

UNIVERSITÉ DE MONTRÉAL



**UN RÉSEAU DE NEURONES À DÉCHARGES POUR LA  
RECONNAISSANCE DE PROCESSUS SPATIO-TEMPORELS**

TUONG VINH HO

DÉPARTEMENT DE GÉNIE ÉLECTRIQUE  
ET DE GÉNIE INFORMATIQUE  
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

THÈSE PRÉSENTÉE EN VUE DE L'OBTENTION  
DU DIPLÔME DE PHILOSOPHIAE DOCTOR (Ph.D.)  
(GÉNIE ÉLECTRIQUE)

JANVIER 1999

© Tuong Vinh Ho, 1999.



National Library  
of Canada

Acquisitions and  
Bibliographic Services

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

Bibliothèque nationale  
du Canada

Acquisitions et  
services bibliographiques

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file* *Votre référence*

*Our file* *Notre référence*

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-42823-0

Canada



UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL



Cette thèse intitulée:

**UN RÉSEAU DE NEURONES À DÉCHARGES POUR LA  
RECONNAISSANCE DE PROCESSUS SPATIO-TEMPORELS**

présentée par: HO Tuong Vinh

en vue de l'obtention du diplôme de: Philosophiae Doctor

a été dûment acceptée par le jury d'examen constitué de:

M. PLAMONDON Réjean, Ph.D, président

M. MERLO Ettore, Ph.D, directeur

M. ROUAT Jean, Ph.D, co-directeur

M. BRAULT Jean-Jules, Ph.D, membre

M. VILLA Alessandro, Ph.D, examinateur externe

## REMERCIEMENTS

**J**e tiens avant tout à remercier très vivement la personne sans laquelle tout ceci n'existerait pas: il m'importe en effet d'exprimer toute ma gratitude à Jean Rouat, mon codirecteur de recherche et professeur à l'Université du Québec à Chicoutimi. Ses questions pertinentes, son ouverture d'esprit et ses conseils m'ont permis de réaliser ce travail. Je tiens à mentionner qu'il m'a accordé une bourse d'études pour la durée de mon doctorat, sans quoi je n'aurais pas pu poursuivre mes études. Je le remercie particulièrement pour avoir corrigé ma thèse.

Je tiens à remercier Ettore Merlo, mon directeur de recherche à l'École Polytechnique de Montréal pour l'attention et le support qu'il m'a accordés.

Je remercie Messieurs les membres du jury: Réjean Plamondon, directeur général de l'École Polytechnique; Jean-Jules Brault, professeur à l'École Polytechnique; Alessandro Villa, directeur du laboratoire de neuro-heuristique à l'Université de Lausanne (Suisse). Je les remercie pour l'attention et la considération qu'ils ont accordées à mon travail.

Le travail est réalisé dans le cadre du programme de doctorat conjoint entre l'École Polytechnique de Montréal et l'Université du Québec à Chicoutimi.

Je remercie l'équipe ERMETIS (Équipe de Recherche en Microélectronique et Traitement Informatiques des Signaux) de l'UQAC où ce travail a été réalisé. J'ai apprécié le soutien technique et l'amitié que les membres de l'équipe m'ont accordés.

Je remercie toutes les personnes qui m'ont aidé lors de la rédaction de cette thèse (Mohamed Sakami, Hassan Ezzaidi, Claudia Fritz).

Je voudrais également remercier mes parents, ma sœur et mes frères qui m'encouragent par de nombreuses longues lettres depuis le début de mes études au Canada.

Pour son courage et sa générosité, j'exprime toute ma gratitude à ma femme Pham Thi Hoa qui, elle, est dans la rédaction de son mémoire de maîtrise en littérature. Grâce à elle, ce travail a pu être achevé à temps.

Un merci tout particulier à mon fils Tuong Minh Ho, âgé de 6 mois, il a donné toute une autre dimension à ma vision de la vie.

Merci à tous !

Công cha như núi Thái sơn  
Nghĩa mẹ như nước trong nguồn chảy ra.  
Một lòng thờ mẹ kính cha  
Cho tròn chữ hiếu mới là đạo con.

Comme le mont Thai Son, ce que l'enfant doit à son père,  
Et à sa mère, un dévouement qui, comme l'eau s'écoulant d'une source, ne tarit jamais.  
De tout son cœur, vénérer à son père et à sa mère,  
Faire ce que dit la piète filiale, ainsi l'enfant se conduit envers ses parents.



## RÉSUMÉ

**M**algré les succès de l'application des réseaux de neurones à la reconnaissance des formes statiques, le traitement ou la reconnaissance des processus dynamiques non stationnaires (où le temps est une composante essentielle) avec les réseaux de neurones est encore un grand défi et est non résolu. Les travaux récents en neurophysiologie indiquent qu'un comportement complexe, comme par exemple, le chaos, la synchronisation groupée, et la synchronisation de décharges de certaines zones cérébrales, pourrait intervenir dans des processus de mémorisation et de perception. De plus, les outils mathématiques liés à l'étude des phénomènes dynamiques et non linéaires permettent d'élargir notre compréhension de certains mécanismes neuronaux. Ces concepts récents ouvrent une nouvelle voie dans le développement de réseaux de neurones (neuromimétiques et formels).

Plutôt que de résoudre un problème donné ou d'améliorer une technique connue, l'idée fondatrice de cette thèse était de concevoir et d'expérimenter une architecture de réseau de neurones fondée sur de nouveaux principes inspirés de la dynamique chaotique (au sens des systèmes dynamiques) et des observations en neurophysiologie. Le but de cette thèse est donc de proposer et d'étudier une architecture neuromimétique novatrice qui permet de traiter l'information spatio-temporelle d'un processus non stationnaire.

Ce travail revêt un caractère multidisciplinaire et fera appel aux outils disponibles suivants: simulateurs de réseaux formels; simulateur de réseaux neuromimétiques; et simulateur du système périphérique auditif.



Les apports de la présente thèse se situent sur différents plans:

- ♦ sur le plan des innovations, nous proposons une nouvelle architecture neuromimétique inspirée de la couche IV du cortex. Cette architecture est dotée de prédispositions appropriées au traitement et à la reconnaissance de processus dynamiques non stationnaires.
- ♦ sur le plan des applications, nous avons expérimentalement montré que le réseau proposé possède la capacité d'effectuer le traitement et la reconnaissance des processus spatio-temporels non stationnaires à travers les tâches telles que la reconnaissance des chiffres bruités, le traitement des séquences temporelles, la détection de mouvement dans des séquences d'images et le traitement des enveloppes à la sortie d'un banc de filtres cochléaires permettant de réaliser un prototype de système d'identification du locuteur.

## ABSTRACT

**I**n spite of the fact that neural networks have been applied successfully in static pattern recognition, their applicability to the processing or the recognition of dynamic non stationary patterns (where time is an essential component) are still a great challenge and unsolved. Recent works in neurophysiology indicate that complex behaviors such as chaos, grouped synchronization, and spiking synchronization of certain cerebral zones could intervene in the processes of memorizing and of perception. Moreover, mathematical tools related to the study of non linear dynamics allow to widen our comprehension of certain neuronal mechanisms. These recent concepts open a new way in the development of neural networks.

Rather than to solve a given problem or to improve a known technique, the fundamental idea of this thesis is to develop and to test a neural network architecture based on new principles inspired from chaotic dynamics and from observations in neurophysiology. The goal of the research presented here is thus to propose and to study an innovative neural network architecture that allows to treat temporal information of non stationary spatio-temporal processes.

This work adopts a multidisciplinary character and uses the following available tools: simulators of artificial neural networks and simulators of the auditory peripheral system.

The contributions of this thesis covers two aspects:

- ♦ on the aspect of innovations, we propose a new neural network architecture inspired of layer IV of the cortex. This architecture has characteristics that are appropriates to the processing and recognition of non stationary dynamic processes.
- ♦ on the aspect of applications, we have experimentally shown that the proposed network is able to process and to recognize non-stationary spatio-temporal processes through tasks such as recognizing noisy digits, processing temporal sequences, detecting movement in sequences of images, processing envelopes obtained from a bank of cochlear filters and realizing a prototype of speaker recognition system.

## TABLE DES MATIÈRES

<b>Remerciements</b> .....	iv
<b>Résumé</b> .....	vii
<b>Abstract</b> .....	ix
<b>Table des matières</b> .....	xi
<b>Liste des tableaux</b> .....	xiv
<b>Liste des figures</b> .....	xv
<b>INTRODUCTION GÉNÉRALE</b> .....	1
<b>CHAPITRE 1 : TRAITEMENT DES PROCESSUS DYNAMIQUES NON STATIONNAIRES DANS LES RÉSEAUX DE NEURONES</b> .....	4
<b>1.1 Dynamique non linéaire et réseaux de neurones</b> .....	7
<b>1.2 Traitement de l'information dans les systèmes nerveux biologiques ...</b>	11
<b>1.3 Point de vue adopté dans cette thèse</b> .....	16
<b>CHAPITRE 2 : MODÈLE DU RÉSEAU DE NEURONES À DÉCHARGES</b> .....	17
<b>2.1 Modèles du neurone</b> .....	19
2.1.1 Modèle avec fréquence instantanée .....	19
2.1.2 Modèle avec un seuil variable .....	21

<b>2.2</b>	<b>Architecture et apprentissage</b> .....	23
<b>2.3</b>	<b>Activité d'auto-organisation</b> .....	27
2.3.1	Conditions de simulation .....	27
2.3.2	Auto-organisation avec un stimulus d'entrée .....	29
2.3.3	Auto-organisation avec une séquence de stimuli d'entrée .....	36
<b>2.4</b>	<b>Application à la reconnaissance des chiffres bruités</b> .....	41
2.4.1	Critère de reconnaissance .....	41
2.4.2	Détection de la nouveauté .....	41
2.4.3	Comparaison avec d'autres réseaux .....	45
<b>2.5</b>	<b>Réseau avec mécanisme de «feedback» avec récompense</b> .....	50
<b>2.6</b>	<b>Discussion et conclusion</b> .....	57
<b>CHAPITRE 3 : TRAITEMENT DES SÉQUENCES TEMPORELLES ET DÉTECTION DE MOUVEMENT</b> .....		60
<b>3.1</b>	<b>Traitement des séquences temporelles</b> .....	61
3.1.1	Méthodologie .....	62
3.1.2	Expériences avec des séquences de caractères .....	63
3.1.3	Expériences avec des séquences de voyelles .....	67
3.1.4	Discussion .....	70
<b>3.2</b>	<b>Détection de mouvement</b> .....	71
3.2.1	Méthodologie .....	71
3.2.2	Résultat de simulation .....	74
3.2.3	Discussion .....	77
<b>3.3</b>	<b>Discussion et conclusion</b> .....	78

<b>CHAPITRE 4 : PROTOTYPE POUR UN SYSTÈME D'IDENTIFICATION DU LOCUTEUR À L'AIDE DU RÉSEAU PROPOSÉ</b>		80
<b>4.1</b>	<b>Analyse de la parole par modulation d'amplitude dans le système auditif</b>	83
<b>4.2</b>	<b>Système d'identification du locuteur</b>	85
<b>4.3</b>	<b>Traitement des enveloppes par le réseau proposé</b>	88
4.3.1	Données de la parole	88
4.3.2	Analyse des enveloppes à la sortie du banc de filtres	91
<b>4.4</b>	<b>Identification du locuteur basée sur les paramètres de sortie du réseau proposé</b>	101
<b>4.5</b>	<b>Discussion et conclusion</b>	106
<b>CONCLUSION ET PERSPECTIVES</b>		109
<b>BIBLIOGRAPHIE</b>		112
<b>ANNEXES : Articles de référence (articles de conférence internationales sélectionnées sur la base de l'article complet)</b>		120
<b>A</b>	<b>Article :</b> Ho T.V. et Rouat J. (1997). Novelty Detector Using a Network of Integrate-and-Fire Neurons. <u>ICANN97</u> , Switserland.	121
<b>B</b>	<b>Article :</b> Ho T.V. et Rouat J. (1998). Novelty Detection Based on Relaxation Time of a Network of Integrate-and-Fire Neurons. <u>IJCNN98</u> , Alaska, USA.	128
<b>C</b>	<b>Article :</b> Rouat J. et Ho T.V. (1998). La détection de nouveauté basée sur le temps de stabilisation d'un réseau de neurones: application possible en reconnaissance de parole ? <u>Journée d'étude de la parole</u> , Switserland.	135

## LISTE DES TABLEAUX

<b>Tableau 2.1</b> Temps de stabilisation (en itérations) du réseau entraîné par les chiffres propres (0-4) et testé sur (0-9) propres et bruités. ....	44
<b>Tableau 2.2</b> Temps de stabilisation (en itérations) du réseau entraîné par les chiffres propres (5-9) et testé sur (0-9) propres et bruités. ....	44
<b>Tableau 2.3</b> Pourcentage de l'erreur en reconnaissance et en détection de nouveauté pour l'expérience 1: apprentissage sur les versions propres (0-4) et test sur les versions propres et bruitées (0-9). ....	48
<b>Tableau 2.4</b> Pourcentage de l'erreur en reconnaissance et en détection de nouveauté pour l'expérience 2: apprentissage sur les versions propres (5-9) et test sur les versions propres et bruitées (0-9). ....	48
<b>Tableau 2.5</b> Comparaison de performance en terme de pourcentage global de l'erreur en reconnaissance et en détection de nouveauté. ....	49
<b>Tableau 3.1</b> Exemple des séquences de temps de stabilisation .....	64
<b>Tableau 4.1</b> Données de la parole échantillonnée à 8 kHz (contexte «and») ....	89
<b>Tableau 4.2</b> Résultats de l'identification du locuteur sur 10 locuteurs .....	103

## LISTE DES FIGURES

<b>Figure 1.1</b>	Exemples de codages temporels possibles pour le traitement cortical de l'information .....	13
<b>Figure 2.1</b>	Architecture avec un contrôleur global: exemples de zones de voisinage du neurone excitateur (a) et inhibiteur (b) avec un rayon de 2 et 1 respectivement. ....	24
<b>Figure 2.2</b>	Images des chiffres propres (colonnes 1 et 4) et bruités (20% de bruit, colonnes 2,3,5 et 6). ....	28
<b>Figure 2.3</b>	L'évolution dynamique du réseau après sa stimulation par: a) le chiffre <b>0</b> et b) le chiffre <b>2</b> . ....	31
<b>Figure 2.4</b>	Les seuils du réseau après sa stimulation .....	33
<b>Figure 2.5</b>	L'activité de décharge du réseau après sa stimulation .....	34
<b>Figure 2.6</b>	L'activité de décharge pendant la phase de transition .....	35
<b>Figure 2.7</b>	Temps de stabilisation du réseau lorsqu'une séquence de cinq stimuli d'entrée (chiffres 0-4) y est présentée 20 fois. ....	40
<b>Figure 2.8</b>	L'activité de décharge du réseau après sa stimulation par une séquence de cinq chiffres (0-4) 20 fois. ....	40
<b>Figure 2.9</b>	Architecture avec un mécanisme de «feedback» avec récompense. ....	51
<b>Figure 2.10</b>	Les seuils du réseau après sa stimulation (réseau avec un mécanisme de «feedback» avec récompense) .....	54
<b>Figure 2.11</b>	L'activité de décharge pendant la phase de transition (réseau avec un mécanisme de «feedback» avec récompense) .....	55
<b>Figure 3.1</b>	Exemples des images des caractères <i>ABCDE</i> .....	63
<b>Figure 3.2</b>	Transformation de l'ensemble des séquences de <b>5</b> caractères .....	66
<b>Figure 3.3</b>	Transformation de l'ensemble des séquences de <b>26</b> caractères .....	67



<b>Figure 3.4</b>	Transformation de l'ensemble des séquences de 5 voyelles .....	69
<b>Figure 3.5</b>	Transformation de l'ensemble des séquences de 11 voyelles .....	69
<b>Figure 3.6</b>	L'activité du réseau lors de la présentation d'une séquence d'images d'entrée .....	73
<b>Figure 3.7</b>	Détection de mouvement par le réseau proposé .....	76
<b>Figure 4.1</b>	Architecture générale d'un système d'identification du locuteur ...	86
<b>Figure 4.2</b>	Architecture du système d'identification du locuteur à l'aide d'une analyse par modulation d'amplitude et d'un traitement par le réseau proposé. ....	86
<b>Figure 4.3</b>	Enveloppes à la sortie du banc de filtres: a) locuteur <i>sp1497</i> et b) locuteur <i>sp1499</i> . ....	90
<b>Figure 4.4</b>	Traitement des enveloppes par le réseau proposé ayant une dimension 6x4 correspondant à 24 canaux à la sortie du banc de filtres cochléaires. ....	93
<b>Figure 4.5</b>	Séquences de potentiels d'action pour le locuteur <i>sp1497</i> avec $U_{int} = -3.0$ . ....	95
<b>Figure 4.6</b>	Séquences de potentiels d'action pour le locuteur <i>sp1497</i> avec $U_{int} = -5.0$ . ....	96
<b>Figure 4.7</b>	Taux de décharge moyen pour le locuteur <i>sp1497</i> .....	99
<b>Figure 4.8</b>	Taux de décharge moyen pour le locuteur <i>sp1499</i> .....	99
<b>Figure 4.9</b>	Taux de décharge moyen pour le locuteur <i>sp1415</i> .....	100
<b>Figure 4.10</b>	Taux de décharge moyen pour le locuteur <i>sp1096</i> .....	100

## Introduction générale

**C**omprendre les mécanismes à l'origine des fonctions supérieures du cerveau est l'objet de recherches au carrefour de la neurobiologie, de la psychologie, de l'informatique et de la physique. Dans ce vaste champ d'étude, le domaine des *réseaux de neurones* s'est tout particulièrement développé durant les années quatre-vingt. Les «neurones» sont des automates élémentaires dont le mode de fonctionnement s'inspire de celui des neurones biologiques. Il y a deux motivations principales: d'une part la modélisation du cerveau, et d'autre part la réalisation d'algorithmes et de machines spécialisées dans des tâches de reconnaissance des formes par exemple, dont les performances pourraient être supérieures à celles des algorithmes et ordinateurs classiques. Malgré les succès de l'application des réseaux de neurones à la reconnaissance des formes statiques, le traitement ou la reconnaissance des processus non stationnaires avec les réseaux de neurones est encore un grand défi et est non résolu. Les travaux récents en neurophysiologie indiquent qu'un comportement complexe, comme par

exemple, le chaos, la synchronisation groupée, et la synchronisation de décharge de certaines zones cérébrales, pourrait intervenir dans des processus de mémorisation et de perception. De plus, les outils mathématiques liés à l'étude des phénomènes dynamiques et non linéaires permettent d'élargir notre compréhension de certains mécanismes neuronaux. Ces concepts récents ouvrent une nouvelle voie dans le développement de réseaux de neurones (neuromimétiques et formels).

Plutôt que de résoudre un problème donné ou d'améliorer une technique connue, l'idée fondatrice de cette thèse était de concevoir et d'expérimenter une architecture de réseau de neurones fondée sur de nouveaux principes inspirés de la dynamique chaotique (au sens des systèmes dynamiques) et les observations en neurophysiologie. Le but de cette thèse est donc de proposer et d'étudier une architecture neuromimétique novatrice qui permet de traiter l'information spatio-temporelle d'un processus non stationnaire.

Ce travail revêt un caractère multidisciplinaire et fera appel aux outils disponibles suivants: simulateurs de réseaux formels; simulateur de réseaux neuromimétiques; et simulateur du système périphérique auditif.

Les **apports** de la présente thèse se situent sur différents plans:

- ♦ sur le plan des **innovations**, nous proposons une nouvelle architecture neuromimétique inspirée de la couche IV du cortex. Cette architecture est dotée de prédispositions appropriées au traitement et à la reconnaissance de processus dynamiques non stationnaires.

- ◆ sur le plan des **applications**, nous avons développé des logiciels de simulation de l'architecture proposée et les avons appliqués aux problèmes de la reconnaissance de séquences temporelles, de la détection de mouvement et de l'identification du locuteur.

Le présent document s'organise en cinq parties. Nous essayons tout d'abord d'établir un lien entre la dynamique chaotique (au sens des systèmes dynamiques), les observations en neurophysiologie et les réseaux de neurones artificiels afin de retirer les grands principes permettant de concevoir des nouveaux modèles de réseaux de neurones capables de traiter l'information spatio-temporelle d'un processus non stationnaire.

La seconde partie est consacrée à la présentation et l'étude du modèle de réseau proposé. Nous y présentons le modèle ainsi que l'analyse de son comportement sous l'optique du traitement de l'information dynamique spatio-temporelle.

Nous présentons dans la troisième partie la capacité de traiter de l'information spatio-temporelle du modèle de réseau à travers deux applications exemplaires: la reconnaissance des séquences temporelles et la détection de mouvement.

La quatrième partie présente un prototype d'un système d'identification du locuteur à l'aide du modèle de réseau proposé.

La conclusion ainsi que la direction pour la recherche à suivre sont proposées dans la dernière partie.

# Chapitre 1

## Traitement de processus dynamiques non stationnaires dans les réseaux de neurones

**L**e problème de la modélisation et de la reconnaissance de processus spatio-temporels non stationnaires est crucial pour les recherches en réseaux de neurones. L'application de ce problème est très vaste, par exemple, dans la reconnaissance des formes spatio-temporelles: les traces radar, la parole en flot continu et multi-locuteurs, l'écriture manuscrite en continu; dans la planification de trajectoires pour les robots intelligents; dans les systèmes experts couplés en temps réel à des processus industriels; c'est à dire des applications où la «dimension de temps» est prédominante. Mais quel moyen utilise le cerveau pour traiter l'information temporelle ?

Il est généralement considéré que le cerveau est un système dynamique non linéaire, d'où l'intérêt de la recherche en neurosciences (neurophysiologie, informatique) pour le développement des réseaux de neurones dynamiques non linéaires. L'idée principale vient du fait que lorsque le cerveau est soumis à certains stimuli, les signaux transmis aux cellules nerveuses sont traités pour aboutir à une réponse du cerveau qui se caractérise par un comportement observable. Or, on peut noter des relations stables entre certains stimuli et certains comportements, ce qui donne à penser que le système nerveux aboutit à des configurations associatives stables. L'introduction des attracteurs dynamiques (par exemple attracteurs de cycle limite, attracteurs chaotiques) au lieu d'attracteurs de point fixe a ouvert une nouvelle approche pour la recherche dans le domaine des réseaux de neurones. Dans le contexte d'utilisation d'attracteurs dynamiques, la présente recherche essaye de comprendre le rôle de la dynamique non linéaire dans les systèmes nerveux biologiques en terme de traitement de l'information. En d'autres termes, il est nécessaire de découvrir la raison d'être de l'existence de cette dynamique non linéaire et ensuite de trouver des manières pour les manipuler (contrôler) afin de concevoir des systèmes artificiels capables de traiter des processus dynamiques non stationnaires.

Dans le domaine de la modélisation des systèmes nerveux biologiques, plusieurs travaux ont indiqué que l'information est traitée à l'aide d'impulsions (appelées «potentiels d'action»). D'un point de vue de traitement de l'information une question importante est de savoir comment les systèmes nerveux biologiques utilisent les potentiels d'action en vue de coder l'information. De plus, l'architecture des systèmes nerveux biologiques a une grande influence sur le comportement de ces derniers concernant le traitement de l'information.

En ce qui concerne notre recherche, nous essayons de retirer à partir des découvertes dans ces deux domaines (neurodynamique et physiologie) des idées (ou des principes) qui nous servent pour la conception de nouveaux modèles de réseaux de neurones capables de traiter l'information spatio-temporelle non stationnaire.

## 1.1 Dynamique non linéaire et réseaux de neurones

Les réseaux de neurones sont des systèmes à auto-organisation dont la dynamique est essentiellement non linéaire. Le sujet des réseaux de neurones étudié en tant que système dynamique non linéaire est référé comme neuro-dynamique (Haykin, 1994).

Un concept particulièrement important lié à l'étude des réseaux dits neuro-dynamiques est celui de la dynamique chaotique. Un système chaotique est imprévisible mais il est parfaitement décrit par des équations simples et déterministes. Le lien entre ces deux notions paradoxales, déterminisme et imprévisibilité, est la propriété de sensibilité aux conditions initiales : deux conditions initiales semblables peuvent conduire à des états très différents du système. Cette propriété est la principale caractéristique des systèmes chaotiques. La dynamique chaotique joue un rôle important dans l'étude des réseaux de neurones, en particulier en termes de traitement de l'information dynamique spatio-temporelle.

L'introduction de connexion asymétrique permet d'augmenter la richesse en terme de dynamique avec les attracteurs à cycles limites et chaotiques. Elle permet en conséquence de surmonter l'insuffisance des réseaux de type Hopfield qui n'ont qu'une solution statique. Amit (1989) a suggéré que le chaos n'empêche pas l'apprentissage de nouveaux patrons («patterns»); sans chaos il y a une forte tendance qu'un des patrons déjà appris soit renforcé plutôt qu'une mémorisation d'un nouveau patron.



Tsuda (1992, 1997) a suggéré un rôle fonctionnel possible du chaos cortical dû aux interactions thalamo-corticales. Dans ce sens, le chaos est considéré comme un outil pour générer des mémorisations à court terme. Il a proposé un modèle de couplage dynamique de mémoire dans lequel une activité chaotique parcourt les états du réseau. Lorsque l'entrée est fortement différente de tous les patrons appris, le réseau réalise une recherche dans l'espace de mémoire. Par contre, lorsque l'entrée est similaire à un des patrons déjà appris, le réseau est dirigé vers un bassin d'attraction du patron le plus similaire à l'entrée. Hayashi (1994) a observé la même situation dans son réseau oscillant: une génération autonome d'un cycle limite près d'un patron mémorisé et une trajectoire chaotique de recherche parmi les patrons mémorisés selon la nature de l'entrée. Ces observations suggèrent un mécanisme de reconnaissance pour les réseaux dynamiques, par exemple le mécanisme de détection de la nouveauté.

Kanenko (1995) a proposé la notion de stockage de l'information par mémoire hiérarchisée à plusieurs attracteurs. Cette architecture peut être liée à l'activité de la catégorisation dynamique du cerveau. Selon l'entrée, le réseau peut se diriger d'un attracteur à l'autre.

Du point de vue de la conception, le traitement de l'information par les réseaux dynamiques chaotiques est constitué d'une association de l'information à traiter avec le comportement dynamique non linéaire de ces réseaux. En d'autre terme, en introduisant les dynamiques chaotiques avec des attracteurs oscillants et chaotiques, on espère que les réseaux chaotiques auraient une dynamique assez riche pour représenter exclusivement ou implicitement la dynamique de l'information spatio-temporelle.

L'idée générale est d'associer l'information spatio-temporelle à des attracteurs de ces réseaux par l'apprentissage (Osana, 1998; Ishii, 1998). Thiran et Hasler (1996) ont donné un excellent survol sur le traitement de l'information par des réseaux de neurones dynamiques. Hayashi (1994) a réussi à associer les patrons transformés continuellement aux attracteurs de cycle limite par l'apprentissage. Matsugu et Yuille (1994) ont proposé un modèle permettant de représenter la dimension spatiale de l'information par les connexions excitatrices et la dimension temporelle par les connexions inhibitrices. Dans ce réseau, les connexions excitatrices aident à stabiliser spatialement le réseau et les connexions inhibitrices permettent de produire et de reconnaître les séquences temporelles.

Une autre approche est d'associer l'information à l'activité de synchronisation (en fréquence ou/et phase) d'une population de neurones. Wang (1995b) a développé un réseau oscillant avec un inhibiteur global permettant de faire la segmentation en regroupant les populations de neurones ayant la même activité. Dans la même optique, Brown et Wang (1997) ont développé des modèles de réseau oscillant pour faire la segmentation de signaux auditifs provenant de la parole par la synchronisation temporelle des neurones.

Malgré les avantages que la dynamique non linéaire (chaotique) peut apporter à la conception des réseaux de neurones, il y a très peu, jusqu'à présent, d'applications de ce type de réseau pour la reconnaissance. Il semblerait qu'on n'a pas encore trouvé de méthode théorique ou pratique efficace pour associer la dynamique de l'information spatio-temporelle à la dynamique du réseau chaotique. Certains travaux ont réussi à représenter l'information spatio-temporelle par des attracteurs dynamiques, mais ces derniers ont souvent des formes complexes. Par conséquent, la reconnaissance

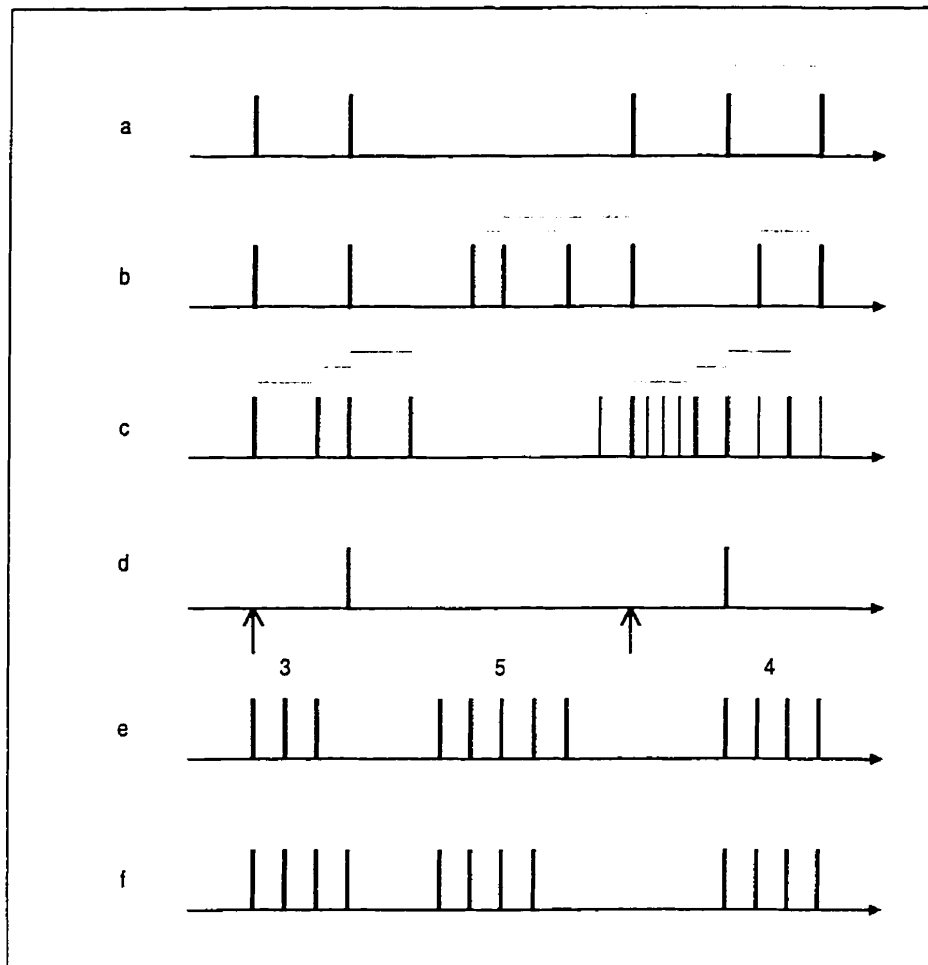
basée sur les attracteurs devient une tâche difficile et parfois plus compliquée que la reconnaissance basée sur l'information de l'entrée originale. Évidemment, il faut trouver d'autres paradigmes d'exploitation de la dynamique chaotique afin de traiter effectivement et efficacement l'information spatio-temporelle. En ce qui concerne notre réflexion, nous proposons des modèles de réseau dynamique dont le temps de stabilisation peut être utilisé comme critère pour la reconnaissance de processus dynamiques spatio-temporels non stationnaires.

## 1.2 Traitement de l'information dans les systèmes nerveux biologiques

Un des défis pour les neurosciences est de déterminer le codage de l'information utilisée dans les systèmes nerveux. Pendant des décennies, la plupart des neurophysiologistes ont considéré que le codage de l'information s'effectue par le taux moyen des potentiels d'action («spike»). Ce codage fonctionne bien si l'entrée («input») est constante ou change dans une échelle temporelle qui est lente par rapport à la durée de la fenêtre temporelle. Cependant, ce codage basé sur une moyenne néglige toute l'information probablement liée au moment exact («timing») des potentiels d'action (Gerstner, 1998). Une approche alternative selon laquelle les patrons de décharge temporels («temporal firing patterns») contiennent l'information est en train de gagner l'attention comme les nouvelles approches théoriques et expérimentales (Ferster, 1995; Gerstner, 1998). Il s'agit d'un codage temporel fin. Différents types de codages temporels sont discutés dans l'article de Cariani (1995): codage par intervalle entre potentiels d'action consécutifs, codage par intervalle entre potentiels d'action consécutifs multiplexés temporellement, codage par forme temporelle complexe, codage par phase (temps d'arrivée par rapport à une référence, codage par nombre de potentiels d'action par groupes («burst»), et codage par écarts entre «bursts» (Figure 1.1).

Jusqu'ici nous avons parlé du train de potentiels d'action d'un seul neurone. Puisqu'il y a tant de neurones dans le cerveau, des milliers de trains de potentiels d'action sont émis constamment par différents neurones. Quelle est l'information contenue dans un tel patron spatio-temporel d'impulsions

(«spatio-temporal pattern of pulse») ? Utilisant plusieurs électrodes à la fois, les neurophysiologistes ont découvert la synchronie dans les patrons de potentiels d'actions qui peut représenter une nouvelle voie de coder l'information (Barinaga, 1998). Cette approche peut aider à résoudre le «binding problem» qui consiste à étudier la façon dont les informations concernant un même objet mais représentées à divers endroits du cortex peuvent être liées et identifiées comme appartenant à un même objet. Le codage par synchronie a été étudié intensivement à la fois par l'expérimentation et par des modèles (Ritz, 1997 pour une synthèse des mécanismes potentiels).



**Figure 1.1** Exemples de codages temporels possibles pour le traitement cortical de l'information (d'après Cariani, 1995; Chappelier, 1996). Les barres larges verticales représentent des potentiels d'action significatifs (*a contrario* ceux en trait fin ne le sont pas). Les barres grisées horizontales soulignent où réside le codage.

- a) *codage par simple intervalle entre potentiels d'actions*
- b) *codage par simple intervalle entre potentiels d'actions multiplexés temporellement*
- c) *codage par forme temporelle complexe*
- d) *codage par phase (temps d'arrivée par rapport à une référence)*
- e) *codage par nombre de potentiels d'action par groupes («burst»)*
- f) *codage par écarts entre «bursts»*

Les travaux concernant le codage de l'information dans les systèmes nerveux ont été à nouveau une source d'inspiration pour la recherche en réseaux de neurones artificiels. Une nouvelle génération de réseaux de neurones est en train d'émerger (Maass, 1996, 1997). Cette génération de réseaux utilise les modèles de neurones inspirés de la biologie que l'on nomme les modèles de neurones à décharge («spiking neurons») ou les modèles de type intégration et décharge («Integrate-and-Fire neurons»). Ce type de neurones reproduit très finement certaines propriétés temporelles des neurones biologiques, par exemple la période réfractaire du neurone (après avoir déchargé, pendant cet intervalle de temps le neurone ne peut pas émettre de décharge, quelque soit le potentiel postsynaptique). Mais les neurones n'ont pas de raison *a priori* de se «mettre à jour» tous en même temps: la dynamique naturelle est donc asynchrone aléatoire. Leur principe est de sommer spatialement et temporellement les entrées. Lorsque cette somme dépasse un seuil qui leur est propre, ils émettent une impulsion (potentiel d'action). En d'autres termes, c'est une activité de décharge («firing activity») du neurone. Il existe de nombreux modèles de neurones de ce type et ils sont d'une certaine façon les modèles simplifiés de celui de Hodgkin et Huxley (Gerstner, 1998).

