



## AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : [ddoc-theses-contact@univ-lorraine.fr](mailto:ddoc-theses-contact@univ-lorraine.fr)

## LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

[http://www.cfcopies.com/V2/leg/leg\\_droi.php](http://www.cfcopies.com/V2/leg/leg_droi.php)

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

**THESE**

présentée à l'Université de NANCY I en vue de l'obtention du titre de

**DOCTEUR DE L'UNIVERSITE**

en Métrologie, Automatique et Electronique

option : Métrologie et Electronique Biomédicales

par

**Naïm Mohamed RAHMANI**

**INSTRUMENTATION POUR LE TRAITEMENT NUMERIQUE DU SIGNAL**

**ELECTROENCEPHALOGRAPHIQUE**

**Application à la reconnaissance automatique, temps réel, des différents stades  
de sommeil et de veille chez le rat.**

soutenue publiquement le 26 septembre 1990 devant la commission d'examen

---

**JURY**

**Rapporteurs :** Mr LHUILLIER J.P. (président)  
Mr LAMOTTE M.

**Directeur de thèse :** Mr PRIEUR G.  
**Examineurs :** Mr LEPOUTRE X.F.  
Mr THIELTGEN A.

*Je remercie vivement le Professeur G. PRIEUR qui a dirigé ce travail et qui m'a soutenu tout au long de ce projet.*

*Je tiens à remercier le Professeur J.P. LHUILLIER qui m'a fait l'honneur de présider cette thèse et d'être le rapporteur de ce travail.*

*Que Monsieur M. LAMOTTE, directeur de recherches au CNRS, soit remercié pour les remarques judicieuses dont il m'a fait part en tant que rapporteur de ce travail.*

*Je suis très reconnaissant à M. X.F. LEPOUTRE pour les nombreux conseils qu'il m'a donné à la lecture de ce manuscrit et d'avoir accepté de faire partie du jury de thèse.*

*Je remercie M. A. THIELTGEN pour l'intérêt qu'il a porté à ce travail et pour ses remarques constructives.*

*M. G.GOVAERT a mis à ma disposition, au sein de l'INRIA Lorraine, une partie des moyens nécessaires à l'aboutissement de ce travail, que lui et son équipe soient vivement remerciés.*

*Mes remerciements vont aussi au Professeur LIMOGES et à son équipe qui m'ont accueilli au laboratoire d'électrophysiologie de la faculté de chirurgie dentaire à Montrouge.*

*Je remercie également le ministère de l'enseignement supérieur du gouvernement Algérien de m'avoir octroyé une bourse d'étude qui m'a permis de mener à bien ce travail.*

*Que tous ceux qui ont contribué de près ou de loin à la réalisation de ce travail trouvent ici mes sincères remerciements.*

*A mon père.*

## **TABLE DES MATIERES**

INTRODUCTION .....	1
--------------------	---

## CHAPITRE I

INTRODUCTION A UNE METHODOLOGIE GENERALE POUR LA CONCEPTION D'UNE INSTRUMENTATION DE RECONNAISSANCE DE FORME .....	4
--	---

1 - 1 INTRODUCTION .....	5
1 - 2 QUELQUES CRITERES POUR LA CONCEPTION D'UNE TELLE INSTRUMENTATION.....	6
1 - 3 CONDITIONS POUR LE SUCCES D'UNE TELLE INSTRUMENTATION .....	6
1 - 4 DEFINITION DU PROJET.....	7
1 - 5 COMMUNICATION PROCESSUS-OPERATEUR HUMAIN .....	8
1 - 5 - 1 NOTION DE PROCESSUS .....	10
1 - 5 - 2 INTERACTION DU PROCESSUS AVEC LE MONDE EXTERIEUR .....	11
1 - 5 - 1 - 1 Acquisition du signal .....	11
1 - 5 - 2 - 2 Réglage .....	11
1 - 5 - 3 OBSERVATION DES PROCESSUS .....	12
1 - 6 EXTRACTION D'INFORMATION.....	13
1 - 7 REPRESENTATION ET INTERPRETATION [6],[44].....	13
1 - 8 ASPECTS SEMANTIQUE ET SYNTAXIQUE D'UN PHENOMENE PHYSIQUE.....	18
1 - 9 PHASE APPRENTISSAGE.....	18
1 - 10 RECONNAISSANCE DE FORMES .....	21
1 - 11 CONCLUSION CHAPITRE I.....	23

## CHAPITRE II

APPLICATION A LA RECONNAISSANCE DES STADES DE SOMMEIL ET DE VEILLE CHEZ LE RAT .....	24
---	----

2 - 1 INTRODUCTION .....	25
2 - 1 - 1 CAHIER DES CHARGES.....	25
2 - 1 - 2 PRESENTATION ET DISCUSSION DE QUELQUES EXEMPLES DE SYSTEMES .....	25
2 - 1 - 3 NOUVELLE APPROCHE.....	26
2 - 2 MESURE DES PHASES DE SOMMEIL ET DE VEILLE .....	28



3 - 2	DEFINITIONS .....	57
3 - 3	ANALYSE DE DONNEES .....	57
3 - 3 - 1	PRATIQUE DE L'ANALYSE DES DONNEES :.....	58
3 - 3 - 2	LES PIEGES DE L'ANALYSE DES DONNEES .....	59
3 - 3 - 3	LES PREMIERES ETAPES D'UNE ANALYSE DE DONNEES .....	60
3 - 3 - 3 - 1	Choix des individus .....	60
3 - 3 - 3 - 2	Définition et choix des variables .....	60
3 - 3 - 4	DEFINITIONS ELEMENTAIRES.....	61
3 - 3 - 4 - 1	Les différents types de variables.....	61
3 - 3 - 4 - 2	Le tableau de données.....	62
3 - 3 - 4 - 3	Quelques éléments descriptifs des variables quantitatives .....	64
3 - 3 - 4 - 4	Centre de gravité d'un nuage de points .....	65
3 - 3 - 4 - 5	Inertie d'un nuage de points par rapport à un point.....	66
3 - 3 - 4 - 6	Inertie interclasse et inertie intraclasse.....	66
3 - 3 - 5	CLASSES ET CLASSIFICATION AU SENS DE L'INERTIE.....	67
3 - 3 - 6	QUELQUES INSTRUMENTS DE L'ANALYSE DE DONNEES .....	67
3 - 3 - 6 - 1	L'analyse en composantes principales .....	67
3 - 3 - 6 - 2	Les nuées dynamiques .....	71
3 - 3 - 7	QUELQUES DONNEES NUMERIQUES MESUREES SUR UN ECHANTILLON DE SIGNAUX EEG CHEZ LE RAT.....	73
3 - 3 - 7 - 1	Description des variables quantitatives.....	74
3 - 3 - 7 - 2	Description des variables qualitatives .....	75
3 - 3 - 7 - 3	Matrice de corrélation .....	76
3 - 3 - 7 - 4	Classification automatique .....	77
3 - 3 - 8	APPRENTISSAGE.....	79
3 - 3 - 8 - 1	Détermination des modèles .....	79
3 - 3 - 8 - 2	Correction de l'apprentissage .....	86
3 - 3 - 8 - 3	Cas particulier de l'éveil .....	87
3 - 3 - 8 - 4	Discussion sur la technique d'apprentissage utilisée.....	87
3 - 3 - 8 - 5	Arrêt sur l'intérêt du filtrage des signaux .....	88
3 - 4	CONCLUSION CHAPITRE III .....	89



## CHAPITRE IV

### DESCRIPTION DU SYSTEME DE RECONNAISSANCE AUTOMATIQUE

DES STADES DE SOMMEIL ET DE VEILLE ..... 91

4 - 1 INTRODUCTION ..... 92

4 - 2 EQUIPEMENT INFORMATIQUE..... 92

4 - 2 - 1 MATERIEL ..... 92

4 - 2 - 2 LOGICIEL ..... 92

4 - 3 PHASE RECONNAISSANCE DE FORME ..... 95

4 - 3 - 1 ALGORITHMES DE RECONNAISSANCE..... 95

4 - 3 - 2 NOTION DE DISTANCE ENTRE  
REPRESENTATION ..... 96

4 - 3 - 3 NOTION DE RESSEMBLANCE ET DE  
DISSEMBLANCE ..... 97

4 - 4 DESCRIPTION DU SYSTEME D'ANALYSE DE SOMMEIL  
CHEZ LE RAT ..... 99

4 - 4 - 1 DESCRIPTION DU SLP..... 99

4 - 4 - 2 DESCRIPTION DE VISU .....101

4 - 4 - 3 REGLAGE DU NIVEAU DES SIGNAUX EEG.....102

4 - 5 EVALUATION.....105

4 - 6 CONCLUSION CHAPITRE IV .....107

CONCLUSION .....109

BIBLIOGRAPHIE.....112

## **INTRODUCTION**

Ce travail a pour objectif de représenter une approche méthodologique globale au problème de traitement automatique, temps réel, des signaux électrophysiologiques; cependant, les techniques utilisées nous permettent d'avancer qu'il pourrait s'adapter à d'autres ensembles d'applications, tels que : les signaux non électrophysiologiques, les diagnostics, la surveillance des processus dynamiques, etc... et dans toutes les tâches où la présence continue d'un opérateur humain est indispensable.

L'origine physiologique de certains signaux leur donne des caractéristiques complexes difficiles à interpréter. L'œil " expert " de l'opérateur humain peut interpréter tel ou tel signal, non sans se tromper dans son jugement de temps à autre. L'opérateur complète un ensemble d'informations objectives, visualisées au niveau des commandes et à la sortie d'un système ( moniteurs cathodiques, dérouleurs graphiques, cadrans, etc...), à l'aide d'observations de grandeurs difficilement mesurables par des moyens techniques, et grâce à une connaissance subjective du comportement de ce système. Une connaissance acquise par apprentissage, et qu'on appelle "l'expérience".

La conception d'une instrumentation en vue d'un traitement automatique de ce type de signaux relève, quand à elle, d'avantage d'une certaine forme d'artisanat que de méthodes universelles et objectives. La gamme très large des signaux rencontrés conduit, en effet, à considérer chaque cas comme un cas particulier. On définira, toutefois, quelques principes méthodologiques fondamentaux.

Le premier chapitre est une introduction à une méthodologie générale pour la conception d'une instrumentation pour la reconnaissance de forme dans un cas général. Il définit quelques notions relatives à la reconnaissance de forme qui seront exploitées dans les chapitres suivants. On insistera sur l'amélioration de la communication entre l'instrument et l'opérateur humain qui s'en sert.

Dans le deuxième chapitre on aborde l'exemple qui a servi de base à l'application de la méthodologie décrite dans le premier chapitre et qui consistait en la réalisation d'un système de reconnaissance automatique, temps réel, des différents stades de sommeil et de veille chez le rat à partir uniquement du traitement d'un signal électroencéphalographique (EEG). En effet dans certains travaux de laboratoire, on avait besoin d'une lecture de l'enregistrement EEG pour déterminer les différents stades de sommeil et de veille chez le rat. Cette tâche étant souvent réalisée par un opérateur humain, on décrit le long de ce travail l'instrumentation mise en œuvre pour remplacer

l'opérateur dans cette tâche qui est souvent pénible. Ce chapitre est essentiellement consacré au traitement numérique du signal électroencéphalographique.

Le troisième chapitre, et suite aux traitements présentés dans le deuxième chapitre, introduit une technique de modélisation des différentes formes du signal EEG, l'apprentissage du système se faisant en parallèle par "professeur" et au moyen d'un classificateur automatique.

On décrit les différents modules composant le système de reconnaissance de forme mis en œuvre et leurs rôles respectifs. Ceci fera l'objet d'un quatrième chapitre.

Une conclusion fait l'inventaire des avantages et des inconvénients d'un tel système, tel qu'il a été conçu, pour ce type particulier d'application et les améliorations à y apporter.

## **CHAPITRE I**

### **INTRODUCTION A UNE METHODOLOGIE GENERALE POUR LA CONCEPTION D'UNE INSTRUMENTATION DE RECONNAISSANCE DE FORME**

## 1 - 1 INTRODUCTION

Dans ce chapitre on présente des techniques et des notions de base relatives à tout système de reconnaissance de forme. Cette présentation évite de reprendre ces définitions ou ces notions dans la suite du manuscrit. Toutefois, chaque fois qu'on aurait besoin d'une définition qui n'est pas donnée dans ce chapitre, et qui est spécifique à une technique bien précise, on l'introduira en son lieu.

Notre démarche visait à la conception d'une instrumentation dite "intelligente", où l'instrument n'est plus seulement l'objet qui permet à l'homme de contrôler son environnement, mais le remplace dans certaines tâches et devient un outil d'aide à la réflexion et à la décision.

Avant d'aller plus en avant on définira, une fois pour toute, la notion d'instrumentation intelligente et le contexte dans lequel cette appellation sera utilisée dans ce manuscrit :

*Instrumentation intelligente :*

Rendre l'instrument capable d'accomplir des tâches qui réclameraient de "l'intelligence", si elles étaient faites par un être humain.

Alors, la question qu'on pourrait se poser dès le départ est : " *Le système informatique autour duquel sera réalisée cette instrumentation est-il un système intelligent ou bien un ensemble de programmes plus classiques ?*".

*Qu'est-ce qui fait l'intelligence d'un système informatique ? [1,2,3,4] :*

Il existe une grande variété de logiciel disponible sur le marché, dont une partie très complexe. La question suivante est à la fois intéressante et très controversée : " *Qu'y a-t-il de spécial dans un système intelligent ?*". Il existe des cas de projets réalisés pour mettre sur pied des systèmes experts où la tâche était considérée, au début, comme très complexe. Cependant, après une énorme acquisition de savoir et un haut degré de raffinement de ce savoir, on a mis au point des systèmes simples dans leur conception. En fait, c'est l'acquisition du savoir qui a été difficile et non les équipements qui en ont découlé, les chapitres 3 et 4 le montrent très bien.

De tels programmes dont la structure est relativement simple sont-ils des systèmes intelligents ou des programmes de résolution de problèmes complexes ? En réalité, cette question ne nous concerne pas dans le contexte de ce manuscrit, car notre principal souci est ici de produire un système utile et le nom qu'on lui donne est, dans une certaine mesure, une question secondaire sur laquelle on ne s'attardera pas.

## 1 - 2 QUELQUES CRITERES POUR LA CONCEPTION D'UNE TELLE INSTRUMENTATION

Quelques-uns de ces critères ont été énoncés par le groupe de DENDRAL<sup>1</sup>, et s'appliquent bien à notre cas :

- Concentration sur un domaine associé à une spécialité pointue qui ne fait pas appel à trop de connaissances générales de bon sens .

- Choisir une tâche qui n'est ni trop simple, ni trop compliquée pour un opérateur humain.

- Définir clairement la tâche.

- La participation d'un véritable expert ainsi que son engagement de longue durée sont essentiels.

## 1 - 3 CONDITIONS POUR LE SUCCES D'UNE TELLE INSTRUMENTATION

Une instrumentation, telle qu'elle va être présentée et conçue, autour d'un système informatique, a le défaut d'être "fragile" au sens de la fiabilité. Pour remédier à cela, il est important de prendre en compte les quelques précautions suivantes :

- L'espace de recherche doit être petit, dans le cas d'un ordinateur de type PC, un grand espace de recherche introduit un temps de calcul très important ce qui a, au moins, la conséquence d'enlever à l'instrumentation la qualité du "temps réel", (voir chapitre IV, paragraphe 4-2-2).

---

<sup>1</sup> DENDRAL : groupe de recherche constitué à l'université de Stanford en 1965 pour travailler à la conception d'un système destiné à aider les chimistes à inférer la structure de composants chimiques à partir de données fournies par leur spectre de masse.

- Le domaine de connaissance qui a permis l'apprentissage doit être fiable et ne doit pas contenir d'erreurs ni d'incohérence, ce qui est souvent très difficile à vérifier au départ.

- Les données fournies par l'utilisateur doivent être sûres et constantes (les données liées à une solution ne doivent pas être invalidées durant le processus de reconnaissance. Un exemple typique est le changement du niveau des signaux au cours du processus de reconnaissance).

#### 1 - 4 DEFINITION DU PROJET

L'instrumentation doit être conçue, au niveau matériel et logiciel, pour être en conformité avec les critères et les consignes fournies par des experts, dans le but de réaliser une tâche bien définie.

Une instrumentation qui marche efficacement sera certainement une grande aide pour l'opérateur humain. Toutefois, son succès ne peut être garanti dès le démarrage de l'étude. Dans les tout premiers stades de développement, il est capital de considérer le projet très soigneusement. Le problème doit être clairement défini, et suffisamment important pour justifier l'étude, tout en étant techniquement traitable.

Parmi les questions à poser, dès le départ, figurent celles-ci [5] :

- Quelle est l'importance de ce problème ?
- Quels seraient les avantages d'une telle instrumentation ?
- Le problème est-il fréquent ?
- Quelle sera son importance dans le futur proche ?
- Le problème peut-il être défini facilement ?
- serait-il pratique d'utiliser un ordinateur ?
- Quelqu'un a-t-il travaillé ailleurs sur un problème similaire ?
- Qui se servirait d'un tel système ? pourquoi ?
- Y a-t-il une "coque" dont nous pourrions nous servir ?
- Existe-t-il une documentation quelconque ?
- Comment l'opérateur humain a-t-il appris à réaliser la tâche ?
- Pouvons-nous nous passer du temps d'un opérateur expert ?
- De quelles ressources avons-nous besoin ?
- Qu'est-ce qui pourrait rendre ce projet difficile à réaliser ?
- Est-ce que les experts sont en désaccord ?



- combien faut-il de temps pour devenir un expert ?
- Y a-t-il des moments où les experts ne sont pas disponibles pour une consultation ?
- Le savoir est-il complexe ? Nécessite-t-il plusieurs mécanismes d'inférence et représentations ?
- Le système devra-t-il être remis à jour fréquemment ? ou est-il définitif ?
- Pouvons nous tolérer une réponse imparfaite ?
- La réalisation de l'interface instrument-opérateur exigera-t-elle de grands efforts ?

Une discussion ainsi orientée aidera à distinguer les difficultés possibles et à parvenir à un accord sur des objectifs raisonnables.

Certaines de ces questions trouverons leurs réponses dans les chapitres suivants. On y trouvera une description du problème de reconnaissance des différents stades de sommeil et de veille chez le rat, le rôle de l'opérateur humain et la manière avec laquelle il détermine les différents stades, les systèmes réalisés dans le domaine, etc...

Certaines idées seront exploitées explicitement (apprentissage, reconnaissance ...), et d'autres implicitement (l'observation, l'extraction d'information, la représentation ...)

## 1 - 5 COMMUNICATION PROCESSUS-OPERATEUR HUMAIN

On considère, tout au long de l'étude, l'opérateur humain "expert" et le processus électrophysiologique, comme faisant partie d'une même chaîne de communication unidirectionnelle, la source étant un appareil et la destination l'opérateur. La figure (1,1) donne le schéma élémentaire d'une chaîne de communication unidirectionnelle (d'après VASSEUR) [6].

De la bonne communication entre un processus et un opérateur humain, dépend en grande partie, l'interprétation et la compréhension de ce processus. On présente dans ce qui suit quelques notions concernant l'amélioration de cette communication.

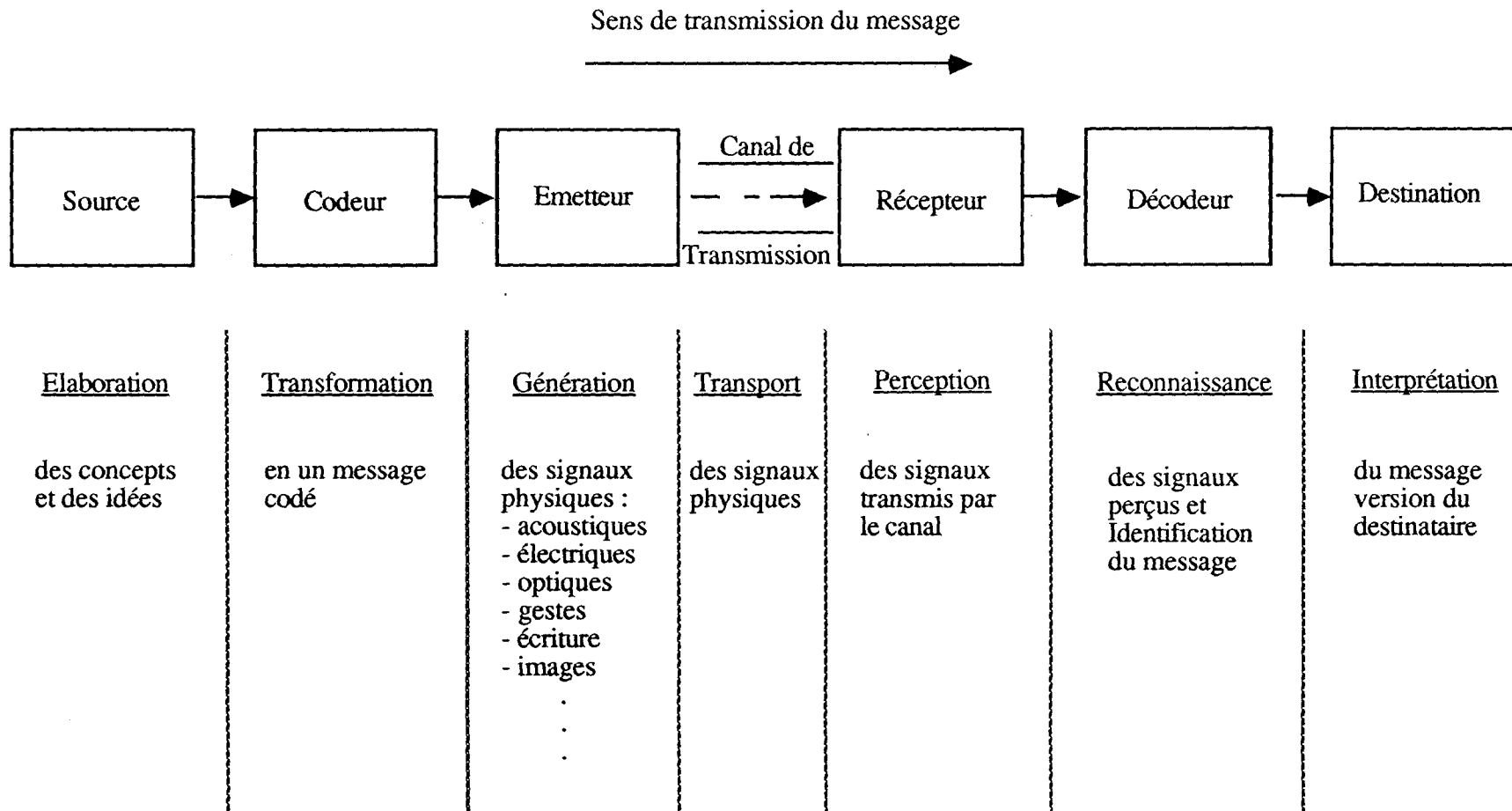
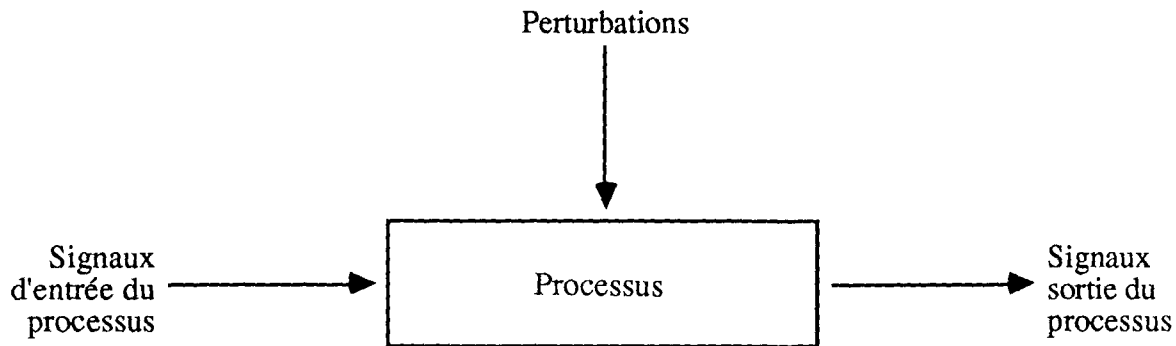


figure (1,1)  
Schéma élémentaire d'une chaîne de communication unidirectionnelle

### 1 - 5 - 1 NOTION DE PROCESSUS

Un processus peut être considéré comme une séquence d'opération ou d'actions qui sont conditionnées par des données d'entrée et qui produisent des signaux de sortie figure(1,2).



figure(1.2)

Représentation schématique d'un processus

- Les données d'entrée peuvent être soit des perturbations, soit des grandeurs de commande.

- Les perturbations sont des signaux d'entrée qui agissent sur le processus, mais qui ne sont pas connues ou sur lesquelles on ne désire pas agir.

- Les grandeurs de commande sont des signaux dont les valeurs sont imposées au processus, par exemple par un opérateur humain.

- Les signaux de sortie représentent l'activité utile du processus. Ils sont les seuls points observables, et ce sont eux qui permettent d'apprécier l'état du processus.

Il existe trois différents types de processus :

- les processus continus
- les processus discrets
- les processus mixtes

Il convient de noter que dans toute l'étude qu'on mènera ici, les processus considérés sont des processus électrophysiologiques continus.

### 1 - 5 - 2 INTERACTION DU PROCESSUS AVEC LE MONDE EXTERIEUR

Les interactions entre un processus et le monde extérieur se situent au niveau des signaux de sortie pour la détermination de l'état du processus, et au niveau des signaux d'entrée pour la commande du processus. On distingue quatre modèles différents d'interaction avec les processus :

- l'acquisition de signaux
- la commande
- le réglage
- la commande d'automatismes

On va détailler un peu plus l'acquisition du signal et le réglage.

#### 1 - 5 - 2 - 1 *Acquisition du signal :*

Dans le cas d'un signal continu, l'acquisition du signal se réduit à la mesure des signaux de sortie du processus. L'interaction avec le processus se ramène à une simple détermination de l'état du processus, afin de le connaître et d'en prévoir son évolution, mais sans agir sur le processus. La collecte de données électrophysiologiques ou météorologiques constituent un exemple d'acquisition de signaux.

L'acquisition du signal est souvent précédée par une opération de numérisation, on utilise pour cela des convertisseurs analogique-numérique pour digitaliser le signal avec le moins de dégradation possible. En effet, il n'est pas toujours possible de se connecter directement sur le phénomène à observer, mais il faudrait le préparer à l'aide d'une interface pour pouvoir, par exemple le relier à un ordinateur.

#### 1 - 5 - 2 - 2 *Réglage :*

Afin de réduire l'influence des perturbations, quelles qu'elles soient, à condition qu'elles restent dans certaines limites, on procède à des réglages qui sont, au sens des automaticiens, des boucles de contre-réaction qui tendent à maintenir l'état du

système à une valeur de consigne. les réglage de la bande passante d'un système, ou l'atténuation du bruit dans un signal sont des exemples de réglage.

### 1 - 5 - 3 OBSERVATION DES PROCESSUS

L'observation des processus physiques s'appuie sur les points suivants :

a - Le fonctionnement du processus se traduit par un ensemble de manifestations extérieures telles que :

- bruits,
- déplacement,
- tensions électriques,
- pressions,
- températures . . .

b - Ces manifestations extérieures constituent autant de message permettant à l'observateur (individu ou instrument) d'entrer en communication avec le processus.

c - l'objectif est de développer les méthodes et les moyens permettant d'améliorer la communication entre le processus et l'observateur.

d - On commence par comprendre les mécanismes de communication entre un processus et un opérateur humain pour arriver, enfin, à remplacer l'opérateur humain par une instrumentation.

Cette observation peut se décomposer en deux phases distinctes, plus ou moins développées :

- La perception, à laquelle est liée la notion de capteur.
- La compréhension, à laquelle sont liées les notions d'interprétation et de signification des objets perçus.

## 1 - 6 EXTRACTION D'INFORMATION

A partir d'un signal électrophysiologique, on voudrait tirer un maximum de données significatives, caractérisant une certaine information véhiculée par ce signal, afin de faciliter la communication entre processus, en l'occurrence électrophysiologique ( sommeil, respiration, etc...), et un observateur ( humain ou instrument ).

Pour cela on dispose de différents outils de traitement du signal qui permettent de conditionner le signal :

- filtrage,
- seuillage,
- amplification,
- atténuation,
- modulation-démodulation, etc... ;

Le conditionnement étant une première étape, figure (1,3) [6], la seconde est l'interprétation. Pour mieux définir ce qui à été écrit, on aborde les notions de représentation et d'interprétation.

## 1 - 7 REPRESENTATION ET INTERPRETATION [6],[44]

Pourvu qu'on leur donne les instructions appropriées, les machines informatiques sont capables d'effectuer n'importe quelle transformation sur une représentation codée. En particulier, on peut penser qu'elles sont capables d'effectuer et d'interpréter des représentations provenant de capteurs physiques, tels qu'une caméra, un microphone, un enregistreur; donc des représentations d'image, de parole et d'activité biologique.

Donnons quelques définitions de termes que nous utilisons dans un contexte informatique de reconnaissance de forme.

- *Information* : Une information est un couple formé d'une représentation matérielle, et de son ou ses interprétations.

- *Représentation* : C'est une chaîne d'éléments d'un alphabet fini, permettant de représenter (de coder) une information; en fait au niveau de la machine une chaîne de

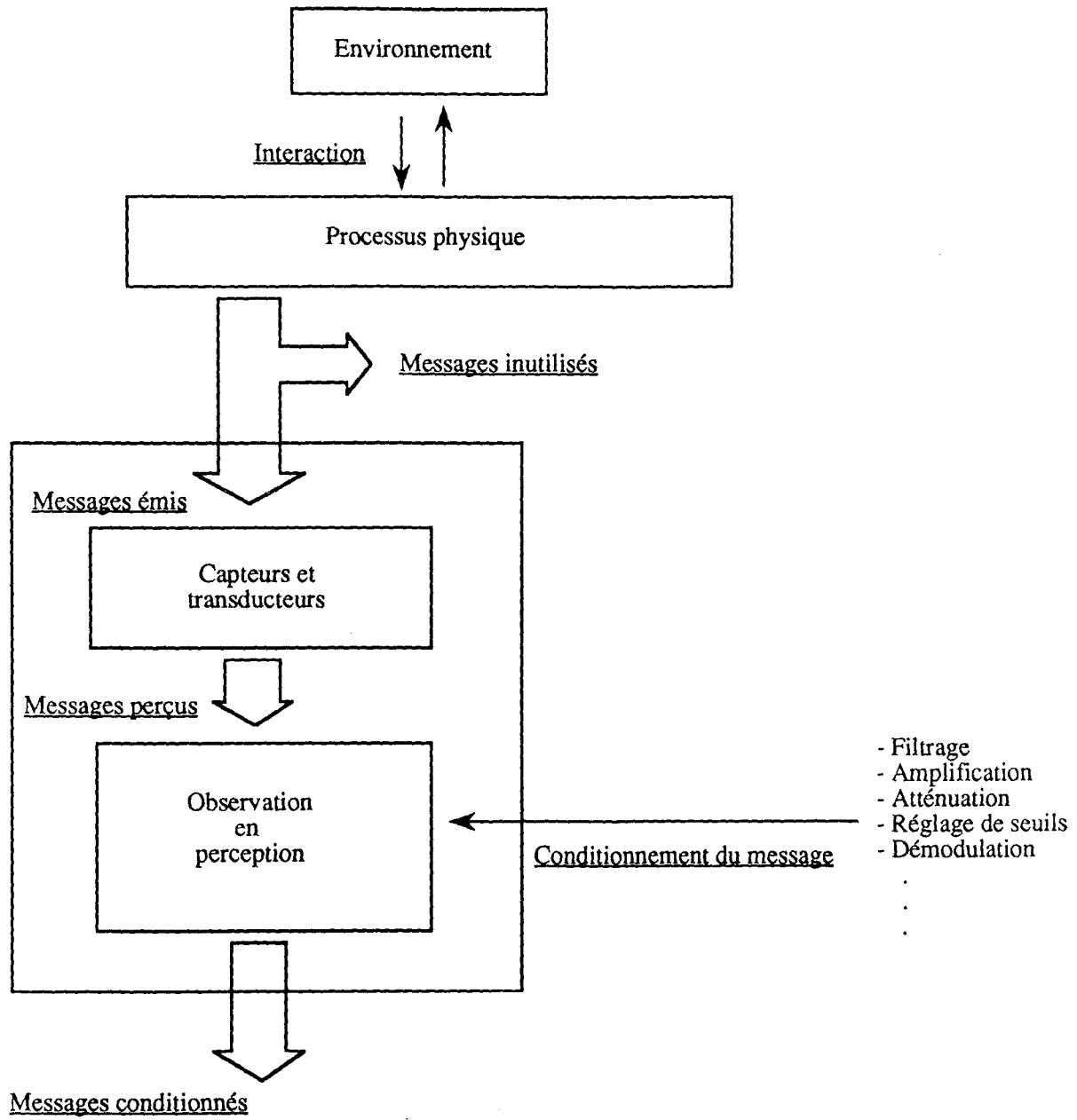


figure (1.3)  
Conditionnement d'un processus physique

bits en général enregistrés en mémoire. L'utilisateur a rarement accès à ce niveau; pour lui, les représentations mises en machine sont le plus souvent sous forme de chaîne de caractères.

