seule.

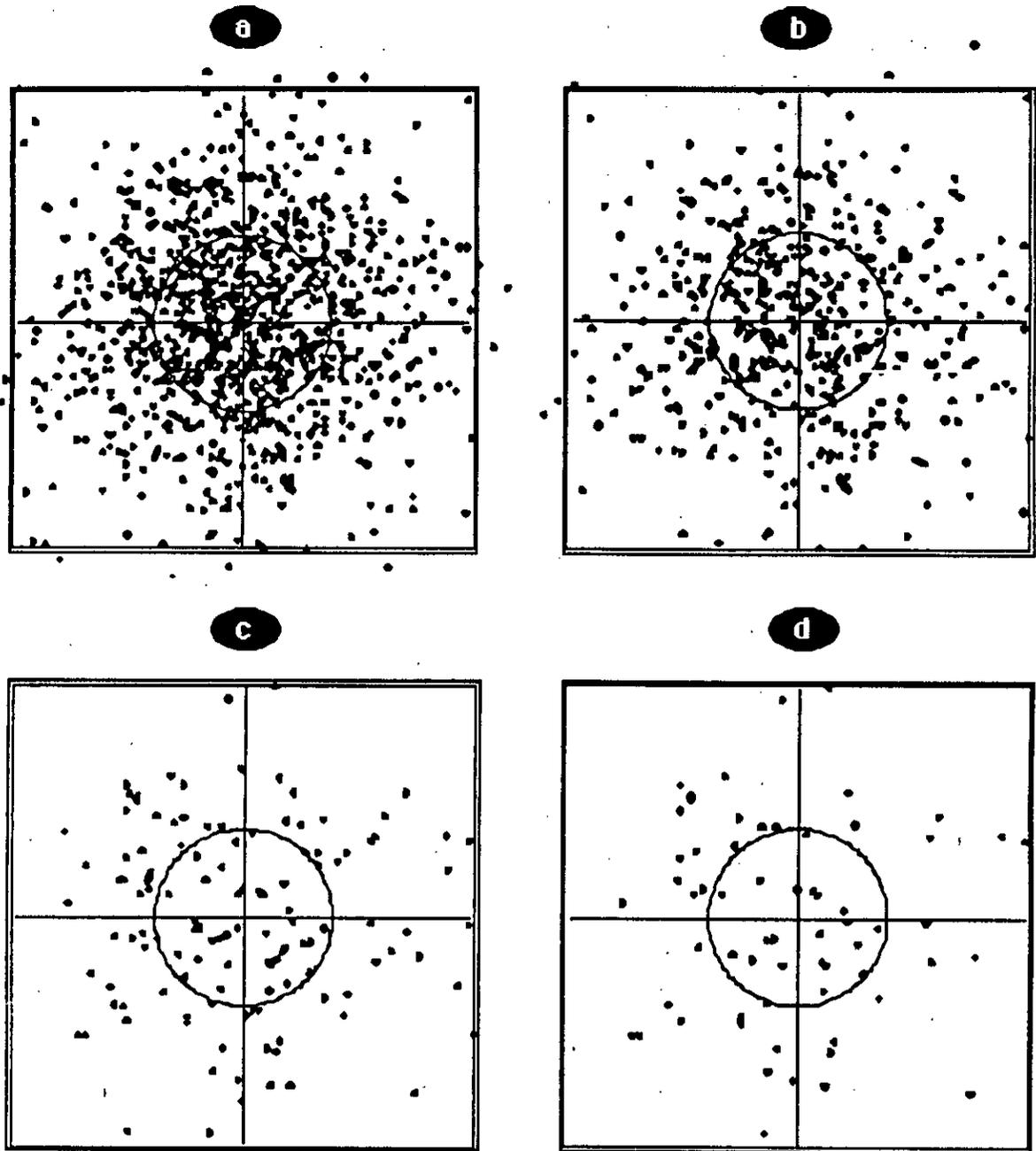


Figure 3.2.1: représentation bidimensionnelle des points de la source d'excitation : a-1024 points ; b- 512 points ; c- 128 points ; d- 64 points.

En effet , il existe un autre procédé basé sur la transformée de Fourier . Cette méthode présente des avantages multiples , on peut citer notamment :

- la Transformée de Fourier Discrète ( TFD ) peut faire partie intégrante du processus de traitement ou du phénomène physique de la transmission [55, 56],

- la TFD possède la propriété de décorrélation , elle est asymptotiquement optimale au sens de la Transformée de Karhunen-loève ( TKL ) [55, 57,58].

Nous présentons ci-dessous les propriétés statistiques de chaque colonne de coordonnées ( tableau 3.2.1, figure 3.2.2 ).

**X1 : Column 1**

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
.029134	.9854	.015397	.971014	3382.345414	4096
Minimum:	Maximum:	Range:	Sum:	Sum Squared:	* Missing:
-3.424244	3.181125	6.605369	119.331374	3979.777133	0
t 95%:	95% Lower:	95% Upper:	* < 10th %:	10th %:	25th %:
.030189	-.001056	.059323	410	-1.227523	-.635032
50th %:	75th %:	90th %:	* > 90th %:	Kurtosis:	Skewness:
.012595	.685612	1.317951	410	-.137102	.02963

**X2 : Column 2**

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
.032379	1.004113	.015689	1.008243	3101.140596	4096
Minimum:	Maximum:	Range:	Sum:	Sum Squared:	* Missing:
-3.376187	3.446303	6.82249	132.623719	4133.05124	0
t 95%:	95% Lower:	95% Upper:	* < 10th %:	10th %:	25th %:
.030762	.001616	.063141	410	-1.265778	-.656274
50th %:	75th %:	90th %:	* > 90th %:	Kurtosis:	Skewness:
.024062	.707026	1.342365	410	-.103002	.053786

Tableau 3.2.1: Caractéristiques statistiques individuelles de chaque type de coordonnée.

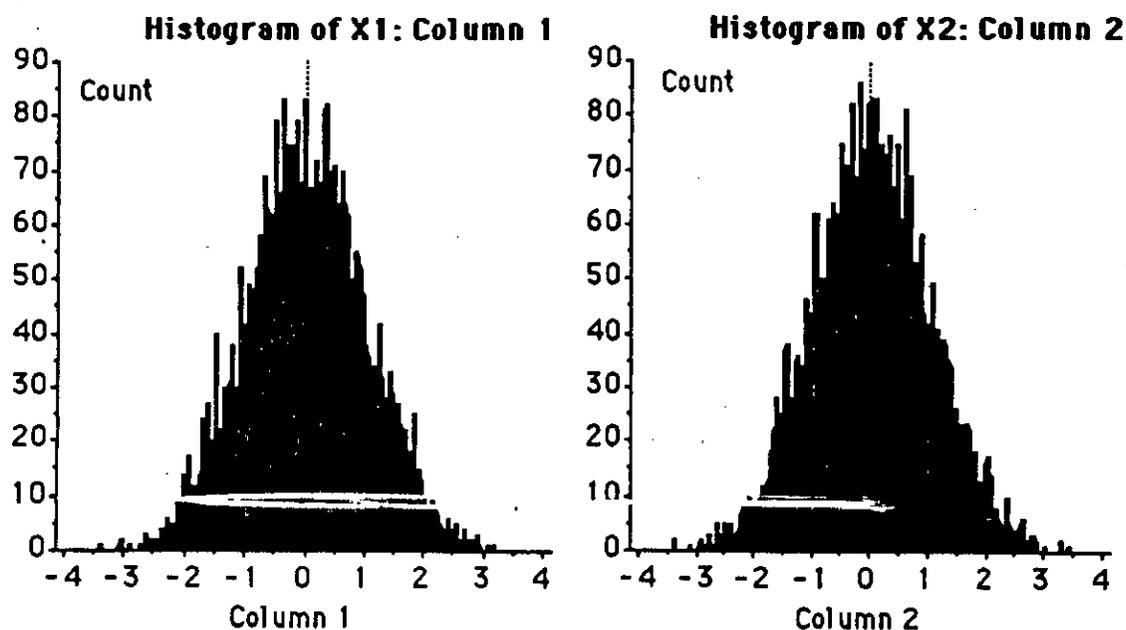


figure 3.2.2: Histogramme par colonne de données.

Les caractéristiques statistiques par paire d'échantillons formant une constellation de 4096 points dans un espace à deux dimensions sont fournies par le tableau 3.2.2.

**Paired t-Test X<sub>1</sub> : Column 1 X<sub>2</sub> : Column 2**

DF:	Mean X - Y:	Paired t value:	Prob. (2-tail):
4095	-.003245	-.147026	.8831

**Corr. Coeff. X<sub>1</sub> : Column 1 X<sub>2</sub> : Column 2**

Count:	Covariance:	Correlation:	R-squared:
4096	-.008135	-.008222	.000068

**Simple Regression X<sub>1</sub> : Column 1 X<sub>2</sub> : Column 2**

Count:	R:	R-squared:	Adj. R-squared:	RMS Residual:
4096	.008222	.000068	-.000177	1.004202

Tableau 3.2.2: Caractéristiques statistiques des paires d'échantillons.

Les mesures effectuées confirment les propriétés statistiques de la source comme nous le montrent les résultats présentés. Un logiciel d'extension par permutation a été conçu, il permet de disposer de  $2^{32}$  vecteurs pour les différents types de tests des quantificateurs.

### 3.3- Optimisation des paramètres de la spirale.

Les trois paramètres qui définissent la spirale d'Archimède sont l'angle initial "têta0", le pas ou l'incrément "delta" qui permet d'obtenir l'angle têta et le facteur d'échelle "gain".

Il s'agira de déterminer les valeurs de ces paramètres pour lesquelles le RSB fourni par le "Spiral Quantizer" est maximal. Dans une première étape, nous avons étudié comment varie le RSB en fonction de ces grandeurs pour une spirale de 16 points. Les résultats sont donnés sous forme graphique (figure 3.3.1 à 3.3.6) :

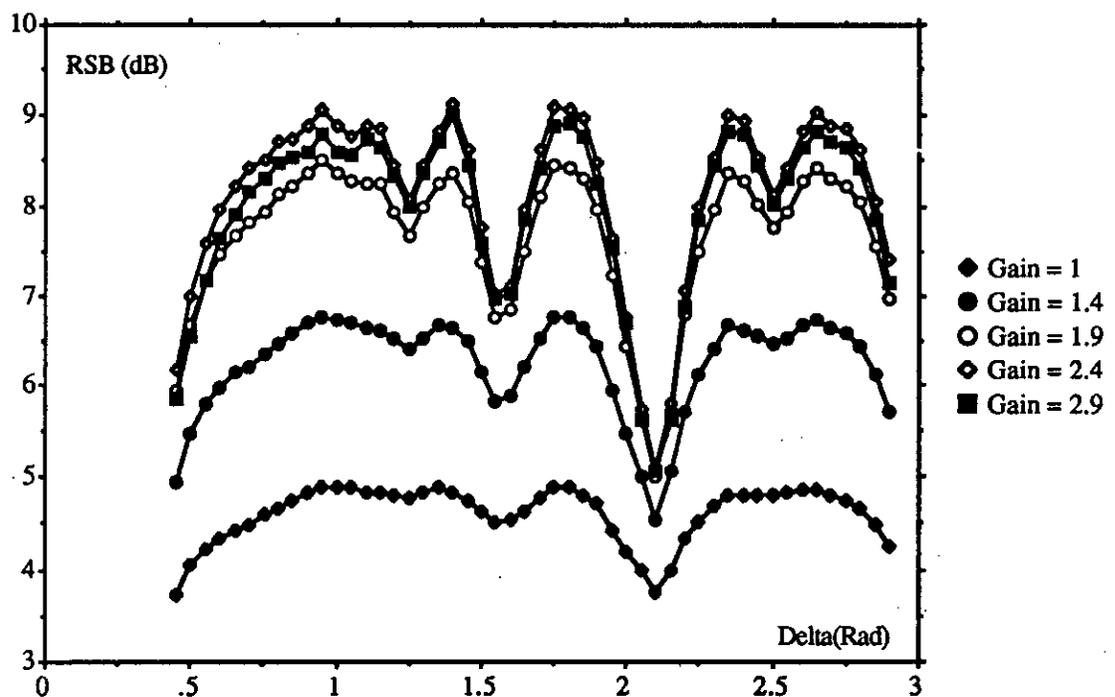


Figure 3.3.1: Variation du RSB (dB) en fonction de l'incrément delta de l'angle têta pour un ensemble de valeurs données du gain.

L'allure des courbes obtenues (figure 3.3.1) montre que l'on peut fixer les valeurs de delta qui maximisent le RSB et affiner les valeurs optimales du gain. On notera que ces valeurs ont été normalisées en utilisant la relation suivante:

$$\text{Gain} = \frac{\xi}{\theta_{\max}} = \frac{\xi}{\theta_0 + (N-1)\Delta} \quad (3.3.1)$$

Où  $\theta_0$  est l'angle initial "têta0"

La figure 3.3.2 donne les valeurs désirées du gain selon l'incrément ou le pas de l'angle  $\Delta$  utilisé dans l'algorithme de calcul.

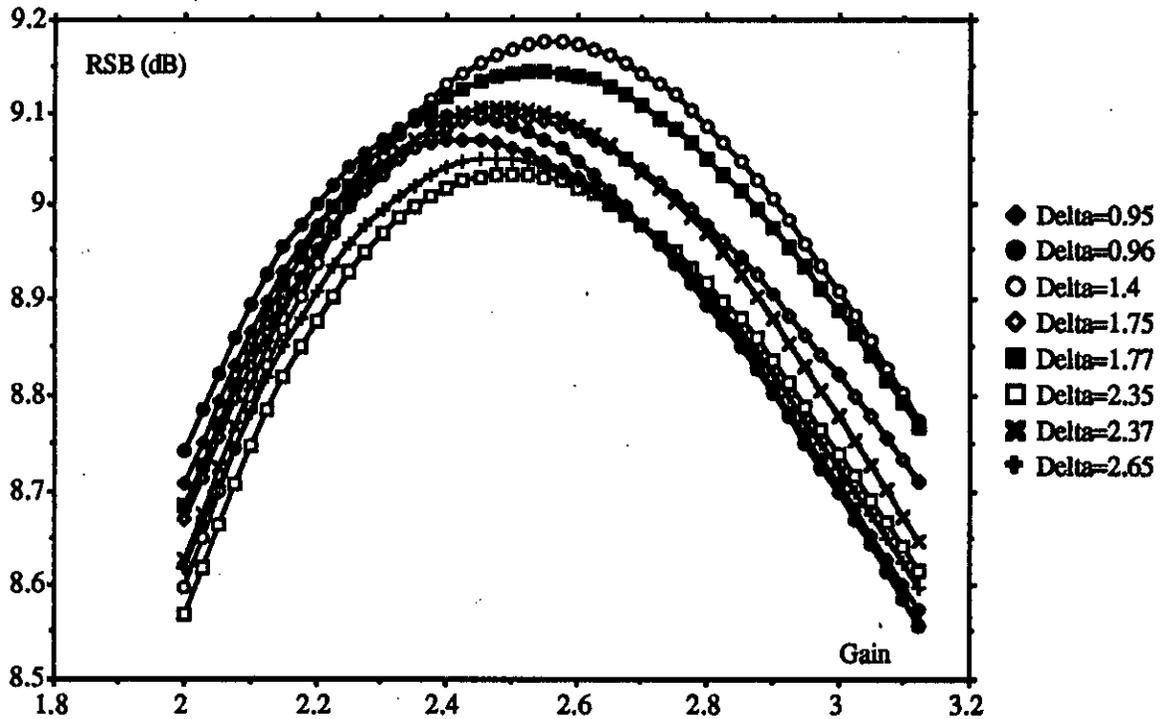


Figure 3.3.2: Variation du RSB(dB) en fonction du gain pour un pas constant .

Le troisième paramètre de la spirale est l'angle initial  $\theta_0$ , le graphe de la figure 3.3.3 donne les valeurs optimales pour un gain considéré.

Une deuxième recherche des valeurs optimales du gain est faite dans le cas où l'angle initial  $\theta_0$  n'est pas nul. Le graphe ci-dessous (figure 3.3.4) illustre cette recherche.

Les figures 3.3.5 et 3.3.6 montrent une certaine périodicité de  $2\pi$  des valeurs extrémales du RSB pour un gain donné.

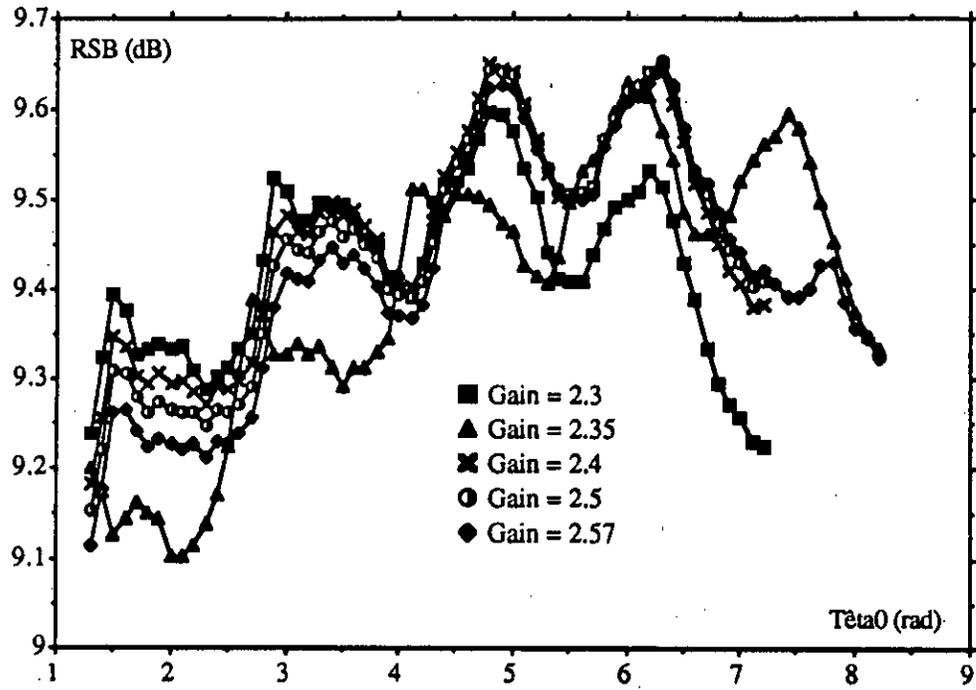


Figure 3.3.3: Variation du RSB en fonction de la valeur de l'angle initial  $t\eta a_0$ .

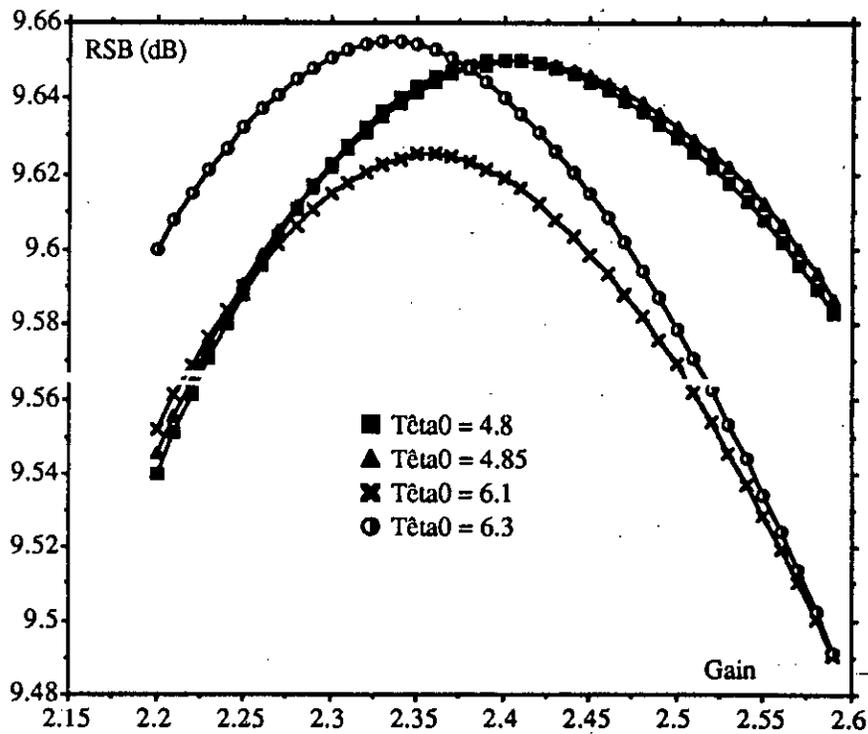


Figure 3.3.4 : Optimisation de la valeur du gain pour  $t\eta a_0$  donné.