Ces modèles peuvent être inclus dans les architectures neuronales classiques et leur donnent des capacités temporelles intéressantes en vue de traitement de l'information spatio-temporelle (Maass, 1996; Ruf et Schmitt, 1997). Ce type de réseaux est appelé réseaux de neurones à décharges («spiking networks»). Maass (1996) a démontré que les réseaux de neurones à décharges sont plus performants en terme de calcul que les réseaux basés sur les modèles de neurones McCulloch Pitts. Hopfield (1995) suggère l'utilisation du moment de décharge pour coder l'information et les réseaux à retard pour traiter cette représentation. Un tel codage permet une grande capacité de calcul et une meilleure vitesse par rapport aux réseaux utilisant un codage basé sur le

taux de décharge. De même, Buonomano (1995) propose un réseau de neurones avec des propriétés réelles qui permet une transformation de l'information temporelle en code spatial d'une manière auto-organisée.

La plupart des travaux liés à cette approche portent actuellement sur des mécanismes ou algorithmes d'apprentissage basés sur le moment de décharge exact (le «timing») de l'ensemble des neurones (Ruf et Schmitt, 1998). L'application de ce type de réseaux pour des problèmes d'ingénierie n'a à notre connaissance pas encore été envisagée.



### 1.3 Point de vue adopté dans cette thèse

Partant de l'étude de la dynamique non linéaire, les réseaux de neurones dynamiques et le codage de l'information dans les systèmes nerveux biologiques, nous essayons de tirer des principes en vue de concevoir un modèle de réseau possédant des propriétés dynamiques et spatio-temporelles. En raison de leurs propriétés temporelles, les modèles de neurone à décharge sont selon notre point de vue un candidat potentiel comme élément de base pour un réseau spatio-temporel. Ils ont une dynamique temporelle au niveau local. De plus, un ensemble de neurones de ce type interconnectés pourrait créer des comportements dynamiques et complexes spatio-temporellement au niveau global dans le sens de la dynamique non linéaire. Au lieu d'utiliser les architectures de réseaux formels classiques, nous cherchons une structure inspirée de celles du cortex biologique. En adoptant ce point de vue, nous espérons pouvoir intégrer les propriétés dynamiques spatio-temporelles de ces deux approches (discutées ci-dessus) afin de concevoir de nouveaux modèles de réseaux de neurones qui possèdent des dispositions appropriées au traitement de l'information dynamique spatio-temporelle non stationnaire.

## Chapitre 2

# Modèle du réseau de neurones à décharges

**N**ous proposons un modèle de réseau de neurones à décharges. L'originalité du travail réside principalement dans la définition d'une règle de modification du seuil de décharge du neurone et dans la création du critère de temps de stabilisation pour la reconnaissance. En effet, la majorité des techniques de reconnaissance des formes qui utilisent les réseaux de neurones (souvent formels) se base sur des critères relativement statiques (minimisation d'une fonction d'erreur, maximisation d'une probabilité, etc.) et codent l'information temporelle de façon statique via la structure spatiale des entrées (Haykin, 1994). Par ailleurs, l'approche adoptée ici ne nécessite pas de supervision du réseau. Celui-ci est en mesure de détecter la nouveauté et de

s'adapter de façon autonome. De plus, il n'y a pas de différence entre apprentissage et reconnaissance.

Dans un premier temps, il est important de tester et de valider ce type d'approche vis-à-vis des systèmes plus classiques afin de s'assurer que le réseau proposé puisse réaliser au minimum des tâches similaires à celles exécutées par les réseaux dits formels. Nous présentons donc une série de tests préliminaires effectués en reconnaissance des chiffres bruités. Ainsi, nous effectuons une étude expérimentale du comportement du réseau afin de mieux comprendre ses caractéristiques et de les exploiter aux chapitres 3 et 4 en vue du traitement de l'information dynamique spatio-temporelle.

## 2.1 Modèles du neurone

Au cours de ce travail, nous avons développé deux modèles de neurone qui sont principalement inspirés du modèle de neurone proposé par Hill et Villa (1995, 1997). Ce dernier est utilisé dans la modélisation du système thalamo-cortical. Le premier modèle tient compte exclusivement de la fréquence instantanée de décharge tandis que le deuxième modèle dispose d'un seuil de décharge qui est variable en fonction de sa propre activité de décharge.

### 2.1.1 Modèle avec fréquence instantanée

Le modèle de neurone est de type intégration et décharge («Integrate-and-Fire» ou «spiking neurons») avec période réfractaire et potentiel postsynaptique décroissant. L'état du neurone à l'instant  $t$  est caractérisé par son potentiel de contrôle  $U$ . La réponse  $S_i(t)$  du neurone  $i$  est donnée par:

$$S_i(t) = \begin{cases} 0 & \text{pour } (t - t_{spike}) < \rho, \\ H[U_i(t) - \theta] & \text{pour } (t - t_{spike}) > \rho, \end{cases} \quad (2.1)$$

où  $\theta$  est le seuil et  $H$  est la fonction de Heaviside définie telle que  $H(x) = 1$  pour  $x > 0$ , sinon  $H(x) = 0$ . La valeur  $t_{spike}$  représente le dernier moment de décharge pour le neurone  $i$ . Ainsi, la valeur  $\rho$  représente la période réfractaire absolue du neurone: après avoir déchargé, pendant cet intervalle de temps le

neurone ne peut pas émettre de décharge quel que soit le potentiel de contrôle  $U_i(t)$ . Le potentiel de contrôle est défini par:

$$U_i(t) = U_{i,int} + U_{i,ext}(t) \quad (2.2)$$

où  $U_{i,int}$  est le potentiel de repos («resting potential») et  $U_{i,ext}(t)$  est le potentiel externe. Dans ce réseau, le potentiel externe  $U_{i,ext}(t+1)$  pour le neurone  $i$  est l'intégration de toutes les réponses afférentes à l'instant  $t$ .

$$U_{i,ext}(t+1) = \sum_j C_{ij}(t) S_j(t) + \delta U_{i,ext}(t) + e_i(t) \quad (2.3)$$

$C_{ij}(t)$  est la connexion du neurone  $j$  vers le neurone  $i$  à l'instant  $t$ .  $e_i(t)$  est le signal externe (entrée du stimulus) pour le neurone  $i$  considéré. En fait, l'entrée de chaque neurone est non seulement une intégration spatiale mais également une intégration temporelle implicite de l'activité de décharge. Le potentiel externe décroît selon la variable cinétique de premier ordre  $\kappa$  avec  $\delta = \exp(-\ln(2)/\kappa)$  (Hill et Villa, 1995). Il est également réinitialisé à 0 après une décharge.

Le temps est discret et mesuré en unités d'une itération. Dans ce modèle, il y a une exécution d'une dynamique parallèle: on évalue à chaque instant tous les paramètres des neurones et du réseau.

Afin de tenir compte de l'influence de l'activité de décharge dans le comportement du neurone, nous introduisons la fréquence instantanée de décharge  $f_i$  dans l'équation 2.3.

$$U_{i,ext}(t+1) = \sum_j C_{ij}(t)S_j(t) + \delta U_{i,ext}(t) + e_i(t) + \mu f_i(t) \quad (2.4)$$

En fait, ce facteur représente un compteur de potentiels d'action du neurone. Il est pondéré par une constante  $\mu$  qui varie entre 0.1 et 0.9. L'introduction de la fréquence instantanée permet de créer une distinction entre des neurones en fonction de leur activité de décharge. Cependant, tel qu'elle est introduite dans l'équation 2.4, ce facteur a une influence non équilibrée par rapport aux autres paramètres. Cette influence devient trop grande si le neurone considéré se décharge assez fréquemment. Autrement dit, le potentiel externe du neurone ne dépend, assez rapidement, que de la fréquence instantanée de décharge.

### 2.1.2 Modèle avec un seuil variable

Afin d'améliorer le premier modèle de neurone, nous cherchons une autre manière permettant d'intégrer l'influence de l'activité de décharge dans le comportement du neurone. Nous introduisons donc un seuil de décharge (du neurone) qui est variable en fonction de sa propre activité de décharge. Notons que les équations 2.1 à 2.3 de la première version sont maintenues et l'équation 2.4 n'est plus utilisée dans ce modèle.

Le seuil de décharge est alors défini tel que:

$$\theta_i(t) = b_i(t) - a_i(t) \quad (2.5)$$

$a_i(t)$  et  $b_i(t)$  sont des facteurs qui correspondent à l'activité de décharge et non-décharge du neurone. A chaque instant, si le neurone considéré se décharge le facteur de décharge  $a_i(t)$  sera augmenté. Autrement, le facteur de non décharge  $b_i(t)$  sera augmenté. En d'autres termes, si le neurone se décharge, son seuil de décharge sera diminué (équation 2.5) et ceci facilite son aptitude à décharger plus tard. Dans le cas contraire, si le neurone ne se décharge pas son seuil sera augmenté et il lui sera donc plus difficile de se décharger ultérieurement. Ces facteurs changent au cours du temps et sont mis à jour par:

$$a_i(t + 1) = a_i(t) + \lambda a_i(t)(1 - a_i(t)) \quad (2.6)$$

$$b_i(t + 1) = b_i(t) + \lambda b_i(t)(1 - b_i(t)) \quad (2.7)$$

$\lambda$  est une constante de mise à jour qui varie entre 0.1 et 0.9. À partir des équations 2.6 et 2.7, on peut constater que les facteurs  $a_i(t)$  et  $b_i(t)$  évoluent dans le temps selon une fonction logistique. La quantité de changements pour ces facteurs est proportionnelle à leur dérivée. L'expression  $a(t)(1-a(t))$  est la dérivée de la fonction logistique  $a(t) = 1 / (1 + \exp(-t))$ . Les facteurs  $a_i(t)$  et  $b_i(t)$  sont initialisés par une valeur positive fixe plus petite que 1. Il est important de noter que ces facteurs ne sont mis à jour que lorsque le neurone est en dehors de sa période réfractaire.

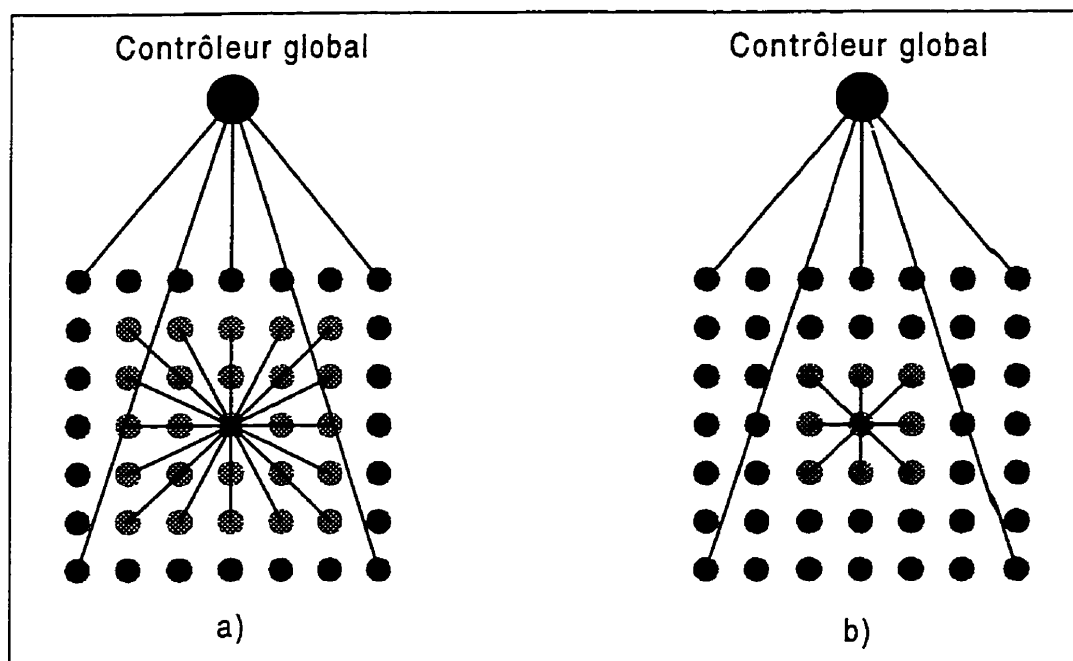
De cette manière, le seuil de décharge du neurone dépend fortement de sa propre activité de décharge. De plus, le fait que ce seuil varie entre  $-1$  et  $+1$  permet un meilleur équilibre en terme d'influence par rapport aux autres paramètres en comparaison avec le modèle de neurone précédent.

## 2.2 Architecture et apprentissage

L'architecture est inspirée du modèle de la couche IV du cortex tel que proposé par Hill et Villa (Hill, 1995, 1997). Ils définissent une couche à 2 dimensions comprenant des neurones inhibiteurs et excitateurs avec récursivité. Les neurones sont positionnés aléatoirement selon une distribution Sobol (Press, 1992) et les connexions entre eux sont constantes. Dans notre modèle de réseau, les connexions sont adaptées selon un mécanisme d'apprentissage. Chaque neurone est interconnecté à ceux qui appartiennent à son ensemble de voisinage. Cet ensemble est défini comme étant carré, centré autour du neurone et comprend une taille différente selon le type du neurone considéré (inhibiteur ou excitateur). En effet, la taille de la zone de voisinage est une variable. Même si le modèle de Hill et Villa aide à observer et étudier l'évolution de l'activité d'auto-organisation spatio-temporelle dans un cortex simulé, l'apprentissage et ses mécanismes n'y sont pas abordés. Dans ce contexte, nous proposons un modèle de réseau inspiré de cette architecture ainsi que le mécanisme d'apprentissage pour ce modèle de réseau à base de neurones à décharges.

Selon la topologie décrite au-dessus on peut dire que le modèle tel que proposé par Hill et Villa utilise une activité avec interactions au niveau local pour créer une évolution d'auto-organisation. Ici, nous voulons modifier le modèle par l'introduction d'une activité avec interactions au niveau global en utilisant un contrôleur dit global (Figure 2.1).





**Figure 2.1** Architecture avec un contrôleur global: exemples de zones de voisinage du neurone excitateur (a) et inhibiteur (b) avec un rayon de 2 et 1 respectivement.

Le contrôleur global est en fait un déclencheur dont l'état est soit actif, dans ce cas, il se décharge, soit inactif selon un mécanisme de contrôle. Ce dernier est basé sur un seuil de contrôle pour le nombre total de neurones de décharge (les neurones qui se déchargent en même temps). Lorsque ce nombre à l'instant  $t$  est supérieur à un seuil de contrôle, le contrôleur global se déclenche et génère un signal de «feedback» négatif (-1) pour tous les neurones du réseau. Par contre, si le nombre de décharges est inférieur au seuil, le contrôleur global génère un signal de «feedback» positif (+1). De cette manière, le contrôleur global joue un rôle de régularisation de l'activité des neurones au niveau global. Étant donné le signal de «feedback»  $h$  ( $h = -1$  ou  $+1$ ), l'équation 2.3 devient:

$$U_{i,ext}(t+1) = \sum_j C_{ij}(t)S_j(t) + \delta U_{i,ext}(t) + e_i(t) + G_i(t)h(t) \quad (2.8)$$

$G_i(t)$  est le poids de connexion entre le contrôleur global et le neurone  $i$ .

Dans cette architecture, il y a deux types de couplage: le premier est celui entre les neurones de voisinage et le deuxième est celui entre le contrôleur global et tous les neurones. Les poids de connexions entre les neurones voisins sont mis à jour selon une règle proposée par (Stassinopoulos et Bak, 1995):

$$C_{ij}(t+1) = C_{ij}(t) + \alpha C_{ij}(t)(1 - C_{ij}(t))S_i(t)S_j(t) \quad (2.9)$$

$\alpha$  est le taux d'apprentissage.

Les poids de connexions entre le contrôleur global et les neurones sont mis à jour par:

$$G_i(t+1) = G_i(t) + \beta G_i(t)(1 - G_i(t))S_i(t)h(t) \quad (2.10)$$

où  $\beta$  est une constante de mise à jour.

Les poids de connexions du réseau sont mis à jour tout au long de la simulation. À chaque instant, les connexions d'un neurone sont mises à jour si celui-ci se décharge. Il reçoit également un signal de «feedback» à partir du contrôleur global. Ce dernier est soit positif ou négatif suivant le nombre de

neurones de décharge qu'il soit supérieur ou inférieur au seuil du contrôleur global.

Toutefois, il faut admettre que la notion de contrôleur global n'existe probablement pas dans les systèmes nerveux.

Dans les sections précédentes, nous avons présenté deux modèles pour le neurone. Afin d'alléger la thèse, nous nous concentrons sur le modèle de réseau à base de neurones ayant un seuil de décharge variable dans la suite de cette thèse. Cependant, quelques résultats préliminaires concernant le modèle de réseau à base de neurones qui utilisent la fréquence instantanée sont présentés à l'annexe A (Ho, 1997).

## 2.3 Activité d'auto-organisation

Dans cette section, nous faisons une étude du comportement et des caractéristiques dynamiques du modèle de réseau proposé dans deux situations suivantes:

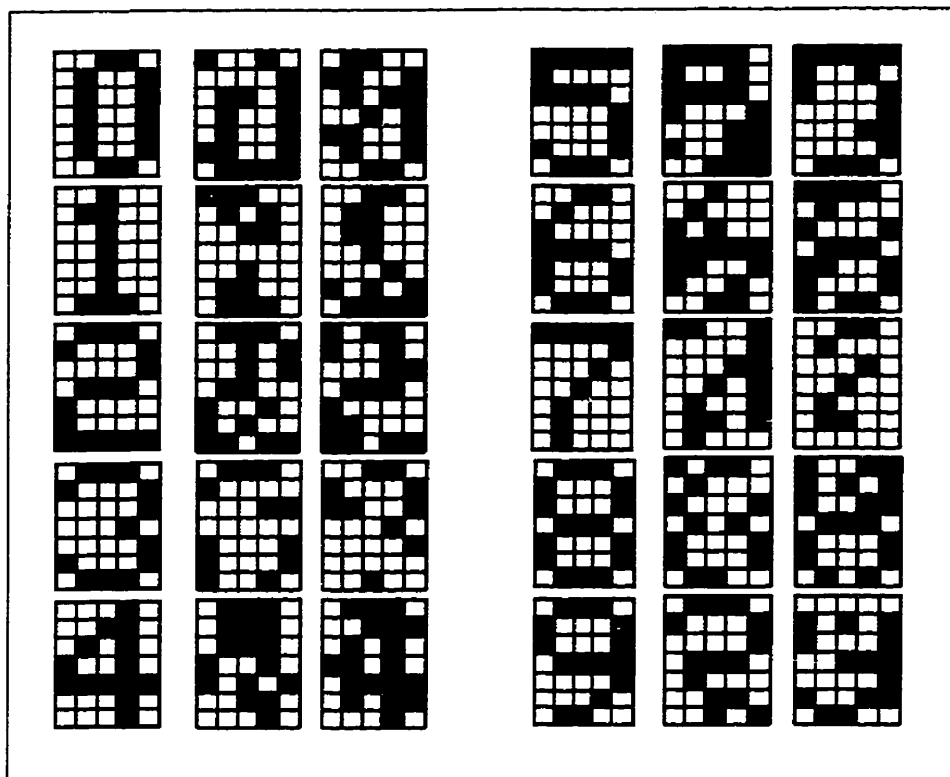
- ◆ lorsqu'il est stimulé par un seul stimulus;
- ◆ lorsqu'il est stimulé par une séquence de stimuli.

Les expériences portant sur ces situations permettent de comprendre l'aptitude du réseau à traiter l'information structurée spatio-temporellement. De plus, nous nous intéressons particulièrement aux points suivants:

- ◆ l'étude sur l'auto-organisation en se basant sur le temps de stabilisation du réseau;
- ◆ l'étude sur l'auto-organisation en se basant sur l'activité de décharge du réseau.

### 2.3.1 Conditions de simulation

Un ensemble des chiffres de 0 à 9 codés sur une matrice binaire de 7 par 5 est utilisé comme stimuli d'entrée (Figure 2.2, colonnes 1 et 4). Même si les images des chiffres utilisées ici sont binaires, le réseau proposé peut manipuler les images analogiques (c'est-à-dire que les nombres réels peuvent être manipulés par le réseau).



**Figure 2.2** Images des chiffres propres (colonnes 1 et 4) et bruités (20% de bruit, colonnes 2, 3, 5 et 6).

Le réseau est composé de 70% de neurones excitateurs et de 30% de neurones inhibiteurs. Il comprend 35 neurones (7 par 5). Le rayon de voisinage est de 2 pour les neurones excitateurs et de 1 pour les neurones inhibiteurs. Les chiffres sont codés sur une matrice de pixels. Chaque pixel est donc présenté sur l'entrée  $e(t)$  d'un neurone correspondant dans la matrice du réseau. Les paramètres du réseau sont initialisés tels que:  $a(0) = 0.1$ ,  $b(0) = 0.1$ ,  $U_{int} = -3.0$ ,  $\kappa = 3.0$ ,  $\rho = 3.0$ ,  $\alpha = 0.2$ ,  $\beta = 0.2$ ,  $\lambda = 0.1$ , et le seuil pour le contrôleur global (c'est-à-dire le nombre de neurones de décharge en même temps) est de 50% du nombre total de neurones du réseau. Les poids de connexions sont initialisés aléatoirement dans un intervalle de 0 à 1. Le temps est discret et mesuré en unités d'une itération.

## 2.3.2 Auto-organisation avec un stimulus d'entrée

### 2.3.2a Activité de stabilisation

L'état du réseau au cours de la simulation ne peut pas être analysé au moyen d'un algorithme simple. Pourtant, des observations de l'état du réseau peuvent aider à étudier les caractéristiques d'auto-organisation du réseau. Après être stimulé par un stimulus, le réseau a un comportement complexe. On dit qu'il se stabilise lorsque les changements des poids sont inférieurs à un seuil préétabli (par exemple 0.001 dans ces expériences). Afin de mieux comprendre le comportement complexe du réseau, nous étudions l'évolution des changements des poids de connexions au cours de la phase d'apprentissage. Par simplicité, la valeur maximale des changements des poids de connexions à chaque instant est utilisée pour caractériser la stabilisation. Cette valeur est définie comme étant une différence maximale de poids, notée  $m$ :

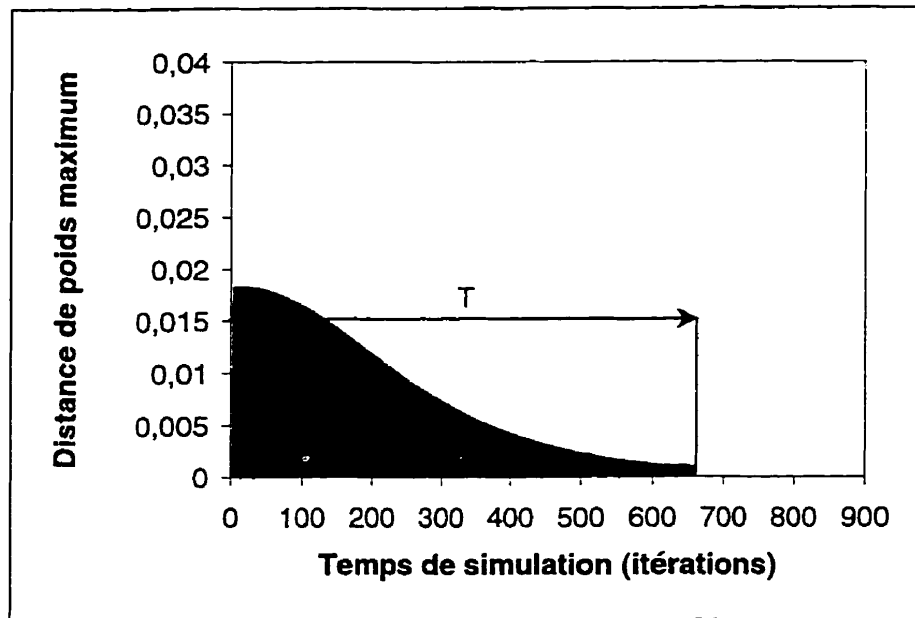
$$m = \max\{|C_{ij}(t+1) - C_{ij}(t)|\} \text{ pour tous les } i, j. \quad (2.12)$$

À partir d'ici, nous utilisons la différence maximale de poids  $m$  comme une représentation de l'évolution dynamique du réseau. Dans le texte, nous appelons  $m$  distance maximale de poids de connexion.

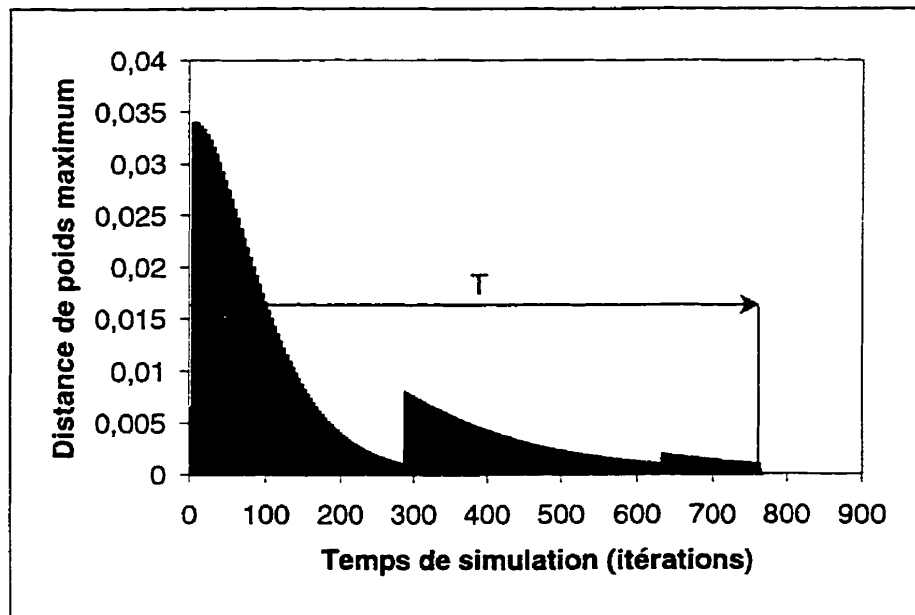
La figure 2.3 montre l'évolution de la distance de poids de connexions du réseau après sa stimulation par les chiffres 0 et 2 respectivement. Les mêmes conditions initiales ont été utilisées avant de présenter 0 ou 2. Nous pourrions identifier une période d'apprentissage pour laquelle la distance maximum de

poids est grande et oscillante. Après une période d'oscillation  $T$ , cette distance ne change plus ou change dans un intervalle très petit. Ainsi, nous considérons que le réseau est stable et l'apprentissage ou la reconnaissance est terminé.

Afin d'observer le comportement d'auto-organisation lorsqu'un stimulus est présenté plusieurs fois au réseau, nous y présentons le chiffre 0 cinq fois. Pour la première fois, le réseau atteint un état d'équilibre après 660 itérations et puis 11 itérations pour la deuxième fois et ainsi de suite. La même expérience est répétée avec le chiffre 2 et il atteint l'état d'équilibre après 768 itérations pour la première fois, et puis 11 itérations. Cette observation démontre que le réseau a des capacités d'adaptation par l'auto-organisation de façon autonome.



a)



b)

**Figure 2.3** L'évolution dynamique du réseau après sa stimulation par: a) le chiffre 0 et b) le chiffre 2.

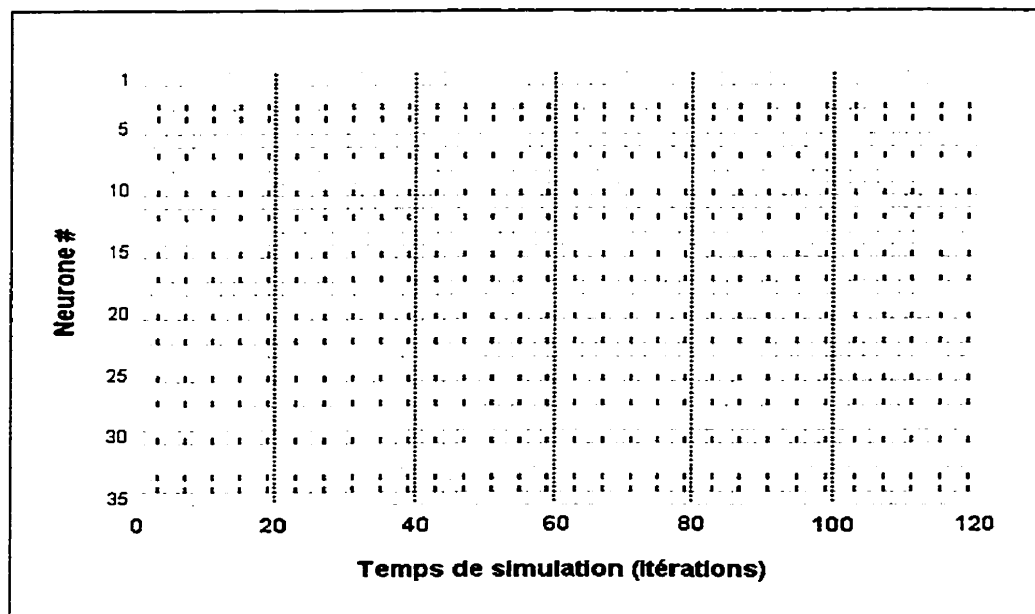


### 2.3.2b Activité de décharge

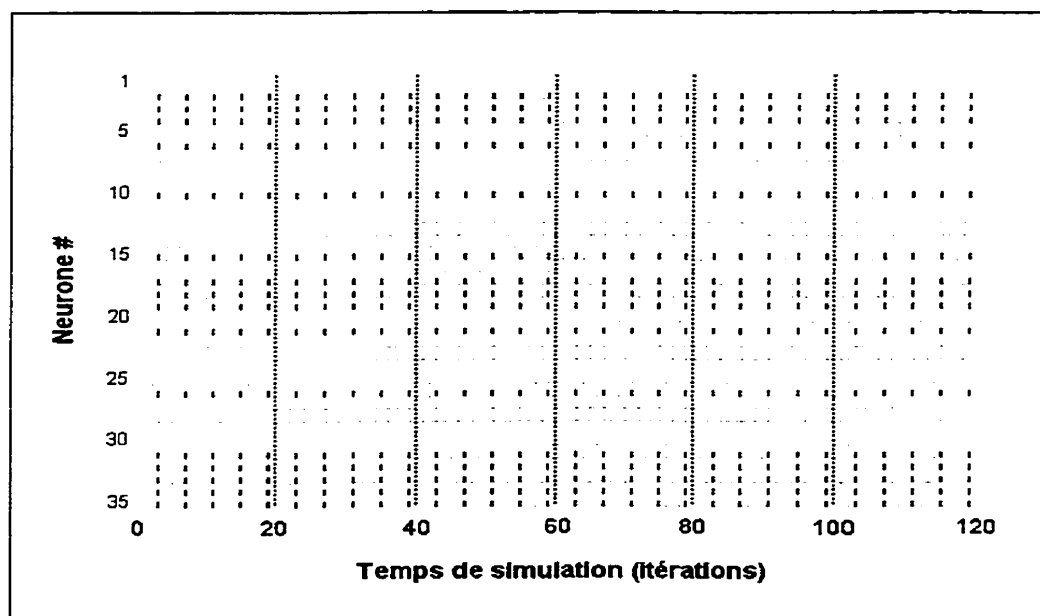
Pour les réseaux à décharge («spiking networks»), l'instant de décharge («firing time») de tous les neurones est souvent considéré comme un patron spatio-temporel ayant une structure temporelle explicite. Nous présentons ici quelques expériences concernant ce problème. Les sections *a* et *c* de la figure 2.4 donnent la valeur du seuil de décharge de tous les neurones lorsque le réseau a atteint l'état final après sa stimulation par les chiffres 0 et 2 respectivement. Les neurones sont ordonnés (1 à 35) de gauche à droite et de haut en bas. Notons que les mêmes conditions initiales ont été utilisées avant de présenter 0 ou 2. Nous observons que l'ensemble des neurones couverts par les stimuli d'entrée (le chiffre 0 ou 2) a une activité synchrone de décharge (Figures 2.5a, b). Ils ont la même phase et fréquence de décharge. Ces neurones ont également un même seuil de décharge  $\theta = -0.9$ . Les autres neurones demeurent silencieux avec un seuil élevé  $\theta = 0.9$ .