- *Interprétation* : Les interprétations d'une représentation peuvent être multiples. Elles sont faites par l'homme ou par l'instrumentation. Il importe qu'elles soient les mêmes, c'est en fait le but des concepteurs de systèmes informatiques : chercher à ce qu'une instrumentation, convenablement programmée, interprète les données (les représentations) comme le ferait un opérateur humain expert. En reconnaissance de forme, on sera particulièrement intéressé par les identifications.

- *Identification* : Acte de "donner un nom " à une représentation, c'est également le résultat de cette action.

- *Quantité d'information* : Par convention, on appelle quantité d'information le volume de mémoire occupé par une représentation. En fait le concept d'information est un concept flou, mal défini; par contre la quantité d'information, telle qu'on l'entend est parfaitement fixée.

- *Codage, prétraitement* : Un capteur physique transforme un signal en un signal dépendant de l'intensité mesurée. Ce signal continu est numérisé à intervalle régulier ce qui donne des "échantillons" de précision donnée. Il y correspond une "quantité d'information" qui peut être considérable. Un *prétraitement* de ces représentations peut permettre de réduire, parfois considérablement, ces quantités d'informations, qui risquent de saturer rapidement les mémoires des systèmes.

- *Description initiale, symbolique* : L'information recueillie est représentée sous la forme d'une description initiale. L'identification automatique est le but le plus fréquent de la reconnaissance de forme; toutefois on peut souhaiter représenter l'information sous une forme plus facile à analyser par un observateur. C'est l'objectif de l'analyse de données, dont il est parlé dans le chapitre III. Une telle représentation s'appelle une "description symbolique".

- *Interprétation de la représentation d'un objet* : Le terme d'objet doit être pris au sens large : la description initiale peut représenter l'image d'un objet, mais aussi une donnée médicale ou économique....



*Une identification* consiste à donner le même nom à des représentations, c'est à dire à ranger dans une même "classe" des représentations. C'est donc typiquement une opération "ensembliste", au sens mathématique. Une *forme caractéristique* ou simplement une *forme* est le résultat d'une identification partielle. Parfois le terme de "forme initiale" est utilisé pour la description-représentation initiale. Ce qui souligne qu'une telle représentation est obtenue du monde extérieur à l'aide de capteurs physiques: une représentation est déjà une interprétation du monde extérieur. D'une façon générale la description symbolique d'une classe peut être :

- L'élément *le plus représentatif*, tel que par exemple le centre de gravité, ou l'élément le plus proche du centre de gravité;

- Les *n* éléments les plus représentatifs;

- La représentation d'une classe, telle qu'un sous-espace vectoriel par exemple, une représentation géométrique, un dessin;

- Un *concept*, "une représentation mentale de quelque chose perçue par les sens";

- Une *déclaration*, une *expression*, une *proposition*, à laquelle on peut parfois attribuer une valeur de vérité (vrai ou faux, voir annexe).

Ces quelques définitions énoncées, et en résumé on peut dire que le conditionnement résultant de l'observation en perception permet d'obtenir une représentation acceptable des messages perçus, c'est ce que réalisent la plupart des appareils de visualisation et de contrôle tels que moniteurs, cadrans analogiques ou numériques, dérouleurs graphiques, etc. . .

Le passage du conditionnement (représentation) à l'interprétation du message contenu dans le signal (figure (1,4) [6]) et l'extraction de l'information, constituent la tâche la plus critique, quand il s'agit du cas où l'observateur n'est plus un opérateur humain mais un instrument.

Ce passage est obtenu, en partie, par diverses transformations mathématiques :

- transformations fréquentielles (Fourier, Walsh ...),

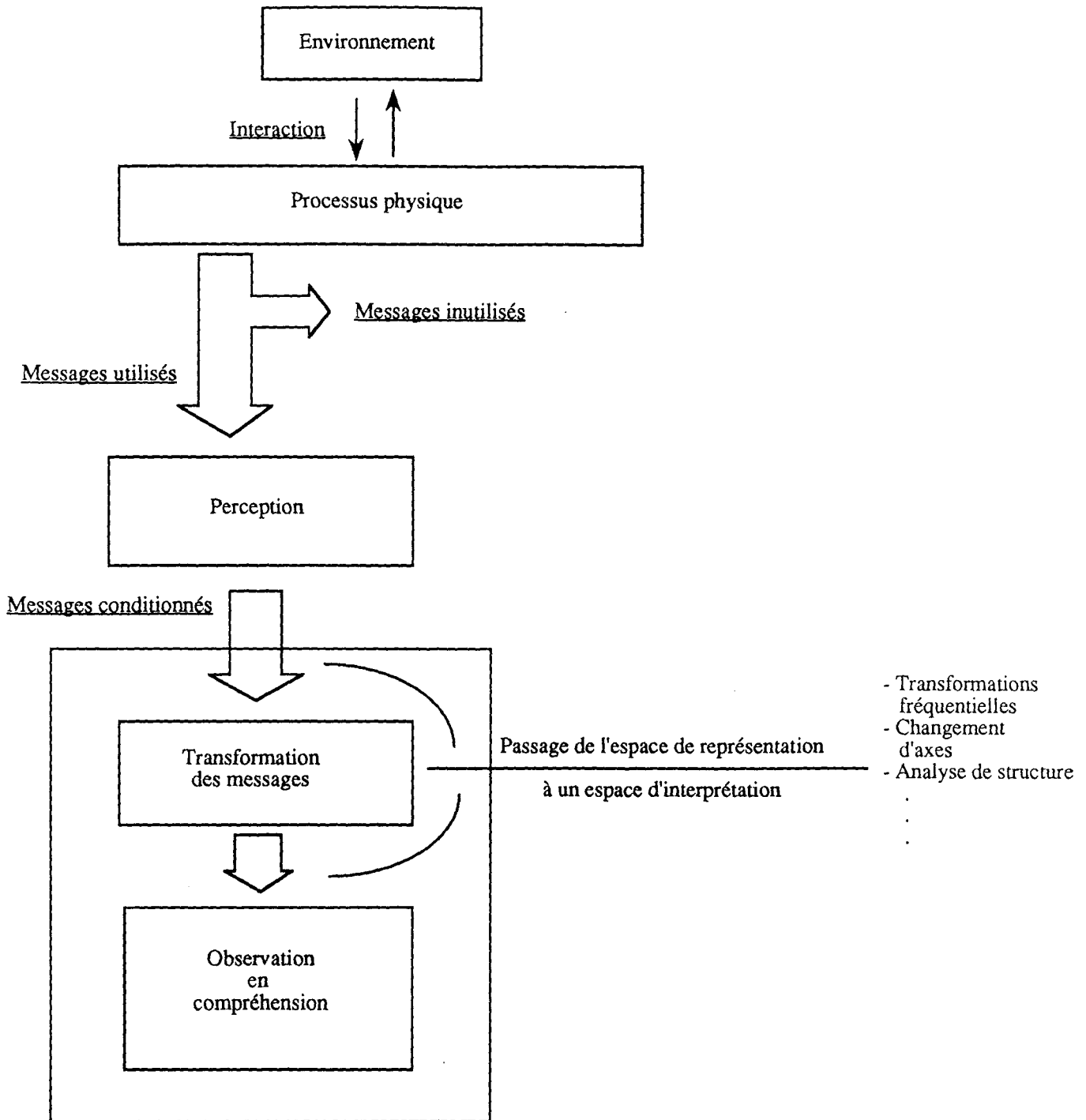


figure (1.4)  
Représentation et interprétation

- extraction de paramètres caractérisant le signal ( nombre de passage par zéro, les amplitudes des extremums, les points de cassures, les rythmes...),

- changement d'axes et recherche des projections principales, etc . . .

## 1 - 8 ASPECTS SEMANTIQUE ET SYNTAXIQUE D'UN PHENOMENE PHYSIQUE

Tout phénomène physique peut être étudié sous deux aspects :

- Celui de sa construction selon des règles parfaitement définies (les phrases sont écrites avec des mots suivant des règles définies). Ceci est l'aspect syntaxique.

- Celui de sa signification, c'est à dire l'information qu'il véhicule (les phrases transmettent des idées). Ceci est l'aspect sémantique.

Le passage à un espace d'interprétation constitue une première forme d'introduction de la sémantique dans la procédure d'observation.

Dans toute opération de reconnaissance l'aspect sémantique du phénomène à reconnaître est important. En effet, on ne va pas étudier de la même manière un signal électroencéphalographique (EEG) et une tension ramenée à zéro à la sortie d'un composant électronique, même si tout deux présentent le même comportement s'ils sont visualisés sur moniteur. Le contexte sémantique des deux cas est différents.

L'interprétation n'est fiable que s'il y a apport d'un "professeur", qui a une bonne connaissance du processus pour entreprendre l'apprentissage de "l'élève", l'élève étant ,dans notre cas, un instrument.

## 1 - 9 PHASE APPRENTISSAGE

L'apport important permettant l'amélioration de la communication processus-opérateur humain est obtenu en faisant appel, donc, à un partenaire ou "professeur". Le rôle de ce professeur est alors de transmettre sa propre connaissance du processus à

l'élève (l'observateur). Cette transmission mettant elle-même en jeu diverses procédures d'apprentissage, figure (1,5).

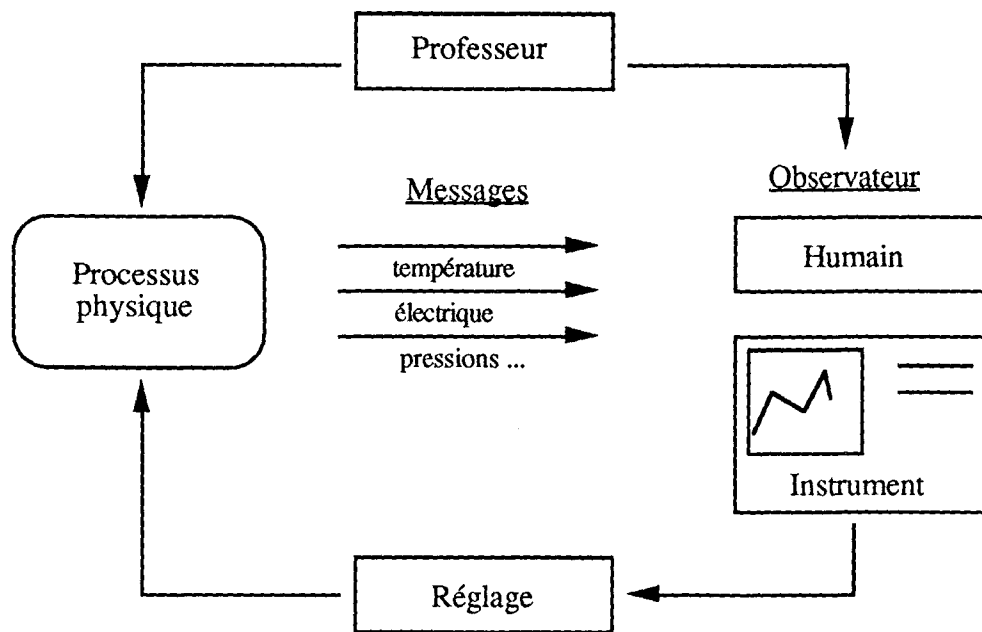


figure (1,5)

Communication processus-observateur

De ce fait, le fonctionnement d'une machine ou d'un instrument doit nécessairement intégrer une phase d'apprentissage. Celle-ci se traduit par une mémorisation des "connaissances", soit sous formes analogiques (enregistrement), soit sous forme numérique (mémoires de microprocesseur).

Le "professeur", ou "expert", doit être encouragé à fournir l'information de la manière qui lui est la plus naturelle; c'est à dire en se servant au maximum des graphiques, des tableaux, des diagrammes, etc....

L'expert ne doit pas être forcé de produire une représentation qu'il n'utiliserait pas ordinairement. Si l'expert est habitué, par exemple, à reconnaître les différents stades de sommeil à partir de l'enregistrement du signal EEG sur papier, il lui serait difficile ou même impossible de les reconnaître à partir d'une représentation mathématique ( histogramme, spectre, etc... ).

L'opération d'apprentissage consiste à améliorer, si possible de façon automatique, les capacités de discrimination entre les différentes classes de formes proposées. Elle repose sur l'exploitation d'un lot de formes judicieusement choisies et servant à l'apprentissage du système.

Une façon de procéder consiste à réserver sur un ensemble de données initiales, une partie à l'apprentissage et l'autre au test du système. Il subsiste toutefois la question relative à la représentativité du lot choisi, figure (1,6).

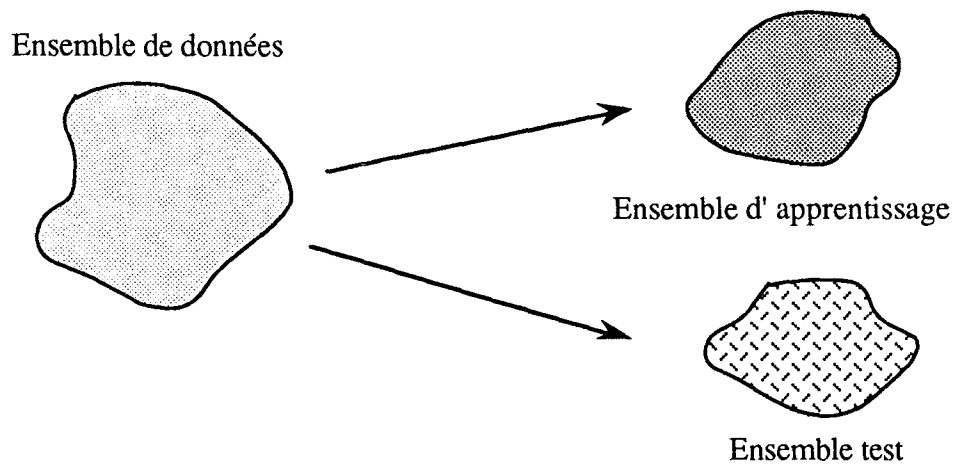


figure (1,6)

Ensembles d'apprentissage

## 1 - 10 RECONNAISSANCE DE FORMES

Un des problèmes essentiels de toute reconnaissance est celui de la classification, c'est à dire de la partition judicieuse de l'ensemble des données à traiter en sous-ensembles cohérents au sens d'un critère qu'il est parfois difficile, pour l'expert, de formaliser de façon immédiate. L'ensemble des données est une représentation mathématique de l'ensemble des signaux physiques à traiter.

Une représentation est le support matériel d'une information. On la note comme la suite suivante :

$$V = (x_1, x_2, \dots, x_i, \dots, x_n)$$

Chaque lettre  $x_i$  représente le résultat d'une mesure, une valeur prise dans un ensemble fini A qualifiée d'*attribut*. Il se peut que nous ayons à considérer une représentation comme une variable V, qualifiée de caractère, notée :

$$V = (x_1, x_2, \dots, x_i, \dots, x_n)$$

L'espace des représentation X est l'ensemble des variables V. Le terme d'espace indique que des lois peuvent opérer sur les variables V, cet espace peut être un espace vectoriel.

L'espace d'interprétation F est un ensemble pourvu de lois d'opération. L'interprétation d'une représentation donne un résultat, par exemple un nom, une forme, un code . . . . L'ensemble des formes F est un espace d'interprétation :

$$F = \{ f_1, f_2, \dots, f_i, \dots, f_n \}$$

L'identification est une application E de l'espace de représentation dans l'espace d'interprétation :

$$\begin{aligned} E : X &\longrightarrow F \\ E : (x_1, x_2, \dots, x_n) &\longrightarrow f_j \end{aligned}$$

Le but de la reconnaissance de forme est d'effectuer cette application de façon constructive, c'est à dire trouver un algorithme qui l'effectue pour toute valeur V.

En d'autres termes, soient :

$M = \{ M_1, M_2, \dots, M_n \}$  l'ensemble des modèles de formes à reconnaître,  $n$  étant le nombre des modèles.

A une forme d'un signal  $S$  à reconnaître;

Il faut donc trouver une fonction discriminante  $R$  qui permette d'affecter la forme  $A$  à un modèle  $M$ , le problème peut se formaliser ainsi :

$$R \circ A = M$$

où "  $\circ$  " est une composition d'opération qui calcule l'effet  $M$  à partir de la cause  $A$  au moyen de la relation  $R$ .

La première étape consiste à retrouver, avec l'aide d'un expert, le nuage de points associés à une même forme, puis à déterminer le prototype caractérisant chaque nuage. Ces prototypes ( modèles ) seront utilisés ultérieurement, au cours de l'étape de reconnaissance d'une forme inconnue évaluée avec les mêmes procédés de mesure ou d'appréciation.

La conception d'une instrumentation qui intègre une phase de reconnaissance, consiste alors, à élaborer un système permettant d'affecter un individu, dont la classe d'origine est inconnue, à l'une des classes déjà répertoriées. Le problème est donc de mettre au point et d'optimiser une procédure de décision prise, en générale, sur des critères de proximité entre l'individu à reconnaître et les nuages représentant chaque classe.

Pour résoudre un problème de reconnaissance de forme, on doit trouver une telle procédure qui réduise la quantité d'information, mais surtout qui coïncide avec les décisions de reconnaissance d'un observateur humain, ceci dans tout l'espace des représentation.

## 1 - 11 CONCLUSION CHAPITRE I

A la fin de ce premier chapitre et en conclusion on dira que même si ce chapitre semble " à part " par rapport aux autres, il nous a semblé nécessaire et même indispensable de l'introduire avant toute description de l'instrumentation réalisée. Les notions introduites dans ce chapitre sont très générales, on essayera le plus souvent possible de les lier à l'aspect pratique de notre application et cela dans les chapitres à venir.



## **CHAPITRE II**

### **APPLICATION A LA RECONNAISSANCE DES STADES DE SOMMEIL ET DE VEILLE CHEZ LE RAT**

## 2 - 1 INTRODUCTION

### 2 - 1 - 1 CAHIER DES CHARGES

Le problème était posé ainsi :

Dans certains travaux de laboratoire, on été amené à mesurer les différents stades de sommeil et de veille, chez le rat, à partir d'un enregistrement EEG. Cette opération devenait pénible quand il s'agissait, pour l'opérateur humain, de lire des enregistrements réalisés sur des durées de plusieurs heures, ou plusieurs jours. En plus de la fatigue que cela induisait, il y avait, au bout d'un certain moment, une " saturation " de l'opérateur, d'où une baisse évidente de fiabilité de l'opération de reconnaissance.

Le but du projet était donc de réaliser, autour d'un micro-ordinateur, une instrumentation capable de remplacer l'expert humain dans cette tâche.

L'intérêt d'une telle instrumentation est que : un instrument ne se fatigue pas (au sens de la fatigue humaine), il n'a pas " d'humeur " et il n'est pas subjectif. Ce dernier point présente, d'autre part, un inconvénient de la part de l'instrument. En effet, dans ce cas d'application, on trouve une grande part de subjectivité et " d'expérience " de la part de l'opérateur humain qui intervient, autant de paramètres difficilement quantifiables au sens physique.

Un instrument ne remplaçant jamais entièrement un être humain, on essaiera modestement de réaliser une instrumentation se rapprochant le plus possible du cheminement de l'acquisition de la connaissance chez l'être humain et dans laquelle l'instrument apporte, entre autres, sa vitesse et sa disponibilité.

### 2 - 1 - 2 PRESENTATION ET DISCUSSION DE QUELQUES EXEMPLES DE SYSTEMES

Depuis des dizaines d'années, la recherche dans le domaine de l'analyse du sommeil a suscité de nombreux travaux [14], que ce soit chez l'Homme [7,13] ou chez le rat [17 à 22].

La grande évolution de la micro-informatique et de la micro-électronique, a permis ces dernières années, la conception de divers systèmes automatiques d'analyse du sommeil avec des approches et des moyens différents. On notera en particulier les analyses du spectre de fréquence de l'EEG [8 à 11] ou même des analyses syntaxiques ou autres [12,15]. La diversité des méthodes utilisées lors de l'élaboration de ces systèmes, et qui restent à notre sens artisanales, ne nous permet pas de reprendre ici la description de la totalité de ces systèmes. Toutefois, on donne l'inventaire des signaux pris en compte pour caractériser l'événement du sommeil.

La mesure des différentes phases du sommeil et de veille chez le rat a été réalisée à partir de l'enregistrement des trois signaux électrophysiologiques suivants : l'EEG cortical, l'EEG hippocampique et l'électromyogramme EMG [18,20,22]. Cependant des travaux ont montré que cette mesure pouvait se faire uniquement avec deux signaux : l' EEG cortical et l' EMG [16,17,19,21].

Les systèmes réalisés avaient un coût élevé ou disposaient d'un équipement informatique vieilli [17 à 21]. En outre, ces systèmes n'étaient pas entièrement automatisés, étant donné que l'opérateur devait déterminer et entrer manuellement des valeurs de seuils pour certains paramètres [17,19 à 22].

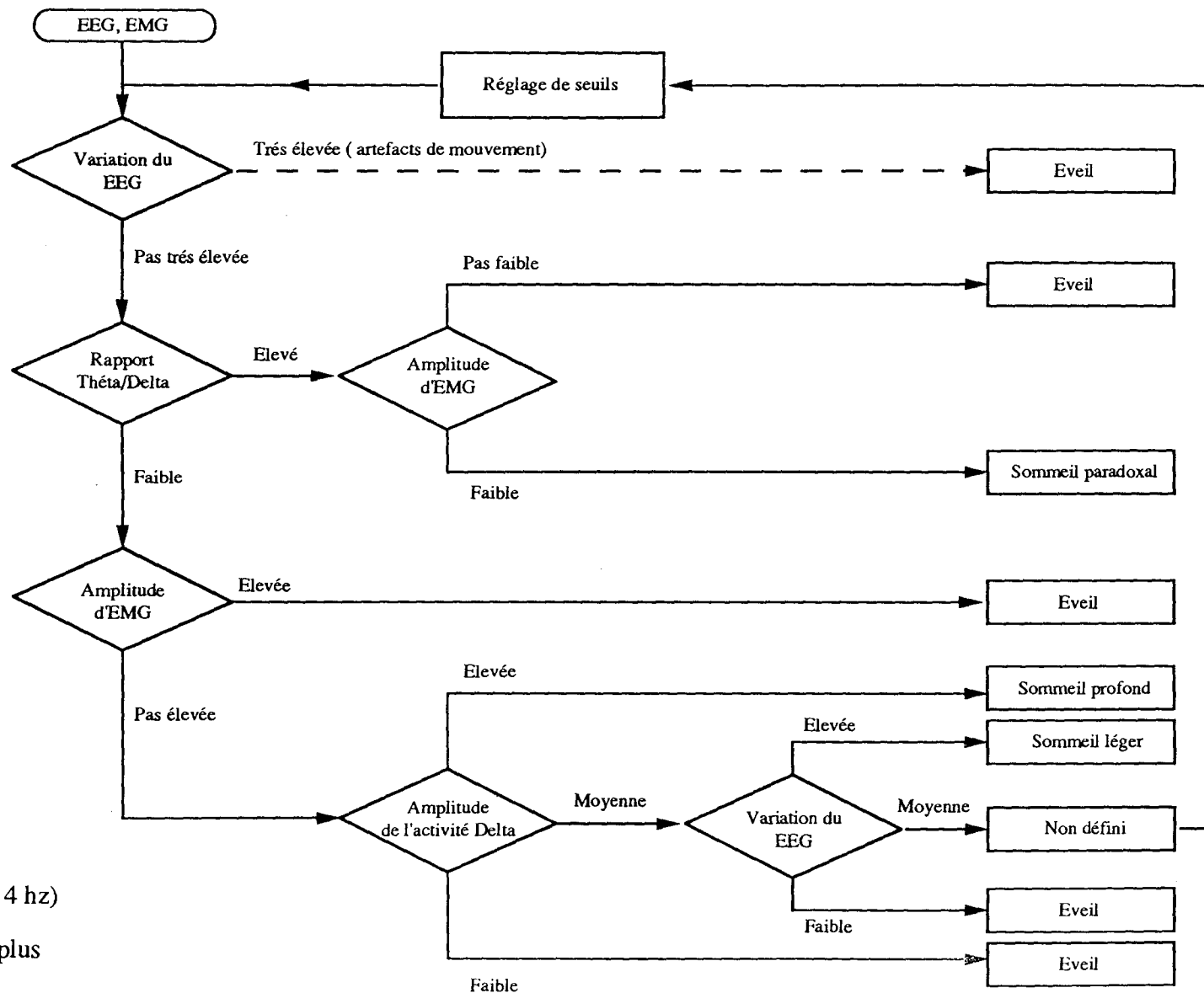
L'organigramme (orga.1) présente un schéma de prise de décision dans la quantification des différents stades chez le rat, quand le signal électromyographique (EMG) est pris en compte. A la lecture de cet organigramme, on conçoit la difficulté d'une telle quantification quand l'EMG n'est pas enregistré, vu le rôle que joue ce signal.

### 2 - 1 - 3 NOUVELLE APPROCHE

A la différence des systèmes existants actuellement, l'approche qu'on va exposer dans ce manuscrit apporte deux éléments nouveaux, au moins, mis à part le coût du système qui est réduit au maximum :

- L' opération de reconnaissance est réalisée, uniquement à partir de la lecture d'un signal EEG à la manière de l'expert humain qui a servi de modèle à la conception de ce système. Disposer d'un enregistrement d' EMG n'étant pas nécessaire.

- Notre ambition était de réaliser un système entièrement automatique, c'est à dire que l'opérateur n'aurait pas à déterminer puis à entrer manuellement et pour chaque rat certains paramètres (seuils, gain, etc...), mais que ces paramètres seraient fixés, une



delta : amplitude des signaux de basse fréquence ( $F \leq 4$  Hz)

théta : amplitude des signaux plus rapides ( $6 \leq F \leq 13$  Hz)

ces amplitudes sont en générale mesurées à la sortie de filtres passe-bande.

(orga.1)

Diagramme de prise de décision pour la classification des différents stades de sommeil et de veille chez le rat

fois pour toute, de manière à répondre au plus grand nombre de cas possible.

Il est évident que ces deux contraintes supplémentaires ne devront pas affecter la fiabilité de l'opération de reconnaissance. On s'efforcera toujours de maintenir le système dans une plage d'erreur tolérée par l'utilisateur, malgré ces contraintes.

L'intérêt de ne prendre en compte qu'un seul signal pour définir une forme ou une information résulte dans le fait que :

- il n'est pas toujours possible de caractériser une forme ou une information à partir de plusieurs sources, en général on ne dispose que d'une source unique ( signal vocal par exemple)

- les difficultés d'ordre pratique sont réduites (préparations des rats, branchement ... )

- le nombre de rat à traiter en parallèle est supérieur dans le cas d'enregistrement d'un seul signal que dans le cas de deux signaux, pour des instrumentations de performances égales.

Dans le chapitre 4, on retrouvera les incidences d'une telle technique surtout sur la séparation des deux stades : éveil et sommeil paradoxal, qui se fait presque toujours sur lecture du signal EMG.

## 2 - 2 MESURE DES PHASES DE SOMMEIL ET DE VEILLE

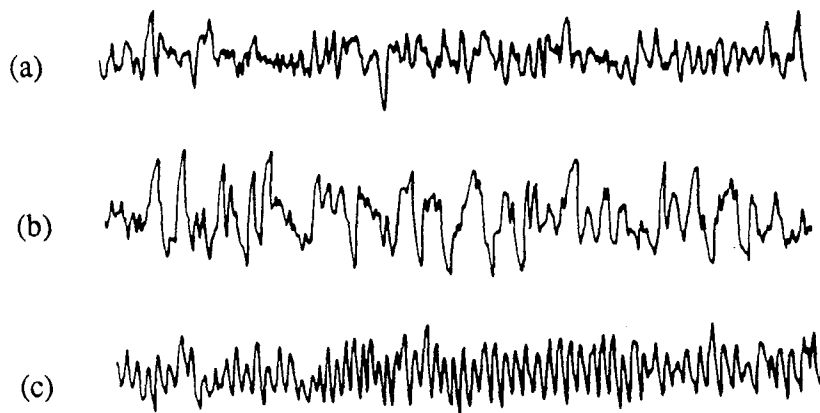
La mesure des différents stades chez le rat, se fait sur des critères électroencéphalographiques [33,34,35], généralement admis et qui se résument comme suit (figure(2,1)):

- L'éveil caractérisé par des ondes irrégulières (8 c/s, ou 8 Hz) et de bas voltage (30 à 50  $\mu$ v). En rapport avec l'activité de l'animal, une activité thêta (8 c/s) peut apparaître, figure (2,1,a) .

- La phase de SOL (sommeil lent), caractérisée par des ondes lentes (3 à 4 c/s) et de grandes amplitudes, figure (2,1,b) , associées à des fuseaux ( 13 c/s ).

- Une phase intermédiaire entre SOL et SP caractérisée par la présence de fuseaux. Les fuseaux ne sont pas visibles sur la figure (2,1) par contre, on les distingue assez bien sur les figures (2,2 a et b) associés au sol

- Une phase de SP (sommeil paradoxale ), caractérisée par l'inscription d'une activité thêta hippocampique (8 c/s). Il s'agit d'une activité très régulière de ( 8 c/s ) et plus ample que l'éveil ( 100  $\mu$ v ). Cette activité est accompagnée d'une atonie musculaire visible sur l'EMG lorsque celui-ci est enregistré, figure (2,1,c).



(vitesse de défilement du papier  $v = 15$  mm/s)

figure (2,1)

Différentes formes d'un signal EEG

Il est très important de noter que les fréquences données ci-dessus ne sont que des fréquences moyenne. En réalité on ne sera jamais en présence d'un signal

propre, régulier et dont la fréquence est constante dans le temps. Cette remarque étant faite, on comprendra mieux les différentes étapes du traitement du signal EEG.

La figure (2,2-a), plus réaliste, montre la succession dans le temps des différents stades de sommeil et de veille. On remarquera, toutefois, que l'activité musculaire très nette sur le tracé de l'EMG correspond à de l'éveil.

Quand à la figure (2,2-b), elle représente une transition SOL-SP-SOL telle qu'elle a été reconnue par l'expert. On notera la différence de niveau entre les deux enregistrements (le haut et le bas) qui représentent le même signal en fonction du réglage de gain sur les deux voies.

## 2 - 3 ESTIMATION DES DIFFERENTS STADES

Dans l'estimation des différents stades de vigilance, on applique les règles suivantes :

- La durée du SOL comprend aussi bien des phases propres de SOL que des phases de transitions SOL-SP (fuseaux intermédiaires).

- Une phase de SP commence juste après la fin de la phase de transition et se termine généralement par un éveil secondaire. Les phases SP de durée inférieure à 10 secondes ne sont pas prises en considération.

- l'éveil est calculé par soustraction du temps passé en SP et en SOL de la durée totale de l'enregistrement.

La quantification de ces différents stades et le dépouillement des enregistrements, se font en général à l'aide de méthodes établies. Toute différence entre deux lectures dépassant un certain seuil (par exemple 3%), nécessite une troisième lecture. D'autre part, les enregistrements qui durent souvent plusieurs jours, sont dépouillés suivant des tranches horaires. Tout cela n'enlève rien à la difficulté de la tâche qui incombe à l'opérateur humain et qu'il se doit d'accomplir [23].