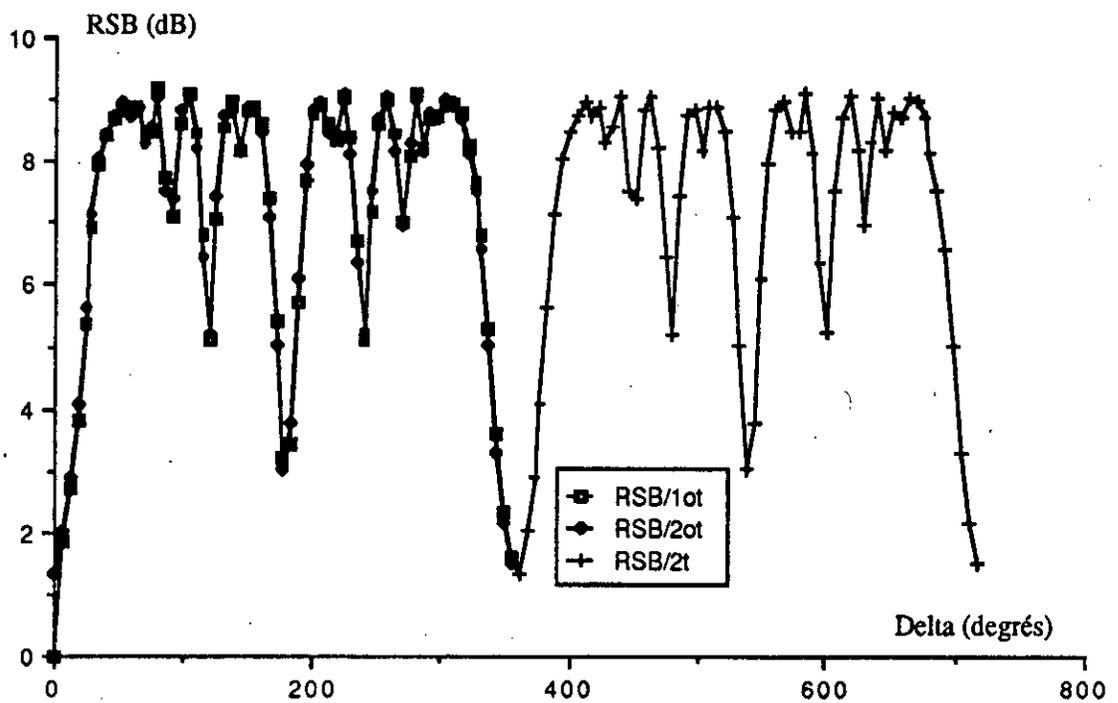


Figure 3.3.5: Variation du RSB (dB) en fonction de l'incrément delta de l'angle tête pour une valeur de gain optimale. RSB/1ot désigne les valeurs de delta allant de 0 à  $2\pi$ , RSB/2ot les valeurs de  $2\pi$  à  $4\pi$  et RSB/2t les valeurs de 0 à  $4\pi$ .

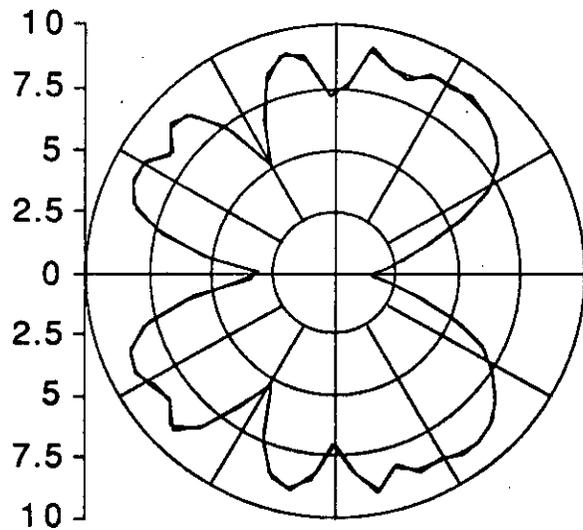


Figure 3.3.6: Particularité de la fonction  $RSB = f(\text{delta})$ .

Les figures 3.3.5 et 3.3.6 montre qu'il existe une certaine périodicité des fonctions tracées. Ainsi la courbe RSB/2 $\theta$  qui, initialement occupe une gamme de valeurs comprises entre  $2\pi$  et  $4\pi$  a été sciemment placée dans la plage des valeurs de 0 à  $2\pi$ . On démontre par là que les valeurs du RSB se répète pratiquement. Cette propriété que confirme le diagramme polaire de manière plus explicite, facilite la recherche des valeurs des paramètres de la spirale donnant les meilleurs RSB en limitant la plage des investigations.

Au cours de l'étude du quantificateur , nous avons considéré les spirales suivantes:

1- une spirale avec un angle initial  $\theta_0$  nul, notée "Spirale Z0".

2- une spirale avec un angle initial non nul, discrétisée en N-1 points à partir de  $\theta_0$  avec rajout d'un point à l'origine et notée "Spirale ZT".

3- et "Spirale P" qui désigne une spirale sans point à l'origine, discrétisée en N points à partir de  $\theta_0$ .

Nous résumons de façon sommaire, en attendant l'étude sur la stabilisation en fonction du nombre de vecteurs du débit, ces premiers résultats de l'optimisation à l'aide du tableau 3.3.1. Des résultats plus complets seront présentés plus loin.

	Delta	Gain	T $\theta_0$ (rad)	RSB (dB)
Spirale avec point à l'origine	1.4	2.57	0	9.176931
	1.4	2.4	6.2	9.651427
Spirale sans point à l'origine	1.4	2.34	6.3	9.654821

Tableau 3.3.1: Tableau récapitulatif des résultats de l'optimisation des spirales à 16 points.

### Chapitre 3

Cette étape d'optimisation des paramètres a été faite pour une spirale de 16 points.

Après une étude plus détaillée de notre quantificateur concernant le débit et des tableaux des valeurs numériques qui nous ont permis de tracer les courbes des graphes précédents, il a été remarqué que les valeurs des paramètres qui apparaissent pour une spirale de 16 points optimales ne le sont pas pour toutes les autres dimensions. Cet état de fait nous a conduit à rechercher ces mêmes valeurs pour les spirales d'un nombre de points allant de 2 à 512.

Nous nous sommes limités à détailler le cas d'un quantificateur à 16 points, à décrire la méthodologie suivie en vue de trouver les valeurs arrondies de sortie qui semblent les plus optimales possibles.

Pour les autres dimensions, les valeurs optimales seront données plus loin, en commun avec d'autres résultats.

Les résultats définitifs de cette investigation ne seront présentés qu'après l'étude de la stabilisation des performances en terme de RSB du quantificateur en fonction du nombre de vecteurs utilisés, ce qui permettra de mieux situer leur importance ( voir tableaux 3.5.1, 3.5.2 et 3.5.3 ).

Par ailleurs ces résultats seront donnés comme suit :

- par catégorie de spirale comme cela a été déjà souligné plus haut.
- avec et sans application de l'algorithme de la K-Moyenne pour chaque cas et catégorie de quantificateur.

#### **3.4- Etude de la variation du RSB en fonction du nombre de vecteurs.**

L'évaluation des performances d'un quantificateur donné est obtenu en effectuant des moyennes sur un certain nombre de résultats de quantification, c'est à dire sur le nombre de vecteurs d'une séquence test.

Afin de déterminer un nombre raisonnable de vecteurs pour la séquence test , nous allons analyser expérimentalement la convergence de l'estimateur moyenne pour le RSB.

Les résultats des variations du RSB en fonction du nombre de vecteurs pour les différents types de spirales avec un nombre de points  $N = 16$ , tels que déjà définis sont illustrés par les figures ci-dessous (figures 3.4.1 et 3.4.2).

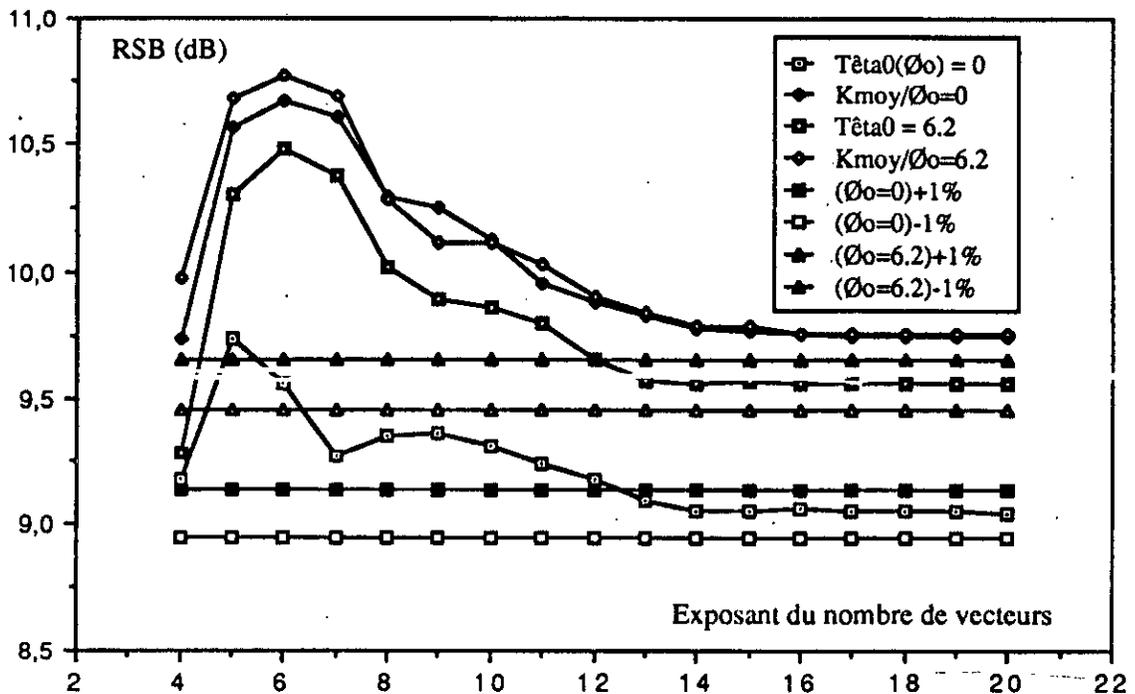


Figure 3.4.1 : Variation du RSB en fonction du nombre de vecteurs pour des spirales de types "Spirale Z0" et "Spirale ZT". Les courbes Kmoy désignent des spirales auxquelles a été appliqué l'algorithme de la K-moyenne .

Les résultats obtenus montrent que dans tous les cas le RSB se stabilise à partir de la valeur 12 de l'exposant du nombre de vecteurs soit à partir de l'utilisation d'un nombre de vecteurs égal ou supérieur à 4096. Par ailleurs ces résultats confirment que les meilleurs quantificateurs sont obtenus lorsque l'angle initial  $têta0$  n'est pas nul.

D'autre part, on peut d'ore et déjà affirmer que le gain en RSB réalisé avec l'application de la K-Moyenne est peu significatif.

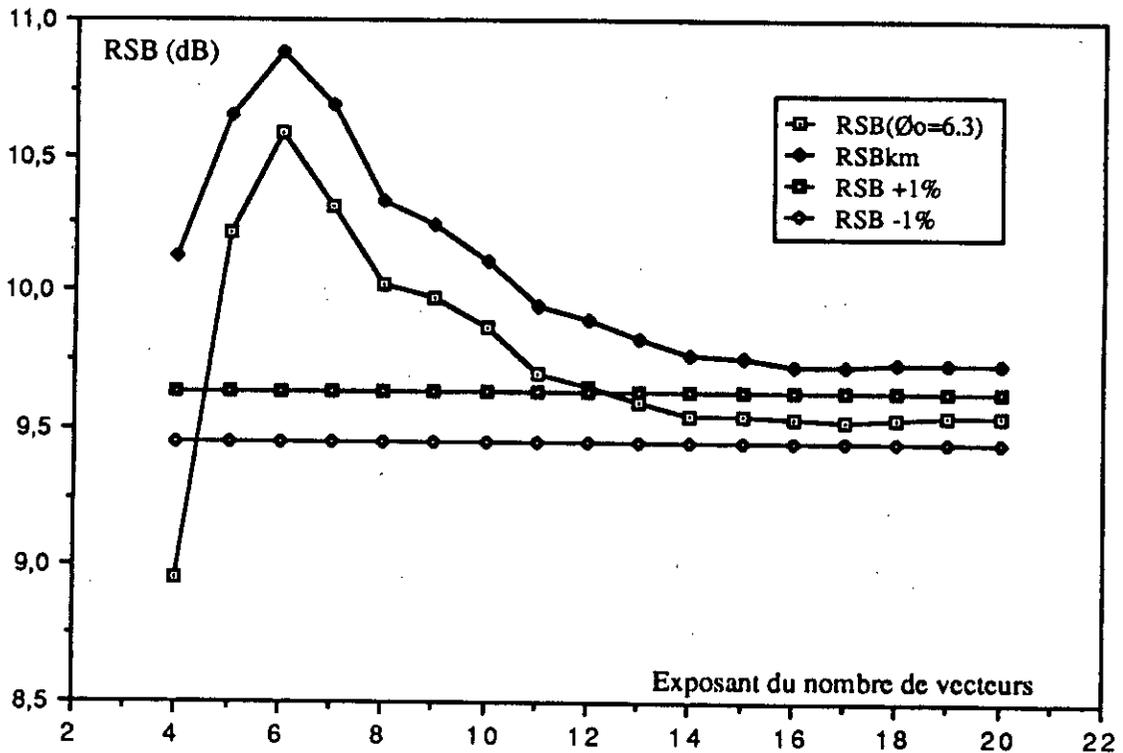


Figure 3.4.2 : Variation du RSB en fonction du nombre de vecteurs pour un "Spiral Quantizer" de type "Spirale P", avec et sans application de l'algorithme de la K-moyenne.

### 3.5-Performances du "Spiral Quantizer".

Nous présentons une autre catégorie de résultats graphiques (de la figure 3.5.1 à la figure 3.5.3) qui montre les variations du RSB en fonction du nombre de points de la spirale. Les graphes illustrent les performances de notre quantificateur et le situe par rapport aux limites théoriques définies par Shannon.

La figure ci-dessous présente les performances en termes de RSB atteintes par:

- a- les limites théoriques de Shannon et appliquées au cas bidimensionnel,
- b- le "Spiral Quantizer" : courbe "RSB"
- c- le "Spiral Quantizer" après application de la K-moyenne :
  - "KM12V" et "KM11V(7,8,9)" où "12" et "11" est l'exposant du nombre de vecteurs pour des spirales sans point à l'origine, "(7,8,9)" sont les exposants du nombre de points.
  - "Kmoy" et "Km/F" concernent les spirales avec point à l'origine.

d-le "Spiral Quantizer" après application de la K-moyenne et à partir de 128 points, 2048 vecteurs ont été utilisés . L'intérêt de ce cas est qu'il présente des performances très proches du "Spiral Quantizer" sans K-moyenne . Les tableaux 3.5.1 à 3.5.3 présentent les résultats chiffrés qui montrent au mieux cet avantage de notre quantificateur .

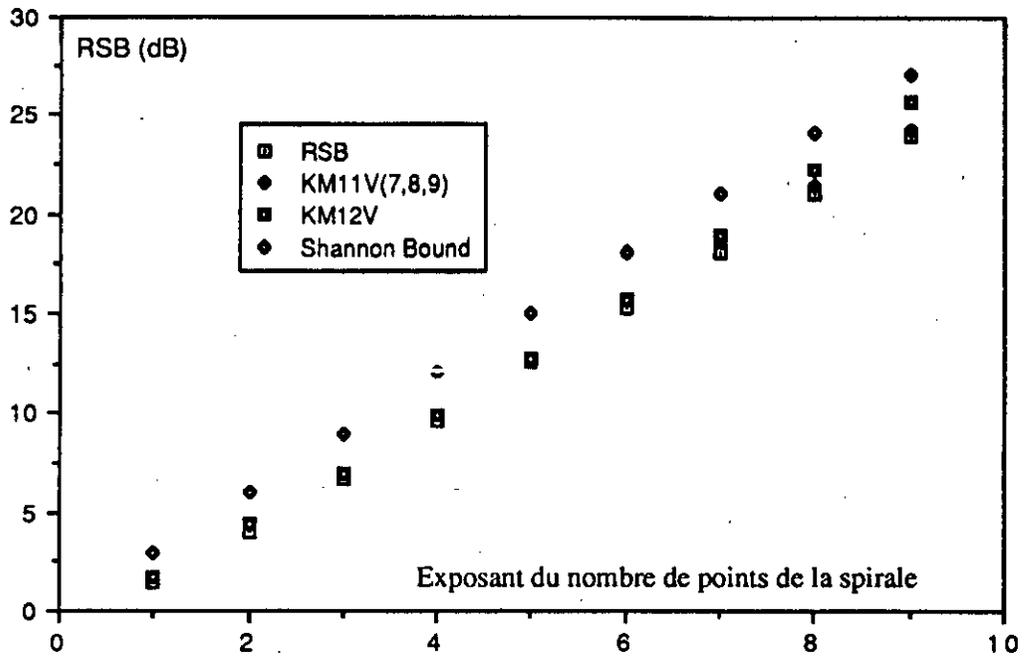


Figure 3.5.1 : Variation du RSB en fonction du nombre de points de la spirale pour un quantificateur de type "Spirale P" avec et sans application de l'algorithme de la K-moyenne.

La figure 3.5.2 illustre les propriétés analogues que la figure précédente dans le cas des spirales passant par l'origine avec un angle initial égal ou différent de zéro. Là encore on remarquera la supériorité du quantificateur avec  $\theta_0$  non nul avec toutefois une différence peu significative. Le tableau 3.5.3 donne les résultats détaillés de ces cas.

La figure 3.5.3 présente les mêmes cas mais en utilisant 2048 vecteurs pour des nombres de points supérieurs à 128. Les mêmes remarques sont à faire et le tableau 3.5.1 montre cet état de fait.

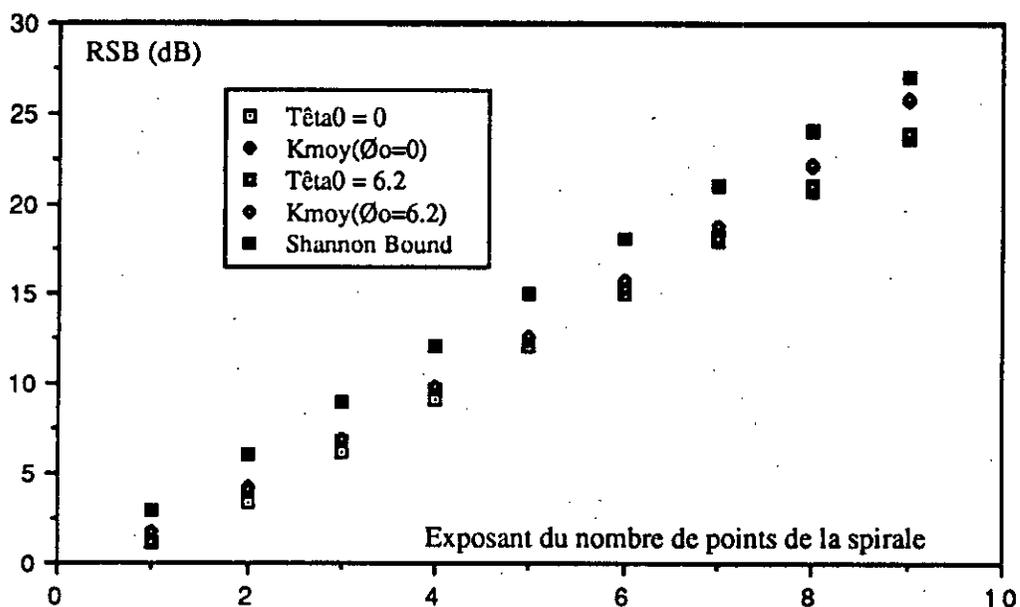


Figure 3.5.2 : Variation du RSB en fonction du nombre de points pour des spirales de types "Spirale Z0" et "Spirale ZT". Les courbes Kmoy désignent des spirales auxquelles a été appliqué l'algorithme de la K-moyenne.

