Détection de ruptures et classification automatique dans un environnement de bruits impulsifs

Georges Linares

15 Novembre 1999

Résumé

Ce mémoire est consacré au problème de la supervision d'environnements réels de bruits impulsifs. L'objectif est de développer une méthode capable de détecter, dans un flux continu de bruits transitoires, les événements acoustiques qui ne s'inscrivent pas dans une certaine régularité du signal.

Nous montrons que les techniques classiques de détection de ruptures sont mal adaptées à la surveillance d'un milieu intrinsèquement irrégulier. Ceci nous amène à adopter une approche de plus haut niveau, basée sur le regroupement en classes des différentes composantes acoustiques de l'environnement.

L'étude bibliographique des algorithmes de classification automatique montre leur difficulté à classifier des formes qui peuvent être superposées. Or le recouvrement de signaux émis par différentes sources acoustiques est probable dans des milieux complexes.

Nous proposons une solution neuro-mimétique originale au problème de la modélisation dynamique du signal. Il s'agit d'une mémoire associative non-supervisée qui adapte sa structure par deux processus inter-dépendants : la modélisation du signal par classification d'unités acoustiques récurrentes et la détection d'événements provoquant la rupture du modèle. L'architecture proposée permet la détection et la classification de formes acoustiques linéairement mélangées. Néanmoins, la co-occurence d'une rupture et d'une forme acoustique référencée provoque une mauvaise formation du modèle dûe au mélange des prototypes des classes.

La séparation des classes mélangées peut être considérée comme un problème de séparation de sources. Ceci nous amène à implanter, dans l'architecture neuronale proposée, un mécanisme d'apprentissage qui s'inspire des méthodes de séparation aveugle de sources. Il s'agit d'un processus d'auto-organisation du classifieur neuronal qui minimise la dépendance statistique des classes.

Les performances de cet algorithme sont évaluées à la fois en détection de ruptures et en classification non-supervisée. Nos tests montrent le bon comportement du modèle dans des situations de fort taux de superposition des formes acoustiques ainsi qu'une dégradation sensible des performances en présence de bruit blanc de forte énergie.

Introduction

Depuis la révolution industrielle, le cinéma et la littérature de science fiction mettent régulièrement en scène des créatures issues de la technologie et dont le comportement ou l'apparence sont semblables à ceux des humains. Ce thème récurrent véhicule l'image d'une science inquiétante, capable de défier les dieux pour créer des monstres maladroits et psychopathes, mais habités néanmoins d'une faible étincelle de vie. Heureusement, ces créatures échappent généralement au contrôle de leur créateur, et finissent par se retourner contre lui.

L'informatique a aussi ses docteurs Franckenstein : au lieu des monstres approximatifs, on trouve des machines au mental fragile, mais capables de parler, de reconnaître des visages, d'analyser une situation et parfois d'échafauder des plans machiavéliques pour atteindre un objectif souvent mystérieux.

L'ambition des chercheurs en Intelligence Artificielle n'est sans doute pas de faire une machine à leur image, mais plutôt d'améliorer un outil en lui donnant des capacités de perception, d'analyse et de décision proches de celles d'un humain. Néanmoins, cette tentative de simulation des mécanismes mentaux repose sur l'hypothèse que l'intelligence peut être simulée par une machine virtuelle. Un ordinateur psychopathe est donc théoriquement possible. Malheureusement, les performances des machines destinées à simuler le comportement humain restent assez faibles, en particulier lorsque la tâche à accomplir implique de bonnes capacités perceptives. Quelles sont les raisons de cet échec ?

Derrière l'apparente simplicité des tâches élémentaires réalisées par un homme (marcher, parler, écouter, reconnaître un visage, saisir un objet, etc.), se cache une très grande complexité, qui peut être propre à la difficulté de la tâche elle-même, mais aussi à la diversité des conditions dans lesquelles elle est réalisée. La résolution spontannée et parfois inconsciente de ces problèmes fait appel à un ensemble de ressources d'origine et de nature diverses : capacités sensorielles et perceptives, connaissance abstraite de l'univers, expérience, etc. Ces ressources reposent généralement sur une grande masse d'informations qui ont été collectées tout au long de la vie de l'individu. Malheureusement, on ne sait pas précisément comment la connaissance est acquise, stockée et exploitée dans le cerveau, ce qui rend difficile la simulation de fonctions basées sur ces mécanismes. Cette méconnaissance des mécanismes biologiques de traitement de l'information a incité les chercheurs à considérer le cerveau biologique comme une source d'inspiration plutôt que comme un modèle. Mais ce changement d'approche n'a pas changé le problème de fond de l'IA qui est resté celui de l'acquisition, de la représentation et de l'exploitation des connaissances.

La difficulté à simuler le comportement humain se retrouve dans les différentes tentatives d'automatisation des tâches qui impliquent la perception, notamment l'analyse de formes visuelles ou acoustiques. La supervision d'un environnement réel se heurte au même type de problèmes. Il s'agit de simuler ce processus permanent qui filtre l'information sensorielle et éveille l'attention lorsqu'un événement anormal se produit. Pour comprendre ce mécanisme, il faudrait savoir comment fonctionne le système attentionnel humain, quel est le mécanisme qui nous permet de distinguer un événement normal d'un événement "alarmant", qu'est-ce qu'un événement "alarmant", etc. Malheureusement, ces processus sont eux aussi mal connus. Ce mémoire traite du problème de la supervision automatique de milieux réels par l'analyse du canal acoustique.

Le premier chapitre expose, de façon détaillée, le problème abordé dans ce mémoire. La définition d'un environnement de bruits impulsifs est donnée et les problèmes spécifiques à ce type de milieux sont évoqués. Cette description du cadre applicatif de notre travail nous conduit à préciser et à justifier la méthodologie adoptée.

Le second chapitre aborde le problème de la définition d'un critère de rupture. Différentes approches rencontrées dans la littérature sont décrites et confrontées. Leur adéquation à notre problème est ensuite discutée. Ce survol des travaux développés dans le domaine de la détection de ruptures montre l'inadaptation de ces méthodes au problème de la surveillance d'environnements transitoires. La nécessité d'une approche de plus haut niveau est mise en évidence. Cette constatation nous amène finalement à reformuler notre problème en termes de modélisation de l'environnement par classification des différentes composantes acoustiques.

Le troisième chapitre décrit la problématique abordée en classification automatique non-supervisée. Différentes méthodes sont présentées. Ce descriptif met l'accent sur les capacités d'auto-adaptation des algorithmes proposés dans la littérature. Enfin, nous discutons de la capacité de ces techniques à modéliser dynamiquement un environnement réel de bruits impulsifs. Ce chapitre nous permet d'esquisser les principes d'une approche neuronale pour la détection de ruptures et la classification d'événements acoustiques transitoires.

Dans le chapitre 4, nous proposons un réseau neuro-mimétique original pour la détection de ruptures et la classification de bruits impulsifs. Son architecture est d'abord décrite dans le détail. Le processus d'adaptation en ligne du modèle est ensuite présenté, et illustré par un exemple. Enfin nous discutons du potentiel de cette méthode à dissocier des événements acoustiques simultanés. Ceci nous amène à poser le problème de la réorganisation du réseau dans les cas de ruptures simultanées du modèle. Nous montrons que la recherche d'un tel processus peut être formulée comme un problème de séparation aveugle de sources.

Le chapitre 5 expose la thématique abordée en séparation aveugle de sources. Après avoir posé rigoureusement le problème traité, nous décrivons différentes approches qui ont été developpées dans la littérature. Enfin, nous situons les approches dans le cadre théorique plus général de la représentation optimale des données dans des espaces multi-dimensionnels. Ce chapitre nous permet d'introduire le principe d'une règle d'adaptation du modèle neuro-mimétique présenté au chapitre précédent.

Le chapitre 6 décrit un processus d'auto-organisation du modèle présenté au chapitre 4. La règle d'apprentissage proposée s'inspire de travaux réalisés dans le domaine de la séparation de sources. La validité théorique de notre approche est montrée, puis illustrée par un exemple. Ses performances sont ensuite évaluées sur des signaux simulés, à la fois en terme de détection de ruptures et de classification.

Le dernier chapitre conclut sur les capacités de la méthode proposée à effectuer une tâche de surveillance d'un environnement de bruits impulsifs. Nous dressons un bilan critique des points forts et des limitations de la méthode proposée. Enfin, les perspectives ouvertes par l'étude réalisée sont passées en revue. Nous montrons que l'approche proposée dans ce mémoire peut être généralisée à d'autres types de problèmes.

Chapitre 1

Position du problème

Ce chapitre précise le thème abordé dans ce mémoire. Le problème de la surveillance d'un environnement réel est présenté d'une façon générale, puis formulé plus concrètement comme un problème de détection de ruptures dans un flux d'événements acoustiques.

Le cadre applicatif de ce travail est présenté par la description des spécificités du signal traité. La notion de bruits impulsifs est définie.

Nous décrivons ensuite les difficultés propres au traitement de l'acoustique dans un milieu réel complexe. L'accent est mis sur la variabilité et la diversité des formes acoustiques engendrées par les phénomènes physiques composant le milieu.

Enfin, nous justifions le choix d'une approche non-supervisée, adaptative et dynamique pour la détection de ruptures dans un environnement réel.

1.1 Objectif

Surveiller un phénomème, c'est l'observer de façon à isoler les événements qui peuvent être considérés comme "alarmants". Le codage d'une notion aussi floue paraît difficile dans le cadre des techniques employées ici. Notre objectif est moins ambitieux mais plus réaliste : dans un flux ininterrompu d'événements acoustiques, nous voulons détecter ceux qui sont en rupture, c'est à dire ceux qui ne s'inscrivent pas dans une certaine régularité du signal. Bien que plus précise, cette formulation laisse la place à de nombreuses interprétations : quelle est la nature du signal à traiter, quels sont les paramètres qui seront observés, quel critère nous permettra de décider de l'irrégularité d'un événement ?

Le signal à traiter est issu d'un capteur audio plongé dans un environnement de bruits impulsifs. Ce choix est motivé par une collaboration avec le ministère de la défense pour la conception d'un système d'aide à la surveillance acoustique du milieu sous-marin. Il s'agit de concevoir un automate capable d'extraire d'un ensemble de bruits impulsifs ceux qui sont significatifs d'un changement d'état de l'environnement et qui doivent, par conséquent, susciter une réaction.

Les difficultés posées par la surveillance d'un environnement de bruits transitoires sont liées à la fois à la nature du signal traité et au principe de supervision d'un environnement complexe par l'étude du canal acoustique. Ces deux points font l'objet des sections suivantes.

1.2 Environnement réel de bruit impulsifs

Un bruit impulsif est un bruit bref dont l'attaque est brutale. La brièveté d'un signal est une caractéristique relative à sa fréquence d'échantillonnage et, par ailleurs, assez subjective. On considère que nos bruits impulsifs durent au plus quelques dixièmes de seconde et, surtout, on négligera leur structure temporelle. Cet aspect est capital dans la mesure où c'est justement la dimension temporelle qui caractérise les problèmes de traitement du signal acoustique. Cependant, cette dimension n'est pas écartée, puisque notre objectif est de traiter un flux ininterrompu d'informations acoustiques.

La violence d'un bruit impulsif est une caractéristique de son enveloppe plutôt que de son énergie. Elle correspond à une attaque brutale du son. Il s'agit donc d'une forme particulière de sa structure temporelle. Cette information étant perdue,

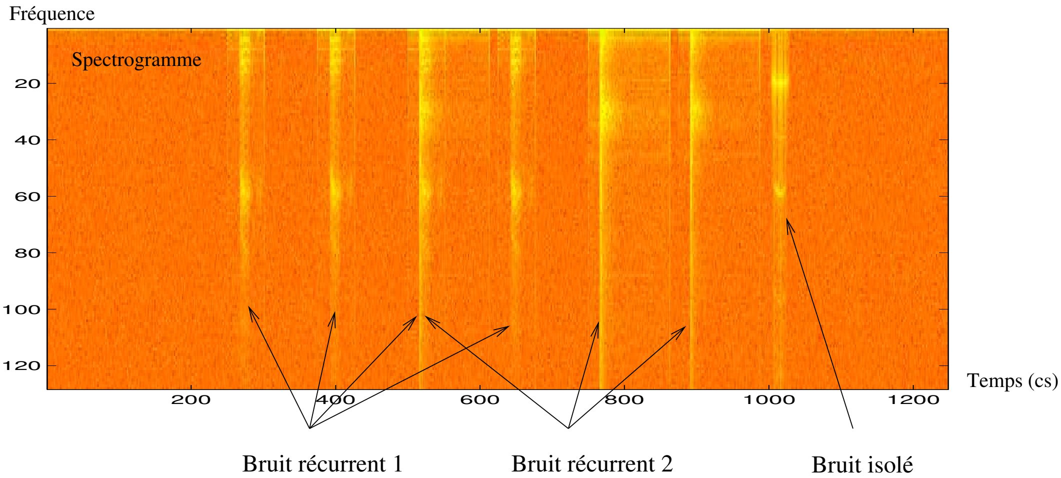

FIG. 1.1: *Exemple de spectrogramme d'un signal capté dans un environnement de bruits impulsifs. Ce signal synthétique est composé de deux séquences de bruits impulsifs récurrents et d'un bruit impulsif isolé. Les événements répétitifs se recouvrent au temps 5.1 s. Ces bruits transitoires sont plongés dans un bruit aléatoire de distribution uniforme.*

cet aspect des bruits impulsifs est négligé.

Le signal traité est composé de divers agents qui produisent des bruits transitoires variés. Certains de ces bruits se répètent de façon relativement fréquente. Dans ce cas, toutes les occurrences d'un même événement acoustique peuvent être regroupées en une classe homogène, dont seul le premier élément doit provoquer une détection de rupture.

D'autres bruits sont rares relativement à l'échelle de temps choisie. Ces événements isolés signalent un phénomène ponctuel qui doit, lui aussi, être détecté par le système.

Bien que notre étude se limite aux seuls signaux impulsionnels, un capteur placé dans un milieu réel perçoit généralement un bruit de fond peu structuré qui peut être, dans le cas de signaux sous-marins, un bruit de mer large bande plutôt présent dans les basses fréquences. Le traitement de ce type de bruit n'est pas l'objet de cette étude. Cependant, nous tiendrons compte de la présence éventuelle d'un bruit de fond dans les signaux étudiés.

La figure 1.1 montre un exemple typique de signal perçu par un capteur plongé dans un environnement de bruits impulsifs. L'image est ici issue d'une analyse de Fourier du signal, réalisée dans une fenêtre glissante. On peut observer la brièveté des événements acoustiques, dont on distingue cependant la structure temporelle. Ce signal synthétique est composé de deux séquences de bruits impulsifs récurrents qui se recouvrent à l'instant $5.1s$. La première d'entre elles apparaît à l'instant $3s$ et dure environ 4 s. Le seconde séquence est composée de trois bruits impulsifs qui se produisent dans l'intervalle de temps de 5.1 à 9 s. Un bruit isolé survient au temps $10s$. Enfin, l'ensemble de ces bruits transitoires est plongé dans un bruit de fond de distribution uniforme.

1.3 Surveillance d'un environnement réel

1.3.1 Image acoustique d'un environnement réel

Un environnement réel est composé d'une multitude de phénomènes physiques qui produisent des signaux acoustiques aléatoires. Le canal acoustique en renvoie une image à la fois incomplète et redondante.

Incomplète parce que certains de ces phénomènes sont silencieux, même si leur présence se traduit indirectement par des effets acoustiques secondaires (un cycliste coupe silencieusement la route d'un automobiliste qui freine, puis l'interpelle). L'environnement est ponctué d'événements acoustiques mais, généralement, les phénomènes physiques qui le composent ne peuvent pas être complètement décrits par l'information acoustique. Cette incapacité à "expliquer" un événement par la seule perception auditive est sans doute une source d'erreurs pour un système de supervision de l'environnement.

Si un phénomène physique n'est pas nécessairement bruyant, l'information portée par le canal acoustique est fortement redondante : à un phénomène peut correspondre un ensemble d'images acoustiques qui peuvent prendre des formes variées, se répéter ou se mélanger à d'autres (crissements de pneus, éclats de voix, etc.).

1.3.2 Séquentialité des sons

Généralement, les systèmes de traitement de l'information acoustique sont perturbés par les différents sons émis par l'environnement. Ceci repose sur l'idée qu'un signal utile est plongé dans un milieu produisant des bruits qui sont considérés comme parasites.

Ici, le problème ne peut pas être posé de cette façon. Il n'y a pas un signal utile à extraire, ou du bruit à effacer. Tout ce qui est perçu participe à une description globale de l'environnement. Il ne faut pas considérer pour autant que le problème du bruit ne se pose pas, au contraire : plutôt que d'avoir une forme acoustique modifiée par le bruit, ce sont plusieurs signaux utiles qui sont superposés. Cette simultanéité éventuelle d'événements acoustiques complique fortement le problème de supervision de l'environnement. Une telle situation étant très probable en conditions réelles, l'hypothèse d'une séquentialité des événements nous a paru trop restrictive. Nous avons donc cherché à développer une méthode capable de gérer correctement de telles situations.

1.3.3 Variabilité

Le signal engendré par un phénomène physique est modifié par une variabilité propre à l'environnement : variabilité des conditions d'acquisition (qualité du capteur, de la ligne, etc.), variabilité de la propogation de l'onde en fonction des caratéristiques du milieu (hydrométrie, réverbérations, etc.), variabilité de l'origine des sources, des réalisations acoustiques elles-mêmes, etc. Cette variété des formes qui peuvent être engendrées par les phénomènes physiques perceptibles par un capteur audio augmente la difficulté du traitement. Par ailleurs, cette variabilité naturelle des réalisations acoustiques et du milieu de propagation est accentuée, dans le cas de signaux sous-marins, par la mobilité éventuelle des capteurs.

1.3.4 Imprévisibilité

De nombreux systèmes de traitement de l'information acoustique reposent sur l'hypothèse que le signal peut être modélisé *a priori*. Cette hypothèse de prédictibilité des données conduit à paramétrer le système en fonction d'une connaissance définitive, en espérant que les données futures se conforment à ce cadre préalablement arrêté. Une telle approche est mal adaptée au problème de supervision d'un environnement réel, essentiellement pour trois raisons :

- on ne peut pas établir une liste exhaustive de tous les événements susceptibles de survenir dans un milieu réel, *a fortiori* si on cherche à détecter ceux qui sont inattendus,
- on ne peut pas fixer *a priori* les formes acoustiques qui seront engendrées par des phénomènes physiques divers (et qui sont eux-mêmes imprévisibles),

- il est difficile de fixer le cadre des évolutions du milieu de propagation, dont un des effets sera de modifier les formes acoustiques perçues.

1.4 Vers un système dynamique

Notre objectif est de développer une méthode automatique de détection de ruptures dans un environnement de bruits impulsifs. Le signal n'est pas segmenté mais constitue un flux ininterrompu d'informations acoustiques. Cette observation nous a conduit à opter pour un traitement en ligne des données. De plus, c'est le mode de fonctionnement naturel d'un système de veille acoustique.

Par ailleurs, on a vu que l'application de la méthode dans un milieu réel limite fortement les hypothèses qui auraient pu simplifier le problème en limitant la généralité des solutions proposées.

Le caractère imprévisible d'un environnement réel rend sa modélisation *a priori* difficile, et parfois irréaliste. Ceci nous a incité à écarter l'utilisation de corpus contenant des exemples des formes acoustiques susceptibles d'être rencontrées. Ceci exclut l'usage d'une méthode d'apprentissage supervisé qui permettrait une paramétrisation préalable du système.

Par ailleurs, l'environnement est variable, à la fois par les paramètres physiques qui conditionnent la propagation de l'onde et par la diversité des phénomènes qui engendrent le signal acoustique. Cette variabilité suggère une approche dynamique et adaptative du problème de la détection de ruptures en milieu réel.

Enfin, nous n'avons pas fait l'hypothèse de séquentialité des événements, ni celle d'une simultanéité improbable. La possibilité de superposition de formes acoustiques augmente fortement la difficulté du traitement, en particulier parce que ceci est conjugué à l'absence de connaissances préalables sur les formes acoustiques qui seront rencontrées. Ce dernier choix a fortement orienté nos recherches, et constitue un point essentiel du travail présenté.

1.5 Conclusion

Notre objectif est donc de développer un système non supervisé pour la détection de ruptures dans un signal composé de bruits impulsifs variables et imprévisibles.