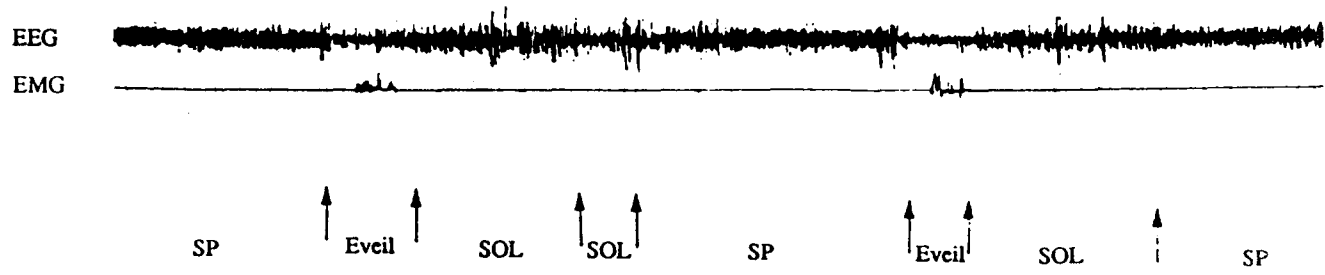


figure (2.2-a)  
 Les différents stades de sommeil et d'éveil détectés sur un enregistrement d'EEG

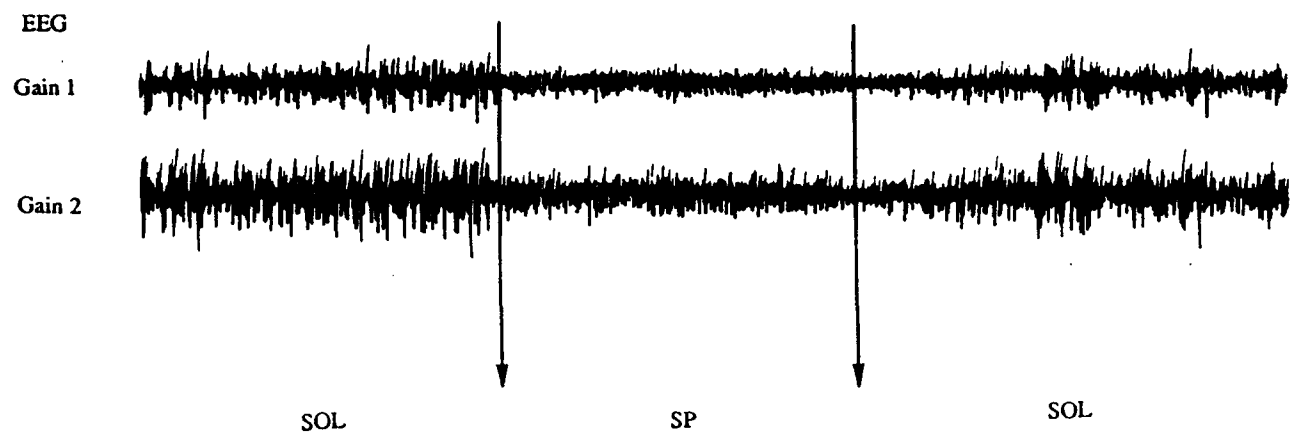


figure (2.2-b)  
 Transitions SOL-SP-SOL

vitesse de défilement du papier  $v = 2,5 \text{ mm/s}$



## 2 - 4 ACQUISITION DU SIGNAL EEG

On devait disposer d'un ensemble de signaux EEG enregistrés sur des rats, afin d'en créer un ensemble d'apprentissage regroupant les signaux représentant au mieux les différents stades de sommeil chez le rat.

La représentativité de l'échantillon recueilli sur l'ensemble des signaux est un élément très important dans la conception de notre système tel qu'il est pensé. Pour remédier à cela on s'est efforcé, dès le départ, à respecter les trois principes suivants :

- faire des enregistrements sur différents rats,
- étaler les enregistrements sur des périodes assez longues,
- ne pouvant pas traiter efficacement tous les enregistrements, un échantillonnage aléatoire est effectué sur l'ensemble des signaux : une procédure simple consiste à attribuer un nombre à chaque élément de l'ensemble, pour sélectionner ensuite les éléments devant constituer l'échantillon au moyen d'une table de nombres aléatoires.

Les enregistrements ont été réalisés au laboratoire d'électrophysiologie de la faculté de chirurgie dentaire à Montrouge.

### 2 - 4 - 1 DESCRIPTION DE L'INSTRUMENTATION

La figure (2,3) schématise l'instrumentation mise en œuvre pour l'acquisition des signaux EEG sur un rat. Auparavant, il fallait une préparation technique du rat permettant de fixer un micro-connecteur sur le crâne du rat et de brancher des électrodes sur des points localisés sur le cerveau [32], pour pouvoir obtenir des signaux EEG dans de bonnes conditions, l'enregistreur papier est doté d'un filtre qui élimine les 50 Hz et un amplificateur qui ramène les niveaux du EEG à l'ordre du 1 V.

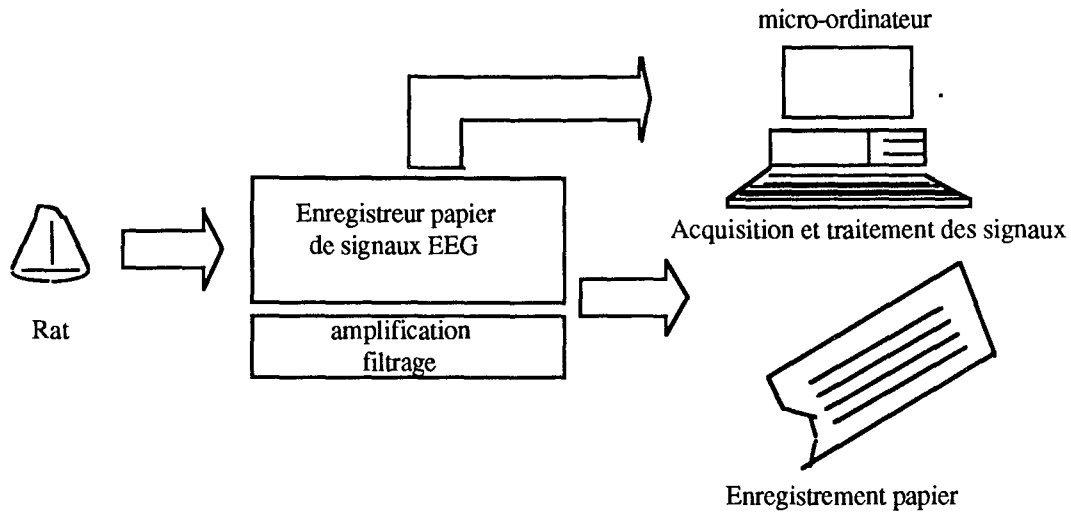


figure (2.3)

Instrumentation pour l'acquisition des signaux EEG

## 2 - 4 - 2 PREPARATION DU RAT

Les rats étaient des rats mâles, adultes de race Wistar (environ 2 mois et 250 à 300 g). Après anesthésie du rat, les étapes de la préparation sont les suivantes :

- 1) mettre le crâne à nu,
- 2) passer la partie supérieure du crâne à l'acide citrique 5%, environ 30 secondes,
- 3) rincer au sérum physiologique,
- 4) percer des trous assez fins sur le crâne pour pouvoir introduire les électrodes d'argent,
- 5) positionner les électrodes sur la dure mère, figure (2,4),
- 6) colmatage des trous à la cire Horsley (cire à trépanation),
- 7) mettre une première couche de résine (super bond) pour l'adhésion du connecteur sur le crâne,
- 8) terminer le chapeau avec de la résine Texton,
- 9) suturer.

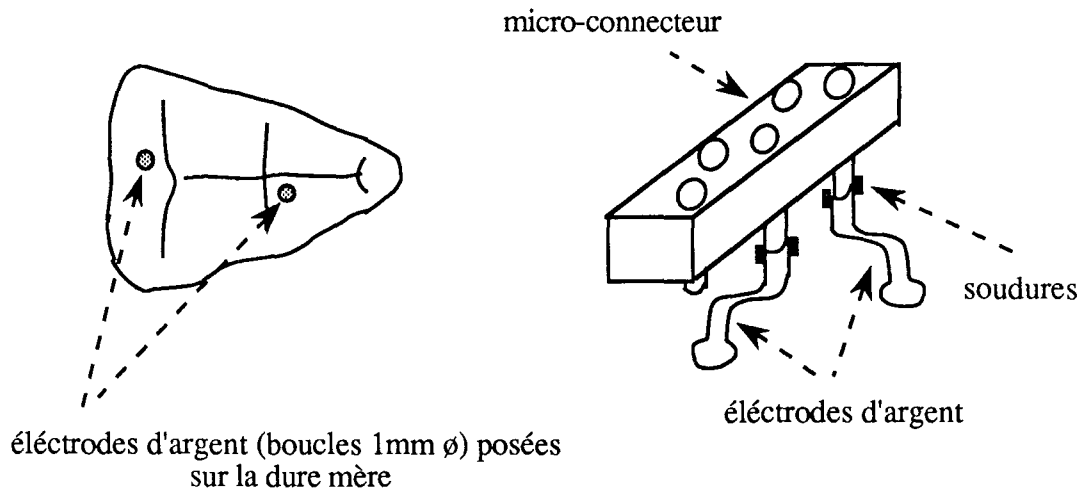


figure (2,4)

Positionnement des électrodes sur le crâne du rat

2 - 4 - 3 CARTE D'ACQUISITION

Bien que le laboratoire d'instrumentation électronique (L.I.E.N) de l'université de Nancy I ait réalisé plusieurs cartes d'acquisition pour différentes applications, dans le travail qu'on devait réaliser, on a utilisé le matériel mis à notre disposition par le laboratoire d'électrophysiologie de la faculté de chirurgie dentaire de Montrouge, entre autre une carte d'acquisition réalisée par " Cambridge Research Systems " : la carte S200. On peut trouver, ces derniers temps, dans toutes les revues spécialisées des études et des réalisations de cartes d'acquisition assez performantes. On ne s'attardera donc pas sur les détails d'une carte d'acquisition.

2 - 4 - 3 - 1 *Description de la S200 :*

La S200 est un système informatique complet, avec sa propre mémoire et ses propres périphériques, permettant son interfaçage avec un autre système informatique "hôte". Les principaux composants de cette carte sont :

a - un micro-processeur intel 80186 avec une horloge de 10 MHz, 3 compteurs/timers à 16 bits, 2 canaux DMA, et un contrôleur d'interruption intégré;

b - 512 Koctets de RAM dynamique pour les programmes et le stockage des données;

c - 64 Koctets d'EPROM contenant les routines de démarrage de la carte et d'initialisation;

d - un Zilog Z8536, unité interface entrées/sorties et compteur;

e - une FIFO Zilog Z8038, pour communication asynchrone rapide entre la S200 et le "host";

f - 4 convertisseurs numérique/analogique à 12 bits;

g - un convertisseur analogique/numérique AD7572 - 5  $\mu$ s à 12 bits, doté d'un multiplexage permettant 8 voies entrées analogiques possibles.

#### 2 - 4 - 3 - 2 *Organigramme de la procédure d'acquisition :*

La procédure d'acquisition peut se résumer par l'organigramme (orga.2).

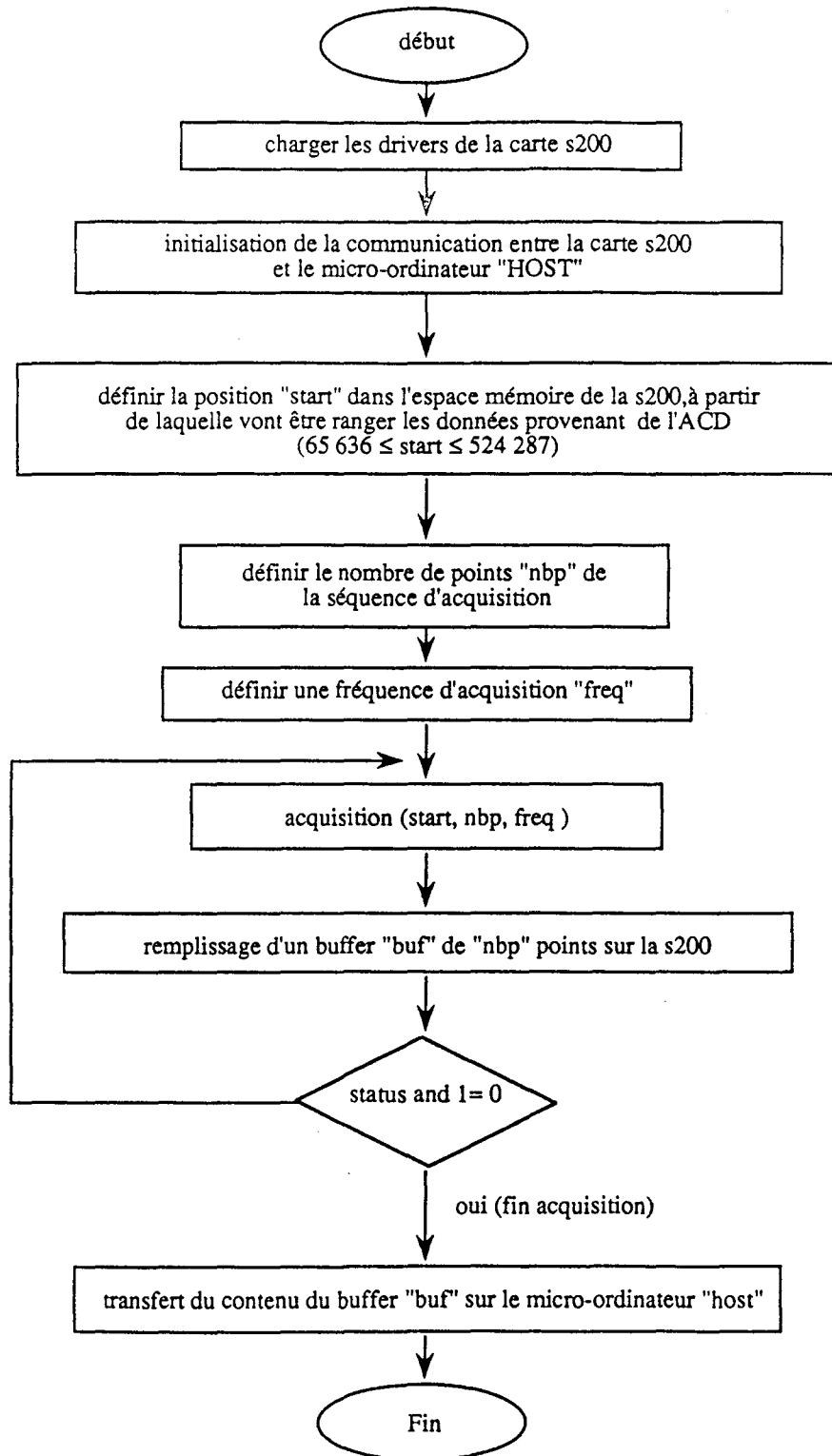
## 2 - 5 TRAITEMENT DU SIGNAL EEG ET EXTRACTION D'INFORMATION

La contrainte du "temps réel", nous a forcé à mettre en œuvre des techniques adaptées à une représentation rapide, mais efficace, du signal EEG qui permettraient d'avoir une information assez complète sur le comportement du signal en fréquence et en amplitude. En effet, utiliser des techniques classiques d'analyse fréquentielle, prendrait un temps de traitement relativement excessif. D'autant plus qu'il faudrait associer toute analyse fréquentielle à un traitement de détection des extremums du signal.

Pour les raisons citées en haut, on a opté pour une technique utilisée, sous une autre forme, dans le traitement automatique de la parole [24,25] et qu'on a adaptée au traitement des signaux électroencéphalographiques.

Bien que l'on ait disposé d'un filtre associé à un amplificateur au niveau de l'enregistreur d'EEG, il nous a semblé prudent d'introduire, avant tout traitement sur les signaux, un filtre numérique passe bas qui éliminerait les fréquences au delà de 13 Hz;

étant donnée que parmi les stades qui nous intéressaient, la fréquence la plus importante était 13 Hz, celle des fuseaux.



Organigramme d'acquisition

(orga.2)

## 2 - 5 - 1 FILTRES NUMERIQUES PASSE-BAS

Tout signal physique provenant d'une instrumentation électronique peut se représenter comme suit :

$$\text{Signal} = \text{Signal utile} + \text{Bruit}$$

Un filtre a pour but de séparer certains signaux utiles, d'autres signaux indésirables, qui leur sont mêlés. La réalisation d'un filtre nécessite donc la connaissance du spectre des fréquences composants le signal utile.

La suppression du bruit peut se faire à l'aide de techniques plus élaborées et plus complexes [26,27,28].

### 2 - 5 - 1 - 1 *Filtre idéal*

Un filtre idéal serait celui qui transmettrait le signal utile sans déformation ni retard, tout en éliminant complètement les signaux indésirables.

Pour chaque filtre, on est amené à définir certains domaines de fréquence transmises sans atténuation, appelées bandes passantes et d'autres pour lesquels l'atténuation doit être élevée : ce sont les bandes coupées.

Les filtres passe-bas ont une bande passante qui s'étend depuis la fréquence nulle jusqu'à une fréquence de coupure  $f_c$ , figure (2,5).

### 2 - 5 - 1 - 2 *Filtre réel*

En pratique, et pour des raisons de temps de calcul qui n'est pas indéfini, il n'est pas possible de concevoir des filtres passe-bas présentant une courbe de réponse semblable à celle représentée figure (2,5). Il est seulement possible de s'en approcher plus ou moins bien.

Les filtres réels présentent trois inconvénients :

- l'atténuation en bande passante n'est pas nulle,

- l'atténuation en bande coupée présente une valeur finie,
- la transition entre les bandes passantes et coupées ne se fait pas d'une manière brusque, mais progressive.

#### 2 - 5 - 1 - 3 *Gabarit d'un filtre réel*

La figure (2,6) montre qu'un gabarit passe-bas est entièrement défini par la connaissance des quatre grandeurs suivantes :

- $A_{max}$ , l'atténuation maximale tolérée en bande passante,
- $A_{min}$ , l'atténuation minimale tolérée en bande coupée,
- $f_p$  et  $f_a$  , fréquences de début et fin de transition respectivement.

Plus le filtre réel se rapproche du filtre idéal et plus les bandes de transition sont étroites, l'atténuation  $A_{max}$  est faible et l'atténuation  $A_{min}$  est élevée.

Il existe une autre grandeur caractéristique d'un gabarit passe-bas, la sélectivité  $K = f_p/f_a$  , qui exprime la raideur de la bande de transition. Plus le filtre s'approche du filtre idéal et plus sa sélectivité se rapproche de 1. Cette grandeur permet d'exprimer le degré de perfection du filtre, et comme on le verra plus en avant, l'ordre  $n$  du filtre.

La courbe de réponse de la fonction de transfert du filtre s'inscrit à l'intérieur du gabarit imposé.

#### 2 - 5 - 1 - 4 *Filtres de TCHEBYSCHEFF*

Les filtres de Butterworth sont optimaux en ce sens que leur réponse est la plus plate possible à l'origine. Les filtres de Legendre sont optimisés pour avoir la coupure la plus raide possible tout en conservant une atténuation continuellement croissante.

Les filtres de Tchébyscheff, quand à eux, sont optimisés de manière à ce que l'atténuation en bande passante oscille le plus grand nombre de fois possible entre 0 et  $A_{max}$  pour un filtre d'ordre donné. Aussi, pour les filtres de Tchébyscheff, l'imperfection que constitue l'atténuation résiduelle en bande passante est uniformément

répartie dans cette bande. Par contre en bande coupée, l'atténuation augmente d'une manière continue.

Les filtres de Tchébyscheff présentent un grand intérêt pratique car, de tous les filtres polynomiaux, ce sont eux qui donnent la coupure la plus raide pour un ordre  $n$  donnée. La figure (2,7) donne à titre d'exemple la réponse comparée de cinq filtres d'ordre 5 [29]:

- Bessel,
- Butterworth,
- Tchébyscheff, atténuation  $A_{\max} = 0.1$  dB,
- Legendre
- Tchébyscheff, atténuation  $A_{\max} = 1$  dB.

On peut constater que la différence est considérable puisque les atténuations respectives (en valeurs absolues) à deux fois la fréquence  $f_p$  sont : 14 dB, 30 dB, 35 dB, 40.5 dB et 45 dB.

Par contre, les filtres de Tchébyscheff n'ont pas une très bonne régularité du temps de propagation de groupe en bande passante et leur comportement transitoire n'est pas aussi bon que celui des filtres de Butterworth ou de Legendre. Ils seront utilisés chaque fois que ce défaut ainsi que l'ondulation en bande passante ne présentent pas d'inconvénients majeurs.

Les courbes de réponse en amplitude et la réponse en régime transitoire des filtres de Tchébyscheff d'ondulation en bandes passante 1 dB et 0.5 dB sont représentées par les figures (2,8-a et b).

#### 2 - 5 - 1 - 5 Conception d'un filtre passe-bas de Tchébyscheff

Parmi les filtres qui ont été calculés, on donne les deux exemple suivants: on désire réaliser un filtre passe-bas de Tchébyscheff satisfaisant au gabarit suivant :

- |                  |                              |                               |
|------------------|------------------------------|-------------------------------|
| - <u>cas 1</u> : | $A_{\max} = 1$ dB            | $A_{\min} = 20$ dB            |
|                  | $f_{np} = 14$ Hz             | $f_{na} = 20$ Hz              |
|                  | $(\omega_{np} = 87.96$ rd/s) | $(\omega_{na} = 125.66$ rd/s) |
|                  | $f_e = 80$ Hz                | $(T_e = 1/f_e = 0.0125$ s.)   |



Le gabarit auquel doit satisfaire le filtre est donc donné par la figure (2,9).

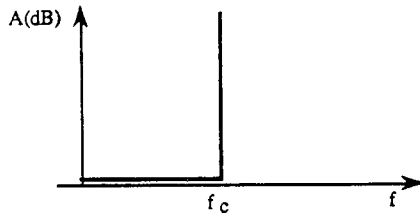


figure (2.5)  
filtre passe-bas

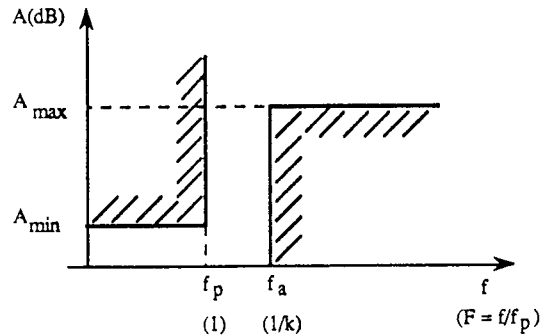


figure (2.6)  
Gabarit d'un filtre passe-bas réel

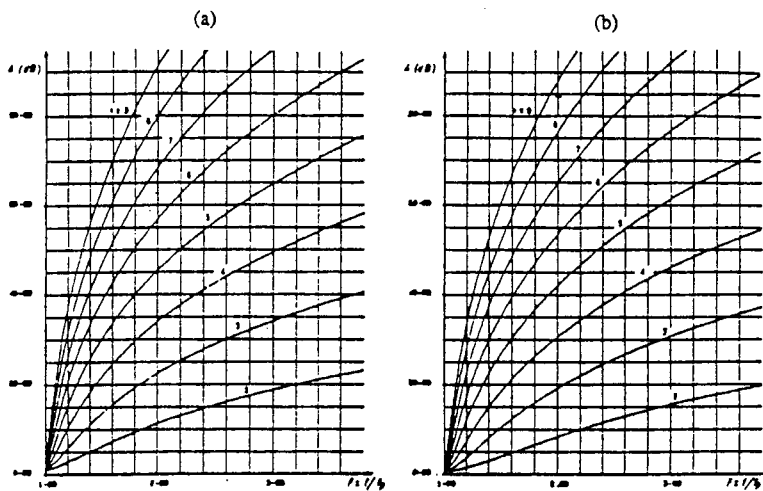


figure (2.8)  
Courbes de l'atténuation en fonction de la fréquence des filtres de Tchébyscheff d'ondulation en bande-passante 1 dB (a) et 0.5 dB (b).

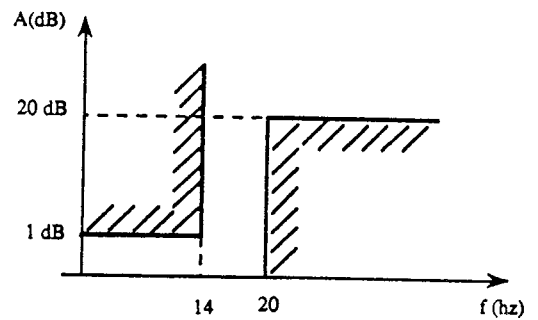


figure (2.9)  
Gabarit du filtre à réaliser dans le cas 1

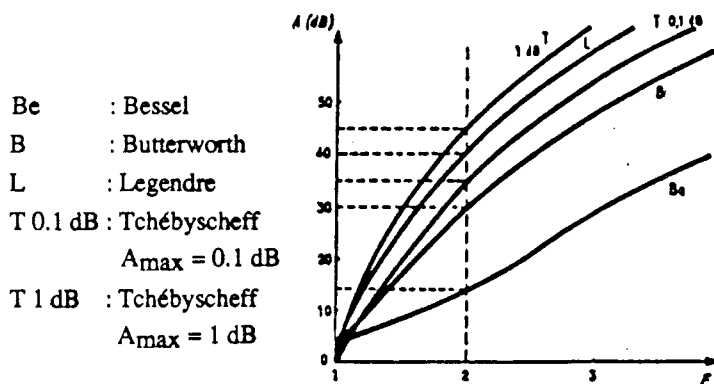


figure (2.7)

Courbes de réponse en amplitude de 6 filtres polynomiaux d'ordre 5

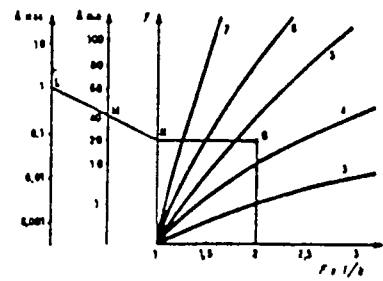


figure (2.10)

Détermination de l'ordre n d'un filtre à l'aide des abaques de Kawakami

Une transposition des fréquences permet de passer des fréquences numériques aux analogiques, qui est en fait le passage du filtre numérique recherché à un filtre lui correspondant dans l'analogique. Ce passage se fait à l'aide de la formule de transposition suivante :

$$\omega_a = \frac{2}{T_e} \cdot \operatorname{tg} \left( \frac{\omega_n T_e}{2} \right)$$

la selectivité  $k$  est donnée par :

$$k = f_{ap} / f_{aa} = \omega_{ap} / \omega_{aa}$$

après calcul, on aura donc :

$$\omega_{ap} = 98.04 \text{ rd/s}$$

$$\omega_{aa} = 159.99 \text{ rd/s}$$

$$k = 0.6127$$

$$1/k = 1.632$$

#### 2 - 5 - 1 - 6 Détermination par abaque de l'ordre $n$ du filtre satisfaisant le gabarit donné :

On pourrait déterminer l'ordre du filtre par calcul mathématique. C'est ce qui a été entrepris pour vérifier mathématiquement tous les filtres qu'on a été amené à utiliser. Dans ce paragraphe on ne développera que la détermination de l'ordre  $n$  du filtre à l'aide des abaques de KAWAKAMI [30], figure(2,10), dont le mode d'emploi est le suivant :

la droite joignant les points  $L$  sur  $A_{\max}$  et  $M$  sur  $A_{\min}$ , rencontre l'axe  $y$  de l'abaque en  $N$ . On cherche le point  $O$  de coordonnées  $N$  et  $1/k$ . Si ce point est situé entre les courbes  $n$  et  $n+1$ , il faudrait au moins un filtre d'ordre  $n+1$  pour satisfaire le gabarit défini par :

$$A_{\max} = L, A_{\min} = M \text{ et } k.$$

Le gabarit défini dans le cas1, impose donc, d'après l'abaque figure (2,11) un filtre d'ordre  $n \geq 4$ .