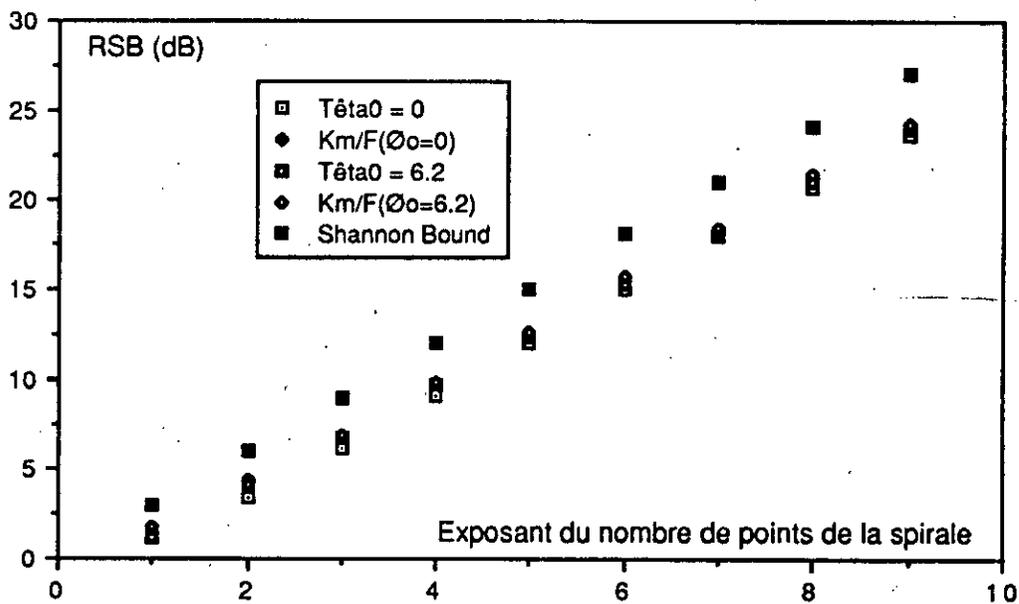


Figure 3.5.3 : Variation du RSB en fonction du nombre de points pour des spirales de types "Spirale Z0" et "Spirale ZT". Les courbes KmF désignent des spirales auxquelles a été appliqué l'algorithme de la K-moyenne avec un nombre de vecteurs différent.

les tableaux ci-dessous résument les résultats obtenus à partir de la discrétisation de la spirale d'Archimède.

Delta	Gain	n.Vect.	n.Pts	MSE	RSB (dB)
1.4	1.28	12	2	.770779	1.130702
1.4	1.28	10 KM	2	.667812	1.753461
1.4	1.28	11 KM	2	.667929	1.752698
1.4	1.28	12 KM	2	.667635	1.754608
4.08	1.8	12	4	.456996	3.400876
4.08	1.8	10 KM	4	.360407	4.432068
4.08	1.8	11 KM	4	.370263	4.314898
4.08	1.8	12 KM	4	.369326	4.3259
3.8	2.25	12	8	.24033	6.191926
3.8	2.25	10 KM	8	.202929	6.926565
3.8	2.25	11 KM	8	.202036	6.94572
3.8	2.25	12 KM	8	.202117	6.943962
3.8	2.25	14 KM	8	.202117	6.943962
1.4	2.569	12	16	.120867	9.176931
1.4	2.569	12 KM	16	.102696	9.884451
1.86	2.61	12	32	.063011	12.005861
1.86	2.61	12 KM	32	.055574	12.55126
5.3	2.88	12	64	.032157	14.927243
5.3	2.88	12 KM	64	.027758	15.56612
2.6	3	12	128	.016302	17.877582
2.6	3	11 KM	128	.014901	18.267774
2.6	3	12 KM	128	.013866	18.58054
1	3.1	12	256	.008403	20.755661
1	3.1	11 KM	256	.007323	21.35339
1	3.1	12 KM	256	.006289	22.013931
1	3.2	12	512	.004244	23.72262
1	3.2	11 KM	512	.003829	24.169183
1	3.2	12 KM	512	.002705	25.678234

Tableau 3.5.1: Tableau récapitulatif des performances du "Spiral Quantizer" avec angle initial nul.

Delta	Gain	n.Vect.	Têta0	n.Pts	MSE	RSB (dB)
1.4	1.27	12	1.39	2	.770773	1.130738
1.4	1.27	10 KM	1.39	2	.667812	1.753461
1.4	1.27	11 KM	1.39	2	.667918	1.752769
1.4	1.27	12 KM	1.39	2	.667642	1.754563
4.08	1.51	12	18.4	4	.404561	3.930159
4.08	1.51	10 KM	18.4	4	.378797	4.21593
4.08	1.51	11 KM	18.4	4	.382089	4.17835
4.08	1.51	12 KM	18.4	4	.387695	4.115099
4.08	1.51	14 KM	18.4	4	.387	4.115099
3.8	2.04	12	14	8	.213515	6.705708
3.8	2.04	10 KM	14	8	.207815	6.823236
3.8	2.04	11 KM	14	8	.205847	6.864564
3.8	2.04	12 KM	14	8	.206411	6.852668
3.8	2.04	14 KM	14	8	.206411	6.852668
1.4	2.4	12	6.2	16	.108357	9.651427
1.4	2.4	12 KM	6.2	16	.102135	9.908236
1.86	2.66	12	11.5	32	.057415	12.409783
1.86	2.66	12 KM	11.5	32	.053683	12.70161
5.3	2.82	12	43	64	.029441	15.31042
5.3	2.82	12 KM	43	64	.026934	15.69695
2.6	3.05	12	40.5	128	.015294	18.15469
2.6	3.05	11 KM	40.5	128	.01427	18.45571
2.6	3.05	12 KM	40.5	128	.012943	18.87969
1	3.19	12	25	256	.007876	21.03675
1	3.19	11 KM	25	256	.007102	21.4865
1	3.19	12 KM	25	256	.006115	22.136091
1	3.27	12	24.5	512	.004046	23.929891
1	3.27	10 KM	24.5	512	.006439	21.91149
1	3.27	11 KM	24.5	512	.003781	24.22378
1	3.27	12 KM	24.5	512	.002614	25.8266

Tableau 3.5.2: Tableau récapitulatif des performances du "Spiral Quantizer" de type "Spirale ZT".

Delta	Gain	n.Vect.	Têta0	n.Pts	MSE	RSB (dB)
3.3	1.2	12	1.39	2	.711723	1.476893
3.3	1.2	11 KM	1.39	2	.667171	1.757629
3.3	1.2	12 KM	1.39	2	.66725	1.757118
3.3	1.2	16 KM	1.39	2	.66725	1.757118
4.08	1.59	12	8	4	.400566	3.973265
4.08	1.59	12 KM	8	4	.357142	4.471595
3.8	2.02	12	10.2	8	.21189	6.73889
3.8	2.02	12 KM	10.2	8	.203338	6.917813
1.4	2.4	12	6.3	16	.108272	9.654821
1.4	2.4	12 KM	6.3	16	.102445	9.895082
2.41	2.64	12	13.87	32	.055842	12.530403
2.41	2.64	11 KM	13.87	32	.053756	12.69575
2.41	2.64	12 KM	13.87	32	.052667	12.78462
5.3	2.82	12	43	64	.029814	15.25584
5.3	2.82	12 KM	43	64	.027002	15.68611
2.6	3.06	12	29	128	.015467	18.10587
2.6	3.06	11 KM	29	128	.013951	18.55395
2.6	3.06	12 KM	29	128	.012883	18.899824
1	3.14	12	15.99	256	.007962	20.98967
1	3.14	11 KM	15.99	256	.007245	21.39985
1	3.14	12 KM	15.99	256	.006064	22.172134
1	3.25	12	19.5	512	.004062	23.91248
1	3.25	11 KM	19.5	512	.00377	24.23639
1	3.25	12 KM	19.5	512	.002649	25.76892

Tableau 3.5.3: Tableau récapitulatif des performances du "Spiral Quantizer" de type "Spirale P".

Les trois derniers tableaux que nous venons de présenter résument la majorité des résultats obtenus pour des spirales ayant un nombre de points ("n.pt") de 2 à 512 points .

La colonne "n.vect". designe l'exposant de "2" du nombre de vecteurs.

### 3.6- Résultats obtenus avec le "Set Partitioning"

#### 3.6.1- Présentation des constellations types

Nous présentons en premier lieu, des résultats obtenus, en utilisant les valeurs optimales des paramètres, pour des quantificateurs configurés en simples constellations de points. Chaque point du "Spiral Quantizer" désigne une valeur arrondie de sortie. Puis nous donnerons ceux des quantificateurs configurés en "set Partitioning" où chaque point de l'espace est lui-même un quantificateur à part entière.

Nous terminerons par des exemples de configurations de quantificateurs de dimension plus importante ( avec un nombre de points plus grand).

La figure 3.6.1 présente le "Spiral Quantizer" configuré en quantificateur simple avec une phase initiale non nulle et sans point à l'origine. Nous donnons ci-après son interprétation.

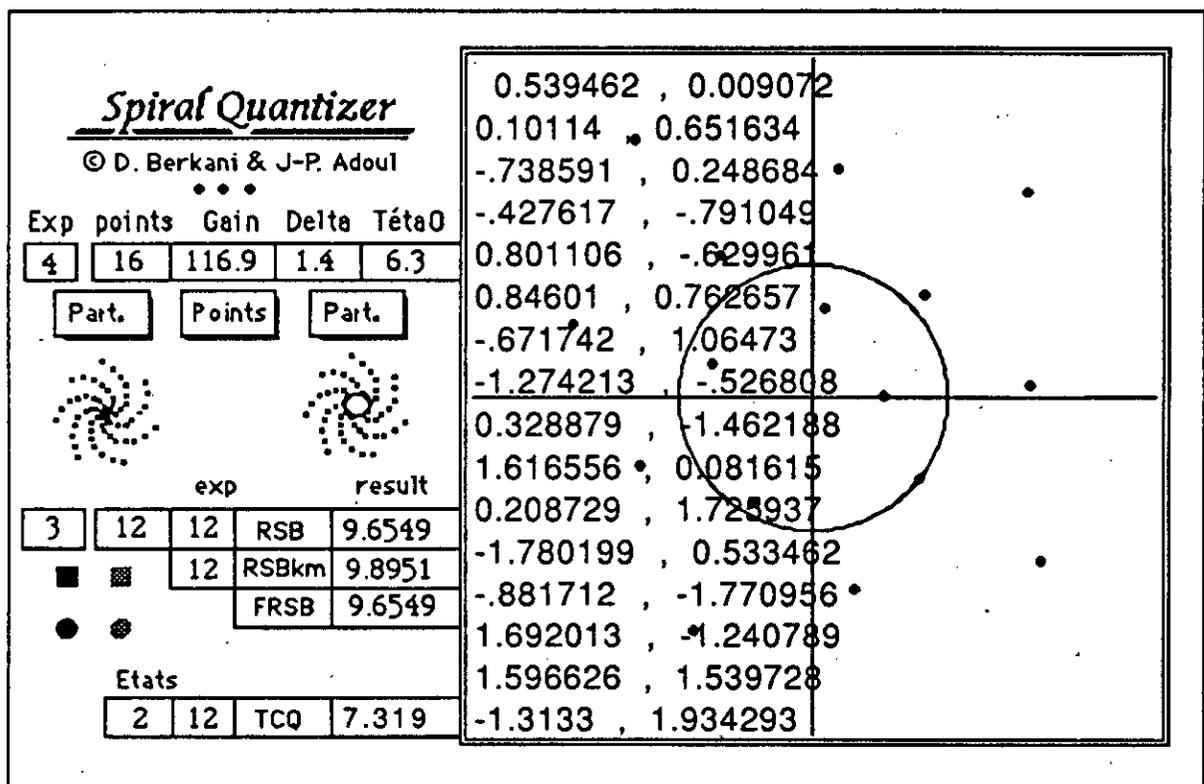


Figure 3.6.1: Quantificateur spiral simple avec phase initiale non nulle de type "Spirale P".

La figure 3.6.1 représente une carte ("carte-logiciel") type de l'environnement "HyperCard" qui a été conçue pour visualiser les constellations des valeurs arrondies de sortie fournies par le "Spiral Quantizer" (ou les quantificateurs élémentaires dans le cas du "Set Partitioning"). Elle permet aussi de piloter d'autres logiciels de calcul, d'effectuer des saisies de données et d'afficher les résultats.

La carte est composée d'un espace réservé au tracé, de champs de données nécessaires au tracé ou au calcul, de champs de résultats et de boutons de commande dont les scripts contiennent des programmes ou font appel à des logiciels extérieurs à l'environnement "HyperCard".

a- Les champs :

Les paramètres définissant la spirale sont indiqués par les champs "Gain", "Delta", "têta0" situés en haut de la carte.

Deux données sont encore nécessaires pour les calculs et tracés, ce sont :

- l'exposant ou la puissance de 2 nombre de points du quantificateur, noté "Exp",
- et le nombre de points de la spirale "points" qui est directement calculé par la carte dès la saisie.

On peut distinguer deux colonnes de champs :

- la colonne de données "exp" qui indiquent les exposants de 2 du nombre de vecteurs ( $2^{12}$  soient 4096 vecteurs),
- et la colonne de résultats "result" qui donne les performances en terme de RSB exprimés en dB pour chaque cas étudié.

Les chiffres "3" et "12" au dessus des quatre points symboliques concernent les conventions du script de ces points dans le cas du "Set Partitioning".

Le champ "Etats" désigne l'exposant de "2" du nombre d'états du treillis dans le cas du TCQ. La valeur maximale prévue de cet exposant est huit (8), ce qui correspond à un treillis de 256 états.

Les coordonnées des points du "Spiral Quantizer" ou des valeurs arrondies de sortie sont fournis par un champ contenu à l'intérieur de l'espace réservé au tracé.

b- Les boutons de commande :

Cinq boutons sont essentiellement destinés au tracé :

- Le bouton "Points" est un bouton de dessin, dans son script se trouve un programme qui lit les coordonnées des points ( ou les valeurs) contenu dans le champ réservé au tracé , il dessine ensuite le quantificateur correspondant. Ce bouton permet ainsi d'importer des quantificateurs quelconques , présentés sous la forme d'un fichier de points, à des fin de comparaisons par exemple ( voir chapitre 4).

- Les boutons notés "Part." pilotent des programmes de tracé dans le cas du "set Partitioning". Ils calculent également les coordonnées des points de sortie du "Spiral Quantizer" pour cette configuration, selon les variantes de spirales telles que déjà définies plus haut.

- Les deux petites spirales symboliques sont des boutons de commande derrière lesquels se trouvent des programmes de calcul des valeurs arrondies de sortie du "Spiral Quantizer". Ces boutons dessinent également les quantificateurs dont les points ont été calculés.

"RSB", "RSBkm", "FRSB" et "TCQ" sont des boutons concernent le calcul des performances des quantificateurs :

- "RSB" est un programme qui permet de déterminer la distorsion (MSE) et le RSB, mais seul le RSB est affiché dans l'environnement "HyperCard" tandis que "RSBkm" est une option de ce logiciel qui exécute l'algorithme de la K-Moyenne. "RSBkm" est un programme externe à la carte , mais il lui est fait appel à partir du bouton de commande correspondant. Il fournit une multitude d'autres résultats dont les nouveaux fichiers de points des valeurs approchées réajustées par l'algorithme de la K-Moyenne. L'intérêt de ces fichiers est qu'ils peuvent guider dans la recherche des meilleures valeurs et permettre par exemple une étude de la répartition.

Ces programmes ( "RSB" et "RSBkm" ) utilisent les fichiers de points fournis par les logiciels de tracé cités plus haut.

- "FRSB" est une version du programme "RSB" exécutable entièrement ( y compris le calcul des valeurs arrondies de sortie générées par le "Spiral Quantizer" ) dans un

environnement extérieur à la carte . "FRSB" nous permet de vérifier les résultats fournis par le programme "RSB".

- "TCQ" est un logiciel de calcul des performances du "Spiral Quantizer" lorsque ce dernier est configuré en "Set Partitioning".

Tous ces programmes utilisent les données fournies par les champs.

Les figures 3.6.2 , 3.6.3 , 3.6.4 , 3.6.5 présentent le cas d'un quantificateur simple et celui d'un quantificateur de type "set Partitioning" avec des phases initiales non nulles.

Les figures 3.6.6 et 3.6.7 présentent des quantificateurs à 16 valeurs arrondies de sortie avec phase initiale nulle et les figures 3.6.8 à 3.6.13 la discretisation des spirales de dimension plus importante ( 64 points ).

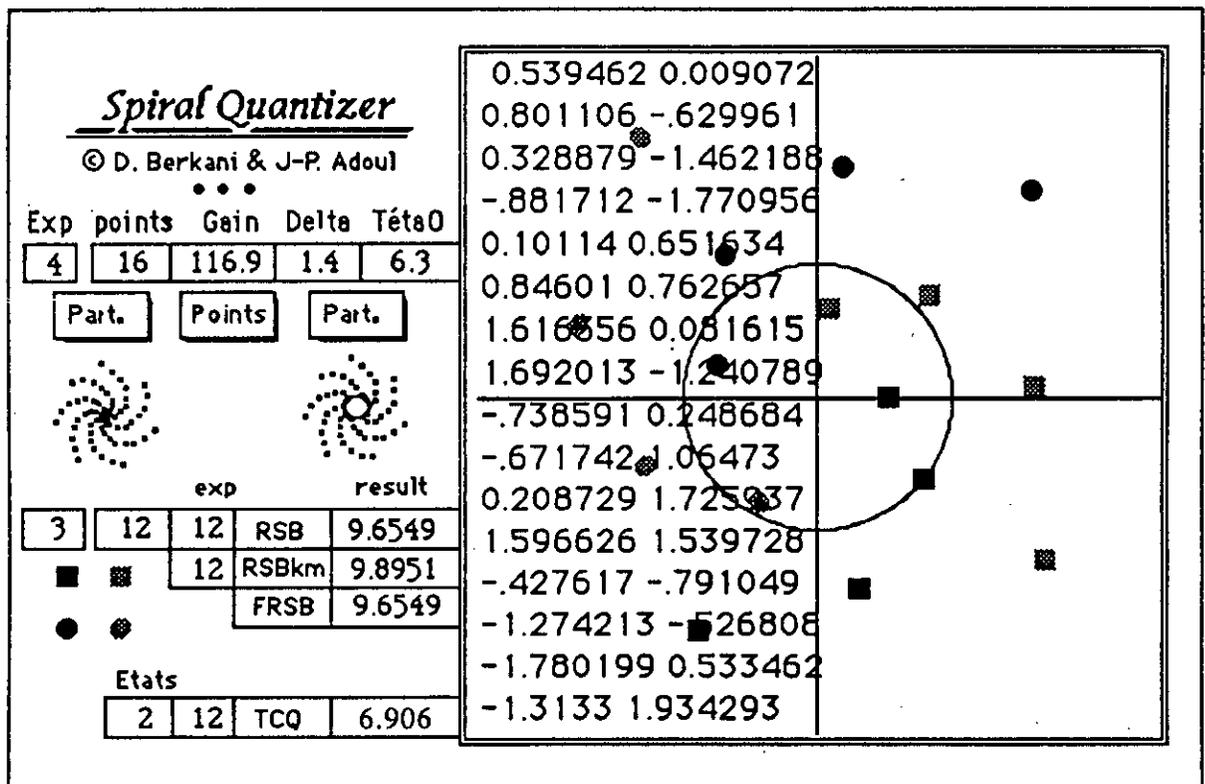


Figure 3.6.2: "Spiral Quantizer" configuré en "Set Partitioning", avec phase initiale non nulle de type "Spirale P".