Cette méthode devra être capable de s'adapter dynamiquement aux évolutions du milieu et de gérer correctement des situations critiques de superposition d'événements acoustiques. Globalement, on cherchera à réaliser un système aussi générique que possible, fondé sur un minimum de connaissance *a priori*. L'approche adoptée consiste à favoriser la capacité d'adaptation au détriment d'une connaissance dont on sait qu'elle finira, tôt ou tard, par se révéler partielle et insuffisante.

Chapitre 2

Détection de ruptures

Ce chapitre décrit la problématique abordée en détection de ruptures. Deux approches distinctes sont présentées : la surveillance d'un paramètre et la comparaison de modèles.

La première approche consiste à observer les variations brutales d'un paramètre descriptif du signal analysé.

La seconde repose sur la comparaison d'un modèle global du signal et d'un modèle évalué dynamiquement sur une fenêtre temporelle de taille réduite.

Enfin, nous discutons de la capacité des méthodes de détection de ruptures à traiter une tâche de surveillance d'un environnement réel de bruits impulsifs.

2.1 Introduction

Les méthodes de détection de ruptures proposent des solutions au problème suivant : comment détecter et localiser des événements acoustiques transitoires dans un signal aléatoire. La recherche de ces points de discontinuité est motivée par le fait qu'ils sont généralement caractéristiques du signal analysé, et souvent significatifs du phénomène physique qui a engendré le signal. Ces méthodes donnent lieu à deux types d'applications :

- la segmentation d'un signal en parties quasi-stationnaires. Dans ce cas, la détection d'irrégularités permet de segmenter un signal globalement transitoire en zones stables. Le codage et, plus généralement, les post-traitements sont facilités par cette stabilité locale du signal.
- la surveillance d'environnements réels. Ce type d'applications concerne plus particulièrement la surveillance de milieux industriels très spécifiques et s'accompagne souvent d'une analyse des points de discontinuité. Typiquement, ces techniques sont utilisées pour la détection et le diagnostic de pannes sur des machines tournantes.

On voit que ces méthodes sont généralement appliquées à des signaux localement transitoires mais néanmoins stables par parties. Notre signal est composé de bruits impulsifs dont le nombre par unité de temps peut être très élevé. Cette forte densité de bruits transitoires provoque une constante instabilité du signal. Dans ce chapitre, nous faisons une étude bibliographique du potentiel des algorithmes de détection de ruptures dans des signaux de ce type.

2.2 Surveillance d'un paramètre et rupture de modèle

Deux approches différentes ont été largement développées dans la littérature.
La première consiste à chercher les changements brutaux de certains paramètres descriptifs du signal analysé. Souvent fondées sur un principe de surveillance d'une statistique, ces méthodes se distinguent par le choix d'un codage du signal et celui d'un critère de rupture.

Une autre approche consiste à confronter un modèle global du signal à des observations locales : une analyse du signal à long terme est faite et comparée à un modèle estimé dans une fenêtre temporelle de taille réduite. Une rupture est détectée lorsque ces deux modèles ne coïncident pas. Se pose alors le problème du choix des modèles et des outils de comparaison.

2.3 Discontinuité d'un paramètre

Il s'agit de déterminer un paramètre décrivant un ou plusieurs aspects du signal analysé, puis de mettre au point un indice à partir duquel la décision de rupture pourra être prise. Généralement, le signal est identifié par un modèle prédictif dont l'erreur de prédiction sera exploitée de façon à obtenir la détection des changements brusques du signal. Plus que dans le modèle adopté, dont le choix influence peu les performances finales des détecteurs ([Ljung 83]), c'est la stratégie d'exploitation de l'erreur de prédiction qui distingue les différentes méthodes.

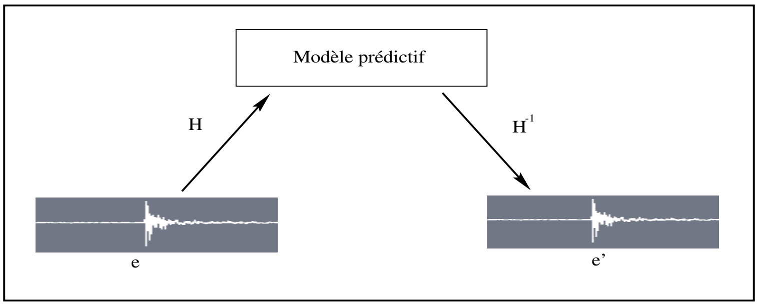

FIG. 2.1: *Détection de rupture par observation de l'erreur de prédiction d'un modèle prédictif : le signal e permet la paramétrisation d'un modèle prédictif H. Le filtrage inverse permet d'estimer l'erreur de prédiction e'.*

2.3.1 Densité d'erreurs de prédiction

Cette méthode, d'abord proposée par [Castanié 83] puis développée par [Daymier 85], consiste à estimer un modèle linéaire prédictif du signal puis à mesurer la densité d'erreurs de prédiction dans une fenêtre temporelle de largeur fixée *a priori*. Une erreur de prédiction a lieu lorsque la distance du signal estimé au signal réel dépasse un seuil. Si un certain nombre d'erreurs de prédiction surviennent consécutivement, une rupture est détectée.

Cette méthode a l'inconvénient d'utiliser trois seuils qui influenceront fortement les probabilités de fausses alarmes et de non-détection. Le choix de ces paramètres du détecteur relève d'une estimation empirique et sans doute peu fiable du signal. On peut néanmoins envisager l'utilisation d'un corpus d'apprentissage sur lequel ces paramètres seraient estimés, bien que cette solution ne corresponde pas à l'approche non supervisée qui a été adoptée ici. De plus, un bruit impulsif est peu prédictible. L'application de cette méthode à des signaux impulsifs conduirait à un très fort taux de fausses alarmes. Enfin, bien que deux événements puissent être prévisibles, leur mélange ne l'est généralement pas. La superposition de formes acoustiques perturberait sans doute la paramétrisation du prédicteur, et introduirait par la suite des erreurs de prédiction injustifiées.

2.3.2 Indépendance de l'erreur de prédiction

Cette méthode, introduite par [Doncarli 83], consiste à évaluer une cohérence locale de l'erreur de prédiction. La corrélation temporelle à court terme des erreurs de prédictions successives est calculée. Nulle dans les phases stables, cette corrélation s'écarte du zéro dans les phases transitoires. C'est ce décentrage des corrélations qui constitue le critère de rupture adopté. Leur estimation est faite par un filtre passe-bas du premier ordre, le paramètre α du filtre étant un terme d'oubli exponentiel qui permet, ici, de régler la sensibilité du détecteur :

$$r_t = \alpha . r_{t-1} + (1-\alpha) . e'_t . e'_{t-1} \ , \quad \alpha \in [0,1]$$

où e'_t est l'erreur de prédiction centrée, à l'instant t. La décision de rupture est prise lorsque la variable aléatoire r_t n'est plus centrée, ce qui se produit lorsque les erreurs de prédiction successives sont linéairement corrélées. Cet algorithme, comme le précédent, suppose que les événements acoustiques composant le signal sont prévisibles à l'échelle de la fenêtre d'analyse, ce qui est généralement faux dans le cas des bruits impulsifs.

2.4 Rupture de modèle

Les méthodes issues de la seconde approche sont basées sur la confrontation de deux modèles. Le premier est un modèle global du signal, qui est estimé à partir d'un ensemble important d'observations. Le second modèle représente une caractérisation locale du signal qui, lorsqu'elle diffère du modèle global, peut permettre la détection de ruptures. Il s'agit donc de la comparaison d'observations à court et à long terme du signal. Ces méthodes se caractérisent essentiellement par le choix des modèles et celui des différentes fenêtres d'analyse.

2.4.1 Distance cepstrale

[Basseville 89] décrit et évalue de nombreuses distances entre modèles qui peuvent permettre la détection de ruptures. La plus simple est une distance quadratique entre spectres à long et à court terme :

$$d(s_0, s_1) = \int_{-\frac{1}{2}}^{\frac{1}{2}} (log(s_0(\nu)) - log(s_1(\nu))^2 d\nu$$

où $s_0(\nu)$ et $s_1(\nu)$ sont les densités spectrales de puissance estimées à court et à long terme, et ν la fréquence normalisée. Lorsque la distance ainsi évaluée dépasse un seuil fixe, une rupture est détectée.

Cette méthode suppose que l'observation globale du signal permet une modélisation qui coïncide généralement avec une analyse locale du signal, sauf aux moments de transition d'un état vers un autre. Notre signal étant supposé perpétuellement irrégulier, l'écart entre l'image spectrale instantanée et l'image spectrale moyenne du signal doit être constamment importante. Cette méthode est donc mal adaptée au traitement de signaux intrinsèquement transitoires.

2.4.2 Rapport de vraisemblance généralisé

Cette méthode consiste à établir deux lois de densité de probabilité $p_0(Y_1^t)$ et $p_1(Y_1^t)$ d'une séquence d'observations Y_1^t pour chacune des hypothèses H_0 et H_1 définies par :

H_0 : le modèle global décrit bien l'ensemble du signal observé.

H_1 : il existe un instant t avant lequel le signal est décrit par un modèle M_0 et après lequel le signal est décrit par un autre modèle $M_1 \neq M_0$.

Le logarithme du rapport de vraisemblance S_1^t entre ces deux hypothèses est ensuite utilisé comme mesure d'irrégularité locale du signal :

$$S_1^t(H_1/H_0) = \sum_{k=1}^{t} log(\frac{p_1(Y_1^t)}{p_0(Y_1^t)})$$

Lorsqu'il existe une hypothèse H_1 pour laquelle ce rapport dépasse un seuil fixé *a priori*, une rupture est détectée.

Cette méthode étant extrêmement coûteuse en temps de calcul, de nombreux auteurs ont proposé des techniques dérivées de complexités inférieures. [Brandt 83b] [Brandt 83a] dissocient les processus de détection et de localisation, affinant la localisation de la rupture par réductions successives de la largeur de la fenêtre d'analyse. D'autres auteurs ([Nikiforov 86], [Basseville 93], [Zhang 92], [Theys 93]) proposent des méthodes tirant parti d'une information *a priori* sur la forme de la rupture à détecter. Néanmoins, la quantité de données nécessaire à la paramétrisation du système resterait problématique dans notre cas. En particulier, la mobilité éventuelle des capteurs interdit l'initialisation du détecteur de ruptures sur une période stable de durée importante.

2.4.3 Méthode du maximum de vraisemblance

Les performances des méthodes fondées sur la comparaison de modèles à court et à long terme sont dépendantes de la qualité de l'estimation de ces modèles. Cette estimation est parfois imprécise, en particulier celle du modèle à court terme qui doit être déterminée à partir d'une faible quantité de données. [Carpentier 92] proposent une méthode qui tient compte de cette imprécision dans l'estimation des modèles : l'erreur d'estimation est considérée comme une variable aléatoire, et la décision de rupture est prise en évaluant la probabilité *a posteriori* qu'une rupture ait eu lieu à un instant donné. Une règle de décision bayésienne, fréquemment rencontrée dans des techniques de classification supervisées, est ici utilisée pour la détection de ruptures.

Comme la méthode précédente, l'usage de cet algorithme pour la détection de ruptures dans un environnement en évolution rapide semble peu réaliste : l'estimation de la probabilité de rupture nécessiterait une stabilité du signal sur un intervalle de temps important. Le manque de plasticité de cette technique la rend peu adaptée au problème qui est le nôtre.

2.5 Détection de ruptures et environnement de bruits transitoires

Que se soit par détection des variations brutales d'un paramètre descriptif du signal ou par comparaison entre modèles à court et à long terme, les méthodes de détection de ruptures cherchent à isoler les moments de transition entre les parties stables d'un signal.Ces méthodes sont inadaptées à la surveillance d'environnements de bruits impulsifs, essentiellement pour deux raisons :

- les signaux qui résultent du mélange de plusieurs sources de bruits impulsifs sont doublement irréguliers : irrégularité due au caractère bref et énergétique des sons eux-mêmes, diversité et variabilité des sources composant l'environnement acoustique. Chercher les variations brutales d'un signal qui est, par nature, aussi instable conduirait inévitablement à la détection permanente de ruptures. Les méthodes décrites dans ce chapitre semblent donc plus adaptées au traitement de signaux constitués d'une succession de parties stables qu'à celui de signaux transitoires.
- surveiller un environnement acoustique, c'est être capable de distinguer un événement des autres par son caractère alarmant. Bien que cette notion soit difficile à préciser dans le cadre d'un système automatique de surveillance, on a vu qu'on cherchait plutôt à détecter un événement inconnu ou improbable dans un contexte donné. Cette idée est exploitée dans certains des détecteurs de ruptures présentés. Les méthodes fondées sur le calcul d'une erreur de prédiction utilisent le caractère inattendu d'un événement acoustique comme indice de rupture. Malheureusement, l'occurrence d'un bruit impulsif n'est pas annoncée dans le signal, et de tels événements sont imprévisibles au niveau d'analyse auquel ces prédicteurs fonctionnent. En effet, bien que la $n^{i\grave{e}me}$ occurrence d'un événement acoustique puisse être prévisible, les méthodes utilisées cherchent à prévoir l'évolution du signal dans une fenêtre temporelle de courte durée relativement à la fréquence du phénomène. À cette échelle, les bruits impulsifs sont non seulement imprévisibles, mais aussi quasiment équiprobables : après une même séquence d'observations, plusieurs bruits impulsifs sont susceptibles d'apparaître, indifféremment. Un prédicteur proposera une valeur moyenne des différents futurs possibles, valeur qui serait sans doute parfaitement improbable. La méthode du maximum de vraisemblance aborde ce problème, lié au

niveau de traitement, de façon plus globale, par l'usage d'un critère de rupture purement probabiliste qui pourrait apporter une réponse intéressante au problème de l'imprévisiblité locale d'un bruit impulsif. Malheureusement, une telle approche semble difficilement compatible avec l'objectif initialement fixé de ne pas utiliser d'hypothèses sur la séquentialité des événements acoustiques : la méthode envisagée doit fonctionner dans des situations critiques, mais probables en milieu très bruité, de simultanéité d'événements acoustiques. Par ailleurs, la quantité de données nécessaires à l'estimation des paramètres du modèle probabiliste aurait rendu son adaptation en ligne problématique.

2.6 Conclusion

Les méthodes décrites fonctionnent donc à un niveau de traitement inadapté au problème spécifique de la surveillance d'un environnement de bruits impulsifs : l'irrégularité intrinsèque du signal, ainsi que son caractère imprévisible et variant, nous ont incités à aborder le problème de façon plus globale.

Plutôt que de chercher dans le signal les instants de transitions entre des parties stables hypothétiques, nous avons considéré qu'il constituait une image à la fois incomplète et redondante de l'environnement physique. Reconstruire une représentation de cet environnement à partir de l'information acoustique peut permettre de déplacer le problème de la détection de ruptures vers celui, plus général, de l'analyse de scènes auditives.

Le signal étudié est composé de bruits isolés ou récurrents, qui signalent des phénoménes physiques ponctuels ou répétitifs. En regroupant ces événements en classes homogènes, on peut construire une image concise et structurée du signal. Chaque classe représenterait alors une composante de l'environnement active dans une fenêtre temporelle de largeur donnée. Cette modélisation du signal en classes d'événements acoustiques peut permettre la détection des ruptures du modèle ainsi constitué, par une mesure de la qualité de la représentation du signal par le modèle.

Le critère de rupture retenu serait donc défini comme un changement d'état du modèle correspondant à la détection d'événements acoustiques mal représentés par ce modèle. L'hypothèse H_1, définie dans les algorithmes de détection de rupture de modèle comme une hypothèse vérifié en présence de ruptures, correspondrait alors à l'apparition, dans le signal, d'un événement qu'on peut difficilement affecter à une des classes référencées.

Dans la perspective de construire un système autonome capable d'apprendre et de mettre à jour dynamiquement une telle représentation de l'environnement, nous avons étudié plusieurs algorithmes de classification non supervisée. Le chapitre suivant décrit la problématique abordée, ainsi que quelques uns des nombreux travaux réalisés dans ce domaine.

Chapitre 3

Méthodes non-supervisées de classification automatique

Dans ce chapitre, nous présentons un panorama des méthodes non-supervisées de classification automatique.

Après avoir précisé la problématique générale du domaine, nous décrivons les approches supervisées et non-supervisées. Cette description met l'accent sur les points qui, rapportés à notre problème spécifique, se révéleront essentiels.

Différentes classes de méthodes sont ensuite présentées, d'abord dans leur principe puis, de façon plus précise, par quelques algorithmes classiques.

Enfin, nous discutons du potentiel de ces techniques pour la classification aveugle de bruits impulsifs de formes et de nombre inconnus.

3.1 Généralités

3.1.1 Problématique

Classifier désigne à la fois l'action de partitionner un ensemble d'objets en groupes homogènes et celle de donner un attribut à un objet. Dans le premier cas, il s'agit de regrouper des éléments et dans le second de les identifier à une étiquette. Cette ambiguïté du mot correspond à une ambiguïté de l'action qui lui est associée : regrouper des objets en classes, c'est donner l'attribut "appartient à la classe i" à l'objet, tandis que donner à chaque élément une étiquette, c'est regrouper les éléments en classes d'objets semblablement identifiés.

Les deux sens du mot se distinguent par le procédé qui conduit à la classification (regroupement ou identification) plutôt que par leur résultat commun, qui est la classification elle-même. Par ailleurs, l'idée sous-jacente qu'un sens peut être associé à une étiquette conduit à considérer l'identification comme un diagnostic tandis que regroupement et interprétation sont clairement dissociés ([Kendall 66], [Jardine 66]).

Ces distinctions restent cependant assez théoriques. Dans la pratique, classifier fait partie des facultés qu'un homme possède et utilise spontanément, sans efforts, et souvent sans savoir précisement quelle information est exploitée ni les mécanismes qui sont mis en jeu. Pourtant, la perception de similarités entre des phénomènes physiques d'une grande variabilité met en jeu des processus d'autant plus difficiles à imiter qu'ils sont mal connus : identifier une voix, reconnaître un visage ou distinguer deux symboles sont des facultés pour nous évidentes, mais dont la simulation pose des problèmes extrêmement complexes.

Le premier d'entre eux est de trouver une représentation des objets à classifier permettant un codage numérique de l'information utile. Ensuite, une mesure de similarité est nécessaire à la comparaison des objets entre eux et, éventuellement, à la comparaison des classes entre elles. D'une certaine façon, la classification d'un ensemble de données est donc un problème de recherche d'une partition minimisant une fonction de coût. C'est principalement par l'algorithme d'optimisation et par la fonction d'énergie que les algorithmes se distinguent. Deux approches très différentes peuvent être dissociées, en fonction de la méthode par laquelle les classifieurs sont paramétrés. Il s'agit des approches supervisées ou non-supervisées.

3.1.2 Méthodes supervisées

Le premier type de méthodes consiste à paramétrer un classifieur à partir d'un ensemble d'exemples et de contre exemples. Les règles de classification apprises sur ce corpus sont ensuite généralisées à l'ensemble des données. Ces techniques peuvent procéder par modélisation ou par discrimination des classes.

Les méthodes discriminantes recherchent une fonction de décision qui maximise le contraste entre les classes. Les paramètres minimaux qui permettent de distinguer chaque classe de ses concurrentes sont isolés dans le corpus d'apprentissage. Une classe est ensuite caractérisée relativement aux autres, par les frontières qui la séparent des autres classes.

Les algorithmes qui procèdent par modélisation construisent, à partir des données d'apprentissage, une image type des classes. Chaque donnée sera ensuite affectée à la classe qui lui sera la plus proche, au sens de la mesure de similitude adoptée. Contrairement aux méthodes discriminantes, les algorithmes de modélisation extraient les caractéristiques communes à tous les individus d'une classe.

Ces techniques sont dites supervisées, parce qu'elles sont guidées par un ensemble d'exemples fournis généralement par un expert humain. Elles correspondent à une classification par identification : les classes sont décrites dans un corpus d'apprentissage où elles sont associées à une étiquette qui permet leur interprétation. Dans cette approche, la connaissance est implicitement apportée au classifieur par l'opérateur qui a collecté — ou produit — les données d'apprentissage.