Ces observations suggèrent que les neurones qui se déchargent de façon synchrone avec un seuil  $\theta$  plus bas appartiennent à un même ensemble. Cet ensemble de neurones pourrait être employé pour caractériser le stimulus d'entrée. Autrement dit, les neurones qui appartiennent à un stimulus d'entrée (c'est-à-dire, un objet) se déchargent collectivement et de façon synchrone. Dans un sens, on pourrait se référer à des chaînes de «synfire» formées par un mécanisme d'auto-organisation biologiquement plausible dans les réseaux de neurones (Abeles, 1982). Nous y constatons également que le réseau a des patrons de décharge finaux différents lorsqu'il est stimulé par différents stimuli. Selon le stimulus, le mécanisme d'auto-adaptation produira différents ensembles de neurones synchrones. Une autre interprétation plausible est que



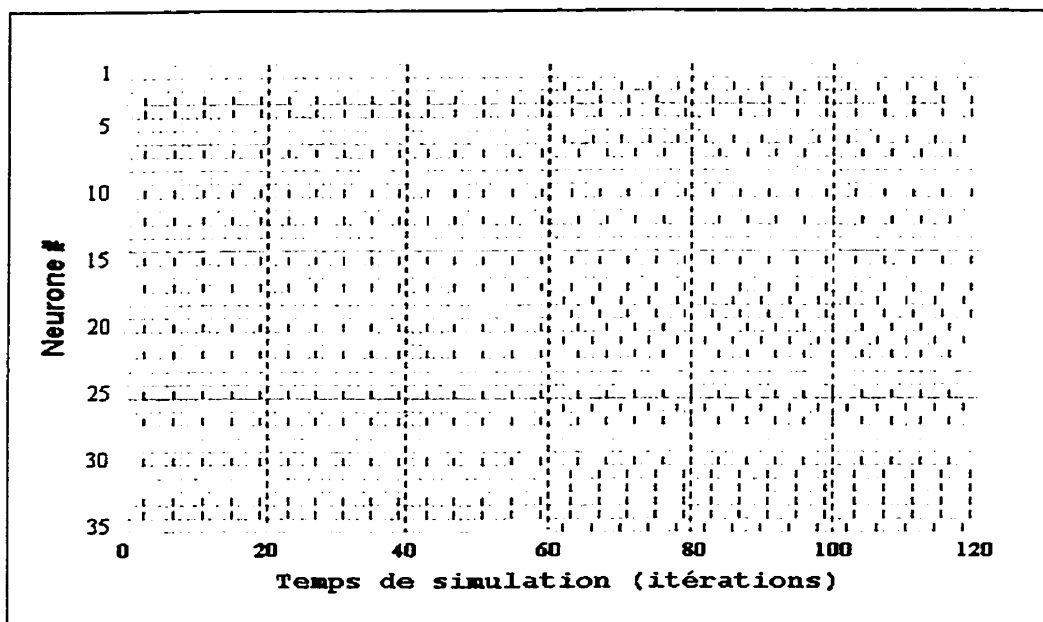


a)

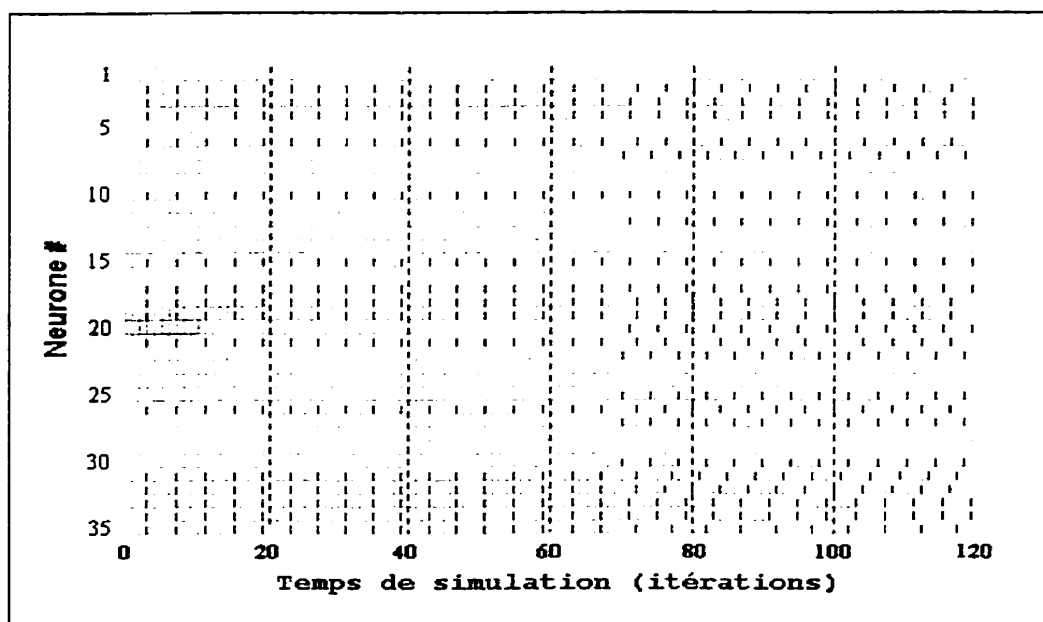


b)

**Figure 2.5** L'activité de décharge du réseau après sa stimulation par: a) le chiffre 0 et b) le chiffre 2. Les conditions initiales sont les mêmes avant la présentation de chaque chiffre.



a)



b)

**Figure 2.6** L'activité de décharge pendant la phase de transition lorsque le réseau a été stimulé par: a) les chiffres 0 et 2 *séquentiellement* et b) les chiffres 2 et 0 *séquentiellement*. Il n'y a aucune réinitialisation des conditions initiales entre la présentation de chaque chiffre.

### 2.3.3 Auto-organisation avec une séquence de stimuli d'entrée

Dans l'expérience précédente, seulement un stimulus est employé pour stimuler le réseau et les poids de connexion sont initialisés aléatoirement. Cependant, nous devons comprendre le comportement du réseau lorsqu'un ensemble des différents stimuli y est séquentiellement présenté comme entrée. C'est-à-dire lorsqu'il n'y a aucune initialisation des poids de connexion entre les présentations des stimuli d'entrée.

#### *2.3.3a Expérience 1 avec une séquence de deux stimuli*

D'abord, le réseau est stimulé par le chiffre 0. Il oscille et atteint un état stable après 660 itérations. Ensuite, le chiffre 2 y est présenté, le réseau est excité, oscille et atteint à nouveau un autre état stable après 748 itérations. Le seuil de tous les neurones est remis à l'état initial avant la présentation du chiffre 2. La section *b* de la figure 2.4 donne la valeur du seuil de tous les neurones à l'état final. L'activité de décharge pendant la phase de transition de 0 à 2 (pour deux intervalles de 60 itérations, pris avant et après la présentation du chiffre 2) est montrée dans la section *a* de la figure 2.6. Nous observons que de nouveaux neurones sont activés tandis que les neurones déjà actifs le demeurent.

Les neurones qui ne sont pas couverts par les chiffres 0 et 2 demeurent silencieux avec  $\theta = 0.9000$ . Par contre, tous les neurones qui ne sont que couverts par le chiffre 0 se déchargent encore de façon synchrone mais leur

phase de décharge est retardée d'une itération (groupe 1: neurones 7, 12, 20, 22, 25, 27, 30) avec  $\theta = -0.8910$ . Les neurones qui sont couverts à la fois par le chiffre 0 et 2 se déchargent eux aussi de façon synchrone avec la même phase qu'avant la présentation du chiffre 2 (groupe 2: neurones 3, 4, 10, 15, 17, 33, 34) avec  $\theta = -0.9000$ . Il est possible que le stimulus déjà appris (chiffre 0) soit encore enregistré dans le réseau et qu'il ne soit pas détruit par la présentation d'un nouveau stimulus.

Les neurones qui ne sont que couverts par le chiffre 2 peuvent être divisés en deux groupes (groupes 3 et 4) en se basant sur leur phase de décharge. Le groupe 3 (neurones 19, 31, 32), avec  $\theta = -0.8708$ , a la même phase de décharge que le groupe 2 tandis que le groupe 4 (neurones 2, 6, 18, 21, 26, 35), avec  $\theta = -0.8813$ , a une phase de décharge différente de tous les autres groupes.

Cette expérience indique que le réseau a un comportement complexe et qu'il pourrait préserver les stimuli déjà appris lors d'apprentissage de nouveaux stimuli. Apparemment, la synchronisation de l'activité de décharge en groupes pourrait être un résultat du changement continu du seuil  $\theta$  au cours de la simulation.

Le réseau a une activité dynamique dépendante du temps. En effet, le comportement complexe de ce réseau résulte de changements temporels de ses états. Ceci signifie que le réseau ne traite pas les stimuli de manière indépendante au cours du temps. Ainsi, l'action produite par un stimulus sera différente s'il est ou non précédé par un autre stimulus. Donc, le réseau peut avoir une capacité à traiter des processus dépendants du temps.

Afin de vérifier cette hypothèse, nous utilisons la séquence des chiffres 2 et 0 pour stimuler le réseau avec les mêmes conditions initiales. Contrairement à l'expérience précédente, le chiffre 2 est d'abord présenté et ensuite le chiffre 0. Les temps de stabilisation correspondant aux chiffres 2 (768 itérations) et 0 (572 itérations) sont différents de ceux dans l'expérience précédente (660 itérations pour le chiffre 0 et 748 itérations pour le chiffre 2). Le seuil du réseau et l'activité de décharge sont aussi différents (Figures 2.4c, d et Figure 2.6b). Nous observons que le réseau a un changement significatif dans son comportement. Bien que le réseau soit sensible à l'ordre temporel des stimuli, il préserve encore les caractéristiques internes semblables des deux stimuli (Figure 2.4b, d et Figure 2.6a, b).

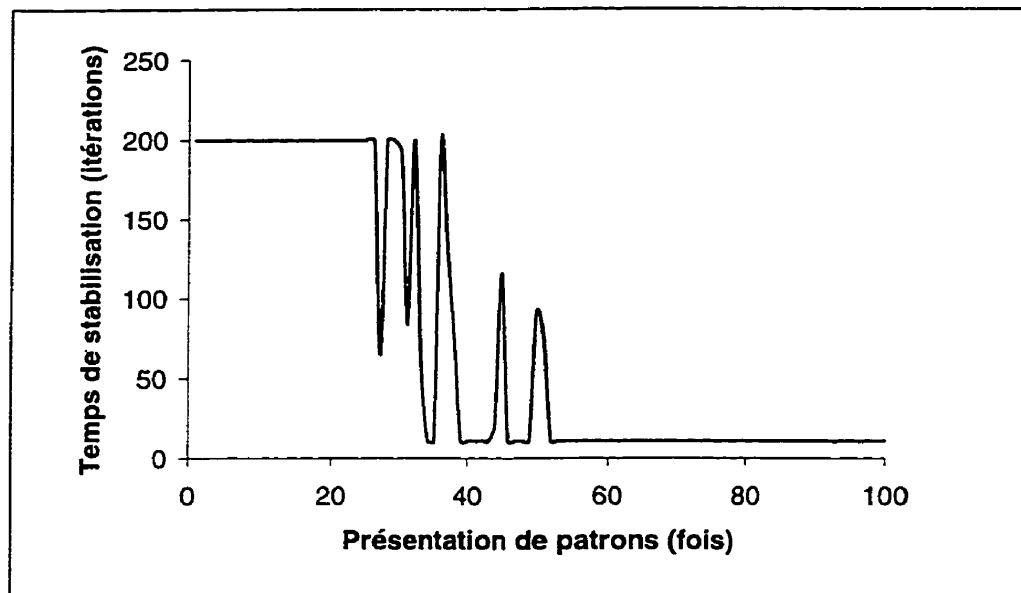
### ***2.3.3b Expérience 2 avec une séquence de cinq stimuli***

Nous utilisons une séquence de 5 chiffres (0-4) comme stimulus d'apprentissage. Chaque stimulus est présenté séquentiellement au réseau. Cependant, la façon de présenter les stimuli n'est pas la même que pour l'expérience précédente. Chaque stimulus est présenté pendant un temps donné  $Z$ . Et ensuite un nouveau stimulus est présenté, peu importe que le réseau atteigne un état stable ou non. En d'autres termes, après avoir été stimulé, le réseau oscille. Nous n'attendons pas qu'il atteigne un état d'équilibre afin de présenter un nouveau stimulus. Ce dernier est présenté tout de suite lorsque le temps  $Z$  est écoulé. Dans la situation où le réseau atteint l'état stable avant  $Z$ , un nouveau stimulus est présenté immédiatement.

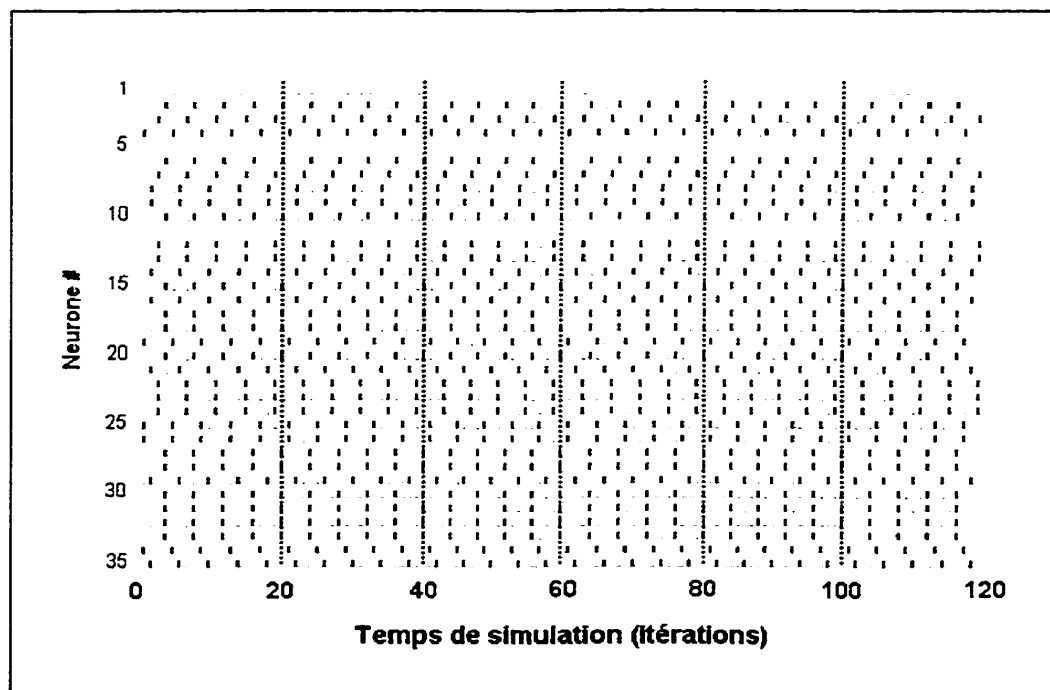
Nous choisissons le temps de présentation  $Z = 200$  itérations. Notons que la capacité de mémorisation du réseau est faible si le temps de présentation du stimulus est trop court ( $Z < 100$  itérations). La séquence de 5 chiffres (0-4) avec un ordre aléatoire est présentée au réseau plusieurs fois. La figure 2.7 montre le temps d'oscillation du réseau lorsque cette séquence de 5 chiffres est présentée 20 fois (c'est-à-dire, 100 présentations des chiffres au total). Pendant les 25 premières présentations des chiffres, le réseau ne peut pas atteindre un état stable. Il est forcé de s'arrêter après chaque intervalle de 200 itérations et un nouveau stimulus est présenté. Après cette période, il commence à entrer dans une phase «d'apprentissage» où il peut atteindre l'état stable avant  $Z$ . Après 60 présentations, il reste dans une phase où il atteint l'état stable dans un temps très court (11 itérations). Dans cette phase, il reconnaît les stimuli d'entrée. Les deux phases sont séparées par une transition très courte. Cette observation montre que le réseau a des capacités d'adaptation pour plusieurs stimuli.

Comme dans l'expérience précédente, nous y observons une synchronisation en groupe de l'activité de décharge (Figure 2.8). Selon la phase de décharge nous pouvons diviser les neurones en quatre groupes: groupe 1 (neurones 2, 6, 10, 15, 18, 20, 27, 28, 30, 31, 32, 33), groupe 2 (neurones 3, 7, 12, 13, 22, 23, 24), groupe 3 (neurones 4, 19, 25, 26, 34), et groupe 4 (neurones 8, 9, 14, 16, 21, 29, 35). Cependant, nous ne sommes pas en mesure d'expliquer théoriquement et expérimentalement comment ces groupes sont formés. Il faut noter que ce phénomène a été observé dans les systèmes nerveux (Eckhorn, 1990). À présent, plusieurs chercheurs suggèrent que la synchronisation de l'activité de décharge soit peut-être un moyen de coder l'information (Barinaga, 1998). Il y a certainement intérêt à pousser l'analyse sur ce phénomène pour le réseau proposé.





**Figure 2.7** Temps de stabilisation du réseau lorsqu'une séquence de cinq stimuli d'entrée (chiffres 0-4) y est présentée 20 fois.



**Figure 2.8** L'activité de décharge du réseau après sa stimulation par une séquence de cinq chiffres (0-4) 20 fois.

## **2.4 Application à la reconnaissance des chiffres bruités**

### **2.4.1 Critère de reconnaissance**

Pour le réseau proposé, il n'y a pas de différence entre apprentissage et reconnaissance (ou détection de la nouveauté). La mise à jour des connexions synaptiques est toujours en cours sauf lorsque le réseau est stable. Après présentation d'un stimulus, le réseau a un comportement complexe qui finit par se stabiliser au cours du temps. Lorsque les changements de poids sont inférieurs à un seuil préétabli, on considère que le réseau est stable et que l'apprentissage ou la reconnaissance est terminé. Le temps nécessaire  $T$  pour atteindre cet état stable est défini comme le temps de stabilisation du réseau. Nous proposons donc l'utilisation du temps de stabilisation comme critère de détection de nouveauté. Ce temps  $T$  permet de caractériser le stimulus d'entrée. Un temps  $T$  très court (de l'ordre de 11 itérations) implique que le réseau a probablement déjà «vu» ce stimulus au préalable. Par contre, un temps  $T$  relativement long implique que le réseau n'a jamais vu ce stimulus. Cette notion permet de créer un système de détection de la nouveauté.

### **2.4.2 Détection de la nouveauté**

Des expériences de détection de la nouveauté ont été réalisées à partir des chiffres de 0 à 9 codés sur une matrice binaire de 7 par 5. Le réseau

apprend uniquement sur les données non bruitées (Figure 2.2, colonnes 1 et 3). Les expériences de la détection de nouveauté portent sur les versions bruitées et propres de chiffres de 0 à 9. Les versions bruitées (Figure 2.2, colonnes 2, 3, 5, et 6) sont créées par l'introduction du bruit aux versions propres. Autrement dit, étant donné une quantité de bruits (en pourcentage par rapport au nombre de pixels de l'image), un nombre de pixels dans une image de chiffre est modifié. Les pixels affectés sont aléatoirement choisis selon une distribution uniforme. Dans ces expériences, le taux de bruit est de 20%, c'est-à-dire que 7 pixels dans chaque image des chiffres (7x5) ont leurs valeurs modifiées. L'apprentissage s'effectue sur des séquences de chiffres sans aucune initialisation des paramètres du réseau. En effet, les systèmes biologiques n'initialisent probablement pas leurs paramètres entre présentations de stimuli. Dans ce qui suit nous présentons en détail ces expériences ainsi qu'une comparaison de performance avec des réseaux classiques tels que réseau de Hopfield, réseau à rétropropagation et réseau DYSTAL.

#### **2.4.2a Expérience 1: Apprentissage des chiffres de 0 à 4**

Les versions propres des chiffres (0-4) sont employées pour entraîner le réseau. Chaque chiffre  $y$  est présenté séquentiellement. Ci-après, chaque chiffre est présenté pour une durée de 200 itérations. Nous n'attendons pas que le réseau atteigne un état stable pour la présentation d'un nouveau chiffre. Un nouveau chiffre est présenté immédiatement lorsque le réseau atteint un état stable avant 200 itérations. La séquence des chiffres de 0 à 4 est présentée au réseau 20 fois avec des ordres aléatoirement différents.

Le test de la détection de nouveauté porte sur les versions bruitées et propres des chiffres de 0 à 9. Tous les chiffres de la figure 2.2 sont présentés au réseau. Selon le critère de reconnaissance basé sur le temps de stabilisation du réseau (*pendant la phase de test*), nous pouvons décider si un chiffre a déjà été «vu» (par le réseau) pendant l'apprentissage. Une réponse «Non» signifie que ce chiffre n'a jamais été vu par le réseau (c'est la détection de la nouveauté). Lors de l'interprétation des résultats, nous utilisons le terme «reconnaissance» pour les chiffres (propres et bruités) qui ont déjà été appris (par exemple, les chiffres 0-4 dans cette expérience); et le terme «détection de nouveauté» pour les chiffres qui n'ont pas été appris (les chiffres 5-9). Le terme «reconnaissance» est utilisé pour indiquer seulement que le réseau a déjà «vu» un chiffre. Au contraire, le terme «détection de nouveauté» est utilisé pour indiquer que le réseau n'a jamais «vu» un chiffre.

Nous considérons que le temps de stabilisation est court lorsqu'il est inférieur à 100 itérations, sinon il est long. Le tableau 2.1 montre que le réseau a un temps de stabilisation court lorsque les chiffres de test sont ceux déjà appris. Le temps de stabilisation est également court vis-à-vis des versions bruitées de ces chiffres. Par contre, le réseau a un temps de stabilisation significativement long lorsque les chiffres de test ne sont jamais vus pendant l'apprentissage (les chiffres de 5 à 9 et leurs versions bruitées). Le réseau a fait des erreurs de reconnaissance sur 2 chiffres (version bruitée 1 du chiffre 0 et version bruitée 2 du chiffre 3). Il y a également des erreurs de détection de nouveauté sur 3 chiffres (version bruitée 1 du chiffre 9, versions bruitées 2 des chiffres 7 et 9). Le pourcentage moyen d'erreur est de 16.7% (5/30).

**Tableau 2.1** Temps de stabilisation (en itérations) du réseau entraîné par les chiffres propres (0-4) et testé sur (0-9) propres et bruités.

Chiffres	0	1	2	3	4	5	6	7	8	9
Version propre	11	11	11	11	11	595	371	595	371	371
Version bruitée 1	199	11	11	11	11	283	583	667	371	11
Version bruitée 2	11	11	11	283	11	667	583	11	667	11

#### 2.4.2b Expérience 2 : Apprentissage des chiffres 5 à 9

Un des points faibles de certains réseaux de neurones est que leur performance de reconnaissance change significativement lorsqu'ils sont entraînés par différents ensembles de stimuli d'entrée. Pour examiner ce problème, une autre expérience dont l'apprentissage est basé sur les chiffres (5-9) est effectuée. Le test de la reconnaissance et de la détection de nouveauté porte également sur les chiffres de 0 à 9. Les résultats sont donnés dans le tableau 2.2. Le réseau a un taux d'erreur plus grand que pour l'expérience précédente: 2 erreurs pour la reconnaissance et 7 pour la détection de nouveauté. Le pourcentage moyen d'erreur est de 30% (9/30).

**Tableau 2.2** Temps de stabilisation (en itérations) du réseau entraîné par les chiffres propres (5-9) et testé sur (0-9) propres et bruités.

Chiffres	0	1	2	3	4	5	6	7	8	9
Version propre	11	3997	45	11	2985	11	11	11	11	11
Version bruitée 1	20	1678	1799	11	1708	346	11	11	11	23
Version bruitée 2	11	2593	149	11	4888	560	11	11	11	11

### 2.4.3 Comparaison avec d'autres réseaux

Les expériences portant sur la reconnaissance des chiffres bruités permettent de montrer que le temps de stabilisation (du réseau) peut effectivement être utilisé comme critère pour la reconnaissance des formes. Par contre, l'étude de la comparaison avec d'autres réseaux tels que le réseau multicouche, le réseau de Hopfield et le réseau DYSTAL aide à évaluer la performance du réseau dans une tâche de reconnaissance (la détection de nouveauté dans ce travail, par exemple). Il est à noter que cette étude n'a pas pour but de comparer la capacité de ces réseaux (y compris le réseau proposé) en reconnaissance des formes en général.

Les réseaux de type Hopfield sont intéressants du point de vue théorique et peuvent être utilisés pour la classification. Nous choisissons un réseau de type Hopfield mis en application dans le logiciel MATLAB (Demuth et Beale, 1996). Dans le domaine des réseaux de neurones, les réseaux multicouches à rétropropagation sont les plus largement utilisés comme systèmes de reconnaissance des formes. Un réseau multicouche supervisé par la règle généralisée standard de delta avec moment (Eberhart et Dobbins, 1992) est choisi pour nos expériences. De plus, nous choisissons un réseau de type DYSTAL («Dynamically Stable Associative Learning») qui est inspiré des caractéristiques de l'apprentissage et de la mémorisation identifiées sur les crassicomis *Hermisenda* et l'hippocampe de lapin (Alkon, 1990). Le réseau apprend associativement des corrélations et des anticorrélations entre les événements temporels qui se produisent dans les neurones présynaptiques. Ce type de réseau a été utilisé dans la reconnaissance de caractères manuscrits (Blackwell, 1992).

Nous utilisons ces réseaux pour réaliser une étude de comparaison avec le réseau proposé. Dans le premier temps, nous essayons de trouver les architectures (ainsi que les paramètres) de ces réseaux afin d'assurer qu'ils pourraient donner la meilleure performance vis-à-vis de la même tâche à la section précédente. Notons qu'il est impossible de trouver des architectures optimales. En essayant plusieurs architectures pour chacun de ces réseaux, nous avons le choix suivant:

- ◆ Le réseau de Hopfield a une dimension 7x7 avec 49 neurones.
- ◆ Le réseau multicouche avec une architecture de 3 couches: une couche d'entrée ayant 35 neurones, une couche cachée ayant 5 neurones et une couche de sortie ayant 5 neurones.
- ◆ Le réseau DYSTAL a une couche d'entrée ayant 35 neurones et une couche de sortie ayant 5 neurones.

On décrit souvent le réseau de Hopfield comme permettant de réaliser des mémoires auto-associatives: un chiffre mémorisé (non bruité) est correctement restitué par une stabilisation du réseau si ce dernier a été stimulé par une version bruitée de ce chiffre. On considère dans ce cas que le réseau a reconnu le chiffre d'entrée. Dans le cas contraire, lorsque le chiffre restitué est différent de ceux mémorisés (en se basant sur une mesure de similarité telle que celle de Hamming par exemple), on considère qu'il n'a pas reconnu le chiffre d'entrée. La détection de nouveauté, par ce réseau, est réalisée dans le sens où le réseau restitue un chiffre qui est totalement différent de ceux mémorisés lorsqu'il est stimulé par un chiffre non appris au préalable.

D'autre part, la détection de nouveauté par le réseau multicouche est réalisée de la façon suivante: on apprend au réseau les chiffres (0-4). Chaque chiffre est associé à une sortie désirée; par exemple (1 0 0 0 0) correspond au

chiffre 0, (0 1 0 0 0) au chiffre 1 et ainsi de suite. Pendant la phase de test, si la réponse du réseau est parmi ces sorties désirées, on considère alors qu'il a reconnu les chiffres déjà appris. Dans le cas où on lui présente des chiffres non appris au préalable et où sa réponse n'est pas parmi ces sorties désirées, on considère qu'il n'a jamais «vu» ces chiffres, c'est-à-dire qu'il a effectué la détection de nouveauté.

Du côté du réseau DYSTAL, la détection de nouveauté est basée sur le critère du maximum de corrélation (notons que la sortie de ce réseau donne le coefficient de corrélation de Pearson). Selon un seuil de corrélation choisi, on peut considérer si un chiffre a été déjà «vu» par le réseau.

Dans ce qui suit, nous utilisons une base de données beaucoup plus grande que celle employée dans la section précédente (celle qui n'a que 30 patrons comme dans la figure 2.2). Nous produisons pour chacun des chiffres (0-9) un ensemble des versions bruitées avec un taux de bruit de 20%. Chaque ensemble comprend 50 versions bruitées. Cela signifie qu'il y a 500 patrons bruités au total pour le test.

Tous les réseaux, y compris celui proposé, effectuent la même tâche telle que dans les 2 expériences précédentes. Notons que l'apprentissage porte toujours sur les versions propres des chiffres (0 à 4) et (5 à 9) respectivement. Par contre, le test s'effectue sur l'ensemble des 500 patrons bruités. Les résultats de la reconnaissance et de la détection de nouveauté sont donnés en terme de pourcentage de l'erreur dans les tableaux 2.3 et 2.4.



**Tableau 2.3** Pourcentage de l'erreur en reconnaissance et en détection de nouveauté pour l'expérience 1: apprentissage sur les versions propres (0-4) et test sur les versions propres et bruitées (0-9).

Chiffres	0	1	2	3	4	5	6	7	8	9
Réseau proposé	52	34	26	26	40	0	28	14	28	20
Réseau Hopfield	48	50	30	60	54	20	4	2	14	8
Réseau multicouche	4	2	2	44	0	70	42	30	62	66
Réseau DYSTAL	0	2	2	2	0	20	12	0	54	22

**Tableau 2.4** Pourcentage de l'erreur en reconnaissance et en détection de nouveauté pour l'expérience 2: apprentissage sur les versions propres (5-9) et test sur les versions propres et bruitées (0-9).

Chiffres	0	1	2	3	4	5	6	7	8	9
Réseau proposé	26	16	46	70	4	40	36	36	30	30
Réseau Hopfield	10	0	30	16	0	58	60	52	48	48
Réseau multicouche	50	36	58	54	42	24	6	8	34	6
Réseau DYSTAL	2	0	28	38	0	2	6	0	6	2

Un résultat synthétisé à partir des tableaux 2.3 et 2.4 est donné dans le tableau 2.5. En terme du pourcentage global de l'erreur, le réseau proposé a une performance comparable à celle du réseau de Hopfield et du réseau multicouche. Cependant, sa performance est encore loin de celle du réseau DYSTAL. Il est important de noter que le réseau DYSTAL est tout à fait adapté aux systèmes basés sur la corrélation des stimuli. Il y aura lieu à une recherche plus poussée pour améliorer la performance du réseau proposé.

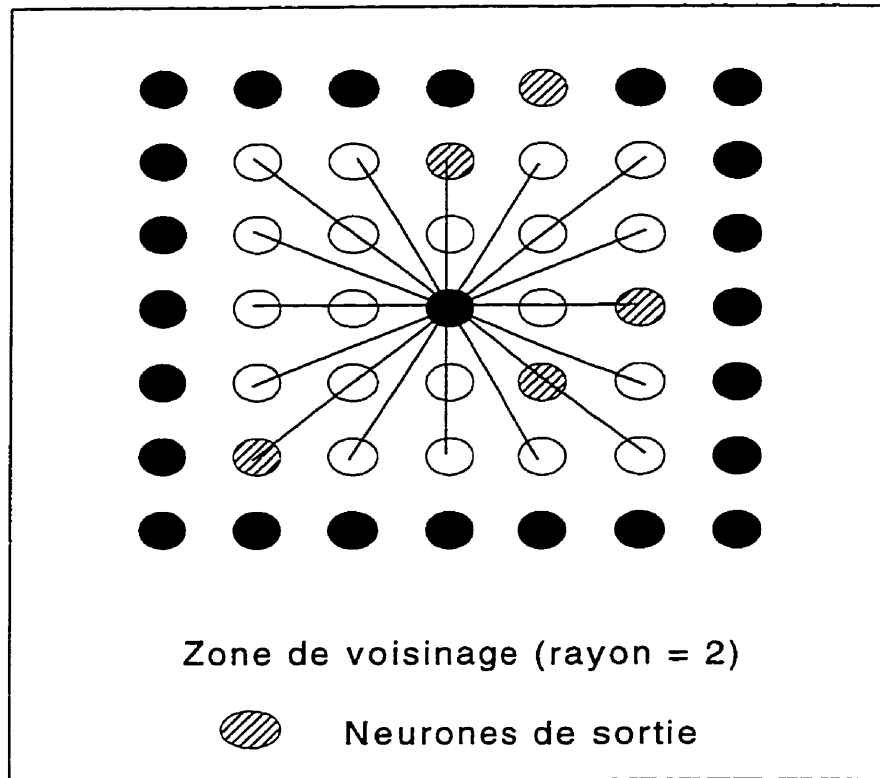
**Tableau 2.5** Comparaison de performance en terme de pourcentage global de l'erreur en reconnaissance et en détection de nouveauté.

Réseau	Apprentissage (0-4)		Apprentissage (5-9)		Pourcentage de l'erreur global
	Test (0-9)		Test (0-9)		
	Reconnaissance	Détection de nouveauté	Reconnaissance	Détection de nouveauté	
Réseau proposé	35.6	18.0	34.4	32.4	30.1
Réseau Hopfield	48.4	9.6	53.2	11.2	30.8
Réseau multicouche	10.4	54.0	15.6	48.0	32.0
Réseau DYSTAL	1.2	21.6	3.2	13.6	9.9

## 2.5 Réseau avec mécanisme de «feedback» avec récompense

Tous les études précédentes portent sur le modèle de réseau avec un contrôleur global. Dans cette section, nous proposons un autre modèle de réseau qui utilise un mécanisme de «feedback» avec récompense pour l'auto-organisation.

La différence principale entre ce réseau et celui présenté précédemment est qu'il n'existe pas de contrôleur global. Autrement dit, il n'y a que les connexions locales entre les neurones du réseau (Figure 2.9). Ceci se justifie plus vis-à-vis de la physiologie. De plus, nous y utilisons un mécanisme d'apprentissage particulier inspiré du travail de Stassinopoulos et Bak (1995) pour la mise à jour des poids de connexion. D'une manière analogue aux techniques comportementales utilisées dans l'entraînement des animaux, les stimuli d'entrée sont présentés au réseau et chacun d'entre eux récompense une action spécifique. Le réseau apprend à identifier tous les stimuli et choisit l'action de récompense correspondante. Il mémorise les réponses correctes. Autrement dit, selon sa réponse (succès ou sans succès) il reçoit une récompense spécifique. La réponse est représentée par l'état de décharge de l'ensemble des neurones de sortie. Ces derniers sont choisis de façon aléatoire à partir des neurones de type excitateurs (Figure 2.9). Notons que dans ce réseau un neurone peut être à la fois un neurone de sortie et d'entrée.



**Figure 2.9** Architecture avec un mécanisme de «feedback» avec récompense.

Pour chaque stimulus d'entrée, l'action du réseau est considérée comme étant un succès si au moins un neurone appartenant à l'ensemble des neurones de sortie se décharge. Dans ce cas, les connexions entre les neurones actifs sont renforcées. Si l'action est sans succès, les connexions entre les neurones actifs sont affaiblies. La mise à jour des connexions se fait selon la règle d'Hebb. Soit  $l$  le signal de feedback de récompense ( $l = +1$  pour le renforcement ou  $-1$  pour affaiblir), la mise à jour des poids entre deux neurones  $i$  et  $j$  est donnée par:

$$C_{ij}(t+1) = C_{ij}(t) + \gamma C_{ij}(t)(1 - C_{ij}(t))S_i(t)S_j(t)l(t) \quad (2.11)$$

avec  $\gamma$  le taux d'apprentissage.

Afin d'étudier le comportement et des caractéristiques dynamiques de ce modèle de réseau, nous avons effectué toutes les expériences présentées aux sections précédentes.

Le réseau est composé de 70% de neurones excitateurs et de 30% de neurones inhibiteurs. Il comprend 35 neurones (7 par 5). Le rayon de voisinage est de 2 pour les neurones excitateurs et de 1 pour les neurones inhibiteurs. Cinq neurones ont été choisis aléatoirement comme neurones de sortie (neurones 2, 7, 19, 23 et 25). Notons que les neurones sont ordonnés (1 à 35) de gauche à droite et de haut en bas sur le plan de 7x5. Les paramètres du réseau sont initialisés tels que:  $a(0) = 0.1$ ,  $b(0) = 0.1$ ,  $U_{int} = -0.2$ ,  $\kappa = 3.0$ ,  $\rho = 3.0$ ,  $\alpha = 0.2$ ,  $\beta = 0.2$ ,  $\lambda = 0.1$ . Les poids de connexions sont initialisés aléatoirement dans un intervalle de 0 à 1.

Afin d'alléger cette section, nous résumons ici les résultats obtenus. Les expériences montrent que ce réseau possède des caractéristiques similaires à celles du réseau avec un contrôleur global. Ainsi, en étudiant l'activité de stabilisation, nous constatons que ce réseau a des capacités d'adaptation par l'auto-organisation de façon autonome. De plus, les expériences avec des séquences temporelles des chiffres 0 et 2 montrent que le réseau est sensible à l'ordre temporel des stimuli: notons par exemple, la différence de seuil entre le cas où le réseau est stimulé par la séquence des chiffres 0-2 et celui où il est stimulé par la séquence des chiffres 2-0 (Figure 2.10).

D'autre part, l'activité de groupement des neurones est observée en se basant sur la valeur du seuil de décharge des neurones (Figure 2.10) ainsi que sur la phase et la fréquence de décharge (Figure 2.11). Cependant, nous constatons une situation différente par rapport au réseau ayant un contrôleur

global. Les neurones qui ne sont que couverts par le premier stimulus (par exemple, neurones 3, 4, 7, 10, 12, 15, 17, 20, 22, 25, 27, 30, 33 et 34 pour le chiffre 0 dans l'expérience présentée à la figure 2.10a) s'éteignent lors de l'introduction du deuxième stimulus. Par contre, ces neurones demeurent actifs dans le cas du réseau avec un contrôleur global. Les neurones couverts à la fois par les deux stimuli (chiffres 0 et 2) se déchargent de façon synchrone avec la même phase telle qu'avant la présentation du deuxième stimulus. Les neurones qui ne sont que couverts par le deuxième stimulus se déchargent eux aussi de façon synchrone mais avec une phase différente de ces derniers.