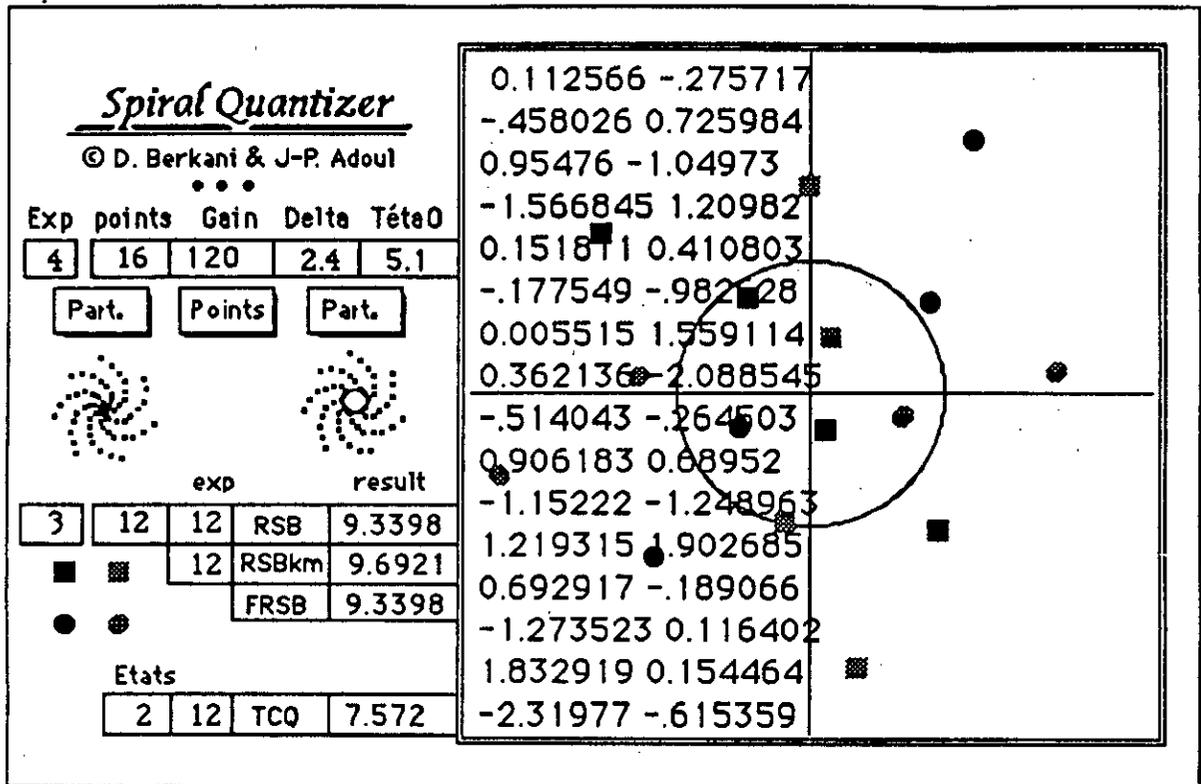


Figure 3.6.3 :Quantificateur précédant, optimisé pour le set Partitioning.

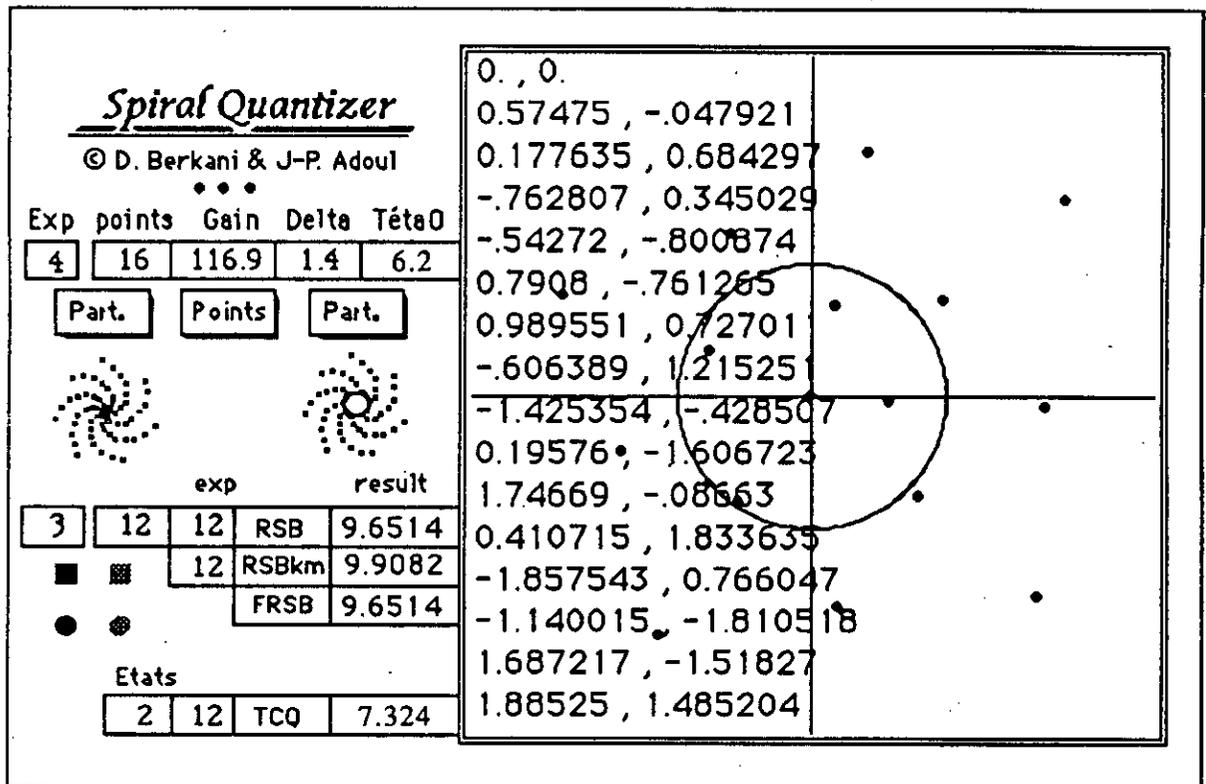


Figure 3.6.4 : "Spirale ZT" configuré en quantificateur simple (ensemble de points).

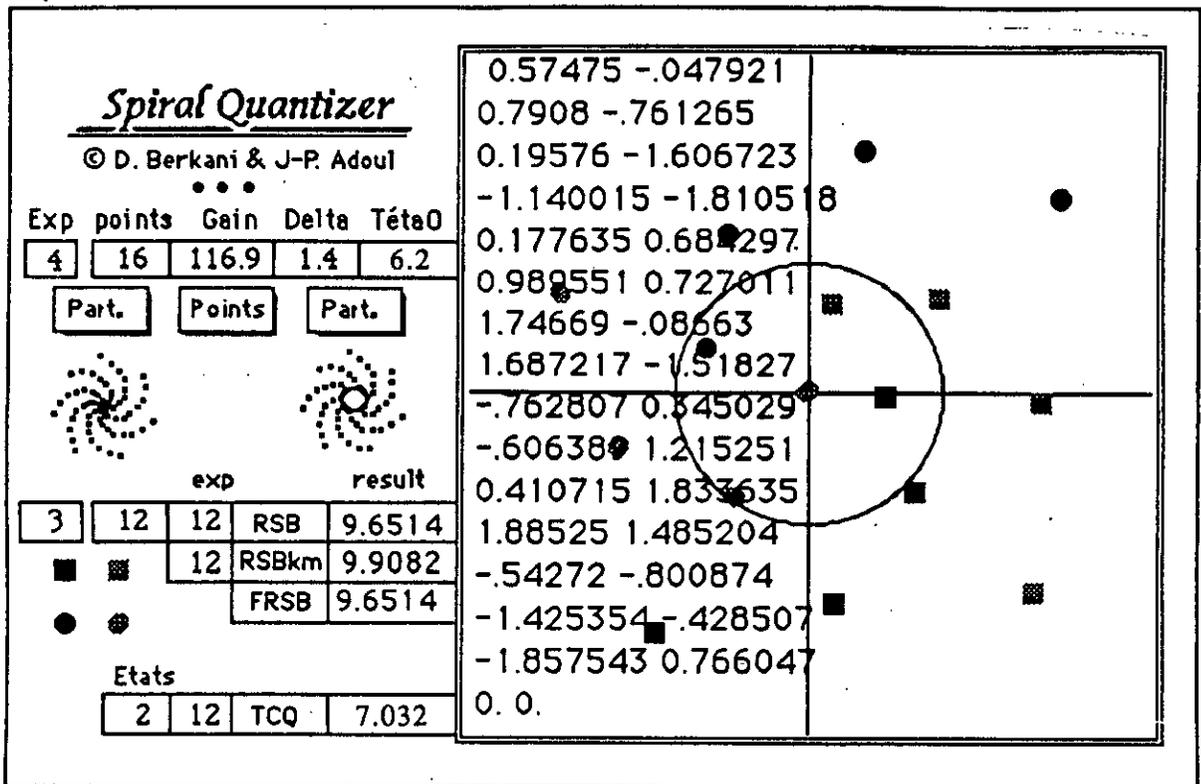


Figure 3.6.5: "Spiral Quantizer" de type "Spirale ZT" configuré en "Set Partitioning".

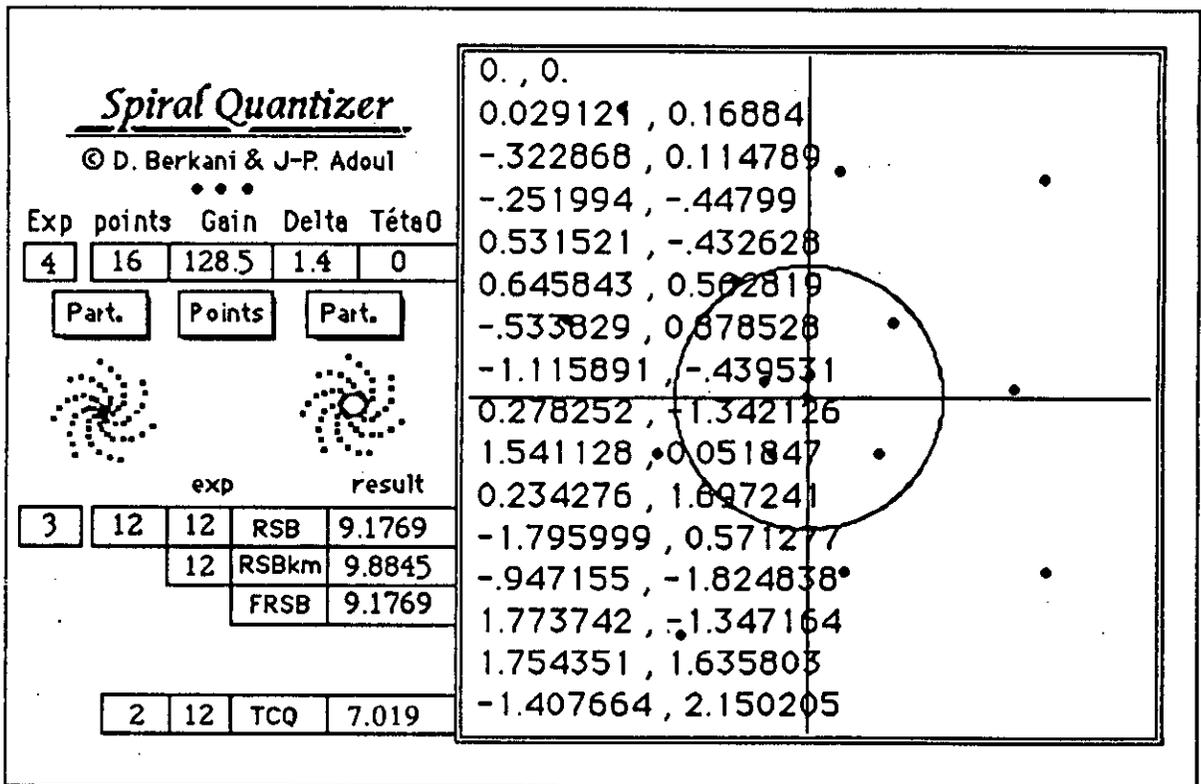


Figure 3.6.6: "Spiral Quantizer" en quantificateur simple avec phase initiale nulle.

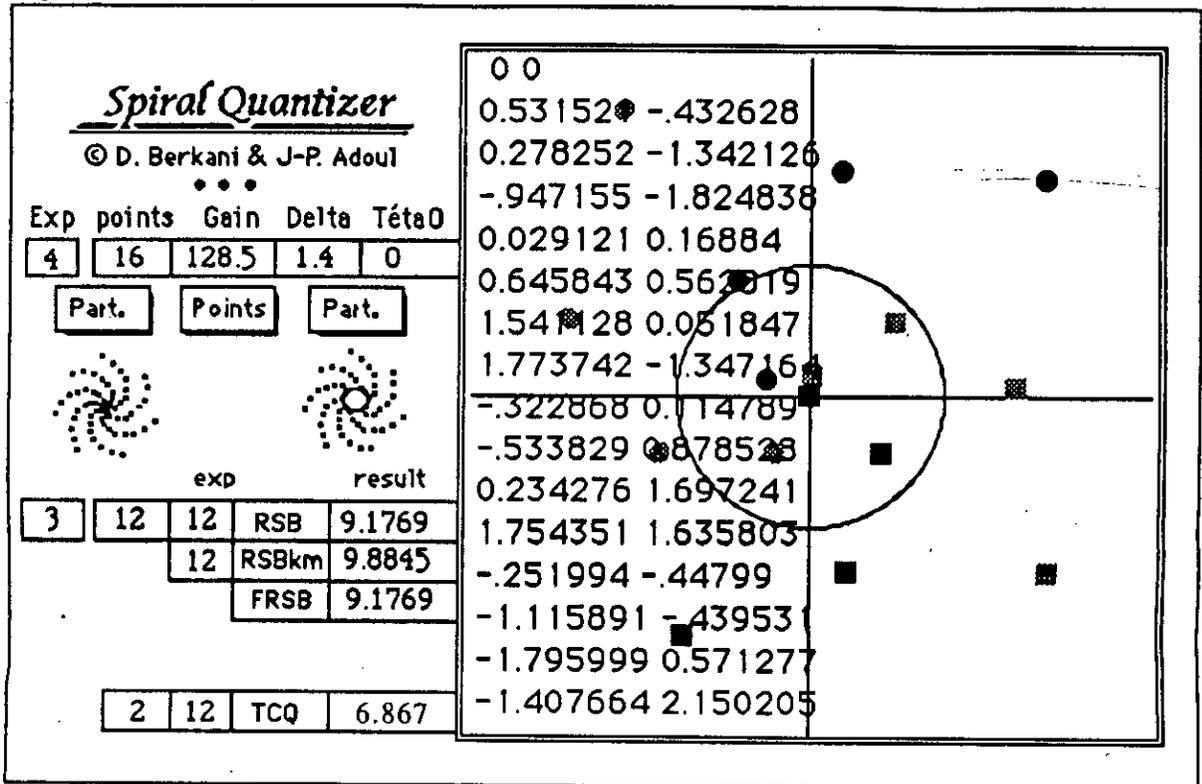


Figure 3.6.7: "Spiral Quantizer" de type "Spirale Z0" en "Set Partitioning".

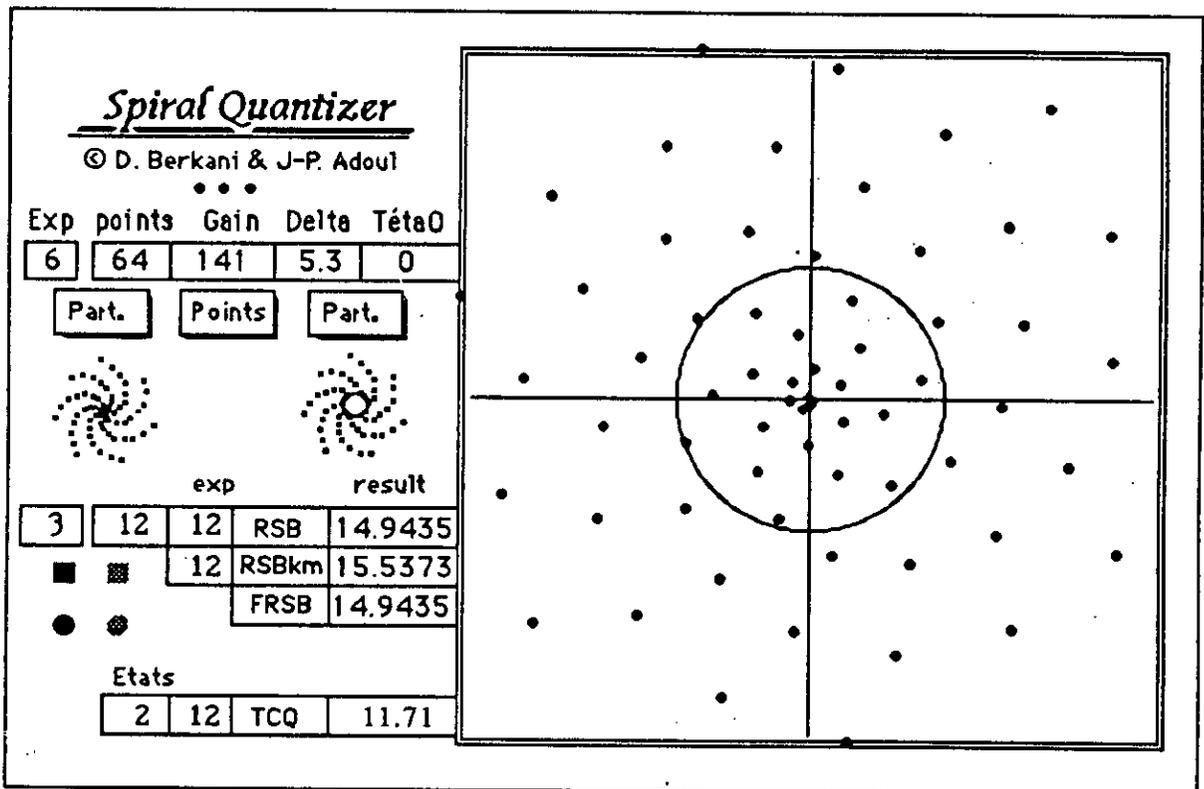


Figure 3.6.8 : "Spiral Quantizer" à 64 points , de type "Spirale Z0".

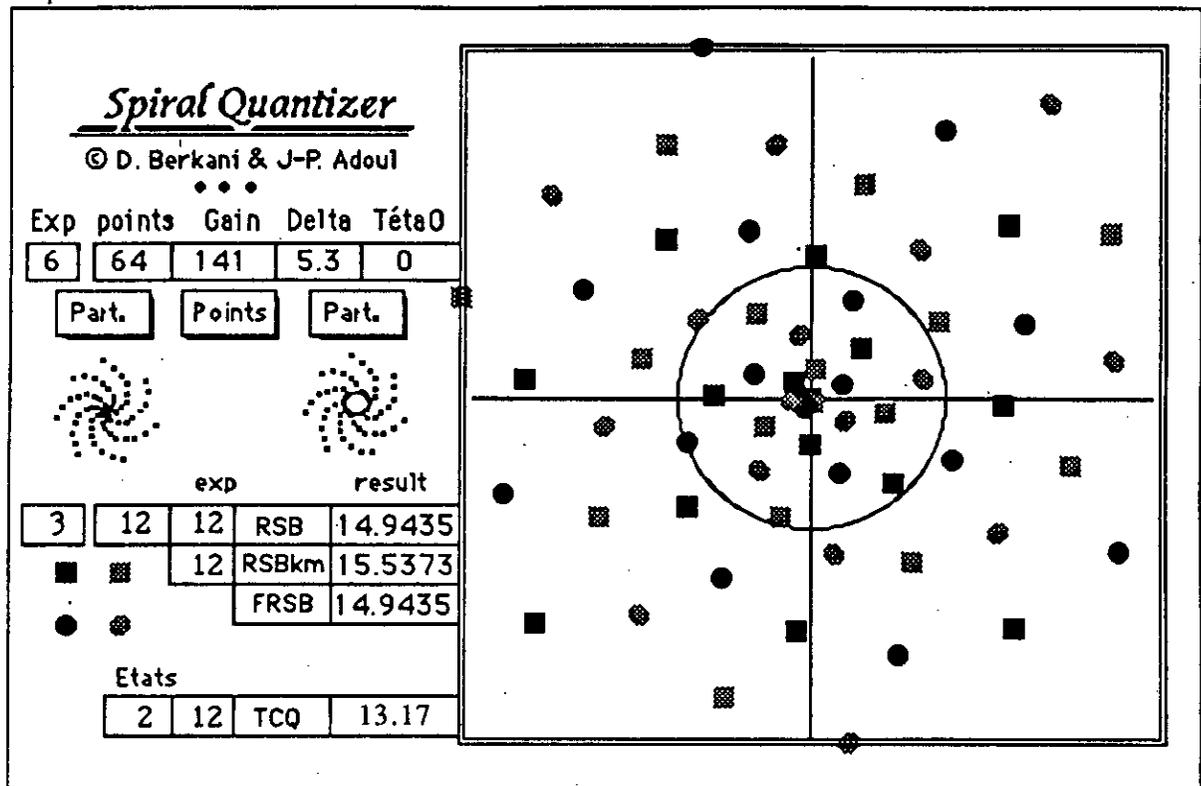


Figure 3.6.9 : Quantificateur précédent configuré en "Set Partitioning".