Cet aspect des techniques supervisées constitue un intérêt majeur. En effet, le fait que l'expert humain puisse réaliser une tâche de classification ne veut pas dire qu'il est capable de décrire la stratégie qui a été employée. Un classifieur supervisé pourra parfois reproduire ce comportement, y compris dans son côté *aveugle*. Cependant, ce mimétisme a des limites qui sont liées à la nature des méthodes elles-mêmes : cherchant à obtenir des résultats identiques par un processus différent, on obtient souvent des résultats différents. Ainsi, les performances de ces techniques sont susceptibles de dépasser celles d'un humain, par exemple parce qu'elles peuvent exploiter des ressources éventuellement différentes. Le cas des systèmes de traitement de la parole est typique de cette situation : les algorithmes de classification de lettres de l'alphabet, ou des reconnaissance des chiffres, obtiennent de meilleurs résultats que l'opérateur humain dans des conditions bien contrôlées. Cela n'est malheureusement pas le cas général en reconnaissance de la parole (par exemple pour le traitement de la parole spontanée ou la vérification du locuteur).

Une difficulté commune aux méthodes de classification supervisées est l'importance du choix du corpus d'apprentissage. L'apprentissage reposant uniquement sur les exemples et les contre exemples du corpus d'entraînement, toutes les régularités sont susceptibles d'être généralisées. Un corpus peu représentatif risque de conduire à des généralisations erronées qui fausseront ensuite la classification. L'opérateur doit donc fournir un corpus contenant toute l'information utile à la classification, bien qu'il ne sache pas exactement quelle est l'information que lui-même exploite pour traiter les données. Une solution consiste à augmenter la taille des corpus jusqu'à obtenir une sur-représentation des classes qui exclue la présence de toute régularité factice dans les corpus. Des méthodes fondées sur une amélioration de la qualité des corpus, ou encore sur une correction *a posteriori* des paramètres du classifieur ont été proposées ([Quelavoine 97]).

Cependant, la composition du corpus d'apprentissage d'un classifieur supervisé ne pose pas seulement un problème de qualité ou de quantité des exemples. Fixer un ensemble fini de données représentatives suppose que cet ensemble existe. Le nombre de classes, leur forme, leur variabilité, tous ces paramètres sont déterminés définitivement pendant la phase de paramétrisation du classifieur. L'hypothèse que les données futures se conformeront au schéma arrêté pendant l'apprentissage est valide si celles ci comportent un haut degré de régularité. Cette contrainte, qui peut être satisfaite dans des environnements stables, devient problématique dans le cas d'environnements réels qui sont, par nature, fortement variables et imprévisibles.

De très nombreux algorithmes de classification supervisée issus de domaine de recherche variés ont été développés : méthodes basées sur les statistiques bayésiennes ou l'analyse de données, modèles neuromimétiques inspirés de la structure des cerveaux biologiques, techniques de minimisation d'une fonction d'énergie, etc. L'application de ces techniques de classification dans des conditions bien contrôlées a montré leur efficacité, notamment dans le domaine de la reconnaisance de formes visuelles ou acoustiques.

3.1.3 Méthodes non-supervisées

L'autre catégorie de méthodes est dite non-supervisée parce que, à l'inverse des techniques supervisées, leur comportement n'est pas dicté par un opérateur qui décrit, par l'exemple, la tâche à effectuer. Cette approche repose sur l'hypothèse qu'une structure sous-jacente existe dans les données, et que cette organisation peut être révélée par l'étude de leurs similitudes et de leurs différences. L'objectif est d'effectuer une synthèse des données minimisant la perte d'information. Les perspectives ouvertes par une telle approche sont multiples.

D'une part, le nombre et la variété des données peuvent gêner leur analyse. Construire automatiquement une image concise et structurée de ces données facilite leur interprétation et peut constituer une étape préalable à un traitement de plus haut niveau. Dans un tel processus, l'absence de biais introduit par l'intervention de l'opérateur humain peut permettre la mise en évidence de propriétés auparavant inconnues. Contrairement aux méthodes supervisées dont l'objectif est d'appliquer une connaissance qui n'est pas formulée explicitement, ces techniques sont susceptibles de générer une connaissance qui est faiblement orientée par l'expert. Par ailleurs, ces méthodes présentent l'intérêt de limiter les hypothèses sur la nature des données. Ne s'appuyant pas sur une connaissance définitive, elles sont généralement douées d'une grande capacité d'adaptation. Le traitement de données dont la diversité est difficile à cerner se trouve facilité par cet aspect ouvert des systèmes non-supervisés. Cependant, l'idée séduisante d'une méthode de classification affranchie de l'influence d'une connaissance explicitement ou implicitement énoncée doit être tempérée : généralement, ces méthodes sont paramétrées par le choix d'une distance, celui d'une fonction de coût, ou encore, plus directement, par le choix de seuils qui déterminent le comportement du classifieur. Par ailleurs, les performances d'un classifieur non-supervisé sont aussi dépendantes non pas d'un ensemble d'exemples étiquetés, mais des données qui ont permis et conditionné l'apprentissage.

La capacité d'adaptation de ces méthodes à des configurations variables et imprévisibles des données n'en reste pas moins supérieure à celle des méthodes supervisées. Cette raison, conjuguée à la difficulté d'établir un corpus représentatif des formes acoustiques composant un environnement réel, nous a incité à choisir cette approche plutôt que la précédente.

Les principes de fonctionnement de ce type d'algorithmes de classification sont très divers, notamment parce que la problématique abordée renvoie à des domaines de recherche variés. Leur objectif commun est de partitionner l'ensemble des données en classes, de façon à obtenir le maximum de distance entre des classes aussi homogènes que possible. Il s'agit par conséquent d'une classification par regroupement des éléments en classes plutôt que par identification.

La section suivante décrit les principales méthodes de classification non supervisées. Différents algorithmes son regroupés en trois classes qui ne sont pas forcément disjointes : les méthodes issues de l'analyse de données, les algorithmes de quantification vectorielle et les mémoires associatives. Bien que présentées par son auteur comme un modèle supervisé, les mémoires associatives de Kohonen sont décrites à la fin de ce chapitre.

3.2 Méthodes issues de l'analyse de données

3.2.1 Centres mobiles

Cette méthode suppose connu le nombre de classes de la partition recherchée. n individus choisis aléatoirement forment les n centres initiaux des classes. Les données sont ensuite regroupées autour de ces centres provisoires, chacune d'elles étant affectée à la classe dont le centre lui est le plus proche. Cette opération revient à diviser l'espace des données en n zones séparées par les hyperplans médiants des centres (cf. figure 3.1). Les centres de gravité des classes ainsi formées constituent les centres de nouvelles classes qui regroupent les éléments en fonction de leur distance aux nouveaux centres. Ce processus est itéré tant que l'inertie intra-classe diminue.

Cette méthode extrêmement simple présente plusieurs défauts importants. D'une part, les surfaces de séparation des classes sont des hyperplans. Il s'agit donc d'une séparation linéaire, ce qui peut ne pas convenir à certains types de problèmes. Par ailleurs, la partition obtenue est très dépendante du choix des centres initiaux. Enfin, la métrique utilisée est élémentaire : deux éléments sont regroupés dans la même classe s'ils sont plus proches de son centre que de celui de toute autre classe.

3.2.2 Méthode des nuées dynamiques

Introduite par [Diday 71] et [Lloyd 82] la méthode des nuées dynamiques est une généralisation de la méthode des centre mobiles. Cet algorithme décrit une classe non pas par son centre, mais par un noyau qui peut être composé de plusieurs individus. Ce groupe caractérise une classe mieux que ne le ferait un unique individu. Par ailleurs, l'interprétation éventuelle des classes est améliorée par leur description plus

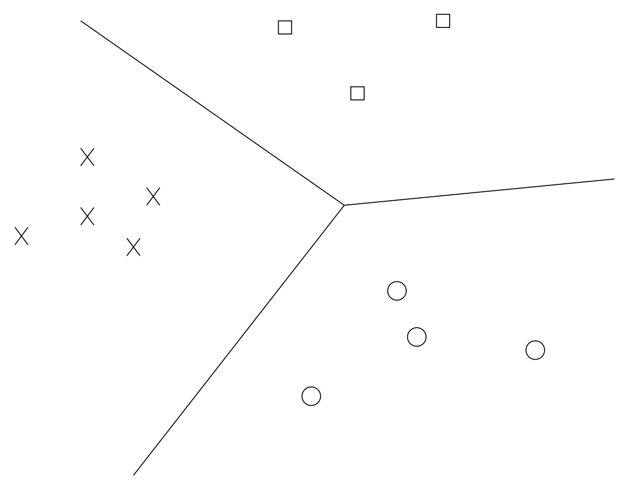

FIG. 3.1: *Partitionnement des données en 3 classes par la méthode des centres mobiles : les classes sont séparées deux à deux par les médiatrices des centres. Les courbes de séparation sont formées par les segments connexes des médiatrices.*

complète. Cette méthode admet de nombreuses variantes qui diffèrent par le choix de trois mesures qui caractérisent l'algorithme :

- une mesure de la proximité d'un individu à un groupe,
- une méthode de construction des noyaux à partir des classes,
- une mesure de la qualité d'une partition.

Une fois ces choix arrêtés, le principe de fonctionnement de l'algorithme est identique à celui des centres mobiles : chaque partition permet de réorganiser les noyaux qui permettent, à leur tour, de recomposer une partition. Cette opération est répétée jusqu'à ce que la qualité des partitions formées ne s'améliore plus.

Bien que cette méthode réalise une séparation linéaire des classes (au sens de la distance adoptée), son pouvoir de représentation des données peut être amélioré par le choix d'une distance apropriée. Elle garde, du moins dans sa version initiale, l'inconvénient d'un nombre de classes fixé *a priori*.

3.2.3 Classification hiérarchique

Contrairement aux deux méthodes décrites plus haut, le nombre de classes n'est pas un paramètre fixe des classifieurs hiérarchiques. Ces techniques composent une suite de partitions emboîtées des données. Chaque classe d'une partition intermédiaire contient une ou plusieurs classes de niveau inférieur, et est contenue dans une classe de niveau supérieur.

La construction de l'arbre de classification peut se faire de façon ascendante ou descendante. Le mode ascendant consiste à initialiser une partition très précise avec un grand nombre de classes élémentaires (par exemple contenant un seul élément), puis à itérer un processus de réunion des classes qui construit un niveau à partir du niveau inférieur. Le mode descendant consiste à fractionner les classes d'un niveau pour obtenir le niveau inférieur. Dans ce cas, la classe initiale est composée de toutes les données à classifier.

Un point caractéristique de ce type de techniques est le critère utilisé pour déterminer les classes qui doivent être fusionnées ou segmentées. La méthode naturelle consiste à regrouper les classes qui sont les plus proches. Encore faut-il disposer d'une mesure de proximité des classes. Ward propose une méthode ascendante dont le critère de fusion repose sur la minimisation de l'inertie intra-classe : les classes regroupées sont celles pour lesquelles la perte d'inertie est minimale. La figure 3.2 illustre l'organisation hiérarchique des données dans un arbre de classification. On voit que le choix d'un niveau (par exemple à l'aide d'un seuil λ sur l'inertie des classes) permet de décider, si c'est nécessaire, de la partition finalement retenue.

3.3 Quantification vectorielle

3.3.1 Principe

Les méthodes de quantification vectorielle associent à un ensemble donné de M vecteurs $(X_i)_{1<i<M}$, un dictionnaire de taille réduite, composé de N vecteurs prototypes $(P_i)_{1<i<N<M}$. Elle se caractérisent par la recherche d'un faible nombre de vecteurs représentant optimalement les données plutôt que par celle d'une partition de l'ensemble des données en classes. Les méthodes des centres mobiles et celle des nuées dynamiques peuvent donc être considérées comme des méthodes de quantification vectorielle. On trouve dans [Gersho 92], un détail des principes et des algorithmes de ce type.

L'objectif de ces techniques est d'optimiser la représentation des données X_i par les prototypes P_i. Pour cela, deux critères peuvent permettre d'évaluer la qualité

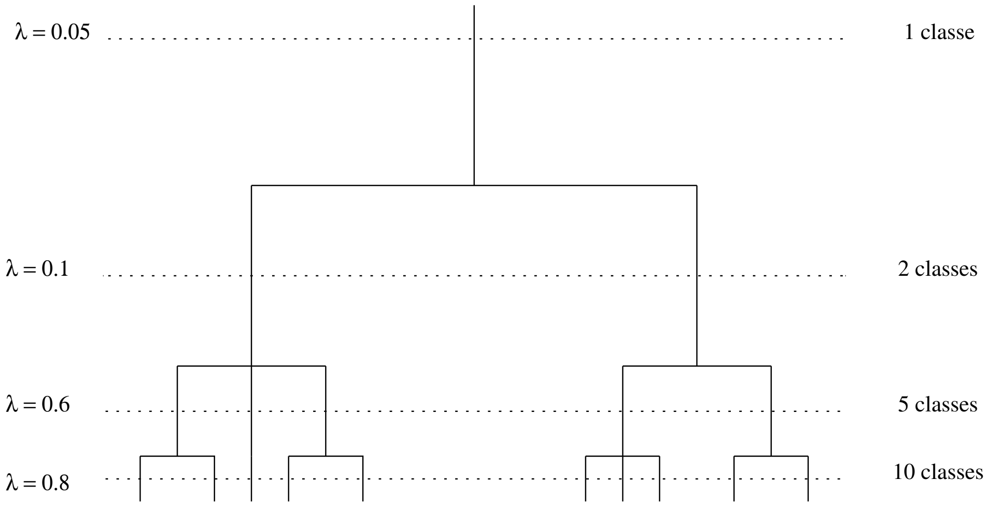

FIG. 3.2: *Arbre de classification hiérarchique : à chaque niveau de l'arbre, correspond une partition de l'ensemble des données. Une partition peut être obtenue par regroupement des classes de niveau inférieur, où par segmentation des classes de niveau supérieur. Le paramètre λ caractérise, dans l'arbre de classification, un niveau qui correspond à une partition de l'ensemble des données.*

d'un dictionnaire : une mesure de complexité, qui doit être minimisée, et un indice de distorsion qui correspond à la qualité de représentation des données par les prototypes.

Ces techniques ont été développées dans des perspectives diverses : compression et caractérisation de données, classification, filtrage de signaux bruités, etc.

Nous présentons ici deux algorithmes clefs qui sont typiques des différentes techniques developpées dans la littérature.

3.3.2 Quantification vectorielle par maximisation d'entropie

Cet algorithme, proposé par [Buhmann 92b] s'inspire de la physique des particules. Il est basé sur la minimisation d'une *énergie libre* qui est à la fois fonction de la complexité du dictionnaire et d'une mesure de distorsion de la représentation. Le nombre de classes est exclusivement fonction du minimum obtenu par l'algorithme d'optimisation.

La distorsion $D_{i,j}$ est la distance d'une donnée X_i à un prototype P_j, pour un entier positif n :

$$D_{i,j} = \|X_i - P_j\|^n$$

On en déduit la distorsion totale associée au dictionnaire B composé de N prototypes :

$$D(B) = \sum_{i=1}^{N} \sum_{j=1}^{M} D_{i,j}$$

où M est le nombre de vecteurs à quantifier.

Buhmann estime la complexité C du dictionnaire à l'aide de son entropie et de la probabilité p_i d'occupation de la classe i :

$$C_i = -ln(p_i)$$

$$C = \sum_{i=1}^{N} p_i C_i$$

D'autres fonctions de complexité peuvent être utilisées, en fonction de l'application visée ou d'un comportement particulier attendu du classifeur.

La fonction de coût $L(B)$ (associée au dictionnaire B) proposée par Buhmann permet de pondérer par un paramètre γ l'importance relative de la complexité et de la distorsion :

$$L(B) = \sum_{i=1}^{M} \sum_{j=1}^{N} p_{i,j}.(D_{i,j} + \gamma C_j)$$

où $p_{i,j}$ est la probabilité que le vecteur i soit quantifié par le prototype j. Le choix du paramètre γ détermine un compromis entre nombre de prototypes et qualité de la représentation des données par le dictionnaire.

Les probabilités $p_{i,j}$ sont estimées par minimisation d'une fonction de coût assimilable à une énergie libre. Un processus d'optimisation permet ensuite d'obtenir un dictionnaire B correspondant à un minimum local de la fonction de coût $L(B)$ ([Buhmann 92b]).

Cette méthode a l'avantage d'être facilement adaptable en fonction de l'application envisagée : la fonction de complexité et la mesure de distorsion utilisées par Bhumann et Kuhnel pour la compression de données peuvent être remplacées par d'autres, plus conformes aux objectifs souhaités ([Pican 95]). Par ailleurs, une version connexionniste de cet algorithme a été proposée par [Buhmann 92a].

Néanmoins, cet algorithme est fondé sur un principe de recherche d'une partition optimale de l'ensemble des données. Nous avons vu qu'une des particularités du problème abordé ici était le mélange éventuel des formes acoustiques perçues par le capteur. La possibilité qu'une donnée puisse résulter de la superposition d'éléments de catégories différentes est difficilement compatible avec ce principe de partionnement en classes. L'adaptation de cette méthode à cette spécificité de notre problème supposerait de reconsidérer la façon dont les probabilités *a posteriori* sont estimées, la fonction de complexité, l'algorithme d'optimisation, etc. Par ailleurs, la classification est déterminée à partir d'une analyse globale des données. Ce mode de traitement ne correspond pas au traitement dynamique et séquentiel que nous envisageons.

3.3.3 Cartes auto-organisatrices de Kohonen

Ce modèle s'inspire du fonctionnement du cortex des vertébrés. Il est basé sur le principe d'une quantification vectorielle préservant la topologie naturelle des données. Cet aspect des cartes topologiques permet d'exploiter le fait que la distance entre stimuli peut porter une information utile au codage des données dans le réseau. En effet, en conservant les relations de voisinage des stimuli, le réseau augmente sa robustesse à une certaine variabilité des classes qui sont caractérisées par une zone d'activation plutôt que par un ou quelques neurones donnés.

Les cartes de Kohonen sont constituées d'une couche de neurones qui sont répartis sur une ligne ou sur une surface plane. Les cellules possèdent deux types de connexions :

- des connexions latérales, dont les valeurs sont fixes,
- des connexions à l'extérieur, qui pondèrent les coefficients du stimulus par des valeurs W_{ij}.

La réponse du réseau à un stimulus $V = (v_i)_{i=1,N}$ se fait en deux phases, qui correspondent aux deux types de liaisons du réseau. Dans la première étape, chaque cellule y_i est excitée par l'ensemble des coefficients du vecteur stimulus selon une fonction de transfert linéaire :

$$y_i = \sum_j W_{ij}.v_j$$

Il s'agit d'une fonction de transfert identique à celle des mémoires associatives linéaires ([Widrow 60] [Rosenblatt 62] [Kohonen 89]).

La deuxième phase consiste à intégrer des relations de proximité entre cellules : les connexions latérales propagent l'activité d'une cellule dans son voisinage immédiat et inhibent l'activité des cellules plus lointaines. Les interactions de cellules très éloignées sont faibles. La figure 3.3 montre la fonction pondérant l'influence d'une cellule sur ses voisines en fonction de leur distance. Les connexions latérales sont récurrentes. Après activation directe des cellules par le stimulus, elles interagissent jusqu'à stabilisation de leurs activités. Cette phase conduit à une augmentation du contraste sur la carte et permet une certaine tolérance à la variabilité du stimulus. En effet, les légers décalages du stimulus sur la carte vont se traduire, par l'effet de la fonction d'activation latérale en chapeau mexicain, par une excitation d'un groupe de cellules adjacentes.

Les liens latéraux ne sont cependant pas modifiables. La largeur de la zone d'influence d'une cellule sur la carte est donc fixée *a priori*.

La dynamique des connexions directes est réglée par l'équation :

$$\Delta W_{ij} = \alpha . y_i . (v_j - Wij)$$

On peut observer que l'ampleur des modifications est proportionnelle à l'activité de la cellule et que l'action de cet ajustement est de renforcer cette activation. Globalement, ce processus de modification des poids synaptiques conduit à augmenter la sensibilité de la cellule la plus réactive au stimulus et correspond à une règle d'apprentissage Hebbienne ([Hebb 49] : le lien des cellules simultanément actives doit être renforcé).

Lorsque la proximité des coefficients d'un vecteur d'entrée correspond à une proximité naturelle des composantes du stimulus, la capacité des cartes de Kohonen à préserver cette topologie des entrées peut constituer un intérêt capital. Ceci est le cas, par exemple, des images numérisées dont la position géographique des pixels porte une information caractéristique de l'image : en réarrangeant les points de l'image, on perdrait sa cohérence globale. C'est aussi le cas des images spectrales d'un son, dont la proximité traduit une similarité temporelle ou fréquentielle.