Le tableau de la figure (2,12), donne alors une fonction de transfert analogique  $H_a(s)$  du filtre d'ordre 4, comme la suivante :

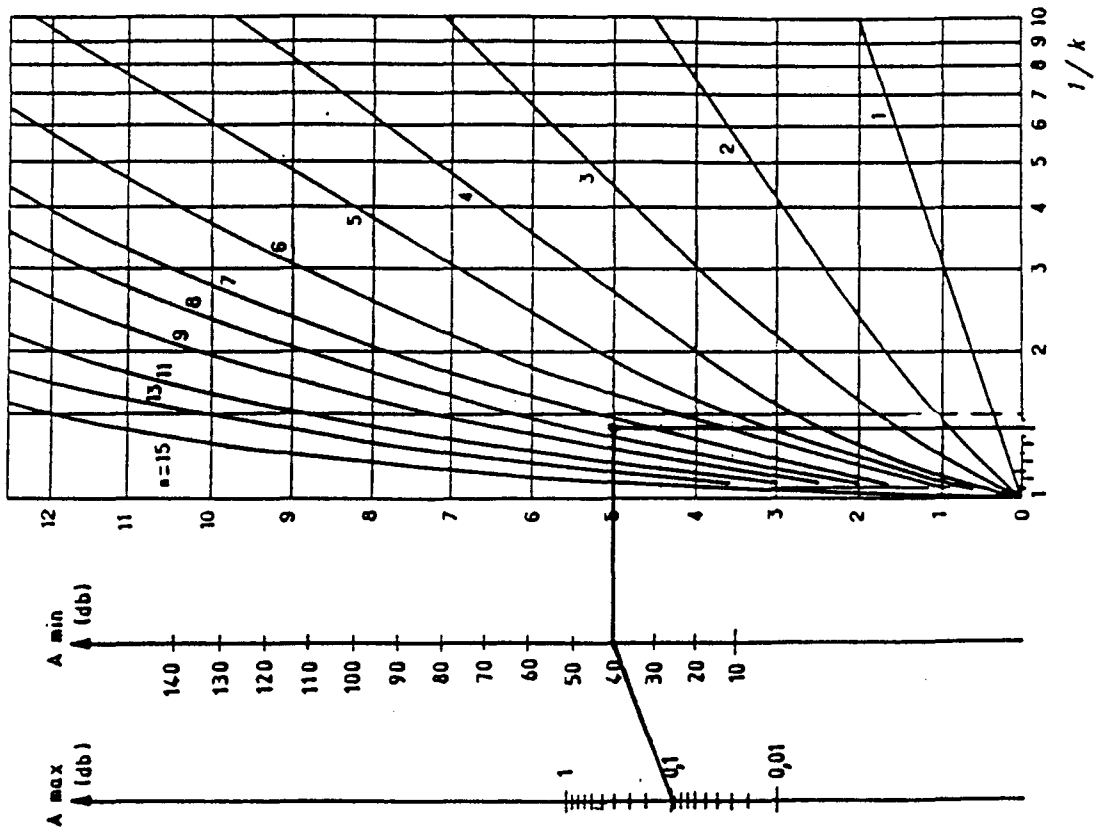


figure (2.11)  
 Abaque permettant de déterminer l'ordre d'un filtre de Tchébyscheff en fonction de la sélectivité  $k$ , l'ondulation en bande passante et l'atténuation minimale en bande coupée

N	CIRCUIT	n	$\epsilon$	$V_m$	$F_m$	FONCTION DE TRANSMISSION
2	1	0.5978	1.8219	1.12	0.707	$(0.9070s^2 + 0.9954p + 1)$
	2	0.2445 2.0233	4.0471	2.08	0.933	$(1.0018s^4 + 0.4970p + 1)$ $(2.0233p + 1)$
4	1	0.1414	7.1466	3.39	0.973	$(1.0136s^4 + 0.2828p + 1)$
	2	1.2056	2.9685	1.01	0.228	$(3.5791s^4 + 2.4113p + 1)$
3	1	0.3455	4.2497	1.49	0.545	$(2.3293s^3 + 1.0911p + 1)$
	2	0.0905	11.1783	5.57	0.986	$(1.0118s^3 + 0.1810p + 1)$ $(3.4548p + 1)$
6	1	1.8408	4.3091	-	-	$(8.0188s^2 + 3.7217p + 1)$
	2	0.0627	18.0820	8.01	0.991	$(1.0093s^4 + 0.1255p + 1)$
	3	0.3046	5.8866	2.23	0.707	$(1.7930s^4 + 0.6092p + 1)$
7	1	0.8030	4.4033	1.40	0.402	$(4.3393s^2 + 1.6041p + 1)$
	2	0.1959	7.8080	3.19	0.787	$(1.5303s^2 + 0.3919p + 1)$
	3	0.0460	21.8775	10.91	0.994	$(1.0073s^2 + 0.0920p + 1)$ $(4.8689p + 1)$
	4	0.4482	-	-	-	$(14.2326s^2 + 5.0098p + 1)$
8	1	2.5049	5.6818	4.29	0.838	$(1.3820s^4 + 0.2755p + 1)$
	2	0.1977	10.0305	2.02	0.544	$(2.9337s^4 + 0.8754p + 1)$
	3	0.4377	6.7022	14.24	0.995	$(1.0058s^4 + 0.0704p + 1)$
	4	0.0352	28.5848	-	-	$(14.2326s^2 + 5.0098p + 1)$
9	1	1.0516	6.6790	1.37	0.312	$(7.0242s^3 + 2.1033p + 1)$
	2	0.2783	8.1830	2.76	0.639	$(2.2401s^3 + 0.5344p + 1)$
	3	0.0278	36.1421	18.03	0.996	$(1.0047s^3 + 0.0354p + 1)$
	4	0.1027	12.1525	5.54	0.873	$(1.2884s^3 + 0.2054p + 1)$ $(6.2748p + 1)$

figure (2.12)  
 Valeurs des éléments et grandeurs de réglages pour les filtres passe-bas et passe-haut de Tchébyscheff d'ondulation 1 dB.

$$H_a(s) = \frac{1}{(1.0136 s^2 + 0.2828 s + 1)} \frac{1}{(3.5791 s^2 + 2.4113 s + 1)}$$

On remarquera que dans le tableau de la figure (2,12) la variable p est déjà réduite, mais en général on note :

$$s = \frac{p}{\omega_{ap}}$$

et la transformation suivante définie comme la transformation bilinéaire :

$$p = \frac{2}{T_e} \cdot \frac{1 - z^{-1}}{1 + z^{-1}}$$

pour retrouver la fonction de transfert du filtre numérique correspondant, il suffit de poser dans  $H_a(s)$  :

$$s = \frac{2}{\omega_{ap} T_e} \cdot \frac{1 - z^{-1}}{1 + z^{-1}}$$

Après calcul on trouvera :

$$H(z) = \frac{0.0166 + 0.0664 z^{-1} + 0.0996 z^{-2} + 0.0664 z^{-3} + 0.0166 z^{-4}}{1.0000 - 1.9963 z^{-1} + 2.1976 z^{-2} - 1.2903 z^{-3} + 0.3548 z^{-4}}$$

$z^{-1}$  représentant un retard on aura donc :

$$y(n) \cdot z^{-k} = y(n-k)$$

cela implique que :

$$y(n) = 1.9963 y(n-1) - 2.1976 y(n-2) + 1.2903 y(n-3) - 0.3548 y(n-4) + 0.0166 x(n) + 0.0664 x(n-1) + 0.0996 x(n-2) + 0.0664 x(n-3) + 0.0166 x(n-4)$$

- cas 2:

$$A_{\max} = 0.1 \text{ dB}$$

$$A_{\min} = 40 \text{ dB}$$

$$f_{np} = 13 \text{ Hz} \quad (\omega_{np} = 81.68 \text{ rd/s})$$

$$f_{na} = 17 \text{ Hz} \quad (\omega_{na} = 106.81 \text{ rd/s})$$

$$f_e = 80 \text{ Hz} \quad (T_e = 1/f_e = 0.0125 \text{ s.})$$

après calcul on a :

$$\omega_{ap} = 89.6 \text{ rd/s}$$

$$\omega_{aa} = 126.13 \text{ rd/s}$$

$$k = 0.710$$

$$1/k = 1.407$$

l'abaque montre que le filtre satisfaisant au gabarit défini dans le cas 2 est d'ordre  $n \geq 8$ . La fonction de transfert analogique sera donc la suivante :

$$H_a(s) = \frac{1}{(6.8675 s^2 + 4.4178 s + 1)} \cdot \frac{1}{(0.9350 s^2 + 0.1196 s + 1)}$$

$$\frac{1}{(1.2517 s^2 + 0.4561 s + 1)} \cdot \frac{1}{(2.4026 s^2 + 1.3103 s + 1)}$$

De la même manière que dans le cas 1, on obtiendra un filtre passe-bas numérique plus sélectif que le précédent.

$y(n) = 4.544 y(n-1) - 10.371 y(n-2) + 14.976 y(n-3) - 14.768 y(n-4) + 10.108 y(n-5)$ $- 4.674 y(n-6) + 1.335 y(n-7) - 0.181 y(n-8)$ $+ 0.00012 [ x(n) + 8 x(n-1) + 28 x(n-2) + 56 x(n-3) + 70 x(n-4) + 56 x(n-5)$ $+ 28 x(n-6) + 8 x(n-7) + x(n-8) ]$
--

## 2 - 6 HISTOGRAMME D'AMPLITUDE ET DES TEMPS DE PASSAGE PAR ZERO

Sur chaque individu on a fait une étude statistique pour extraire des informations sur le nombre de passage par zéro et les temps de passage par zéro  $\Delta t(i)$ , donc une information sur la fréquence, puis des informations sur les amplitudes  $a(i)$  des extremums  $e(i)$ , figure (2,13).

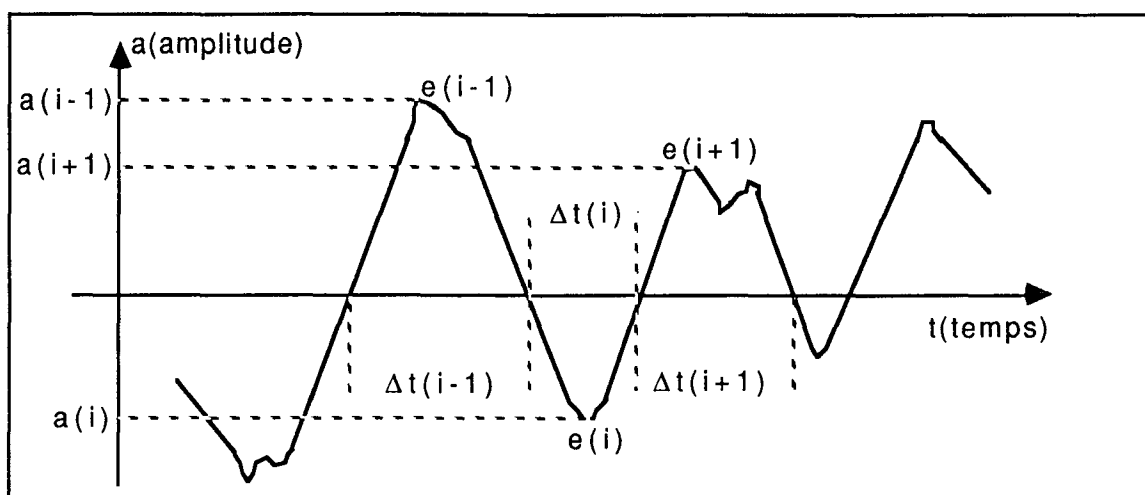


figure (2.13)

### Etude statistique sur le signal EEG

Chaque signal sera donc représenté par un certain nombre de valeurs caractérisant sa fréquence et son amplitude. Pour cela, on va élaborer trois histogrammes:

- le premier est celui des temps de passage par zéro noté : HZC,
- le deuxième est celui des amplitudes d'extremums noté : HEXT,

- le troisième est une juxtaposition de HZC et HEXT, dont la première composante sera le nombre de passage par zéro NBZC. Cet "histogramme" qu'on appellera "hybride" sera noté HEEG. On remarquera que cet "histogramme hybride" n'a aucun sens mathématique, on utilisera le terme "histogramme" pour uniquement une

facilité de langage. Dans le chapitre suivant on parlera de "vecteur" qui est un terme plus correct pour désigner cette représentation

## 2 - 6 - 1 HISTOGRAMME DES TEMPS DE PASSAGE PAR ZERO HZC

Cet histogramme est celui des durées de temps entre deux passages par zéro du signal EEG.

Un passage par zéro est défini quand la relation suivante est vérifiée

$$a(j).a(j+1)<0$$

les  $\Delta t(i)$  sont calculés ainsi , figure (2,14), avec :

$$\text{delt1} = |a(j+1)| / (|a(j)| + |a(j+1)|);$$

$$\text{delt2} = |a(l)| / (|a(l)| + |a(l+1)|);$$

Supposons une suite de points, d'amplitudes  $a(0), a(1), \dots, a(n), a(n+1)$  tels que :

$$- a(0) . a(1) < 0$$

$$- a(i) . a(i+1) > 0 , i=1 \dots n-1 \text{ et } n \geq 2$$

$$- a(n) . a(n+1) < 0$$

on a donc :

$$\Delta t = [ \text{delt1} + \text{delt2} + (n-1) ] . 1/fe$$

$fe$  : étant la fréquence d'échantillonnage du signal EEG,

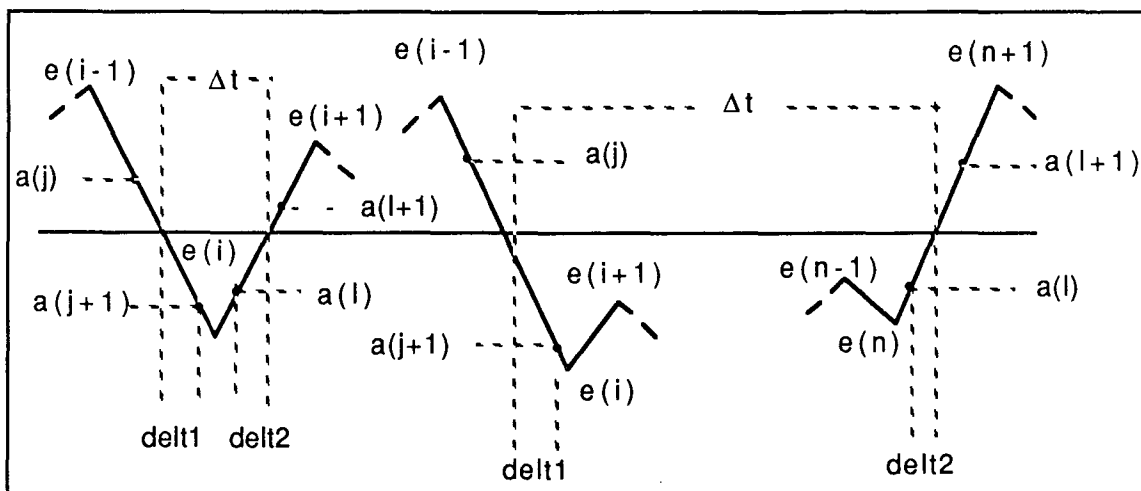


figure (2.14)

Calcul des temps de passage par zéro

On construit donc, l'histogramme HZC de ces temps qui est un vecteur  $V$  à  $n$  dimensions dont chaque composante représente le nombre de temps de passage par zéro du signal appartenant à la  $i$ ème classe ( $i \leq n$ ). On choisira  $n = 5$  pour des considérations qui vont être détaillées dans le paragraphe 2-6-3-1.

Remarquons que cette utilisation des passages par zéro fait perdre toute l'information concernant leur ordre d'apparition. Ces histogrammes ne peuvent donner qu'une image limitée des caractéristiques du signal.

- Le niveau du zéro :

On ne peut pas être certain de la stabilité au cours du temps de la position absolue des amplitudes nulles. C'est à dire que la moyenne des amplitudes avec signe du bruit de fond peut être différente de zéro.

Le problème se pose effectivement avec les équipements électroniques "grand public" et il est très important d'y remédier puisque nous comptons travailler sur les passages par zéro du signal EEG.

On calcule une valeur approchée du niveau zéro du signal par :



$$a_z = \frac{1}{N} \sum_{i=1}^N a_i$$

Dans toute la suite des calculs, on sera obligé de rectifier les valeurs des amplitudes  $a_j$  par :

$$a_{jr} = a_j - a_z$$

RICE [31] a établi des relations liant le nombre moyen de passages par zéro du signal ou de sa dérivé aux moments du spectre de puissance du signal. Les démonstrations de cela sont intéressantes mais assez longue ce qui ne nous permet pas de les reprendre dans ce manuscrit.

## 2 - 6 - 2 HISTOGRAMME DES AMPLITUDES D'EXTREMUMS HEXT

A chaque temps de passage par zéro  $\Delta t$ , on associe l'amplitude  $A$  de l'extremum  $e$ , dont la projection sur l'axe des temps est comprise dans l'intervalle  $\Delta t$ , figure (2,14).  $A$  est défini comme suit :

$$A = \text{Max} ( |a(1)|, \dots, |a(n)| ).$$

## 2 - 6 - 3 CHOIX PRATIQUE DES INTERVALLES POUR CHAQUE HISTOGRAMME

Le théorème de Nyquist dit qu'un signal numérisé contient toute l'information nécessaire à la reconstruction du signal analogique à partir des échantillons du signal numérique, si le signal est numérisé avec une fréquence au moins égale à deux fois la plus grande fréquence présente dans le signal analogique. En pratique, il est toujours nécessaire, de choisir une fréquence d'échantillonnage supérieure à deux fois la plus grande fréquence présente dans le signal.

Le fait de prendre une portion du signal supposé ne contenir qu'une seule fréquence  $f_0$  et d'essayer de retrouver sa fréquence par un comptage des temps de passage par zéro, introduit une erreur  $\Delta f$ .

On veut estimer la fréquence  $f_0$  d'une portion d'onde sinusoïdale échantillonnée avec une fréquence  $f_e = 80$  Hz. Si on compte le nombre d'échantillon  $N$  existant entre 3 passages successifs par zéro du signal numérisé, on aura la relation suivante:

$$f_0 = f_e / N.$$

on démontre facilement donc, que l'erreur en fréquence  $\Delta f$  est donnée par :

$$\frac{\Delta f}{f_0} = \frac{f_0}{f_e}$$

par exemple un signal de fréquence 13 Hz (fuseaux) sera mesuré avec une erreur  $\Delta f$  telle que :

$$- 2.11 \text{ Hz} \leq \Delta f \leq + 2.11 \text{ Hz}$$

Donc les signaux de fréquences comprise entre 10.89 Hz et 15.11 Hz seront estimés comme ayant la même fréquence.

Dans notre cas d'application, les différentes formes de signaux à reconnaître ayant des fréquences différentes, ceci suggère d'utiliser plusieurs fréquences d'échantillonnage en parallèle propres à la détection de chaque forme, afin de minimiser pour chacune d'elles l'erreur d'estimation  $\Delta f$ . Une telle mise en œuvre nécessite le matériel adéquat, on s'est contenté donc d'une fréquence d'échantillonnage unique  $f_e=80\text{Hz}$ , avec une erreur maximale égale à 2.11 Hz correspondant à un signal de fréquence 13 Hz. Les autres signaux qui nous intéressent ne dépassant pas en fréquences les 13 Hz, l'erreur sur l'estimation de celles-ci sera moindre.

#### 2 - 6 - 3 - 1 *Choix des intervalles pour HZC*

On raisonnera dans ce qui suit en nombre d'échantillons  $N$ , ou bien en nombre de pas d'échantillonnage  $\Delta T$ . Dans les deux cas le raisonnement est le même.

L'histogramme des temps de passage par zéro doit fournir une information sur les différentes fréquences suivantes : 4, 8 et 13 Hz. En nombre de points par cycle, ces fréquences correspondent respectivement à : 20, 10, 6.15. On a choisi donc les intervalles suivants centrés sur ces valeurs :

[ 0 , 4 ]

] 4 , 8 ]

] 8 , 12 ]

] 12 , 16 ]

et

] 16 ,  $\infty$  [

à chaque intervalle sont associées, respectivement, les variables suivantes :

DZC1, DZC2, DZC3, DZC4 et DZC5.

### 2 - 6 - 3 - 2 *Choix des intervalles pour HEXT*

Pour définir les intervalles des amplitudes d'extremums, on procède ainsi :

- on détermine les extremums du signal, figure(2,13),

- on sélectionne le maximum des extremums  $e_{\max}$ ,

- on vérifie que ce maximum n'est pas "très" isolé de ses successeurs, sinon on prend celui qui le suit directement. Le terme "très" peut être formulé mathématiquement : Si le maximum détecté est  $\geq 2$  fois l'extremum qui le suit, on prend celui-ci comme maximum. En fait, c'est une précaution qu'on prend pour éviter des extremums dus à des "pic" de tension ou à des parasites qui n'ont rien à voir avec le signal EEG,

- le maximum étant trouvé, il suffit de définir un pas  $\Delta e$

$$\Delta e = e_{\max} / \text{nombre d'intervalles.}$$

On a choisi le nombre d'intervalles égal à 7 sur des considérations expérimentales, après avoir testé 5, 10, ... Si le nombre d'intervalles est égal à 7 et  $e_{\max} = 350$  (unités), les intervalles seront les suivants :

[ 0 , 50 ]  
] 50 , 100 ]  
] 100 , 150 ]  
] 150 , 200 ]  
] 200 , 250 ]  
] 250 , 300 ]

et

] 300 ,  $\infty$  ].

A chaque intervalle est associé une variable, ces variables respectives sont :

EXT1, EXT2, EXT3, EXT4, EXT5, EXT6 et EXT7.

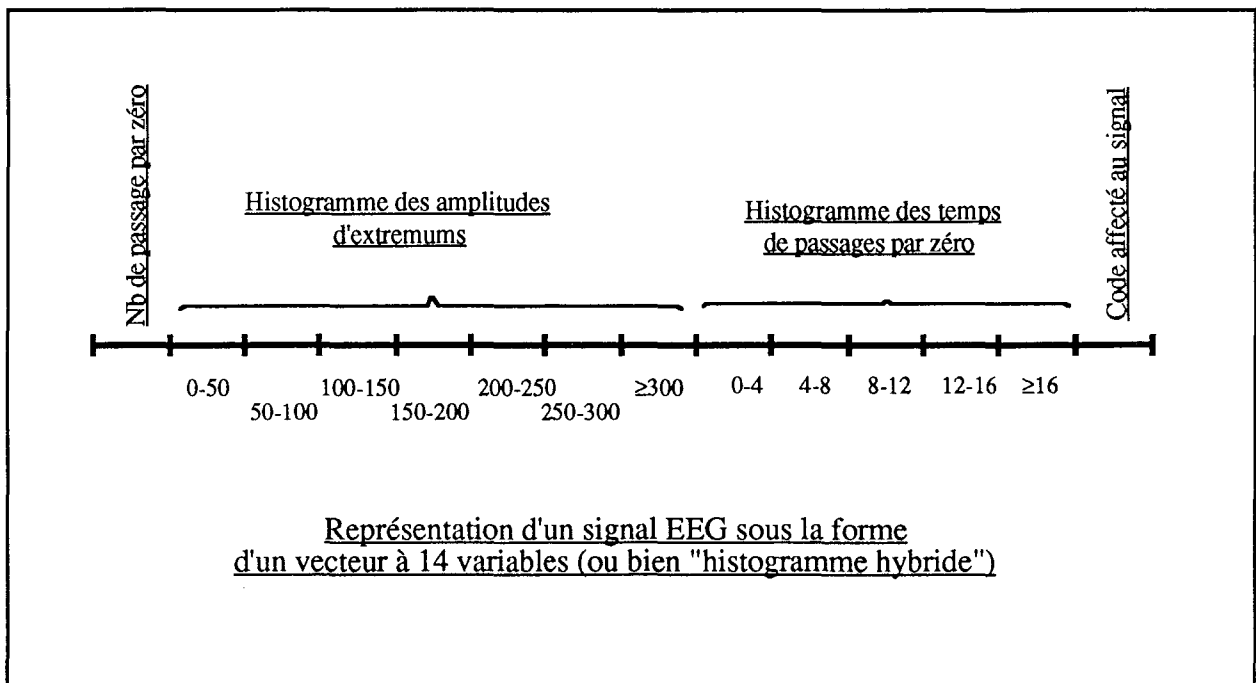
On remarquera après, que l'unité des amplitudes (comme celle des temps de passage par zéro) n'est pas importante, étant donné que le but de ce travail est d'aboutir à une comparaison entre "histogrammes hybrides" dans lesquels l'unité physique n'intervient pas, pour cela il faudrait qu'ils soient normalisés. On abordera le problème de normalisation qui est étroitement lié au réglage de niveau des signaux EEG dans le chapitre IV.

#### 2 - 6 - 4 HISTOGRAMME HYBRIDE HEEG

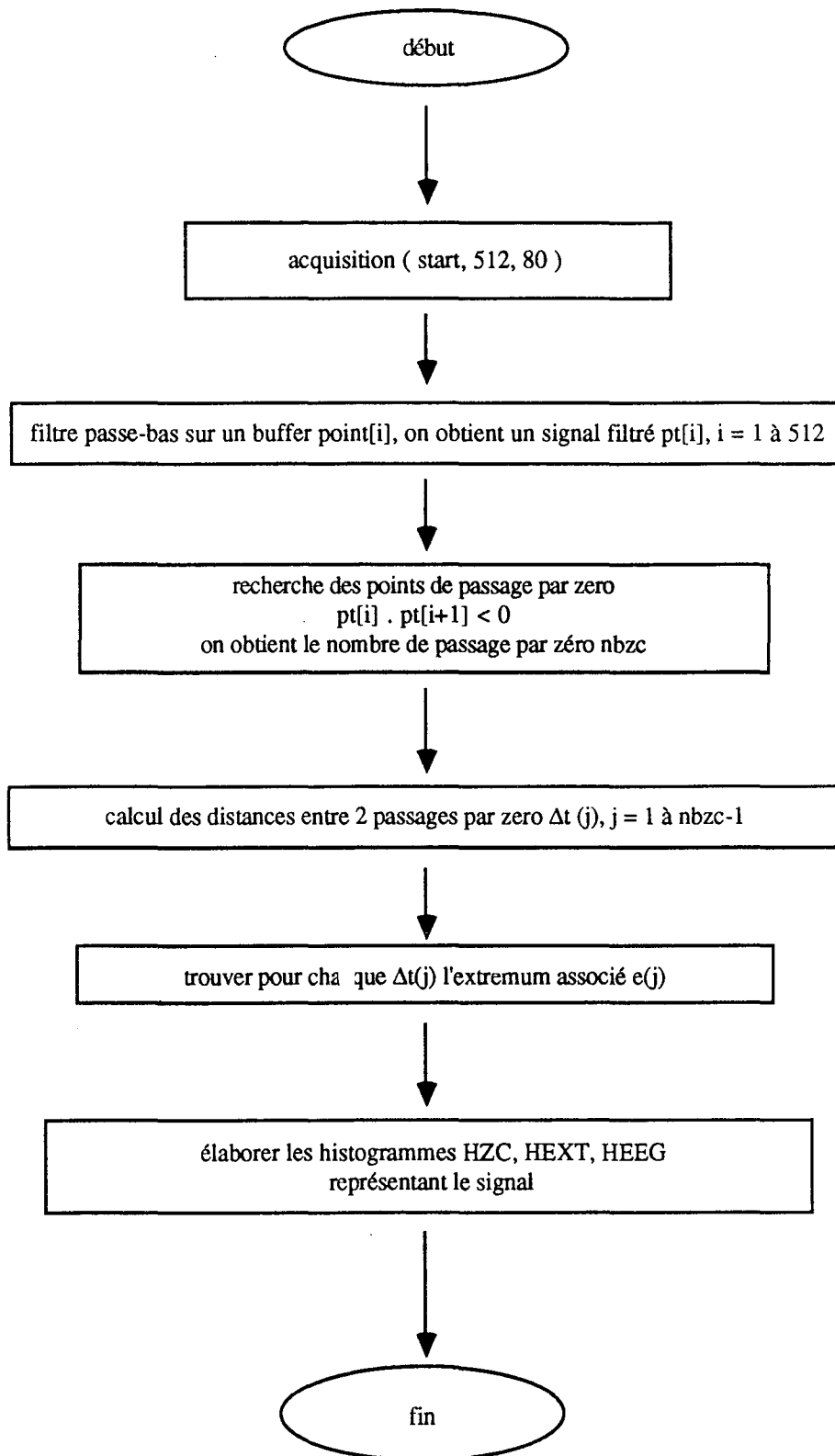
A partir des deux histogrammes des distributions d'amplitudes, et celles des temps de passage par zéro, on fabrique une forme "d'histogramme hybride " de 13 variables. Un individu est caractérisé donc, par un vecteur de 13 variables : la première étant le nombre de passage par zéro, de la 2<sup>ième</sup> à la 8<sup>ième</sup> on retrouve l'information sur les amplitudes, et enfin de la 9<sup>ième</sup> à la dernière l'information sur les temps de passage par zéro.

On peut avancer que cette représentation donne des informations sur la fréquence du signal (nombre de passage par zéro), l'ampleur du signal (amplitude d'extremums) et la régularité du signal (temps de passage par zéro).

La procédure de traitement du signal EEG et de l'extraction de l'information contenue dans celui-ci est résumée par l'organigramme (orga.3). Le résultat d'une telle procédure est repris par le graphique suivant.

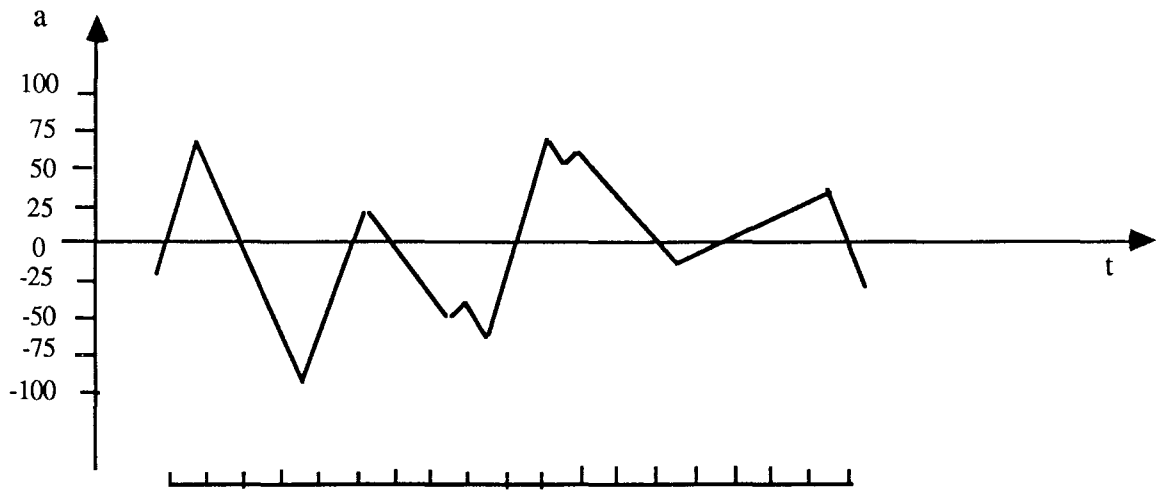


Un signal schématisé par la figure (2,15) sera donc représenté par "l'histogramme", figure(2,16). Pour des facilités d'illustration les intervalles choisis sont différents de ceux décrits précédemment.



(orga.3)

Organigramme du traitement du signal EEG



figure(2.15)  
Un exemple de signal

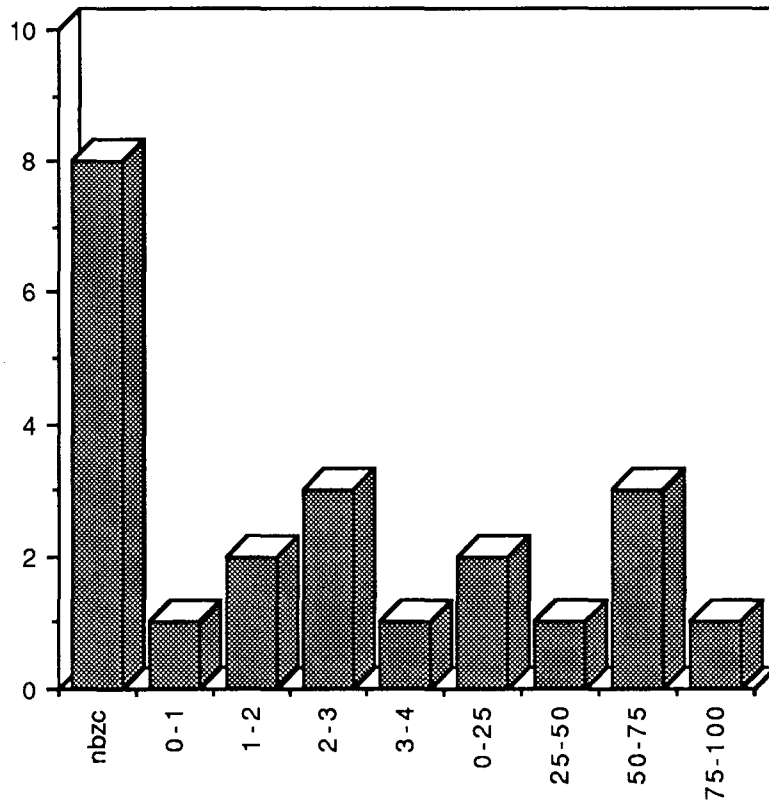


figure (2.16)  
Représentation graphique -après traitement- du signal

## 2 - 6 - 5 CODAGE DES SIGNAUX EEG

Un expert reconnaissait les différents stades, à partir de l'enregistrement sur papier, et les notait. Par conséquent les signaux stockés sur disque, et préalablement divisés en 908 échantillons de 400 points chacun ( 5 secondes ), ont pu être codés de 1 à 5 comme suit :

### Codage des signaux EEG

Signal	fuseaux	Sp	éveil	Sol	non-reconnu
Code	1	2	3	5	4

Dans la catégorie non reconnu, nous avons inclus aussi bien les stades non reconnus par l'expert, que les stades qui ont semblés ne pas être très représentatifs ou sur lesquels il y avait un doute.

## 2 - 7 CONCLUSION CHAPITRE II

Dans la première partie de ce chapitre, on a défini la problématique du projet d'une manière breve, ceux qui voudrons élargir leurs connaissance dans le domaine du sommeil doivent consulter les ouvrages consacrés à l'électrophysiologie.

Dans la deuxième partie, on cherche à obtenir une représentation rapide et efficace du signal EEG. Cette représentation étant obtenue, on abordera dans le chapitre suivant comment obtenir des représentations-modèles pour définir chaque stades de sommeil et de veille.



### **CHAPITRE III**

## **RECHERCHE ET MODELISATION DES SIGNAUX REPRESENTATIFS DE CHAQUE STADE DE SOMMEIL ET DE VEILLE CHEZ LE RAT**

### 3 - 1 INTRODUCTION

Les problèmes de reconnaissance sont étroitement liés à ceux de l'apprentissage. Il faut être capable de produire un ensemble de formes modèles qui serviront à l'apprentissage du système de reconnaissance.

Dans ce chapitre on abordera une phase très importante dans la conception de tout système de reconnaissance de forme : la modélisation ou la représentation judicieuse de chaque stade par une ou plusieurs représentations qui seront la référence dans toutes les opérations de reconnaissance.

### 3 - 2 DEFINITIONS

- Un modèle est une représentation d'un phénomène qui vise à en faciliter la compréhension, en mettant en évidence certains aspects ou certaines parties et on ignorant les autres. Un modèle est d'autant plus imparfait qu'il ignore plus de détails de la réalité.

- La modélisation est l'activité qui consiste à bâtir des modèles. L'activité de modélisation est sans doute aussi ancienne que l'homme lui-même. Modéliser c'est se représenter une chose, le modèle n'est pas la chose mais un résumé de ses propriétés principales. La représentation mentale, le souvenir, la description dans le discours, la règle empirique ou la loi scientifique, la maquette et le prototype, le plan et les spécifications techniques, tout cela relève de l'activité de modélisation.

Cependant pour chaque phénomène à modéliser il existe un multiple jeu de représentation, il revient au concepteur de bien choisir les paramètres à prendre en compte et le mode de représentation adéquat ( voir chapitre précédent ).

### 3 - 3 ANALYSE DE DONNEES

Le grand nombre de données à traiter imposait l'utilisation de techniques puissantes, adaptées au traitement de grands ensembles de données. L'emploi de certaines techniques d'analyse de données fut nécessaire. Tous les traitements, les

description de données, l'analyse, et la classification ont été réalisés sur station de travail APOLLO. Les traitements ont été réalisés grâce au deux système SICLA\* et SPAD\*\* .