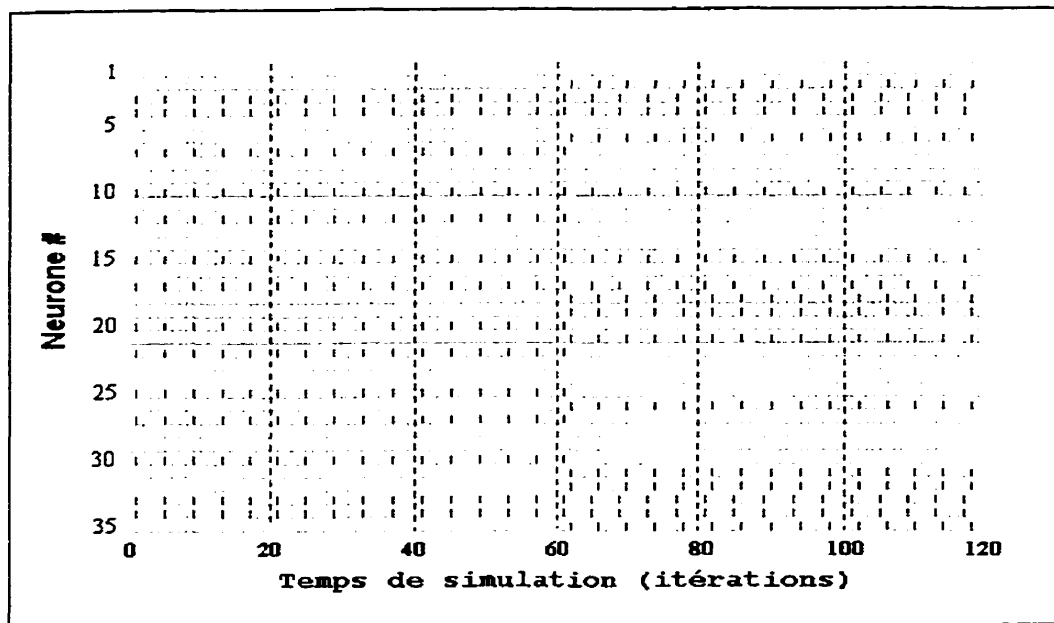
En utilisant le même critère de reconnaissance (basé sur le temps de stabilisation), nous effectuons les expériences portant sur la détection de nouveauté avec des chiffres. Ce réseau a un taux d'erreur global de 34.7% tandis que le réseau avec un contrôleur global a un taux d'erreur de 30.1%.

**Figure 2.10** Les seuils du réseau après sa stimulation par: a) le chiffre 0; b) les chiffres 0 et 2 séquentiellement (il n'y a aucune réinitialisation des connexions entre la présentation des chiffres); c) le chiffre 2; d) les chiffres 2 et 0 séquentiellement. Chaque valeur correspond au seuil de décharge de chaque neurone sur le plan du réseau.

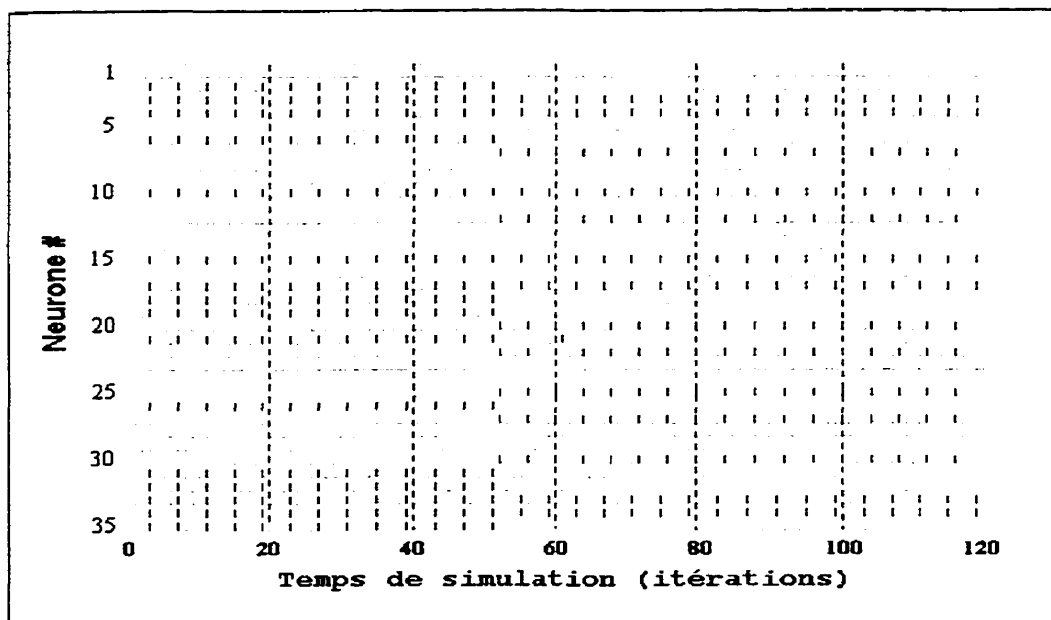
		a)				b)			
1	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000
11	-0.8998	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000
21	0.9000	-0.8996	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000
31	0.9000	0.9000	-0.8996	0.9000	0.9000	0.9000	-0.8996	0.9000	0.9000
5	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000
15	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000
15	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000
25	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000
25	-0.8998	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000
31	-0.8998	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000
35	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000

		c)				d)			
1	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000
11	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000
21	-0.8998	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000
31	-0.8998	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000
5	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000
15	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000
15	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000
25	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000
25	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000
31	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000
35	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000



a)



b)

**Figure 2.11** L'activité de décharge pendant la phase de transition lorsque le réseau a été stimulé par: a) les chiffres 0 et 2 *séquentiellement* et b) les chiffres 2 à 0 *séquentiellement*. Il n'y a aucune réinitialisation des conditions initiales entre la présentation de chaque chiffre.



Il faut rappeler que le réseau (avec un mécanisme de «feedback» avec récompense) étudié ici est à base de neurones avec un seuil de décharge variable. Dans un autre travail, nous essayons également d'étudier ce réseau à base de neurones utilisant la fréquence instantanée. Les expériences de reconnaissance de données statiques (les chiffres corrompus) indiquent que le réseau est en mesure de faire un travail comparable à un réseau de Hopfield et à un réseau à rétropropagation (annexe B, Ho, 1998). Les expériences portant sur la reconnaissance des séquences temporelle (les séquences de voyelles) permettent de mettre en valeur l'aptitude du réseau à traiter l'information temporelle (annexe C, Rouat et Ho, 1998). En effet, la réponse du réseau (temps de stabilisation) à une voyelle dépend de ses états précédents. Pour une même voyelle, les caractéristiques dynamiques sont liées aux présentations antérieures. Autrement dit, la réponse du réseau dépend de l'ordre temporel des séquences des stimuli d'entrée.

Les expériences portant sur ce réseau avec un mécanisme de «feedback» avec récompense sont ici relativement limitées et il y aura lieu d'approfondir le travail afin de bien connaître les limites du réseau et son potentiel en reconnaissance de patrons spatio-temporels. Cependant, les contraintes de temps relatives à ma thèse m'obligent à m'arrêter à ce stade et à laisser cette direction ouverte pour les recherches futures.

Dans les chapitres 3 et 4, nous n'utilisons que le modèle de réseau ayant un contrôleur global et à base de neurones avec un seuil de décharge variable.

## 2.6 Discussion et conclusion

Le modèle proposé de réseau de neurones est un résultat d'une tentative d'associer des principes des systèmes dynamiques non linéaires et des réseaux neuromimétiques dans le but du traitement des processus spatio-temporels non stationnaires. Les observations en neurophysiologie sont également une source d'inspiration pour la conception du modèle de réseau dont l'architecture est inspirée de la topologie de la couche IV du cortex. En ce qui concerne le modèle du neurone à décharge (ce qui est utilisé dans la modélisation du système thalamo-cortical), l'introduction d'un seuil de décharge variable qui dépend de sa propre activité de décharge a effectivement un grand effet sur le comportement du réseau proposé. Cette façon de faire permet d'intégrer les états du neurone au cours du temps, toujours au niveau local du neurone. À notre connaissance, aucun travail ne propose un modèle du neurone dont le seuil varie en fonction de sa propre activité de décharge. Certains proposent un seuil variable (du neurone) qui dépend plutôt de l'activité globale du réseau (Stassinopoulos et Bak, 1995). Nous ne sommes pas en mesure de justifier ce problème. Cependant, nous pensons qu'une activité temporelle au niveau local (du neurone) peut créer un comportement d'auto-organisation spatio-temporelle au niveau global (du réseau). Par conséquent, ce comportement pourrait représenter la structure spatio-temporelle du stimulus d'entrée.

Comme nous l'avons analysé expérimentalement, le réseau possède des caractéristiques qui peuvent être exploitées en vue du traitement des formes dynamiques et organisées spatio-temporellement. L'activité de décharge dépend fortement du stimulus d'entrée et caractérise vraisemblablement une

représentation interne de ce dernier. Il y a plusieurs groupes de neurones qui se déchargent de façon synchrone et périodiquement. Chaque groupe représente probablement une caractéristique du stimulus d'entrée. Notons qu'un phénomène semblable est observé dans les systèmes nerveux et est considéré comme un moyen pour coder de l'information. De plus, le réseau est sensible à la structure temporelle de la séquence des stimuli d'entrée. Ces caractéristiques sont très intéressantes parce qu'elles se produisent à partir de la dynamique interne du réseau.

Au sens des systèmes dynamiques non linéaires, nous montrons par simulation que le réseau proposé pourrait être considéré comme un système à attracteurs multiples. Chaque attracteur peut effectivement être associé à un stimulus d'entrée. Au lieu de caractériser le comportement complexe du réseau par des attracteurs, nous proposons l'utilisation du temps de stabilisation du réseau. Par conséquent, le traitement basé sur le temps de stabilisation devrait être moins compliqué que celui basé sur la caractérisation des attracteurs. Les expériences sur la détection de nouveauté à partir d'une base de données des chiffres bruitées montrent que le temps de stabilisation peut effectivement représenter le stimulus d'entrée.

Cette approche ne nécessite pas de supervision du réseau. Celui-ci est en mesure de détecter la nouveauté et de s'adapter de façon autonome. Vis-à-vis des réseaux classiques, l'apprentissage et la reconnaissance (du réseau) sont deux aspects d'un même processus dynamique. Le réseau effectue «aveuglément» la même tâche en tout temps. En effet, les systèmes nerveux semblent fonctionner de cette manière (Il n'y a aucune séparation entre apprentissage et reconnaissance).

Il est à noter que le modèle de réseau proposé possède de nombreux degrés de liberté (en termes de paramètres). Le choix de ces paramètres a certainement une influence importante sur le comportement du réseau. Une étude approfondie de l'effet de variation des nombreux paramètres est donc essentielle et souhaitable afin de mieux comprendre le comportement du réseau dans des conditions variables. Étant donné que les paramètres (du réseau) sont assez nombreux, l'étude de l'effet de leurs variations est sans doute une tâche complexe. Théoriquement, nous n'avons pas encore trouvé un moyen mathématique permettant de réaliser cette étude. Expérimentalement, nous avons réalisé un nombre limité d'expériences et une analyse systématique de la façon suivante:

- ◆ nous modifions chacun des paramètres en leur assignant une série de valeurs appartenant à une plage des valeurs.
- ◆ pour chacune de ces valeurs, nous observons le comportement du réseau en termes de stabilisation, de synchronisation de l'activité de décharge, de mémorisation des stimuli appris, etc. À partir de ces observations, nous pouvons identifier l'effet de la variation de chaque paramètre sur chaque caractéristique du réseau tel que la stabilisation, la synchronisation de décharge et la mémorisation de stimuli du réseau.

Cependant, la contrainte de temps dans le cadre de la thèse ne nous a pas permis de mener une série d'expériences adéquate. À ce stade, nous ne sommes pas en mesure de donner des commentaires concernant l'effet des variations des paramètres du réseau. Cette étude constitue un objectif pour un travail futur.

## Chapitre 3

# Traitement des séquences temporelles et détection de mouvements

**L**e but de ce chapitre est d'évaluer les capacités du modèle de réseau proposé à effectuer le traitement de l'information spatio-temporelle: c'est-à-dire, la classification / détection des formes dynamiques organisées spatio-temporellement. Ce problème a des applications en reconnaissance de la parole et du mouvement. Nous présentons dans ce chapitre l'application de ce modèle de réseau pour les tâches suivantes:

- ◆ le traitement des séquences temporelles,
- ◆ la détection de mouvement.

### 3.1 Traitement des séquences temporelles

Le traitement de l'information temporelle est particulièrement difficile parce que l'information nécessaire est incluse dans le temps (donc il y a une dynamique inhérente) et n'est pas simultanément disponible. Néanmoins, ce sujet a été étudié par nombre d'investigateurs dans le domaine des réseaux de neurones (Wang, 1995b; Haykin, 1994).

La classification statique des formes n'est pas une méthode suffisante lorsque la dynamique des formes constitue une caractéristique importante de l'information, c'est-à-dire, lorsque l'information passée est nécessaire pour l'interprétation de l'information actuelle. Notons par exemple en reconnaissance de la parole où l'information contextuelle, sur différentes échelles de temps, est cruciale pour la compréhension. Entre autre, il s'agit du problème de l'ordre temporel des stimuli d'entrée. Par exemple, la séquence *A-B-C* est différente de *C-B-A*. Le traitement de séquences temporelles consiste en l'apprentissage, la reconnaissance / identification, la production, et la représentation des séquences. Nous nous intéresserons à la représentation et à l'identification de séquences temporelles.

Dans cette partie, nous proposons un mécanisme par lequel les séquences temporelles de stimuli d'entrée peuvent être représentées par le temps de stabilisation du réseau proposé. En fait, l'information sur l'ordre temporel des séquences de stimuli est transformée en séquences de temps de stabilisation du réseau.

### 3.1.1 Méthodologie

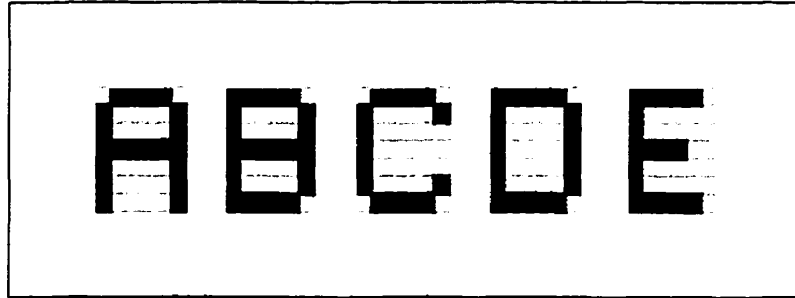
Dans le chapitre précédent, nous avons montré que le réseau proposé a des capacités d'auto-organisation lorsqu'il est stimulé par une séquence de stimuli d'entrée. Chaque stimulus est présenté de façon séquentielle au réseau. Le réseau oscille et se dirige vers un état stable chaque fois qu'il est activé par un stimulus. Il n'y a pas d'initialisation du réseau entre les présentations successives des stimuli. On procède ainsi pour l'ensemble des stimuli de la séquence d'entrée.

Le temps de stabilisation du réseau proposé est considéré comme sa réponse lorsqu'on lui présente une entrée. En utilisant ce mécanisme, nous pouvons obtenir une séquence de temps de stabilisation du réseau correspondant à la séquence des stimuli d'entrée. En d'autres termes, nous avons effectué une «projection» («mapping») de la séquence des stimuli d'entrée sur une séquence de temps de stabilisation tout en réduisant la dimension des paramètres de la séquence d'entrée.

En effet, la séquence de temps de stabilisation résulte de changement des états du réseau dans le temps. Chaque état du réseau dépend de ses états précédents. Ceci signifie que le réseau ne traite pas les stimuli de manière indépendante au cours du temps. Ainsi, l'action produite par un stimulus sera différente s'il est ou non précédé par un autre stimulus. Dans ce sens, les séquences de temps de stabilisation sont des transformations de l'information structurée temporellement des stimuli. Dans ce qui suit, nous présentons une série d'expériences portant sur les données des caractères et de la parole concernant cette transformation.

### 3.1.2 Expériences avec des séquences de caractères

Les caractères majuscules (de *A* à *Z* avec une dimension 7x6) sont utilisés dans ces expériences (Figure 3.1). L'architecture du réseau a une dimension 7x6 ayant 42 neurones dont chaque pixel de l'image d'entrée est présentée sur l'entrée  $e_i$  d'un neurone.



**Figure 3.1** Exemples des images des caractères *ABCDE*

Les expériences portent sur les séquences ayant une longueur différente afin de savoir comment celle-ci influence le résultat de la transformation.

#### 3.1.2a Expérience 1: transformation des séquences de cinq caractères

Étant donné un ensemble de 5 patrons distincts (*ABCDE*), il est possible de produire  $5! = 120$  autres séquences par permutation des 5 patrons. Nous présentons chaque séquence une fois au réseau. Pour chaque séquence, ses



caractères y sont successivement présentés. Ainsi, nous obtenons 120 séquences de temps de stabilisation correspondant à 120 séquences de caractères d'entrée. À titre d'exemple, nous présentons les séquences de temps de stabilisation correspondant à quelques séquences des caractères dans le tableau 3.1. On peut constater que la lettre «A» dans la séquence (*EACDB*) a un temps correspondant de stabilisation (804 itérations) différent de celui (588 itérations) de la lettre «A» dans la séquence (*CDEAB*). Il est clair que l'information de l'ordre temporel des séquences d'entrée a une influence importante dans le comportement du réseau en terme de temps de stabilisation.

Afin de savoir s'il y a plusieurs occurrences de la même séquence de temps de stabilisation parmi les 120 séquences, nous comparons chaque séquence avec les autres. Aucune occurrence des mêmes séquences n'a été observée. Ceci signifie que le réseau a 120 réponses différentes (c'est-à-dire 120 séquences distinctes de temps de stabilisation) pour 120 séquences d'entrée du même ensemble de 5 caractères. Autrement dit, le réseau semble avoir un degré élevé de sensibilité à l'ordre temporel des stimuli.

**Tableau 3.1** Exemple des séquences de temps de stabilisation

Séquence de caractères	Séquences de temps de stabilisation (en itérations)				
	Élément 1	Élément 2	Élément 3	Élément 4	Élément 5
EACDB	576	804	1597	1612	711
CDEAB	448	275	1097	588	500
CEDBA	448	933	903	651	199
ABCDE	576	1368	100	1824	663

Cependant, une question se pose: la transformation a-t-elle une influence importante sur la séparabilité des données. En d'autres termes, quel est le degré de séparabilité des données avant et après la transformation ? Pour répondre à cette question, nous avons choisi le coefficient de corrélation de Pearson (King et Julstrom, 1982) comme mesure de la séparabilité.

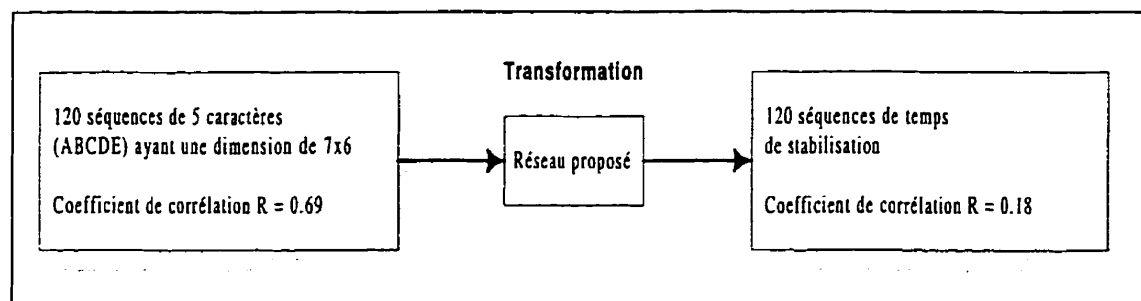
Étant donné  $n$  paires de données  $(X_i, Y_i)$ , le coefficient de corrélation de Pearson  $R$  est défini par:

$$R(X_i, Y_i) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\left( \sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2 \right)}} \quad (3.1)$$

Ce coefficient de corrélation a été utilisé comme mesure de corrélation dans le réseau DYSTAL (Alkon, 1990) puisqu'il évalue la structure des stimuli, plutôt que l'intensité absolue. Cette mesure a une valeur de -1.0 à +1.0: +1.0 indique une corrélation positive parfaite (similitude); -1.0 indique une corrélation négative parfaite et 0.0 indique une indépendance complète entre une paire de stimuli.

Nous calculons la corrélation de Pearson pour les deux ensembles de 120 séquences d'entrée et de 120 séquences correspondantes de temps de stabilisation. Chaque séquence est comparée à toutes les autres séquences du même ensemble de données afin de calculer le coefficient de corrélation. Par simplicité, seulement les valeurs moyennes globales de cette mesure sont présentées ici pour la comparaison (Figure 3.2). L'ensemble de temps de stabilisation a un coefficient de corrélation ( $R = 0.18$ ) inférieur à celui des

données initiales ( $R = 0.69$ ). Ceci signifie que l'ensemble des séquences de temps de stabilisation a un degré de séparabilité plus élevé que celui de l'ensemble des caractères originaux. La raison est probablement due à l'information sur l'ordre temporel des séquences qui est extraite par la transformation.

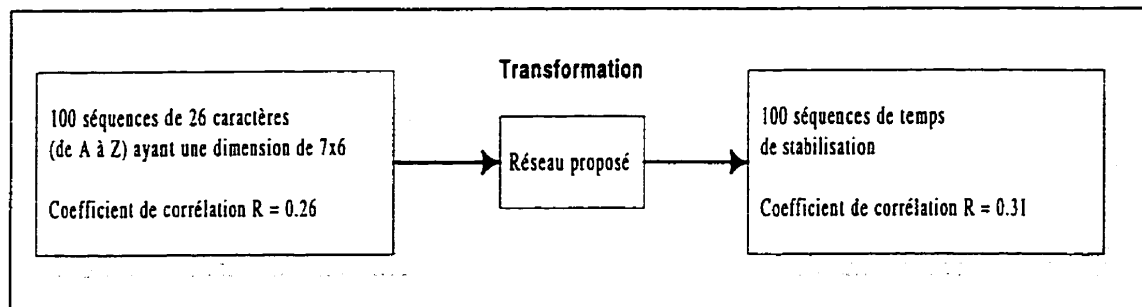


**Figure 3.2** Transformation de l'ensemble des séquences de 5 caractères

### **3.1.2b** Expérience 2: transformation des séquences de 26 caractères

Cette expérience est effectuée pour vérifier si le réseau peut également avoir une sensibilité élevée aux séquences temporelles ayant une longueur beaucoup plus importante. Nous utilisons un ensemble de 26 caractères (de  $A$  à  $Z$ ). En raison du très grand nombre possible de séquences générées par permutation de l'ensemble de 26 caractères, nous réalisons le test sur seulement 100 séquences de 26 caractères. En utilisant le même procédé que dans l'expérience précédente, nous ne trouvons aucune occurrence d'une même séquence de temps de stabilisation. Il semble que le réseau peut traiter

des séquences de stimuli dont la longueur est assez importante. Pourtant, l'ensemble de temps de stabilisation a un degré de séparabilité légèrement plus bas que celui des données initiales (Figure 3.3).



**Figure 3.3** Transformation de l'ensemble des séquences de 26 caractères

### 3.1.3 Expériences avec des séquences de voyelles

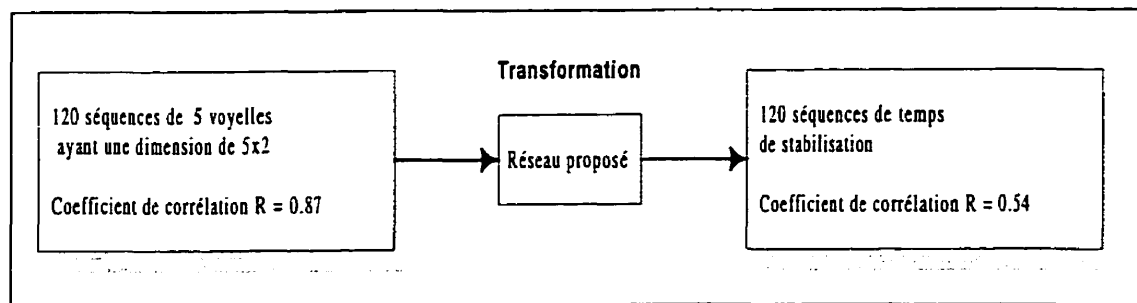
Le but de ces expériences est de vérifier si le réseau peut fonctionner avec des données obtenues à partir de signal de parole. Les données de parole utilisées sont disponibles sur le site WEB de (Merz, 1998). Nous utilisons les voyelles prononcées par un seul locuteur pour onze mots anglais: *heed* (i), *hid* (I), *head* (E), *had* (A), *hard*(a:), *hud* (Y), *hod* (O), *hoard* (C:), *hood* (U), *who'd* (u:) et *heard* (3:). Le signal d'entrée est filtré passe-bas à 4.7 kHz puis il est échantillonné à 10 kHz et quantifié sur 12 bits. Une analyse LPC d'ordre 12 est réalisée afin d'extraire 10 coefficients par fenêtre de signal. Une fenêtre de Hamming de 512 points est centrée au préalable sur la zone stable de chaque voyelle. On utilise donc 10 paramètres LPC par voyelle. Le réseau

comprend 10 neurones (une matrice de 5x2). Chaque paramètre est présenté sur l'entrée  $e_i$  d'un neurone.

### 3.1.3a Expérience 1: transformation des séquences de cinq voyelles

Une séquence de 5 voyelles des mots anglais *heed* (*i*), *hid* (*I*), *head* (*E*), *had* (*A*) et *hard*(*a:*) est choisie. À partir de cette séquence, nous générons 120 autres séquences par permutation des 5 voyelles. En utilisant le même procédé de transformation qu'aux expériences précédentes, nous obtenons 120 séquences de temps de stabilisation. Nous analysons ces dernières et trouvons quelques occurrences de la même séquence. Pour chaque séquence, nous calculons tout d'abord le taux moyen pour lequel cette séquence se produit dans l'ensemble entier des 120 séquences de temps de stabilisation. Basé sur le taux d'occurrence moyen de chaque séquence, nous calculons le taux d'occurrence moyen global pour l'ensemble entier de 120 séquences de temps de stabilisation. Ce taux d'erreur est de l'ordre de 0.36%.

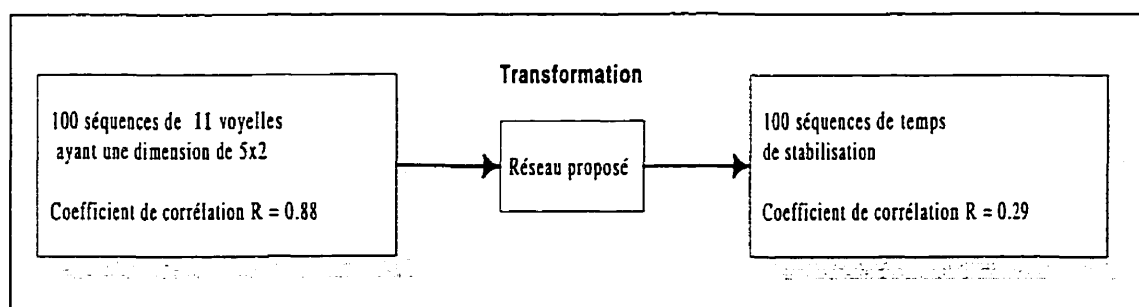
L'ensemble des temps de stabilisation a un coefficient de corrélation ( $R = 0.54$ ), et par conséquent un degré de séparabilité plus élevé que celui observé à partir des vecteurs LPC de la parole ( $R = 0.87$ ) (Figure 3.4).



**Figure 3.4** Transformation de l'ensemble des séquences de 5 voyelles

### 3.1.3b Expérience 2 : transformation des séquences de 11 voyelles

Nous utilisons une séquence de 11 voyelles des mots anglais *heed* (*i*), *hid* (*I*), *head* (*E*), *had* (*A*), *hard*(*a:*), *hud* (*Y*), *hod* (*O*), *hoard* (*C:*), *hood* (*U*), *who'd* (*u:*) et *heard* (*3:*). Lors du test, nous utilisons un ensemble de séquences de voyelles qui comprend 100 combinaisons des 11 voyelles choisies. Le taux d'occurrence est de 0% (aucune occurrence d'une même séquence). L'ensemble de temps de stabilisation a également un degré de séparabilité ( $R = 0.29$ ) plus élevé que celui ( $R = 0.88$ ) des vecteurs LPC (Figure 3.5).



**Figure 3.5** Transformation de l'ensemble des séquences de 11 voyelles

### 3.1.4 Discussion

Le travail présenté dans cette section est basé sur l'idée d'utiliser le comportement dynamique du réseau proposé en tant que moyen d'effectuer une «projection» des séquences de stimuli d'entrée en séquences de temps de stabilisation (du réseau). Cette projection permet non seulement la préservation de l'information temporelle, mais aussi la réduction de la dimension des stimuli d'entrée. En effet, les résultats montrent qu'il est possible d'employer le réseau pour transformer l'information structurée spatio-temporellement en vecteur dont les composantes sont discrètes (temps de stabilisation). Cette transformation se fait par le réseau de façon auto-organisée. Nous montrons expérimentalement que le vecteur obtenu après transformation a souvent un degré de séparabilité plus élevé que celui des stimuli d'entrée initiaux. Cette transformation facilite la tâche de reconnaissance des formes dynamiques spatio-temporelles. Ainsi, le réseau proposé peut être utilisé comme pré-traitement pour les systèmes de reconnaissance spatio-temporelle où la structure temporelle de l'information est cruciale. Les expériences présentées sont relativement limitées mais l'aptitude du réseau proposé à traiter l'information structurée dans le temps laisse à penser que ce réseau est un bon candidat pour la mise au point de systèmes de reconnaissance des formes dynamiques.

## 3.2 Détection de mouvement

La détection de mouvement est un des sujets de recherche dans le domaine de la vision par ordinateur. Plusieurs approches sont proposées dans la littérature (Davis et Bobik, 1997; Chappelier, 1996; Pallbo, 1993). En général, la tâche principale est la détection d'un objet (la présence, le lieu, etc.) qui effectue le mouvement dans un environnement observé. Nous présentons ici l'utilisation du modèle de réseau proposé comme un filtrage spatio-temporel en vue de la détection de mouvement. Basé sur l'évolution du seuil de décharge des neurones, il est possible d'identifier un ensemble de neurones (considéré comme la sortie du réseau) qui soit représentatif de la présence et du lieu d'un objet en mouvement dans une séquence d'images d'entrée. Ceci est en effet une autre exploitation possible et originale des caractéristiques dynamiques et spatio-temporelles du réseau proposé.

### 3.2.1 Méthodologie

L'expérience suivante illustre le principe d'utilisation de l'activité de décharge des neurones pour représenter le mouvement à l'aide du réseau proposé. L'environnement observé a deux objets: une maison et un arbre (Figure 3.6a). Le mouvement à détecter est supposé provenir d'un objet (une personne, Figure 3.6b) qui apparaît par la suite dans la scène observée. Au sens de la détection de mouvement, le réseau doit identifier l'objet (la personne) dans la scène par l'intermédiaire de son activité de décharge lorsqu'une

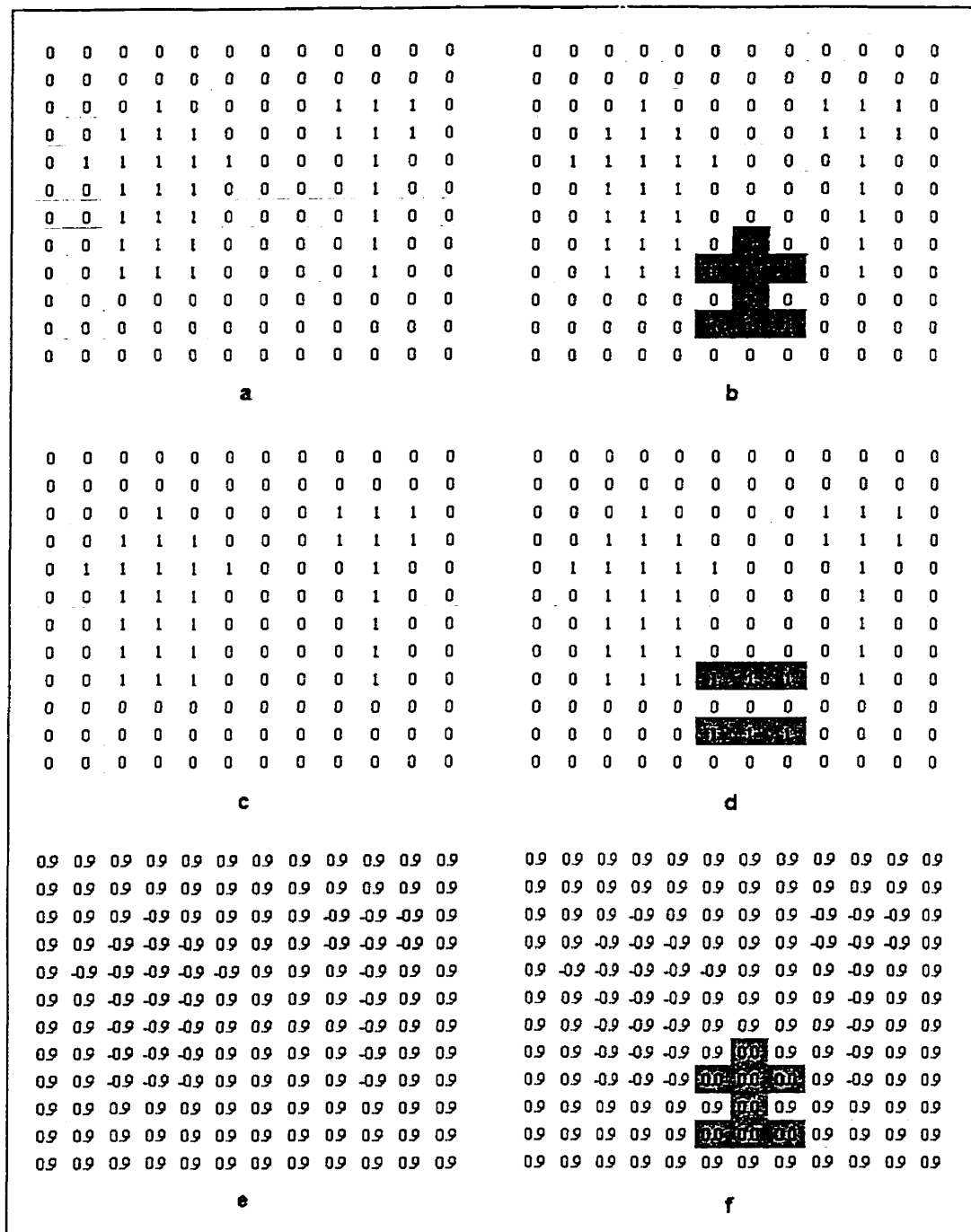


séquence d'images (par exemple, une séquence de deux images Figure 3.6a et b) lui est présentée successivement.