Néanmoins, cet algorithme semble difficilement utilisable pour la classification de mixtures de données. En effet, son apprentissage est basé sur l'augmentation du contraste des réponses. Ce processus doit conduire à améliorer la sélectivité des cellules. Mais cette compétition n'a plus de sens lorsque le stimulus est composé de plusieurs données mélangées : la sélectivité forcée des neurones conduira sans doute à l'excitation de l'un d'entre eux, et à l'inhibition des autres. Or l'ambiguïté des

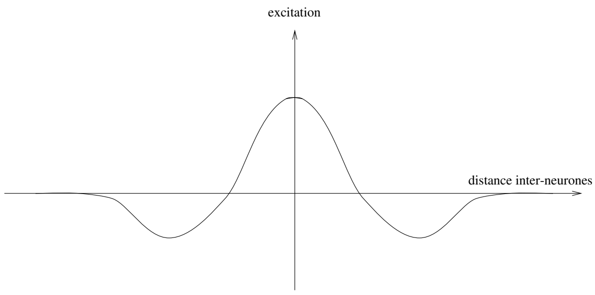

FIG. 3.3: *Fonction d'activation latérale en chapeau mexicain : l'excitation d'une cellule provoque l'excitation de ses voisines, et l'inhibition des cellules plus lointaines. La taille de la zone d'influence d'une cellule sur la carte est néanmoins limitée.*

réponses peut signaler l'ambiguïté naturelle d'un stimulus résultant du mélange de plusieurs composantes. D'une façon générale, ceci constitue un obstacle majeur à l'utilisation d'algorithmes compétitifs pour la classification de mixtures de données.

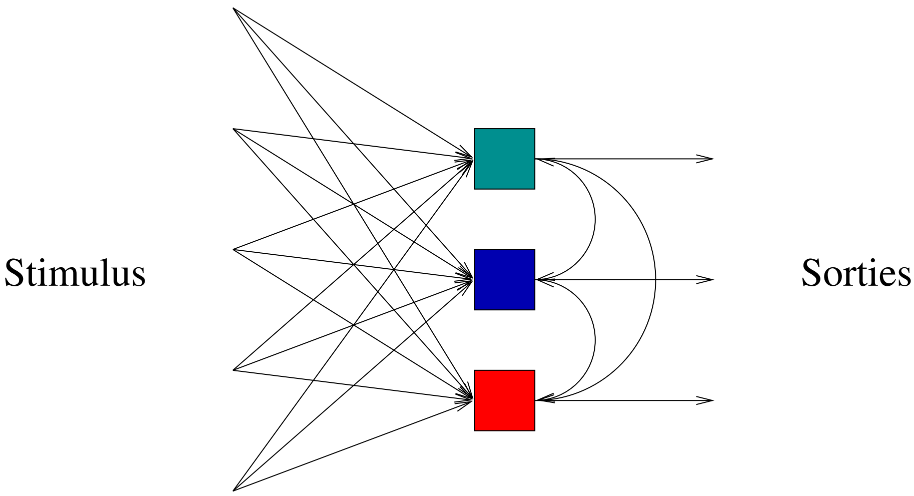

FIG. 3.4: *Architecture des cartes auto-organisatrices de Kohonen : chaque cellule est connectée à l'extérieur par des liaisons valuées. De plus, toutes les cellules sont connectées entre elles, les valeurs associées à ces liens étant déterminées par une fonction de proximité du type chapeau mexicain.*

3.4 Mémoires associatives

3.4.1 Principe

La mémoire d'un ordinateur est composée d'un ensemble fini d'unités capables de stocker une quantité limitée d'information. Ces unités élémentaires sont assimilables à des lieux *géographiques* qui sont accessibles par leur adresse. La gestion de la mémoire consiste à stocker l'information au bon endroit, à maintenir son intégrité, et, finalement, à y accéder. Bien que le fonctionnement des mémoires biologiques soit mal connu, on peut penser que cette "métaphore de l'esprit bibliothèque" ne correspond pas à la façon dont l'information est codée puis restituée par un cerveau biologique. D'abord, les capacités mnémoniques ne sont pas indépendantes des autres facultés cérébrales. Le rappel d'une information ne se fait pas par l'usage d'une clef indépendante de l'information elle-même, mais plutôt par des indices qui lui sont liés. La mémoire est donc adressable par son contenu. D'autre part, la destruction d'une partie du support physique de la mémoire se traduit par l'atténuation des concepts mémorisés plutôt que par leur disparition totale. Enfin, si des zones du cerveau sont spécialisées dans des tâches particulières, sa structure est massivement parallèle et l'information y est distribuée.

Ces observations ont conduit au développement de modèles neuromimétiques capables d'associer, à une version bruitée ou incomplète d'une forme mémorisée, la forme originelle. Ces modèles de mémoires associatives sont donc des mémoires adressables par leur contenu. Les premiers algorithmes de ce type ont été introduits dans les années 1960 par [Widrow 60] [Rosenblatt 62], suivis de [Anderson 84] et de [Kohonen 89].

Nous présentons, dans cette section, deux modèles : celui de Hopfield, basé sur une analogie avec la physique statistique et le modèle de Kohonen.

3.4.2 Modèle de Hopfield

Cette architecture a été introduite par [Hopfield 82]. Elle s'inspire de la physique statistique, en particulier de lois d'interaction entre particules orientées.

Il s'agit d'un réseau de neurones totalement interconnectés (cf. figure 3.5). Contrairement aux modèles *feedforward*, qui sont traversés par un stimulus qu'ils transforment, les réseaux de Hopfield sont globalement initialisés par une stimulation qui est perçue par toutes les cellules. La réponse est obtenue après une phase de relaxation, par convergence globale du réseau vers un état stable correspondant au minimum d'une fonction d'énergie.

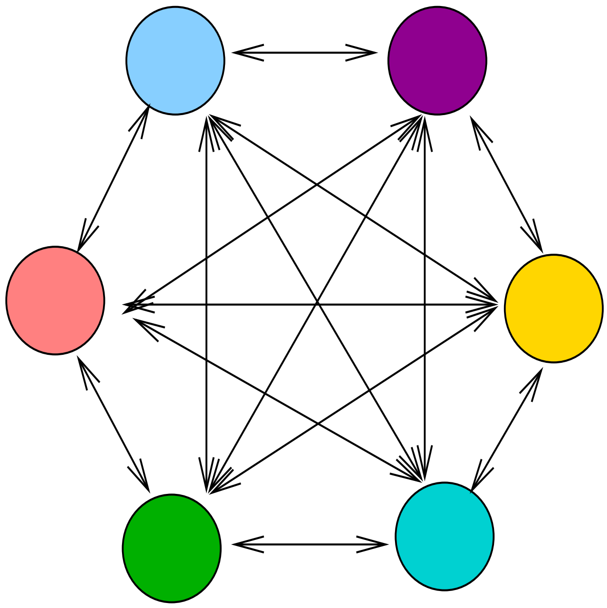

FIG. 3.5: *Architecture du modèle de Hopfield : toutes les cellules sont interconnectées par des liaisons valuées. Il n'y a pas de direction privilégiée : les valeurs des cellules sont toutes initialisées simultanément, puis modifiées séquentiellement ou simultanément.*

Chaque cellule du réseau de Hopfield peut prendre deux états, 1 ou -1 (qui correspondent à l'orientation de la particule dans le modèle physique analogue). Les connexions valuées sont bi-directionnelles et symétriques. À chaque instant, l'état du réseau est donné par les valeurs de ses nœuds.

L'état d'une cellule i est calculé en fonction de l'activité des autres cellules V_j et des poids synaptiques W_{ij} par une règle assez simple :

- si $\sum_j W_{ij}.V_j > 0$, alors $V_i = 1$,
- sinon $V_i = -1$.

L'état final du réseau est obtenu en appliquant cette règle sur les cellules, dans un ordre éventuellement aléatoire, jusqu'à convergence globale du réseau.

Considérons la fonction d'énergie $E(V)$:

$$E(V) = \frac{-1}{2} V^t W^t V$$

où $V = (v_i)$ est le vecteur associé à l'état global du réseau et W la matrice des poids du réseau.

Le processus séquentiel de mise à jour des cellules se fait à énergie décroissante. La stabilisation se produit lorsque le réseau se trouve dans un minimum local de la fonction d'énergie. En choisissant de façon adéquate la matrice des poids, on peut obtenir une fonction d'énergie dont les minima locaux correspondent à un vecteur d'état particulier.

Cette utilisation du modèle de Hopfield permet d'associer à tout état initial une forme de référence codée dans la matrice des poids du réseau. Un stimulus proche d'un état stable conduira à la convergence vers cet état de référence. On obtient alors une mémoire associative dont les références sont codées explicitement dans les liens synaptiques.

Cependant, s'il est possible de déterminer une matrice des poids dont la fonction d'énergie est localement minimale pour des états choisis *a priori*, la capacité de mémorisation de cette structure est limitée. Cela se traduit par l'existence de minima locaux parasites. [Peretto 84] évalue qu'en choisissant au plus $0,15.N$ vecteurs à mémoriser dans un réseau de taille N, on limite le risque d'avoir des minima indésirables dans la fonction d'énergie.

Indépendament de cette faible capacité de mémorisation des réseaux de Hopfield, le mode de paramétrisation de ce classifieur limite sa plasticité, et semble mal adapté au traitement en ligne qui est envisagé.

3.4.3 Mémoires associatives de Kohonen

Algorithme élémentaire

Il s'agit d'un réseau à une couche, connecté à l'extérieur par des liaisons valuées. A chaque cellule, est associée une valeur réelle. La propagation d'un stimulus dans le réseau est faite par une fonction de transfert linéaire. L'activité y_i d'une cellule i excitée par un stimulus x est obtenue par :

$$y_i = \sum_j W_{ij} x_j$$

où W_{ij} est le lien de la cellule i à la cellule j. W est donc la matrice des poids du réseau.

La première règle d'apprentissage proposée par Kohonen est supervisée : des couples de vecteurs (X, Y) devant être associés sont présentés au réseau lors d'une

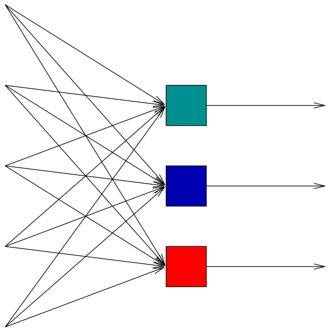

FIG. 3.6: *Mémoires associatives de Kohonen : il s'agit d'une architecture à une dimension qui est traversée par un stimulus qu'elle transforme en fonction d'une règle linéaire.*

phase d'apprentissage durant laquelle les connexions sont modifiées suivant la règle :

$$\delta W_{ij} = y_i.x_j$$

L'apprentissage consiste à exécuter cette opération pour chacun des N couples (x^k, y^k) qui doivent être associés. On obtient alors la matrice des poids correspondant à un corpus d'aprentissage composé :

$$W_{ij} = \sum_{k=1}^{N} y_i^k.x_j^k$$

Une fois cette paramétrisation du modèle achevée, les poids du réseau sont définitivement fixés. La réponse y à un stimulus x est obtenue par la fonction de transfert (3.1), qui peut être exprimée en fonction de l'ensemble d'apprentissage par :

$$y_i = (W.x)_i = (\sum_{k=1}^{N} y_i^k.x_j^k).x_j$$

$$y_i = (\sum_{k=1}^{N} y_i^k.x_j^k).x_j$$

Le terme $x^k.x$ est un produit scalaire qui peut constituer une mesure de similarité entre le vecteur stimulus x et la *clef* mémorisée x^k. Si les vecteurs x^k sont orthogonaux, alors le réseau renvoie un vecteur d'autant plus proche du prototype mémorisé que le stimulus est proche de la clef correspondante. Cette propriété permet le rappel des formes de références qui ont été codées dans la matrice des poids pendant l'apprentissage.

Une variante de ce modèle consiste à confondre les clefs x^k et les prototypes y^k. La mémoire est, dans ce cas, auto-associative : un exemple d'aprentissage est associé à lui-même dans la perspective de retrouver, à partir d'une version bruitée ou incomplète d'un prototype, le vecteur mémorisé.

Avec ce réseau auto-associatif, Kohonen met en œuvre un processus d'accès à l'information qui relève d'un principe biologiquement plausible d'adressage de la mémoire par le contenu. Malheureusement, l'algorithme proposé comporte des défauts importants. D'une part, sa capacité de mémorisation est faible puisqu'on a, au mieux, autant de prototypes que de cellules. Les prototypes étant orthogonaux, ce nombre est majoré par la dimension de l'espace de représentation. De plus, l'algorithme n'effectue qu'une séparation linéaire des classes formées par les versions bruitées des prototypes.

Kohonen propose de contourner la contrainte d'orthogonalité par un pré-processus d'orthogonalisation par la méthode de Graham-Schmit, puis par une extension du modèle initial basée sur une généralisation de la notion d'inversion de matrice. C'est ce modèle que nous décrivons dans la section suivante.

Mémoires associatives optimales

Cet algorithme est une extension des mémoires associatives simples au cas non-orthogonal.

Le principe est de chercher la matrice W qui minimise l'erreur moyenne de rappel, qui est la distance entre l'état du réseau après la présentation d'un stimulus x^k et l'état souhaité qui est défini par le vecteur y^k. L'erreur de rappel E s'exprime en fonction de la matrice des poids W par :

$$E = \sum_{k=1}^{N} (y^k - W.x^k).(y^k - W.x^k)^t$$

La matrice W minimisant ce critère est le produit de la matrice des prototypes Y et de la pseudo-inverse X^+ de la matrice des clefs X :

$$W = Y.X^+$$

X est la matrice dont chaque ligne est un vecteur clef. Sa pseudo-inverse X^+ peut être calculée explicitement, ou par un algorithme itératif (par exemple par l'algorithme de Gréville ou encore celui de projection du Gradient qui sont décrits dans [Kohonen 89]).

La contrainte d'orthogonalité des clefs est donc ramenée à celle, plus faible, d'indépendance linéaire. Néanmoins, la version optimale des mémoires associatives de Kohonen conserve le défaut d'une séparation linéaire des classes. Ses performances en classification automatique sont donc inférieures aux techniques de classification supervisées classiques (comme les perceptrons muticouches).

Un aspect des mémoires associatives linéaires a été peu développé dans la littérature. C'est leur sensibilité à la nouveauté : la composante d'un stimulus qui se trouve hors de l'espace défini par les prototypes mémorisés ne suscitera aucun rappel. Kohonen suggère l'utilisation de cette propriété pour la détection de nouveautés. Nous verrons, dans le chapitre suivant, de quelle manière nous avons exploité cette sensibilité des mémoires associatives à la présence d'une forme qui ne peut pas être associée à un prototype.

3.5 ART : Théorie de la résonance adaptative

Le modèle ART (Adaptative Resonance Theory) a été introduit puis développé par [Grossberg 76]. Il s'inspire d'observations biologiques et comportementales. L'objectif de Grossberg était de trouver des mécanismes susceptibles de résoudre le problème de plasticité et de stabilité des réseaux auto-organisateurs. Plus précisément, il s'agit du compromis à établir entre la capacité d'adaptation d'un réseau (sa plasticité) et sa stabilité. En effet, une trop forte stabilité du réseau empêche l'évolution des classes, tandis qu'une trop forte plasticité provoque un traitement à court terme dont l'instabilité nuit à la qualité de la classification produite. Plusieurs réseaux ART ont été développés. Le premier, ART-1, est utilisé pour la classification de formes à composantes binaires. ART-2 [Carpenter 76] est l'extension de ce modèle à la classification de vecteurs à composantes réelles.

Nous allons décrire seulement le fonctionnement de ART-1 (cf. figure 3.7), ART-2 étant fondé sur le même principe. ART-1 est composé de deux couches, chacune d'elles effectuant des opérations différentes au cours des divers cycles de traitement. La couche d'entrée reçoit le vecteur à classer, le transmet à la seconde couche à travers une série de connexions valuées qui constituent le filtre adaptatif descendant. La seconde couche dont chaque nœud correspond à une classe de vecteurs, accentue la réponse du nœud le plus actif. Elle renvoie ensuite le vecteur d'activité à la première couche à travers le filtre adaptatif ascendant. Cette couche joue alors le rôle

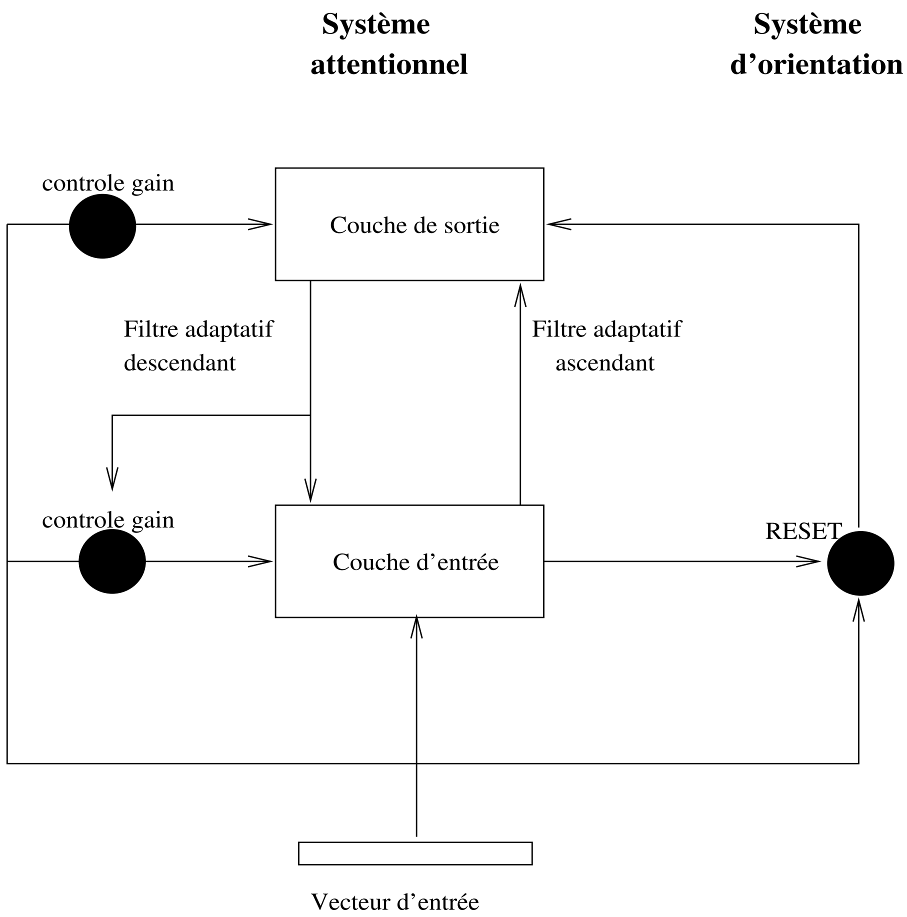

FIG. 3.7: *Architecture de ART-1 : deux couches sont reliées par un filtre ascendant et un filtre descendant. Le stimulus oscille entre l'une et l'autre, la couche de sortie constiuant finalement la couche classifiante.*

de couche de comparaison, qui permet de tester l'hypothèse d'appartenance à une classe. Si cette hypothèse est validée, les prototypes sont adaptés au nouveau stimulus sinon, une phase de recherche est déclenchée. L'hypothèse émise est, dans ce cas, remise en cause et d'autres sont testées. Si aucune d'entre elles n'est satisfaisante, alors une nouvelle classe est créée. Ce type de réseau est donc capable d'apprendre une classification par un algorithme incrémental, de façon totalement automatique. Cependant, le seuil à partir duquel une hypothèse d'appartenance à une classe est retenue est fixé *a priori* (S. Grossberg appelle ce paramètre *seuil de vigilance*). Par ailleurs, la classification produite s'adapte aux évolutions des stimuli, et est sensible à l'apparition d'éléments nouveaux.

Cette idée utilisée par Grossberg d'un processus de détection de nouveauté provoquant la mise à jour du classifieur nous a paru particulièrement intérressante. Elle peut permettre l'adaptation rapide du modèle à une évolution brutale du signal, provoquée par exemple par l'apparition d'une nouvelle catégorie d'événements acoustiques, ou encore par la modification brutale d'une des classes référencées. Cependant, l'utilisation de cet algorithme pour la modélisation d'un environnement de bruits transitoires poserait quelques problèmes importants.

D'une part, en implémentant un processus de création de classes, on suppose que ce nombre reste stable. Or la diversité des stimuli qui peuvent être perçus dans un milieu réel complexe n'est pas bornable *a priori*. La création d'un grand nombre de classes sans destruction de celles qui ne sont plus représentatives du signal conduirait sans doute à un engorgement du classifeur, et, à terme, à une perte de sensibilité ou encore à un taux de confusion élevé, provoqué par la proximité excessive des classes.