*- SICLA est un système interactif de classification automatique réalisé par l'INRIA, permettant la gestion, la description élémentaire et le traitement des données par des méthodes d'analyse multidimensionnelle. Son objectif principal est la diffusion de méthodes issues de recherches récentes, principalement celles concernant la classification automatique et notamment les nuées dynamiques, mais il permet également d'accéder aux méthodes factorielles de la bibliothèque SPAD et aux méthodes de la bibliothèque de programmes MODULAD : bibliothèque Fortran pour l'analyse des données [42].*

*- SPAD est un système d'analyse de données interfaçable au système SICLA.*

Le grand nombre de données traitées : différentes fréquences d'échantillonnage, différents filtrages, différentes représentation du signal, etc..., ne nous permet pas de reprendre la totalité des résultats dans ce manuscrit, on a choisi un ensemble de données représentatif pour illustrer ce chapitre.

Tout en présentant les résultats obtenues par le traitement des données concernant le signal EEG, on introduira certaines notions de base relatives à l'analyse et à la classification automatique des données, dans le but de rendre ces résultats plus significatifs.

### 3 - 3 - 1 PRATIQUE DE L'ANALYSE DES DONNEES :

L'analyse des données est devenue, ces dernières années, une science à part entière, grâce aux progrès du calcul électronique, et à la masse de données croissante provenant de divers domaines et disciplines de recherche, allant des travaux de laboratoires jusqu'aux sondages d'opinion, et qui demandent à être traitées.

Il est certain que ceux qui veulent approfondir leur réflexion sur l'analyse de données doivent lire certains ouvrages consacrés, d'où on a tiré la majorité des définitions présentées ici [36 à 41]. Mais on peut trouver des moyens commodes qui permettent à l'utilisateur praticien que nous sommes d'accéder à une compréhension des principaux instruments de l'analyse des données : c'est ce que nous nous efforcerons de faire.

---

\* SICLA est sous le copyright(c) de l'INRIA, Rocquencourt, 78153 le Chesnay cedex.

\*\* SPAD est un produit du CESIA, 82 rue de Sevres 75007 Paris

Nous avons donc décidé de présenter chaque instrument comme réponse à une question d'ordre pratique, à laquelle nous sommes confrontés. La première question que se pose le praticien est : comment il peut étudier de façon satisfaisante tel type de tableau de données ?, c'est seulement après qu'il peut percevoir les relations que les diverses méthodes ont entre elles, et peut découvrir qu'elles peuvent s'articuler à partir de quelque principe unique.

On a recouru dans certaines démonstrations à des " raisonnements " imagés ou analogiques qui, s'ils n'ont pas de valeur démonstrative, ont le mérite de guider l'intuition dans la mise au point d'un formalisme. Cependant une fois celui-ci introduit, nous le déroulerons avec le plus de rigueur possible.

### 3 - 3 - 2 LES PIEGES DE L'ANALYSE DES DONNEES

Connaissant les méfaits de l'utilisation aveugle, de quelque technique que ce soit, nous avons abordé les instruments de l'analyse de données avec beaucoup de précautions et de réserves. Une telle utilisation peut comporter quatre phases :

- on prend un tableau construit n'importe comment;
- on lui applique n'importe quelle méthode;
- on obtient des " résultats " qui ne veulent rien dire;
- on commente des résultats absurdes dénués de sens, ou alors on exprime des doutes sur l'utilité de l'analyse des données.

Nous verrons par la suite que nous ne nous sommes prononcé sur les résultats de classification que quand ceux-ci ont été concordant avec la classification de l'expert humain. Nous avons en fait, pour plus de rigueur, utilisé deux type d'apprentissage : avec professeur (codage des signaux de 1 à 5), et sans professeur (classification automatique). Les résultats du deuxième concordant avec ceux du premier, nous les avons adoptés.

### 3 - 3 - 3 LES PREMIERES ETAPES D'UNE ANALYSE DE DONNEES

Les premières étapes d'une analyse de données sont essentiellement la définition des " individus " et des " variables ", la construction d'un tableau de données, les descriptions élémentaires de ce tableau à l'aide d'outils que nous détaillerons plus en avant. Pour effectuer ces différentes étapes, l'utilisateur doit faire un certain nombre de choix de base. Il s'agit essentiellement du choix de l'ensemble des " individus " et de l'ensemble des " variables ".

#### 3 - 3 - 3 - 1 *Choix des individus*

Le terme " individu " peut désigner, selon les cas, l'employé d'une entreprise, un client, une forme, un signal ....Il s'agit toujours de l'unité de base sur laquelle l'observateur réalise un certain nombre de mesures.

Le choix de l'individu peut s'imposer de façon plus ou moins évidente au spécialiste; il s'agit le plus souvent d'un échantillon de la population étudiée; on peut espérer que cet échantillon soit représentatif de cette population, relativement aux tendances principales dans le comportement qu'on désire dégager; mais c'est une condition qu'on ne peut contrôler. Il se produit parfois que l'ensemble des individus traités soit exhaustif, mais dans les cas les plus fréquents y compris le nôtre, ce n'est qu'un échantillon de population plus vaste.

On ne se tardera pas longtemps sur la façon dont les individus ont été recueillis étant donné que la majorité des méthodes d'analyse de données ne nécessitent pas d'hypothèse sur cela et ont pour objectif essentiel de traiter les données telles qu'elles sont fournies.

#### 3 - 3 - 3 - 2 *Définition et choix des variables*

A chaque paramètre choisi par l'utilisateur pour décrire les individus, on peut associer une ou plusieurs variables. Le choix des variables est étroitement lié au problème posé. C'est au spécialiste que revient cette tâche. Cependant, l'analyse des données peut permettre d'en vérifier la validité en faisant ressortir, par exemple les redondances et la pertinence de certaines variables.

Si on note  $\Omega$  l'ensemble des individus, une variable est définie par :

1°) Une application  $v$  de  $\Omega$  sur un ensemble d'arrivée  $O$ ,

2°) Une structure algébrique  $S$  sur  $O$  pouvant se transporter sur  $\Omega$  par  $v$ .

On donnera le même nom à l'application  $V$  et à la variable  $v$  par le schéma suivant :

$$\begin{array}{c} v \\ \Omega \longrightarrow O \text{ muni de la structure } S. \end{array}$$

### 3 - 3 - 4 DEFINITIONS ELEMENTAIRES

#### 3 - 3 - 4 - 1 *Les différents types de variables*

Les variables, appelées aussi caractères, peuvent être quantitatives ou qualitatives.

##### a) *Les variables quantitatives :*

Une variable est quantitative quand elle prend ses valeurs sur une échelle numérique, plus précisément, une variable est quantitative lorsque l'ensemble des valeurs qu'elle prend sur les individus est inclus dans l'ensemble des nombres réels  $R$ , c'est à dire que l'ensemble d'arrivée  $O = R$ . Sur ces variables, on peut effectuer les opérations algébriques habituelles : addition, multiplication par une valeur constante, calcul de moyenne, etc...

Dans la pratique, on distingue les types de variables quantitatives suivants:

- quantitatif mesurable (taille, poids, salaire ...);
- quantitatif d'ordre (rang, note ...);

- quantitatif de comptage (fréquence, contingence ...);

-quantitatif binaire (lourd-léger, présence-absence ...);

b) *Les variables qualitatives :*

L'ensemble d'arrivée  $O$  est fini. Les éléments de  $O$  sont appelés modalités de la variable. Une variable est dite qualitative lorsqu'elle prend des modalités non numériques.

Les modalités d'une variable qualitative peuvent être ordonnées (niveau hiérarchique, niveau de satisfaction ...), on dit alors que la variable est qualitative ordinale. Sinon, on dit qu'elle est qualitative nominale (couleur, sexe, lieu géographique...), ou bien qualitative textuelle (nom d'auteur, titre de film ...). On remarquera que sur des variables qualitatives les opérations algébriques n'ont aucun sens.

### 3 - 3 - 4 - 2 *Le tableau de données*

Un tableau de données est un tableau rectangulaire qui se déduit de la définition de l'ensemble des individus et des variables. Soit  $n$  le nombre d'individus et  $p$  le nombre de variables, le tableau de données relatif aux définitions du précédent chapitre sera représenté par le tableau suivant. On y retrouve les échantillons de signaux mentionnés précédemment, ainsi que les caractères s'y rapportant.

		VARIABLES					
		NBZC	EXT1	..	VAR j	..	DZC5
INDIVIDUS	1	$x_{11}$	$x_{12}$	..	$x_{1j}$	..	$x_{1p}$
	2	$x_{21}$	$x_{22}$	..	$x_{2j}$	..	$x_{2p}$
	3	$x_{31}$	$x_{32}$	..	$x_{3j}$	...	$x_{3p}$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	i	$x_{i1}$	$x_{i2}$	...	$x_{ij}$	...	$x_{ip}$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	n	$x_{n1}$	$x_{n2}$	...	$x_{nj}$	...	$x_{np}$

Tableau de données

On note  $\{ \omega_1, \dots, \omega_n \}$  l'ensemble  $\Omega$  des individus et  $\{ v_1, \dots, v_p \}$  l'ensemble  $V$  des variables, le tableau des données associé est donc :

$$X = ( x_{ij}; i = 1, \dots, n; j = 1, \dots, p ) \quad \text{où} \quad x_{ij} = v_j(\omega_i) .$$

Chaque individu est considéré comme un élément de  $\mathbb{R}^p$  appelé espace des individus. La  $i$ ème ligne de la matrice  $X$  représente l'ensemble des  $p$  valeurs prises par le  $i$ ème individu. A chaque individu  $\omega_i$  de  $\Omega$  est associé le vecteur  $x_i$  :

$$x_i = \left\{ \begin{array}{c} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{array} \right\}$$



De même est introduit l'espace des variables  $R^n$ . Chaque variable  $v_j$  est associée à une colonne de  $X$  définie par le vecteur  $x^j$  :

$$x^j = \begin{Bmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{Bmatrix}$$

### 3 - 3 - 4 - 3 Quelques éléments descriptifs des variables quantitatives

Les éléments descriptifs de chaque variables sont :

- le *minimum* :  $m_j = \min_i x_{ij}$

- le *maximum* :  $M_j = \max_i x_{ij}$

- l'*étendue* :  $M_j - m_j$

- si la variable  $x$  prend les valeurs  $x_1, \dots, x_i, \dots, x_n$ , la *moyenne* est :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

si les observations sont munies d'un poids  $p_i > 0$  avec  $\sum_{i=1}^n p_i = 1$  on a :

$$\bar{x} = \sum_{i=1}^n p_i x_i$$

- pour quantifier la dispersion des valeurs, on utilise la *variance* :

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{ou} \quad s^2 = \sum_{i=1}^n p_i (x_i - \bar{x})^2$$

- l'*écart-type* est caractéristique de la dispersion autour de la moyenne :

$$s = \sqrt{s^2}$$

- la covariance entre les caractères x et y est :

$$s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})$$

ou, lorsque les individus sont pondérés :

$$s_{xy} = \sum_{i=1}^n p_i (x_i - \bar{x}) \cdot (y_i - \bar{y})$$

- on montre alors que le coefficient de corrélation r s'obtient par :

$$r(x, y) = \frac{s_{xy}}{s_x s_y}$$

cette corrélation s'interprète comme le cosinus de l'angle des deux vecteurs x et y dans l'espace des variables. La corrélation est un paramètre qui permet de juger de la dépendance entre les différentes variables

- l'ensemble des coefficients de corrélation est regroupé dans la matrice de corrélation R, dont les termes diagonaux valent 1 puisque  $r(x, x) = 1$  :

$$R = \begin{pmatrix} 1 & r_{12} & \dots & r_{1p} \\ & 1 & \dots & \dots \\ & & \dots & \dots \\ & & & 1 \end{pmatrix}$$

R est une matrice d'ordre p, symétrique car  $r_{ij} = r_{ji}$ . On pourra donc se contenter d'écrire seulement la moitié des termes de ces matrices.

### 3 - 3 - 4 - 4 Centre de gravité d'un nuage de points

On résume chacune des p variables par sa moyenne et son écart type. L'individu, en général fictif, dont les variables auraient pour valeurs les moyennes respectives des p variables, est appelé le centre de gravité g du nuage :

$$g = (\bar{x}^1, \bar{x}^2, \dots, \bar{x}^p)$$

### 3 - 3 - 4 - 5 Inertie d'un nuage de points par rapport à un point

L'espace  $RP$  étant muni d'une métrique euclidienne  $d$ , on appelle inertie du nuage de points  $N(\Omega)$  par rapport à un point  $a$  de  $RP$  la quantité:

$$I_a = \sum_{i=1}^n p_i d^2(x_i ; a)$$

les  $p_i$  étant les poids respectifs des points du nuage  $N(\Omega)$ .

Si  $g = \sum_{i=1}^n p_i x_i$  est le centre de gravité du nuage  $N(\Omega)$  on a:

$$I_a = I_g + \left( \sum_{i=1}^n p_i \right) d^2(a ; g)$$

$I_g$  est l'inertie totale du nuage  $N(\Omega)$  et sera notée  $I$ .

### 3 - 3 - 4 - 6 Inertie interclasse et inertie intraclasse

soient  $I_1, I_2, \dots, I_k$  les inerties de chaque classe par rapport à leurs centres de gravité respectifs  $g_1, g_2, \dots, g_k$  et notons  $P_1, P_2, \dots, P_k$  les poids de ces classes :  $P_j$  est la somme des poids des individus de la classe  $n^o j$ . La moyenne de ces inerties est appelée *inertie intraclasse* et est notée  $I_W$  :

$$I_W = \sum_{j=1}^k P_j I_j$$

Considérons maintenant l'ensemble des  $k$  centres de gravité  $g_1, g_2, \dots, g_k$  leur dispersion autour de  $g$ , centre de gravité du nuage total des  $n$  individus, est appelée *inertie interclasse* et est notée  $I_B$ :

$$I_B = \sum_{j=1}^k P_j d^2(g_j ; g)$$

$I_W$  et  $I_B$  sont reliées par une importante relation généralisant le théorème de HUYGHENS :

$$I_W + I_B = I$$

où  $I$  est *l'inertie totale* du nuage des  $n$  points.

### 3 - 3 - 5 CLASSES ET CLASSIFICATION AU SENS DE L'INERTIE

Si on peut considérer les individus comme des points d'un espace euclidien, le problème de la classification peut se définir comme la recherche d'une partition d'un nuage de  $n$  points en  $k$  sous-nuages. La dispersion d'un nuage de points étant caractérisée par son inertie, une classe sera donc d'autant plus homogène que son inertie sera faible. Il est donc souhaitable que l'inertie intraclasse  $I_W$  soit la plus faible possible pour avoir un ensemble de classes très homogènes. Pour avoir une bonne séparation des classes il conviendra que l'inertie interclasse  $I_B$  soit la plus grande possible.

Rendre maximale  $I_B$  est donc équivalent à rendre minimale  $I_W$  puisque leur somme  $I$  est constante. Si l'on considère l'inertie comme critère de classification, il suffirait donc de caractériser les meilleures partitions possibles en  $k$  classes comme celles qui rendent minimale  $I_W$ . La meilleure partition possible est celle où chaque individu constitue une classe car dans ce cas là  $I_W = 0$  puisque chaque point est confondu avec le centre de gravité de sa classe.

### 3 - 3 - 6 QUELQUES INSTRUMENTS DE L'ANALYSE DE DONNEES

Sans trop rentrer dans les détails théoriques, on présente ici deux des importants instruments de l'analyse des données que nous avons eu l'occasion de rencontrer dans nos traitements, et qui sont : l'analyse en composantes principales et les nuées dynamiques.

#### 3 - 3 - 6 - 1 *L'analyse en composantes principales*

L'analyse en composantes principales (ACP) est la plus ancienne des méthodes de l'analyse des données, sa mise au point par HOTELLING date de 1933. On

aura souvent intérêt à réaliser sur les données une analyse en composantes principales qui fournira de premières indications. L'analyse en composantes principales présente donc encore un réel intérêt pratique, même si elle n'apparaît plus comme la seule méthode possible d'analyse factorielle, ni même comme la plus performante.

Cette méthode a pour objet la description des données contenues dans un tableau individus-caractères numériques :  $p$  caractères sont mesurés sur  $n$  individus.

a) *Présentation de la méthode :*

Lorsqu'on est en présence de deux caractères (ou variables)  $x_1$  et  $x_2$ , il est facile de représenter, sur un graphique plan, l'ensemble des données. Chaque individu  $e_i$  est alors un point de coordonnées  $x_{i1}$  et  $x_{i2}$  et un simple examen visuel de l'allure du nuage permet d'étudier l'intensité de la liaison entre  $x_1$  et  $x_2$  et de repérer les individus ou groupes d'individus présentant des caractéristiques voisines, figure (3,1) :

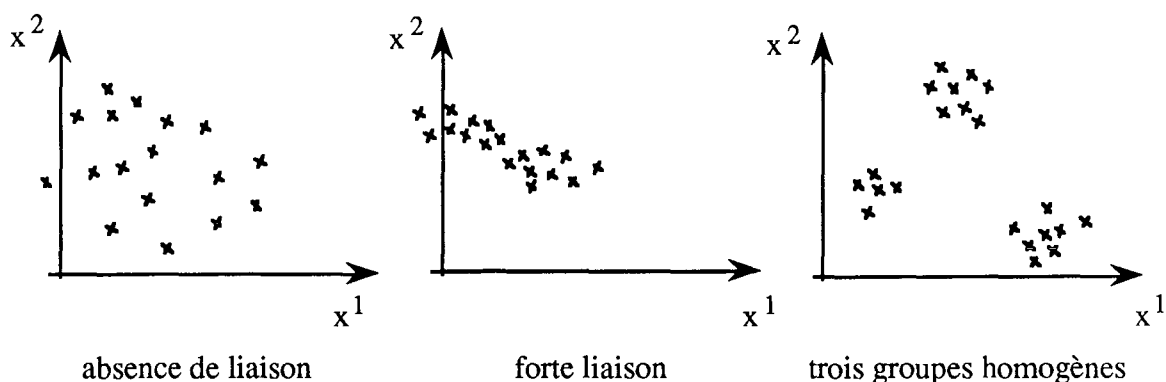


figure (3,1)

représentation d'un nuage de point dans un espace à deux dimensions

S'il y a 3 caractères, l'étude visuelle est encore possible en faisant de la géométrie dans l'espace. Mais dès que le nombre  $p$  de caractères devient supérieur ou égal à 4, cela devient impossible.

Dans notre cas, comme nous l'avons vu précédemment, chaque individu est représenté par 12 caractères (mis à part le 13ième qui est qualitatif et auquel on ne s'intéresse pas pour le moment). Les individus, plus de 900, forment un nuage dans un

espace à 12 dimensions, puisqu'il y a 12 coordonnées. Il est bien évident qu'une représentation d'un tel espace ne serait pas concevable pour nous si ce n'est par des projections dans des plans.

Si l'on veuille représenter les individus sur un graphique plan. Ce que l'on verra sur le dessin ne sera qu'une représentation déformée de la configuration exacte : l'opération de projection déformant ainsi les distances entre individus. Il y a des distorsions qu'il faudrait rendre minimum.

b) *Géométrie de l'ACP :*

Le graphique s'obtient en projetant les points individus  $e_1, e_2, \dots, e_n$  sur un plan comme le montre la figure ci-dessous, figure (3,2) :

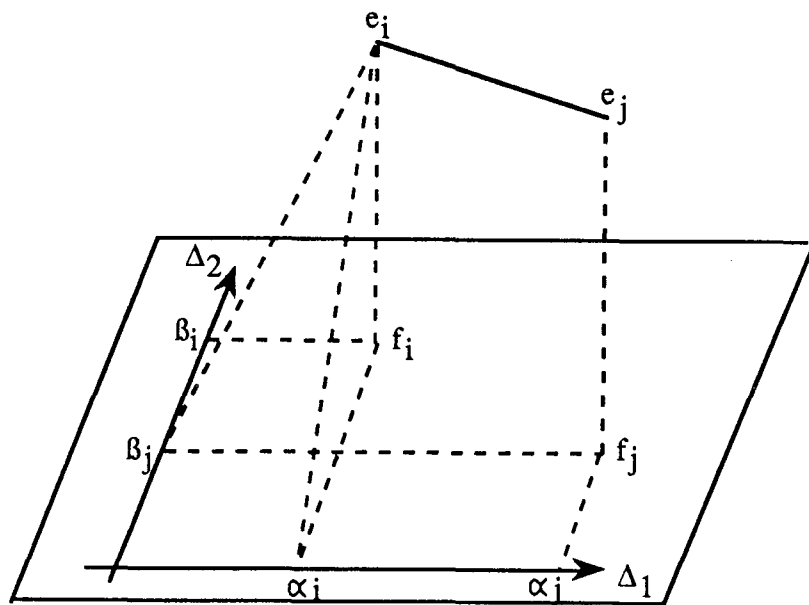


figure (3,2)  
géométrie de l'ACP

Il faut choisir le plan de projection sur lequel les distances seront en moyenne le mieux conservées, comme l'opération de projection raccourcit les distances on a :

$$d(f_i, f_j) \leq d(e_i, e_j)$$

On se fixera pour critère de rendre maximale la moyenne des carrés des distances entre les projections  $f_1, f_2, \dots, f_n$ .

c) *Plan principal* :

Pour déterminer ce plan il suffit de trouver deux droites  $\Delta_1$  et  $\Delta_2$ , si  $\Delta_1$  et  $\Delta_2$  sont perpendiculaires on a :

$$d^2(f_i ; f_j) = d^2(\alpha_i ; \alpha_j) + d^2(\beta_i ; \beta_j)$$

où les  $\alpha_j$  et les  $\beta_j$  sont les projections des  $e_j$  et  $f_j$  sur  $\Delta_1$  et  $\Delta_2$  respectivement. La méthode consiste alors à chercher  $\Delta_1$ , rendant maximale la moyenne des  $d^2(\alpha_i ; \alpha_j)$  puis  $\Delta_2$  perpendiculaire à  $\Delta_1$ , rendant maximale la moyenne des  $d^2(\beta_i ; \beta_j)$ . En d'autres termes le but de l'ACP est d'obtenir une représentation du nuage de point  $N(I)$  de  $R^p$  dans un espace de dimension réduite de telle manière que l'inertie portée par cet espace soit la plus grande possible.

d) *Axes principaux* :

On peut continuer en dehors du plan et on trouvera alors,  $\Delta_3, \Delta_4, \dots, \Delta_p$  perpendiculaires entre elles : les  $\Delta_i$  sont les axes principaux du nuage.

e) *Composantes principales* :

En projetant  $e_i$  qui avait pour coordonnées initiales  $(x_{i1}, x_{i2}, \dots, x_{ip})$  sur les axes principaux on obtient de nouvelles coordonnées  $(c_{i1}, c_{i2}, \dots, c_{ip})$ . On construit ainsi de nouveaux caractères  $(c^1, c^2, \dots, c^p)$  que l'on appelle les composantes principales. Chaque composante  $c^k$ , qui n'est autre que la liste des coordonnées des  $n$  individus sur l'axe  $\Delta_k$ , est une combinaison linéaire des caractères initiaux :

$$c^k = u_{1k} x^1 + u_{2k} x^2 + \dots + u_{pk} x^p$$

Les coefficients  $(u_{1k}, u_{2k}, \dots, u_{pk})$  forment le  $k$ -ième facteur principal  $u^k$ .

Si on prend seulement les  $q$  premières composantes principales, on obtiendra une meilleure représentation des données au moyen de  $q$  caractères ( $q < p$ ).

L'ACP est donc une méthode de réduction du nombre de caractères permettant des représentations géométriques des individus et des caractères, à condition que les  $p$  caractères initiaux ne soient pas indépendants donc qu'ils n'aient pas des coefficients de corrélation nuls.

L'ACP est une méthode linéaire, l'utilisation des notions de combinaisons linéaires, de distances, de projection conduit à raisonner selon l'hypothèse suivante : les individus et les variables (caractères) sont des éléments de deux espaces vectoriels euclidiens à  $p$  et  $n$  dimensions respectivement. Les outils mathématiques utilisés seront ceux de l'algèbre linéaire et du calcul matriciel.

### 3 - 3 - 6 - 2 *Les nuées dynamiques*

Cette méthode, due à E. DIDAY [41], fait partie de ces méthodes dites de partitionnement, qui ont l'avantage d'être rapide et de permettre la classification de population très nombreuses.

L'application de cette technique nécessite tout d'abord la définition d'un mode de représentation de tout groupe d'individus.

#### a) *Choix d'un mode de représentation :*

Cette représentation dite " noyau ", pour un groupe d'individus donné, peut être, par exemple, figure (3,3) :

- une droite,
- un groupe de point de la population,
- un centre de gravité.



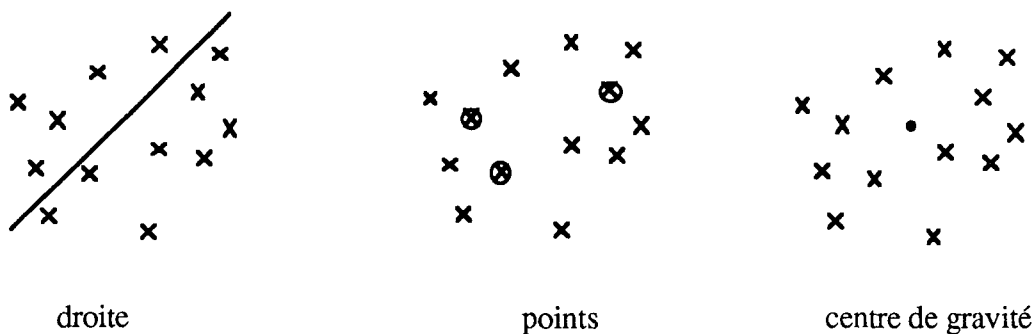


figure (3.3)

Les différents modes de représentation de groupe d'individus

b) *Description de l'algorithme des nuées dynamiques :*

Supposons donné un ensemble  $E$  sur lequel sont définies une distance  $d(x,y)$  entre éléments et une distance  $D(X,Y)$  entre sous-ensemble.

On se fixe à priori le nombre de classe  $k$  que l'on désire obtenir. Si on prend au hasard  $k$  paquets de  $p$  points chacun dans  $E$ , on appellera chacun de ces paquets un "noyau". Ces noyaux permettent de définir une partition de  $E$  en  $k$  classes, chaque classe comprenant les éléments qui sont plus proches d'un des noyaux que de tout les autres noyaux, ce qui permet de calculer une nouvelle famille de noyaux en associant à chaque classe de la partition l'ensemble de  $p$  points qui en est le plus proche. Le processus est réitéré jusqu'à ce que le classement soit stable c'est à dire : quand les noyaux ne se modifient que très peu. Mais cette construction contient une part d'arbitraire : la partition obtenue dépend du choix initial des noyaux. pour essayer de compenser cet arbitraire on applique la méthode plusieurs fois de suite, en partant à chaque fois d'une famille différente de noyaux tirés au hasard. Ceci permet de faire apparaître les nuées les plus significatives (formes fortes) qui sont stables d'un tirage à l'autre. Les formes faibles changent, par contre, aisément de classe et ne doivent pas être considérées comme des partitions fiables. Les formes fortes qui ne comprennent qu'un élément ne présentent pas d'intérêt, car elles concernent des éléments qui semblent inclassables. Par contre, les formes fortes qui comprennent beaucoup d'éléments sont intéressantes.

On schématise, dans ce qui suit, l'algorithme des nuées dynamiques dans le cas simple où le noyau est un point, figure (3,4) :

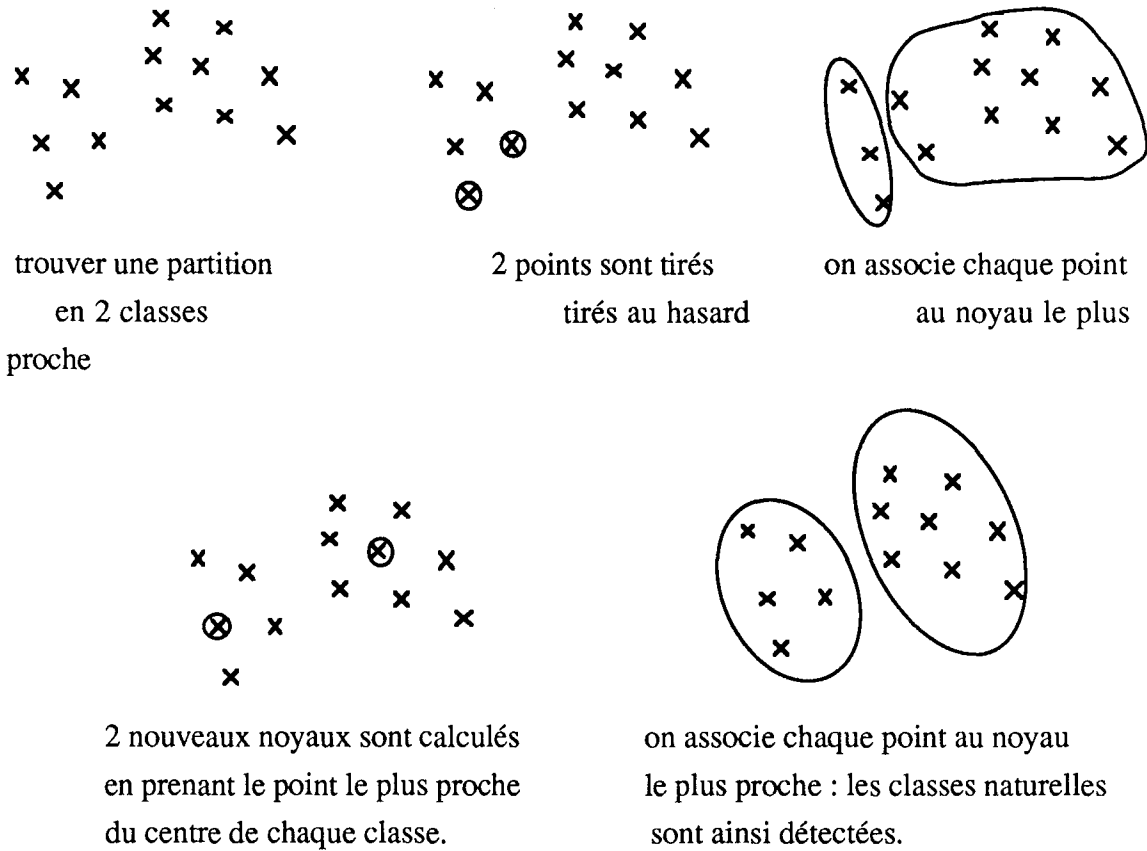


figure (3.4)

Algorithme des nuées dynamiques

3 - 3 - 7 QUELQUES DONNEES NUMERIQUES MESUREES SUR UN ECHANTILLON DE SIGNAUX EEG CHEZ LE RAT

Chaque individu (signal) est un vecteur  $V$  de 14 variables :

- 13 quantitatives où on retrouve "l'histogramme" HEEG,
- 1 qualitative où on retrouve le code affecté à l'individu par l'expert (1, 2, 3, 4 ou bien 5).

Le nombre énorme d'échantillons traités (plusieurs milliers), ne permet pas de reprendre la totalité des résultats obtenus durant cette étude, et ceci pour les raisons suivantes :

- Certains résultats ont été jugé non satisfaisants, de par la mauvaise qualité des enregistrements EEG due, en général, à une mauvaise implantation du micro-connecteur sur le rat (voir chapitre 2).

- Certains filtres ne permettaient pas, après analyse, une bonne séparation des différentes classes de signaux. En effet, on avait utilisé plusieurs filtres numériques avec des sélectivités différentes dont certains n'ont rien donné.

- Sur le même ensemble d'échantillons, il est arrivé de faire plusieurs traitements qui finalement ont abouti aux mêmes conclusions, en particulier sur le choix de filtre ou des intervalles.

- Certains choix des intervalles pour l'élaboration des différents histogrammes caractérisant le signal EEG, n'ont pas abouti à des résultats exploitables.

- Les acquisitions des signaux ont été réalisées avec différentes fréquences d'échantillonnage, pour arriver à la fin au choix d'une fréquence de 80 Hz.