Le réseau comprend 144 neurones qui sont distribués dans un plan de dimension 12x12. Il est composé de 70% de neurones excitateurs et de 30% de neurones inhibiteurs. Le rayon de voisinage est de 2 pour les neurones excitateurs et de 1 pour les neurones inhibiteurs. Chaque neurone reçoit le signal d'entrée du pixel correspondant dans l'image d'entrée. Dans cette expérience, chaque image est présentée au réseau de façon séquentielle pour une durée de 500 itérations par image. Il n'y a pas d'initialisation du réseau entre les présentations successives.

L'activité de décharge du réseau peut être représentée par la sortie des neurones qui est 1 pour un état de décharge ou 0 pour un état de non décharge. Cette activité peut également être représentée par le seuil des neurones. Comme nous avons montré au chapitre précédent, ces paramètres peuvent caractériser des stimuli d'entrée. Nous présentons dans les sections *c*, *d*, *e*, et *f* de la figure 3.6 ces paramètres obtenus après la présentation de l'image 3.6a (la scène observée) et de l'image 3.6b (qui contient l'objet en mouvement) successivement.

En analysant la sortie des neurones (Figure 3.6c et d), nous constatons que le réseau représente complètement la scène observée qui est toujours la même et statique. Par contre, l'objet en question (la personne) n'est pas complètement identifié.



**Figure 3.6** L'activité du réseau lors de la présentation d'une séquence d'images d'entrée : a) l'image de la scène observée; b) l'image avec l'apparition d'un objet (la personne); c) et d) la sortie des neurones après la présentation de l'image 3.6a et 3.6b précédée par l'image 3.6a respectivement; e) et f) le seuil des neurones après la présentation de l'image 3.6a et 3.6b précédée par l'image 3.6a respectivement.

D'autre part, l'ensemble des seuils des neurones représente très distinctement le fond, les objets du premier plan de la scène (la maison et l'arbre) et la personne. L'ensemble des neurones qui appartiennent au fond a un même seuil  $\theta = 0.9$  (la valeur est arrondie ici). Il en est de même pour les neurones appartenants aux objets du premier plan. Ces neurones ont un même seuil  $\theta = -0.9$ .

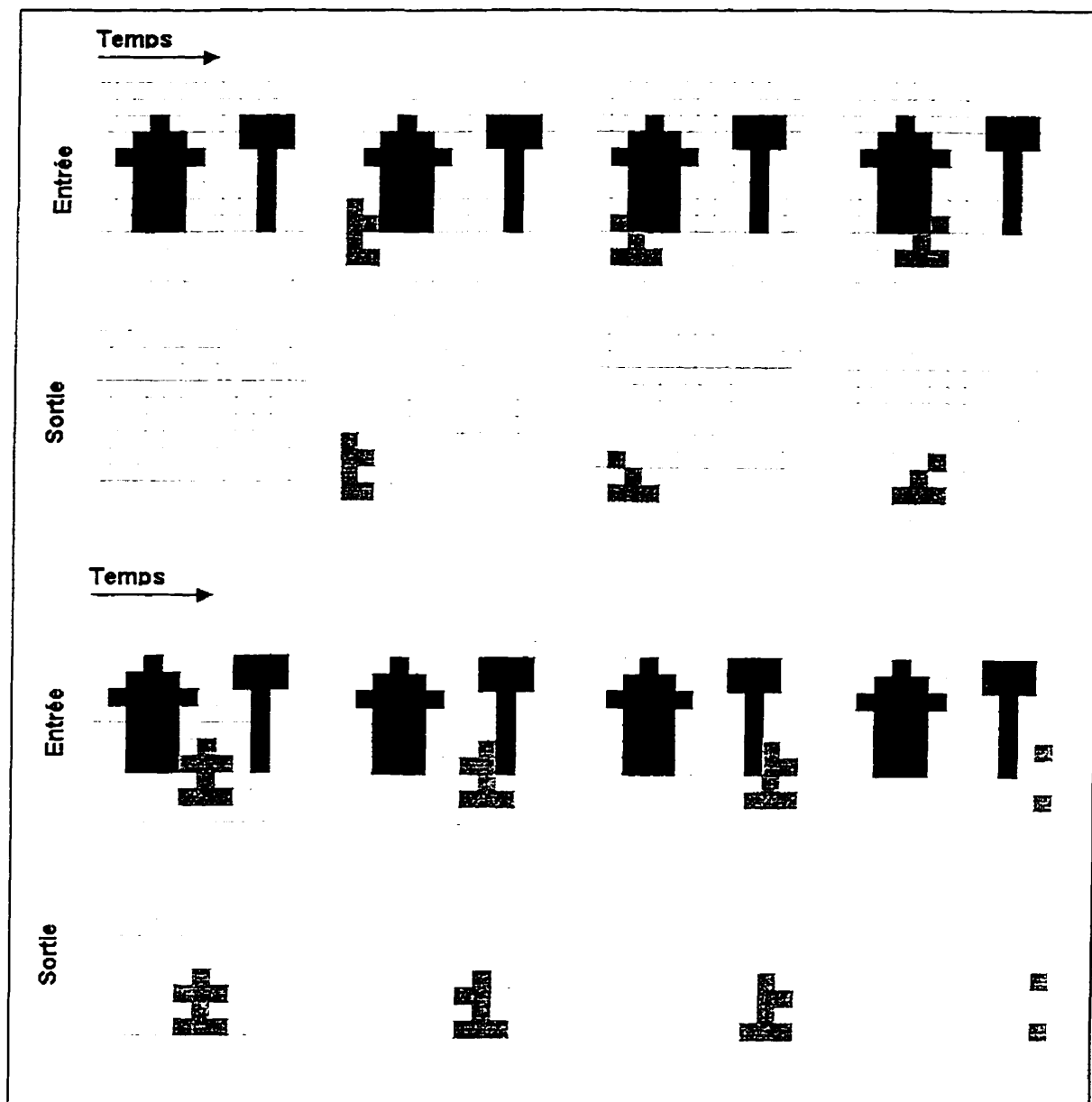
Les neurones qui appartiennent à la personne ont un même seuil dont la valeur est très distincte de celle des autres objets ( $\theta = 0.0$ ). Nous constatons que les neurones représentant des objets statiques dans la scène observée préservent toujours le même seuil (soit 0.9 ou  $-0.9$ ). Par contre, l'objet en mouvement est représenté par un ensemble de neurones dont le seuil change de 0.9 à 0.0. Autrement dit, les neurones qui représentent un objet en mouvement ont un même seuil  $\theta = 0.0$  après la présentation de l'image d'entrée. Ce seuil est distinct de ceux des neurones qui représentent les objets statiques. Basé sur le seuil des neurones, nous pouvons donc identifier les neurones qui représentent l'objet en mouvement. Ce dernier est détecté par sa présence et son lieu (c'est-à-dire sa position dans la scène observée).

### 3.2.2 Résultat de simulation

Dans l'expérience suivante, une séquence continue d'images binaires est utilisée comme entrée. La sortie du réseau est une séquence d'un ensemble de neurones dont le mouvement se produit à leur position. Ces neurones sont identifiés en se basant sur leur seuil de décharge  $\theta$ . Ici, l'entrée est une

séquence d'images d'une personne se déplaçant devant une maison et un arbre. Ces données sont inspirées du travail de Chappelier (1996).

La séquence des images d'entrée et la séquence correspondante des sorties du réseau sont présentées à la figure 3.7. Chaque image d'entrée y est présentée de façon séquentielle pour une durée de 500 itérations. Il n'y a pas d'initialisation du réseau entre les présentations successives. Nous constatons que la sortie du réseau (basée sur les seuils des neurones) représente effectivement la personne en mouvement.



**Figure 3.7** Détection de mouvement par le réseau proposé: le «film» présenté au réseau représente, de façon schématique, une personne passant devant une maison et un arbre. La séquence des entrées et des sorties du réseau se déroule de gauche à droite et de haut en bas: l'entrée est présentée au dessus de la sortie correspondante pour chaque pas de temps. On observe en sortie les neurones qui représentent la personne en mouvement. Le réseau effectue donc une détection du mouvement de la partie dynamique (la personne).

### 3.2.3 Discussion

Les expériences présentées sont relativement limitées et il y a lieu d'approfondir le travail afin de bien connaître les limites du réseau en détection de mouvement. Ce travail est assez loin de notre préoccupation de recherche dans le cadre de cette thèse. Le but ici est de présenter un principe avec lequel on peut envisager une reconnaissance de mouvement à l'aide du réseau proposé. Une étude de performance concernant la nature des séquences d'image d'entrée, par exemple la durée de présentation d'une image d'entrée (c'est-à-dire la vitesse du «film» d'entrée), la taille des objets dans l'image, est recommandée pour un travail futur. En particulier, les expériences de la détection de mouvement portant sur les images vidéos sont essentielles en vue de mettre en valeur cette approche. Cependant, le présent travail démontre que le réseau proposé représente effectivement la relation spatio-temporelle dans un signal dynamique non stationnaire. Ceci permet la détection d'objets en mouvement dans une séquence d'images.

### 3.3 Discussion et conclusion

Dans ce chapitre, nous avons présenté deux approches qui permettent d'exploiter les caractéristiques du réseau proposé afin de représenter l'information spatio-temporelle dans les signaux dynamiques non stationnaires. Dans une approche, nous proposons l'utilisation du réseau comme moyen d'effectuer une transformation des séquences de signaux d'entrée en séquences de temps de stabilisation du réseau. Cette transformation non seulement permet une préservation de l'information structurée spatio-temporellement mais aussi une réduction de la dimension des signaux d'entrée. L'application possible de cette approche est plausible dans les tâches de reconnaissance pour lesquelles l'information contextuelle est cruciale, par exemple en reconnaissance de parole, en identification de séquences temporelles et en reconnaissance de signature.

Dans une autre approche, nous proposons l'utilisation du réseau comme filtrage spatio-temporel permettant la détection d'objets en mouvement dans une séquence d'images. Le principe sous-jacent à la détection de mouvement (par le réseau proposé) réside dans l'utilisation de l'évolution du seuil de décharge des neurones comme critère pour identifier l'objet en mouvement. Celui-ci est représenté par l'ensemble des neurones dont le seuil varie distinctement de ceux qui sont représentatifs de la scène observée (y compris les objets statiques). Expérimentalement, nous avons montré que ce principe permet la détection d'objets en mouvement dans une séquence d'images.

Même si les expériences présentées ici sont relativement limitées, elles montrent l'aptitude du réseau proposé à traiter l'information spatio-temporelle résidante dans des processus dynamiques non stationnaires. Ceci est en effet le but visé de la recherche dans cette thèse.



## **Chapitre 4**

# **Prototype pour un système d'identification du locuteur à l'aide du réseau proposé**

**L**a reconnaissance du locuteur est un terme générique pour discriminer parmi plusieurs personnes en fonction de leur voix. On distingue en général l'identification et la vérification du locuteur. L'identification consiste à reconnaître un locuteur appartenant à une population de locuteurs. La vérification consiste à accepter ou à refuser une identité proclamée par un locuteur. On distingue également la reconnaissance indépendante du texte et la reconnaissance effectuée sur la base d'un texte imposé.

L'expression vocale est une caractéristique propre d'un locuteur; ainsi est-il possible, dans des conditions normales, de reconnaître son correspondant au cours d'une conversation téléphonique.

Les variations individuelles entre locuteurs ont deux origines essentielles. En premier lieu, les caractéristiques physiques de l'appareil de phonation influencent les formants, la valeur moyenne de la fréquence de glotte, etc, et cela indépendamment de la phrase prononcée. D'autre part, une même phrase n'est pas prononcée de la même façon par deux locuteurs. On observe des différences dans les débits d'élocution, dans l'étendue des variations de la fréquence de glotte, etc.

La mise en place de système de reconnaissance versatile fait souvent appel à des analyses perceptives inspirées de modèles du système auditif (Patterson, 1995). Ces analyses codent l'information liée à la parole dans un espace à représentation spatio-temporelle complexe et difficile à caractériser via les algorithmes couramment utilisés en reconnaissance. En effet, ces représentations spatio-temporelles sont structurées à la fois spectralement et temporellement. L'information est en général non stationnaire. Il y a donc lieu de mettre au point des paramètres et des algorithmes de reconnaissance des formes capables d'exploiter cette information temporelle. Or, une majorité des algorithmes contemporains assument une certaine stationnarité du signal et des paramètres d'analyse. De plus, ces systèmes requièrent un apprentissage supervisé souvent long et fastidieux. Nous pensons qu'il y a lieu d'étudier l'élaboration de nouvelles techniques d'analyse et de reconnaissance adaptées à l'exploitation de l'information temporelle fine telle que générée par exemple via des analyses d'inspiration perceptive. Une information temporelle fine serait par exemple celle qui est liée à l'enveloppe des signaux à la sortie d'un banc de filtres cochléaire (Rouat, 1997b).

Il apparaît intéressant d'élaborer de nouveaux outils aptes à traiter une information dynamique et non stationnaire. Ce type d'outil devrait pouvoir être utilisé en reconnaissance de formes afin de traiter une information codée dans

le temps (parole, par exemple). Rappelons que ceci est la motivation principale de notre recherche dans le cadre de cette thèse.

Dans ce chapitre, nous présentons un prototype d'un système d'identification du locuteur à l'aide du modèle de réseau proposé et d'une analyse non linéaire de la parole (la sortie d'un banc de filtres cochléaire). Les expériences portant sur des données de parole téléphonique bruitée ici sont indispensables et permettent d'évaluer le modèle de réseau proposé.

#### 4.1 Analyse de la parole par modulation d'amplitude dans le système auditif

Actuellement, la plupart des paramètres acoustiques utilisés dans les systèmes de reconnaissance de parole et du locuteur sont extraits à partir de la représentation spectrale de la parole. On doit donc supposer que le signal est stationnaire dans la fenêtre d'analyse. Ceci donne une représentation de la parole qui est en fait une estimation des valeurs moyennes des paramètres dans le temps. Par conséquent, la structure à court terme de la parole est partiellement cachée par l'analyse et les structures fines de la modulation d'amplitude ne peuvent pas être détectées. Nous définissons ici les structures à court terme de la parole comme des caractéristiques de la parole observées par un système auditif périphérique sur des échelles très courtes (quelques ms). Nous sommes intéressés par la structure à court terme observée à la sortie d'un banc de filtres cochléaires en termes de modulation d'amplitude. Cette structure est représentative et caractéristique des modèles d'audition (Rouat, 1997b). De plus, elle permet de préserver l'information temporelle et ne nécessite pas l'hypothèse de la stationnarité du signal.

Nous utilisons un banc de 24 filtres centrés dans un intervalle de 330 Hz à 4700 Hz. Cette gamme de fréquences peut contenir la partie la plus importante de l'information fréquentielle de la parole. La sortie de chaque filtre est un signal filtré passe-bande centré autour d'une fréquence centrale  $f_i$  où l'indice  $i$  correspond au canal  $i$ . Le signal de sortie  $s_i(t)$  du canal  $i$  peut être considéré comme un signal modulé en amplitude et en phase dont la fréquence porteuse est  $f_i$ .

$$s_i(t) = A_i(t) \cos([\omega_i(t) + \phi_i(t)]) \quad (4.1)$$

$A_i(t)$  est l'amplitude modulée (enveloppe) de  $s_i(t)$ ,  $\phi_i(t)$  est la phase modulée et  $\omega_i = 2\pi f_i$ .

L'enveloppe des signaux  $s_i(t)$  est calculée par:

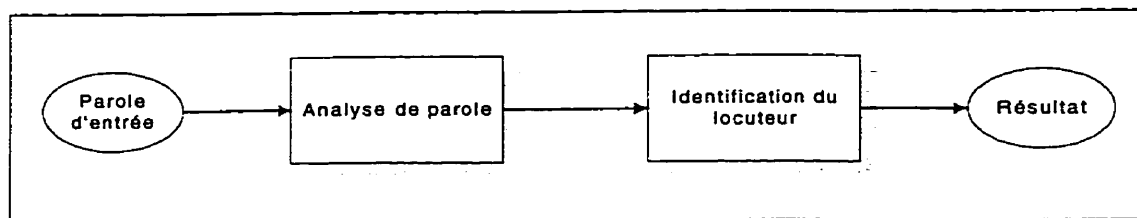
$$A_i(t) = \sqrt{s_i(t)^2 + s_i(t)_q^2} \quad (4.2)$$

où  $s_i(t)_q$  est la transformée de Hilbert de  $s_i(t)$ .

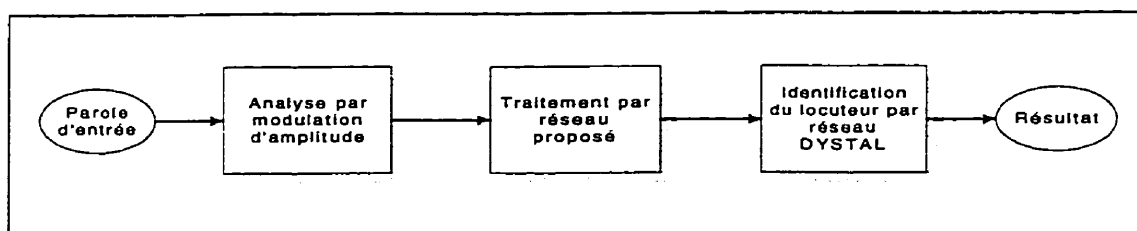
## 4.2 Système d'identification du locuteur

En principe, un système d'identification du locuteur comprend les modules suivants (Figure 4.1):

- ◆ un module d'acquisition des signaux.
- ◆ un module d'analyse du signal dont le rôle est d'extraire du signal acoustique les paramètres robustes et pertinents afin de caractériser le locuteur. Les analyses conventionnelles (par exemple, la transformation de Fourier et le LPC «Linear Predictive Coding») supposent la stationnarité des signaux de parole à l'intérieur d'une fenêtre d'analyse, en conséquence l'information temporelle est «écrasée». C'est sur ce point que nous proposons d'utiliser une analyse par modulation d'amplitude et un traitement par le réseau proposé dans le but de préserver et d'exploiter plus efficacement l'information temporelle.
- ◆ un module de reconnaissance dont la tâche est d'identifier le locuteur à partir des paramètres fournis par le module d'analyse. Plusieurs techniques de reconnaissance peuvent être employées ici telles que les techniques statistiques, les réseaux de neurones, les modèles de Markov, etc.



**Figure 4.1** Architecture générale d'un système d'identification du locuteur



**Figure 4.2** Architecture du système d'identification du locuteur à l'aide d'une analyse par modulation d'amplitude et d'un traitement par le réseau proposé.

Dans ce travail, nous proposons un prototype d'un système d'identification du locuteur à l'aide d'une analyse par modulation d'amplitude et d'un traitement par le réseau proposé (Figure 4.2). L'analyse par modulation d'amplitude emploie un banc de 24 filtres (décrit dans la section 4.1) et calcule l'enveloppe à la sortie de chacun des 24 filtres. Notons que cette analyse ne fait pas partie de notre préoccupation de recherche dans le cadre de cette thèse. Nous nous intéressons principalement à l'analyse des enveloppes par le réseau proposé. Notre tentative d'utilisation du réseau de neurones à décharges est en effet, à notre connaissance, la première dans ce genre d'analyse (par réseaux de neurones). La tâche est donc d'extraire, à partir des enveloppes, des paramètres qui sont représentatifs du locuteur. En d'autres termes, le traitement des enveloppes par le réseau doit fournir de nouveaux paramètres afin de

caractériser le locuteur. Ces paramètres seront ensuite traités dans le module d'identification du locuteur. N'importe quelle technique de reconnaissance peut être utilisée dans ce module. Nous y choisissons le réseau DYSTAL en raison de la simplicité de son utilisation et de sa performance pour ce type d'application.



## 4.3 Traitement des enveloppes par le réseau proposé

### 4.3.1 Données de la parole

La base de données de parole téléphonique bruitée SPIDRE («Speaker IDentification REsearch») (Godfrey, 1992) est choisie dans notre expérience. Cette base de données (27 hommes et 18 femmes) se compose de 4 enregistrements pour chaque locuteur provenant de 3 combinés téléphoniques différents. Deux enregistrements utilisent le même combiné téléphonique. Les enregistrements sont effectués à partir de conversations dans des conditions réelles. Par conséquent, le signal de parole ici est corrompu par le bruit et les interférences à travers les lignes et les combinés téléphoniques.

Pour les expériences, nous choisissons les 10 locuteurs suivants: *sp1497*, *sp1499*, *sp1575*, *sp1415*, *sp1436*, *sp1096*, *sp1528*, *sp1130*, *sp1531* et *sp1007*. Parmi ces locuteurs, il y a trois femmes: *sp1415*, *sp1096* et *sp1007*. Nous considérons tous les locuteurs dans un même contexte, c'est-à-dire que tous les signaux sont extraits dans un même mot choisi dans les conversations. Le mot choisi est le mot anglais «*and*». La fréquence d'échantillonnage est de 8 kHz. La durée du signal (c'est-à-dire le nombre d'échantillons) n'est pas la même pour tous les signaux<sup>1</sup> (Tableau 4.1).

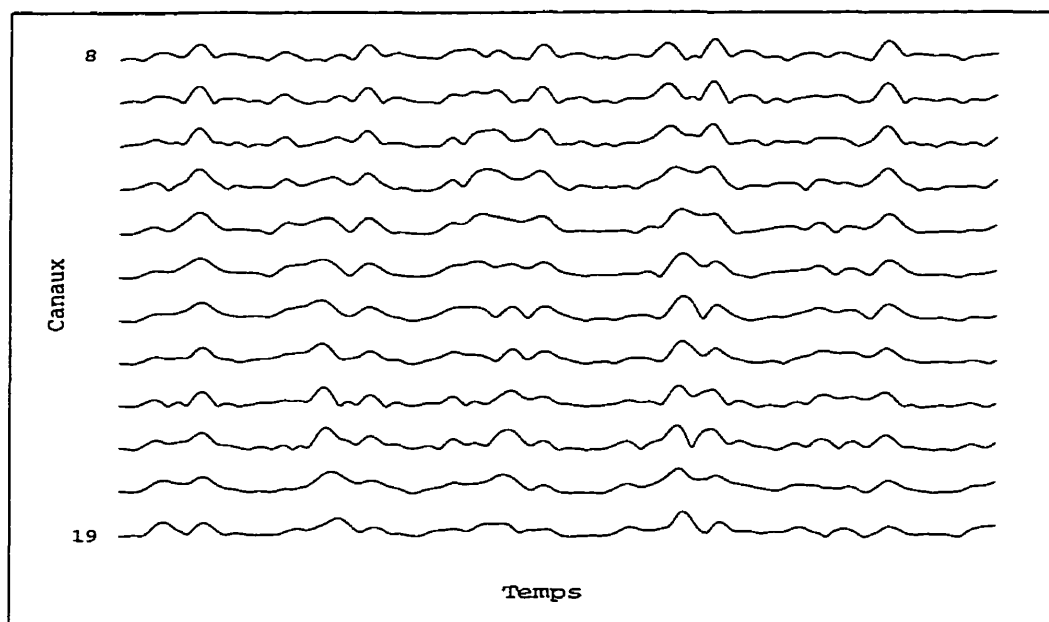
---

<sup>1</sup> Un merci tout particulier à Ezzaidi H. (ERMETIS, UQAC) pour la préparation de ces données.

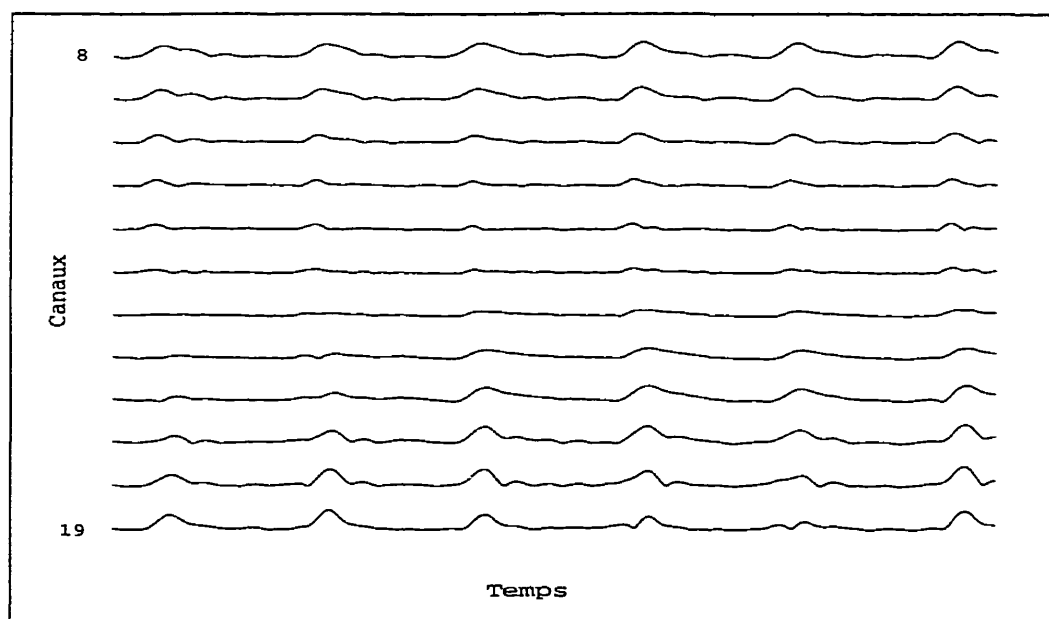
**Tableau 4.1** Données de la parole échantillonnée à 8 kHz (contexte «and»)

No	Locuteur	Nombre d'échantillons			
		Combiné 1	Combiné 1	Combiné 2	Combiné 3
1	<i>sp1497</i>	17575	13497	45359	44407
2	<i>sp1499</i>	3326	23344	18135	26301
3	<i>sp1575</i>	9170	6930	12853	3887
4	<i>sp1415*</i>	21979	48239	44793	44793
5	<i>sp1436</i>	32793	27673	20543	25024
6	<i>sp1096*</i>	13893	21830	33199	59580
7	<i>sp1528</i>	30306	14621	15658	20701
8	<i>sp1007*</i>	18614	Non disponible	15333	15332
9	<i>sp1130</i>	7089	7837	38950	9570
10	<i>sp1531</i>	12372	7012	24624	5088

\* Femme



a)



b)

**Figure 4.3** Envelopes à la sortie du banc de filtres: a) locuteur *sp1497* et b) locuteur *sp1499*. La durée est de 50 ms (400 échantillons à 8 kHz). Les enveloppes des signaux cochléaires sont ordonnées pour les canaux de 8 (basses fréquences) à 19 (hautes fréquences). Les 24 canaux ne sont pas représentés afin d'alléger les figures.

### 4.3.2 Analyse des enveloppes à la sortie du banc de filtres

Les enveloppes sont obtenues à partir d'un banc de 24 filtres cochléaires. À titre d'exemple, nous présentons à la figure 4.3 les enveloppes de 12 canaux obtenues pour les locuteurs *sp1497* et *sp1499* respectivement, la durée du signal est de 50 ms (400 échantillons à 8 kHz). Pour les expériences effectuées le réseau traite 2.2 secondes (17575 échantillons) du locuteur *sp1497* et 2.3 secondes (18135 échantillons) du locuteur *sp1499*. En pratique, le réseau peut traiter n'importe quelle longueur de signal.

La question qui se pose *a priori* est la pertinence de l'information résidante dans ces enveloppes pour l'identification du locuteur. À notre connaissance, aucune étude ne peut répondre à cette question. D'après les figures (Figure 4.3a, b), nous constatons que nous pouvons les analyser de l'une des manières suivantes:

- ◆ *horizontalement*: l'analyse se fait exclusivement pour chaque canal. L'information alors est peut-être due à la forme des pics, à la distribution des pics (l'intervalle de temps entre les pics, la fréquence des pics, etc.) dans le temps, etc.
- ◆ *verticalement*: l'analyse se fonde sur la position relative des pics à travers les canaux et dans le temps. L'information réside donc dans la synchronisation ou le décalage en terme de phase des pics dans le temps.
- ◆ *globalement*: l'analyse se fait à la fois horizontalement et verticalement. De cette manière, nous espérons exploiter l'information de façon plus complète. Ceci permet de tenir compte de

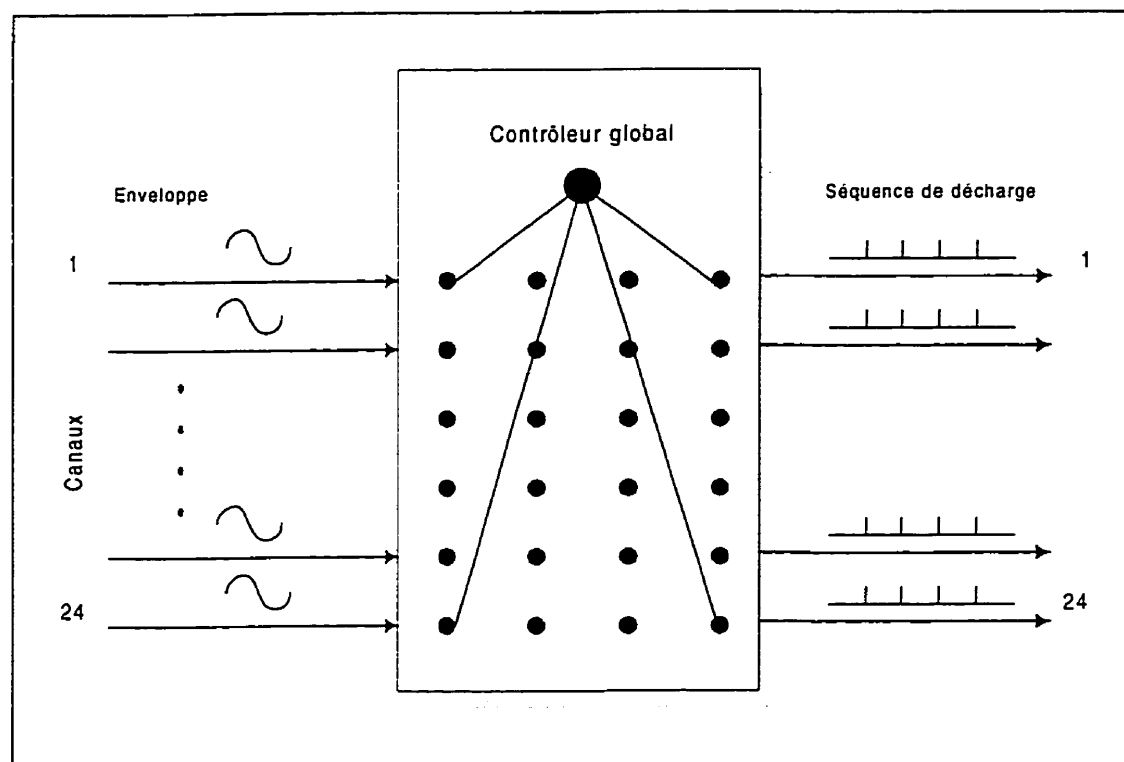
l'interférence entre les pics à travers les canaux. Cette interférence serait représentée par l'activité d'interaction entre les neurones dans le réseau. Nous adopterons donc cette manière d'analyser pour le traitement des enveloppes par le réseau proposé.

L'idée sous-jacente est que les enveloppes à la sortie du banc de filtre cochléaires sont introduites au réseau pour être traitées. Après le traitement, nous considérons qu'un ensemble de paramètres du réseau (que nous détaillons plus tard) est représentatif des enveloppes. C'est-à-dire que le réseau effectue une paramétrisation des enveloppes. En d'autres termes, le réseau traite les enveloppes à l'aide de son comportement dynamique complexe et génère une représentation «interne» de l'information spatio-temporelle résidante dans les enveloppes à travers les canaux. Nous souhaitons étudier cette représentation pour l'identification du locuteur.

Le principe général est présenté à la figure 4.4. Le réseau d'une dimension 6x4 comprend 24 neurones correspondant aux 24 canaux à la sortie du banc de 24 filtres cochléaires. Les neurones sont ordonnés de gauche à droite et de haut en bas. Chaque neurone reçoit le signal du canal correspondant de façon continue (en temps réel). C'est-à-dire qu'il reçoit un échantillon du canal correspondant à chaque instant durant la présentation des enveloppes. On procède ainsi pour tous les neurones. De cette manière, le réseau reçoit les enveloppes parallèlement à travers tous les canaux.

Il est à noter que le choix de l'ordonnement des neurones de gauche à droite et de haut en bas ne reflète pas la physiologie. Il est plutôt un choix d'ingénieur pour cette étude exploratoire. Pour un travail futur, on peut envisager d'utiliser d'autres ordonnancements inspirés de la physiologie (par

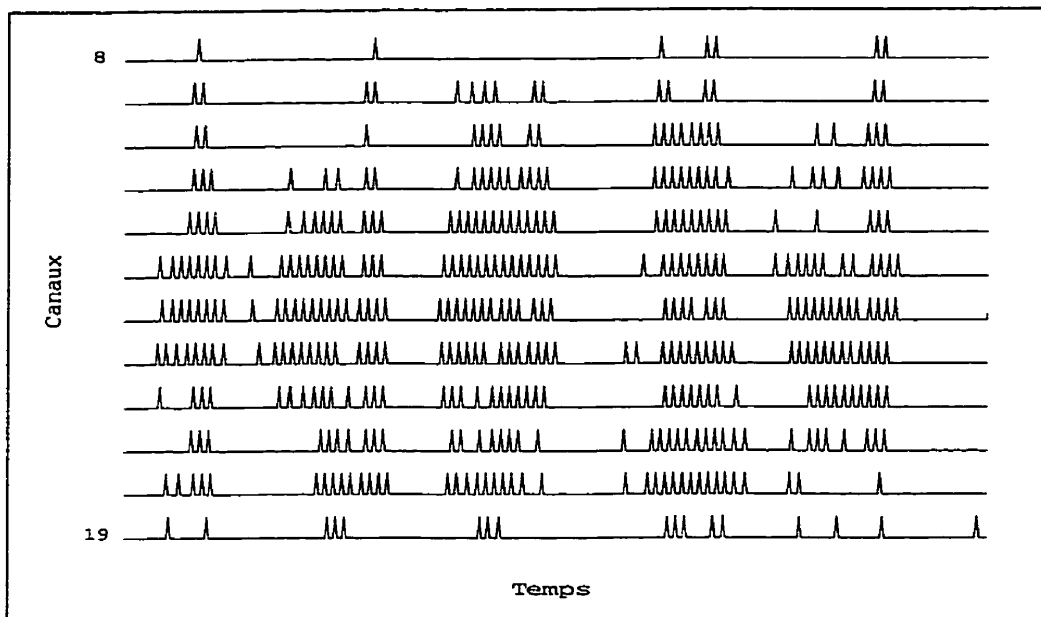
exemple, en se basant sur la distribution des fréquences centrales des filtres cochléaires).