Un autre obstacle à l'utilisation de ce modèle pour la supervision d'environnements de bruit impulsifs est le comportement du classifieur dans des situations de superposition d'événements acoustiques. L'apprentissage compétitif du modèle conduirait sans doute à décider de l'appartenance d'un stimuli composite à une classe qui ne le représente que partiellement.

3.6 Classification aveugle d'un nombre inconnu de formes inconnues

Les catégories de méthodes décrites dans ce chapitre ne sont pas forcément disjointes. Elles correspondent à l'approche adoptée, mais aussi à un étiquetage qui dépend du contexte dans lequel une méthode est apparue, de la façon dont elle a été présentée et peut être de la manière dont la communauté a reçu et exploité la méthode. Ainsi, les états stables du modèle de Hopfield constituent un ensemble de prototypes de taille éventuellement réduite, ce qui correspond au principe de quantification vectorielle ; l'algorithme des nuées dynamiques est un algorithme de quantification vectorielle ; les cartes auto-organisatrices de Kohonen sont des mémoires associatives sur lesquelles on a implanté une contrainte de préservation de la topologie des entrées.

Néanmoins, les méthodes présentées dans ce chapitre sont basées sur des principes assez divers et, par conséquent, reposent sur des hypothèses qui sont elles aussi assez variées. Centres mobiles et nuées dynamiques supposent connu le nombre de classes composant le partitionnement. Le choix d'un seuil est nécessaire pour fixer le niveau de partionnement dans un arbre de classification. Dans le modèle de Hopfield, comme dans les mémoires associatives de Kohonen, les prototypes des classes sont connus.

La nécessité de disposer de ces connaissances préalablement à la paramétrisation des algorithmes représente, pour nous, un inconvénient majeur. En effet, l'objectif d'un système autonome destiné à modéliser dynamiquement un environnement réel est difficilement compatible avec ce type de contraintes : comment évaluer le nombre de classes qui composeront l'environnement acoustique, comment déterminer *a priori* l'ensemble des formes susceptibles d'apparaître dans le milieu ou encore comment estimer le niveau de bruit ?

ART s'auto-organise en fonction d'un seuil de vigilance qui détermine le degré de nouveauté d'un stimulus, ce qui est une contrainte plus compatible avec nos objectifs. Par ailleurs, ce modèle représente un compromis entre la stabilité du réseau et sa capacité d'adaptation, ce qui pourrait constituer un réponse adaptée aux spécificités de notre problème. En particulier, le principe d'une auto-modification de la structure du réseau nous a semblé particulièrement intéressant. Toute la dynamique d'apprentissage de ce modèle est fondée sur une augmentation de la sélectivité de cellules de sortie. L'apprentissage est donc intrinsèquement compétitif. Cette augmentation forcée du contraste entre les cellules de sortie suppose qu'un stimulus soit affecté à une classe unique et, par conséquent, que l'occurrence d'événements acoustiques simultanés soit improbable. Malheureusement, l'hypothèse de séquentialité des événements acoustiques n'est pas réaliste dans le cas d'environnements fortement bruités (cf. *Position du problème*).

Peu adaptées à notre problème, les méthodes décrites dans ce chapitre nous ont inspiré un modèle neuromimétique original, correspondant aux objectifs initialement fixés. Cette architecture, fondée sur les mémoires associatives et s'inspirant du processus d'adaptation structurelle du modèle de Grossberg, est décrite dans le chapitre suivant.

Chapitre 4

Réseau projectif pour la détection et la classification de transitoires

Dans ce chapitre, nous présentons une architecture neuronale inédite pour la modélisation adaptative d'un environnement de bruits impulsifs.

Inspiré des mémoires auto-associatives de Kohonen, ce modèle regroupe automatiquement en classes les différentes composantes de l'environnement acoustique. Cette classification est acquise et mise à jour de façon totalement dynamique, par un double mécanisme de création de nouvelles cellules et de suppression des cellules obsolètes.

Ce processus en ligne de modélisation dynamique de l'environnement nous permet ensuite de dégager un critère de rupture fondé sur une mesure de l'adéquation du modèle au signal.

Après avoir décrit la structure et les aspects dynamiques de l'architecture proposée, nous discutons des points essentiels qui la distinguent des modèles développés antérieurement.

4.1 Introduction

Nous avons vu que les performances en classification des mémoires associatives linéaires sont faibles, en particulier parce qu'elles ne peuvent résoudre que des problèmes de discrimination de classes linéairement séparables (cf. 3.4.1). Néanmoins, nous nous sommes intéressés à ce type de méthodes, essentiellement pour deux raisons.

D'abord, il s'agit d'une architecture simple et générique, qui peut être associée à de très nombreux algorithmes d'apprentissage.

De plus, la linéarité de ce modèle, qui est généralement considérée comme un défaut, peut être exploitée d'une façon totalement originale pour résoudre le problème de la classification de mixtures additives de formes acoustiques. Cet argument est lié aux objectifs que nous nous sommes fixés : il s'agit de réaliser un système autonome capable d'extraire et de mettre à jour une *image* d'un environnement de bruits impulsifs. Ce modèle d'environnement doit permettre de détecter l'apparition d'événements nouveaux et de regrouper en classes des événements récurrents du signal d'entrée. Dans un environnement réel composé de multiples sources acoustiques, le signal d'entrée est un mélange de ce qui est émis par chacune des sources. Le problème est donc de savoir comment extraire et mettre à jour quelques sons de référence permettant de décrire l'environnement acoustique, comment déterminer ensuite les contributions des différentes classes de sons dans le signal résultant du mélange et, enfin, comment détecter les événements nouveaux dans ce flux composite.

Une des difficultés majeures qui se pose est le traitement de signaux mélangés. Une mixture linéaire de n formes représentées par n vecteurs P_i s'écrit :

$$S(t) = \sum_{i=1}^{n} \alpha_i(t) P_i$$

On peut retrouver les contributions α_i des formes P_i dans le mélange en projetant S dans la base des vecteurs P_i, à condition qu'ils soient linéairement indépendants. C'est la formulation connexionniste de cette propriété algébrique simple qu'a proposée Kohonen dans la version optimale de son modèle de mémoire auto-associative ([Kohonen 89]). Cependant, Kohonen utilise cette capacité des réseaux projectifs à extraire une forme référencée d'un mélange linéaire pour réduire le niveau de bruit ou reconstruire un signal incomplet. Nous utilisons cette faculté de manière différente, pour dissocier des formes acoustiques à partir de leur mixture.

4.2 Espace de représentation des entrées

Le signal échantillonné est représenté par une suite de vecteurs acoustiques obtenus par analyse de Fourier dans une fenêtre glissante de largeur T. Cela signifie que chaque coefficient d'un vecteur de paramètres correspond à l'énergie du signal dans une bande de fréquence. Notre signal temporel est donc échantillonné en une suite d'images spectrales instantanées.

Ce choix d'une représentation temps-fréquence a l'avantage essentiel, dans le cadre de notre étude, d'être linéaire, ce qui peut-être montré très rapidement. En effet, la transformée de Fourrier T_F d'un mélange linéaire de deux fonctions f et g s'écrit :

$$\begin{aligned} T_F(\alpha f + \beta g)(\nu) &= \int_t^{t+T} (\alpha f(x) + \beta g(x)) e^{-2i\pi\nu x} dx \\ &= \int_t^{t+T} \alpha f(x) e^{-2i\pi\nu x} dx + \int_t^{t+T} \beta g(x) e^{-2i\pi\nu x} dx \\ &= \alpha T_F(f) + \beta T_F(g) \end{aligned}$$

La transformée de Fourier permet donc de conserver les caractéristiques de linéarité du mélange des signaux acoustiques. Ceci signifie que l'image spectrale du mélange de signaux acoustiques sera le mélange des images spectrales des signaux. La capacité de notre réseau à détecter et à dissocier des mixtures d'événements acoustiques représentées sous forme d'images temps-fréquence repose sur cette caractéristique de la méthode de paramétrisation. Un inconvénient de ce choix d'une paramétrisation conservant la linéarité des mixtures est que l'on ne pourra pas utiliser une échelle logarithmique pour la représentation des énergies dans les bandes de fréquences. La dynamique "naturelle" des entrées doit être conservée bien qu'elle ne corresponde pas à l'échelle perceptive, qui est partiellement linéaire et partiellement logarithmique. Ce problème peut engendrer une mauvaise exploitation des zones "informatives" de variation des paramètres, et un masquage de l'information utile.

Nous avons choisi des vecteurs de 256 coefficients calculés toute les 10 ms dans une fenêtre d'analyse de 24 ms. Ces choix suscitent plusieurs commentaires :

- bien que nous ayons essayé de limiter les hypothèses sur les signaux d'entrée, la taille de la fenêtre d'analyse, celle du vecteur acoustique et, dans une moindre mesure, la valeur du taux de recouvrement des fenêtres ont été choisis de façon relativement empirique. Nous avons déterminé la valeur de ces paramètres à partir de l'observation des bruits impulsifs rencontrés sur quelques fichiers d'exemples. De nombreux travaux ont montré l'intérêt des méthodes d'analyse multi-résolutions du signal [Malvar 90], ou encore de techniques de segmentation automatique d'un signal non-stationnaire en parties quasi-stationnaires ([Wesfreid 92]) pour la représentation du signal acoustique. Malheureusement, la recherche d'une stratégie efficace pour le réglage de ces paramètres nous aurait entrainés certainement assez loin du thème principal de ce travail.
- l'incidence des options retenues ne peut être évaluée qu'à la lumière des traitements réalisés en aval. Ces traitements seront décrits plus loin, et nous verrons comment le biais apporté par la méthode de paramétrisation peut être limité (cf. *Conclusion*).

4.3 Architecture

4.3.1 Structure

Notre réseau est composé de deux couches, totalement connectées entre elles, et ne possédant pas de connexions latérales. La couche d'entrée comporte n cellules destinées à recevoir le stimulus (constitué ici d'un vecteur acoustique). Chaque cellule d'entrée recevra, pour chaque déplacement de la fenêtre d'analyse, l'énergie du signal dans la bande de fréquence qui lui est propre. La taille de la couche d'entrée est déterminée par les paramètres de l'analyse temps fréquence réalisée.

La couche de sortie contient un nombre variable de cellules, chacune d'elle étant associée à une classe d'événements acoustiques. La valeur d'un nœud correspond à l'activité de la classe. Cette activité doit témoigner de la présence d'un événement de la classe dans le vecteur acoustique stimulus. Chacune de ces cellules "classifiantes" est connectée à toutes les cellules de la première couche du réseau par une liaison valuée.

Par ailleurs, une cellule est exclusivement dédiée à la mesure du degré de nouveauté du stimulus. L'activité de la cellule de nouveauté est calculée à partir de celle des cellules classifiantes, des valeurs des poids synaptiques et du stimulus lui-même. La figure 4.1 montre l'architecture effective de notre réseau.

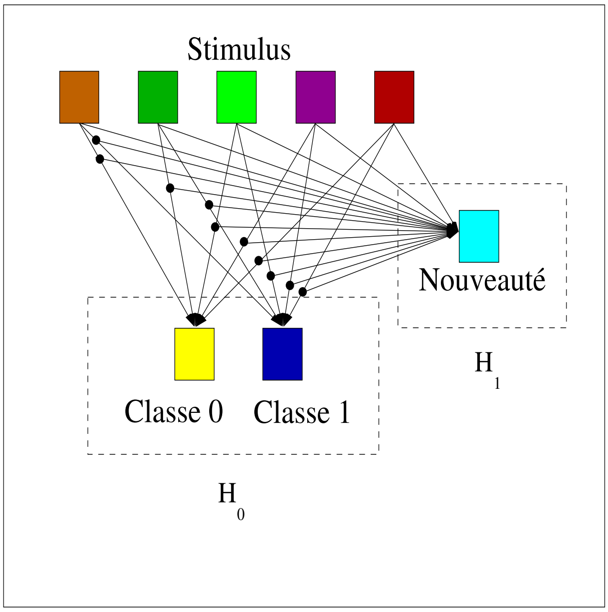

FIG. 4.1: *Structure du réseau projectif : la couche de sortie est composée de cellules classifiantes ; chacune de ces cellules détecte la présence d'un événement de la classe dans le stimulus. L'activité de ces cellules témoigne de la validité de l'hypothèse H_0 (le modèle décrit bien le signal observé). L'activité de la cellule de nouveauté est calculée à partir des poids synaptiques, de l'activité des cellules classifiantes et du stimulus. L'excitation de cette cellule est significative d'une rupture du modèle (hypothèse H_0).*

4.3.2 Fonction de transfert

Cellules de détection des classes

À chaque cellule classifiante, est associé un vecteur qui doit représenter une image spectrale type des éléments de la classe. L'activité des cellules de sortie est calculée en fonction de ces vecteurs, par projection du stimulus dans la base des prototypes. Cela signifie que chaque vecteur d'entrée, qui est considéré comme l'agglomérat d'un certain nombre de formes acoustiques instantanées, est décomposé dans l'espace des prototypes. Par conséquent, la réponse d'une cellule à un stimulus est calculée par une fonction linéaire dont les coefficients sont estimés à partir des prototypes des classes.

La forme de la fonction d'activation d'un neurone reste classique. En effet, la réponse $a_i(t)$ de la cellule i au stimulus $V = (V_j)_{j=1,n}$ est la somme des V_j pondérée par la valeur des liens W_{ij} de la cellule de sortie i à la couche d'entrée :

$$a_i(t) = \sum_{j=0}^{n} W_{ij}.V_j(t)$$

En notation matricielle, l'équation précédente s'écrit :

$$A(t) = W.V(t)$$

La fonction de transfert est donc linéaire, et déterminée par la matrice des poids. Ici, c'est la matrice de projection dans la base des prototypes, et donc la pseudo-inverse de la matrice P des prototypes (c'est à dire la pseudo-inverse de la matrice dont chaque ligne est un vecteur prototype).
Les vecteurs prototypes étant linéairement indépendants, W peut être calculée en fonction de P par :

$$W = (P^+)^t = (P.P^t)^{-t}.P$$

où l'opérateur $^{-t}$ désigne la transposée de la pseudo-inverse. La fonction d'activation des cellules classifiantes est donc identique à celle proposée par Kohonen pour les mémoires auto-associatives optimales. Kohonen ne donne pas de règle permettant d'affecter un stimilus à une classe. Cependant, il utilise les capacités d'adressage par le contenu de la mémoire associative pour retrouver une donnée mémorisée à partir d'une version bruité ou incomplète de la clef (cf. 3.4.3). Nous considérons que l'activité simultanée de plusieurs cellules de sortie traduit la présence de multiples événements acoustiques mélangés dans le stimulus. Cette alternative à la règle bien connue du *winner takes all* nous permet d'associer une forme acoustique à plusieurs références, non pas pour affecter un élément à plusieurs classes, mais pour dissocier des événements superposés. Cette approche est très différente de l'approche classique en classification automatique qui consiste à chercher, parmis les classes disponibles, celle qui représente le mieux la donnée. De plus, la possibilité qu'une donnée puisse être décomposée en éléments appartenant à des catégories distinctes conduit à reconsidérer la façon dont la décision d'appartenance à une classe peut être prise.

Cellule de détection de nouveautés

L'activité de la cellule de nouveautés est donnée par la norme de la composante du stimulus qui est dans l'orthogonal de l'espace-prototype. Le principe est de partitionner l'espace des entrées en deux sous-espaces, celui des prototypes et son complémentaire. Les n cellules classifiantes sont associées à autant de vecteurs qui forment une famille génératrice de l'espace des formes acoustiques pouvant être classifiées par le réseau. L'ensemble des prototypes est supposé "couvrir" l'espace-signal et l'espace des nouveautés est son complémentaire. Un élément en rupture est un événement qui se décompose mal dans l'espace prototype, et qui est, par conséquent, mal modélisé par le classifieur.

L'activité $a_r(t)$ de la cellule de nouveauté est calculée explicitement par :

$$a_r(t) = \frac{\|V(t) - \sum_{i=0}^{n} a_i(t).P_i\|}{\|V(t)\|}$$

Cette valeur est obtenue en enlevant du stimulus la somme des prototypes pondérée par les activités des classes. Elle représente une mesure de validité du modèle appris. La détection de rupture réalisée est basée sur la surveillance de ce coefficient. Il s'agit par conséquent d'une détection de rupture de modèle, bien que la nature de notre modèle diffère sensiblement des modèles physiques ou statistiques proposés généralement dans les méthodes de détection de ruptures issus du domaine du traitement du signal (cf. chapitre 2).

4.3.3 Prototype d'une classe

Chaque classe associée à une cellule de sortie est caractérisée par un prototype qui représente l'image spectrale type des éléments de la classe considérée. Il s'agit d'un vecteur acoustique dont chaque coefficient correspond à l'énergie du signal dans une bande de fréquence. Ces coefficients sont codés, de façon implicite, dans les poids synaptiques du réseau. En effet, on a vu que la matrice des poids était la pseudo-inverse de la matrice des prototypes (cf. 4.3.2).

Habituellement, la dispersion d'une classe est calculée à partir des distances entre ses individus, ou des individus au centre de la classe. Ici, une telle mesure n'aurait pas réellement de sens : la décision d'appartenance d'une donnée à une classe n'est pas réellement prise, la population des classes n'est comptabilisé, et on ne peut pas dresser une "cartographie" de l'ensemble des données. Par contre, on peut déterminer une zone de l'espace de représentation qui peut être associée à une classe. Cette zone est celle pour laquelle la cellule correspondante est fortement réactive : tout stimulus ayant une forte composante dans cette région provoque une forte excitation de la cellule associée. Le centre de cette zone de sensibilité est déterminée par le prototype de la classe : un stimulus colinéaire au protoype provoquera une activité maximale de la cellule. Par contre, la largeur de cette zone dépend non pas du prototype de la classe, mais de sa position par rapport à celle des autres prototypes : plus les prototypes des classes sont proches, plus les zones de sensibilité sont étroites. Par exemple, en dimension deux, l'angle des prototypes des deux classes concurrentes influence l'intensité de la réponse des cellules associées, comme le montre la figure 4.2.

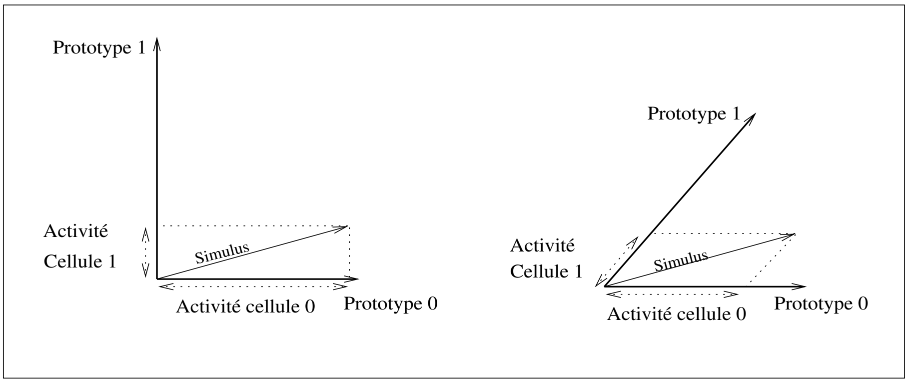

FIG. 4.2: *Influence de la position des prototypes sur l'activité d'une cellule : la proximité des prototypes des classes 0 et 1 réduit la réponse de la cellule 0 à un même stimulus.*

La figure 4.3 montre l'activité d'une classe en fonction de la distance euclidienne d'un stimulus au prototype de la classe. La première courbe illustre le cas où le prototype de la classe concurrente est orthognonal à celui de la classe activée, tandis que la seconde courbe illustre le cas ou les deux prototypes forment un angle de mesure $\frac{\pi}{4}$.

On voit que la proximité du prototype d'une autre classe réduit la largeur de la zone d'activation de la classe observée. La variabilité des classes est donc déterminée par les positions relatives des vecteurs prototypes.