Pour ces raisons et pour d'autres considérations pratiques on s'est retrouvé, à la fin de l'étude, avec un ensemble d'échantillons les plus représentatifs de 968 individus de durée 5 secondes chacun ce qui représente donc, 400 points pour chaque signal si on considère une fréquence d'échantillonnage égale à 80 Hz.

Cet ensemble a servi comme ensemble d'apprentissage au système mis en œuvre.

Les résultats obtenus, après filtrage de ces signaux et après analyse, sont illustrés à titre d'exemple dans tout ce qui suit.

### 3 - 3 - 7 - 1 *Description des variables quantitatives*

On retrouve dans les tableaux de la figure (3,5) quelques descriptions élémentaires des variables quantitatives : minimum, maximum, moyenne et écart-type.

nombre total de variables : 14  
-----

variables quantitatives : 13  
variables qualitatives : 1  
nombre total de modalités : 5

```
*****
* nu * id * type * mod * libelles des variables *
*****
* 1 * zerc 1 1 *nb zero crossing *
* 2 * ext1 1 1 *extremums 0-50 *
* 3 * ext2 1 1 *extremums 50-100 *
* 4 * ext3 1 1 *extremums 100-150 *
* 5 * ext4 1 1 *extremums 150-200 *
* 6 * ext5 1 1 *extremums 200-250 *
* 7 * ext6 1 1 *extremums 250-300 *
* 8 * ext7 1 1 *extremums > 300 *
* 9 * dzc1 1 1 *dist zero crossing 0-4 *
* 10 * dzc2 1 1 *dist zero crossing 4-8 *
* 11 * dzc3 1 1 *dist zero crossing 8-12 *
* 12 * dzc4 1 1 *dist zero crossing 12-16 *
* 13 * dzc5 1 1 *dist zero crossing > 16 *
* 14 * code 2 5 *codage des formes reconnues *
*****
```

```
*****
* variables * minimum * maximum * moyenne * ecart-type *
*****
* zerc * 29.000 * 95.000 * 68.229 * 7.9823 *
* ext1 * 1.0000 * 39.000 * 12.028 * 5.4816 *
* ext2 * 1.0000 * 34.000 * 13.385 * 6.2823 *
* ext3 * 1.0000 * 32.000 * 13.463 * 5.1028 *
* ext4 * 0.00000 * 24.000 * 10.331 * 3.5771 *
* ext5 * 0.00000 * 18.000 * 6.8060 * 3.6128 *
* ext6 * 0.00000 * 18.000 * 4.3158 * 3.2992 *
* ext7 * 0.00000 * 79.000 * 6.8999 * 6.6548 *
* dzc1 * 0.00000 * 4.0000 * 0.75232 * 0.88244 *
* dzc2 * 0.00000 * 27.000 * 9.5015 * 4.1554 *
* dzc3 * 1.0000 * 26.000 * 10.450 * 3.5873 *
* dzc4 * 0.00000 * 20.000 * 7.4953 * 3.0679 *
* dzc5 * 0.00000 * 13.000 * 5.6687 * 2.3512 *
*****
```

figure(3.5)

Description des variables quantitatives

3 - 3 - 7 - 2 *Description des variables qualitatives*

Le tableau, figure (3,6), reprend les différents codes affectés aux différentes formes reconnues sur le signal EEG. Ces codes sont appelés ici des modalités et le tableau suivant donne la part de chaque modalité par rapport à l'ensemble des échantillons.

```
1. variable :code codage des formes reconnues
-----

mod1 : fuseaux
mod2 : sp
mod3 : eveil
mod4 : non reconnu
mod5 : sol

*****
* mod1 * nbre * 4 *
*****
* mod1 * 380 * 39 * *****
* mod2 * 81 * 8 * *****
* mod3 * 57 * 6 * *****
* mod4 * 270 * 28 * *****
* mod5 * 181 * 19 * *****
*****
```

figure(3.6)  
Description des variables qualitatives

### 3 - 3 - 7 - 3 Matrice de corrélation

L'étude de la matrice de corrélation permet d'avoir un aperçu sur la dépendance deux à deux, dans la variation, entre les variables considérées. Cette étude permet de réduire le nombre de ces variables, dans le cas où certaines parmi celles-ci présentent une "forte corrélation". La notion de "forte corrélation" est liée à un seuil au delà duquel on considère que deux variables x et y ont un coefficient de corrélation  $r(x,y)$  tel que l'étude de la variable x uniquement nous informerait sur le comportement de la variable y.

On remarquera sur le tableau de la figure (3,7) qui donne la matrice de corrélation des variables : ZERC, EXT1, . . . , DZC5, que la plus forte corrélation entre deux variables différentes est :

$$r(\text{ZERC}, \text{DZC2}) = 0,7522.$$

Ce coefficient est d'ailleurs le seul à être aussi élevé. Les corrélations entre les variables quantitatives n'étant pas très fortes on préfère garder toutes les variables.

MATRICE DE CORRELATION

	zerc	ext1	ext2	ext3	ext4	ext5	ext6	ext7	dzc1	dzc2	dzc3	dzc4	dzc5
zerc	1.0000												
ext1	0.4848	1.0000											
ext2	0.4337	0.6069	1.0000										
ext3	0.4316	0.4609	0.5037	1.0000									
ext4	0.2728	-0.1178	-0.1300	0.0535	1.0000								
ext5	0.0385	-0.4826	-0.5709	-0.4239	0.1044	1.0000							
ext6	-0.0349	-0.5408	-0.6018	-0.5329	0.0241	0.5414	1.0000						
ext7	-0.0857	-0.5828	-0.6441	-0.6404	-0.0894	0.4483	0.5877	1.0000					
dzc1	0.2896	0.3591	0.1092	0.1240	0.0428	-0.0541	-0.0728	-0.1037	1.0000				
dzc2	0.7522	0.2767	0.1533	0.1297	0.1590	0.1616	0.1449	0.1911	0.0567	1.0000			
dzc3	0.5692	0.1900	0.2853	0.3129	0.1888	0.0268	-0.0146	-0.0891	0.0878	0.0960	1.0000		
dzc4	0.0808	0.2604	0.3919	0.3680	0.0499	-0.2957	-0.3685	-0.4585	-0.0082	-0.2104	-0.1485	1.0000	
dzc5	-0.6840	-0.3996	-0.4838	-0.4638	-0.1779	0.1205	0.1906	0.2561	-0.1077	-0.3784	-0.5671	-0.4853	1.0000

figure(3,7)

Matrice de corrélation des variables quantitatives

3 - 3 - 7 - 4 *Classification automatique*

La technique des nuées dynamiques a permis d'obtenir une partition en 10 classes, de l'ensemble des individus. Le principe consistait à faire plusieurs essais et choisir la meilleure partitions au sens des critères énoncés dans le paragraphe (3 - 3 - 6).

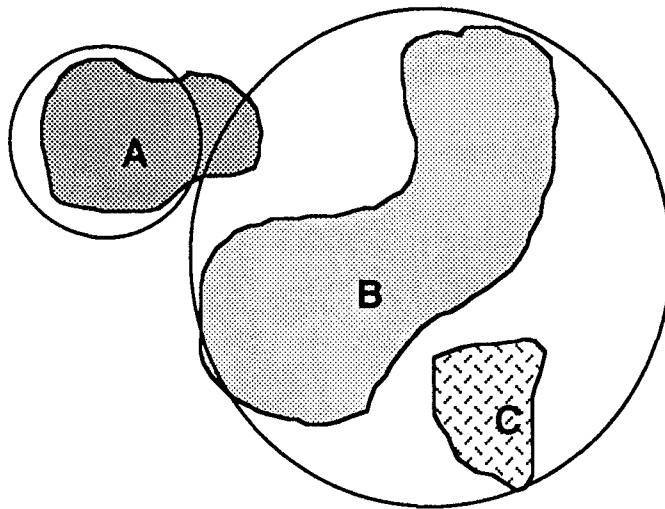
Le choix de 10 classes de partition au lieu de 3, car il y a trois formes à reconnaître (Sol, Sp, Eveil), a semblé plus judicieux, en effet :

- Avec un nombre de classes moins élevé, on risquerait de se retrouver avec des classes "géantes" regroupant des individus qui en réalité ne représenteraient pas la même forme.

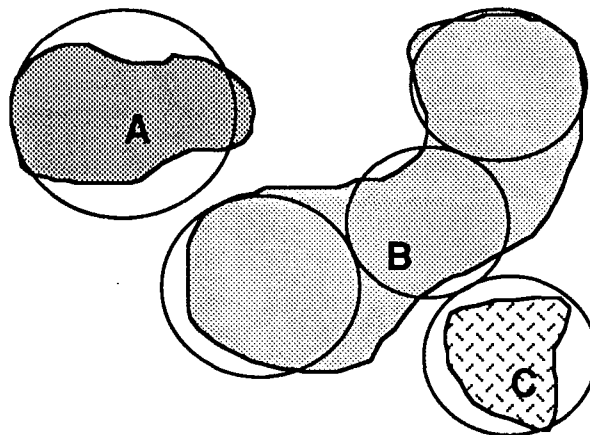
- Pour des raisons de "finesse", il est préférable de prendre un grand nombre de classe, même si à la fin on se retrouve avec des classes séparées mais qui représentent en réalité la même forme. Ceci reviendrait à découper une grande classe en sous-classes, ce raisonnement peut se poursuivre jusqu'à dire que la meilleure partition est celle où on considère chaque individu comme une classe à lui tout seul.

- Le système qui a été utilisé a l'inconvénient de se représenter chaque classe comme une sphère de centre O et de rayon R, une raison technique de plus pour prendre un maximum de partitions possibles. La figure (3,8) résume cela.

A, B, C : classes réelles



Partitions en deux classes sphériques



Partitions en 5 classes sphériques

Figure (3.8)

Partition d'un ensemble d'individus en classes

Après partition en 10 classes et plusieurs essais, on obtient la partition suivante, figure (3,9)

Remarque

Les individus, concernés par ces classes, représentent, en fait, le Sp et le Sol. Le Sol regroupant les signaux de très basse fréquence (4 c/s, Sol pur) ainsi que des signaux de plus grande fréquence (13 c/s, fuseaux), qui naturellement seront classés dans des classe différentes. L'éveil n'est pas présent sur la partition de la figure (3,9) pour des raisons qu'on verra plus en avant.

Dans le tableau de la figure (3,10) on peut lire les centres de gravité de chaque classe. Ces centres de gravité sont les signaux les mieux représentatifs de chaque classe.

Les tableaux des figure (3,11) et (3,12) détaillent un peu plus la description des classes :

- écart-type des variables pour chaque classe de la partition,
- distance entre les centres de gravités des classes,
- distance moyenne de la classe à son centre de gravité.

### 3 - 3 - 8 APPRENTISSAGE

Le but de cette analyse, dès le départ, était de disposer d'un ensemble de modèles représentant au mieux les différentes formes du signal EEG. La phase apprentissage consiste donc, en partie, à "mettre un nom" sur chaque classe de partition. Ce qui veut dire attribuer, à chaque forme du signal, un ou plusieurs modèles.

#### 3 - 3 - 8 - 1 Détermination des modèles

Dans le paragraphe (2 - 6 - 5) du chapitre II, on a abordé le codage des signaux EEG avec l'aide d'un expert. Le codage est le suivant :

- 1 - fuseaux,    2 - Sp,    3 - Eveil,    5 - Sol,    4 - non reconnu.



variables selectionnees :

zerc ext1 ext2 ext3 ext4 ext5 ext6 ext7 dscl drc2  
drc3 drc4 drc5

parametres de la methode :

nombre d'essais : 10  
nombre de classes : 10  
nombre iterations : 10  
valeur du test d'arret : .100E-01  
Choix d'initialisation : 3 tirage au hasard d'une partition  
Choix de l'edidion : 4 resultats succincts  
partition  
coordonnees des centres de gravite et ecart-types des variables dans les classes  
distances entre les centres des classes

inertie des donnees : .2616E+06  
|\*\*\*\*\*|\*\*\*\*\*|

essai numero : 7

la convergence a ete atteinte en 5 iterations

les pourcentages d'inertie expliquee en 10 classe a chaque iteration :

1	2	3	4	5
56.04	60.78	62.15	62.83	63.17

partition obtenue :

elements de chaque classe par ordre d'inertie residuelle en 1/10\*\*4 :

(inertie residuelle d'un element = rapport entre la distance au centre de gravite et l'inertie intraclassées de la population)

classe numero 1 (effectif= 67)

837 ( 3.)	515 ( 4.)	784 ( 5.)	550 ( 5.)	905 ( 5.)
223 ( 5.)	782 ( 6.)	717 ( 6.)	200 ( 6.)	85 ( 6.)
241 ( 6.)	221 ( 7.)	546 ( 7.)	682 ( 7.)	906 ( 7.)
752 ( 7.)	634 ( 7.)	869 ( 7.)	565 ( 8.)	659 ( 8.)
286 ( 8.)	268 ( 8.)	222 ( 8.)	209 ( 8.)	511 ( 8.)
551 ( 8.)	779 ( 8.)	805 ( 9.)	530 ( 10.)	891 ( 10.)
528 ( 10.)	516 ( 10.)	523 ( 10.)	867 ( 10.)	560 ( 10.)
829 ( 10.)	186 ( 11.)	244 ( 11.)	199 ( 11.)	647 ( 11.)
531 ( 11.)	395 ( 12.)	87 ( 12.)	862 ( 13.)	280 ( 13.)
203 ( 13.)	661 ( 14.)	202 ( 15.)	774 ( 15.)	117 ( 15.)
196 ( 15.)	205 ( 15.)	783 ( 16.)	522 ( 16.)	553 ( 16.)
874 ( 17.)	563 ( 17.)	231 ( 17.)	58 ( 20.)	552 ( 21.)
649 ( 22.)	813 ( 23.)	633 ( 24.)	574 ( 26.)	148 ( 39.)
101 ( 44.)	547 ( 46.)			

classe numero 2 (effectif= 59)

47 ( 3.)	49 ( 4.)	171 ( 5.)	307 ( 5.)	584 ( 5.)
71 ( 6.)	300 ( 6.)	804 ( 7.)	882 ( 7.)	64 ( 7.)
773 ( 7.)	732 ( 7.)	74 ( 8.)	8 ( 8.)	55 ( 8.)
79 ( 8.)	99 ( 8.)	728 ( 8.)	330 ( 9.)	579 ( 9.)
65 ( 9.)	46 ( 10.)	801 ( 10.)	750 ( 10.)	308 ( 10.)
638 ( 11.)	762 ( 11.)	710 ( 11.)	824 ( 11.)	146 ( 11.)
859 ( 11.)	107 ( 12.)	434 ( 12.)	790 ( 12.)	235 ( 12.)
846 ( 13.)	315 ( 13.)	827 ( 13.)	729 ( 14.)	533 ( 14.)
457 ( 14.)	95 ( 14.)	80 ( 16.)	194 ( 16.)	627 ( 17.)
66 ( 17.)	145 ( 17.)	170 ( 18.)	137 ( 19.)	812 ( 19.)
597 ( 19.)	643 ( 21.)	538 ( 21.)	529 ( 22.)	616 ( 23.)
90 ( 23.)	645 ( 26.)	84 ( 27.)	803 ( 35.)	

classe numero 3 (effectif= 155)

125 ( 1.)	689 ( 2.)	701 ( 2.)	56 ( 2.)	54 ( 2.)
391 ( 3.)	428 ( 4.)	394 ( 4.)	332 ( 4.)	70 ( 4.)
857 ( 4.)	589 ( 4.)	381 ( 4.)	401 ( 4.)	626 ( 4.)
6 ( 5.)	740 ( 5.)	900 ( 5.)	321 ( 5.)	622 ( 5.)
707 ( 5.)	583 ( 5.)	4 ( 5.)	578 ( 5.)	114 ( 6.)
324 ( 6.)	727 ( 6.)	111 ( 6.)	686 ( 6.)	596 ( 6.)
42 ( 6.)	593 ( 6.)	698 ( 6.)	126 ( 6.)	896 ( 6.)
350 ( 6.)	97 ( 6.)	820 ( 6.)	191 ( 6.)	603 ( 6.)
745 ( 7.)	2 ( 7.)	825 ( 7.)	497 ( 7.)	413 ( 7.)
468 ( 7.)	601 ( 7.)	858 ( 7.)	702 ( 7.)	17 ( 7.)
899 ( 7.)	7 ( 7.)	417 ( 8.)	444 ( 8.)	743 ( 8.)
901 ( 8.)	168 ( 8.)	749 ( 8.)	369 ( 8.)	429 ( 8.)
746 ( 8.)	491 ( 8.)	387 ( 8.)	364 ( 8.)	691 ( 8.)
83 ( 8.)	712 ( 8.)	798 ( 8.)	725 ( 8.)	492 ( 9.)
644 ( 9.)	128 ( 9.)	592 ( 9.)	155 ( 9.)	881 ( 9.)
484 ( 9.)	123 ( 9.)	367 ( 9.)	894 ( 9.)	16 ( 9.)
346 ( 9.)	751 ( 9.)	495 ( 9.)	430 ( 9.)	336 ( 9.)
613 ( 10.)	21 ( 10.)	709 ( 10.)	108 ( 10.)	724 ( 10.)
678 ( 10.)	62 ( 10.)	166 ( 10.)	680 ( 10.)	414 ( 10.)
771 ( 10.)	18 ( 10.)	761 ( 10.)	893 ( 10.)	809 ( 11.)
78 ( 11.)	420 ( 11.)	617 ( 11.)	127 ( 11.)	848 ( 11.)
895 ( 11.)	713 ( 11.)	485 ( 11.)	325 ( 11.)	819 ( 12.)
331 ( 12.)	730 ( 12.)	692 ( 12.)	744 ( 12.)	789 ( 12.)
599 ( 12.)	342 ( 12.)	625 ( 13.)	136 ( 13.)	822 ( 13.)
67 ( 13.)	736 ( 13.)	496 ( 13.)	467 ( 14.)	769 ( 14.)
454 ( 14.)	465 ( 14.)	544 ( 14.)	641 ( 14.)	614 ( 14.)
382 ( 14.)	606 ( 15.)	537 ( 15.)	134 ( 15.)	48 ( 15.)
542 ( 17.)	459 ( 16.)	494 ( 16.)	20 ( 16.)	110 ( 16.)
193 ( 18.)	368 ( 17.)	9 ( 17.)	427 ( 17.)	41 ( 17.)
399 ( 19.)	794 ( 20.)	800 ( 18.)	509 ( 19.)	447 ( 19.)
		714 ( 20.)	737 ( 21.)	823 ( 27.)

classe numero 4 (effectif= 109)

98 ( 2.)	59 ( 3.)	116 ( 3.)	313 ( 4.)	536 ( 4.)
897 ( 4.)	345 ( 4.)	505 ( 4.)	510 ( 4.)	541 ( 4.)
480 ( 5.)	526 ( 5.)	656 ( 5.)	190 ( 5.)	863 ( 5.)
767 ( 5.)	844 ( 5.)	36 ( 6.)	877 ( 6.)	781 ( 6.)
239 ( 6.)	455 ( 6.)	799 ( 6.)	362 ( 6.)	396 ( 6.)
735 ( 6.)	385 ( 6.)	153 ( 6.)	318 ( 6.)	714 ( 7.)
679 ( 7.)	69 ( 7.)	799 ( 7.)	810 ( 7.)	23 ( 7.)
808 ( 7.)	143 ( 7.)	739 ( 7.)	320 ( 7.)	372 ( 7.)
317 ( 8.)	639 ( 8.)	878 ( 8.)	734 ( 8.)	720 ( 8.)
5 ( 8.)	40 ( 9.)	636 ( 9.)	722 ( 9.)	631 ( 9.)
89 ( 9.)	392 ( 9.)	240 ( 9.)	96 ( 9.)	674 ( 9.)
876 ( 9.)	757 ( 9.)	854 ( 10.)	63 ( 10.)	520 ( 10.)
39 ( 10.)	581 ( 10.)	708 ( 10.)	723 ( 10.)	879 ( 10.)
788 ( 10.)	685 ( 10.)	655 ( 10.)	353 ( 10.)	871 ( 11.)
568 ( 11.)	165 ( 11.)	303 ( 11.)	53 ( 11.)	105 ( 11.)
228 ( 11.)	106 ( 11.)	807 ( 11.)	100 ( 11.)	576 ( 11.)
585 ( 11.)	786 ( 11.)	670 ( 11.)	787 ( 12.)	889 ( 12.)
344 ( 12.)	232 ( 12.)	57 ( 12.)	842 ( 12.)	532 ( 12.)
637 ( 12.)	755 ( 12.)	860 ( 13.)	885 ( 13.)	792 ( 13.)
872 ( 13.)	543 ( 13.)	347 ( 14.)	474 ( 14.)	120 ( 15.)
630 ( 15.)	845 ( 15.)	518 ( 15.)	162 ( 15.)	580 ( 17.)
764 ( 18.)	629 ( 21.)	312 ( 21.)	521 ( 22.)	



coordonnees des centres de gravite des variables pour chaque classe de la partition :

variable (effectif)	population ( 908)	classe 1 ( 67)	classe 2 ( 59)	classe 3 ( 155)	classe 4 ( 109)	classe 5 ( 35)	classe 6 ( 114)	classe 7 ( 120)
zerc	68.4	78.6	79.3	68.0	71.9	76.8	66.5	55.3
ext1	12.0	23.6	12.3	8.95	11.9	13.5	14.4	6.83
ext2	13.3	23.6	11.8	9.05	12.7	16.8	16.1	6.87
ext3	13.2	18.4	13.4	9.18	13.7	20.1	18.7	7.68
ext4	10.2	8.28	11.9	10.7	13.2	13.3	10.2	8.23
ext5	6.86	2.51	10.1	9.27	8.50	6.86	3.99	7.23
ext6	4.50	.851	7.54	6.77	5.22	2.89	1.32	5.71
ext7	7.33	.478	11.2	13.2	5.64	2.40	.825	11.8
dzc1	.754	1.24	1.15	.677	.881	.914	.842	.367
dzc2	9.75	14.0	16.1	10.4	11.0	10.6	7.30	5.67
dzc3	10.5	12.2	12.2	10.7	11.4	17.1	9.65	6.83
dzc4	7.22	8.12	5.85	5.85	7.44	6.43	10.0	5.70
dzc5	5.72	3.48	4.15	6.19	4.95	3.14	5.19	8.84

variable (effectif)	population ( 908)	classe 8 ( 122)	classe 9 ( 36)	classe 10 ( 91)
zerc	68.4	62.7	76.8	70.8
ext1	12.0	9.70	7.75	16.7
ext2	13.3	10.9	8.33	23.5
ext3	13.2	12.3	9.58	16.0
ext4	10.2	9.48	8.92	8.71
ext5	6.86	7.27	10.3	3.14
ext6	4.50	5.16	8.33	1.13
ext7	7.33	6.87	22.6	.648
dzc1	.754	.574	.667	.736
dzc2	9.75	7.45	15.7	8.93
dzc3	10.5	8.97	12.0	11.5
dzc4	7.22	6.93	5.28	9.52
dzc5	5.72	7.22	4.58	4.47

figure(3.10)  
Centres de gravité des différentes classes

carte-types des variables pour chaque classe de la partition :

variable (effectif)	( 908)	( 67)	( 59)	( 155)	( 109)	( 35)	( 114)	( 120)	( 120)
zerc	8.08	4.29	3.38	3.26	2.55	4.60	3.46	6.01	
ent1	5.53	4.34	3.92	2.96	3.12	3.96	3.50	2.33	
ent2	6.41	4.70	3.43	2.85	3.14	4.27	2.96	2.61	
ent3	5.09	3.50	3.73	2.67	3.02	3.67	3.94	2.33	
ent4	3.55	2.85	3.43	3.34	3.47	3.25	3.22	2.81	
ent5	3.65	2.13	2.86	3.20	2.36	3.23	2.21	2.61	
ent6	3.30	1.30	3.69	2.53	2.40	1.82	1.33	2.01	
ent7	6.66	.920	3.03	3.39	2.37	2.55	1.29	3.43	
dec1	.887	1.11	.988	.786	.966	.867	.951	.618	
dec2	4.13	3.53	3.39	2.42	2.87	4.61	2.32	2.51	
dec3	3.57	3.39	2.80	2.61	2.80	3.79	2.71	2.33	
dec4	2.85	2.34	2.15	2.17	2.55	2.14	2.78	2.12	
dec5	2.31	1.72	1.46	1.34	1.56	1.61	1.73	1.77	

variable (effectif)	( 908)	( 122)	( 36)	( 91)	( 36)	( 36)	( 91)	( 91)
zerc	8.08	3.04	5.12	2.96				
ent1	5.53	3.31	3.02	3.37				
ent2	6.41	3.24	3.06	3.87				
ent3	5.09	3.14	3.11	3.21				
ent4	3.55	3.16	2.67	2.62				
ent5	3.65	2.87	3.26	1.88				
ent6	3.30	2.42	3.21	1.31				
ent7	6.66	3.03	10.3	1.29				
dec1	.887	.735	.745	.875				
dec2	4.13	2.34	3.82	2.41				
dec3	3.57	2.72	3.30	3.24				
dec4	2.85	2.29	2.42	2.50				
dec5	2.31	1.52	1.55	1.61				

figure(3.11)  
Description des classes de partition

distances entre les centres de gravite des classes :

	classe 1	classe 2	classe 3	classe 4	classe 5	classe 6	classe 7
classe 1	0.00	23.03	29.95	20.97	15.64	18.70	39.03
classe 2	23.03	0.00	14.48	11.22	15.65	22.39	29.10
classe 3	29.95	14.48	0.00	11.36	21.73	20.28	15.18
classe 4	20.97	11.22	11.36	0.00	11.84	12.76	22.44
classe 5	15.64	15.65	21.73	11.84	0.00	14.68	32.31
classe 6	18.70	22.39	20.28	12.76	14.68	0.00	24.21
classe 7	39.03	29.10	15.18	22.44	32.31	24.21	0.00

	classe 8	classe 9	classe 10
classe 8	28.18	34.38	11.95
classe 9	20.42	13.94	22.58
classe 10	10.17	14.38	23.93
classe 4	11.63	20.58	15.68
classe 5	21.00	27.21	13.91
classe 6	13.45	30.91	9.78
classe 7	11.79	27.31	30.21
classe 8	0.00	23.85	19.55
classe 9	23.85	0.00	32.17
classe 10	19.55	32.17	0.00

distance moyenne de la classe a son centre de gravite :

numero de classe	distance
1	10.6
2	10.8
3	9.5
4	9.5
5	11.7
6	9.2
7	9.5
8	9.5
9	12.0
10	9.0

figure (3.12)  
Description des classes de partition

Pour déterminer à quelle forme correspond chaque classe parmi les 10, il suffit de voir qu'elle est la forme prédominante dans chaque classe. La prédominance d'une forme dans une classe n'est pas toujours vérifiée. En effet, il faut distinguer deux cas :

- 1er cas :

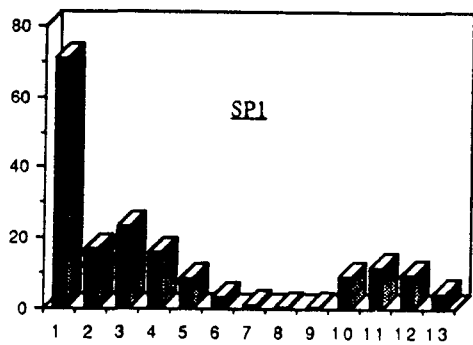
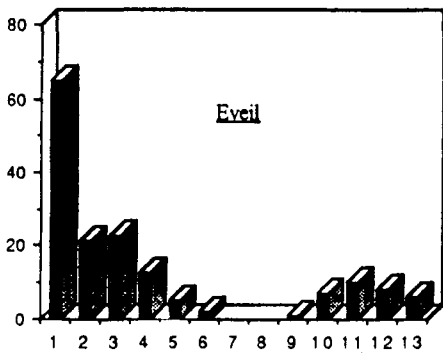
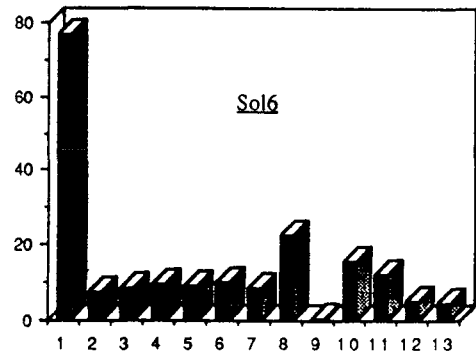
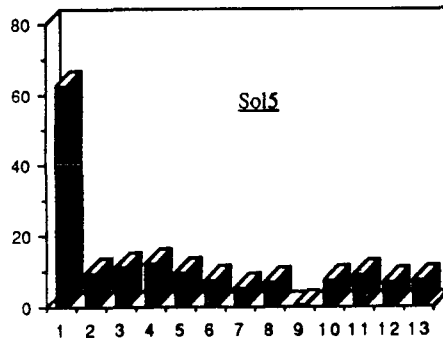
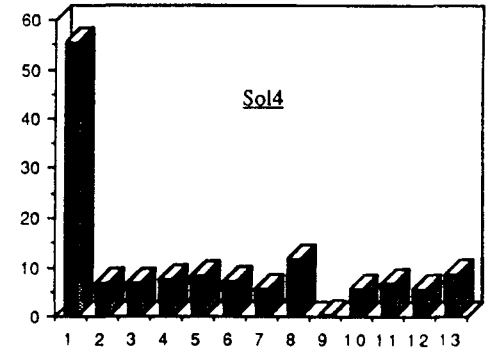
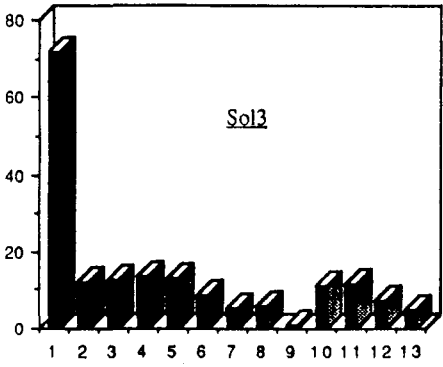
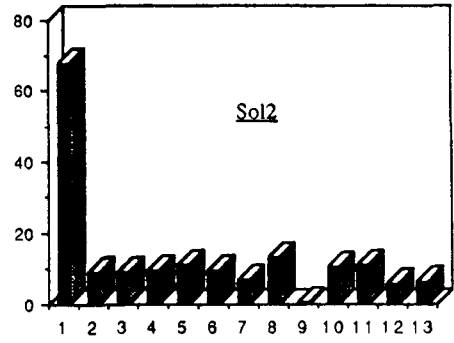
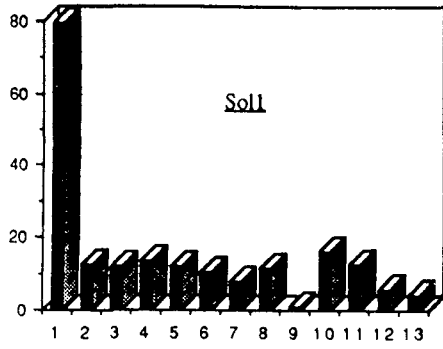
La classe obtenue est un mélange pratiquement égal de plusieurs formes, on ne peut considérer cette classe comme représentative d'une forme. Ce sont, souvent, des classes qui se trouvent aux frontières de classes de formes fortes et qui regroupent, suivant les essais de partitions, des individus de classes avoisinantes. Ces classes sont à rejeter, elle ne peuvent, en aucun cas, servir de modèle.

- 2ieme cas

La classe présente une prédominance d'une forme particulière (Sp par exemple) et son centre de gravité sera considéré comme modèle représentant cette forme. La définition de la prédominance d'une forme est liée à la présence de cette forme dans une classe par rapport aux autres formes. Par exemple, si dans une classe, on est en présence de plus de 75 % d'une forme, cette forme sera prédominante dans cette classe.

Un utilitaire, fort simple, permet de vérifier pour chaque classe, la présence en pourcentage des différentes formes. Suivant les critères énoncés dans les cas 1 et 2, les classes données par le tableau de la figure (3,10) sont affectées aux formes suivantes:

classe1 : rejetée  
classe2 : Sol1  
classe3 : Sol2  
classe4 : Sol3  
classe5 : rejetée  
classe6 : rejetée  
classe7 : Sol4  
classe8 : Sol5  
classe9 : Sol6  
classe10 : Sp1



figure(3.13)  
Représentation graphique des différents modèles

Le grand nombre de classes affectées au Sol est normal étant donné que cette forme inclut des signaux différents et dont le mélange est aléatoire, en fonction de la séquence d'acquisition du signal.