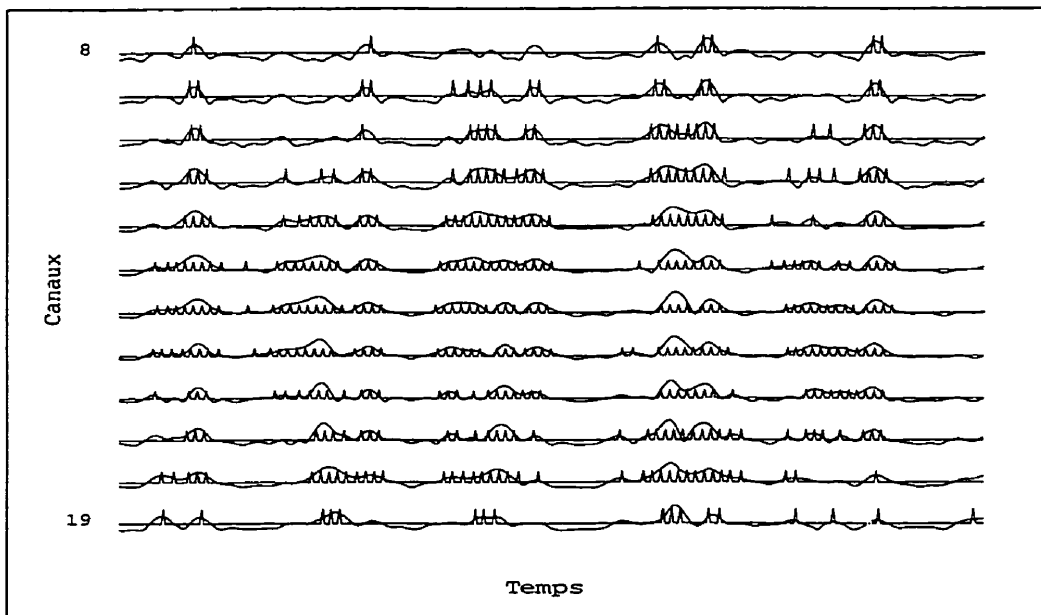
**Figure 4.4** Traitement des enveloppes par le réseau proposé ayant une dimension 6x4 correspondant à 24 canaux à la sortie du banc de filtres cochléaires. Les neurones sont ordonnés de gauche à droite et de haut en bas. Chaque neurone reçoit le signal du canal correspondant de façon continue (c'est-à-dire, en temps réel). La sortie du réseau est une séquence de décharge pour chaque canal.

Un exemple de l'activité de décharge du réseau est présenté à la figure 4.5 lors du traitement des enveloppes du locuteur *sp1497* (Figure 4.3a). En effet, les sorties du réseau sont des séquences de potentiels d'action. Ces dernières sont liées aux signaux d'enveloppe à travers les canaux. Nous observons que les séquences de potentiels d'action représentent effectivement la forme d'une montée de l'enveloppe (la forme est représentée par la durée de

la montée, c'est-à-dire le nombre de pics dans cet intervalle) ainsi que la position relative des pics. Nous constatons également que le choix des paramètres internes du réseau comme  $U_{int}$  dans l'exemple à la figure 4.6 a une influence sur les séquences de potentiels d'action. Ces séquences représentent dans cette figure les enveloppes plus finement qu'à la figure 4.5, ce qui indique que nous avons un contrôle relatif sur la formation de la représentation interne du réseau vis-à-vis des enveloppes.



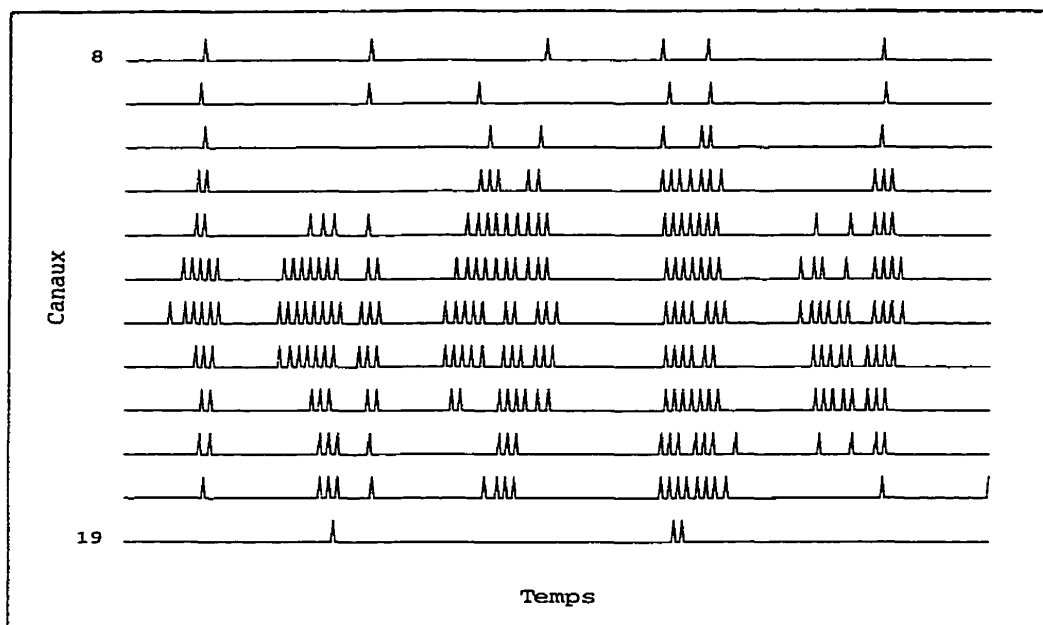
a)



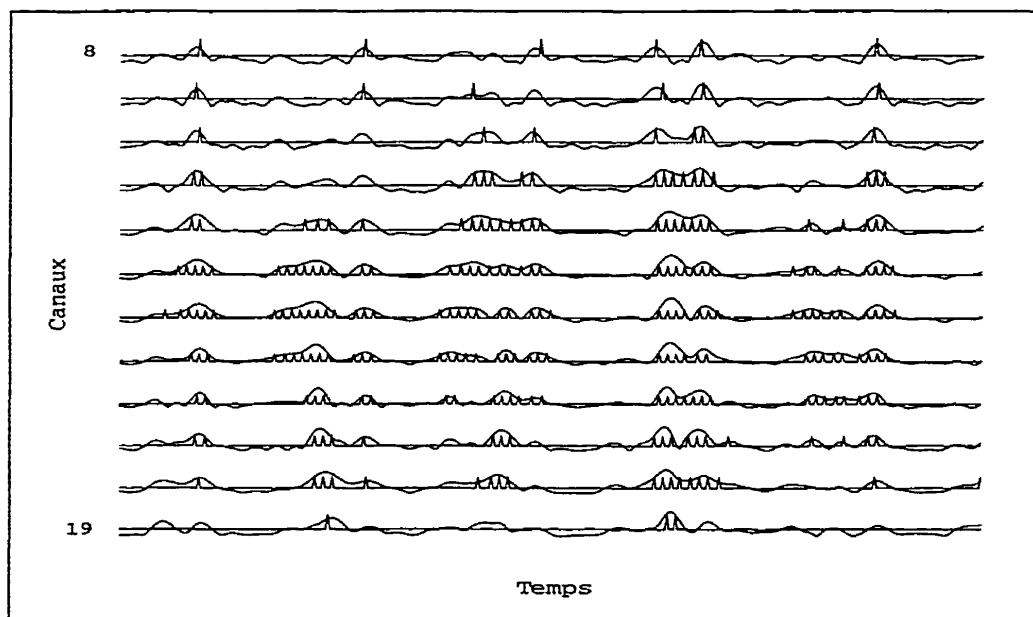
b)

**Figure 4.5** Séquences de potentiels d'action pour le locuteur *sp1497* avec  $U_{int} = -3.0$  (a). On a superposé les enveloppes correspondantes pour mieux illustrer la concordance (b).





a)



b)

**Figure 4.6** Séquences de potentiels d'action pour le locuteur *sp1497* avec  $U_{int} = -5.0$  (a). On a superposé les enveloppes correspondantes pour mieux illustrer la concordance (b).

L'activité de décharge du neurone est non seulement influencée par le signal d'enveloppe de son canal correspondant, mais aussi par les neurones voisins. Cette influence est particulièrement forte lorsqu'il y a une synchronisation des décharges entre les neurones. La synchronisation des décharges intervient lorsqu'il y a synchronisation des enveloppes à travers les canaux. En principe, nous considérons que l'activité de décharge des neurones peut créer une représentation «interne» de l'information spatio-temporelle résidante dans les enveloppes.

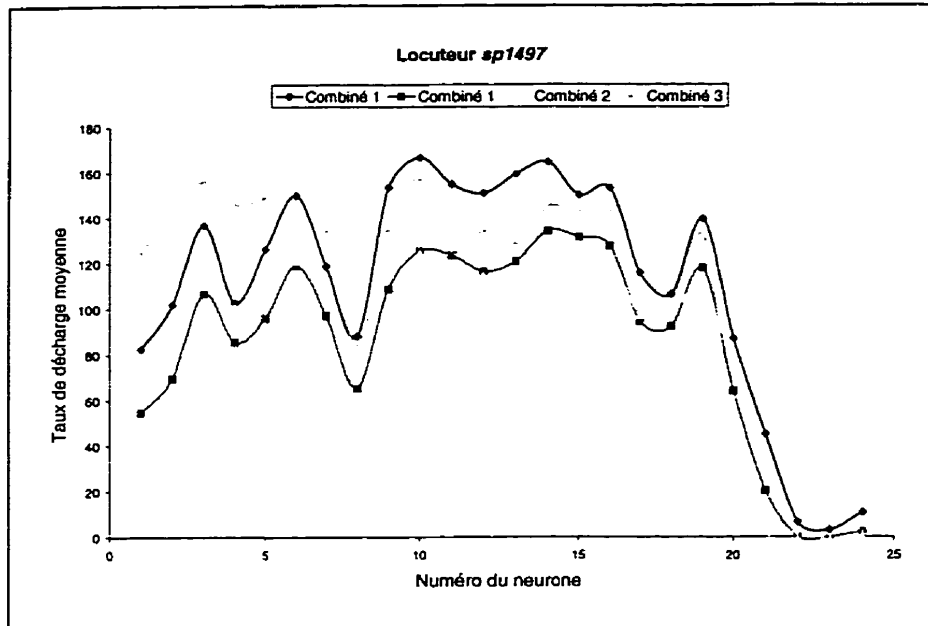
Une question se pose: comment peut-on déterminer cette représentation interne ? En d'autres termes, quels paramètres du réseau peuvent être utilisés pour caractériser cette représentation ?

Puisque les séquences de potentiels d'action à la sortie du réseau représentent effectivement les enveloppes, on peut envisager une analyse de ces séquences afin d'obtenir l'information représentative des enveloppes. On pourrait alors se baser sur les mécanismes de codage temporel (décrits dans le chapitre 1) pour effectuer l'analyse des séquences de potentiels d'action. En raison de la contrainte de temps, nous choisissons dans ce travail le codage basé sur le taux de décharge moyen. C'est en effet le codage le plus simple et le plus utilisé pour l'analyse des séquences de potentiels d'actions. L'analyse basée sur d'autres mécanismes de codage temporel serait souhaitable pour un travail futur.

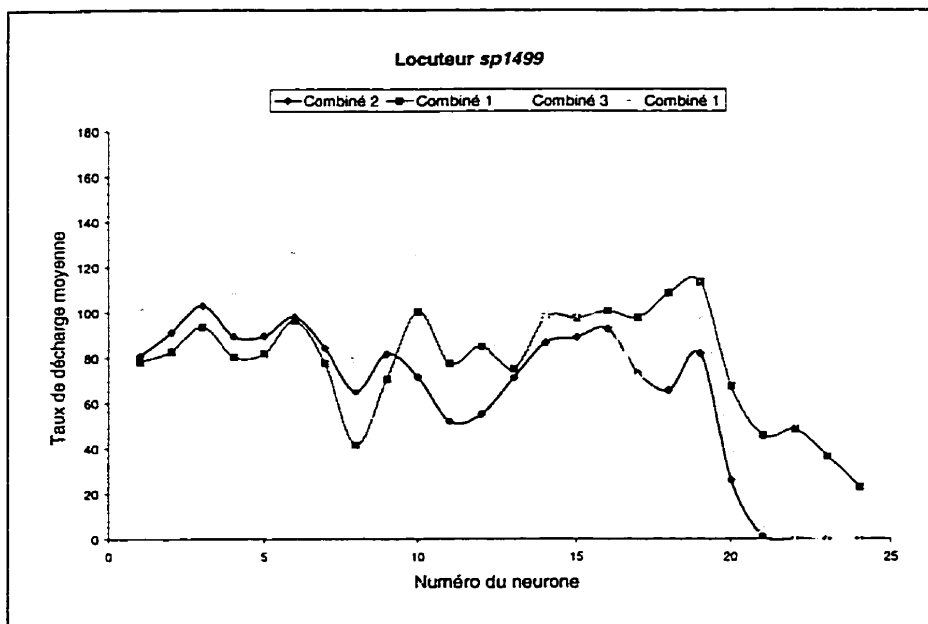
Nous effectuons le traitement des enveloppes par le réseau ayant une dimension de 6x4 (24 neurones) pour dix locuteurs choisis. Rappelons que chaque locuteur a enregistré 4 conversations à partir de 3 combinés téléphoniques différents. Nous traitons le signal (enveloppe) de chaque conversation de façon indépendante. C'est-à-dire que le réseau a les mêmes

conditions initiales (y compris les poids des connexions initiales qui sont initialisées aléatoirement) pour chaque traitement. Le taux de décharge moyen des neurones est considéré comme étant les paramètres de sortie du réseau après le traitement. À titre d'exemple, nous présentons sous formes graphiques le taux de décharge moyen obtenu pour les 4 locuteurs: *sp1497*, *sp1499*, *sp1415* et *sp1096* (Figures 4.7-4.10). Le taux de décharge est estimé pour une durée de 100 ms (c'est-à-dire, 800 échantillons à 8 kHz). En abscisse, on reporte les indices des neurones et en ordonnée le taux de décharge moyen pour chaque neurone.

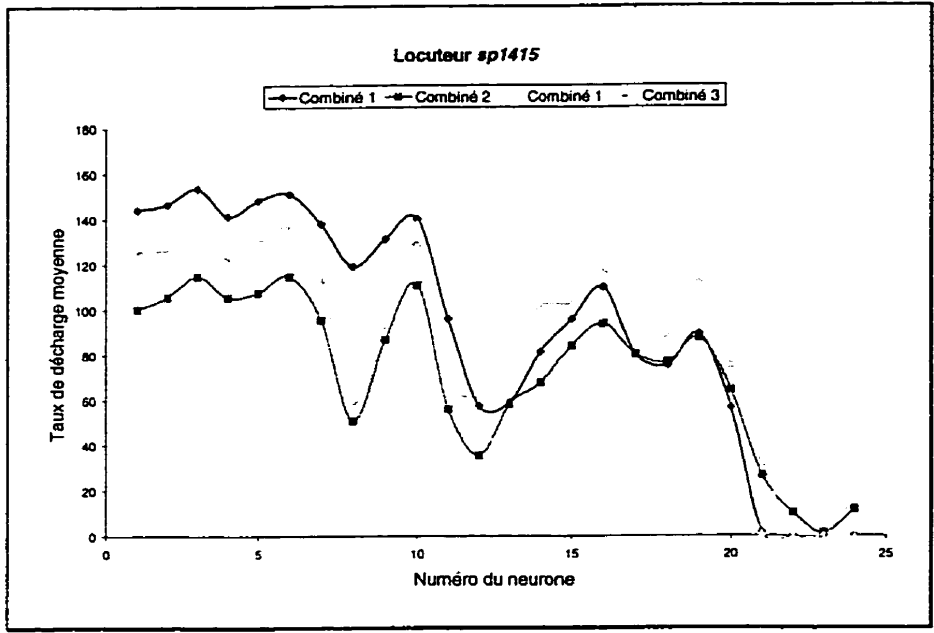
En général, les courbes sont similaires pour un même locuteur et pour des combinés téléphoniques différents. Il est possible de distinguer visuellement les locuteurs à partir de ces courbes. Il est important de souligner que les paramètres obtenus sont vraisemblablement indépendants du combiné et de la durée du signal, ce qui est le but visé. Autrement dit, les paramètres choisis peuvent caractériser les locuteurs et ils sont peu sensibles aux environnements d'opération. Ces observations nous permettent d'envisager l'utilisation des paramètres du taux de décharge moyen (des neurones) en vue de l'identification du locuteur.



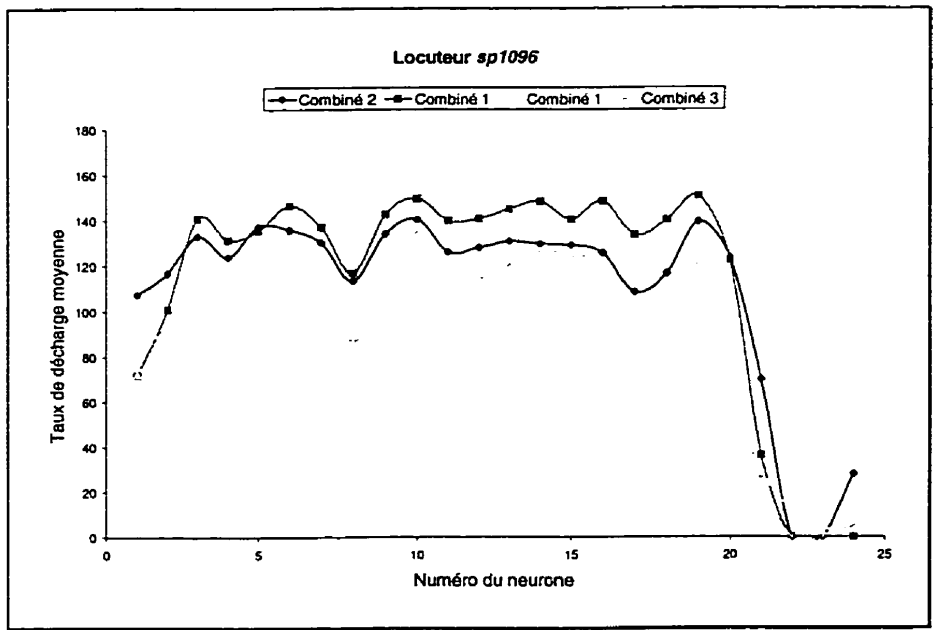
**Figure 4.7** Taux de décharge moyen pour le locuteur *sp1497*



**Figure 4.8** Taux de décharge moyen pour le locuteur *sp1499*



**Figure 4.9** Taux de décharge moyen pour le locuteur *sp1415*



**Figure 4.10** Taux de décharge moyen pour le locuteur *sp1096*

#### 4.4 Identification du locuteur basée sur les paramètres de sortie du réseau proposé

Le traitement des enveloppes à la sortie d'un banc de filtres cochléaires par le réseau proposé est en quelque sorte une transformation des enveloppes d'entrée en paramètres de sortie du réseau. Les enveloppes (sur 24 canaux) de chaque signal enregistré dans une conversation sont transformées en un vecteur qui se compose de 24 paramètres de sortie du réseau. Ces derniers peuvent être le taux de décharge moyen des neurones du réseau. En pratique, nous pouvons considérer qu'un vecteur composé de 24 paramètres est un «patron» qui représente l'enveloppe à l'entrée. Ainsi, nous pouvons effectuer l'identification du locuteur en utilisant ces patrons. Par la suite, nous effectuons des expériences sur l'identification du locuteur à l'aide d'un réseau DYSTAL. Notons que la tâche de reconnaissance ici est assez simple et n'importe quelle technique de reconnaissance peut être utilisée dans cette étape. Le réseau DYSTAL est choisi en raison de sa simplicité d'utilisation. L'identification du locuteur est réalisée en employant les paramètres du taux de décharge moyen.

Puisque chaque locuteur a enregistré 4 conversations, il y a donc 4 patrons correspondants (après le traitement par le réseau) que l'on peut nommer respectivement *patron 1*, *patron 2*, *patron 3* et *patron 4*. Afin de tester sur ces quatre patrons, nous effectuons une rotation de ces patrons pour former les données d'apprentissage et de test. Par exemple, la première expérience utilise le *patron 1* pour le test et les trois autres patrons pour l'apprentissage. La deuxième expérience utilise le *patron 2* pour le test et les trois autres pour l'apprentissage et ainsi de suite. Bref, chaque patron est choisi successivement pour le test et les trois autres sont choisis pour l'apprentissage.

Les expériences sur l'identification du locuteur portent sur dix locuteurs décrits précédemment. Les résultats obtenus sont présentés dans le tableau 4.2. En général, tous les locuteurs sont identifiés avec un taux de succès allant de 25% à 100%. Globalement, le taux de succès est de l'ordre de 57%. Les femmes (*sp1415*, *sp1096*, et *sp1007*) sont identifiées avec un taux de succès très élevé entre 75% et 100%. Par contre, les hommes sont identifiés avec un taux plus faible de 25% à 50%, sauf le locuteur *sp1436* (75%). Nous ne trouvons pas encore la raison pour laquelle les femmes sont identifiées avec un taux de succès beaucoup plus élevé que celui des hommes. Ne serait-il pas parce que les femmes ont la fréquence de glotte la plus élevée et qu'en conséquence le nombre d'événements synchronisés est plus grand ? La réponse serait très intéressante puisqu'elle peut nous permettre de savoir quelle information caractéristique du locuteur est extraite par le réseau proposé lors du traitement des enveloppes.

**Tableau 4.2** Résultats de l'identification du locuteur sur 10 locuteurs

No	Locuteur	Identification du locuteur	
		Succès / tests	Taux (%)
1	<i>sp1497</i>	2 / 4	50
2	<i>sp1499</i>	2 / 4	50
3	<i>sp1575</i>	2 / 4	50
4	<i>sp1415*</i>	3 / 4	75
5	<i>sp1436</i>	3 / 4	75
6	<i>sp1096*</i>	4 / 4	100
7	<i>sp1528</i>	1 / 4	25
8	<i>sp1007*</i>	3 / 3	100
9	<i>sp1130</i>	1 / 4	25
10	<i>sp1531</i>	1 / 4	25
<b>Moyenne</b>			<b>57.5</b>

• *Femme*



Dans le but de faire une comparaison de performance, nous utilisons des techniques contemporaines en traitement de la parole pour effectuer l'identification des locuteurs via leur fréquence de glotte. Nous traitons les mêmes locuteurs et les mêmes fichiers de données qu'aux expériences effectuées précédemment avec le réseau. L'apprentissage est effectué à partir des fichiers «training» de la base SPIDRE et l'identification à partir d'un fichier test par locuteur. Le fichier test utilise un combiné téléphonique différent de ceux utilisés lors de l'apprentissage. Notons que nous ne faisons pas dans ce cas une rotation successive des fichiers de données comme dans les expériences précédentes.

Nous générons de façon automatique la fréquence de glotte à l'aide d'un logiciel de suivi de la hauteur tonale (Rouat, 1997a). Nous estimons ensuite les histogrammes des fréquences de glotte pour chaque locuteur et pour chaque combiné. La moyenne et la variance sont calculées à partir des histogrammes en supposant que la distribution est gaussienne. De nouveaux histogrammes sont générés en intégrant sur l'axe des abscisses les fréquences glottiques au sein d'un intervalle («bin») dont la largeur correspond à 10% de la fréquence centrale de l'intervalle considéré. La plage de fréquence glottique comprise entre 60Hz et 600Hz est donc caractérisée par 24 intervalles («bins») à partir desquels on compte le nombre de fois qu'une fréquence de glotte se trouve dans l'un entre eux.

Les histogrammes sont utilisés pour effectuer la classification à l'aide du réseau DYSTAL<sup>1</sup>. Lorsque la reconnaissance se base sur le critère du maximum de corrélation (la sortie de DYSTAL donne le coefficient de

---

<sup>1</sup> Nous remercions M. Lapointe (ERMETIS, UQAC) pour la réalisation de ces expériences.

corrélation de Pearson), 3 locuteurs sont reconnus sur 10. Ce sont les locuteurs *sp1007*, *sp1528* et *sp1497*.

Avec les mêmes données apprentissage et test, l'identification du locuteur à l'aide du réseau proposé donne un taux de succès de 4/10. Les quatre locuteurs suivants sont reconnus: *sp1007*, *sp1096*, *sp1436* et *sp1499*. L'identification du locuteur basée sur les paramètres de sortie du réseau donne donc pour l'instant de meilleures performances vis-à-vis de celles basées sur la fréquence de glotte extraite à l'aide des techniques contemporaines en traitement de la parole.

Il est intéressant de souligner que les trois des quatre locuteurs reconnus à l'aide du réseau proposé sont différents de ceux reconnus à partir de la fréquence de glotte. Une combinaison des deux techniques serait vraisemblablement intéressante.

## 4.5 Discussion et conclusion

Nous avons présenté un prototype de système d'identification du locuteur. L'originalité du travail réside principalement dans l'utilisation du réseau à décharge comme outil d'analyse vis-à-vis des enveloppes à la sortie d'un banc de filtres cochléaires. Ce travail vise à utiliser directement les enveloppes des signaux du banc de filtres cochléaires sans avoir à définir *a priori* de paramètres spécifiques. Les enveloppes codent l'information liée à la parole dans un espace à représentation spatio-temporelle complexe et difficile à caractériser via les algorithmes couramment utilisés en reconnaissance de la parole. L'information est en général non stationnaire. Le traitement des enveloppes par le réseau est effectué simultanément pour tous les canaux du signal d'entrée en temps réel. Ce traitement permet non seulement d'extraire l'information résidant dans chaque canal, mais il permet aussi d'exploiter l'information liée à l'interaction entre les canaux. Il faut noter que ce traitement ne demande pas beaucoup de calculs par rapport à des traitements basés sur des techniques d'analyse de signal conventionnelles. En effet, les techniques couramment utilisées en identification du locuteur nécessitent un apprentissage long et fastidieux. Par conséquent, notre système peut être efficace pour l'application en temps réel.

L'expérience effectuée est préliminaire en ce sens que les enveloppes sont présentées directement au réseau qui reproduit de façon partielle le comportement de la couche IV du cortex auditif. L'approximation est en fait trop grossière (c'est un peu comme si on connectait directement l'oreille périphérique au cortex sans transiter par les noyaux nerveux intermédiaires). L'expérience, quoique grossière, montre que l'identification du locuteur basée

sur les paramètres extraits par le réseau proposé donne de meilleures performances vis-à-vis de celles basées sur la fréquence de glotte extraite à l'aide de techniques plus classiques en traitement de la parole.

Nous pensons qu'une étude de comparaison plus approfondie avec les méthodes conventionnelles est souhaitable. Pourtant, il est très difficile de dresser un bilan sur les performances des systèmes d'identification du locuteur proposés dans la littérature scientifique, car la durée de test et les bases de données utilisées sont la plupart du temps différentes (chiffres, mots isolés, séquences de mots, communications téléphoniques, parole propre, parole bruitée). Les systèmes actuels de reconnaissance du locuteur offrent une performance intéressante pour une population de 630 locuteurs (hommes et femmes), à condition que la parole soit propre et non corrompue par le bruit et les interférences (Reynolds, 1995). Le taux de succès est de 99.5% pour la base de données TIMIT (avec signaux de parole propre) et de 60.7% pour la base de données NTIMIT (mêmes signaux de la base TIMIT mais enregistrés sur la ligne téléphonique). Par contre, lorsque ces systèmes opèrent dans des conditions réelles, leur performance se dégrade fortement en raison de plusieurs sources de problèmes non résolus.

Reynolds (1996) a étudié les effets de la variabilité des combinés téléphoniques sur la performance de l'identification du locuteur. Les expériences ont été effectuées sur des signaux de parole de la base de données SPIDRE (rappelons que cette base est également utilisée dans notre travail). Il a obtenu un taux de succès de l'ordre de 55%. Les paramètres du signal de parole utilisés sont les coefficients Mel cepstraux et la distribution de ces paramètres est représentée par un modèle de mixture gaussienne.

De notre côté, nos résultats à ce stade ne sont pas au niveau de ceux de Reynolds. Pourtant, il faut rappeler que notre méthode n'est pas encore au point. En effet, les paramètres basés sur le taux de décharge ne sont probablement pas les meilleurs paramètres pour représenter les enveloppes à la sortie du banc de filtres cochléaires. De plus, la présente architecture du réseau proposé n'est pas convenablement adaptée aux traitements des canaux du banc de filtres cochléaires. Il devrait être intéressant dans un futur proche d'approfondir ce type d'approche. En effet, cette méthode est très rapide à mettre en œuvre en comparaison aux méthodes conventionnelles, par exemple celle utilisée dans le travail de Reynolds. Ceci est d'autant plus vrai que nos paramètres sont extraits implicitement par le réseau sans avoir à définir *a priori* de paramètres spécifiques.

La mise en place d'un prototype de système d'identification du locuteur permet d'évaluer le potentiel du réseau dans l'analyse des enveloppes. Ce type d'analyse nous paraît intéressant car il peut aider à une meilleure compréhension des paramètres liés à l'analyse spectro-temporelle des enveloppes. Ceci a en effet un grand intérêt pour l'application en reconnaissance de la parole, par exemple dans la détection automatique de la hauteur tonale (fréquence fondamentale) du signal de parole et plus généralement dans les systèmes de dialogue homme-machine automatiques.

## Conclusion et perspectives

**L**a plupart des techniques de traitement et de reconnaissance des formes supposent que le processus dynamique à traiter ou à reconnaître est stationnaire. Leur application à des problématiques où le temps est une composante essentielle est peu développée actuellement. Dans ce contexte, le but de la présente recherche était de concevoir et d'expérimenter un nouveau modèle de réseau de neurones approprié à la reconnaissance des processus spatio-temporels non stationnaires. Le travail est inspiré d'une part des observations en neurophysiologie, d'autre part des principes de la

dynamique non linéaire. L'originalité du travail réside principalement dans les propositions suivantes:

- ◆ Un modèle du neurone de type «Integrate-and-Fire» dont le seuil de décharge dépend pleinement de ses activités de décharge précédentes. Il y a donc une prise en compte de la relation temporelle au niveau local du neurone.
- ◆ Les règles d'apprentissage, qui permettent une régularisation de façon autonome de l'activité de décharge des neurones. Ceci permet donc de tenir compte de la relation spatio-temporelle entre les neurones dans le réseau au niveau global.
- ◆ Un nouveau critère pour la reconnaissance basé sur le temps de stabilisation du réseau au sens des systèmes dynamiques non linéaires. Ce critère permet d'éviter la tâche très difficile d'identification des attracteurs de formes complexes que l'on rencontre souvent dans la reconnaissance par réseaux dynamiques contemporains.

L'approche adoptée ici ne nécessite pas de supervision du réseau. Celui-ci est en mesure de détecter la nouveauté et de s'adapter de manière auto-organisée. De plus, il n'y a pas de différence entre apprentissage et reconnaissance.

Nous avons pu remplir les objectifs fixés en début de ce travail, c'est-à-dire réaliser un logiciel de simulation, effectuer les premières études et tester quelques applications de ce modèle de réseau. Expérimentalement, nous avons montré qu'il possède la capacité d'effectuer le traitement et la reconnaissance des processus spatio-temporels non stationnaires à travers les tâches telles que la reconnaissance des chiffres bruités, le traitement des séquences temporelles,

la détection de mouvement dans des séquences d'images, le traitement des enveloppes à la sortie d'un banc de filtres cochléaires permettant de réaliser un prototype de système d'identification du locuteur. Les expériences présentées sont relativement limitées et il y aura lieu d'approfondir le travail afin de bien connaître les limites du modèle du réseau et son potentiel en reconnaissance des processus spatio-temporels non stationnaires.

Le cadre de recherche que nous avons choisi pour cette thèse est relativement large. Nous n'avons aucunement la prétention de l'avoir totalement abordé. De nombreux travaux voire programmes de recherche peuvent être envisagés dans cette perspective. Au niveau du modèle de réseau, nous suggérons d'améliorer le présent modèle en s'inspirant des modèles de systèmes thalamo-corticaux tant au niveau architectural que neuronal en raison de leur capacité à traiter l'information spatio-temporelle. Au niveau des applications, on peut envisager dans l'utilisation en reconnaissance de trajectoires temporelles, en détection de la hauteur tonale de la parole (fréquence fondamentale), en détection de défauts dans des séquences de signaux, etc.

Il est à noter que notre travail est le résultat d'un esprit de recherche multidisciplinaire qui est encore peu adopté dans la recherche en ingénierie.



## BIBLIOGRAPHIE

1. ABELES M. (1982). Local Cortical Circuits: An Electrophysiological Study, Studies of Brain Function. Springer-Verlag.
2. ADACHI M. et al. (1993). Analysis of Associative Dynamics in a Chaotic Neural Network with External Stimulation. Proc. of IJCNN, Nagoya, Japan, Vol. 1, pp. 409-412.
3. ALKON D.L. et al. (1990). Pattern-Recognition by an Artificial Network Derived from Biologic Neuronal Systems. Biological Cybernetics, Vol. 62, pp. 363-376, Springer-Verlag.
4. AMIT D. J. (1989). Modeling Brain Function: the World of Attractor Neural Networks. Cambridge University Press.
5. BABLOYANTZ et al. (1994). Computation with Chaos: A Paradigm for Cortical Activity. Proc. of the National Academy of Science USA, Vol. 91, pp. 9027, Sept.
6. BARINAGA M. (1998). Listening In on the Brain. Science, Vol. 280., pp. 376-378.
7. BERSINI H. et al. (1994). Hopfield Net Generation, Encoding and Classification of Temporal Trajectories. IEEE Trans. on Neural Networks, Vol. 5, No. 6, Nov.
8. BERTILLE J. et Perez J.C. (1990). A Spatio-Temporal Novelty Using Fractal-Chaos Model. IJCNN Conf., Washington, Jan.
9. BLACKWELL et al. (1992). A new approach to hand-written character recognition. Pattern Recognition, Vol. 25, No. 6, pp. 655-666.
10. BOURRET P., REGGIA J., SAMUELIDES M. (1991). Réseaux Neuronaux, TEKNEA.
11. BROWN G. J. et WANG D.L. (1997). Modeling the perceptual segregation of double vowels with a network of neural oscillators. Neural Networks, 10, pp.1547-1558.

12. BROWN G. J. et al. (1996). Are Neural Oscillations the Substrate of Auditory Grouping ? Workshop on the Auditory Basis of Speech Perception, Keele University (UK), 174-179.
13. BRUNEL N. (1994). Dynamics of an attractor neural network converting temporal into spatial correlations. Network: Computation in Neural Systems, UK, 5, 449-470.
14. BRURKITT (1994). Attractor Neural Networks with Excitatory Neurons and Fast Inhibitory Interneurons at Low Spike Rates. Network: Computation in Neural Systems, UK, 5 (1994), 437-448.
15. BUONOMANO D.V. et al. (1995). Temporal Information Transformed into a Spatial Code by a Neural Network with Realistic Properties. Science, Feb., Vol. 267, pp. 1028-1030.
16. CAMPBELL S. et WANG D. (1998). Synchrony and Desynchrony in Integrate-and-Fire Oscillators. Proc. of the Intl' IJCNN98, Alaska, USA, pp. 1498-1503.
17. CARIANI P. (1995). As if time really mattered : Temporal strategies for neural coding of sensory information. CC-AI 12 (1-2).
18. CHAPPELIER J.C. (1996). RST: Une architecture connexionniste pour la prise en compte de relations spatiales et temporelles. Ph.D Thesis, Computer Science Department, TELECOM Paris.
19. DAVIS, J. et BOBICK, A. (1997). The representation and recognition of human movement using temporal templates. Computer Vision and Pattern Recognition, pp. 928-934.
20. DAYHOFF J.E. (1994). Short course on Neurodynamics of Temporal Processing, WCCN'94, San Diego, Jun.
21. DAYHOFF J.E. et al. (1994). Developing Multiple Attractors in a Recurrent Neural Network. Proc. of WCCN'94, San Diego, Jun., Vol. 4, pp. 710-715.
22. DEMUTH H. et BEALE M. (1996). Neural Network Toolbox for Use with MATLAB. The Math Works Inc.