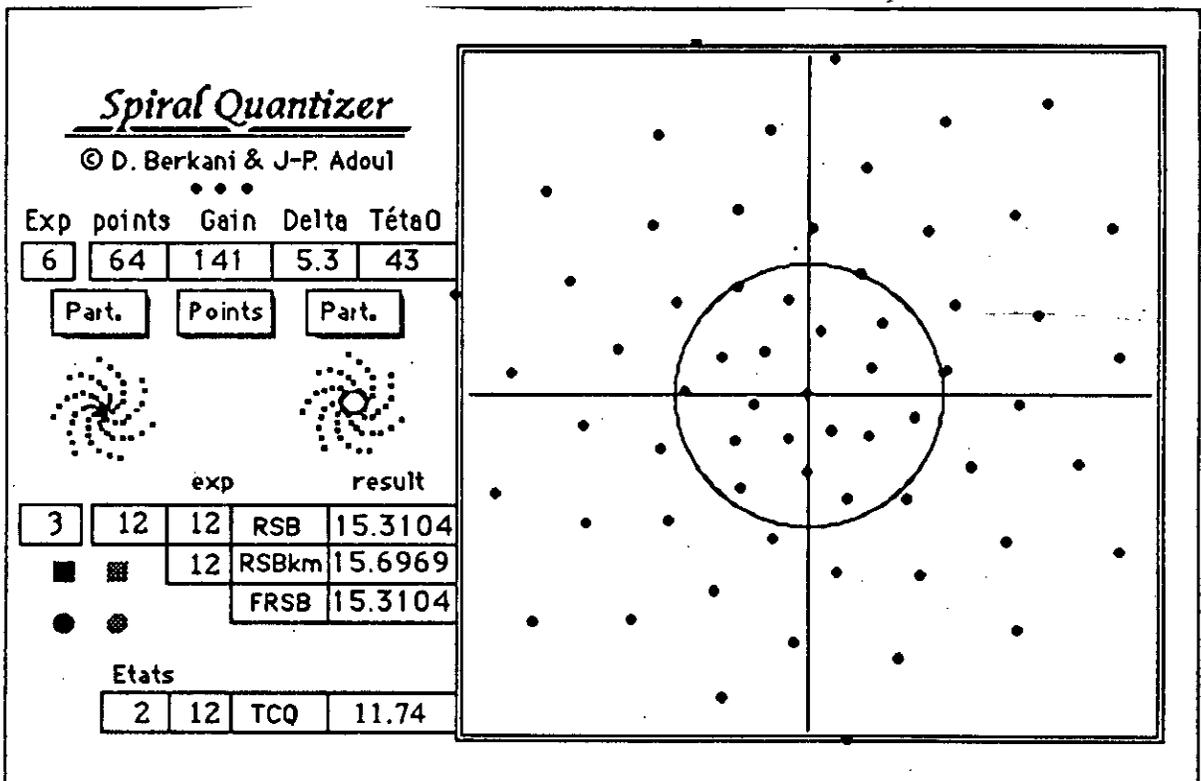


Figure 3.6.10: "Spiral Quantizer" de type "Spirale ZT" configuré en un ensemble de points.

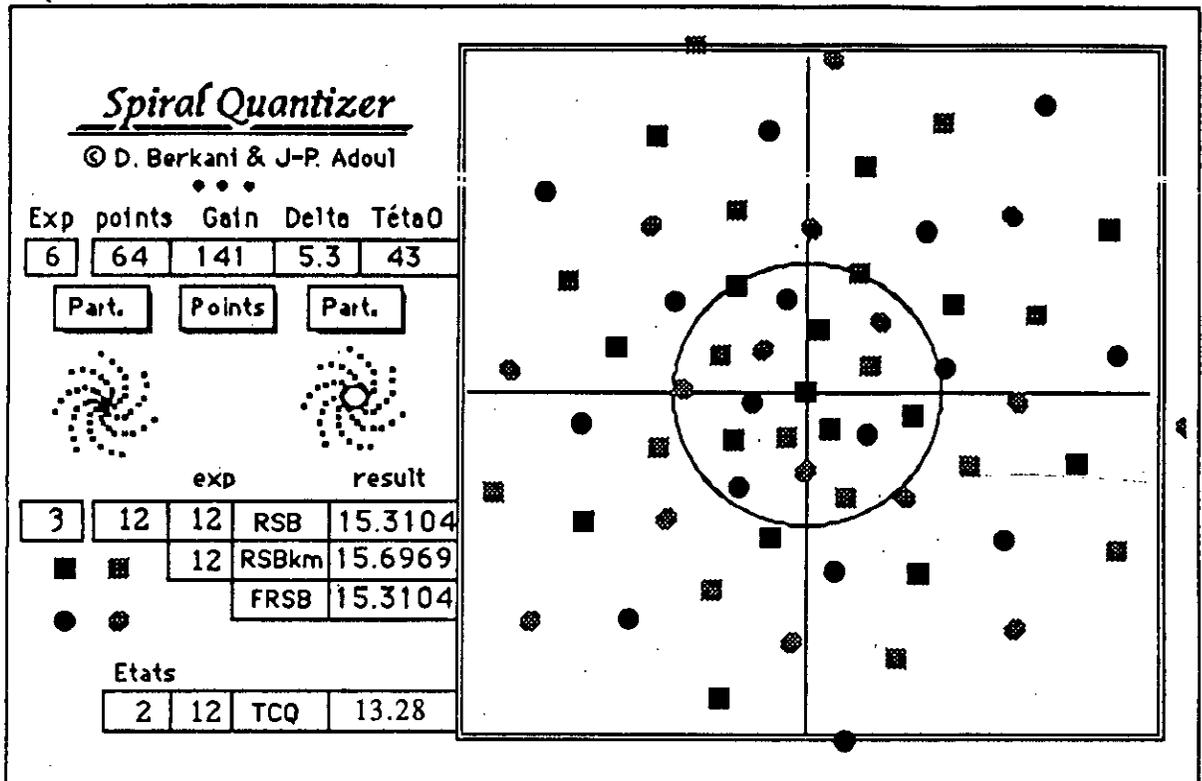


Figure 3.6.11 : "Spiral Quantizer" précédant configuré en "Set Partitioning".

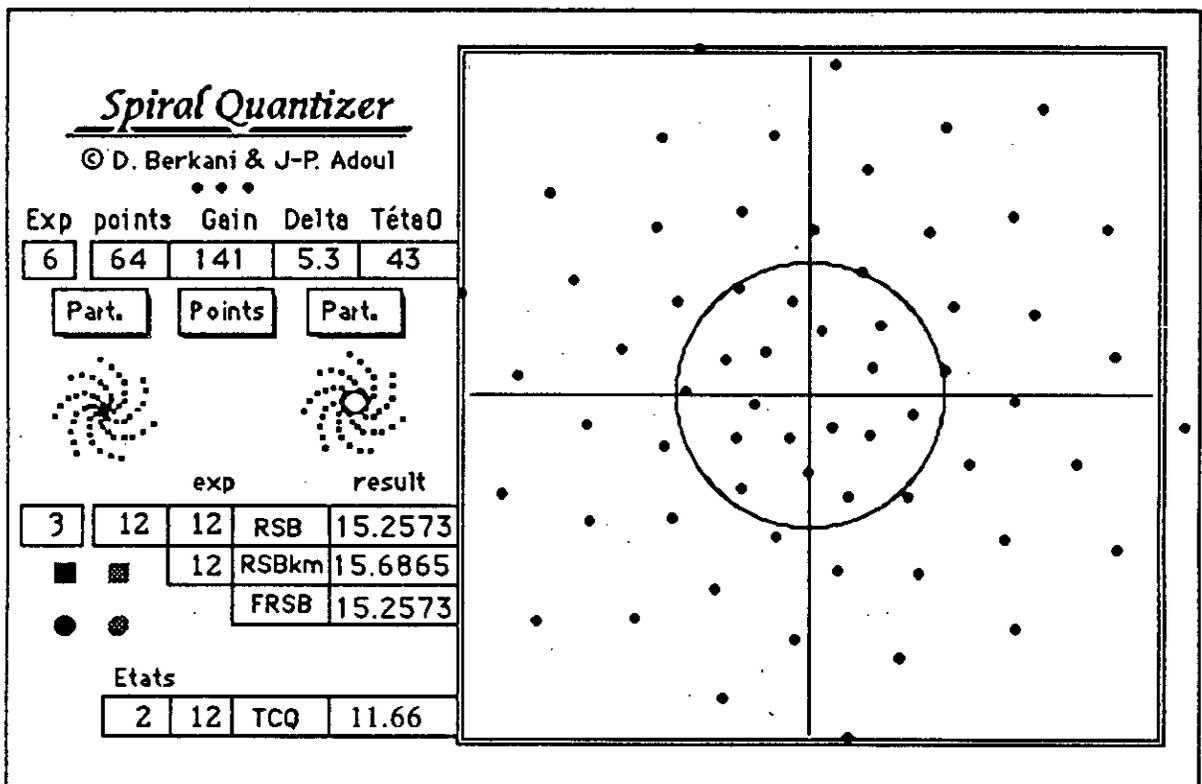


Figure 3.6.12 : "Spiral Quantizer" configuré en quantificateur simple, de type "Spirale P".

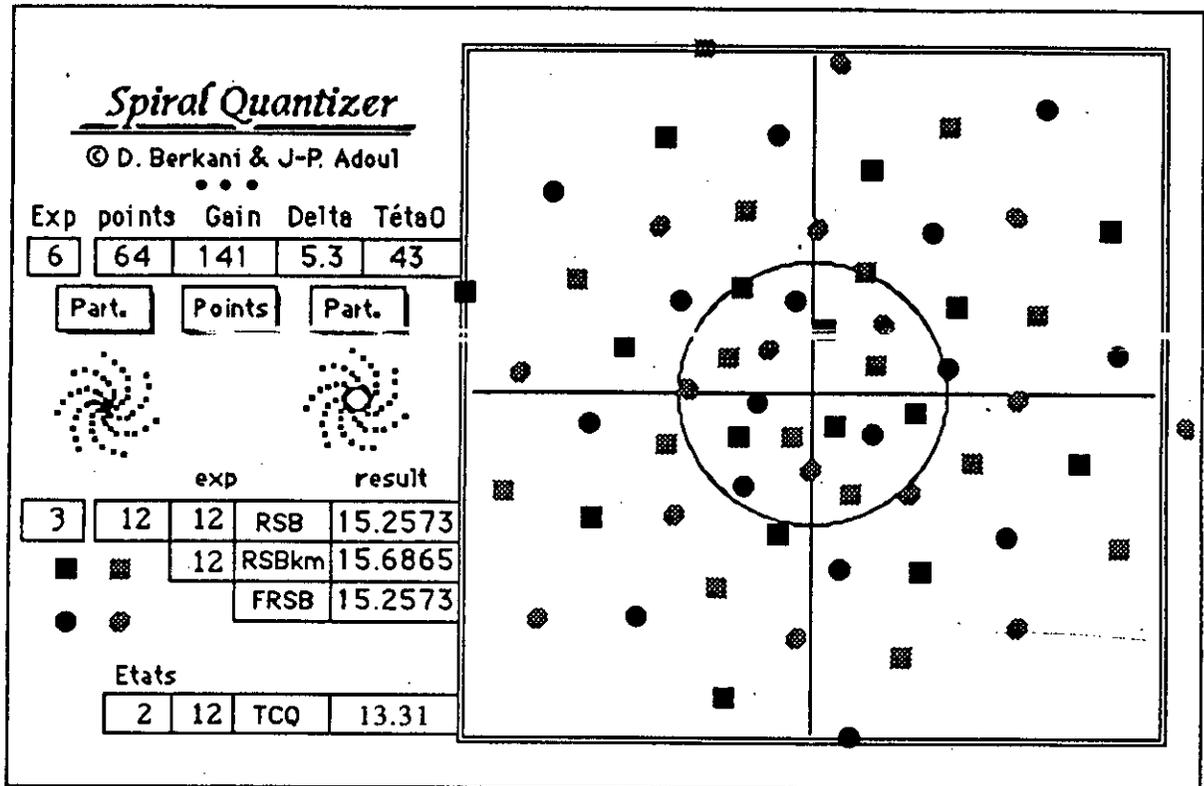


Figure 3.6.13 : "Spiral Quantizer" de type "Spirale P" sans quantificateur à l'origine configuré en "Set Partitioning".

### 3.6.2- Performances obtenues en "Set Partitioning".

La figure 3.6.14 montre comment varie le RSB d'un quantificateur configuré en "Set Partitioning" en fonction de l'accroissement "delta" pour un débit par échantillon donné. Cet exemple est typique pour les différentes spirales étudiées. Les résultats obtenus en "Set Partitioning" confirment ceux des spirales présentées sous forme de constellations de points au paragraphe 3.3.

La figure 3.6.15 donnent les performances en terme d'erreur quadratique moyenne des différentes variantes de spirales. Les essais effectués montrent que les trois types de spirale présentent des performances sensiblement égales.

Le tableau 3.6.1 fait ressortir la faible différence entre ces quantificateurs et permet ainsi de les particulariser.

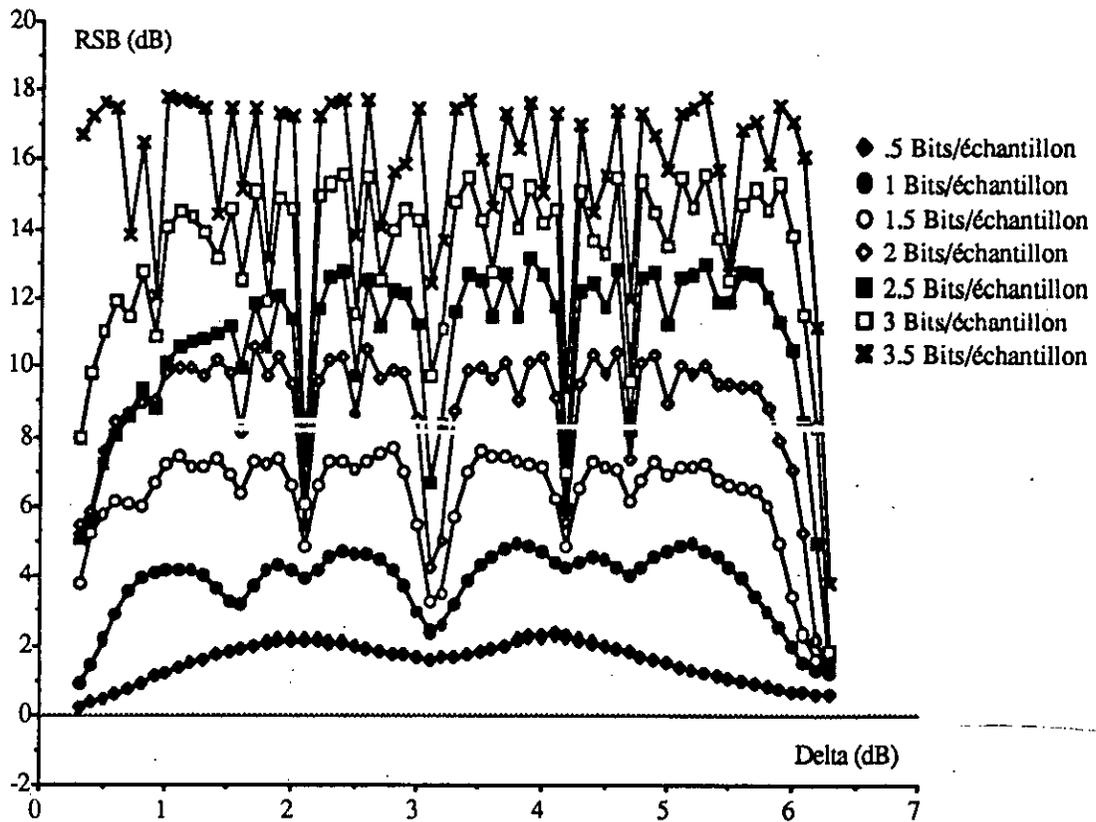


Figure 3.6.14 : Performances des Spirales avec  $\theta_0$  non nul et passant par l'origine.

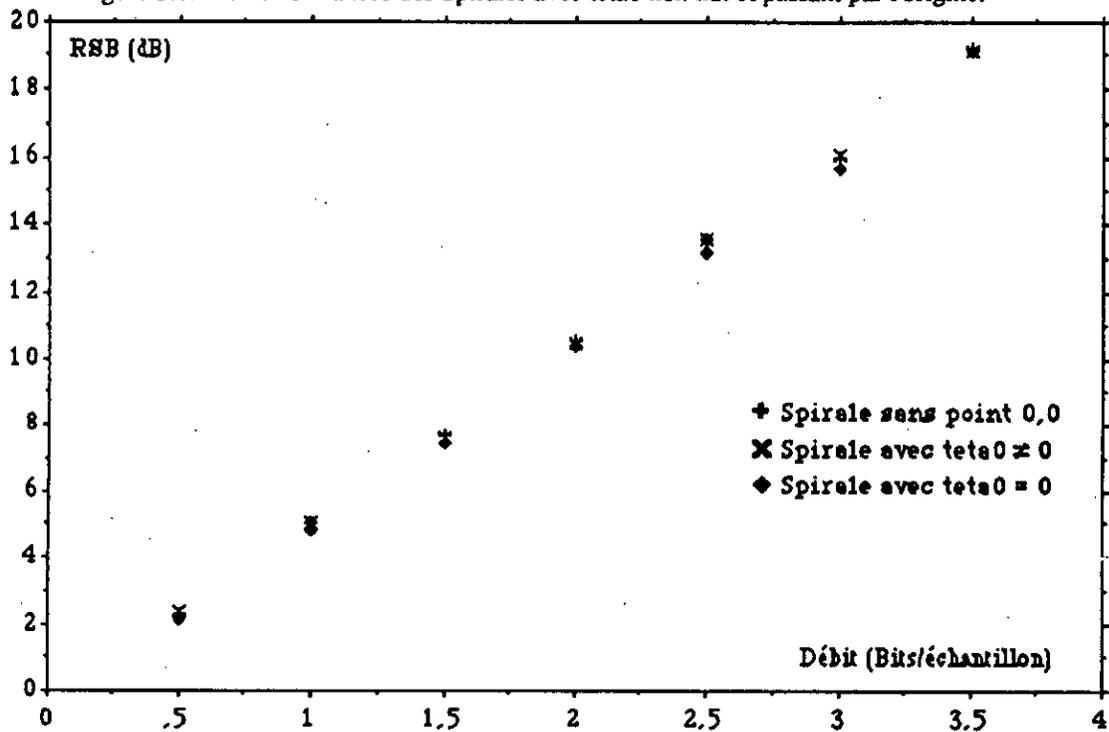


Figure 3.6.15 : Performances des configurations en Set Partitioning du "Spiral Quantizer".

Débit (Bits/échan.)	RSB (dB) pour des spirales avec:		
	Téta0 = 0	Téta0 ≠ 0	Sans point 0,0
.5	2.126	2.376	2.265
1	4.775	5.015	5.007
1.5	7.495	7.612	7.714
2	10.42	10.49	10.52
2.5	13.18	13.54	13.52
3	15.70	16.03	15.97
3.5	19.11	19.12	19.19

Tableau 3.6.1 : Résumé des performances des différentes variantes de quantificateurs.

**Conclusion:**

Dans ce chapitre, nous avons présenté les résultats de notre travail :

- dans une première étape, la méthodologie suivie pour discrétiser la spirale d'Archimède, l'optimisation de ses paramètres et les conditions d'essai de notre quantificateur,

- dans une seconde étape, les performances du "Spiral Quantizer".

On peut remarquer que les diverses configurations et catégories de spirales donnent des performances intéressantes puisque celles-ci sont proches des limites théoriques formulées par la théorie de la distorsion (de Shannon, voir chapitre 2 et 3).

En relâchant les contraintes en vue d'une application de l'algorithme itératif de la K-Moyenne , on remarque qu'il n'y a pas une amélioration sensible des performances en terme de RSB même pour des dimensions relativement élevées.

Le RSB se stabilise pratiquement à un nombre de vecteurs égal à 4096 pour toutes les spirales d'un nombre de points N égal à 16.

Concernant les ensembles de vecteurs générés par le "Spiral Quantizer", on peut dire que les quantificateurs offrant les meilleures performances sont à base de spirales ayant un

Chapitre 3

angle initial différent de zéro. Une légère différence existe entre les spirales sans point à l'origine et avec un point à l'origine .

L'interprétation de tous les résultats est présentée au chapitre suivant.

## Chapitre 4

### Interprétation des résultats

#### Introduction

Dans ce chapitre nous reprenons et expliquons les résultats des différents essais effectués sur notre quantificateur et nous interpréterons les performances obtenues.

Auparavant nous présentons une étude comparative du quantificateur à base de la spirale d'Archimède. Cette étude nous permettra également de prendre connaissance des performances réalisées par certaines constellations de points

Afin de mieux saisir l'importance des performances obtenues par le "Spiral Quantizer", il semble plus utile de revenir sur certaines techniques de quantification.