Les histogrammes (au sens graphiques uniquement) de la figure (3,13) reprennent la représentation des centres de gravité des classes retenues ainsi que de l'éveil obtenu par une analyse à part .

### 3 - 3 - 8 - 2 *Correction de l'apprentissage*

L'opérateur peut reprendre l'apprentissage pour corriger, éventuellement, la reconnaissance. Il peut supprimer les modèles qui ne donnent pas satisfaction, ou rajouter d'autres modèles.

En effet, les modèles obtenus par classification automatique ont été utilisés par le système qui sera décrit dans le chapitre suivant. Après évaluation, et sur les commentaires d'un expert, le modèle représentant la forme Sp n'était pas "l'idéal" car il ne permettait pas de trancher catégoriquement entre les deux formes : Eveil et Sp. Ce modèle fut remplacé par celui de la figure (3,14), beaucoup plus distinct de celui de l'Eveil. On a obtenu ce modèle synthétique de la manière suivante :

Dans l'ensemble des signaux codés 2 ( reconnus par l'expert comme Sp), on a sélectionné les signaux qui ont semblé être les plus réguliers et les plus représentatifs de la forme Sp. A partir de ces signaux on a élaboré le modèle de la figure (3,14), ce qui revient à dire que dans l'ensembles de ces signaux on a choisi uniquement les "formes fortes" et rejeté les "formes faibles" qui étaient prises en considération pour le calcul du modèle initial.

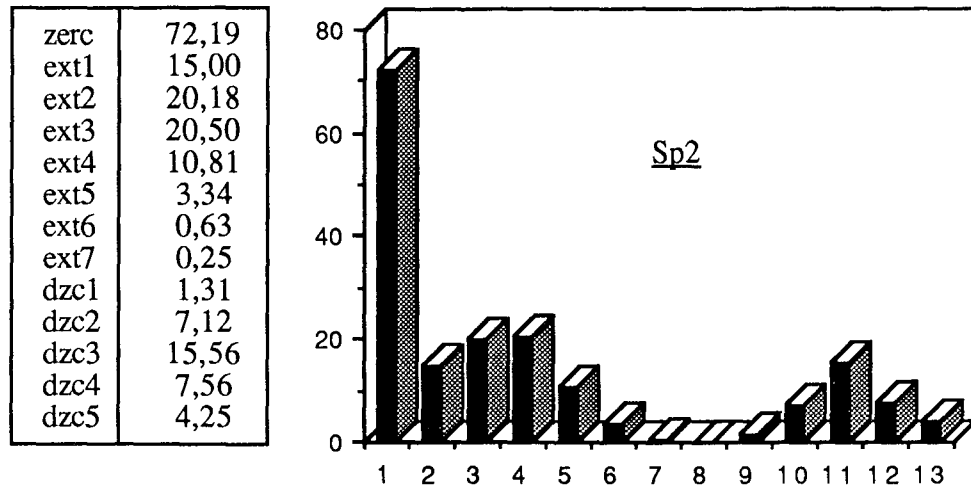


Figure (3.14)  
Représentation graphique du modèle Sp

### 3 - 3 - 8 - 3 *Cas particulier de l'éveil*

L'absence de l'éveil dans le tableau de la figure (3,10) est expliqué par le fait que dans le système initial tel qu'il a été conçu, l'éveil est déduit par élimination : si le signal n'est ni du Sol ni du Sp il ne peut être que de l'éveil. C'est de cette manière que les différents stades de sommeil et de veille étaient souvent mesurés par l'opérateur humain.

Cependant, si on applique cette règle on retrouvera classé comme Eveil aussi bien les vrais signaux Eveil que les signaux non reconnus.

Dans les versions ultérieures, on a préféré, pour plus de rigueur, établir un modèle de l'éveil au même titre que les autres formes. Ce modèle apparaît dans la figure (3,13).

### 3 - 3 - 8 - 4 *Discussion sur la technique d'apprentissage utilisée*

Cette forme d'apprentissage qu'on peut appeler apprentissage parallèle ou double, est un apprentissage qui se fait de deux manières différentes qui se rejoignent à la fin. En effet, le codage des signaux EEG, par un expert, est une première opération d'apprentissage (apprentissage supervisé, ou avec professeur), complétée par une seconde : la classification automatique (apprentissage non supervisé, ou sans



professeur), qui apporte la rigueur du calcul numérique dans la conception des modèles, mais qui reste sans intérêt si elle n'est pas recoupée avec la première.

### 3 - 3 - 8 - 5 Arrêt sur l'intérêt du filtrage des signaux

L'analyse en composantes principales appliquée à l'ensemble des échantillons, ne permet pas de percevoir des classes bien séparées et évidentes (c'est souvent le cas en pratique). Cependant on peut percevoir une certaine séparation entre les classes et l'évolution de celle-ci en fonction de l'efficacité du filtre appliqué aux données.

Les figures (3,15 - a,b,c), sont assez représentatives de l'intérêt de la mise en œuvre d'un filtre passe-bas qui améliorerait les signaux à traiter en éliminant un maximum de bruit possible.

La figure (3,15-a), présente une projection sur les axes principaux des signaux "bruts" (sans aucun filtrage). On remarque une grande masse d'individus regroupés en un énorme nuage. Cette représentation est sans aucun intérêt.

Sur le graphique de la figure (3,15-b), on retrouve les mêmes individus traités par un filtre passe-bas moyen de quatrième ordre. On distingue, déjà, deux nuages (l'un à droite en bas et l'autre à gauche en haut), et peut-être une pseudo-frontière entre ces deux nuages.

La figure (3,15-c), quand à elle, toujours avec les mêmes individus, sépare encore mieux les deux nuages qui semblent plus compacts et une frontière peut paraître plus évidente. Ces signaux sont filtrés par un filtre de huitième ordre, plus sélectif que le précédent.

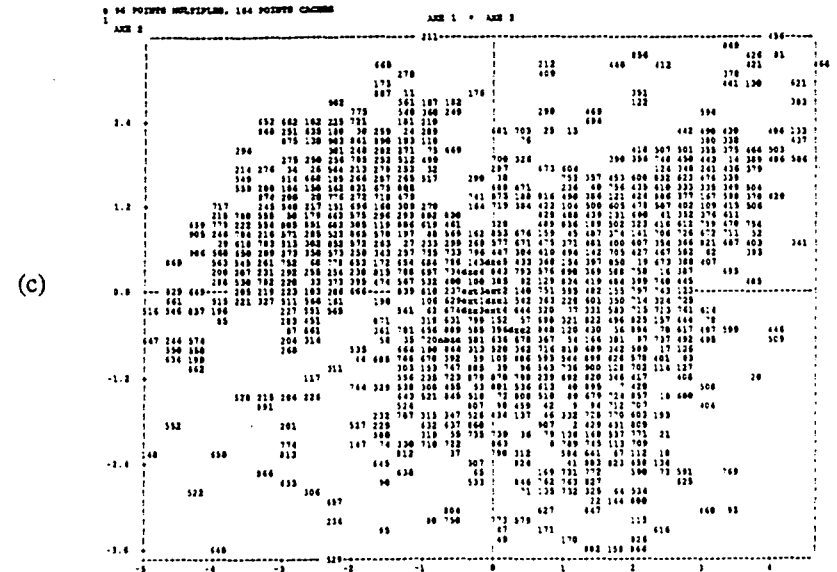
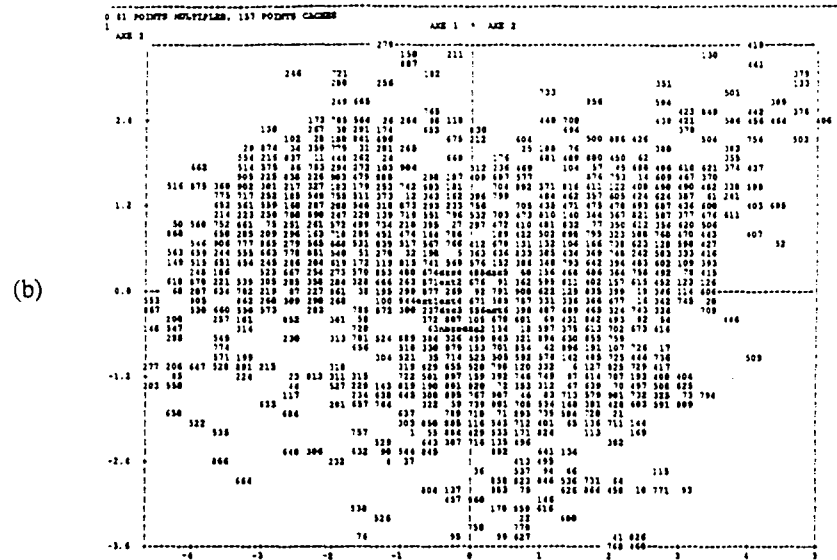
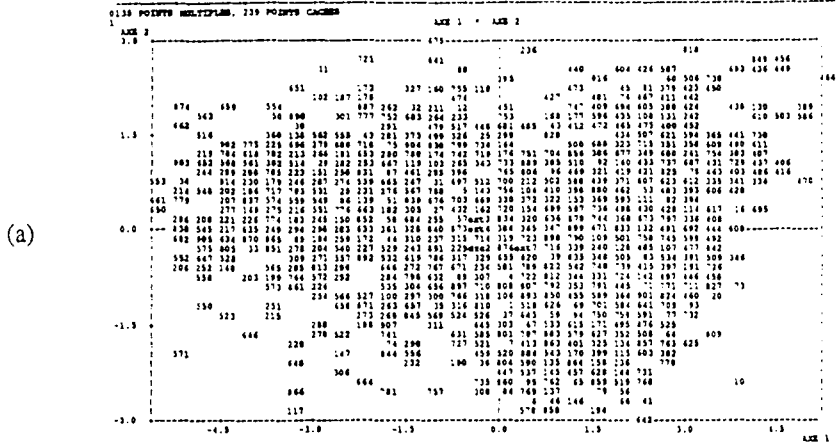
On remarquera, dans tous les cas, des individus éloignés et non classables qui, évidemment, ne seront pas représentatifs.

Après ces constatations on serait tenté de dire : si on augmente l'ordre du filtre et sa sélectivité, on arriverait peut être à obtenir des classes distinctes et bien séparées. Ceci n'est pas aussi évident en effet, si les signaux proviennent d'une instrumentation qui induit un bruit et une déformation du signal, un filtre quel qu'il soit ne peut que limiter l'influence du bruit et non pas le supprimer totalement.

### 3 - 4 CONCLUSION CHAPITRE III

On résume le besoin de ce chapitre par la question suivante : Si on dispose d'un lot de signaux EEG représentés par un lot de vecteurs ("histogrammes hybrides"), comment obtenir les modèles les mieux représentatifs de chaque forme ?. On a répondu à ce besoin par l'étulisation des techniques de l'analyse de données, le choix de la technique des nuées dynamique nous a apparu judicieux étant donnée qu'au départ on suppose que ces signaux sont inconnus, et ce n'est qu'à la fin de la classification que l'on recoupe les résultats avec ceux de l'expert.

Les modèles obtenus sont indispensables à la phase d'apprentissage lors de la conception du système de reconnaissance décrit dans le chapitre suivant.



figure(3.15)  
ACP appliquée à un même ensemble d'individus

## **CHAPITRE IV**

### **DESCRIPTION DU SYSTEME DE RECONNAISSANCE AUTOMATIQUE DES STADES DE SOMMEIL ET DE VEILLE CHEZ LE RAT**

## 4 - 1 INTRODUCTION

Ce chapitre est la suite logique des deux précédents. Le gros du travail consistait en l'élaboration de modèles représentant au mieux les différentes formes du signal EEG.

On abordera, ici, la phase reconnaissance de forme dont l'efficacité dépend en grande partie de la qualité de l'apprentissage. Avant cela on introduit une description de l'équipement informatique, matériel et logiciel, mis en œuvre pour cette réalisation.

## 4 - 2 EQUIPEMENT INFORMATIQUE

### 4 - 2 - 1 MATERIEL

Le système réalisé est piloté par un micro-ordinateur de type PC/AT avec une horloge de 10 MHz. Un modèle qui compte déjà parmi les plus anciens sur le marché actuellement et qui pourrait être amélioré, en vitesse notamment.

Le type de traitement choisi et le nombre limité de modèles font que la mémoire vive qui équipe ce type de machine en standard ( 640 Ko) est suffisante.

Le rôle du micro-ordinateur est de prendre en charge le traitement des données, le dialogue avec la carte S200 et de gérer les différentes composantes du système (acquisition, reconnaissance, visualisation, impression . . .).

### 4 - 2 - 2 LOGICIEL

Le fonctionnement du système est assuré grâce à son logiciel, c'est à dire grâce à l'ensemble des programmes de fonctionnement dont il dispose. Les performances d'un système dépendent étroitement de la qualité du logiciel qu'il exploite, aussi bien sur le plan de la vitesse d'exécution que de la souplesse d'emploi et de la fiabilité.

Toutes les procédures de gestion et de commande du système ont été écrites en langage turbo Pascal 4.0, mais elles pouvaient l'être en n'importe quel langage de programmation.

Quelques précautions ont été prises lors de la rédaction des programmes d'application :

*- Fonctionnement en temps réel*

Le système doit pouvoir gérer en temps réel la tâche qui lui incombe. Cette tâche si elle est réalisée ne doit pas prendre un temps excessif. On doit être en mesure de trouver un juste milieu entre l'application de techniques compliquées et "lourdes" et la perte d'un minimum d'informations sur le signal à reconnaître. En effet, le signal EEG étant un signal continu, le temps passé dans les traitements est égale à la portion du signal EEG qui n'est pas prise en compte, donc une perte d'information. Dans notre cas d'application on perd dans le pire des cas 1/6 du signal, ce temps correspond à celui des traitements et calculs. En effet, pour un signal de 5 secondes, les opérations de filtrage, extraction d'informations, reconnaissance prennent moins d'une seconde sur le matériel informatique décrit précédemment .

*- Environnement*

L'environnement dans lequel fonctionne le système est particulier. L'opérateur chargé de l'exploitation du système est en général peu au courant de l'informatique. Ceci conduit à adopter au niveau du matériel et au niveau du logiciel des solutions particulières susceptibles d' être mise en œuvre facilement par l'opérateur. Une interface utilisateur simple et conviviale est importante.

*- Fiabilité du logiciel [45]*

Les applications informatiques deviennent de plus en plus diverses et s'infiltrent dans presque tous les domaines. Il est donc évident que la caractéristique dynamique la plus importante de la plupart des logiciels est leur fiabilité.

La fiabilité d'un système quelconque (pas seulement d'un logiciel) dépend de la qualité de la conception, de la qualité de la réalisation et de la fiabilité des composants qui constituent le système.

Il est difficile de définir précisément la fiabilité d'un logiciel. On pourrait dire qu'un logiciel est fiable s'il est correct, c'est à dire s'il répond aux spécifications initiales. Cependant, cette définition ne prend pas en compte la possibilité que les spécifications du logiciel soient incomplètes et incorrectes, ce qui est très souvent le cas.

En fait, un logiciel vérifié mathématiquement est parfois moins fiable qu'un logiciel testé mais non vérifié. Les vérifications de programmes ne prennent pas suffisamment en compte l'environnement dans lequel ces programmes doivent s'exécuter. EN particulier, les vérifications font trop souvent l'hypothèse que cet environnement est correct. En revanche, les effets des fautes ou des incohérences dues à l'environnement sont mises en évidence beaucoup plus facilement par le tests.

Une définition plus réaliste de la fiabilité d'un programme est que ce programme doit répondre aux spécifications, ne jamais produire de résultats erronés, quelle que soit l'entrée, ne jamais entrer dans un état incohérent, réagir de façon sensée et utile en présence d'une situation inattendue et n'être mis en défaut que lorsqu'il n y a plus rien à faire.

En résumé, la fiabilité d'un système logiciel est une mesure de sa capacité à fournir les services attendus par les utilisateurs. Pour obtenir un haut degré de fiabilité, il est nécessaire de rajouter au système une bonne quantité de code, souvent redondant, afin d'effectuer les vérifications nécessaires. Cela réduit la vitesse d'exécution et augmente l'espace mémoire nécessaire à l'exécution du programme. Cependant, au fur et à mesure que les utilisateurs deviennent plus expérimentées, leur principal critère de qualité est la fiabilité plutôt que l'efficacité. Il existe un certain nombre de raisons à cela :

- Les équipements deviennent de plus en plus rapides et de bon marché. Il y a donc moins de raisons d'optimiser l'utilisation des équipements aux dépens du confort des utilisateurs.

- un logiciel peu fiable sera rejeté par les utilisateurs, quelles que soient ses performances. Il deviendra très vite inutile.

- L'efficacité d'un système peut être améliorée parce que la plupart du temps d'exécution se passe dans des régions relativement petites du programme. Il est en revanche difficile d'améliorer la fiabilité d'un système peu fiable, car l'absence de fiabilité tend à être répartie sur l'ensemble du programme.

- L'inefficacité d'un programme peut être prédite. Ce n'est pas le cas du manque de fiabilité.

Pour obtenir un bon degré de fiabilité, on devra comprendre l'environnement dans lequel le logiciel opère et préparer des spécifications qui définissent, dans la mesure du possible, le rôle du logiciel dans cet environnement.

Ces quelques points étant énoncés, on aborde la phase reconnaissance de forme.

#### 4 - 3 PHASE RECONNAISSANCE DE FORME

Dans toute reconnaissance de forme le modèle de référence privilégié est naturellement le comportement humain. La reconnaissance de forme peut être définie par l'ensemble des méthodes et moyens permettant au moins de reproduire, et si possible de dépasser, les moyens naturels de perceptions et de compréhension des êtres vivants [43].

Dans la notion de reconnaissance de forme intervient les notions de recueil des données, de traitement et analyse de données et de classification. Une excellente méthode de reconnaissance ne permettra pas de traiter efficacement des données initiales erronées ou non pertinentes.

On remarquera que les connaissances initiales (apprentissage) jouent, dans ce cadre, un très grand rôle. Cette notion de conditions initiales est aussi importante que dans le domaine de l'automatique. On sait que l'état final d'un système, l'efficacité d'un correcteur . . . dépendent des conditions initiales. L'absence de connaissances préalables, servant de base au système, complique énormément la phase de reconnaissance.

##### 4 - 3 - 1 ALGORITHMES DE RECONNAISSANCE

Un signal S est représenté par un vecteur V,

$$V = \{ x_1, x_2, \dots, x_{13} \}$$

l'ensemble des modèles est noté M

$$M = \{ M_1, M_2, \dots, M_9 \}$$

M<sub>9</sub> est un modèle supplémentaire caractérisant le SP. On dispose donc, en réalité, de deux modèles pour la forme SP.



chaque modèle a une représentation graphique sous forme d'histogramme

$$\text{HEEG} = \{ \text{HEEG}_{M1}, \text{HEEG}_{M2}, \dots, \text{HEEG}_{M9} \}$$

La reconnaissance se traduit par une comparaison du signal S avec les différents modèles  $M_1, M_2 \dots$ . Le signal S est reconnu comme appartenant à la même classe de forme du modèle auquel il se rapproche le plus.

#### 4 - 3 - 2 NOTION DE DISTANCE ENTRE REPRESENTATION

On admet, généralement, que les points figuratifs d'une forme donnée sont plus proches les uns des autres que des autres points figuratifs de toutes les autres formes différentes. C'est l'hypothèse de *compacité*, qui n'est malheureusement pas toujours vérifiée, car dépendant de la représentation.

On admet aussi, souvent, que les classes sont bornées, et qu'il n'y a pas de recouvrement. C'est l'hypothèse de *séparabilité*. En fait, cette hypothèse est bien souvent infirmée en pratique.

La notion de distance permet d'évaluer la similitude entre les représentations. Il existe plusieurs approches de cette notion.

Dans le cas d'une distance métrique, on considère un espace métrique  $\{E, d\}$ . On appelle distance d'un point  $P \in E$ , à une classe  $C_0 \subset E$ , la quantité  $d_1$  telle que :

$$d_1(P, C_0) = \inf\{d(P, M) / M \in C_0\}$$

La distance entre deux classes est définie par  $d_2$  telle que :

$$d_2(C_1, C_2) = \inf\{d(P, M) / P \in C_1, M \in C_2\}$$

Plus la distance considérée est petite, plus on admet que la ressemblance, au point de vue de la reconnaissance des formes, est grande.

### 4 - 3 - 3 NOTION DE RESSEMBLANCE ET DE DISSEMBLANCE

La reconnaissance peut être faite en comparant l'histogramme HEEG, construit à partir d'un signal EEG, aux histogrammes modèles.

Une mesure simple et qui donne de bons résultats consiste à utiliser l'aire commune à deux histogrammes. Cette mesure est une mesure de "ressemblance" car elle est maximale lorsque les histogrammes sont identiques et nulle lorsqu'ils sont complètement différents.

On peut aussi définir la mesure de "dissemblance" associée, qui est une distance. Elle utilise l'aire non commune au deux histogrammes à comparer, on l'appellera D, figure (4,1).

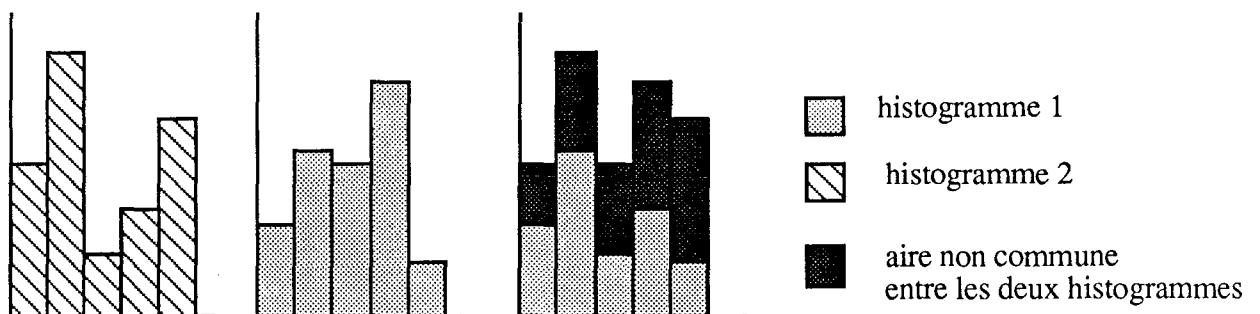


figure (4,1)  
Notion de dissemblance

Soient:

HEEG<sub>M</sub>: histogramme modèle,

HEEG<sub>S</sub>: représentation du signal S à reconnaître.

La dissemblance D est donnée par :

$$D_j = \sum_{i=1}^n | \text{HEEG}_S(i) - \text{HEEG}_{M_j}(i) | \quad \begin{array}{l} n = 1 \dots 13 \\ j = 1 \dots \text{nombre de modèles} \end{array}$$

On aura donc :

D = 0 lorsque les histogrammes sont identiques

D ≠ 0 lorsque les histogrammes sont différents

L'opération de reconnaissance consiste donc, à trouver une dissemblance minimale entre un modèle et le signal. On définit un seuil tel que :

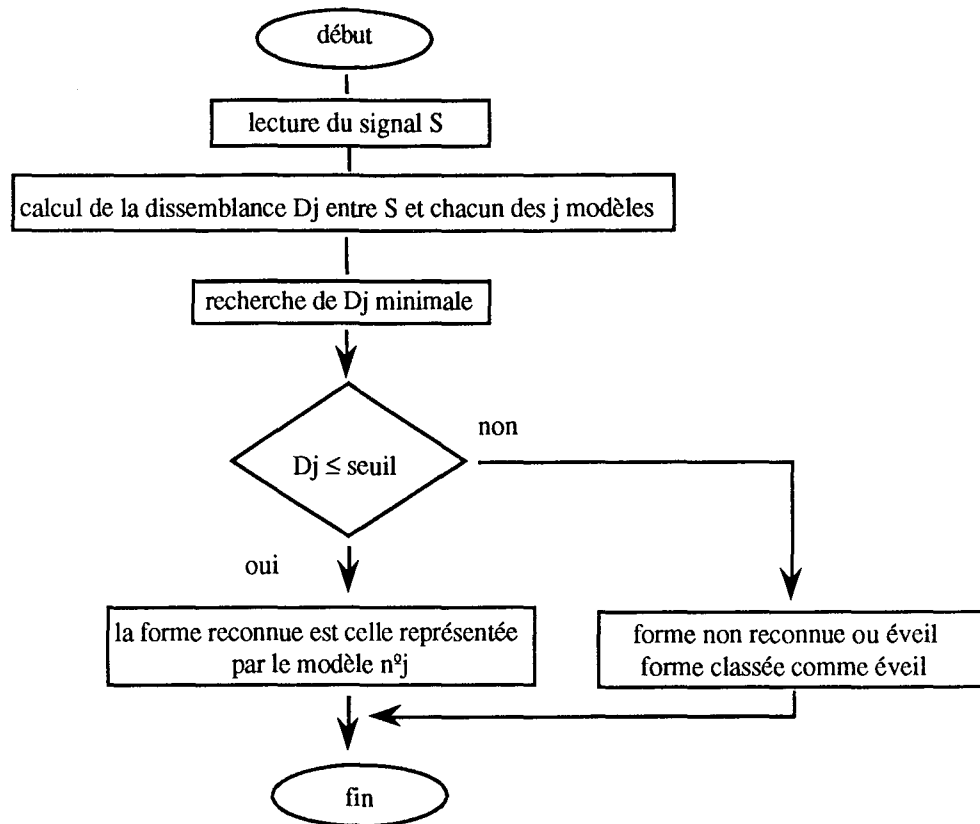
- si D<sub>j</sub> est la dissemblance minimale entre le signal S et les j modèles
- et
- D<sub>j</sub> ≤ seuil

alors :

On considère que la forme reconnue est celle figurée par le modèle n° j.

Si D<sub>j</sub> ≥ seuil, la forme est reconnue et classée comme éveil. On retrouvera donc, aussi bien des formes non reconnues que de l'éveil classés comme éveil.

L'organigramme (orga.4) résume cette opération :



(orga.4)

Organigramme de l'algorithme de reconnaissance

#### 4 - 4 DESCRIPTION DU SYSTEME D'ANALYSE DE SOMMEIL CHEZ LE RAT

On abordera la partie logiciel du système qui est composée de deux parties complémentaires : SLP et VISU.

##### 4 - 4 - 1 DESCRIPTION DU SLP

Les opérations effectuées par le SLP sont les suivantes :

- gestion de l'interface utilisateur
- acquisition du signal
- filtrage
- traitement et représentation du signal
- visualisation, temps réel, du résultat de reconnaissance
- enregistrement des résultats.

Le SLP est construit autour d'une interface utilisateur très simplifiée mais assez complète, on dispose d'un "tableau de bord" qui fournit les possibilités suivantes, figure (4,2) :

- programmation du nom de fichier où seront éventuellement stockés les résultats de reconnaissance,

- programmation de la durée d'enregistrement (de 1 à 24 heures ou plus),

- les commandes du système à partir du clavier du micro-ordinateur dont les touches (S,H,B,Q) sont affectées aux opérations suivantes :

- B (*begin*) : commencer le traitement,

- Q (*quit*) : arrêt du système,

- S (*store*) : sauvegarde des résultats,

- H (*help*) : une commande d'aide à l'utilisateur.

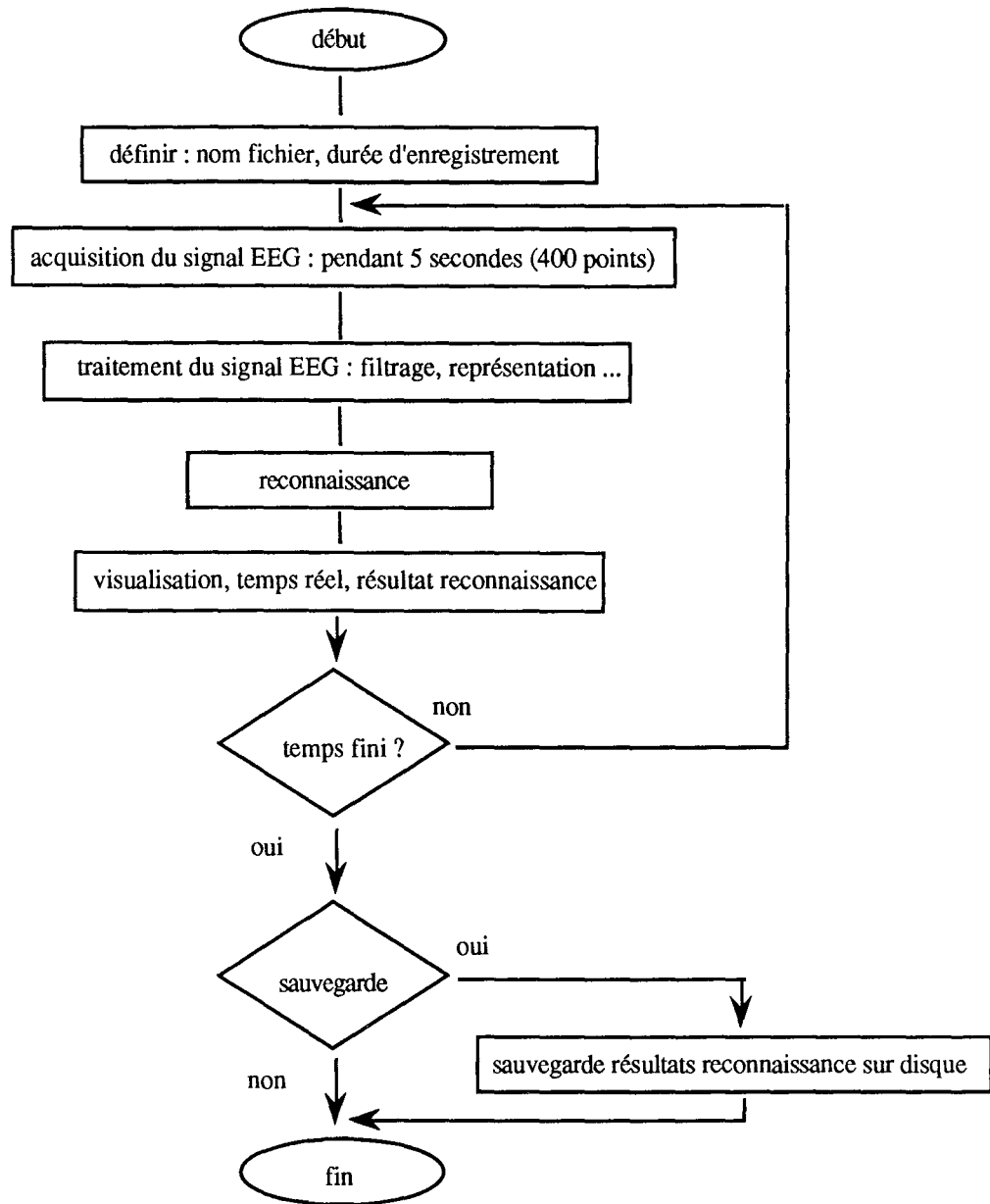
- un chronomètre visible sur le tableau de bord donnant le temps écoulé depuis le début du traitement.

- une alarme qui permet d'avoir en continu, sous les yeux, l'état du système :

- stop : arrêt,

- work : traitement en cours.

L'organigramme de fonctionnement du SLP est le suivant (orga.5) :



(orga.5)

Organigramme de fonctionnement du SLP

#### 4 - 4 - 2 DESCRIPTION DE VISU

La partie VISU permet d'aller chercher les résultats du traitement à n'importe quel moment, de les visualiser par page d'une heure sous forme d'hypnogramme, d'imprimer les hypnogrammes et de donner, par conséquent, la part de chaque forme dans l'ensemble des signaux ou sur une durée limitée d'enregistrement, figure (4,3).

*Remarques :*

*- les interfaces peuvent être personnalisées suivant l'utilisateur,*

*- il existe une version du système incluant des unités regroupant les différentes procédures et accessibles à partir du compilateur du turboPascal 4.0 . Celles-ci permettent à l'utilisateur de développer sa propre application.*

#### 4 - 4 - 3 REGLAGE DU NIVEAU DES SIGNAUX EEG

Il est indispensable, dans ce type de système, que toutes les données soient évaluées dans les mêmes conditions. On dispose d'un ensemble de modèles évalués dans des conditions définies, et si on veut optimiser l'opération de reconnaissance il est évident que ces conditions doivent être requises à chaque utilisation.