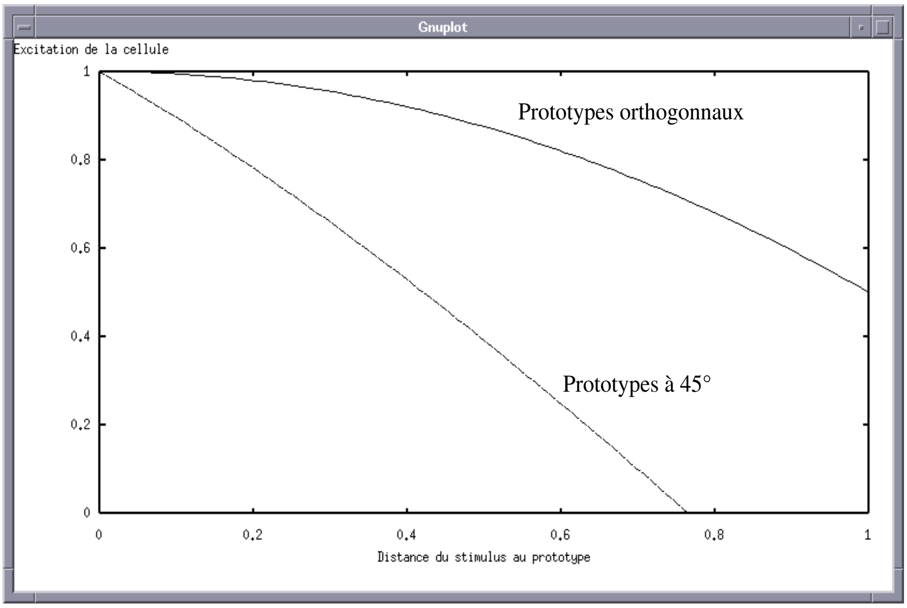

FIG. 4.3: *Activité d'une classe en fonction de la distance quadratique du stimulus au prototype : la réponse d'une cellule en fonction de la distance du stimulus au prototype dépend des prototypes concurrents. Ici, la cellule est plus sensible au stimulus lorsque les prototypes concurrents sont orthogonaux que lorsqu'ils sont séparés par un angle de mesure* $\frac{\Pi}{4}$.

4.4 Adaptation dynamique

L'adaptation du réseau est faite par modification de sa structure en fonction de l'activité des cellules de sortie ainsi que de la valeur de la cellule de nouveauté. Cette adaptation structurelle consiste à ajouter ou supprimer des cellules de sortie au réseau. Ces mécanismes d'adaptation par création et suppression de neurones reposent sur une mesure de la qualité globale de la modélisation effectuée, ainsi que sur un indice de validité des classes référencées.

4.4.1 Création des classes

Initialement, aucune classe n'est connue et la couche de sortie est composée uniquement de la cellule de nouveauté. À chaque fois que l'activité de celle-ci dépasse un seuil de vigilance, fixé *a priori*, on considère qu'une rupture est détectée, et une nouvelle classe est créée. Une cellule est alors ajoutée à la couche de sortie. Enfin, la matrice des poids est recalculée avec le vecteur ayant déclenché l'alarme comme prototype de la nouvelle classe. Les classes intégrées au système correspondent donc toutes à des événements ayant provoqué la détection d'une rupture dans un passé plus ou moins lointain. Ce processus d'acquisition dynamique du modèle mène à une incrémentation du nombre de classes au fur et à mesure que le signal d'entrée couvre l'espace de représentation. De façon implicite, cette incrémentation s'auto-régule : l'augmentation du nombre de classes se traduit par une augmentation du taux de recouvrement de l'espace d'entrée par l'espace système et cette augmentation réduit la dimension du sous-espace des nouveautés. L'augmentation du nombre de classes connues provoque une certaine perte de sensibilité du détecteur et, par conséquent, réduit la probabilité d'intégration d'une nouvelle classe. Inversement, un nombre très faible de classes va se traduire par une dimension très faible de l'espace système, et donc une dimension importante de l'espace des nouveautés. À l'extrême, si le système est vide, tout stimulus sera considéré comme nouveau et immédiatement intégré. Bien entendu, l'influence de ce processus est très directement dépendante du taux de recouvrement. Ce dernier indice est fonction à la fois de l'espace de représentation choisi et de la dimension effective de l'espace-signal.

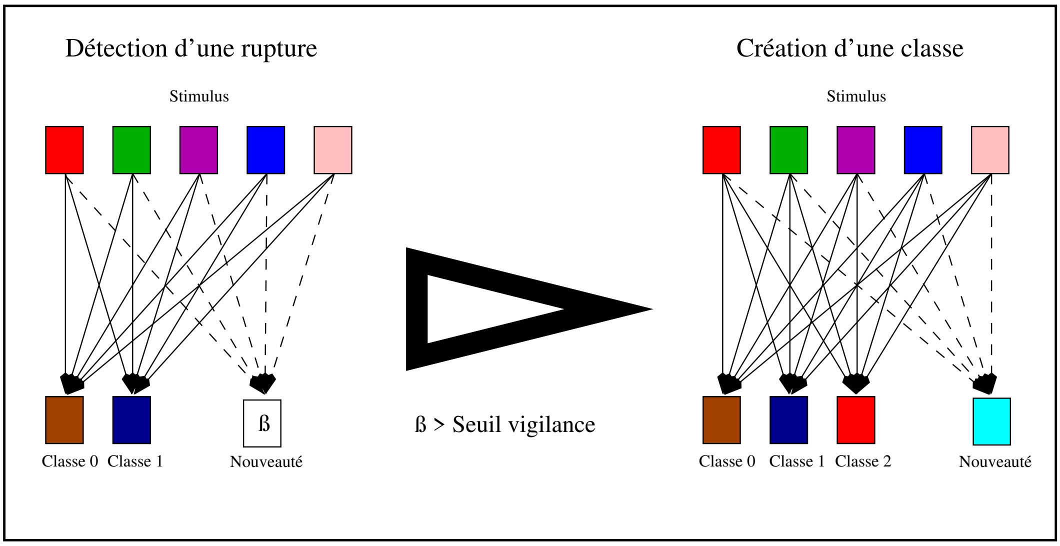

FIG. 4.4: *Création d'une classe à la suite d'une détection de rupture : lorsque l'activité β de la cellule de sortie dépasse le seuil de vigilance, alors une cellule est créée.*

Le corollaire génant de ce mécanisme d'auto-limitation est la perte de sensibilité du système dans le cas d'un fort taux de recouvrement. Dans la situation extrême où l'espace signal a recouvert l'ensemble de l'espace d'entrée, plus aucune détection n'est possible, l'activité de la cellule de nouveauté étant constamment nulle. On voit qu'au processus d'apprentissage et de mémorisation que représente l'intégration des nouveautés au système, il est nécessaire d'associer un mécanisme d'oubli indispensable à tout processus d'adaptation.

4.4.2 Suppression de classes

Le mécanisme de suppression des classes est destiné à stabiliser la dimension du système tout en augmentant sa capacité d'adaptation. Il consiste à supprimer les classes qui ont suscité une activité marginale dans la fenêtre temporelle d'observation. Cela peut correspondre à plusieurs cas de figure :

- événements rares, à l'échelle de l'empan temporel choisi. Un événement rare est un événement isolé, ou dont la fréquence d'apparition est faible en regard de la taille de la fenêtre temporelle choisie.
- événements peu représentatifs : une classe peut être créée à la suite de l'apparition d'une instance fortement bruitée d'un événement d'une autre classe. Dans ce cas, elle modélise un aspect de la variabilité d'un son suffisamment inhabituel pour ne pas en être réellement significatif.
- événements obsolètes : une classe peut représenter un ensemble d'événements qui ont eu lieu dans un passé relativement lointain. La disparition des événements de ce type rend la classe correspondante obsolète.

Par ailleurs, un indicateur de validité peut servir à remettre en cause des classes dont la représentativité n'est pas suffisante. Ceci se produit lorsqu'une classe ne correspond pas réellement à une catégorie de formes acoustiques, par exemple parce qu'elle a été mal constituée. Dans ce cas, l'indice de validité peut permettre de contrôler et éventuellement de corriger le processus de formation des classes.

Enfin, notre objectif est de faire un système de classification auto-adaptatif capable de modéliser dynamiquement un environnement acoustique en constante évolution. Dans un tel milieu, des classes de sons peuvent apparaître et disparaître. Pour que le système donne une image fidèle de l'environnement, il doit s'adapter aux disparitions des catégories acoustiques en supprimant les classes correspondantes.

Un processus d'élimination des cellules classifiantes basé sur une mesure de validité des classes comporte donc de multiples avantages, essentiellement dus à l'amélioration de la qualité de la modélisation.

Cependant, pour mettre en œuvre un tel mécanisme, il nous faut une mesure de validité des classes.

L'inertie est souvent utilisée, dans les méthodes de classification automatique, comme une mesure de représentativité d'une classe. Ce paramètre peut être utilisé pour pondérer la vitesse d'évolution de la classe dans un système dynamique, ou encore comme un indice de confiance destiné à contrôler la qualité d'une classe.

Malheureusement, un tel indice ne peut pas être calculé ici de cette façon, en particulier parce que la décision d'appartenance d'un stimulus à une classe n'est pas effectivement prise. En effet, chaque vecteur d'entrée est décomposé dans l'espace-système suivant la base des prototypes. L'éloignement d'un individu au centre de la classe n'est pas considéré comme un effet de la diversité des événements acoustiques de cette catégorie. Elle est interprétée comme étant due à la présence simultanée d'événements d'autres classes. Dans notre classifieur, il n'y a pas à proprement parler de variabilité intra-classe, la zone de sensibilité d'une cellule étant déterminée non pas par la dispersion de ses individus mais par la position de son prototype relativement à ceux des autres classes (cf. 4.3.3).

Cet aspect du modèle proposé rend difficile l'utilisation explicite d'un critère d'inertie des classes comme mesure de validité. On peut néanmoins utiliser un coefficient de température calculé à partir de l'activité moyenne d'une classe comme un indicateur de validité semblable, dans son principe, à un critère d'inertie fondé sur sa population. Les classes qui sont éliminées sont celles dont la température est inférieure à un seuil s fixé *a priori*. La durée de vie minimale d'une cellule détermine alors la persistance de la mémoire du réseau, qui peut être exprimée comme le temps au bout duquel une cellule ponctuellement activée, puis inactive, sera éliminée. Cette valeur correspond, en réalité, à la fréquence minimale d'un phénomène modélisé.

Le seuil d'élimination, ainsi que les paramètres éventuels de la fonction de température, permettent de régler la persistance de la mémoire du réseau.

Néanmoins ce coefficient de température doit satisfaire des contraintes relativement fortes :

- (1) la température d'une classe doit être calculée en ligne, comme les autres paramètres du classifieur. Ce calcul ne sera pas réalisé *a posteriori* ou à partir d'un historique qui serait maintenu constamment à jour. La température au temps $t + dt$ d'une classe est fonction de sa température au temps t et de son activité au temps $t + dt$.

- (2) il ne doit pas y avoir de différence sensible de température entre une classe active dans un passé immédiat et une autre active dans un passé *assez proche*. En effet, il ne s'agit pas de mesurer les fréquences basses d'un phénomène, mais d'avoir un indicateur pouvant prendre deux états correspondant aux deux situations possibles : la classe correspond à un phénomène effectivement présent dans l'environnement, ou elle n'est pas représentative d'une composante du signal active dans la fenêtre temporelle choisie, et doit être éliminée.
- (3) la température T_t doit être une fonction décroissante du temps. La température d'une classe qui n'est pas activée régulièrement doit diminuer au cours du temps et converger vers 0.
- (4) la fonction de température f_γ doit être croissante pour ses deux variables (température au temps $t - dt$ et l'activité de la cellule au temps t) : plus l'activité d'une cellule est forte, plus sa température augmente (ou moins elle réduit). De la même façon, une plus forte température au temps t doit provoquer une plus forte température au temps $t + dt$.
- (5) bien que fixée *a priori*, la taille de l'empan temporel utile doit être facilement paramétrable.

Un filtre passe-bas du premier ordre satisfait certaines de ces contraintes. Il s'agit d'une fonction du type :

$$T_{t+dt} = \alpha T_t + (1 - \alpha)a(t)$$

Une telle fonction de température comporte cependant deux inconvénients. D'une part, il s'agit d'une fonction des fréquences basses du phénomène, et non d'un indicateur binaire permettant une prise de décision de suppression facile. La figure 4.5 montre le refroidissement d'une cellule ponctuellement active, puis inactive. On voit que le choix du seuil d'élimination est déterminant dans la durée de vie de la cellule. Cette fonction nécessite donc deux seuils dont les fonctions sont partiellement redondantes : le paramètre α, qui règle la loi de refroidissement des cellules, et le seuil d'élimination. De plus, la modification du coefficient α influence non seulement la loi de refroidissement des cellules, mais aussi leur réchauffement : l'augmentation de température provoquée par l'activité ponctuellement élevée d'une classe est aussi déterminée par α. En augmentant la persistance de la mémoire, on diminue l'impact d'une excitation ponctuelle. En effet, l'influence d'une excitation ponctuelle serait affaiblie par le filtre et la cellule pourrait alors être supprimée à tort. Cet inconvénient pourrait cependant être contourné en réinitialisant la température dans le cas d'un activité significative de la cellule.

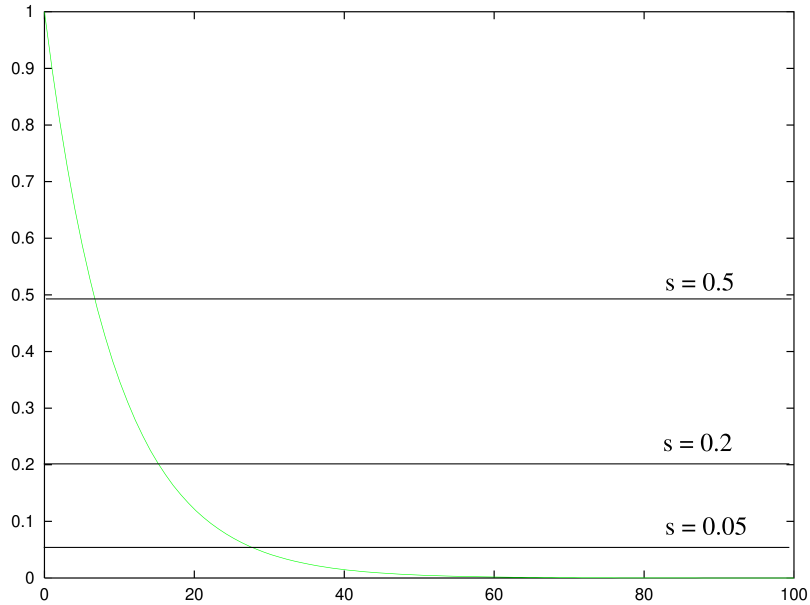

FIG. 4.5: *Calcul de la température par un filtrage passe-bas. La courbe montre l'évolution de la température d'une cellule ponctuellement activée. Le choix d'un seuil d'élimination s détermine le moment de la suppression de la cellule inactive.*

Nous avons plutôt utilisé une famille de fonctions f_γ qui ne présentent pas ces inconvénients et qui offrent quelques propriétés intéressantes :

$$f_\gamma(x) = x\left(\frac{1}{(1+e^{-\beta(x-\frac{1}{2})})}\right)^{\frac{1}{\gamma}}$$

avec $\beta > 1$ et $\gamma > 1$. La figure 4.6 montre la courbe représentative de f_2 avec $\beta = 20$ dans $[0, 1]$.

f_γ est bien croissante sur $[0, 1]$ quel que soit $\gamma > 1$. Elle est stable dans cet intervalle, c'est à dire que si $x \in [0, 1]$, alors $f_\gamma(x) \in [0, 1]$. De plus, $f_\gamma(x)$ est

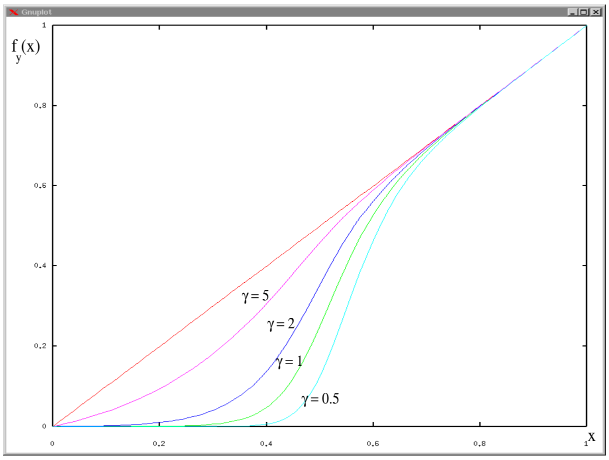

FIG. 4.6: *Fonction de température pour $\gamma = 0.5, 1, 2, 5$. Cette fonction donne l'augmentation de la température d'une cellule en fonction de son activité. Le choix du paramètre γ permet de déterminer l'inertie "thermique" des cellules : un γ élevé augmente la vitesse d'échauffement des cellules.*

strictement inférieure à 1 dans $[0,1]$. On n'a donc pas de point stationnaire dans cet intervalle, ce qui est nécessaire pour satisfaire la condition (3).

La température $T_i(t+dt)$ de la classe i au temps $t+dt$ est obtenue par :

$$T_i(t+dt) = f_\gamma(\alpha T_i(t) + (1 - \alpha.T_i(t))a_i(t+dt)) \quad , \quad \alpha \in [0,1]$$

Le terme $\alpha T_i(t)+(1-\alpha.T_i(t)a_i(t+dt)$ pondère l'influence de l'activité de la cellule dans la variation de température en recadrant cette activité dans l'intervalle $[0, 1 - \alpha.T_i(t)]$. De cette façon, l'activité ponctuellement maximale de la cellule permet la réinitialisation à 1 de sa température.

Par ailleurs, la suite récurrente (F_n^γ) définie par :

$$\begin{aligned} F_{n+1}^\gamma &= f_\gamma(\alpha F_n^\gamma) \\ F_0^\gamma &= 1 \end{aligned}$$

est décroissante et converge vers 0 quelquesoit $\gamma > 1$. Cette suite décrit la température d'une cellule qui aurait connu ponctuellement une activité maximale et une activité nulle ensuite. La figure 4.7 montre l'évolution de la suite F_n^γ en fonction de n.

Enfin, le paramètre γ permet de régler la persistance de la température en fonction du temps. Cet indice est utilisé dans un processus d'élimination des classes

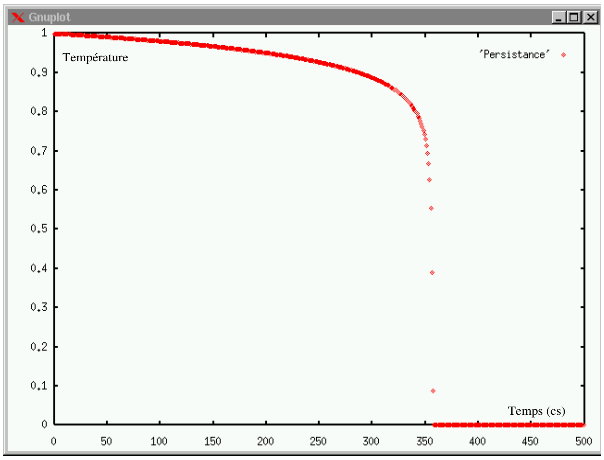

FIG. 4.7: *Température d'une classe ponctuellement activée : une cellule de température maximale puis jamais activée refroidit d'abord lentement puis brutalement, jusqu'à atteindre une température proche de 0.*

(cf. 4.4). Le choix d'une valeur critique en deçà de laquelle la classe sera éliminée influence très peu la mémoire du réseau. En effet, la courbe de refroidissement d'une cellule activée ponctuellement (cf. figure 4.7) montre que ce refroidissement est rapide dans l'intervalle de temps [0.1,0.8] : la cellule passe brutalement d'une température maximale (égale à 1) à une température quasiment nulle. La valeur du seuil choisi dans cet intervalle change peu le moment de la suppression. La figure 4.8 montre la faible influence du choix S de ce seuil, s'il est dans l'intervalle de temps $[0.1, 0.8]$, sur la durée de vie minimale d'une cellule (cf. figure 4.8).

La figure 4.9 montre le nombre minimum d'itérations nécessaires pour obtenir $F_n^\gamma < 0.5$, en fonction de γ. On voit que cette fonction est quasiment linéaire. En effet, le temps nécessaire au refroidissement d'une cellule ponctuellement activée est d'environ mille fois la valeur de γ (avec $\beta = 20$). Le choix de γ permet d'ajuster la "mémoire" du réseau très facilement.