#### 4.1.- Quantification par réseaux réguliers

Nous avons déjà évoqué brièvement les questions afférentes à l'utilisation en quantification de réseaux réguliers de points, des codes correcteurs d'erreurs et des codes sphériques ( voir chapitre 2).

A partir du moment où l'on a établi d'après les travaux de Nyquist et Shannon que la conception d'un code pour un canal à bande limitée revient à un problème d'empilement de sphères dans un espace euclidien, les chercheurs se sont attelés à exploiter les travaux des mathématiciens sur l'empilement des sphères.

Ainsi DeBuda montre la voie en démontrant un théorème de codage d'un canal basé sur les réseaux réguliers [59]. Conway et Sloane, après avoir montré le lien entre la théorie des réseaux et la théorie de codage, ont constitué une véritable encyclopédie des réseaux réguliers [18]. L'application de ces techniques ne s'est pas fait attendre sur le marché des télécommunications.

Les réseaux les plus préférés possèdent des algorithmes rapides [60] de décodage à maximum de vraisemblance (Maximum Likelihood decoding ou ML decoding ). Forney a formulé le principe de leurs constructions [ 61, 62].

Cependant, en quantification, l'utilisation des réseaux réguliers avait stagné au stade des études théoriques. Gherso [15] a montré que pour des débits élevés , la quantification vectorielle donne une distorsion minimale et ce indépendamment de la source à quantifier. Il a été remarqué que les réseaux réguliers constituent de très bons quantificateurs vectoriels (codage de source) pour des variables aléatoires uniformément réparties dans l'espace.

Or il se trouve que les sources naturelles des signaux sont loin d'avoir une répartition uniforme. L'image et le son (ou la parole) [4], en tant que moyen de communication, en sont des exemples types qui intéressent beaucoup les chercheurs.

La proposition d'étendre les techniques de compresseur-expandeur du cas scalaire [63] à la quantification vectorielle a été faite par Gersho [15], Barnes et Sloane. Mais aucune méthode de conception de compresseur vectoriel rendant uniforme la distribution de la source n'a été mise au point.

Cet échec explique en partie le peu de succès en codage de source des réseaux réguliers. Adoul et al. [64] étudièrent la quantification algébrique sphérique et ont utilisé certains réseaux pour le codage d'une source gaussienne, puis la quantification vectorielle du résidu de parole et la modélisation du dictionnaire d'excitation du codeur CELP ( Code Excited Linear Prediction ) [22, 55].

Cependant Barnes et Sloane ont développé un quantificateur optimal en trois dimensions basé sur le réseau cubique à face centrée [65], mais la question essentielle demeure sans réponse : comment trouver la fonction "*compression*" et la fonction inverse ou réciproque "*expansion*" ?

La fonction de compression associe à tout vecteur donné  $x$  appartenant à l'espace  $R^k$  une fonction vectorielle  $f(x)$  appartenant à l'hypercube unité. Dans le modèle du quantificateur uniforme (optimal), les vecteurs arrondis doivent être distribués à l'intérieur de l'hypercube de façon uniforme.

Un quantificateur construit à base d'un réseau doit posséder un ensemble de

points régulièrement espacés ou une constellation uniforme [46]. Toute translation de son origine vers un autre point génère le même réseau. Ainsi les régions de Voronoï de tous les points du réseau considéré sont identiques. Ces régions remplissent tout l'espace sans aucun chevauchement.

Il est connu que les conditions d'optimalité d'un quantificateur exigent l'existence de partitions de Dirichlet et les vecteurs arrondis de sortie doivent être des centroïdes de ces partitions [46,66]. Mais avec l'application d'un opérateur ou fonction de compression  $f(x)$ , il est généré lors de l'application de l'opérateur inverse une nouvelle partition qui n'est pas nécessairement une partition de Dirichlet. Pour préserver la partition de Dirichlet, donc l'optimalité du quantificateur il faut que l'opérateur  $f(x)$  soit conforme, ce qui est loin d'être une opération triviale ; c'est pour cela que la conception d'un tel quantificateur reste un art en soit. Cette difficulté constitue une véritable restriction en vue d'une généralisation à l'espace multidimensionnel, de la méthode de compression-expansion proposée par Bennet [63].

La technique de codage d'une source gaussienne LPQ (Lattice Phase Quantization) proposée pour la première fois par Adoul [67] consiste en fait à utiliser la transformation de coordonnées cartésiennes en coordonnées polaires comme des opérateurs de compression et d'expansion. Le module et la phase sont codés séparément. En effet, comme nous l'avons déjà vu, cette transformation aboutit à une répartition uniforme de la phase et une distribution de Rayleigh pour l'amplitude. L'importance est dans le fait que la phase est distribuée de manière uniforme, ce qui permet d'exploiter les réseaux réguliers.

Cette propriété a été exploitée à bas débit, pour des réseaux donnés en réalisant une transformation non linéaire sur la distribution de Rayleigh de l'amplitude en vue de la rendre uniforme [17], un exemple de réalisation ainsi que tout le développement ont été présentés dans ce travail de recherche.

Notons enfin en ce qui concerne le "Spiral Quantizer" qu'il est en fait basé sur structure régulière, c'est à dire sur une règle simple, capable de générer l'ensemble des vecteurs. Aussi, une forme de quantification algébrique pourrait être envisagée avec certains avantages [7] et que pourrait éventuellement exploiter, de manière systématique notre quantificateur. Ce qui rendrait l'utilisation du "Spiral Quantizer" fort intéressante.

#### 4.2.- Etude Comparative du "Spiral Quantizer":

L'étude comparative du quantificateur en spirale que nous présentons s'est limitée aux quantificateurs possédant 16 valeurs arrondies de sortie . Quelques cas de quantificateurs à 8 points ont été également vus au cours de cette étude et nous avons noté des remarques similaires que pour les quantificateurs précédents. En effet, nous disposons uniquement de ces types de constellations en ce qui concerne les quantificateurs autres que celui se basant sur la spirale d'Archimède.

Ainsi, nous avons repris l'étude de ces différentes constellations avec deux types de programmes de calcul du RSB : celui avec lequel elles ont été conçues la première fois et celui qui est utilisé constamment pour le "Spiral Quantizer".

On a remarqué une légère différence pour ce qui est des performances qui ne peut s'expliquer que par le fait que les sources de vecteurs des deux logiciels ne sont pas strictement identiques.

Les constellations que nous présentons à des fins de comparaisons sont en majorité les meilleures de part leurs performances, dans leur catégorie. Ainsi, par exemple pour les constellation sphériques, celle de rayon égale à l'unité s'avère être la meilleure en terme de RSB.

Les figures 4.1 à 4.5 représentent les constellations types et leurs performances avec et sans application de l'algorithme de la K-Moyenne. Les figures 4.6, 4.7 et 4.8 donnent les représentations de quelques quantificateurs après une application de la K-Moyenne.

Ces représentations ne sont en fait qu'une partie des investigations qui ont été faites sur plusieurs constellations qui ont attiré l'attention des chercheurs. L'étude comparative consiste dans une première étape en une mise en condition d'expérimentation identique des différents quantificateurs par rapport au "Spiral Quantizer".

L'analyse des résultats obtenus montre que la majorité des constellations dont certaines n'ont pas été représentées ici, présentent des performances inférieures à celles du quantificateur à base de la spirale d'Archimède. De plus, on peut affirmer que leurs performances, après une application de l'algorithme de la K-Moyenne sont du même ordre, sinon encore inférieures à celle du "Spiral Quantizer".

Cependant, cette étude demeure restrictive dans le sens où elle ne concerne essentiellement que les constellations composées de 16 points . Certains quantificateurs ne peuvent évoluer effectivement vers des structures à nombre de points différent de celui qui a été considéré initialement ; par exemple, un quantificateur donné peut présenter des performances données avec une structure à 8 points mais ne peut posséder de constellation de points "équivalente" de taille différente. De ce fait, on ne peut étendre cette étude.

Malgré ce qui a été dit, l'étude comparative permet de situer les performances du Spiral Quantizer par rapport à quelques structures particulières.

Enfin, l'étude comparative complète les résultats obtenus au chapitre précédant où les performances ont été comparées aux bornes données par la théorie de la distorsion et énoncées par Shannon. Ces résultats montrent que le "Spiral Quantizer" présentent des performances proches de celle du quantificateur optimal.

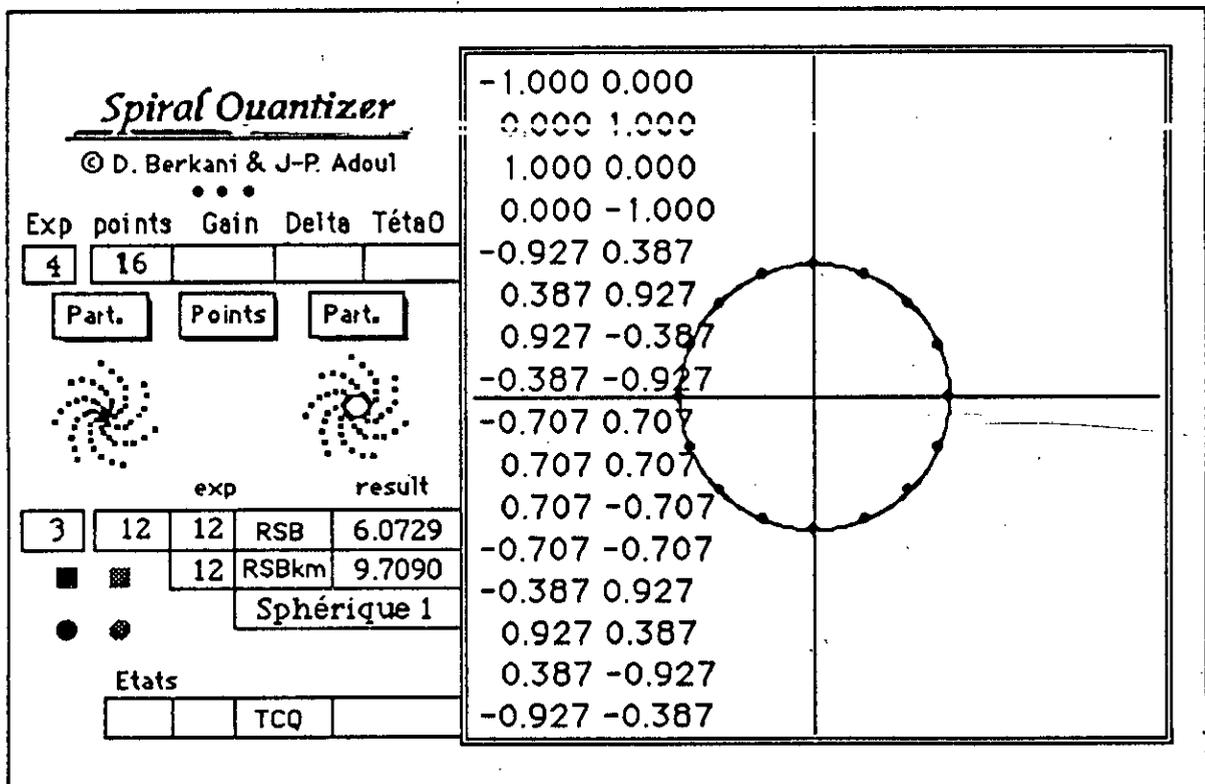


Figure 4.1: Constellation sphérique de rayon égal à l'unité.

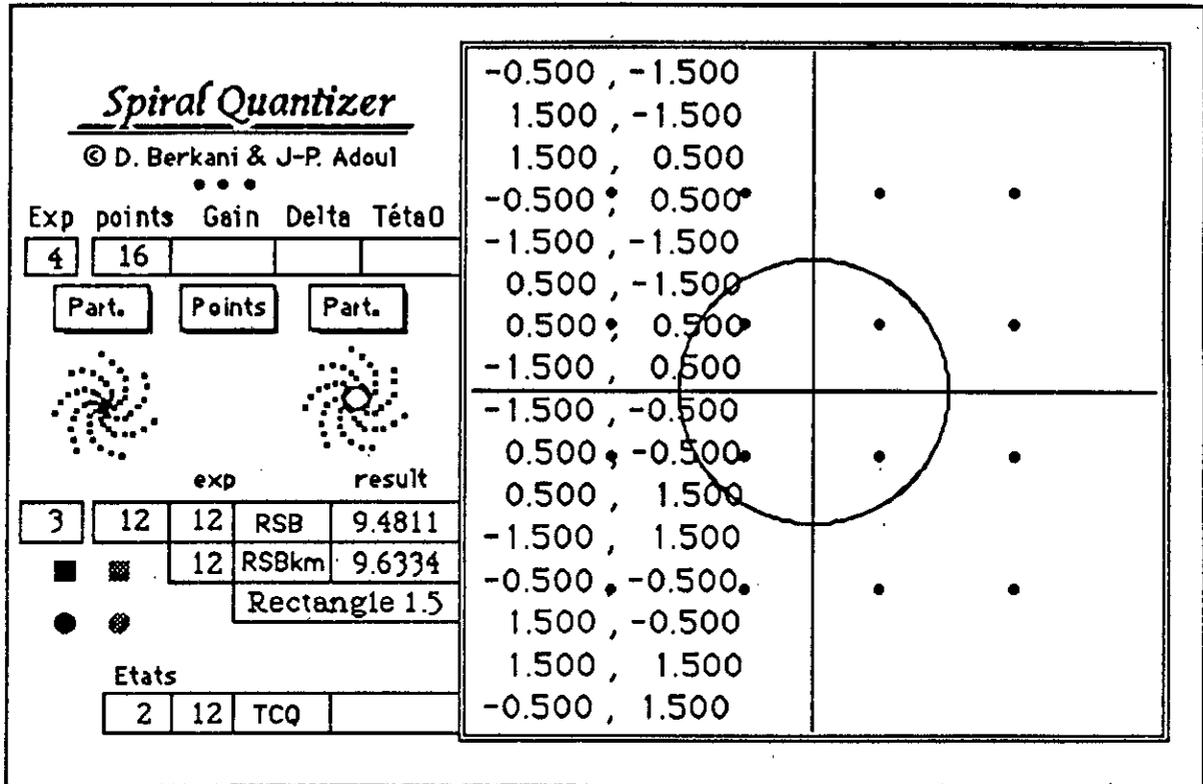


Figure 4.2: Constellation rectangulaire.

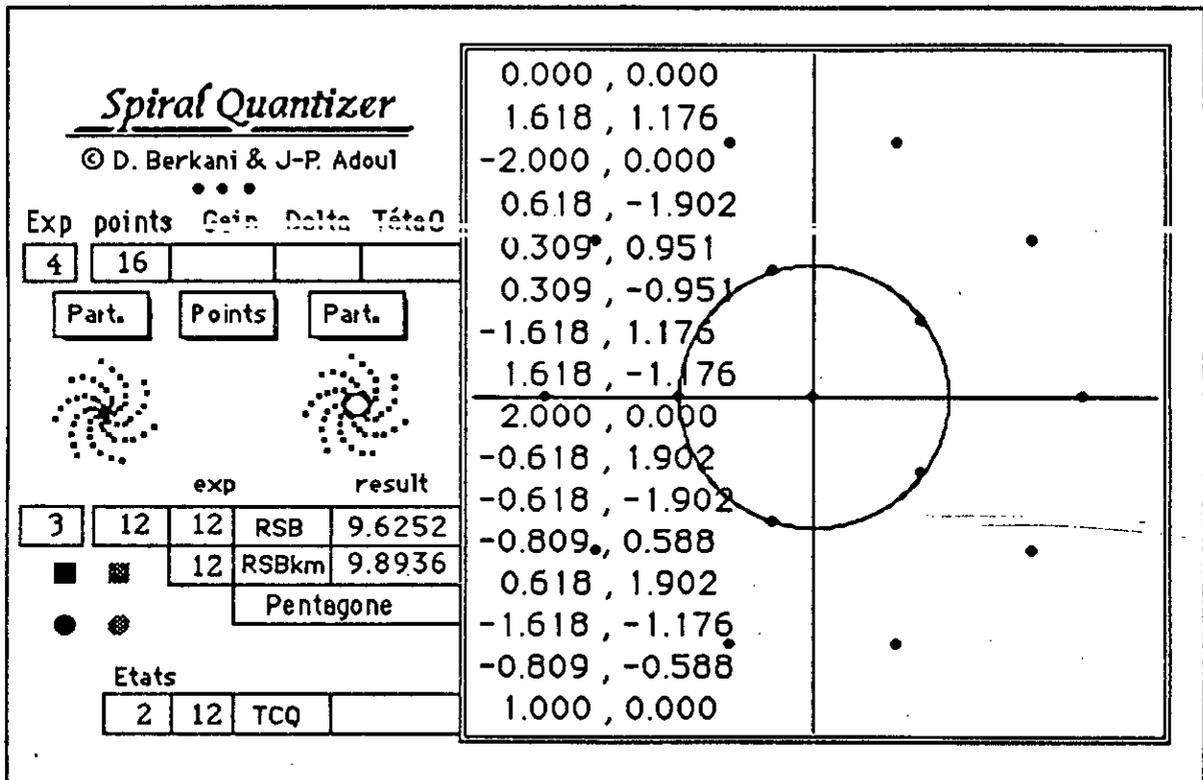


Figure 4.3: Constellation "Pentagone".

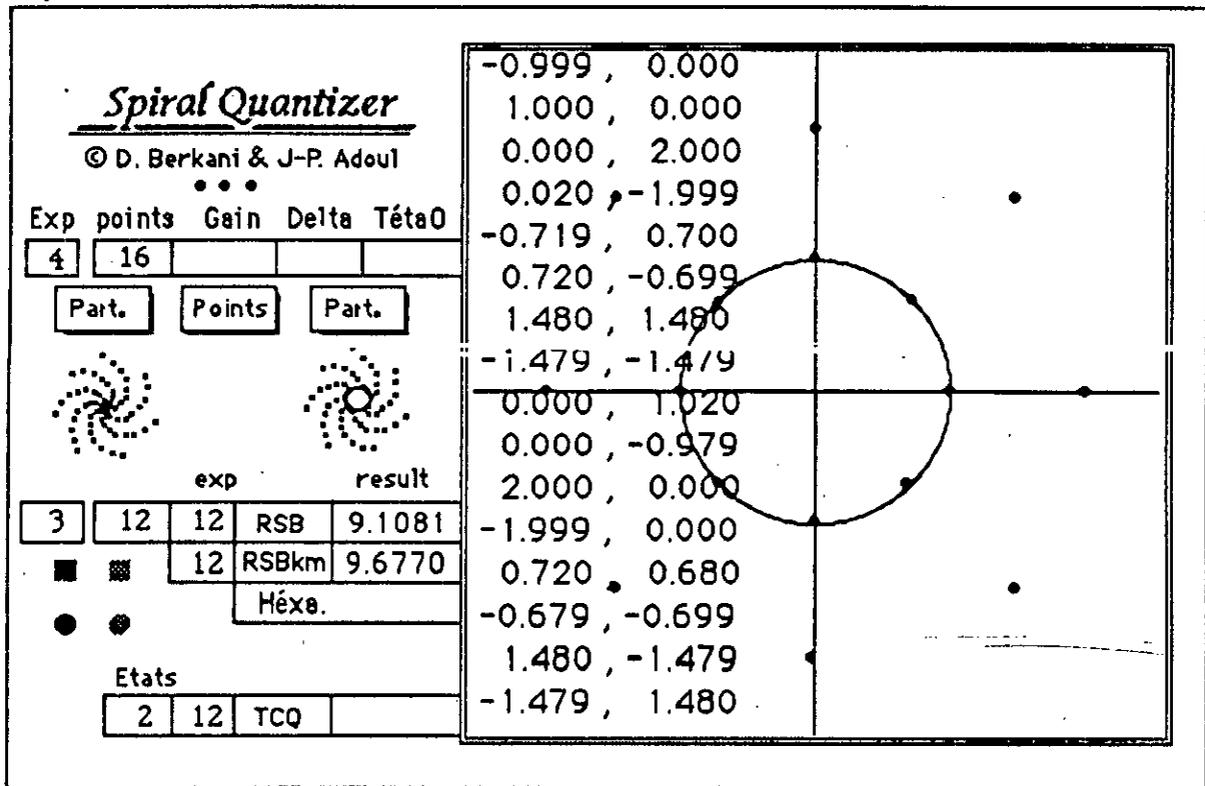


Figure 4.4: Constellation "Héxa".