Les niveaux des signaux EEG ne sont pas fixes mais dépendent de certains paramètres difficiles à contrôler ( épaisseur de la boîte crânienne, disposition des électrodes . . . ); Cependant, on propose à l'utilisateur deux solutions :

a) - *réglage manuel du gain :*

Avant chaque utilisation et pour chaque rat on doit adapter le niveau du signal EEG en réglant le gain de l'amplificateur dont est doté l'enregistreur des signaux EEG. Pour cela il suffit d'ajuster le gain de manière à ce que l'hypnogramme du SLP indique la même forme que celle reconnue par l'utilisateur par lecture de l'enregistrement sur papier. Cette opération, à part le fait qu'elle est pénible, suppose que l'utilisateur a une bonne connaissance de ces signaux ce qui n'est pas souvent le cas.

Il existe une autre manière de procéder : visualiser sur moniteur les deux lignes des niveaux le plus bas et le plus haut correspondant à ceux mesurés sur le signal éveil qui a servi à l'apprentissage du système. Le réglage du gain consiste donc, à maintenir tout signal éveil dans cette plage en ajustant le gain manuellement, pour cela le rat doit être éveillé avant chaque utilisation.

On perd, dans ces cas, la qualité d'une instrumentation entièrement automatisée.

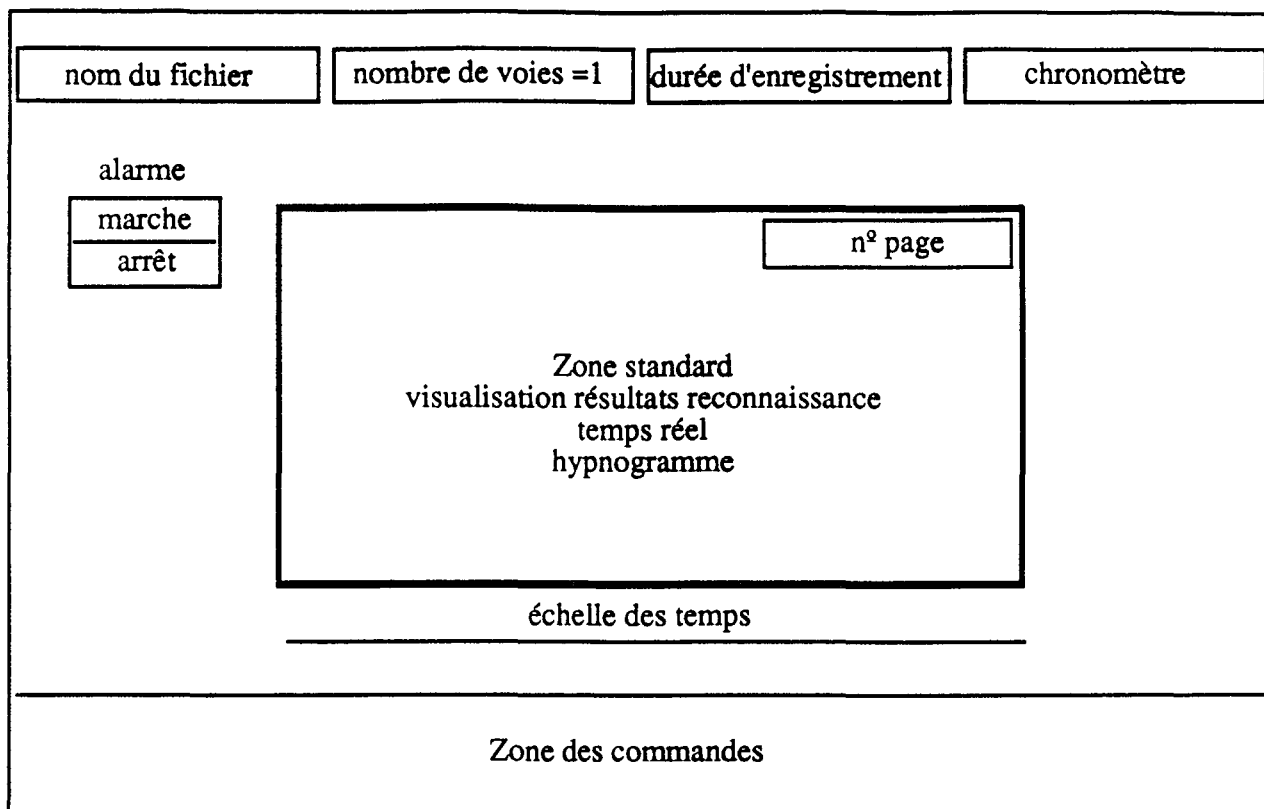


figure (4.2)  
Schéma de l'interface utilisateur pour SLP

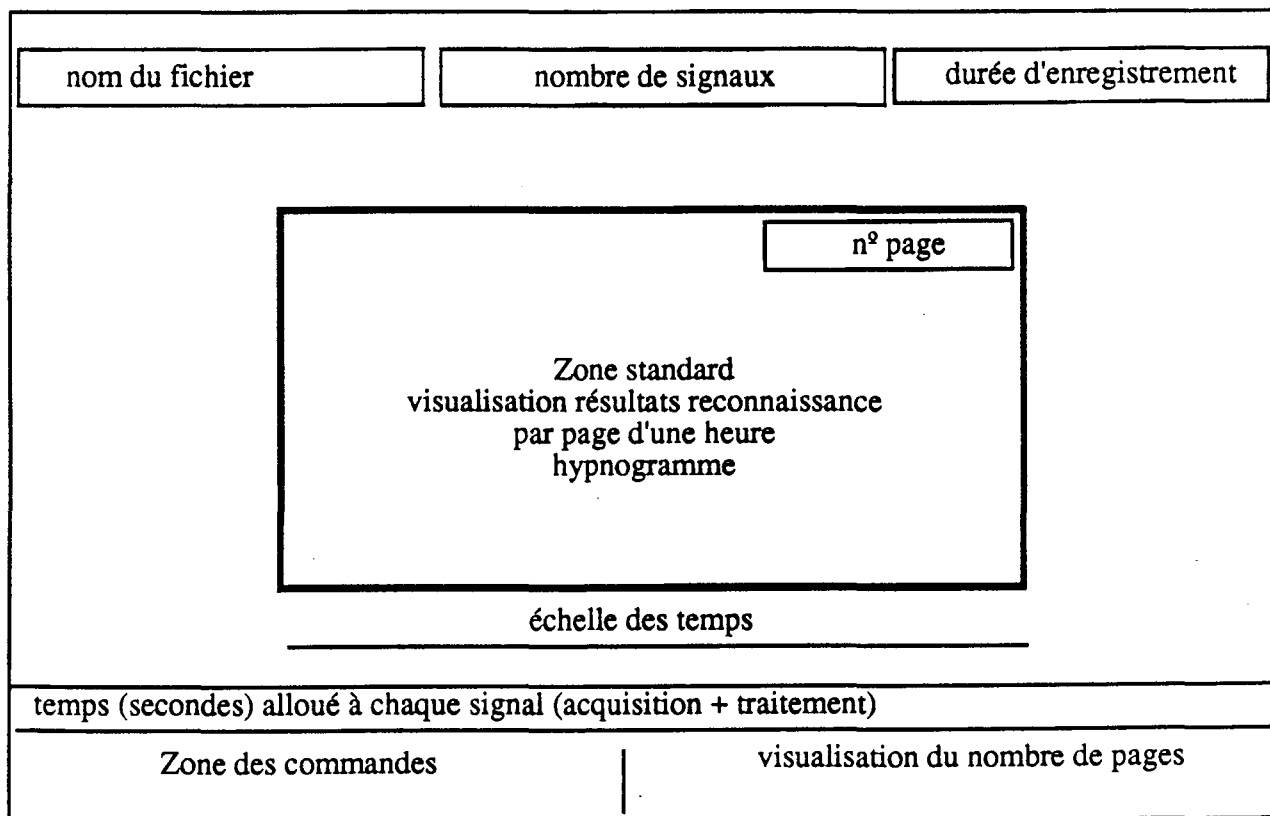


figure (4.3)  
Schéma de l'interface utilisateur pour VISU



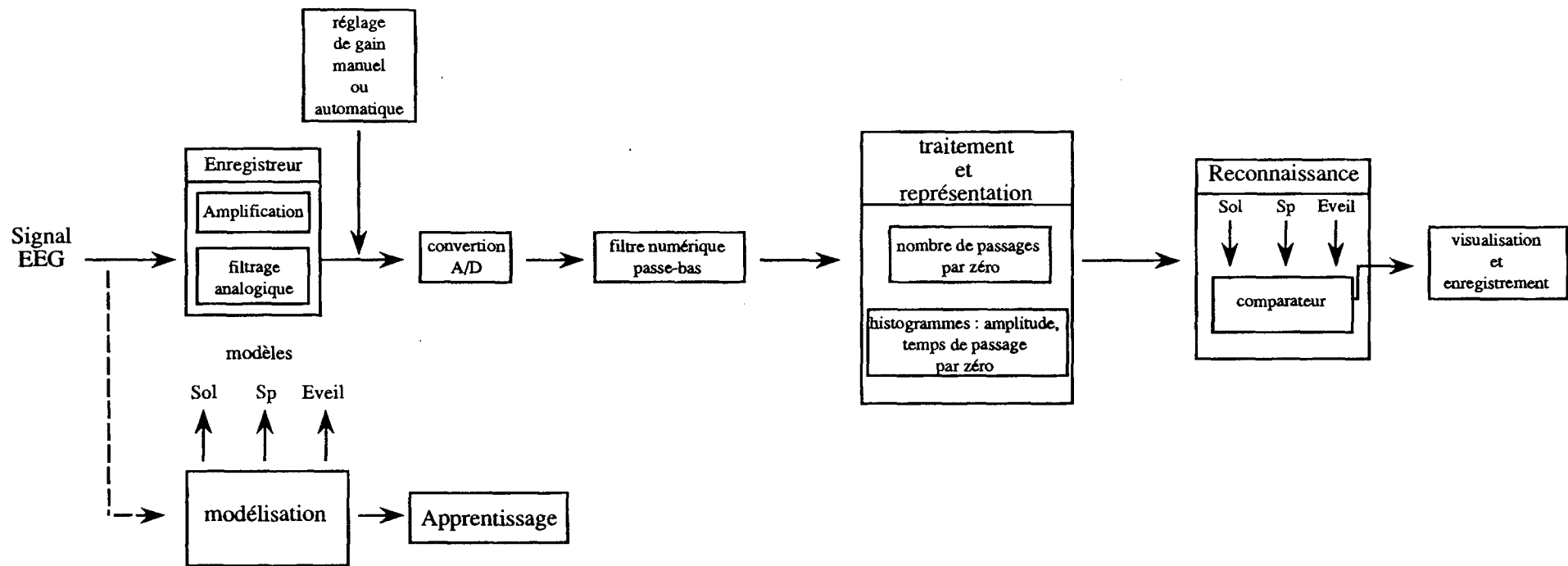


figure (4.4)  
 Synoptique du système de reconnaissance

b) - *réglage de gain automatique (normalisation)* :

L'adaptation des niveaux se fera par logiciel. On procède à une acquisition du signal éveil avant le début de traitement, on calcule l'amplitude moyenne des extremums du signal  $M_s$  puis on la compare à celle de l'éveil modèle  $M_m$ , on obtient un coefficient  $C$  tel que :

$$C = M_s / M_m$$

Il suffit donc d'ajuster tous les extremums détectés sur le signal à reconnaître par le coefficient  $C$  (extremum normalisé = extremum /  $C$ ). On ramène donc, tous les signaux, quelque soit le gain, au niveau des signaux modèles.

En conclusion, on dira que les bons réglages resteront toujours la preuve du bon utilisateur. La figure (4,4) reprend le schéma synoptique du système dans sa totalité.

#### 4 - 5 EVALUATION

L'évaluation a été entreprise suivant le concept énoncé dans le paragraphe (1-8), figure(1,4), c'est à dire qu'un ensemble test est créé à partir de l'ensemble des données reconnues par l'expert. L'évaluation porte sur le nombre de formes communes reconnues par l'expert et le système.

Le système de base étant le même, les versions testées diffèrent soit par le filtre passe-bas mis en œuvre, soit par les valeurs de seuils (voir paragraphe 4-3-3), ou bien par l'implantation de modèles supplémentaires (éveil, nouveau Sp...). Ces différentes versions sont notées : SLP3, SLP4, SLP5, SLP6, SLP7, SLP61, SLP62, SLP63, SLP64. Le tableau suivant résume les différences entre ces versions. On y trouve, le taux (%) de reconnaissance exacte pour les différentes versions:

nom version	seuil (d = )	filtre passe-bas	taux de recon. exactes (%)	remarques particulières
SLP3	30	ordre 4	63.71	sans modèle éveil
SLP4	40	ordre 4	88.47	..
SLP5	30	ordre 4	61.06	premier modèle SP
SLP6	40	ordre 8	89.72	modèle éveil + nouveau SP
SLP7	45	ordre 8	92.99	..
SLP61	30	ordre 8	67.08	..
SLP62	25	ordre 8	42.43	..
SLP63	20	ordre 8	63.55	sans les temps de passage par zéro
SLP64	30	ordre 8	67.29	1 modèle Sp supplémentaire

On remarquera qu'une simple valeur de seuil peut changer le comportement du système. Le choix de ces paramètres est des plus critiques dans ce type de système. L'évaluation par un expert est une étape décisive dans le choix d'une version ou d'une autre.

Si on veut expliquer l'importance du choix de la valeur du seuil, on peut avancer les deux cas suivants :

- on choisit un seuil assez important, on risque de prendre en compte des signaux qui se rapprocheraient plus ou moins d'un modèle sans vraiment appartenir à la même forme.

- on choisit un seuil très faible, on rejeterait des formes qui devraient être reconnues mais qui seraient classées comme éveil, (on rappelle que la classe éveil comprend aussi bien les "vrais" éveil que les signaux non reconnus).

On choisira alors, une valeur de seuil intermédiaire qui ne prendrait pas en compte n'importe quelle forme et qui ne rejeterait pas toutes les formes qui ne seraient pas une copie exacte du modèle.

Il est intéressant de regarder les deux hypnogrammes figure (4,5), résultant du traitement du même ensemble de signaux par SLP6 et SLP61, pour constater une différence dans la classification de certains signaux. En effet, on remarque que SLP6 (en haut) est plus "souple" dans son jugement car il considère certaines formes comme reconnues, alors que SLP61 (en bas) les rejette et les classe, par conséquent, comme éveil.

#### 4 - 6 CONCLUSION CHAPITRE IV

A la fin de ce chapitre et en conclusion on fera les remarques suivantes :

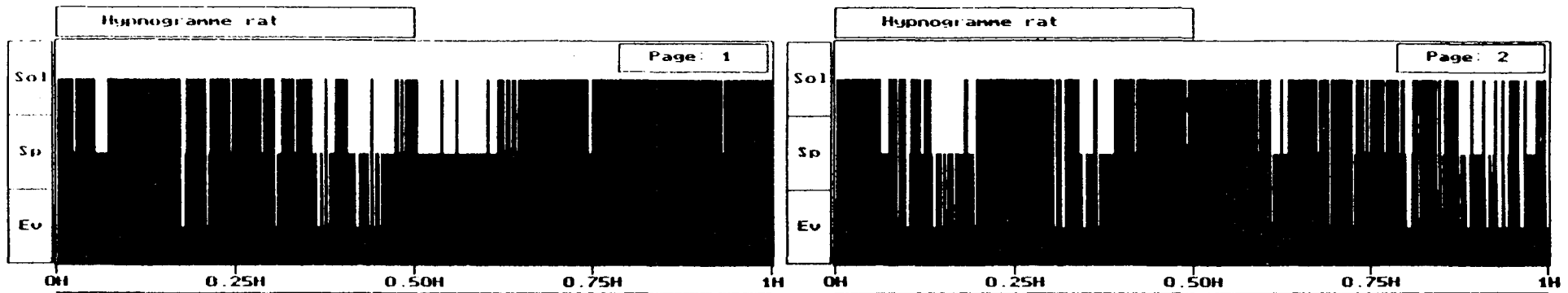
- Si on prend en compte l'évaluation présentée par le tableau précédent, on s'attendrait à ce que l'utilisateur choisisse la version SLP7 (92,99 % de reconnaissance exactes) comme celle donnant les meilleurs résultats en pratique, mais en réalité il préfère la version SLP6 (89,72 %). Ceci s'explique par le fait que la personne qui a évalué le système et celle qui a entrepris l'apprentissage du système sont différentes. Il est donc, tout à fait normal qu'il y est une divergence dans les avis.

- *Limites du système* : un expert humain connaît ses limites, alors que jusqu'à présent les systèmes automatiques réalisés ne fonctionnent pas très bien à cette égard. Ils ont toujours tendance à donner une réponse et par conséquent, il y a souvent un "sur-diagnostic". Ceci peut poser des problèmes et il faut insister sur le fait que le plus raisonnable est d'utiliser ces système en tant qu'outils pour aider plutôt que pour remplacer les experts.

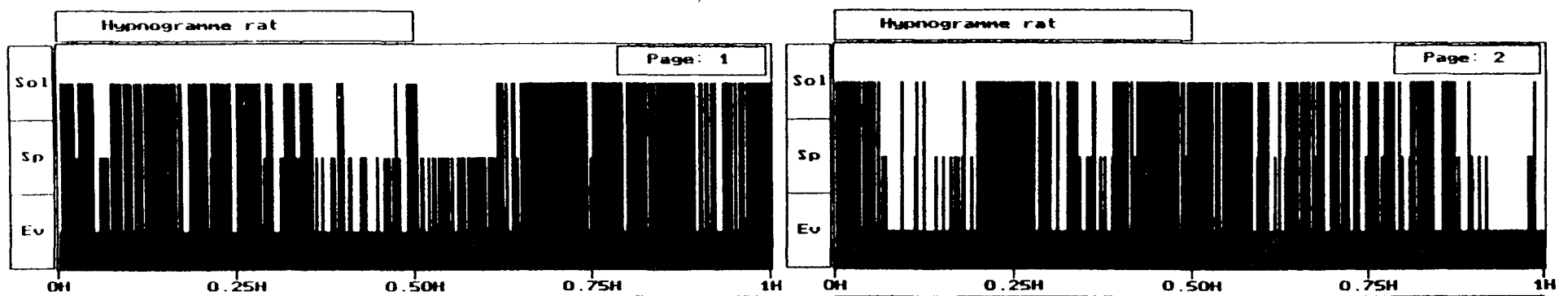
- De nombreux systèmes informatiques sont mis en route sans avoir été complètement testés. Les erreurs sont souvent détectées sur le terrain, alors que les programmes sont déjà en cours. Tester un système présente de réelles difficultés. Les utilisateurs ne sont pas toujours sûrs de la façon dont les systèmes devraient se comporter et donc ils ne peuvent pas les tester à fond.

- *Acceptabilité* : tout le monde n'accepte pas de faire confiance à un ordinateur ou même de s'en servir. Les experts eux-mêmes sont parfois sceptiques à l'égard de ces systèmes, ceci est vraie même quand les systèmes marchent bien et en accord avec les experts.

Non fichier: slp6      Nb signaux: 962      Durée(Heures): 2



Non fichier: slp61      Nb signaux: 962      Durée(Heures): 2



figure(4.5)  
Hypnogrammes d'un même ensemble de données obtenus par SLP6 (haut) et SLP61 (bas)

**CONCLUSION**

Le but du travail était d'appliquer une méthodologie et de montrer la faisabilité d'une instrumentation de reconnaissance des différents stades de sommeil et de veille chez le rat, par la lecture d'un signal EEG unique. Contrairement aux systèmes réalisés jusqu'à présent, à notre connaissance, et qui prennent en compte l'activité électromyographique (EMG) en particulier dans la séparation des deux stades : sommeil paradoxal et éveil.

On a démontré qu'une telle instrumentation pouvait être mise en œuvre même si des améliorations restent à apporter au système :

- augmenter le nombre de rats à traiter en parallèle,
- augmenter le nombre de modèles pour une reconnaissance encore plus fiable,
- introduire une procédure efficace pour le contrôle automatique des niveaux du signal EEG suivant chaque rat,
- introduire une procédure de vérification automatique de la qualité des enregistrements du signal EEG. . .

Ces améliorations peuvent être apportées en grande partie à l'aide d'outils informatiques plus performants, particulièrement en temps de traitement et grâce à la disponibilité totale d'un expert en électrophysiologie pendant toute la durée du projet.

Le temps réel nous a imposé, dès le départ, certaines restrictions :

- fréquence d'échantillonnage, une fréquence plus importante que celle utilisée aurait permis une évaluation plus précise de la fréquence du signal (paragraphe 2-6-3), mais il aurait fallu pour une même portion du signal faire face à un flot plus important de données,
- choix du traitement et de la représentation du signal, un choix rapide et efficace,
- choix des procédures de reconnaissance, une perte minimum d'information entre deux acquisitions du signal EEG, réduire au maximum le temps de traitement . . .

On regrette, toutefois, certains points qui nous ont fait défaut dans ce travail :

- la non disponibilité, durant ce projet, d'un expert ayant une bonne connaissance des signaux électrophysiologiques et pouvant intervenir à n'importe quel moment, en particulier pendant la phase apprentissage. En effet, n'étant pas de formation bio-médicale il se peut que certains détails nous aient échappé,

- la quantité des enregistrements lue par l'expert par rapport à la masse totale des enregistrements est relativement faible,

- les modèles ont été réalisés suivant la reconnaissance des différentes formes établie par une seule personne, il aurait fallu faire lire les enregistrements par plusieurs experts, de manière à réaliser un système qui conviendrait à un maximum d'utilisateur,

- l'évaluation du système a été entreprise par une autre personne que celle qui a aidé à l'apprentissage de celui-ci.

On notera, enfin, que les signaux ont été traités et analysés sans grand rapport avec leur origine physiologique, c'est à dire qu'on ce qui nous concerne l'activité physiologique qui génère ces signaux a moins d'importance que les signaux eux-mêmes. En effet, l'aspect physiologique ainsi que l'interprétation des résultats de reconnaissance font partie du travail de spécialistes. Ceci nous permet d'avancer que les techniques élaborées peuvent facilement être appliquées à des signaux non électrophysiologiques, par exemple à des signaux géophysiques ( séismographie).

En conclusion on dira que l'extraction automatique de caractéristiques pertinentes d'une forme ne pourra pas rivaliser avant longtemps avec la puissance de l'intelligence humaine, cependant on doit tenir compte du caractère pénible d'une telle tâche pour tempérer un tel jugement. La combinaison des procédures décrites dans ce travail avec un système expert doté d'une base de connaissance très complète et où l'aspect sémantique du signal intervient, nous semble être une voie très prometteuse dans ce type d'applications.



**BIBLIOGRAPHIE**

- [1] - ELITHORN A.; BANERJI R. "Artificial and human intelligence", Elsevier, North Holland, 1984.
- [2] - O'SHEAT T.; EISENSTADT M. "Artificial intelligence tools, Techniques and applications", Harper and Row, New York, 1984.
- [3] - RICH E. "Artificial intelligence", McGraw-Hill, 1984.
- [4] - WINSTON P. H. "Artificial intelligence", Addison-Wesley, 1984.
- [5] - HAYES ROTH F.; WATERMAN D.A.; LENAT D.B. "Building experts systems", Addison-Wesley, 1983.
- [6] - VASSEUR C. "La notion d'événement dans les systèmes dynamiques : détection, classification temps réel et application à la conception d'une instrumentation distribuée", Thèse docteur ès-sciences physiques, université de Lille, 1982.
- [7] - PRINCIPE JOSE C.; SMITH JACK R. "SAMICOS - A sleep analyzing microcomputer system.", IEEE Trans. Biomed. Eng., 33, 935-941, 1986.
- [8] - GRASS A.M.; GIBBS F.A. "A Fourier transform of the electroencephalogram", J. Neurophysiol. I, 521-526, 1938.
- [9] - QUINT S.R.; MICHAELS D.F. "A real-time system for the spectral analysis of the EEG", Computer Methods and programs in Biomedicine, 28, 11-22, 1989.
- [10] - MOLINARI L.; DUMERMUTH G. "Robust spectral analysis of the EEG", Neuropsychology, 15, 208-218, 1986.
- [11] - YAKUSH S.A.; WILLEY T.J.; PARK T.C.; MAEDA G. "Multiprocessor architecture for rapid EEG frequency analysis", IEEE Trans. Biomed. Eng., 29, 55-56, 1982.
- [12] - GIESE DOUGLAS A.; BOURNE J.R. "Syntactic analysis of the electroencephalogram", IEEE Trans. Syst. Man and Cybern., 9, 429-435, 1979.
- [13] - FLOOH E.; KORNER E. "Computer evaluation of sleep", Eur. Neurol., 25, 46-52, 1986.
- [14] - "Bibliography of recent literature in sleep research.", Sleep, 9, 527-38, 1986.
- [15] - LIBERSON W.T. "Intelligence, cognition, confusion, dementia and the law of three and a half : Toward a new mathematical equation of sleep", Electromyogr. Clin. Neurophysiol., 26, 309-314, 1986.
- [16] - CLARK F.M.; RADULOVACKI M. "An inexpensive Sleep-Wake state analyzer for the rat", Physiol. and Behav., vol 43, 681-683, 1988.
- [17] - CHOUVET G.; ODET P.; VALATX J.L.; PUJOL J.F. "An automatic sleep classifier for laboratory rodents.", Wak. Sleep, 4, 9-31, 1980.
- [18] - GOTTESMANN C.L.; KIRKHAM P.A.; LACOSTE G.; RODRIGUES L.; ARNAUD C.H. "Automatic analysis of the sleep-waking cycle in the rat recorded by miniature telemetry.", Brain Res., 132, 562-568, 1977.

- [19] - JOHNS T.G.; PIPER D.C.; JAMES G.W.L.; BIRTLEY R.D.N.; FISCHER M. "Automated analysis of sleep in the rat.", *Electroencephalogr. Clin. Neurophysiol.*, 43, 103-105, 1977.
- [20] - MENDELSON W.B.; VAUGHN W.J.; WALSH M.J.; WYATT R.J.A. "Signal analysis approach to rat sleep scoring instrumentation.", *Wak. Sleep.*, 4, 1-8, 1980.
- [21] - NEUHAUS H.U.; BORBELY A.A. "Sleep telemetry in the rat. II. Automatic identification and recording of vigilance states.", *Electroencephalogr. Clin. Neurophysiol.*, 44, 115-119, 1978.
- [22] - VIVALDI E.A.; PASTEL R.H.; FERNSTROM J.D.; HOBSON J.A. "Long term stability of rat sleep quantified by microcomputer analysis.", *Electroencephalogr. Clin. Neurophysiol.*, 58, 253-265, 1984.
- [23] - DANGUIR J. "Influence des apports nutritifs et leurs conséquences métaboliques sur le sommeil chez le rat.", thèse doctorat ès-sciences naturelles, Paris-sud, 1980.
- [24] - BAUDRY M. "Etude du signal vocal dans sa représentation amplitude-temps", thèse doctorat ès-sciences mathématiques, Paris VI, 1978.
- [25] - DUPEYRAT B. "Reconnaissance automatique de la parole, méthode des passages par zéro", thèse doctorat 3ième cycle, Paris VI, 1975.
- [26] - LANSFORD J.L. "Robust noise suppression techniques for neural signals", *IEEE eng. in medicine and biology*, 11th annual intern. conf., 681-682, 1989.
- [27] - PRINCIPE J.C.; SMITH J.R. "Design and implementation of linear phase FIR filters for biological signal processing", *IEEE trans. biomed.eng.*, vol BME-33, 550-559, 1986.
- [28] - PRINCIPE J.C.; SMITH J.R.; BALAKRISHNAN S.K.; PAIGE A. "Microcomputer based digital filters for EEG processing", *IEEE trans. acoust., speech, signal processing*, vol ASSP-27, 697-705, 1979.
- [29] - BILDSTEIN P. "Filtres actifs", éditions radio, 1980.
- [30] - KAWAKAMI M. "Nomographs for Butterworth and Tchébyscheff filters", *IEE trans. circuit theory*, juin 1963.
- [31] - RICE S.O. "Mathematical analysis of random noise", *Bell syst. tech. journal*, vol. 23, 282-339, july 1944, vol. 24, 46-156, january 1945.
- [32] - KRIEG W.J.S. "Connection of the cerebral cortex. I. The albino rat. Topography of the cortical areas", *J.Comp.Neurol.* 84, 221-275, 1946.
- [33] - TIMO-IARIA C., NEGRAO N. "Phases and states of sleep in the rat", *Physiol. and Behav.*, vol 5, 1057-1062, 1970.
- [34] - ROLDAN E., WEISS T. "The cycle of sleep in the rat", *Biol. Inst. Est. Med. Biol, Mex* 20, 155-164, 1962.

- [35] - SOULAIRAC A.; GOTTESMANN C.; THANGAPKEGASSAM M.J." Etude électrophysiologique des différentes phases de sommeil chez le rat", Archsital. Biol., 103, 469-482, 1965.
- [36] - CELEUX G.; DIDAY E.; GOVAERT G.; LECHEVALIER Y.; RALAMBONDRAINY H."Classification automatique des données", ed. Dunod informatique, 1989.
- [37] - VOLLE M."Analyse des données", ed. Economica, 1985.
- [38] - FOUCART T."Analyse factorielle, programmation sur micro-ordinateur", ed. Masson, 1984.
- [39] - BENZECRI J.P. et collaborateurs."Analyse des données", Dunod, 1973.
- [40] - DE LAGARDE J."Initiation à l'analyse des données", ed. Dunod, 1983.
- [41] - DIDAY E.et collaborateurs."Optimisation en classification automatique", ed. INRIA, 1980.
- [42] - " SICLA, système interactif de classification automatique, manuel de l'utilisateur", ed. INRIA, 1987.
- [43] - FAURE A."Perception et reconnaissance des formes.", ed. Test.
- [44] - SIMON J.C."La reconnaissance des formes par algorithmes", ed. MASSON, 1985.
- [45] - SOMMERVILLE I."Le genie logiciel et ses applications", ed. InterEditions, 1988.

## **RESUME DE THESE**

L'origine physiologique de certains signaux leur donne des caractéristiques complexes difficiles à interpréter. L'oeil "expert" de l'opérateur humain peut interpréter tel ou tel signal, non sans se tromper dans son jugement de temps à autre. En effet, la reconnaissance de forme de signaux électrophysiologiques par l'opérateur humain, puis la décision restent très subjectives et peuvent varier d'un opérateur à un autre. Ce travail a donc, pour objectif de présenter une approche méthodologique globale au problème de traitement automatique des signaux électrophysiologiques, autour d'une instrumentation pour la reconnaissance automatique, temps réel, des différents stades de sommeil et de veille chez le rat à partir du traitement d'un signal électroencéphalographique cortical : à partir d'un enregistrement du signal EEG, reconnu par un expert, on a élaboré des représentations-modèles des différents stades auxquelles on a associé des procédures de reconnaissance de forme en temps réel.

## **MOTS-CLES**

**TRAITEMENT DU SIGNAL, CLASSIFICATION AUTOMATIQUE, EEG, RECONNAISSANCE DE FORME, SOMMEIL.**

## **RESUME DE THESE**

L'origine physiologique de certains signaux leur donne des caractéristiques complexes difficiles à interpréter. L'oeil "expert" de l'opérateur humain peut interpréter tel ou tel signal, non sans se tromper dans son jugement de temps à autre. En effet, la reconnaissance de forme de signaux électrophysiologiques par l'opérateur humain, puis la décision restent très subjectives et peuvent varier d'un opérateur à un autre. Ce travail a donc, pour objectif de présenter une approche méthodologique globale au problème de traitement automatique des signaux électrophysiologiques, autour d'une instrumentation pour la reconnaissance automatique, temps réel, des différents stades de sommeil et de veille chez le rat à partir du traitement d'un signal électroencéphalographique cortical : à partir d'un enregistrement du signal EEG, reconnu par un expert, on a élaboré des représentations-modèles des différents stades auxquelles on a associé des procédures de reconnaissance de forme en temps réel.

## **MOTS-CLES**

**TRAITEMENT DU SIGNAL, CLASSIFICATION AUTOMATIQUE, EEG, RECONNAISSANCE DE FORME, SOMMEIL.**