23. DOYON B. et al. (1993). Control of the Transition to Chaos in Neural Networks with Random Connectivity. Int. Journal of Bifurcation and Chaos, Vol. 3, No. 2, pp. 279-291.
24. EBERHART R.C. et DOBBINS R.W. (1992). Neural Networks PC Tools: a practical guide. Academic Press. San Diego, USA.
25. ECKHORN R. et al. (1990). Feature linking via synchronization among distributed assemblies: simulations of results from cat visual cortex. Neural Computation. Vol. 2(3), pp. 293-307.
26. FERSTER D. et SPRUNSTON N. (1995). Cracking the Neuronal Code. Science, Vol. 270, Nov., pp. 74-75.
27. FREEMAN W. J. (1992a). Neural Networks and Chaos. INNS Above Threshold, Fall 1992, pp. 8-10.
28. FREEMAN W. J. (1992b). Tutorial on Neurobiology: from Single Neurons to Brain Chaos. Int. Journal of Bifurcation and Chaos, Vol.2, No. 3, pp. 451-482.
29. FREEMAN W. J.(1991). The Physiology of Perception. Scientific American, Feb., pp. 78-85.
30. GERSTNER W. (1998). Pulsed Neural Networks. Mass W. and Bishop C.M. (Editors), MIT press.
31. GLASS L. (1995). Chaos in Neural Systems. Handbook of Brain Research and Neural Networks. The MIT Press, pp. 186-192.
32. GODFREY J. et al. (1992). Switchboard: Telephone Speech Corpus for Research and Development. Proc. ICASSP '92, San Fransisco, pp. 517-520.
33. HAYASHI Y. (1994). Oscillatory Neural Networks and Learning of Continuously Transformed Patterns. Neural Networks, Vol. 7, No 2, pp.219-231.
34. HAYKIN S. (1994). Neural Networks – A Comprehensive Foundation. Macmillan College Publishing Company, Inc.
35. HERTZ J. et PRUGEL-BENNETTE A. (1996). Learning Short Synfire Chains by Self-Organization. Network: Computation in Neural Systems, UK, 7, 357-363.

36. HILL S. et VILLA A. (1997). Dynamic transitions in global network activity influenced by the balance of excitation and inhibition. Network: Computation in Neural Systems, UK, Vol. 8, 2, 165-184.
37. HILL S. et VILLA A. (1995). Global Spatiotemporal Activity Influenced by Local Kinetics in a Simulated "Cortical" Neural Network. Workshop on Supercomputing in Brain Research. World Scientific, 371-375.
38. HO T.V. et ROUAT J. (1998). Novelty Detection Based on Relaxation Time of a Network of Integrate-and-Fire Neurons. Proc. of the 1998 IEEE Intl' Joint Conference on Neural Networks, Alaska, USA, May, pp. 1524-1529.
39. HO T.V. et ROUAT J. (1997). Novelty Detector Using a Network of Integrate-and-Fire Neurons. Proc. of 7th Intl' Conference in Artificial Neural Networks, Lausanne, Switzerland, Oct., 103-108.
40. HO T.V. et ROUAT J. (1998). A spiking neural network for spatio-temporal detection. Submitted to IEEE Trans. on Neural Networks, Sept. 1998.
41. HOPFIELD J.J. (1995). Pattern Recognition Computation Using Action Potential Timing for Stimulus Representation. Nature, Jul., Vol. 376, pp. 33-36.
42. ISHII S. et SATO M. (1998). Associative memory based on parametrically coupled chaotic elements. To appear in Physica D.
43. KANENKO K. (1995). Cooperative Behavior in Networks of Chaotic Elements. Handbook of Brain Research and Neural Networks, The MIT Press, pp. 258-261.
44. KING R.S. et JULSTROM B. (1982). Applied Statistics using the computer. Alfred Publishing Co., Inc., USA.
45. MAASS W. (1996). Networks of Spiking Neurons: the Third Generation of Neural Network Models. Proc. of the 7<sup>th</sup> Australian Conf. On Neural Networks. Canberra, pp. 1-10.
46. MAASS W. (1997). Networks of Spiking Neurons Can Emulate Arbitrary Hopfield Nets in Temporal Coding. Network: Computation in Neural Systems, UK, Vol. 8(4), 355-372.

47. MATSUGU M. et YUILLE A.L. (1994). Spatiotemporal Information Storage in a Content Addressable Memory Using Realistic Neurons. Neural Networks, Vol. 7, No. 3, pp. 419-439.
48. MATSUNO T. et KIKKAWA Y. (1994). Periodic Signal Learning and Recognition in Couple Oscillators. Journal of The Physical Society of Japan, Vol. 63, No. 3, pp. 1194-204, Mar.
49. MERZ C.J. et MURPHY P.M. (1998). UCI Repository of machine learning databases. [<http://www.ics.uci.edu/mlearn/MLRepository.html>]. Irvine, CA: Univ. of California, Dep. of Information and Computer Science.
50. NAKANO T. et MORIYAMA O. (1994). Effects of temporary synaptic strengthening and residual cell potential in the retrieval of patterns. Network: Computation in Neural Systems, UK, 5, 229-239.
51. NUTZEL K. et al. (1994). Dynamics of Diluted Attractor Neural Networks with Delays. Biological Cybernetics, Vol. 70, pp. 553-561.
52. OSANA Y. et HAGIWARA K. (1998). Successive Learning in Chaotic Neural Network, Proc. of Intl' IJCNN 98, pp. 1510-1515, Alaska.
53. PALLBO R. (1993). Visual Motion Detection Based on a Cooperative Neural Network Architecture. Scandinavian conference on artificial intelligence '93, pp. 193-201. IOS Press.
54. PALMADESSO P. et DAYHOFF J. (1995). Attractor Locking in a Chaotic Network: Stimulus Patterns evoke Limit Cycles. Proc. of WCNN'95, Vol. 1, 254-257.
55. PATTERSON, R.D. et ALLERHAND M.H. (1995). Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. Journal Acoust. Soc. Amer., Vol. 98(4), pp. 1890-1894.
56. PEREZ J.C (1989). De Nouvelles Voies vers L'intelligence Artificielle Pluridisciplinaire. Editions Masson, Paris.

57. PRESS et al. (1992). Numerical Recipes in C. Cambridge University Press, Cambridge, 2nd edition.
58. REYNOLDS D.A. (1996). The effects of handset variability on speakers recognition performance : experiments on the switchboard corpus. IEEE-ICASSP'96, Vol. 1, pp. 113-117.
59. REYNOLDS D.A. et al. (1995). The Effects of Telephone Transmission Degradations on Speaker Recognition Performance. IEEE-ICASSP'95, Vol. 1, pp. 329-333.
60. RITZ R. et SEJNOWSKI T.J. (1997). Synchronous oscillatory activity in sensory systems : new vistas on mechanisms. Current Opinion in Neurobiology, 7 :536-546.
61. ROUAT J. et HO T.V. (1998). La détection de nouveauté basée sur le temps de stabilisation d'un réseau de neurones: application possible en reconnaissance de parole? . XXIIèmes Journées d'études sur la parole, Suisse, Jun, pp. 413-416.
62. ROUAT J. et al. (1997a). A pitch determination and voiced/unvoiced decision algorithm for noisy speech. Speech Communication Journal, Vol. 21 (3), 191-207.
63. ROUAT J. (1997b). Spatio-Temporal Pattern Recognition with Neural Networks: Application to Speech. Proc. of 7th Intl' Conference in Artificial Neural Networks, Lausanne, Switzerland, Oct., pp. 43-48.
64. RUF B. et SCHMITT M (1997). Learning temporally encoded patterns in networks of spiking neurons. Neural Processing Letters 5(1), pp. 9-18.
65. RUF B. et SCHMITT M (1998). Self-organization of spiking neurons using action potential timing. IEEE Transactions on Neural Networks, to appear.
66. SERRA R et ZANARINI G. (1990). Complex Systems and Cognitive Process. Springer-Verlag.
67. STASSINOPOULOS D. et BAK P. (1995). Democratic reinforcement: A principle for brain function. Physical Review E, Vol. 51, No. 5, 5033-5039.

68. STOECKER M. et al. (1995). A Neural Network for Scene Segmentation by Temporal Coding. Neurocomputing, 11, 123-134.
69. SZU H. (1994). Short course on Spatiotemporal Information Processing, INNS, WCNN-94, San Diego, Jun.
70. THIRAN P. et HASLER M. (1996). Information storage using stable and unstable oscillations: an overview, Int. Journal of Circuit Theory and Applications, Vol. 24,57-67.
71. TSUDA I. (1997). A new type of self-organization associated with chaotic dynamics in neural networks. Special Issue on the Role and Control of Random Events in Biological Systems, World Scientific Publishing Company, pp. 451-459.
72. TSUDA I (1992). Dynamic Link of Memory-Chaotic Memory Map in Nonequilibrium Neural Networks. Neural Networks, Vol. 5, pp. 313-326.
73. VILLA A. (1992a). Temporal Aspect of Information Processing in the Central Nervous System. Annales du Groupe CARNAC (Swiss Federal Institute of Technology), 5:15-42.
74. VILLA A. (1992b). Les catastrophes cachées du cerveau. Le nouveau Golem (Revue du Groupe Golem, réflexions et recherches outredisciplinaires), IDERIVE (Swiss).
75. WANG D. (1995a). Temporal Pattern Processing. Handbook of Brain Theory and Neural Networks. The MIT Press, pp. 967-970.
76. WANG D. et TERMAN D. (1995b). Locally Excitatory Globally Inhibitory Oscillatory Networks. IEEE Transactions on Neural Networks, 6: 283-286.
77. WANG D. (1993). Pattern Recognition: Neural Networks in Perspective. IEEE Expert, pp. 52-60.
78. YAO Y., FREEMAN W.J. et al. (1991). Pattern Recognition by a Distributed Neural Network: an Industrial Application. Neural Networks, Vol. 4, pp. 103-121.
79. YAO Y. et FREEMAN W.J. (1990). Model of Biological Pattern Recognition with Spatially Chaotic Dynamics. Neural Networks, Vol. 3, pp. 153-170.

80. ZAK M. (1991). An Unpredictable-Dynamic Approach to Neural Intelligence, IEEE Expert, Aug., pp. 4-10.



## ANNEXES

Articles de conférences internationales  
(sélectionnées sur la base de l'article complet)

- A. Ho T.V. et Rouat J. (1997). Novelty Detector Using a Network of Integrate-and-Fire Neurons. Proc. of 7th Intl' Conference in Artificial Neural Networks, Lausanne, Switzerland, Oct., 103-108.
- B. Ho T.V. et Rouat J. (1998). Novelty Detection Based on Relaxation Time of a Network of Integrate-and-Fire Neurons. Proc. of the 1998 IEEE Intl' Joint Conference on Neural Networks, Alaska, USA, May, pp. 1524-1529.
- C. Rouat J. et Ho T.V. (1998). La détection de nouveauté basée sur le temps de stabilisation d'un réseau de neurones: application possible en reconnaissance de parole? . XXIIèmes Journées d'études sur la parole, Suisse, Jun, pp. 413-416.

**Annexe A** (6 pages)

Ho T.V. et Rouat V. (1997). Novelty Detector Using a Network of Integrate-and-Fire Neurons. Proc. of 7th Intl' Conference in Artificial Neural Networks, Lausanne, Switserland, Oct. 1997, 103-108.

# A Novelty Detector Using a Network of Integrate-and-Fire Neurons

Tuong Vinh HO<sup>1,2</sup> and Jean ROUAT<sup>2</sup>

<sup>1</sup> Ecole Polytechnique de Montreal, Canada, Dept. of Génie Informatique

<sup>2</sup> Université du Québec à Chicoutimi, Canada, Dept. des Sciences Appliquées  
email: vho@uqac.quebec.ca Jean.Rouat@uqac.quebec.ca

**Abstract.** Information in the nervous system has often been considered as being represented by simultaneous discharge of a large set of neurons. We propose a learning mechanism for neural information processing in a simulated cortex model. Also, a new paradigm for pattern recognition by oscillatory neural networks is proposed. The relaxation time of the oscillatory networks is used as a criterion for novelty detection.

## 1 Introduction

Representation of information in the nervous system has often been considered as being contained in simultaneous discharge of a large set of neurons. How does a neural system use that kind of information representation while performing learning and pattern recognition? Recent studies on nonlinear cooperative complex dynamics in neural systems provided various kinds of models that described the cooperative behavior such as synchronization and chaos. In [6], Thiran and Hasler present a valuable overview on this approach. Hayashi [2] present an interesting characteristic of an oscillatory network: a limit cycle near a memory pattern (memory retrieval with ambiguous fluctuation) for an input closed to it, and a chaotic orbit wandering among memory patterns (autonomous search) for an input far from them. It is not easy to identify dynamical behavior. Stassinopoulos and Bak [5] propose a self-organizing model with a capability to interact with the surrounding environment. Self-organizing behavior arises by interaction between non-fixed threshold neurons and by feedback from environment. Although the model displays a rich dynamical behavior, it is still not clear how to associate patterns to the network's states. Dayhoff [1] proposes a learning mechanism that allows a Hopfield network having a rich dynamic behavior including fixed point, limit cycle and chaotic attractors. She also shows that the network can have many attractors and it overcomes the limitation of the original Hopfield model. However, we still do not know how to associate patterns to these behaviors, in other words, how to apply this network model for recognition problems. In fact, we need a way to manipulate the chaotic behaviors.

Hill and Villa [3] developed neural models to study the spatiotemporal pattern generation properties in a simulated "cortical neural network". The model

uses integrate-and-fire neurons as elementary units. Furthermore, the topology is inspired from that of layer IV in the cortex. Although this model helps to observe the evolution of spatiotemporally organized activity in a simulated cortex, the learning rule is not yet proposed. In this paper, we propose a learning mechanism for neural information processing in the simulated cortex model. Along with the learning mechanism, we propose a new paradigm for pattern recognition by oscillatory neural networks. The relaxation time of oscillatory behavior was used as a criterion for novelty detection.

## 2 Neuronal Model

Our neuron model was inspired from the integrate-and-fire neuronal model proposed by Hill and Villa [3] with refractory period and post-synaptic potential decay. The state of the neuron at time  $t$ , is deterministically modeled by a control potential,  $U$  as:

$$S_i(t) = \begin{cases} 0 & \text{if } (t - t_{spike}) < \rho, \\ \mathcal{H}[U_i(t) - \theta] & \text{otherwise,} \end{cases} \quad (1)$$

where  $\mathcal{H}$  is the Heaviside function defined as  $\mathcal{H}[x] = 1$  for  $x > 0$  otherwise  $\mathcal{H}[x] = 0$ . The value  $t_{spike}$  represents the last firing time for unit  $i$ . The value  $\rho$  denotes the absolute refractory period. Refractoriness corresponds to the period following the production of a spike or action potential, during which the cellular biochemical mechanisms cannot generate another signal, regardless of the strength of the stimulation. The control potential is defined as the integration of all afferent postsynaptic potential at time  $t$ :

$$U_i(t+1) = \sum_i C_{ij} S_j(t) + U_i(t) + s_i \quad (2)$$

where the indices  $i$  and  $j$  indicate the units,  $C$  is the connection strength, and  $s$  is the input signal.

In order to introduce the influence of the firing frequency into the neuron's behavior, we added a variable firing frequency factor  $f$  to the neuron model. Simulating experiments showed that this factor has a strong influence in the neuron's behavior. Thus, the equation (2) becomes

$$U_i(t+1) = \sum_i C_{ij} S_j(t) + U_i(t) + s_i + f_i. \quad (3)$$

## 3 Network Architecture

The network architecture was inspired from an oversimplified model of cortical layer IV [3]. This model defines a single two-dimensional sheet of excitatory and inhibitory neurons with recurrent connections. The layer consists of two populations of neurons interspersed within the plane. These neurons are positioned

according to a space-filling pseudo-random Sobol distribution. Each neuron has a set of interconnections chosen according to a square neighborhood, centered at the neuron itself and with a radius depending on whether the neuron is excitatory or inhibitory. From this topology, we can say that this model uses an interactivity at local level to create a self-organizing evolution. Here, we want to modify the model by introducing an interactivity at global level with a global inhibitor (Fig. 1). By this approach, we can create an interactivity between all neurons in the network. The global inhibitor is actually a trigger whose state is either active, i.e. firing, or inactive depending on a control mechanism. The control mechanism is based on a threshold for the total number of firing neurons. Whenever the number of firing neuron at a time  $t$  is above the threshold, the global inhibitor fires and it generates a negative feedback signal to every neuron in the network. Otherwise, if the number of firing neurons is below the threshold, the global inhibitor generates a positive feedback signal. Thus, the global inhibitor plays a role of regulating neuron activity at global level. With a feedback signal  $h$ , the equation (3) becomes:

$$U_i(t+1) = \sum_i C_{ij} S_j(t) + U_i(t) + s_i + f_i + h(t). \quad (4)$$

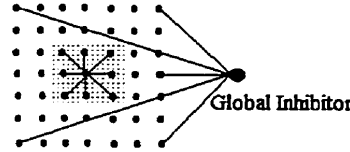


Fig. 1. The architecture of the neural network

#### 4 Learning Rule

The learning rule, that modifies the coupling strengths, is widely used in many kinds of neural network models. In our model, there are two kinds of coupling: one is between neighboring neurons and the other is between the global inhibitor and all neurons. The Hebbian rule was used as updating rule for the coupling weights. The following equation is used for the coupling weights between a pair of neurons:

$$C_{ij}(t+1) = C_{ij}(t) + \alpha C_{ij}(1 - C_{ij}) S_i(t) S_j(t) \quad (5)$$

where  $\alpha$  is the learning rate. The coupling weights between each neuron and the global inhibitor are updated according to:

$$G_i(t+1) = G_i(t) + \beta G_i(t)(1 - G_i(t)) S_i(t) h(t) \quad (6)$$

where  $G_i$  is the coupling weight of the neuron  $i$ ,  $\beta$  is the update rate and  $h$  is the feedback signal from the global inhibitor.

Learning phase starts when the network is stimulated by an input signal. The network begins to oscillate. At each instant, the coupling weights of a neuron are updated if this neuron fires. It receives also a feedback signal. The later is either negative or positive depending on if the number of firing neuron is above or below the threshold of firing neurons. The network is considered to reach a stable state when its local coupling weights do not change anymore or they change in a very small given range. When the network reaches a stable state, the learning phase terminates.

## 5 Novelty Detection by this Model

The proposed network model is a non linear dynamical system. How can the evolution of dynamical systems be associated with the execution of cognitive tasks ? We need to find a paradigm that can be used to characterize dynamical evolution inside the system so that it can be applied to pattern recognition. We observed from simulating experiments, we have observed that the dynamical network can reach a stable state very quickly if the input signal has already been seen. From this observation, we propose a new paradigm for novelty detection by this network model. The paradigm is comprised of two phases:

+ *Learning phase*: the network with randomly initialized connection strengths is trained by training patterns. It reaches an equilibrium state after learning.

+ *Novelty detection phase*: patterns are introduced to the trained network. The network reaches an equilibrium state after a relatively small number of iterations if these patterns have been learned before. Otherwise, it takes a long time for the network to reach an equilibrium state. Based on the relaxation time, novelty detection can be done by our neural network model.

In the following, we show an example of this paradigm to novelty detection by our neural network model. A set of 0-9 digits is used as a pattern set in this paper. As seen in Fig. 2, each digit pattern is coded by using a 7x5 binary pixel matrix. In order to test the robustness of the network, a set of noisy patterns obtained from the original patterns is also used. The noisy patterns are created by adding a certain amount of noise to the original pattern images [Fig 2., right]. In other words, given an amount of noise (by percentage), a number of pixels in the patterns are changed. The pixels are randomly chosen with a uniform probability distribution. As in this experiment, with a 20% of noise, 7 pixels in each 7x5 pixel image have their value changed. Though the pattern images used herein are binary images, our network can manipulate analog images (i.e. real numbers can be manipulated by the network). For the experiments in this paper, we used a 7x7 dimension network with 70% population being excitatory neurons and the remaining 30% being inhibitory neurons. The pattern image is positioned at the center of the network plane.

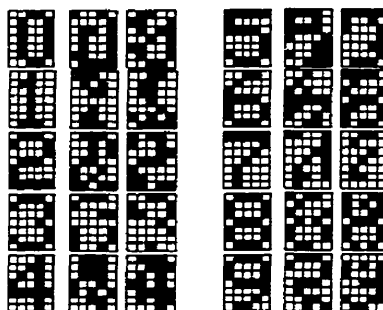


Fig. 2. Patterns of 0-9 digits with 20% noise

The digit patterns from 0 to 4 are used to train the network. The pattern is presented sequentially into the network. The network oscillates and reaches an equilibrium state. When the network reaches a stable state, a new pattern is fed into it. In this experiment, the set of 5 patterns (0-4 digits) is presented to the network only one time. Oscillation times of patterns 0-4 are 351, 290, 321, 11, 307 iterations respectively. Note that oscillation time of the network is often dependent on the sequence of training data. After learning phase, we use either the noisy versions of learned patterns or a set of "never seen" patterns (5-9 digits) to test the ability of novelty detection of the network. According to the proposed paradigm, we use the relaxation time (in term of number of iterations) of the network during testing phase as a criterion to decide whether a pattern is "seen" or "never seen" by the network. A short relaxation time means that the pattern has been seen before. Otherwise, the pattern has never been seen before. Table 1 shows that the network has a short relaxation time (11 iterations) when the testing patterns are either the learned patterns or the noisy patterns of the learned patterns. In contrast, the network has a significant long relaxation time (271 or 162 iterations) when the testing patterns have never seen before. The network made recognition mistakes on 3 patterns (noisy version 2 of patterns 3, 7 and noisy version 1 of pattern 9), i.e. with an error rate of 10% (3/30).

In order to examine the network's performance, another test based on the initial training on digit [5-9] and testing on digits [0-4] was also done. During training, oscillation times of patterns 5-9 are 183, 132, 148 11 and 11 iterations respectively. The testing result is given in Table 2. The network made more recognition mistakes than previous test with an error rate of 23% (7/30). Ongoing works are focuses to improve recognition performance of the network. In addition, theoretical analysis is left as a future work.

## 6 Conclusion

A new paradigm for pattern recognition by non linear systems is proposed in this study. This paradigm is based on a criterion that is the time of oscillation of the network when a pattern is injected into it. In other words, the relaxation time is

Table 1. Relaxation time of the network trained on digits [0-4] and tested on [0-9]

Patterns	0	1	2	3	4	5	6	7	8	9
Original version	11	11	11	11	11	271	271	162	271	271
Noisy version 1	11	11	11	11	11	271	271	162	271	11
Noisy version 2	11	11	11	271	11	271	271	11	271	271

Table 2. Relaxation time of the network trained on digits [5-9] and tested on [0-9]

Patterns	0	1	2	3	4	5	6	7	8	9
Original version	11	152	153	11	143	11	11	11	11	11
Noisy version 1	153	152	153	11	152	153	11	11	11	153
Noisy version 2	11	153	153	11	143	11	11	11	11	11

used to decide whether a pattern has ever been seen before. A short relaxation time implies that the pattern has been already seen. Otherwise, a long relaxation time implies that the pattern has never been seen. This paradigm allows us to develop novelty detection systems based on the proposed network model with capability against noise as well as spatiotemporal transformation.

### Acknowledgments

This work has been supported by the NSERC of Canada, by the "Fondation" from Université du Québec à Chicoutimi. We would like to thank Alessandro Villa and Sean Hill (Université de Lausanne, Swiss) for their cooperative discussion concerning this work.

### References

1. Dayhoff J.E. et al.: "Developing Multiple Attractors in a Recurrent Neural Networks", *Proc. of WCCN'94*, San Diego, Jun. 1994, Vol. 4, pp. 710-715.
2. Yukio Hayashi: "Oscillatory Neural Networks and Learning of Continuously Transformed Patterns", *Neural Networks*, 1994, Vol. 7, No 2, pp. 219-231.
3. Hill S., Villa A.: "Global Spatiotemporal Activity Influenced by Local Kinetics in a Simulated "Cortical" Neural Network, *Workshop on Supercomputing in Brain Research: from topography to neural networks*, 1995, World Scientific, pp. 371-375.
4. Matsuno Tet al.: "Periodic Signal Learning and Recognition in Coupled Oscillators", *Journal of Physical Society of Japan*, Vol. 63, No. 3, March, 1994, pp. 1194-1204.
5. Stassinopoulos D., Bak P.: "Self-organization in a Simple Brain Model", *Proc. of WCCN'94*, San Diego, Jun, 1994, Vol. 1, pp. 4-26.
6. Thiran P., Hasler M.: "Information storage using stable and unstable oscillations: an overview", *Int. Journal of Circuit Theory and Applications*, Vol. 24, 57-67, 1996.



**Annexe B** (6 pages)

Ho T.V. et Rouat J. (1998). Novelty Detection Based on Relaxation Time of a Network of Integrate-and-Fire Neurons. Proc. of the 1998 IEEE Intl' Joint Conference on Neural Networks, Alaska, USA, May 1998, pp. 1524-1529.

# Novelty Detection Based on Relaxation Time of a Network of Integrate-and-Fire Neurons

Tuong Vinh HO<sup>1,2</sup> and Jean ROUAT<sup>1</sup>

<sup>1</sup> Université du Québec à Chicoutimi, Canada, Dept. des Sciences Appliquées

<sup>2</sup> Ecole Polytechnique de Montréal, Canada, Dept. de Génie Informatique  
email: vho@uqac.quebec.ca Jean\_Rouat@uqac.quebec.ca

## Abstract

*We propose a neural network model inspired from a simulated cortex model. Also, a new paradigm for pattern recognition by oscillatory neural networks is presented. The relaxation time of the oscillatory networks is used as a criterion for novelty detection. We compare the proposed Neural Network with Hopfield and back-propagation networks for a noisy digit recognition task. It is shown that the proposed network is more robust. This work could be a possible bridge between nonlinear dynamical systems and cognitive processes.*

## 1 INTRODUCTION

The processing or the recognition of non stationary noisy process with Neural Networks is a challenging and yet unsolved issue. Most of the contemporary pattern processing or recognition techniques assume that the pattern or the time series to be recognized are stationary. Furthermore, in real life applications, the information is most of the time corrupted, partial or noisy (image, speech, etc.). Therefore, the pattern recognizers have also to be robust.

Among the many neural network works that are reported in the literature, we can find neural networks with complex behavior or dynamics. These networks are sometimes called 'chaotic neural networks' by some authors. We are interested in evaluating neural networks with complex dynamics as potential recognizer systems of non stationary noisy processes. In the present work we propose an oscillatory network and we evaluate here the system on a noisy limited task.

Representation of information in the nervous system has often been considered as being contained in simultaneous discharges of a large set of neurons. How does a neural system use that kind of information representation while performing learning and pattern recognition? Recent studies on nonlinear cooperative complex dynamics in neural systems provide various kinds of

models that describe the cooperative behavior such as synchronization and chaos (Buonomano and Merzenich (1995), Destexhe (1994), Brown et al. (1996), Brunel (1994), Burkitt (1994), Matsuno (1994), Wang (1995)). Especially, Thiran and Hasler (1996) give a valuable overview on some principles for information storage and retrieval based on oscillations in dynamical systems.

Synfire chains have been proposed as a mechanism for neural information processing in the cortex by Abeles (1982). Hertz and Prugel-Bennette (1996) investigated whether synfire chains can be formed through a biologically plausible self-organizing mechanism. They proposed a network model of cortical neurons capable of learning synfire chains by introducing a Hebbian learning mechanism. However, this type of network is unstable against the formation of long synfire chains. Hill and Villa (1995, 1997) developed neural models to study the spatiotemporal pattern generation properties in a simulated "cortical neural network". The model uses integrate-and-fire neurons as elementary units. Furthermore, the topology is inspired from that of layer IV in the cortex. Although this model helps to observe the evolution of spatiotemporally organized activity in a simulated cortex, the learning rule is not yet proposed.

In this paper, we propose a neural network model that allows to study neural information processing in the cortex. The network model has the architecture inspired from that of layer IV in the cortex. Learning is based on a rewarding feedback mechanism. The system dynamics and the self-organizing process exhibit robustness against highly noisy input patterns. Along with the neural network model, we present a new paradigm for pattern recognition by oscillatory neural networks. The relaxation time of oscillatory behavior is used as a criterion for novelty detection. When input is similar to one of learned patterns, the network takes a very short time to go to an equilibrium state. In contrast, when input is different from any learned

patterns, the network takes a long time to go to an equilibrium state.

## 2 NEURONAL MODEL

Our neuron model is inspired from the integrate-and-fire neuronal model proposed by Hill and Villa (1995) with refractory period and post-synaptic potential decay. The state of the neuron at time  $t$ , is deterministically modeled by a control potential,  $U$  as:

$$S_i(t) = \begin{cases} 0 & \text{if } (t - t_{spike}) < \rho, \\ \mathcal{H}[U_i(t) - \theta] & \text{otherwise,} \end{cases} \quad (1)$$

where  $\mathcal{H}$  is the Heaviside function defined as  $\mathcal{H}[x] = 1$  for  $x > 0$  otherwise  $\mathcal{H}[x] = 0$ . The value  $t_{spike}$  represents the last firing time for unit  $i$ . The value  $\rho$  denotes the absolute refractory period. Refractoriness corresponds to the period following the production of a spike or action potential, during which the cellular biochemical mechanisms cannot generate another signal, regardless of the strength of the stimulation. In order to simplify the network's behaviour, the refractory period of every neuron is the same. It means that for every neuron, the time for which it can fire is the same.

The control potential is defined as the integration of all afferent postsynaptic potentials at time  $t$ :

$$U_i(t+1) = \sum_j C_{ij} S_j(t) + U_i(t) + s_i \quad (2)$$

where the indices  $i$  and  $j$  indicate the units,  $C_{ij}$  is the connection strength, and  $s_i$  is the input signal.

In order to introduce the influence of the firing frequency into the neuron's behavior, we add a variable firing frequency factor  $f$  to the neuron model. Simulating experiments showed that this factor has a strong influence in the neuron's behavior. Thus, equation (2) becomes

$$U_i(t+1) = \sum_j C_{ij} S_j(t) + U_i(t) + s_i + f_i. \quad (3)$$

Self-reference at local level is also used in our network model to increase the potential of self-organization of the network. The input signal is maintained until the network reaches a stable state. However, one can construct systems where the input signal is presented to the system at a short instant then disappears. In our model, the presenting time of an input signal is a variable parameter. It is demonstrated by experiments that the network is more robust when the input signal is maintained during a reasonable time.

Also, the fact that input signal is maintained creates an interference between the input signal and all dynamical states of the network. It is somewhat a temporal summation at local level.

## 3 LEARNING WITH REWARDING FEEDBACK MECHANISM

The idea is inspired from the work of Stassinopoulos and Bak (1994) by which a global feedback signal is used as a rewarding feedback mechanism to stabilize the self-organizing activity of the network. In a fashion analogous to the behaviorist techniques used in the training of animal, the network is introduced with a set of external signals each of which rewards a specific action. The system learns to recognize all signals and choose the corresponding rewarding action. Learning and retrieving are two aspects of the same dynamical process. In the following, we apply this learning mechanism to the our neural network model.

The network architecture is inspired from an oversimplified model of cortical layer IV (Hill & Villa (1995, 1997)). This model defines a single two-dimensional sheet of excitatory and inhibitory neurons with recurrent connections. The layer consists of two populations of neurons interspersed within the plane. These neurons are positioned according to a space-filling pseudo-random Sobol distribution. Each neuron has a set of interconnections chosen according to a square neighborhood, centered at the neuron itself. Excitatory and inhibitory neurons can have different neighboring radius. The rewarding feedback mechanism is implemented based on the firing activity of a set of neurons that are defined as output neurons. Output neurons are randomly chosen from the population of excitatory neurons (Figure 1). The output signal is the firing state of the set of output neurons. For each input signal, the network's action is considered successful if one or more neurons belonging to the set of output neurons are firing. If the network's action is successful, the feedback signal is positive. Thus, every neuron who is firing is reinforced with a positive feedback signal (+1). Otherwise, if the action is unsuccessful the feedback signal is negative and every firing neuron is reinforced with a negative feedback signal (-1).

In other words, if the network's action, stimulated by an input signal, is successful, all connections of firing neurons are reinforced, whether or not they participated in creating a successful action; if the action is unsuccessful, the connections of firing neurons are weakened. For updating the connection weights, the Hebbian updating rule is applied in this network model.

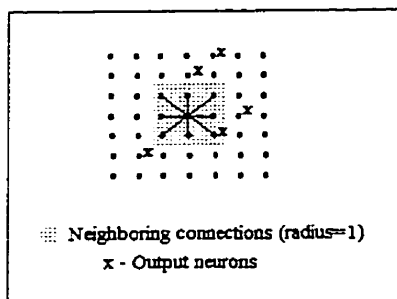


Figure 1. The architecture of the neural network.

Let  $h$  be the feedback signal ( $h = +1$  or  $-1$ ), the following equation is used to update the connection weights between a pair of neurons:

$$C_{ij}(t+1) = C_{ij}(t) + \alpha C_{ij}(1 - C_{ij})S_i(t)S_j(t)h(t) \quad (4)$$

where  $\alpha$  is the learning rate.

In our work, we try to integrate the dynamics of firing activity of neurons and the interactivity between neurons. In fact, the input of each neuron is not only a spatial combination but implicitly a temporal combination of firing activity too. Along with a combination created by the feedback control signal, the network model has a spatiotemporally complex combination of signals. We hopefully think that the complex dynamics of the network can help to catch and to manipulate the dynamics of input signals.

## 4 NOVELTY DETECTION

### 4.1 Novelty detection based on relaxation time

The proposed network model is a non linear dynamical system. How can the evolution of non linear dynamical systems be associated with the execution of cognitive tasks? We need to find a paradigm that can be used to characterize dynamical evolution inside the systems so that they can be applied to pattern recognition. We observed from simulating experiments that the dynamical network can reach a stable state very quickly if the input signal has already been seen. From this observation, we have proposed a new paradigm for novelty detection by oscillating network models (Ho & Rouat (1997)). The paradigm is comprised of two phases:

+ *Learning phase*: the network with randomly initialized connection strengths is trained with learning patterns. It reaches an equilibrium state after learning.

+ *Novelty detection phase*: patterns are introduced to the trained network. The network reaches an equilibrium state after a relatively small number of iterations if these patterns have been learned before. Otherwise, it takes a long time for the network to reach an equilibrium state. Based on the relaxation time, novelty detection can be done by our neural network model.