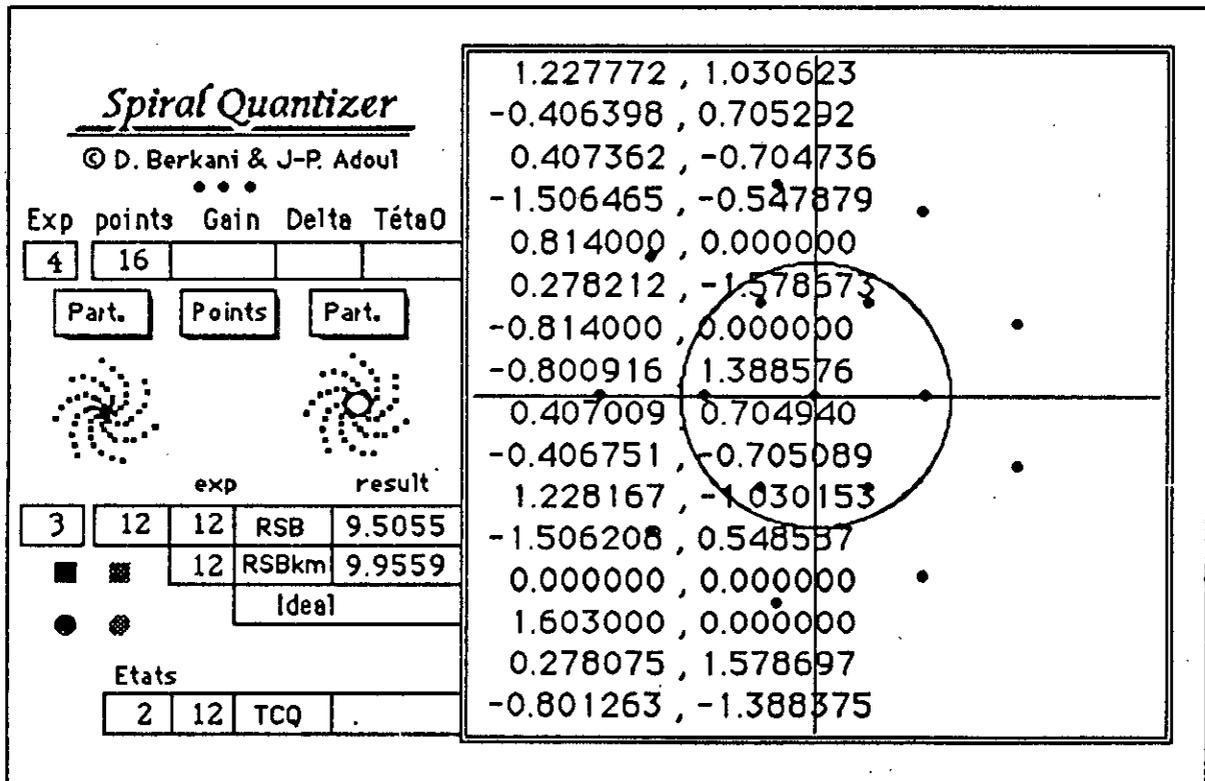


Figure 4.5: Constellation "Idéal".

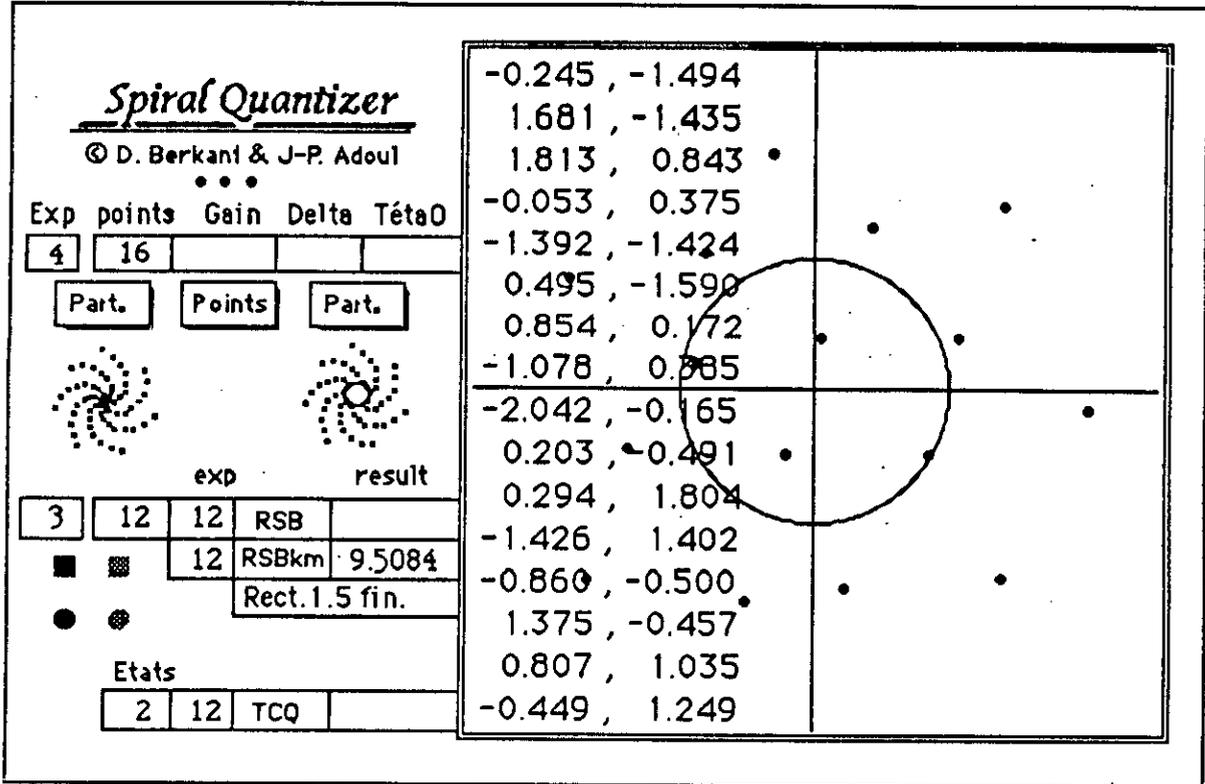


Figure 4.6: Constellation "rectangulaire" avec application de l'algorithme de la K-Moyenne.

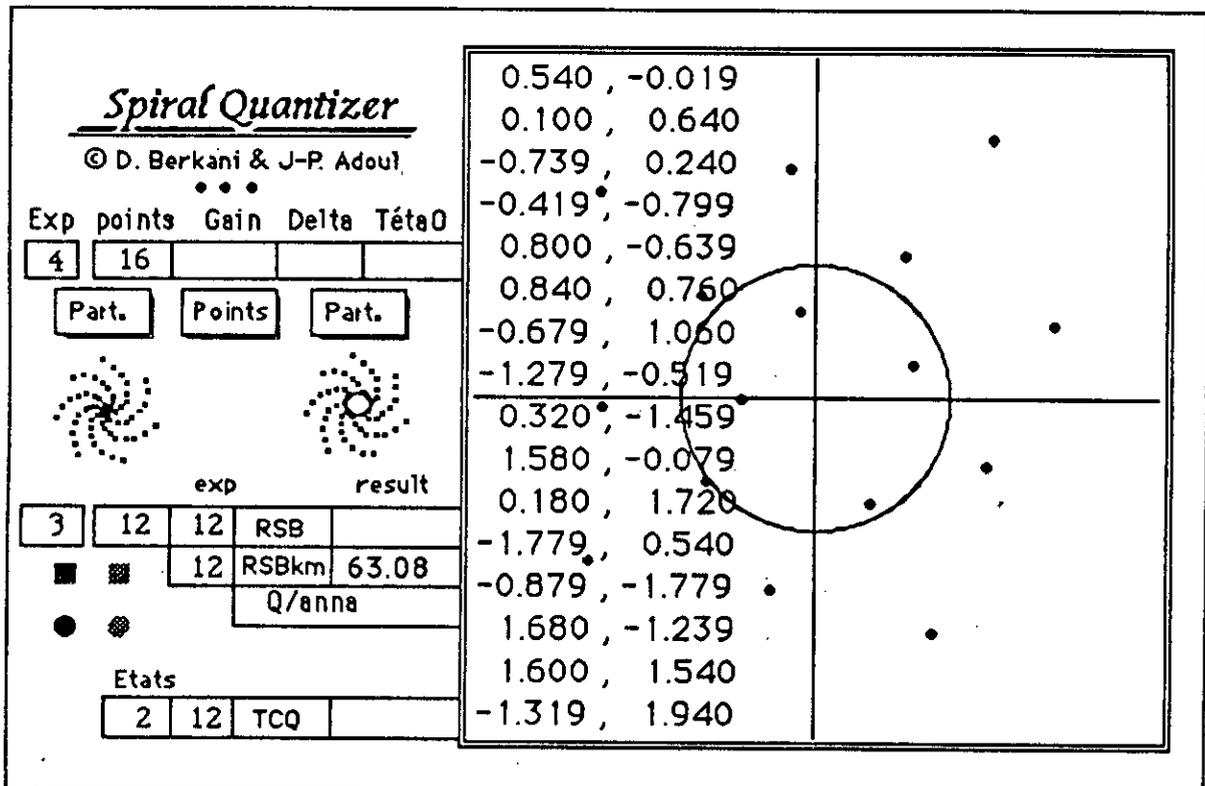


Figure 4.7: Constellation "Q/anna".

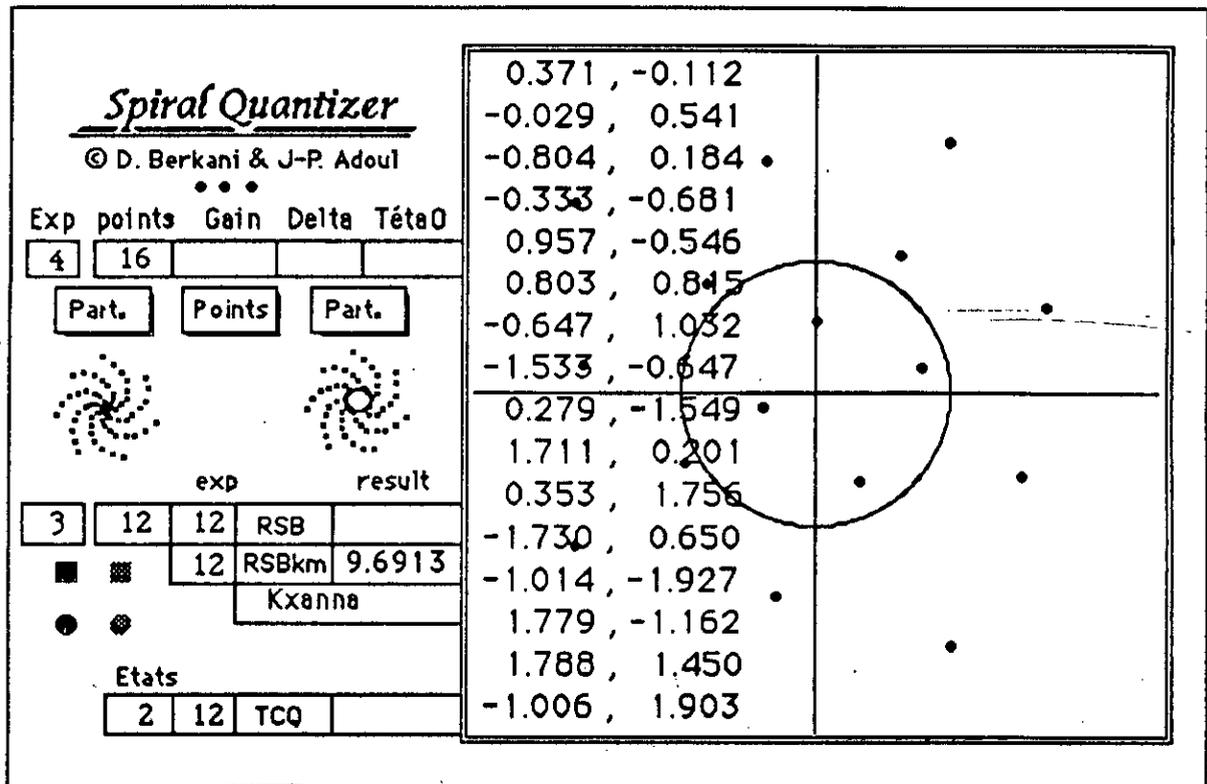


Figure 4.8: Constellation " Kxanna".

### 4.3- Commentaires sur les résultats obtenus.

#### 4.3.1- La discrétisation de la spirale.

La manière dont varie le RSB en fonction du pas d'incrément de l'angle "delta" pour des valeurs données du facteur d'échelle "Gain" a beaucoup simplifié la recherche des valeurs optimales des paramètres de la spirale. En effet, les valeurs de "delta" offrant les meilleures performances restent proches les unes des autres pour des valeurs de "Gain" différentes comme le montre la figure 3.3.1 .

Par ailleurs la "quasi-périodicité" des variations du RSB en fonction de l'incrément de l'angle "delta" permet de limiter la plage de variation de ce dernier (entre 0 et  $2\pi$ ) pendant la recherche pour une meilleure discrétisation de la spirale.

On constate que les variations des performances en fonction de l'angle initial "téta0"

pour un "Gain" constant sont plus complexes. Nous avons été amenés à faire varier l'angle initial dans de très larges plages afin d'éviter toute surprise. Nous avons vérifié les résultats en visualisant à l'aide des programmes de tracé comment évolue la répartition des points du "Spiral Quantizer". Il a été constaté qu'avec l'accroissement monotone de cet angle les points de la spirale s'éloignent de l'origine de façon continue. Cet état de fait ne peut évidemment pas améliorer les performances du "Spiral Quantizer".

En plus de ce qui a été mentionné dans le chapitre 3, il est à remarquer que l'étape de discrétisation s'est déroulée en deux fois et les résultats finaux n'ont été obtenus qu'après l'étude de la stabilité du "Spiral Quantizer". Il n'y a pas d'autre alternative puisque notre quantificateur nous était inconnu de part ses performances .

#### 4.3.2- Stabilité du "Spiral Quantizer".

Les droites horizontales tracées dans les figures 3.4.1 et 3.4.2 délimitent une plage permise des fluctuations. Celle-ci, chiffrée dans les graphes, nous permet d'affirmer que les valeurs du RSB du "Spiral Quantizer" se stabilisent à partir d'un nombre de vecteurs égal ou supérieur à 4096.

Il a été constaté que les valeurs du RSB après une application de l'algorithme de la K-Moyenne se stabilisent aussi à partir de ce même nombre de vecteurs.

#### 4.3.3- Les performances en terme de RSB du "Spiral Quantizer".

Les résultats obtenus, comparés aux résultats théoriques prédits par la théorie de la distorsion (bornes théoriques de Shannon décrites aux chapitres précédants), montrent que les performances réalisées par le "Spiral Quantizer" sont excellentes.

L'algorithme de la K-Moyenne n'a pas donné un apport important pour une spirale configurée en quantificateur . Ce phénomène est de fait logique car les performances de notre quantificateur sont déjà proches de celles du quantificateur optimal ; le "Spiral Quantizer" ne peut atteindre les limites théoriques de Shannon.

Par ailleurs, on peut remarquer une propriété intéressante de notre quantificateur qui est de pouvoir présenter des structures de taille différente, obtenues à partir d'un seul procédé de développement de la spirale. De plus, ces ensembles de structures sont pratiquement tout aussi performants les uns que les autres.

Une possibilité d'augmenter encore le RSB :

Les caractéristiques statistiques de notre source montrent que celle-ci peut être considérée comme une source gaussienne à moyenne nulle.

Cependant une recherche similaire concernant les valeurs arrondies de sortie fournies par le "Spiral Quantizer" et exprimées sous la forme  $x = [x_1 ; x_2]$  font ressortir que ces valeurs de sortie n'obéissent pas aux mêmes lois probabilistes.

Nous présentons ci-dessous un échantillon des mesures effectuées sur les valeurs arrondies de sortie du "Spiral Quantizer" (tableau 4.1, tableau 4.2 et figure 4.9).

X <sub>1</sub> : Column 1					
Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
-.001131	1.313511	.027753	1.72531	-1.16114E5	2240
Minimum:	Maximum:	Range:	Sum:	Sum Squared:	# Missing:
-3.174695	3.190437	6.365132	-2.533945	3862.972124	0
t 95%:	95% Lower:	95% Upper:	Kurtosis:	Skewness:	
.05443	-.055561	.053298	-.393834	-.002157	

Tableau 4.1: Caractéristiques statistiques de la première colonne de valeurs arrondies fournies par le "Spiral Quantizer".

De ce fait, on peut en déduire qu'il existe une possibilité d'augmenter les performances du quantificateur à base de la spirale. Le principe consisterait à rechercher des constellations de points générés par la spirale et qui approcheraient par leur répartition la loi Normale de distribution.

X <sub>2</sub> : Column 2					
Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
.00628	1.316221	.02781	1.732437	20960.35995	2240
Minimum:	Maximum:	Range:	Sum:	Sum Squared:	# Missing:
-3.257764	3.206972	6.464736	14.06624	3879.015059	0
t 95%:	95% Lower:	95% Upper:	Kurtosis:	Skewness:	
.054542	-.048262	.060821	-.397345	.00384	

Tableau 4.2: Caractéristiques statistiques de la deuxième colonne de valeurs arrondies fournies par le "Spiral Quantizer".

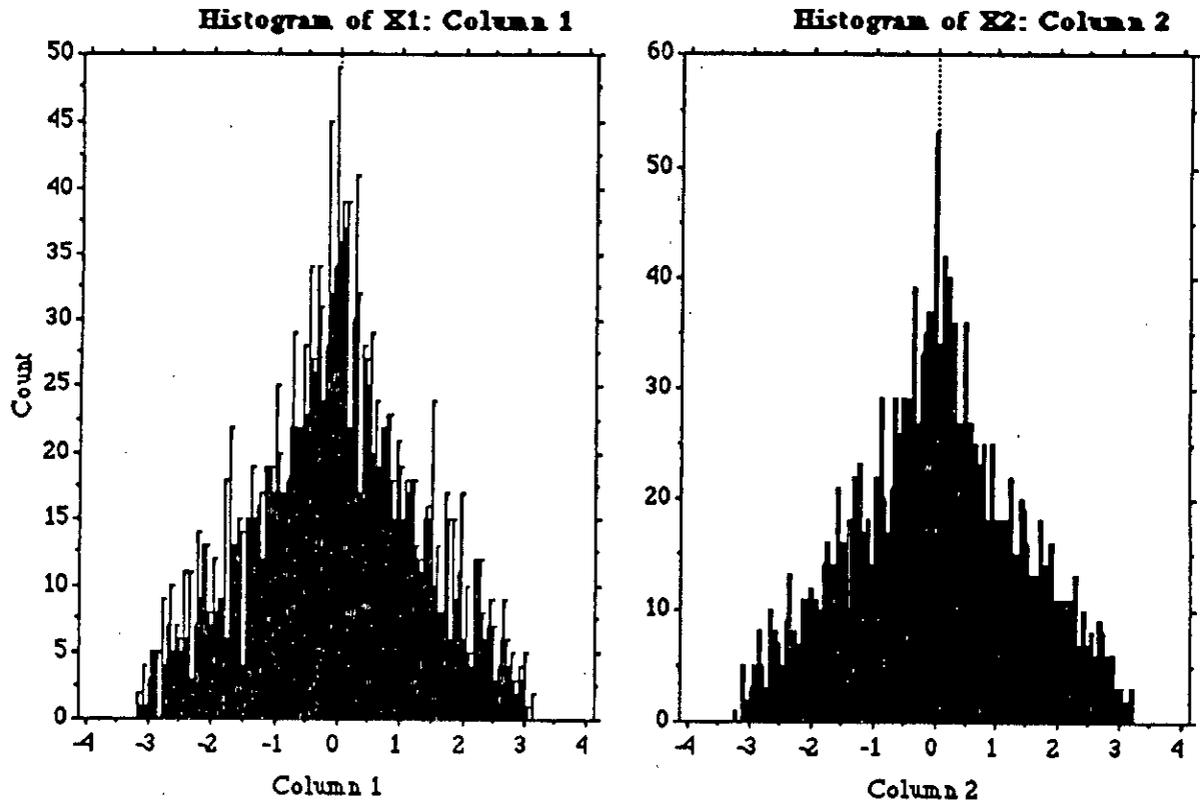


Figure 4.9 : Répartition des valeurs de sortie par colonne.