Comme l'intégration des classes, le processus d'élimination a tendance à se réguler automatiquement au voisinage des valeurs limites du taux de recouvrement : l'énergie du signal est répartie entre les différentes classes et l'espace nouveauté. Un nombre élevé de classes va se traduire par une activité moyenne plus faible, et donc une température moyenne plus faible. Dans ce cas, la probabilité d'élimination d'une classe sera élevée. Inversement, la température moyenne des classes sera généralement plus élevée si leur nombre est faible. Là encore, il s'agit

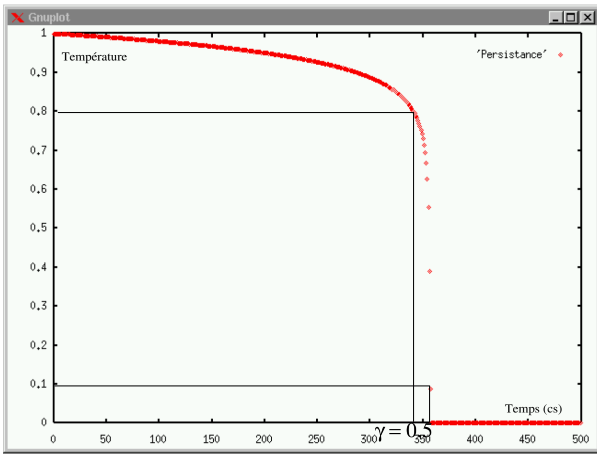

FIG. 4.8: *Influence du seuil de suppression sur la mémoire du réseau : choisir le seuil de suppression des classes entre 0.1 et 0.8 influence peu le moment de la suppression, qui est compris entre 375 cs et 390 cs. La durée de vie d'une cellule ponctuellement activée dépend peu du choix de ce seuil.*

d'un processus influant seulement pour des valeurs limites du taux de recouvrement, et complètement dépendant à la fois de l'espace de représentation et de l'espace-signal effectif. Ce mécanisme régule les paramètres du système dans des configurations extrêmes (nombre de classes proche de 0, ou proche de la dimension de l'espace de représentation). Cependant, il ne garantit pas la stabilité du modèle. Cette stabilité est déterminée par les seuils de vigilance, d'élimination de classes et d'inertie thermique des cellules, qui doivent permettre la compatibilité des deux processus d'incrémentation et de décrémentation du système. Néanmoins, cette stabilité dépend aussi des paramètres fondamentaux du signal, en particulier des fréquences caractéristiques des phénomènes.

4.5 Exemples

La figure 4.10 illustre ce processus d'adaptation par détection des nouveautés, intégration et élimination de classes. Il s'agit d'un signal composé de deux séries de bruits impulsifs récurrents. Le premier type de choc occupe les 3 premières secondes de l'exemple. Il y a ensuite un silence d'environ 2 secondes qui est rompu par l'apparition d'une nouvelle séquence de chocs d'un type différent. Enfin, deux occurrences de chocs du premier type réapparaissent aux temps 7.7 et 8.0 s.

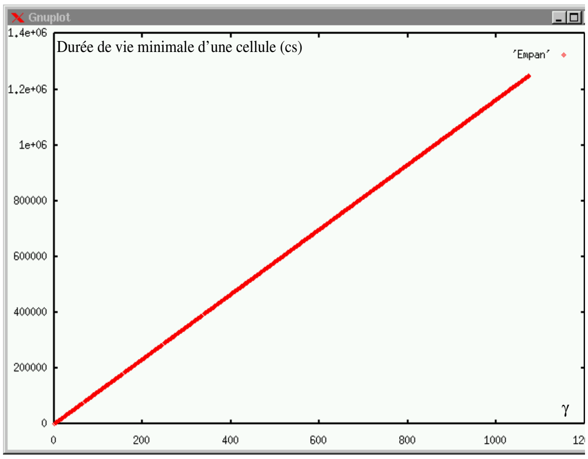

FIG. 4.9: *Largeur de l'empan temporel en fonction de γ. Cette courbe montre la durée de vie d'une cellule ponctuellement activée en fonction du coefficient calorimétrique γ. On voit que la persistance de la mémoire du réseau est une fonction quasiment linéaire de γ.*

La première ligne du schéma représente le signal acoustique, la seconde le spectrogramme correspondant. La troisième ligne montre l'évolution temporelle du taux de recouvrement, c'est à dire du rapport du nombre de classes à la dimension totale de l'espace de représentation. Cette courbe représente un nombre de classes normalisé par la dimension des vecteurs d'entrée. Enfin, la dernière ligne montre l'activité de la cellule de nouveauté.

On voit que les premières occurrences des deux types de chocs suscitent une forte excitation de la cellule de nouveauté (temps 0.1 et 5.4 s) et la création de classes modélisant les événements détectés. La disparition des chocs de la première catégorie provoque la baisse de la température des classes associées puis, avec un certain retard, la disparition des cellules correspondantes (temps 2.9 et 3 s). La réapparition de ces bruits autour de la huitième seconde de signal suscite à nouveau l'excitation de la cellule de nouveauté et la création de classes : ces bruits avaient été "oubliés" par le système et leur présence après une absence suffisamment longue provoque une nouvelle détection de rupture (7.7 et 8 s).

4.6 Conclusion

Comme les techniques classiques de détection de ruptures, le système proposé détecte les variations brusques d'un descripteur du signal. Mais contrairement à elles, il ne s'agit pas d'un descripteur de bas niveau, comme du signal acoustique ou une image temps-fréquence, mais d'une représentation plus abstraite et plus concise de l'environnement. Cette modélisation procède par regroupement en classes des différentes composantes du milieu acoustique. Deux mécanismes inter-dépendants permettent l'adaptation du système aux évolutions du signal : la détection d'événements acoustiques mal modélisés par le classifieur, et la mise à jour du classifieur après la détection d'une rupture ou disparition d'une classe d'événements acoustiques. Les variations brutales du modèle ont lieu lorsque celui ci ne correspond plus aux observations, ce qui se produit lors de l'apparition d'une nouvelle catégorie de formes acoustiques. La rupture du modèle est déclenchée par l'apparition d'une nouveauté : c'est la première occurrence d'un événement acoustique qui provoque la mise à jour du classifieur, le rendant ainsi "insensible" aux occurrences suivantes.

Les avantages de la méthode proposée sont multiples. D'une part, il s'agit d'une méthode fondée sur le minimum de connaissance *a priori*. Son aspect auto-adaptatif permet un fonctionnement autonome et un traitement en ligne des données. Par ailleurs, une alternative à la règle du *winner takes all* est proposée : plutôt que de mettre en compétition les cellules de sortie du réseau classifeur de manière à affecter le stimulus à une classe unique, nous considérons que l'activité synchrone de plusieurs cellules correspond aux présences simultanées d'événements de catégories différentes. Cette exploitation inédite des propriétés des mémoires auto-associatives linéaires permet la détection et la classification d'événements acoustiques superposés. Malheureusement, la capacité du réseau projectif à distinguer des formes acoustiques mélangées n'est effective que si les prototypes des classes correspondantes sont bien connus. L'apparition d'une nouveauté simultanément à un événement d'une classe référencée provoque la création d'une cellule dont le prototype est le stimulus ; en l'occurrence, il s'agira du vecteur résultant de la superposition de deux formes acoustiques, l'une inconnue, et l'autre appartenant à l'espace des prototypes. Le mélange des formes spectrales se produit aussi lorsque deux ruptures coïncident : le prototype intégré au système est dans ce cas le mélange des vecteurs ayant provoqué la rupture du modèle.

Ce mélange des prototypes conduit à un mélange des activités des classes correspondantes. Avant de se poser le problème de la recherche d'un processus qui permettrait la dissociation des prototypes, on peut se poser celui de la séparation des activités des classes. Ce type de problème a été abordé dans le cadre différent de la séparation aveugle de sources. Le chapitre suivant décrit rapidement la problématique de ce domaine. Cette étude nous permettra d'apporter au problème des mixtures de prototypes une solution qui sera décrite dans le chapitre 6.

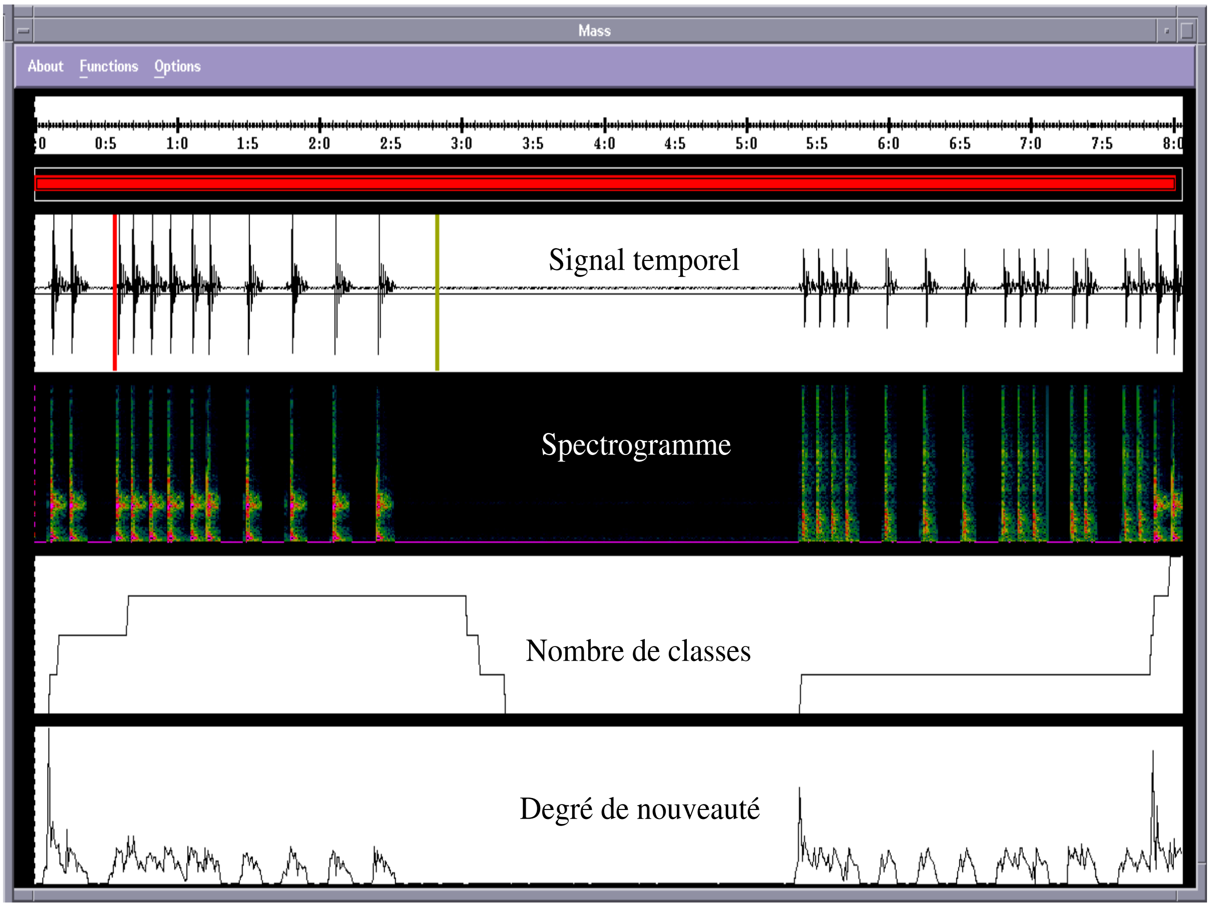

FIG. 4.10: *Adaptation par intégration et supression de classes : signal temporel, spectrogramme correspondant, nombre de classes et degré de nouveauté. On voit le processus de création de classes dans les périodes de 0 à 0.7 sec, puis entre les instants 7.7 et 8 sec. Ces créations de classes correspondent à des pics d'activité de la cellule de nouveauté (activité représentée sur la dernière ligne de ce schéma). Le silence, dans la période 2.5-3.5 sec, provoque un refroidissement des cellules qui se traduit par l'élimination de certaines classes (temps 3.1, 3.2 et 3.4 sec).*

Chapitre 5

Séparation aveugle de Sources

Ce chapitre précise d'abord le problème traité en séparation aveugle de sources.

Les points essentiels des diverses méthodes sont ensuite décrits dans leur principe, puis par l'exposé de quelques uns des principaux algorithmes developpés dans ce domaine.

Nous situons enfin ces travaux dans le cadre plus général des recherches, menées en théorie de l'information, sur la distribution de l'information dans des espaces multi-dimensionnels.

5.1 Introduction

Ce domaine de recherche aborde le problème de la dissociation d'un certain nombre de signaux dont les mélanges sont perçus par différents capteurs. Ces signaux sont inconnus, mais supposés statistiquement indépendants. L'objectif des algorithmes de séparation de sources est de reconstituer les signaux émis par chaque source à partir de la seule observation de leurs mixtures.

Ce sujet a suscité des études dans des disciplines assez variées : traitement du signal, neuro-biologie, connexionnisme, théorie de l'information, etc.. La première formulation du problème a été présentée par Hérault et Jutten ([Hérault 85]). Il s'agissait de dissocier, dans le signal nerveux de contrôle du mouvement, des informations relatives à la vitesse d'étirement d'un muscle de celles relatives à l'étirement de lui-même. La solution proposée consistait en un modèle neuro-mimétique auto-adaptatif capable de capturer l'opérateur de mélange inverse seulement à partir de l'information fournie par les capteurs.

Cette architecture, que nous détaillerons plus loin, en a inspiré beaucoup d'autres. L'amélioration des méthodes initiales a conduit au développement de modèles fondés sur des statistiques d'ordre supérieur ou sur des résultats de théorie de l'information. Ces recherches ont abouti à l'émergence de nouvelles méthodes regroupées sous le nom d'Analyse en Composantes Indépendantes (ACI, ICA en anglais). Ces méthodes généralisent le problème initial de la séparation de sources à celui de la répartition de l'information dans des espaces multi-dimensionnels.

5.2 Problématique

5.2.1 Formulation algébrique

Supposons qu'un nombre n connu de sources S_i émettent simultanément des signaux $S_i(t)$ de moyenne temporelle nulle et supposés statistiquement indépendants. Un nombre n de capteurs perçoivent n signaux temporels X_i qui résultent du mélange des signaux S_i originellement émis par les sources. La mixture de ces signaux peut être exprimée, en notation matricielle, par :

$$X(t) = A.S(t)$$

$$avec\, X(t) = (X_1(t), ..., X_n(t))^T \;\; et \;\; S(t) = (S_1(t), ..., S_n(t))^T$$

L'objectif des techniques de Séparation Aveugle de Sources (S.A.S.) est de retrouver les signaux sources S_i à partir de la seule observation des X_i et en faisant le minimum d'hypothèses sur la nature des signaux d'origine et sur les caractéristiques du mélange.

On cherche donc un opérateur $A'(t)$ tel que :

$$S'(t) = A'(t).X(t)$$

avec $S'_i(t) = \lambda_i S_{\theta(i)}(t), \quad \lambda_i$ scalaire indépendant de t et $\theta(.)$ bijection de {1,..,N} dans lui même. La présence des coefficients λ_i traduit le fait que la puissance des signaux d'origine ne peut pas être restituée, tandis que la fonction $\theta(.)$ met en relation des signaux de sortie et des signaux sources. En effet, l'ordre des coefficients de $X(t)$ étant indéterminé, la correspondance entre les sources et les signaux de sortie est aléatoire. Globalement, on cherche une classe de solutions équivalentes à un changement d'échelle et à une matrice de permutation près.

5.2.2 Formulation connexionniste

L'approche connexionniste du problème de SAS consiste à chercher une architecture neuromimétique capable de restituer les signaux d'origine en réalisant un mélange inverse des signaux d'entrée.

Le modèle générique introduit par Hérault et Jutten est un réseau à une couche. Deux catégories de connexions lient les cellules soit à l'extérieur, soit à elles-mêmes.

Chaque canal d'entrée reçoit les données issues d'un capteur. Les connections latérales règlent l'interaction récurrente des cellules du réseau. Les canaux de sortie portent n signaux supposés statistiquement indépendants, au sens d'une fonction de coût caractéristique de l'algorithme adopté. L'adaptation du réseau consiste à modifier ses poids synaptiques pour obtenir la convergence du modèle vers un état stable réalisant la minimisation de la fonction de coût.

De nombreuses méthodes ont été développées ces dernières années sur ce thème, ([Féty 88], [Comon 89], [Cardoso 89], [Karhunen 94] [Oja 92]. Toutes ces méthodes sont fondées sur l'hypothèse d'indépendance statistique des signaux *sources*, ainsi que sur la diversité spatiale des sources acoustiques et des capteurs. Elles diffèrent essentiellement par les modèles de mélanges adoptés, les mesures de dépendance statistique et les règles d'apprentissage utilisées.

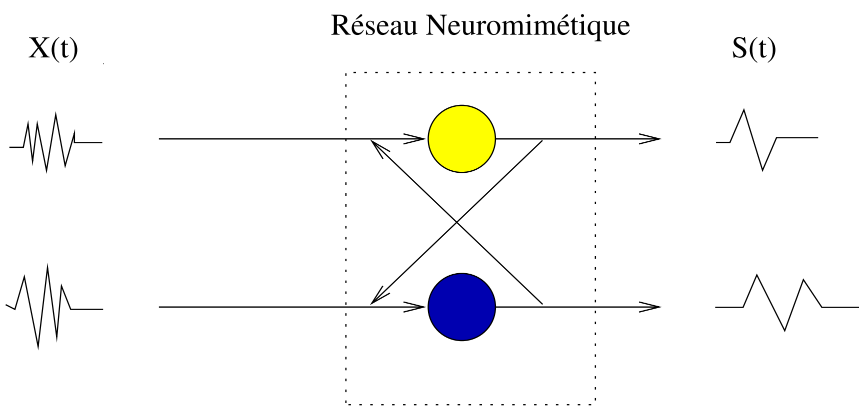

FIG. 5.1: *Architecture neuronale pour la séparation aveugle de deux sources : il s'agit d'un réseau à connections récurrentes, connecté à l'extérieur par des liaisons valuées. L'apprentissage du réseau se fait par modification de ces valeurs à partir des sorties du réseau : des connexions inhibitrices lient les sorties du réseau aux entrées.*

Les solutions connexionistes au problème de SAS sont adaptatives, ce qui correspond bien à l'approche séquentielle et dynamique que nous avons adoptée tout au long de ce travail. C'est pourquoi nous nous sommes orientés vers ce type de méthodes plutôt que vers celles fondées sur des calculs par blocs, bien que certains de ces algorithmes soient, eux aussi, adaptatifs.

5.3 Modèles de mixtures

La complexité du problème de séparation de sources est liée, entre autres, au choix du modèle de mélange. D'un point de vue physique, deux ondes acoustiques qui se propagent simultanément se perturbent mutuellement. Une première approximation consiste à supposer le mélange linéaire. Des solutions neuronales au problème plus difficile de la déconvolution de signaux ont néanmoins été proposées, et ont permis d'obtenir de meilleurs résultats en conditions réelles.

5.3.1 Modèle de mélanges linéaires

Dans ce cas, on a :

$$Y(t) = AX(t) + \nu(t)$$

où $X(t)$ est le vecteur des signaux émis par les sources, A est la matrice de mélange (inconnue) et $Y(t)$ est le vecteur des signaux reçu par les capteurs. Ici, A est un opérateur linéaire.

Retrouver $X(t)$ suppose que A soit inversible, c'est à dire que les mixtures de signaux sources forment une famille libre dans l'espace des mixtures. Ceci traduit en réalité des contraintes sur la diversité spatiale des sources et des capteurs.

Le bruit additif $\nu(t)$ est souvent négligé. Néanmoins, [Karhunen 95a] et [Therrien 92] abordent le problème de SAS avec ν non nul.

5.3.2 Modèle de mélanges convolutifs

Le modèle linéaire du mélange des signaux sources est une approximation relativement grossière, et l'application des méthodes de séparation de sources en conditions réelles nécessite un modèle convolutif, plus fidèle à la réalité.

Le modèle de mélange convolutif est :

$$Y(t) = A * X(t) + \nu(t)$$

où A est une matrice de filtres.

Quelques méthodes susceptibles de déconvoluer des signaux indépendants ont été développées. [Nguyen 92] montre l'amélioration des performances en conditions réelles apportée par cette modélisation plus fine du mélange.

5.4 Critères d'indépendance statistique

On a vu que le problème de S.A.S. peut être formulé comme un problème de minimisation d'une fonction d'énergie représentant une mesure de dépendance statistique des signaux restitués.

Deux événements aléatoires a et b sont mutuellement indépendants si la connaissance de l'un n'apporte aucune information sur l'autre. Ceci peut être exprimé par :

$$p(a) = p(a|b)$$

L'indépendance de n variables aléatoires x_i est satisfaite si et eulement si la fonction F de densité de probabilité du vecteur $X = \{x_i\}_{i=1,..,N}$ est égale au produit des fonctions marginales :

$$P_s(x) = \prod_{i=1}^{N} p_{si}(x_i) \qquad (1)$$

L'utilisation directe d'une telle règle n'est généralement pas réalisable, à moins de connaitre les fonctions de densité de probabilité des x_i.