In the following, we present examples of this paradigm to novelty detection by our neural network model. A set of 0-9 digits is used as a pattern set in this paper. As seen in Figure 2, each digit pattern is coded by using a 7x5 binary pixel matrix. In order to test the robustness of the network, a set of noisy patterns obtained from the original patterns is also used. The noisy patterns are created by adding a certain amount of noise to the original pattern images [Figure 2., right]. In other words, given an amount of noise (by percentage), a number of pixels in the patterns are changed. The pixels are randomly chosen with a uniform probability distribution. As in this experiment, with a 20% of noise, 7 pixels in each 7x5 pixel image have their value changed. Though the pattern images used herein are binary images, our network can manipulate analog images (i.e. real numbers can be manipulated by the network).

For the experiments in this paper, we used a 7x7 dimension network with 70% population being excitatory neurons and the remaining 30% being inhibitory neurons. The neighborhood radius is 2 for excitatory neurons and 1 for inhibitory neurons. The pattern image is positioned at the center of the network plane. Every neuron which is covered by the pattern image will receive input signal, even if it had been chosen as output neuron. Other neurons which are not covered by the pattern image will not receive any input signal (It means that input signal for these neurons is set to 0).

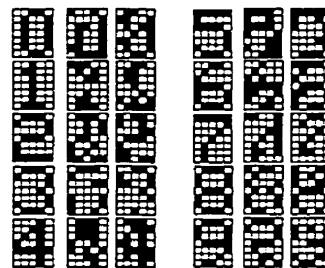


Figure 2. Patterns of 0-9 digits with 20% noise.

## 4.2 Experiment 1: Learning digit patterns [0-4]

### 4.2.1 Novelty detection by the network model

The digit patterns from 0 to 4 are used to train the network. Each pattern is presented sequentially to the network. The network oscillates and reaches an equilibrium state. When the network reaches a stable state, a new pattern is fed into it. In this experiment, the set of 5 patterns (0-4 digits) is presented to the network only one time. Oscillation times of the network stimulated by the patterns 0-4 are 332, 266, 347, 11 and 307 iterations respectively.

After learning phase, we use either the noisy versions of learning patterns or a set of "never seen" patterns (5-9 digits) to test the ability of novelty detection of the network. According to the proposed paradigm, we use the relaxation time (in term of number of iterations) of the network during testing phase as a criterion to decide whether a pattern has been "seen" or "never seen" by the network. A short relaxation time means that the pattern has been seen before. Otherwise, the pattern has never been seen before. Table 1 shows that the network has a short relaxation time (11 iterations) when the testing patterns are either the learned patterns or the noisy patterns of the learned pattern. In contrast, the network has a significant long relaxation time (271 or 170 iterations) when the testing patterns have never been seen before. The network made recognition mistakes on 3 patterns (noisy version 2 of patterns 3, 7 and noisy version 1 of pattern 9), i.e. with an error rate of 10% (3/30). Note that the network used in this experiment has a 7x7 dimension. It seems here that it can memorize about 5 patterns, i.e.  $0.1N$  where  $N$  is the number of neurons of the network.

Table 1. Relaxation time of the proposed network trained on 'clean' digits [0-4] and tested on [0-9].

Patterns	0	1	2	3	4	5	6	7	8	9
Original version	11	11	11	11	11	271	271	170	271	271
Noisy version 1	11	11	11	11	11	271	271	170	271	11
Noisy version 2	11	11	11	271	11	271	271	11	271	271

### 4.2.2 Comparison with a Hopfield network

Hopfield networks are interesting from a theoretical standpoint and can be used for classification. Therefore, comparing with them may give readers a view

about the data set as well as about our model. We use the Hopfield network implemented in the MATLAB package (Demuth and Beale (1996)). Hopfield networks can act as vector categorization networks. Input vector are used as the initial conditions to the network, which recurrently updates until it reaches a stable output vector. This type of network may be used to store the exemplars or training patterns.

We use the same training pattern set (digits [0-4]) and testing pattern set (digit [0-9]) as in the previous test. We tried the Hopfield network with several dimensions: 35, 49 and 81 neurons. The network has an error rate of 27% (8/30), 23% (7/30) and 23% (7/30) respectively. We report in Table 2 the recognition result of the Hopfield network with 49 neurons. Indeed, this network has the same dimension as our network model.

Table 2. Novelty detection using a Hopfield network trained on 'clean' digits [0-4] and tested on digits [0-9].

Patterns	0	1	2	3	4	5	6	7	8	9
Original version	y	y	y	y	y	n	y	y	y	y
Noisy version 1	y	y	y	y	n	y	y	y	y	n
Noisy version 2	y	y	y	n	n	n	y	y	y	n

'y' indicates that the pattern is correctly classified by the network. Otherwise, 'n' indicates that the pattern is not correctly classified. The unlearned patterns are considered being correctly classified if their corresponding recall patterns are different from the learned patterns.

### 4.2.3 Comparison with a backpropagation network

Multilayer networks with the backpropagation learning are most widely used as pattern recognizers in the field. This comparison might let readers gain further understanding about our model and the nature of pattern data in this paper. Multilayer networks under the standard generalized delta rule with momentum (Eberhart & Dobbins (1992)) are used for our experiments.

The same training [0-4] and testing [0-9] pattern sets are used with backpropagation networks. Regarding the data set, a network with one input layer (35 neurons), one hidden layer and one output layer (5 neurons) is chosen. We tried the network with the hidden layer having 3, 5 and 10 neurons respectively. The network has an error rate of 30% (9/30), 27% (8/30) and 30% (9/30). Table 3 presents novelty detection result

by the backpropagation network with the hidden layer comprising 5 neurons.

Table 3. Novelty detection using a backpropagation network trained on 'clean' digits [0-4] and tested on digits [0-9].

Patterns	0	1	2	3	4	5	6	7	8	9
Original version	y	y	y	y	y	n	y	y	n	y
Noisy version 1	y	y	y	y	y	y	y	y	n	n
Noisy version 2	y	y	y	y	y	n	y	n	n	n

'y' indicates that the pattern is correctly classified by the network. Otherwise, 'n' indicates that the pattern is not correctly classified.

Based on the experiment on learning 'clean' digit patterns [0-4], we conclude that our model made the smallest error rate of 10% while the Hopfield and backpropagation network made the error rate of 23% and 30% respectively.

### 4.3 Experiment 2: Learning digit patterns [5-9]

#### 4.3.1 Novelty detection by the network model

One of the weak points of neural networks is that their recognition performance is effectively changed when they are trained by different input sets. In order to examine our network model, other test based on the initial training on digits [5-9] and testing on digits [0-9] was also performed. Simulation conditions are the same as in previous test. Oscillation times of the network stimulated by these patterns are 223, 340, 368, 313 and 11 iterations respectively during 'training'. Results are reported in Table 4. The network has more recognition mistakes than previous test with an error rate of 30% (9/30).

Table 4. Relaxation time of the proposed network trained on 'clean' digits [5-9] and tested on [0-9].

Patterns	0	1	2	3	4	5	6	7	8	9
Original version	11	11	286	11	168	11	11	11	11	11
Noisy version 1	281	168	286	169	168	281	11	11	11	281
Noisy version 2	11	281	286	11	168	168	169	11	11	11

#### 4.3.2 Comparison with a Hopfield network

We also performed novelty detection using the Hopfield network with several dimensions: 35, 49 and 81 neurons. The network has an error rate of 40% (12/30), 43% (13/30) and 40% (12/30) respectively. Novelty detection results using the Hopfield network with 49 neurons are reported in Table 5.

Table 5. Novelty detection using a Hopfield network trained on 'clean' digits [0-4] and tested on digits [0-9].

Patterns	0	1	2	3	4	5	6	7	8	9
Original version	n	y	n	n	y	y	y	y	y	y
Noisy version 1	n	y	n	y	y	y	n	y	n	n
Noisy version 2	y	y	y	n	y	y	n	n	n	n

'y' indicates that the pattern is correctly classified by the network. Otherwise, 'n' indicates that the pattern is not correctly classified. The unlearned patterns are considered being correctly classified if their corresponding recall patterns are different from the learned patterns.

#### 4.3.3 Comparison with a backpropagation network

We performed novelty detection using the network with the hidden layer comprising 3, 5 and 10 neurons respectively. The network made an error rate of 33% (10/30), 30% (9/30) and 37% (11/30). Novelty detection results using the network with the hidden layer comprising 5 neurons are reported in Table 6.

Table 6. Novelty detection using a backpropagation network trained on 'clean' digits [5-9] and tested on digits [0-9].

Patterns	0	1	2	3	4	5	6	7	8	9
Original version	y	y	n	n	y	y	y	y	y	y
Noisy version 1	y	n	n	y	y	y	y	y	n	y
Noisy version 2	n	y	y	n	y	y	n	n	y	y

'y' indicates that the pattern is correctly classified by the network. Otherwise, 'n' indicates that the pattern is not correctly classified.

With the experiment on learning 'clean' digit patterns [5-9], our network model has the same error rate of 30% as the backpropagation network while the Hopfield network has the error rate of 40%.

## 5 CONCLUSION

Our neural network model is a result of an attempt to associate non linear dynamical systems with neural networks in order to treat spatiotemporal patterns. A new paradigm for pattern recognition by non linear systems is presented in this study. This paradigm is based on a criterion that is the time of oscillation of the network when a pattern is introduced into it. In other words, the relaxation time is used to decide whether a pattern has ever been seen before. A short relaxation time implies that the pattern has been already seen. Otherwise, a long relaxation time implies that the pattern has never been seen. This paradigm allows us to develop novelty detection systems based on the proposed neural network model with capability against noise as well as spatiotemporal transformation. However, theoretical analysis is still left as a future work. Ongoing works are investigating the use of this kind of network model for recognition of temporal sequences. Preliminary experiments show that the network has ability to perform recognition of temporal sequences.

## ACKNOWLEDGMENTS

This work has been supported by the NSERC of Canada and the "Fondation" from Université du Québec à Chicoutimi, Canada.

## REFERENCES

- [1] Abeles M. (1982). Local Cortical Circuits. An Electrophysiological Study. Spring-Verlag.
- [2] Buonomano Dean V., & Merzenich Michael M. (1995). Temporal Information Transformed into a Spatial Code by a Neural Network with Realistic Properties. *Science*, Vol. 267, 1028-1030.
- [3] Brown Guy J. et al. (1996). Are Neural Oscillations the Substrate of Auditory Grouping ? . *Proc. of the Workshop on the Auditory Basis of Speech Perception*, Keele University (UK), Worth Printing Ltd., ISSN 1018-4554, pp. 174-179.
- [4] Burkitt Anthony N. (1994). Attractor Neural Networks with Excitatory Neurons and Fast Inhibitory Interneurons at Low Spike Rates. *Network: Computation in Neural Systems*, UK, 5, 437-448.
- [5] Brunel Nicolas (1994). Dynamics of an attractor neural network converting temporal into spatial correlations. *Network: Computation in Neural Systems*, UK, 5, 449-470.
- [6] Demuth H. & Beale M. (1996). Neural Network Toolbox for Use with MATLAB, *The Math Works Inc.*
- [7] Destexhe Alain (1994). Oscillations, complex spatiotemporal behavior and information transport in networks of excitatory and inhibitory neurons. *Physical Review E*, Vol. 50, No. 2.
- [8] R. C. Eberhart, R. W. Dobbins. (1992). Neural Networks PC Tools: a practical guide. Academic Press. San Diego, USA.
- [9] Hertz .J, Prugel-Bennette A. (1996). Learning Short Synfire Chains by Self-Organization. *Neural Computation in Neural Systems*, UK, 7, 357-363.
- [10] Hill S., & Villa A. (1997). Dynamic transitions in global network activity influenced by the balance of excitation and inhibition. *Network: Computation in Neural Systems*, UK, Vol. 8, 2, 165-184.
- [11] Hill S., Villa A. (1995). Global Spatiotemporal Activity Influenced by Local Kinetics in a Simulated "Cortical" Neural Network. *Workshop on Supercomputing in Brain Research*, World Scientific, 371-375.
- [12] Ho T.V., Rouat J. (1997). A Novelty Detector Using a Network of Integrate and Fire Neurons. *7th Int. Conf. on Artificial Neural Networks*, Lausanne, Switzerland, 8-10 Oct. 1997, Lecture Note on Computer Science (1327), Springer, pp. 103-108.
- [13] Matsuno .Tet al. (1994). Periodic Signal Learning and Recognition in Coupled Oscillators. *Journal of Physical Society of Japan*, Vol. 63, No. 3, March, 1194-1204.
- [14] Stassinopoulos D., Bak P. (1994). Self-organization in a Simple Brain Model. *Proc. of WCNN'94*, San Diego, Jun, Vol. 1, pp. 4-26.
- [15] Thiran P., Hasler M (1996). Information storage using stable and unstable oscillations: an overview. *Int. Journal of Circuit Theory and Applications*, Vol. 24, 57-67.
- [16] Wang D., & Terman D. (1995). Locally Excitatory Globally Inhibitory Oscillatory Networks. *IEEE Transactions on Neural Networks*, 6: 283-286.

**Annexe C** (4 pages)

Rouat J. et Ho T.V. (1998). La détection de nouveauté basée sur le temps de stabilisation d'un réseau de neurones: application possible en reconnaissance de parole? .  
XXIIèmes Journées d'études sur la parole, Suisse, Juin, pp. 413-416.



## La détection de nouveauté basée sur le temps de stabilisation d'un réseau de neurones: application possible en reconnaissance de parole?

Jean ROUAT<sup>1</sup>, Tuong Vinh HO<sup>1,2</sup>

<sup>1</sup> ERMETIS, DSA, Université du Québec à Chicoutimi  
555 boul. de l'Université, CHICOUTIMI, Québec, CANADA, G7H 2B1  
Tél.: 1 418 545 5011 x5642 - Fax: 1 418 545 5012

<sup>2</sup> Ecole Polytechnique de Montréal, Canada  
e-mail: Jean\_Rouat@uqac.quebec.ca, vho@uqac.quebec.ca

### ABSTRACT

We propose a spiking neural network model inspired from a simulated cortex model. Also, a new paradigm for pattern recognition by neural networks with complex dynamics is presented. The 'relaxation' time of the network is used as a criterion for novelty detection. We compare the proposed neural network with Hopfield and backpropagation networks for a noisy digit recognition task. It is shown that the proposed network is more robust. We also design a limited experiment based on the recognition of temporal sequences of vowels and we show that the network is able to perform the recognition with a rate of 100% (sequences of 5 and 11 vowels). Regarding speech and pattern recognition tasks, the proposed spiking network seems to have a strong potential.

### 1. INTRODUCTION

Il existe des systèmes de reconnaissance de parole indépendants du locuteur pour des vocabulaires limités et qui offrent des performances intéressantes à condition que l'environnement ne soit pas trop corrompu par le bruit et les interférences [Dup97]. Toutefois, l'interrogation de bases de données à distance, (téléphone, radio, etc.), la transcription des nouvelles télévisées ou radio par exemple, nécessitent de concevoir des systèmes polyvalents et peu sensibles aux conditions et aux environnements d'opérations ainsi qu'aux bruits ambiants (téléphone cellulaire, cabine téléphonique, voitures, avions, camions, etc.) [Ezz97][Rob97]. Or le traitement de la parole en milieu bruyant est loin d'être résolu et les systèmes de reconnaissance actuels ne peuvent fonctionner dans de tels environnements sans observer de dégradation significative des performances [Gau97][Woo97][Bak97].

La mise en place de système de reconnaissance versatile fait souvent appel à des analyses perceptives inspirées de modèles du système auditif [Pat95]. Ces analyses codent l'information liée à la parole dans un espace à représentation spatio-temporelle complexe et difficile à caractériser via les algorithmes couramment utilisés en reconnaissance de parole. En effet, ces représentations spatio-temporelles sont structurées à la fois spectralement et dans le temps. L'information est en général non stationnaire. Il y a donc lieu de mettre au point des paramètres et des algorithmes de reconnaissance de formes capables d'exploiter cette information temporelle. Or une ma-

jeorité des algorithmes contemporains assument une certaine stationnarité du signal et des paramètres d'analyse. De plus, ces systèmes requièrent un apprentissage supervisé souvent long et fastidieux. Nous pensons qu'il y a lieu d'étudier l'élaboration de nouvelles techniques d'analyse et de reconnaissance adaptées à l'exploitation de l'information temporelle fine telle que générée par exemple via des analyses d'inspiration perceptive. Une information temporelle fine serait par exemple celle qui est liée à l'enveloppe des signaux à la sortie d'un banc de filtres cochléaires [Rou97].

### 2. MOTIVATION

Il apparaît intéressant d'élaborer de nouveaux outils aptes à traiter une information dynamique et non stationnaire. Ce type d'outil devrait pouvoir être utilisé en reconnaissance de formes afin de traiter une information codée dans le temps (parole, par exemple). Nous proposons un système qui devrait être en mesure de remplir ces critères.

Il s'agit d'un système de détection de la nouveauté basé sur le temps de stabilisation d'un réseau de neurones (dit bio-inspiré) dont l'architecture est inspirée de la couche IV du cortex. L'originalité du travail réside principalement dans la définition d'une règle de modification des connexions synaptiques et dans la création du critère de temps de stabilisation pour la reconnaissance. En effet, la majorité des techniques de reconnaissance des formes utilisant les réseaux de neurones (souvent formels) se base sur des critères relativement statiques (minimisation d'une fonction d'erreur, maximisation d'une probabilité, etc.) et codent l'information temporelle de façon statique via la structure spatiale des entrées [Hay94]. Par ailleurs, l'approche adoptée ici ne nécessite pas de supervision du réseau. Celui-ci est en mesure de détecter la nouveauté et de s'adapter de façon autonome. De plus, il n'y a pas de différence entre apprentissage et reconnaissance.

Dans un premier temps, il est important de tester et de valider ce type d'approche vis-à-vis des systèmes plus classiques afin de s'assurer que le réseau proposé puisse réaliser au minimum des tâches similaires à celles exécutées par les réseaux dits formels. Nous présentons donc une série de tests préliminaires effectués en reconnaissance de chiffres bruités ainsi qu'en reconnaissance de séquences de voyelles puis nous discuterons du potentiel de ce travail en reconnaissance de parole.

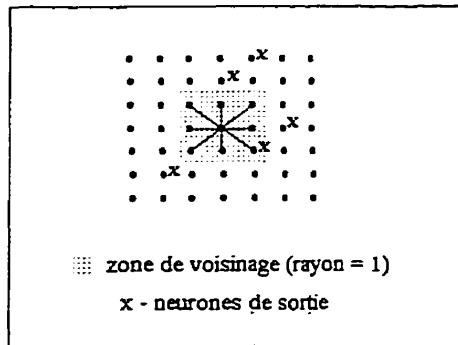


Figure 1: Architecture du réseau de neurone.

### 3. LE RÉSEAU DE NEURONES

#### 3.1. Modèle du neurone

Le neurone est de type intégration et décharge ('Integrate and Fire') avec période réfractaire et potentiel postsynaptique décroissant. L'état du neurone à l'instant  $t$  est caractérisé par son potentiel interne  $U$ . La réponse  $S_i(t)$  du neurone  $i$  est donnée par:

$$S_i(t) = \begin{cases} 0 & \text{pour } (t - t_{spike}) < \rho, \\ \mathcal{H}[U_i(t) - \theta] & \text{pour } (t - t_{spike}) \geq \rho \end{cases} \quad (1)$$

$\mathcal{H}$  est la fonction d'Heaviside. Le potentiel de contrôle  $U_i(t+1)$  pour la cellule  $i$  est l'intégration de toutes les réponses afférentes à l'instant  $t$  et tient compte de la fréquence instantanée  $f_i$  de décharge de la cellule  $i$ .

$$U_i(t+1) = \sum_j C_{ij}(t)S_j(t) + U_i(t) + e_i(t) + f_i(t) \quad (2)$$

$C_{ij}(t)$  est la connexion de la cellule  $j$  vers la cellule  $i$  à l'instant  $t$ .  $e_i(t)$  est le signal externe (entrée du stimulus) pour la cellule  $i$  considérée.

#### 3.2. Architecture et apprentissage

L'architecture est inspirée du modèle de la couche IV du cortex tel que proposé par Hill & Villa [Hil97]. Ils définissent une couche à 2 dimensions comprenant des neurones inhibiteurs et excitateurs avec récursivité. Chaque neurone est interconnecté à ceux qui appartiennent à son ensemble de voisinage. Cet ensemble est défini comme étant carré, centré autour du neurone et comprend une taille différente selon le type de connexion (inhibitrice ou excitatrice). Nous choisissons les neurones de sortie du réseau de façon aléatoire à partir de neurones de type excitateurs (figure 1). Le signal de sortie est caractérisé par l'état de décharge de l'ensemble des neurones de sortie.

L'apprentissage est inspiré du travail de Stassinopoulos et Bak [Sta94]. Pour chaque signal d'entrée, l'action du réseau est considérée comme étant un succès si au moins

Table 1: Taux d'erreurs pour un apprentissage sur les chiffres propres [0-4] et reconnaissance sur les chiffres [0-9] propres et bruités

Type de réseau	proposé	Hopfield	multicouches
Taux d'erreur (%)	10	21	27

Table 2: Taux d'erreurs pour un apprentissage sur les chiffres propres [5-9] et reconnaissance sur les chiffres [0-9] propres et bruités

Type de réseau	proposé	Hopfield	Multicouches
Taux d'erreur (%)	30	40	30

un neurone appartenant à l'ensemble des neurones de sortie se décharge. Dans ce cas, les connexions entre les neurones actifs sont renforcées. Si l'action est sans succès, les connexions entre les neurones actifs sont affaiblies. La mise à jour des connexions se fait selon la règle d'Hebb. Soit  $h$ , le signal de feedback ( $h = +1$  pour le renforcement ou  $-1$  pour affaiblir). La mise à jour des poids entre deux neurones est donnée par:

$$C_{ij}(t+1) = C_{ij}(t) + \alpha C_{ij}(1 - C_{ij})S_i(t)S_j(t)h(t) \quad (3)$$

avec  $\alpha$  le taux d'apprentissage.

#### 3.3. Critère de reconnaissance

Il n'y a pas de différence entre apprentissage et reconnaissance. La mise à jour des connexions synaptiques est toujours en cours sauf lorsque le réseau est stable. Après présentation d'un stimulus, le réseau a un comportement complexe qui finit par se stabiliser au cours du temps. Lorsque les changements de poids sont inférieurs à un seuil préétabli, on considère que le réseau est stable et que l'apprentissage ou la reconnaissance sont terminés. Le temps nécessaire  $T$  pour atteindre cet état stable est utilisé comme critère de détection de nouveauté. Ce temps  $T$  permet de caractériser le signal d'entrée. Un temps  $T$  très court (de l'ordre de 11 itérations) implique que le réseau a probablement déjà 'vu' ce signal au préalable. Cette notion permet de créer un système de détection de la nouveauté ayant un bon degré de robustesse vis-à-vis du bruit.

## 4. APPLICATION À LA RECONNAISSANCE DE CHIFFRES BRUITÉS

Des expériences de reconnaissance ont été réalisées à partir des chiffres 0 à 9 codés sur une matrice binaire de 7 par 5. Le réseau apprend uniquement sur les données non bruitées de 0 à 4 ou de 5 à 9 (colonne 1 ou 4 de la figure 2). Les expériences de reconnaissance portent sur les versions bruitées et propres des chiffres de 0 à 9 (toutes les colonnes de la figure 2). Le bruit est de type uniforme avec 20% des pixels modifiés. Le réseau est composé pour 70% de cellules excitatrices et pour 30% de

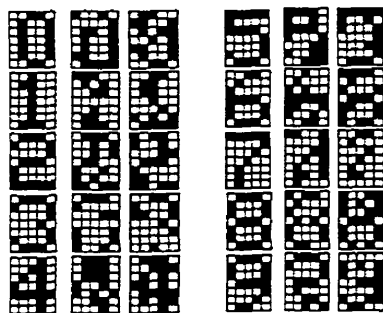


Figure 2: Images des chiffres propres (colonnes 1 et 4) et bruités (20% de bruit, colonnes 2,3,5 et 6).

cellules inhibitrices. Il comprend 49 cellules (7x7). Les chiffres sont codés sur une matrice de pixels 7x5. Chaque pixel est présenté sur l'entrée  $e_i$  d'une cellule choisie au hasard (i.e. 35 cellules sur les 49). Les tables 1 et 2 permettent de comparer le réseau proposé avec un réseau de Hopfield [Dem96] et un réseau multicouches utilisant la règle d'apprentissage généralisée avec moments [Ebe92]. Après un apprentissage à partir des chiffres propres de 0 à 4 (table 1) ou des chiffres propres de 5 à 9 (table 2), on présente l'ensemble des chiffres tel qu'illustrés à la figure 2. Le système doit ensuite différencier les chiffres qu'il connaît (ceux qui étaient représentés dans la séquence d'apprentissage) de ceux pour lesquels il n'a jamais 'vu' de représentant. Pour la série apprise à partir de [0-4], le réseau proposé est nettement supérieur (table 1). Pour la série apprise à partir de [5-9], les performances sont similaires à celles du réseau multicouches.

## 5. RECONNAISSANCE DE SÉQUENCES DE VOYELLES

Le but de ces expériences est d'illustrer la possibilité d'identifier des séquences de voyelles déjà apprises et de les isoler des autres séquences pour lesquelles on retrouve les mêmes voyelles dans un ordre différent.

### 5.1. Les données de parole et la taille du réseau

Les données de parole sont disponibles sur le site de [Mer98]. Nous utilisons les voyelles prononcées par un seul locuteur pour onze mots anglais: heed(i), hid(I), head(E), had (A), hard (a:) hud (Y), hod(O), hoard(C:), hood (U), who'd(u:) et heard(3:). Le signal est filtré passe-bas à 4.7 kHz puis il est échantillonné à 10 kHz et quantifié sur 12 bits. Une analyse LPC d'ordre 12 est réalisée afin d'extraire 10 coefficients par fenêtre de signal. Une fenêtre de Hamming de 512 points est centrée au préalable sur la zone stable de chaque voyelle. On utilise donc 10 paramètres LPC par voyelle. Le réseau comprend 10 cellules (matrice de 5x2). Chaque paramètre est présenté sur l'entrée  $e_i$  d'une cellule.

### 5.2. Méthodologie

Une première séquence de voyelles est utilisée comme séquence d'apprentissage. Chaque voyelle de cette

séquence est présentée au réseau de façon séquentielle pendant un intervalle de temps choisi  $T$ . Par exemple, la première voyelle est présentée au réseau pendant  $T$  itérations. Ensuite, la deuxième voyelle est présentée pour la même durée  $T$ . Il n'y a pas d'initialisation du réseau entre les présentations successives. On procède ainsi pour l'ensemble des voyelles de la séquence.

Après apprentissage, le comportement du réseau permet de savoir si celui-ci a identifié la séquence déjà apprise parmi les séquences qui lui sont présentées. Nous utilisons le temps de relaxation du réseau comme étant caractéristique du comportement. Chaque voyelle est présentée de façon séquentielle au réseau. Lorsque la première voyelle est présentée, le réseau oscille et se dirige vers un état d'équilibre. Ensuite, la deuxième voyelle est présentée, le réseau oscille à nouveau et atteint un autre état d'équilibre. On procède ainsi pour toutes les voyelles de la séquence. Après présentation de la séquence de voyelles, on obtient une série de temps de relaxation correspondant à la séquence de voyelles. En d'autres termes, nous avons effectué une 'projection' de la séquence de voyelles sur une séquence de temps de relaxation tout en réduisant la dimension des paramètres. L'analyse de cette séquence de temps permet d'indiquer si la série de voyelles a déjà été vue par le réseau.

Afin de faciliter l'analyse des séquences de temps, nous utilisons une méthode de codage très simple. Chaque intervalle de temps est codé par un chiffre. Par exemple, si le temps  $T$  se situe dans l'intervalle de 0 à 100 itérations ( $0 < T \leq 100$ ), le chiffre associé est 0. De même, si le temps  $T$  se situe dans l'intervalle de 100 à 200 itérations ( $100 < T \leq 200$ ), le code associé sera de 1, etc. De cette façon, on peut coder une séquence de temps de relaxation en une séquence de nombres entiers. En comparant les séquences de nombres correspondant aux séquences de voyelles de test avec la séquence déjà apprise, on peut savoir si une séquence de voyelles a déjà été 'vue' par le réseau. Les expériences suivantes illustrent ce principe.

### 5.3. Expérience 1: reconnaissance de séquences de 5 voyelles

Une séquence de 5 voyelles des mots anglais heed(i), hid(I), head(E), had (A) et hard (a:) est utilisée comme séquence d'apprentissage. Lors de l'apprentissage, chaque voyelle est présentée au réseau pour une durée de 200 itérations. Ensuite, cette même séquence est présentée au réseau afin de connaître le comportement de celui-ci face aux données déjà 'vues'. La séquence de temps de relaxation du réseau est de: 461, 11, 11, 245 et 145 itérations. En appliquant le codage précédent, on obtient la séquence de nombres suivante: 4, 0, 0, 2, 1. Pour les tests, nous utilisons un ensemble de séries de voyelles qui est composé de toutes les combinaisons possibles (120 combinaisons) des 5 voyelles déjà présentées au réseau lors de l'apprentissage. A titre d'exemple, la séquence de chiffres obtenue pour la série de voyelles (had (A), hid(I), head(E), heed(i), et hard (a:)) est 0, 0, 0, 5, 1. En comparant cette séquence à celle de l'apprentissage on peut conclure qu'elle n'avait jamais été vue par le réseau. Le taux d'erreur de reconnaissance des séquences est de 0% (au-

cune séquence en erreur).

#### 5.4. Expérience 2: reconnaissance de séquences de 11 voyelles

Une séquence de 11 voyelles a été obtenue à partir des mots anglais heed(i), hid(I), head(E), had (A), hard (a:) hud (Y), hod(O), hoard(C:), hood (U), who'd(u:) et heard(3:)). Elle est utilisée comme séquence d'apprentissage. Chaque voyelle est présentée au réseau pendant une durée de 200 itérations. Lors des tests, nous utilisons un ensemble de séries de voyelles qui comprend 100 combinaisons des 11 voyelles d'apprentissage. Le taux d'erreur est de 0% (aucune séquence en erreur).

### 6. DISCUSSION ET CONCLUSION

Les expériences de reconnaissance de données statiques (les chiffres corrompus) indiquent que le réseau est en mesure de faire un travail aussi bon et souvent meilleur qu'un réseau de Hopfield et un réseau à rétropropagation.

Les expériences portant sur les séquences de voyelles permettent de mettre en valeur l'aptitude du réseau à traiter des séquences. En effet, la réponse du réseau (temps de stabilisation) à une voyelle dépend de ses états précédents. Pour une même voyelle, les caractéristiques dynamiques sont liées aux présentations antérieures.

Les expériences présentées sont relativement limitées et il y a lieu d'approfondir le travail afin de bien connaître les limites du réseau et son potentiel en reconnaissance de parole. Toutefois, sa robustesse au bruit ainsi que son aptitude à traiter de l'information structurée dans le temps, laissent à penser que ce système est un bon candidat pour la mise au point de techniques de reconnaissance de parole.

### REMERCIEMENTS

Ce travail a été financé par le Conseil National de la Recherche en Sciences Naturelles et en Génie du Canada (CRSNG) ainsi que par la fondation de l'Université du Québec à Chicoutimi. Un merci tout particulier à Alessandro Villa pour les discussions stimulantes et enrichissantes en regard de ce travail.

### BIBLIOGRAPHIE

- [Bak97] R. Bakis, S.Schen, P. Gopalakrishnan, R. Gopinath, S. Maes and L. Polymenakos (1997). Transcription of broadcast news - system robustness issues and adaptation techniques. *Proc. of the IEEE-ICASSP*, Vol. 2, 711-714.
- [Dem96] Demuth H., Beale M. (1996). Neural Network Toolbox for Use with MATLAB, *The Math Works Inc.*
- [Dup97] S. Dupont, H. Bourlard, O. Deroo, V. Fontaine and J.M. Boite (1997). Hybrid HMM/ANN systems for training independant tasks: Experiments on Phonebook and related improvements. *Proc. of the IEEE-ICASSP*, Vol. 3, 1767-1770.
- [Ebe92] R. C. Eberhart, R. W. Dobbins. (1992). *Neural Networks PC Tools: a practical guide*. Academic Press. San Diego, USA.
- [Ezz97] H. Ezzaidi, I. Bourmeyster and J. Rouat (1997). A new algorithm for double talk detection and separation in the context of digital mobile radio telephone. *Proc. of the IEEE-ICASSP*, Vol. 3, 1897-1900.
- [Gau97] J.L. Gauvain, G. Adda, L. Lamel and M. Adda-Decker (1997). Transcribing Broadcast News Shows. *Proc. of the IEEE-ICASSP*, Vol. 2, 715-718.
- [Hay94] S. Haykin (1994). *Neural Networks – A Comprehensive Foundation*. IEEE Computer Society Press and Macmillan College Publishing Company, Inc., 1994.
- [Hil97] Hill S., Villa A. (1997). Dynamic transitions in global network activity influenced by the balance of excitation and inhibition. *Network: Computation in Neural Systems*, UK, Vol. 8, 2, 165-184.
- [Ho97] Ho T.V., Rouat J. (1997). A Novelty Detector Using a Network of Integrate and Fire Neurons. *7th Int. Conf. on Artificial Neural Networks*, Lausanne, Switzerland, 8-10 Oct. 1997, Lecture Note on Computer Science (1327), Springer, pp. 103-108.
- [Ho98] Ho T.V., Rouat J. (1998). Novelty Detection Based on Relaxation Time of a Network of Integrate-and-Fire Neurons. *International Joint Conference on Neural Networks*, Alaska, May 1998.
- [Mer98] Merz, C.J., Murphy, P.M. (1998). UCI Repository of machine learning databases. [<http://www.ics.uci.edu/mllearn/MLRepository.html>]. Irvine, CA: Univ. of California, Dep. of Information and Computer Science.
- [Pat95] R.D. Patterson and M.H. Allerhand (1995). Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. *J. Acoust. Soc. Amer.*, Vol. 98 (4), pp.1890-1894.
- [Rob97] Robust speech recognition for unknown communication channels. *ESCA-NATO Tutorial and Research Workshop*, Pont-à-Mousson, France, 17-18 avril, 1997, éditions ESCA.
- [Rou97] Spatio-temporal Pattern Recognition with Neural Networks: Application to Speech. *Artificial Neural Networks-ICANN'97, Lecture Notes in Comp.Sci.*,1327, Springer, 43-48.
- [Rub95] A. J. Rubio and J. M. López (eds.) (1995). *Speech Recognition and Coding*. New Advances and Trends. *NATO ASI Series*, Springer.
- [Sta94] Stassinopoulos D., Bak P. (1994). Self-organization in a Simple Brain Model. *Proc. of WCNN'94*, San Diego, Jun. Vol. 1, pp. 4-26.
- [Woo97] P. C. Woodland, M. J.F. Gales, D. Pye and S. J. Young (1997). Broadcast News Transcription Using HTK. *Proc. of the IEEE-ICASSP*, Vol. 2, 719-722.