Différentes alternatives peuvent être envisagées, l'une d'elle pourrait être la recherche d'une pondération du facteur d'échelle "Gain" qui se transformerait, dans ce cas, en une fonction non linéaire. Cette solution complique quelque peu l'équation de la spirale génératrice de l'ensemble des valeurs de sortie.

A partir de ce qui a été annoncé plus haut, nous avons procédé à un essai d'optimisation par une utilisation successive de l'algorithme de la K-Moyenne. Cette opération nous a permis d'obtenir le résultat que nous présentons à la figure 4.10 .

La performance réalisée par la constellation est excellente. Peut-on conclure qu'on n'est pas à l'abris de valeurs optimales locales des paramètres ? Ceci remet-il en cause les résultats obtenus ?

La réponse est négative car l'algorithme de la K-Moyenne ne fait pas augmenter de façon substantielle les performances en terme de RSB de notre quantificateur; même les

performances réalisées après une application de cet algorithme itératif demeurent très proches de celles du "Spiral Quantizer" dont les paramètres sont seulement au voisinage des valeurs estimées optimales.

Par contre la solution qui consiste à rechercher une fonction non linéaire de modulation du facteur d'échelle reste à considérer.

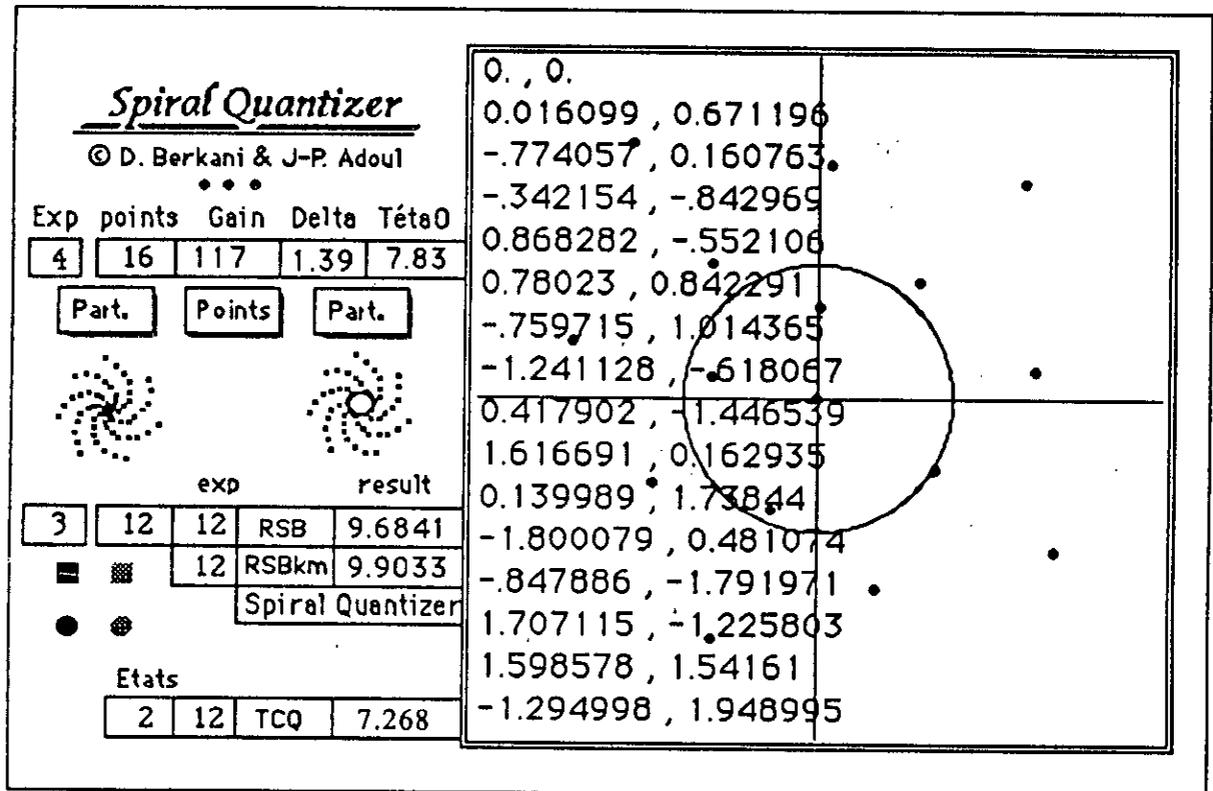


Figure 4.10 : "Spiral Quantizer" réalisé par application successive de la K-Moyenne.

#### 4.3.4- Le Set Partitioning.

Les résultats obtenus avec cette méthode sont meilleurs et confirment les propriétés et les performances théoriques obtenues en modulation (TCM) et en quantification par TCQ.

On constate par ailleurs que la répartition des quantificateurs élémentaires de l'ensemble, qui suit un principe propre à la configuration en "Set Partitioning" [35-37] n'est

pas nécessairement adaptée pour un quantificateur constitué par un ensemble de points (quantificateur simple).

Ainsi, une constellation optimisée composée de points simples n'est pas nécessairement la meilleure pour une conception en "Set Partitioning".

### Conclusion

Dans cette partie, un bref retour à la quantification par réseaux réguliers nous permis de prendre connaissance en partie de la problématique de la quantification uniforme, du quantificateur optimal et des conditions de son existence, de constater l'échec, pour l'heure actuelle, de l'extension des méthodes scalaires au cas de la quantification vectorielle.

Par ailleurs, il est connu que l'utilisation des méthodes de codage par réseaux et codes correcteurs d'erreur offrent plusieurs avantages [7]. On sait maintenant évaluer l'importance et l'apport d'un quantificateur à structure algébrique au problème du codage de canal et de source. Le "Spiral Quantizer" de part ses performances comparées aux limites théoriques de Shannon (chapitre 3) présente des avantages qu'il tire de l'exploitation de ces techniques algébriques de quantification en dehors de ses propres particularités.

Aussi, l'étude comparative n'a été présentée que pour situer les performances des quantificateurs existants.

L'étude statistique des valeurs arrondies du "Spiral Quantizer" montre qu'il existe une possibilité d'augmenter le RSB. C'est une des perspectives qui peut être mise à profit à l'avantage de notre quantificateur dans le futur.

## Conclusion

Nous sommes parti de l'idée générale de mettre au point un algorithme de design d'un jeu de quantificateurs qui optimise la performance sur une séquence d'apprentissage.

Les résultats originaux obtenus montrent que les objectifs ont été plus qu'atteints.

Une issue heureuse a été trouvée à nos efforts et notre travail de recherche a abouti à la mise en œuvre d'un nouveau quantificateur : le "Spiral Quantizer", construit à partir de la discrétisation de la spirale d'Archimède. C'est la première fois que la spirale est utilisée pour générer des ensembles de vecteurs en vue d'une quantification ou d'une modulation.

Dans cette thèse, nous avons présenté ce nouveau quantificateur. L'étude du "Spiral Quantizer" montre que ce dernier présente plusieurs avantages en particuliers pour la quantification d'une source gaussienne bidimensionnelle.

Notre quantificateur vectoriel met en jeu un ensemble indexé de vecteurs capable d'approcher les performances optimales théoriques avec comme critère de fidélité l'erreur quadratique moyenne.

L'application à une séquence d'apprentissage, d'un algorithme de type K-Moyenne ou algorithme LBG n'améliore pas de façon significative les performances initiales du "Spiral Quantizer"; cette propriété reste valable pour des dimensions ou débits par échantillons différents. Ce qui est dû essentiellement au fait que ses performances, en terme d'erreur quadratique moyenne pour une source gaussienne à moyenne nulle, sont déjà assez proches des performances optimales prédites par la théorie de la distorsion.

Cette particularité reste importante car elle permet de se passer de l'algorithme itératif qui est long à exécuter, de garder la structure initiale du quantificateur et toute les autres propriétés qui en découlent, en particuliers:

#### Conclusion

- le jeu de vecteurs peut être généré à partir de deux paramètres et du facteur d'échelle.

- il n'y a aucune nécessité à stocker des vecteurs dans le dictionnaire.

Configuré en "Set Partitioning", le "Spiral Quantizer" exploite les propriétés de la dualité entre la modulation et la quantification, et la combinaison "partitionnement" et structure en treillis pour présenter d'excellentes performances en terme de RSB.

Un des autres avantages très particuliers et propre au "Spiral Quantizer" est la réduction du débit de stockage : un point donné de l'espace considéré est entièrement défini par la distance à l'origine, de ce fait on peut considérer seulement un des paramètres soit l'angle soit la distance à l'origine.

Cette particularité influe également sur la recherche du plus proche voisin donc sur la rapidité et la structure de l'algorithme de recherche .

Les perspectives de recherche futures restent très largement ouvertes. Cette thèse ne constitue qu'une première incursion dans l'environnement des spirales. Aussi les recherches à entreprendre peuvent être orientées vers des directions ou horizons des plus divers.

Parmi les axes les intéressants, il semble que l'étude du comportement du "Spiral Quantizer" en présence de sources réelles telles que l'image et la parole demeure en premier lieu la plus urgente, ce qui se résumerait à une application pratique. Ici, on pourrait éventuellement considérer une autre métrique ou critère de mesure de distorsion.

Une autre orientation non moins intéressante serait de considérer un quantificateur construit à partir d'une discrétisation d'un ensemble de spirales d'Archimède, de spirales modifiées...

## Bibliographie

- [1] T. Berger, "Rate distortion theory, a mathematical basis for a data compression," Englewood Cliffs, N.J.; Prentice-Hall, 1971.
- [2] C.E. Shannon, "A mathematical theory of communication," Bell Syst. Tech. Journal, vol. 27, pp 623-656, 1948.
- [3] R.M. Gray, "Source coding theory," Kluwer, Academic Press, 1990.
- [4] N. S. Jayant and P. Noll, "Digital coding of waveforms," Prentice Hall, 1984.
- [5] R. E. Blahut, "Principles and practice of information theory," Addison Wesley, 1987.
- [6] A. Gersho, "Principles of quantization," IEEE Trans. on circuits and syst., vol. CAS 25, pp 427-436, July 1978.
- [7] J.P. Adoul, "La quantification vectorielle des signaux : approche algébrique," Annales des télécommunications vol 41, 1986.
- [8] R.M. Gray, "Vector quantization," IEEE ASSP Magazine, vol., pp 4-29, April 1984.
- [9] J.P. Adoul, "Speech-coding algorithms and vector quantization," in "Advanced digital communications", Ed. K. Feher, 1987.
- [10] S.P. Lloyd, "Least squares quantization in PCM," IEEE Trans. on Information theory, vol. IT-28, pp 129-137, March 1982.
- [11] J. Max, "Quantizing for minimum distortion," IRE Trans. on Information theory, vol. IT-6, pp7-12, Mar. 1960.
- [12] M.D. Paez, T.H. Glisson, "Minimum mean squared error quantization in speech PCM and DPCM systems," IEEE Trans. on Communications, pp 225-230, April 1972.
- [13] W.A. Pearlman and G.H. Senge, "Optimal quantization of the Rayleigh probability distribution," IEEE Trans. on Communications, vol. COM-2, pp 101-112, Jan. 1979.
- [14] Y. Yamada, S. Tazaki, R.M. Gray, "Asymptotic performance of block quantizers with difference distortion measures," IEEE Trans. on Information theory, vol. IT-26, pp 6-14, January 1980.

Bibliographie

- [15] A. Gersho, " Asymptotically optimal block quantization," IEEE Trans. on Information theory, vol. IT-25, pp 373-380, July 1979 .
- [16] R.M. Gray, A.H. Gray Jr, " Asymptotically optimal quantizers," IEEE Trans. on Information theory , vol. IT-23 , pp 143-154 , January 1977.
- [17] Z. Ben Néticha, "Nouvelles techniques de quantification vectorielle algébrique par réseaux réguliers, codes et treillis," Ph.D dissertation, Sherbrooke, 1990.
- [18] J.H. Conway and N.J.A. Sloane, "Sphere packing-lattice and groupe," New York, Springer Verlag, 1987.
- [19] F.J. MacWilliam and N.J.A. Sloane, " The theory of error correcting codes," Amsterdam, North-Holland, 1977.
- [20] J.H. Conway and N.J.A. Sloane, " Fast quantizing and decoding algorithms for lattices quantizers and codes, " IEEE Trans. on Information theory, vol. IT-28, pp 227-232, March 1982.
- [21] J.H. Conway and N.J.A. Sloane, " Soft decoding techniques for codes and lattices codes , including the Golay code and the Leech lattice," IEEE Trans. on Information theory, vol. IT-32, pp 41-50, January 1986.
- [22] J.P. Adoul and C. Lamblin, "A comparison of some algebraic structures for CELP coding of speech," Proceeding of the IEEE ICASSP' 87, pp 1953-1956, Nov. 1987.
- [23] J.P. Adoul, "Fast ML decoding algorithm for the Nordstrom-Robinson code," IEEE Trans. on Information theory, vol. IT-33, pp 931-933, November 1987.
- [24] M.R. Schroeder and N.J.A. Sloane, " New permutation codes using Hadamard unscrambling," IEEE Trans. on Information theory, vol. IT-33, pp 144-146, January 1987.
- [25] T. Berger, F. Jelinek, J.K. Wolf, " Permutation codes for sources," IEEE Trans. on Information theory, vol. IT-18, pp 160-169, January 1972.
- [26] T. Berger, " Optimum quantizers and permutation codes," IEEE Trans. on Information theory, vol. IT-18, pp 759-765, November 1972.
- [27] T. Berger, "Minimum entropy quantizers and permutation codes," IEEE Trans. on Information theory, vol. IT-28, pp 149-166, March 1982.
- [28] W.A. Pearlman, "Polar quantization of a complex random variable," IEEE Trans. on Communications, vol. COM-27, pp 892-899, June 1979.
- [29] Y. Linde, A. Buzo and R.M. Gray, "An algorithm for vector quantizer design," IEEE Trans. on Communications, vol. COM-28, pp 84-95, Jan. 1980.
- [30] J. Mac Queen, "Some methods for classification and analysis of multivariate observations," The 5th. Berkley Symp. Mathematics, statistics and probability, 1967
- [31] J. P. Adoul & al., "Spectral distance measure applied to the optimum design of DPCM with L predictors," ICASSP' 80, pp 512-515, April 1980.

Bibliographie

- [32] J. P. Adoul, " Backward adaptative reencoding : a technique for reducing the bit rate of m-law PCM transmissions," IEEE Trans. on Communications, vol. COM-30, pp 581-592, April 1982.
- [33] R.T. Fisher and R.M. Dichary, " Vector quantizer design for memoryless Gaussian, Gamma and Laplacian sources," IEEE Trans. on Communications, vol. COM-32, September 1984.
- [34] P. Mabillean, " La quantification vectorielle et son application au codage de la parole," PhD dissertation, Université de Sherbrooke, Qué., Canada, 1983.
- [35] G. Ungerboeck, " Channel coding with multilevel/phase signals," IEEE Trans. on Information theory, vol. IT-28, pp.55-67, Jan. 1982.
- [36] G. Ungerboeck, " Treillis coded modulation with redundant signal sets - PartI: Introduction," IEEE Communications Magazine, vol. 25, pp 5-11, Feb. 1987.
- [37] G. Ungerboeck, " Treillis coded modulation with redundant signal sets - Part II : State of art," IEEE Communications Mag., vol. 25 , pp 12-21 , Feb. 1987.
- [38] A.R. Calderbank, N.J.A. Sloane, " New treillis codes based on lattices and cosets," IEEE Trans. on Information theory, vol. IT-33, pp 177-195, March 1987.
- [39] L. F. Wei, " Treillis coded modulation with multidimensional constellations," IEEE Trans. on Information theory, vol. IT-33, pp 483-501, July 1987.
- [40] G.D. Forney Jr., R.G. Gallager, G.R. Lang, F.M. Longstaff, S.U. Qureshi, " Efficient modulation for band-limited channels," IEEE J. Select. Areas Comm., vol. SAC-2, pp 632-647, September 1984 .
- [41] A.J. Viterbi and J.K. Omura, "Principles of digital communication and coding," Mc Graw-Hill Book Company, New York, 1979.
- [42] M.W. Marcellin and T.R. Fischer, " Treillis coded quantization of memoryless and Gauss-Markov sources," Submitted to IEEE Trans. on Information theory, 1988 .
- [43] A.J. Viterbi, " Error bounds for convolutional codes and asymptotically optimum decoding algorithm," IEEE Trans. on Information theory, vol. IT-13, pp 260-269 , March 1967.
- [44] G.D. Forney Jr., " The Viterbi algorithm," Proceeding of IEEE, vol. 61, pp 268-276, March 1973.
- [45] P.F. Swaszek, " Quantization, " Published by V.N.R. Comp., N.Y, 1985.
- [46] A. Gersho, " On the structure of vector quantizers, " IEEE Trans. on Information theory, vol. IT-28, pp 157-166, March 1982.
- [47] J.J.Y. Huang and P.M. Schultheiss, " Block quantization of correlated Gaussian random variables , " IEEE Trans. on Communications Syst., vol. CS-11, pp 289-296, September 1963.

Bibliographie

- [48] J.A. Bucklew and N.C. Gallagher Jr., "Quantization schemes for bivariate Gaussian random variables," *IEEE Trans. on Information theory*, vol. IT-25, pp 537-543, September 1979.
- [49] S.G. Wilson, "Magnitude/phase quantization of independent Gaussian variates," *IEEE Trans. on Communications*, vol. COM-28, pp 1924-1929, Nov. 1980.
- [50] P.F. Swaszek and J.B. Thomas, "Optimal circularly symmetric quantizers," *J. Franklin Inst.*, vol.313, pp 272-384, June 1982.
- [51] P.F. Swaszek and T Ku, "Asymptotic performance of unrestricted polar quantizers," *Proceeding of the 1984 Conference on Information Sciences and Systems*, pp 581-586, March 1984, Princeton University.
- [52] P.F. Swaszek, J.B. Thomas, "Multidimensional spherical coordinates quantization," *IEEE Trans. on Information theory*, vol. IT-29, pp 570-576, 1983.
- [53] D. Berkani, A.Chekima, J.P. Adoul, "La spirale discrète: applications en quantification vectorielle et modulation," soumis au Coll. GRETSI sur le traitement du signal et des images, Septembre 1991.
- [54] N.C. Gallagher Jr., "Quantizing schemes for the Discrete Fourier Transform of random times series," *IEEE Trans. on Information theory*, vol. IT-24, pp 156-163, March 1978.
- [55] W.A. Pearlman, K.M. Gray, "Source coding of the Discrete Fourier Transform," *IEEE Trans. on Information theory*, vol. IT-24, pp 683-692, November 1978.
- [56] J.P. Adoul and Al., "CELP algorithm with algebraic structures for low bit-rate speech coding," Submitted to ICASSP.
- [57] M. Hamidi and J. Pearl, "On the residual correlation of finite-dimensional the Discrete Fourier Transform of stationary signal," *IEEE Trans. on Information theory*, vol. IT-21, pp 480-482, July 1975.
- [57] N.C. Gallagher Jr., "Discrete spectral phase coding," *IEEE Trans. on Information theory*, vol. IT-22, pp 622 - 624, September 1976.
- [59] R. De Buda, "The upper bound of a new near optimal code," *IEEE Trans. on Information theory*, vol. IT-21, pp 441-445, 1975.
- [60] J.H. Conway and N.J.A. Sloane, "Fast encoding method for lattice codes and quantizers," *IEEE Trans. on Information theory*, vol. IT-29, pp 820-824, November 1983.
- [61] G.D. Forney Jr., "Coset codes PartI: Introduction and geometrical classification," *IEEE Trans. on Information theory*, vol. IT-34, pp 1123-1151, September 1988.
- [62] G.D. Forney Jr., "Coset codes II: Binary lattices and related codes," *IEEE Trans. on Information theory*, vol. IT-34, pp 1152-1187, September 1988.
- [63] W.R. Bennett, "Spectra of quantized signals," *Bell Syst. Tech. J.*, vol. 27, pp 446-472, July 1948.

Bibliographic

- [64] J.P. Adoul, M. Barth, "Nearest neighbor algorithm for spherical codes from the Leech lattice," IEEE Trans. on Information theory, vol. IT-34, pp 1188-1192, September 1988.
- [65] E.S. Barnes, N.J.A. Sloane, "The optimal lattice quantizer in three dimensions," SIAM J. Alg. Disc. Math., vol. 4, pp 30-41, March 1983.
- [66] J.A. Bucklew, "Companding and random quantization in several dimensions," IEEE Trans. on Information theory, vol. IT-27, pp 207-211, March 1981.
- [67] J.P. Adoul, "Spherical-code construction from lattices with efficient ML decoding," IEEE Information theory symposium, Kobe, Japan, 1988.