Les méthodes de SAS utilisent donc plutôt des approximations de cette mesure théorique, ou encore des fonctions de coût dont les minimums doivent correspondre à un état d'indépendance des signaux de sortie. Deux approches peuvent être distinguées.

Dans la première, le critère à optimiser est basé sur des statistiques d'ordre supérieur. Notamment, les cumulants croisés de n variables aléatoires x_i peuvent être utilisés, sous certaines contraintes, comme critères d'indépendance. Le cumulant croisé d'ordre n des variables aléatoires $(x_i)_{i=1..N}$ s'écrit :

$$Cum(X) = \sum_{p=1}^{n} (-1)^n (p-1)! E(\prod_{i \in v_1} x_i) ... E(\prod_{i \in v_p} x_i)$$

où $E(\prod_{i \in v_p} x_i)$ sont les moments des variables x_i pour les partitions V_p de 1,..,N.

Si tous les cumulants de tous les ordres sont nuls, alors les variables x_i sont indépendantes. L'utilisation partielle de cette propriété permet d'obtenir des mesures de dépendance basées soit sur des heuristiques ([Jutten 87], [Karhunen 94], [Oja 92], [Cichocki 92]) soit sur des *fonctions de contraste* ([Comon 91a], [Cardoso 94]).

Cette approche a été adoptée par de très nombreux chercheurs. Un panorama des diverses méthodes proposées dans la littérature peut être trouvé dans [Comon 94] [Karhunen 96] [Haykin 96].

Une autre approche est basée sur des résultats de la théorie de l'information qui permettent la comparaison de lois de probabilité. Cela consiste à reformuler l'équation (1) en utilisant la mesure de divergence de Kullback, qui évalue l'information d'une loi P sur une loi Q :

$$K(P,Q) = \begin{cases} \int p.log\frac{p}{q}.d\nu & si \quad P << Q \\ \infty & sinon \end{cases}$$

Cette mesure permet d'évaluer l'information mutuelle de deux variables aléatoires X et Y. En effet, cette mesure est donnée par l'information de Kullback de la loi $F_{X,Y}$ de (X,Y) sur le produit des lois marginales :

$$I(X,Y) = K(F_{X,Y}, F_X.F_Y)$$

Ceci peut s'exprimer en fonction de l'entropie H des variables aléatoires X et Y et de l'entropie de X conditionnellement à Y :

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

avec

$$H(X|Y) = -\int p(x,y) log p(x|y)$$

L'information mutuelle de deux variables aléatoires est nulle si et seulement si elles sont indépendantes. Sinon, elle est strictement posisitive.

De plus, en minimisant la redondance d'information des sorties d'un réseau séparateur de sources, on minimise aussi leur dépendance statistique. [Nadal 94] montrent que la minimisation de l'information mutuelle des signaux de sortie peut être obtenue par une maximisation de l'information mutuelle entre l'entrée du réseau et sa sortie. Ce critère d'indépendance statistique est utilisé dans des méthodes de SAS récemment introduites ([Deco 95], [Nadal 97], [Bell 95], [Taleb 97]).

5.5 Quelques algorithmes de séparation aveugle de sources

Le problème abordé dans ce chapitre a connu, ces dernières années, des développements très variés. Nous allons seulement décrire trois méthodes qui relèvent d'approches assez différentes : l'algorithme de Hérault et Jutten (HJ), basé sur des statitiques d'ordre supérieur, celui de Bell et Sejnowski fondé sur une minimisation de l'information mutuelle entre entrées et sorties du réseau, et enfin l'algorithme de Matsuoka, inspiré du modèle initial de HJ mais qui s'applique à des sources non-stationnaires. Nous utiliserons cet algorithme pour la réorganisation du classifieur auto-adaptatif que nous avons developpé et qui est décrit dans le chapitre suivant. C'est pourquoi l'exposé de cette méthode est sensiblement plus détaillé que celui des méthodes précédentes.

5.5.1 L'algorithme Hérault-Jutten

L'algorithme proposé en 1985 par Hérault, Jutten et Ans traite du problème de SAS pour un mélange linéaire instantané de sources stationnaires. Les signaux sources doivent être indépendants et de moyennes nulles.

Le réseau proposé par ces auteurs correspond à l'architecture illustrée par la figure 5.1. La fonction de transfert est linéaire et les connexions latérales sont inhibitrices. Les signaux de sortie sont obtenus par :

$$s_k(t) = x_k(t) - \sum_{i=1}^{n} c_{ik} s_i(t) \quad , \qquad c_{ii} = 0$$

Lorsque le réseau est stable, cette règle devient, en notation matricielle :

$$S(t) = (C + I)^{-1} X(t)$$

L'apprentissage du réseau consiste en une modification en ligne des poids synaptiques. Cet ajustement doit conduire à une minimisation d'un critère de dépendance statistique des sorties. L'indice de dépendance retenu est dérivé de la condition d'indépendance basée sur les cumulants d'ordre supérieurs (5.4). En effet, si les moments croisés $M_{i,j}$ des signaux x_i, x_j sont tous nuls, alors ceux ci sont mutuellement indépendants :

$$M_{n,m}(x_i, x_j) = E(x_i^n x_j^m) = E(x_i^n) E(x_j^m)$$

avec n et m des entiers positifs et impairs. Dans le cas de signaux numérisés par N échantillons, les moments croisés peuvent être estimés avec :

$$M_{n,m}(x_i, x_j) = \frac{\sum_{k=0}^{N} \sum_{l=0}^{N} x_i^n(k) x_j^m(l)}{N}$$

De façon à minimiser ce critère, les auteurs proposent une règle d'apprentissage générique :

$$\Delta w_{ki} = a f(y_k).g(y_i')$$

où y' désigne le vecteur y centré, f et g deux fonctions non-linéaires et impaires et a un coefficient réglant le pas d'apprentissage. Des fonctions f et g non-linéaires et impaires permettent d'instancier la règle d'apprentissage pour obtenir une règle basée sur les moments d'ordre supérieurs. En choisissant :

$$f(x) = \sum_{k=0}^{\infty} a_k x^{2.k+1}$$

$$g(x) = \sum_{k=0}^{\infty} b_k x^{2.k+1}$$

on obtient la règle d'apprentissage :

$$c_{i,j}(t + dt) = c_{i,j}(t) + \mu \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} a_n x_i(t)^{2n+1} b_m x_j(t)^{2m+1}$$

L'usage de ces fonctions permet d'expliciter un état stable en fonction des moments décrits plus haut :

$$E(c_{i,j}(t + dt) - c_{i,j}(t)) = \mu \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} a_n b_m E(x_i(t)^{2n+1} x_j(t)^{2m+1}) = 0$$

La convergence de l'algorithme est assurée lorsque les signaux à séparer sont sous-gaussiens et de fonctions de densité de probabilité symétriques. Cependant, cette convergence n'est pas théoriquemnt garantie dans le cas général. [Sorouchyari 91] propose une analyse des points d'équilibre de l'algorithme. L'étude théorique de son comportement reste néanmoins extrêmement difficile. En effet, la particularité de ce modèle neuromimétique est que la modification des poids du réseau est faite en ligne, et que cet ajustement utilise les sorties du réseau qui sont elles-mêmes dépendantes des poids. À chaque instant, les liens c_{ij} sont modifiés en fonction d'un terme dynamique. Par conséquent, l'analyse théorique du comportement du réseau suppose la manipulation d'équations différentielles non-linéaires. Ces équations peuvent difficilement être approximées par des fonctions linéaires, puisque c'est justement ce caractère fortement non-linéaire qui caractérise l'algorithme.

Cette technique a fait l'objet de nombreuses études théoriques ([Fort 91], [Comon 91b], [Sorouchyari 91]). Il a par ailleurs été étendu à la séparation de mélanges convolutifs de signaux ([Platt 91] [Nguyen 95]).[Cichocki 94] proposent un modèle similaire mais dont les connexions d'une cellule à elle-même sont adaptées de façon à améliorer la robustesse de l'algorithme.

5.5.2 Algorithme de Bell et Sejnowski

Cet algorithme, présenté par [Bell 95], s'inspire du travail de Linsker ([Linsker 92]). Il s'agit d'une méthode fondée sur la maximisation de l'information mutuelle entre entrées et sorties du réseau. Contrairement au modèle précédent, la fonction de transfert n'est pas linéaire :

$$S = g(W.X + W_0)$$

où W_0 est une matrice constante et $g(.)$ est une fonction sigmoïdale, par exemple $g(.) = tanh(.)$, qui permet d'optimiser la transmission de l'information lorsque la pente de la sigmoïde correspond aux zones très informatives du signal d'entrée ([Laughlin 81]).

La règle d'apprentissage consiste à modifier les poids w_{ij} de manière à maximiser $I(X, Y)$. Or l'entropie $H(X, S)$ de X conditionnellement à S n'est pas dépendante des poids, ce qui permet d'écrire :

$$\frac{\delta}{\delta w} I(X, S) = \frac{\delta}{\delta w} H(S)$$

Maximiser l'information mutuelle revient donc à maximiser l'entropie des sorties du réseau. Lorsque $g(.)$ est monotone, les fonctions de densité de probabilité des entrées et des sorties sont liées suivant la relation :

$$f_X(X) = \mid J \mid .f_Y(S)$$

où J est le déterminant de la matrice D des dérivées partielles ($D_{i,j} = \frac{\delta s_i}{\delta x_j}$). Pour maximiser l'entropie des sorties, il faut maximiser le jacobien $\mid J \mid$. Ceci peut être fait à l'aide d'une méthode du gradient stochastique. Bell propose la règle d'apprentissage, pour une fonction d'activation $g(u) = 1/(1 + e^{-u})$:

$$\Delta W = W^{-1t} + (1 - 2S)X^t$$

Cette méthode a fait l'objet d'études ultérieures : le choix de $g(.)$ est discuté dans [Bell 95]. Une alternative à la règle d'apprentissage initiale permet d'éviter l'inversion de la matrice W ([Wang 95]). Récemment introduite, elle a l'avantage de reposer sur des bases théoriques claires. Cependant, elle requiert une quantité de calculs assez importante et surtout, elle suppose les signaux non gaussiens. Cette hypothèse est cependant nécessaire dans la plupart des méthodes de SAS.

5.5.3 Algorithme de Matsuoka

Principe

Les méthodes précédentes s'appuient généralement sur des statistiques non gaussiennes. La principale originalité de l'algorithme proposé par [Matsuoka 95] est qu'il s'applique à des signaux centrés non-stationnaires, gaussiens ou non-gaussiens. Cette caractéristique peut se revéler extrêmement intéressante dans le cas où les signaux à traiter sont mal connus, ou proches de signaux gaussiens. C'est souvent le cas des signaux aléatoires réels (par exemple du signal de parole), ce qui rend difficile, dans ce cas, l'estimation des paramètres statistiques.

On a vu que les méthodes neuronales de SAS utilisaient généralement des statistiques d'ordre supérieur à 2. L'algorithme de Matsuoka utilise les moments croisés d'ordre 2 mais nécessite que tous les signaux soient non-stationnaires et indépendants, ce qui signifie que les rapports de leurs variances doivent être variables.

En conditions réelles et pour des signaux acoustiques, on peut supposer que cette hypothèse est souvent vérifiée.

Architecture

La structure du réseau proposé par Matsuoka est en tout point semblable à celle de HJ. Le réseau est donc un réseau totalement interconnecté (cf. figure 5.1).

La dynamique des activités des cellules est déterminée par l'équation :

$$T\frac{dS(t)}{dt} + S(t) = x(t) - CS(t)$$

Cette équation différentielle peut être approximée par l'équation linéaire liant entrées et sorties du réseau :

$$S(t) = (I + C)^{-1}x(t)$$

Apprentissage

Matsuoka montre que, sous les hypothèses énoncées plus haut, la matrice C est une matrice de séparation si et seulement si l'opérateur $Q(C, R(t))$ défini ci-dessous est constamment égal à 0 :

$$Q(C, R(t)) = \frac{1}{2}\sum_{i=1}^{N} log < S_i^2(t) > -log \; |< S(t)S(t)^t >|$$

Cette relation est utilisée pour réduire la recherche d'une matrice inverse C à un problème de minimisation de Q. Matsuoka propose que cette optimisation soit réalisée par une méthode de descente du gradient. Il en déduit la règle d'adaptation des poids synaptiques :

$$T\frac{dC_{ij}}{dt} = -\frac{Q(C, R(t)}{\delta c_{ij}}$$

qui s' écrit, en développant le terme de droite :

$$T\frac{dC}{dt} = (I + C)^{-1}\{(diag < S(t)S^t(t) >)^{-1}S(t)S^t(t) - I\}$$

Le paramètre T permet d'ajuster la vitesse d'évolution des poids. Matsuoka montre que les positions d'équilibre de la loi d'apprentissage correspondent bien à des situations d'indépendance des signaux de sortie du réseau. Par ailleurs, si les signaux restent localement non-stationnaires, alors ces états d'équilibre sont uniformement stables.

Exemples

La figure 7.1 montre la séparation de signaux artificiellement mélangés. Ces signaux sont constitués d'un bruit de hall de gare mélangé artificiellement à un morceau de musique de Miles Davis. La figure montre, pour les deux sources considérées, le signal issu du réseau (sorties 1 et 2), les signaux émis par les sources, et ceux captés par les micros. Ces derniers sont donc les signaux d'entrée du réseau. Après quelques secondes, la convergence est atteinte. Elle correspond à une stabilité du réseau pour laquelle les sorties correspondent aux signaux émis par les sources. Ici, on n'a pas cherché un réglage du réseau qui accélérerait la convergence. Par ailleurs, un bruit de hall de gare n'est pas issu d'une source ponctuelle. Cependant, le signal a été mélangé artificiellement et peut donc être considéré comme ponctuel et instantané. En réalité, si l'ensemble des sources qui composent ce bruit de hall de gare sont lointaines relativement à la distance entre les micros, alors elles sont perçues avec la même intensité par les deux micros. Toutes ces sources seront donc vues comme une source unique, puisque ce n'est pas l'origine géographique de la source qui est utilisée, mais le fait que les micros perçoivent le même signal à des niveaux différents.

5.6 Distribution optimale de l'information

Les algorithmes de séparation de sources cherchent une combinaison de n signaux d'entrée de façon à reconstituer n signaux indépendants. La toute première solution consistait à minimiser les corrélations des entrées, ce qui peut être réalisé par des modèles relativement simples. En effet, minimiser les corrélations des données revient à en faire une Analyse en Composantes Principales (ACP). La recherche de règles non-supervisées pour l'adaptation des modèles neuro-mimétiques effectuant une ACP a été l'objet de très nombreuses études ([Karhunen 95b]). Malheureusement, la corrélation est une mesure de dépendance statistique très incomplète. Elle ne permet pas, dans le cas général, de séparer des signaux complexes. Cette insuffisance des règles d'apprentissage linéaires pour la décorrélation des sorties d'un réseau a conduit au développement de méthodes plus générales, capables de minimiser des critères d'indépendance proches de la mesure théorique d'indépendance statistique (cf. 5.4). Du problème initial de la séparation de sources, a alors émergé le nouveau concept d'Analyse en Composantes Indépendantes (ACI) ([Jutten 87], [Comon 94]).

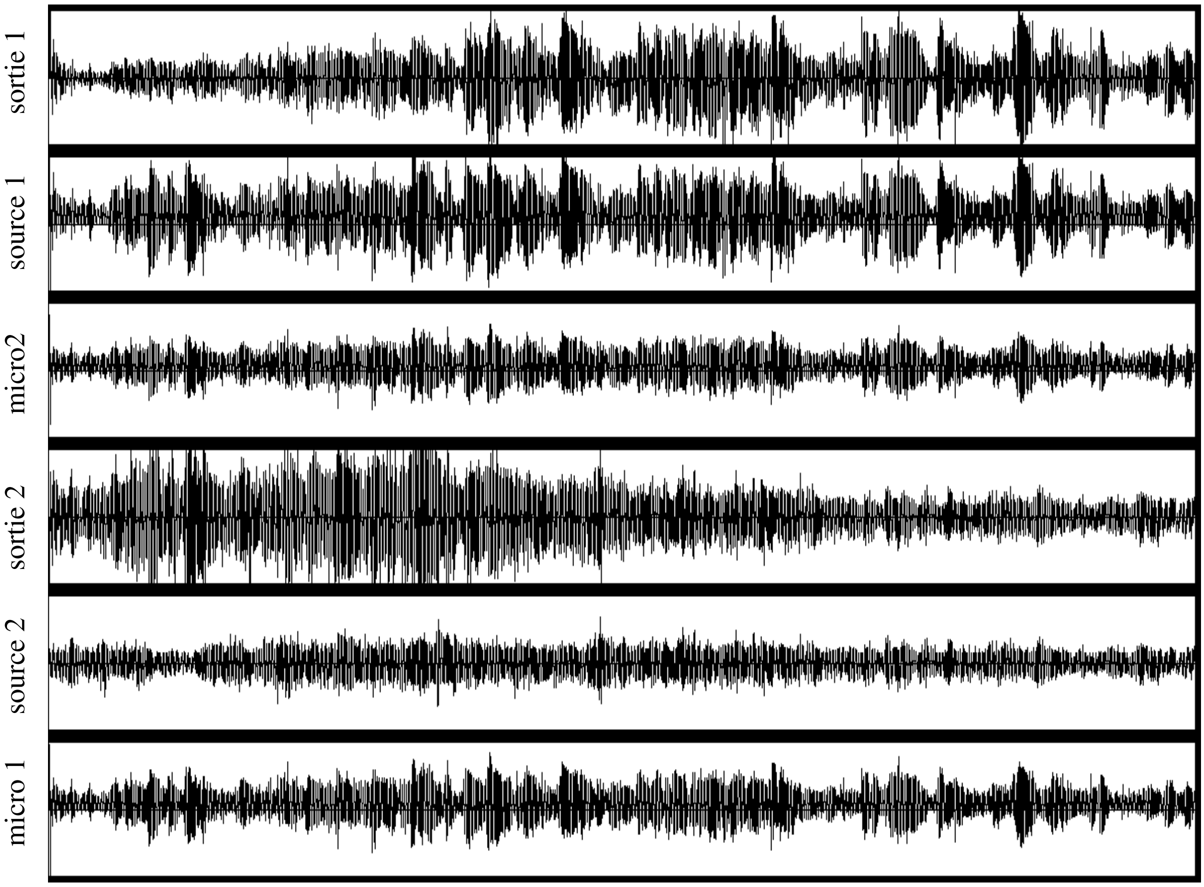

FIG. 5.2: *Exemple de séparation aveugle de deux sources acoustiques : les signaux d'entrée (micro1 et micro2) sont issus du mélange des signaux-sources (source 1 et source 2). Ces mixtures sont dissociées par le réseau qui produit des signaux de sortie (sortie 1 et sortie 2) correspondant, après stabilisation, aux deux sources.*

L'ACI a été présentée comme une généralisation des réseaux réalisant des analyses en composantes principales et, d'une certaine façon, comme une amélioration des systèmes de séparation de sources à fonction de coût linéaire. Introduite par Hérault et Jutten en 1985, elle est à mettre en relation avec des travaux menés sur des thématiques *a priori* différentes : architectures neuromimétiques et analyse de données ([Oja 85]), étude du codage sensoriel, théorie de l'information. En particulier, le problème de la recherche d'espaces minimisant la dépendance des paramètres avait été étudié sous une formulation différente, dans le domaine de l'étude du codage sensoriel et de l'optimisation de la représentation de l'information par réduction des redondances. Après l'*économie de codage* de Attneave ([Attneave 54]), Barflow avait introduit l'idée de réduction des redondances, qui lui semblait être une stratégie biologiquement plausible de traitement bas niveau des signaux sensoriels. Cette notion a ensuite été développée pour aboutir au codage factoriel ([Barflow 89], [Redlich 93]). Linsker a proposé une approche directement inspirée de la théorie de l'information, qui consiste à maximiser l'information mutuelle entre signaux perçus et les sorties du réseau ([Linsker 89]). Cette voie a ensuite donné lieu à de nombreuses études ([Linsker 92], [Schuster 92], [Bell 95], [Nadal 94], [Parra 95]).

Issue de travaux en traitement neuromimétique du signal, l'ACI aborde donc le problème plus général, de la répartition des données dans un ensemble de canaux d'information. Les méthodes développées dans ce cadre pourraient constituer une alternative intéressante à l'ACP, essentiellement pour deux raisons. La première est que les composantes indépendantes constituent une base de l'espace de représentation dans laquelle les données sont statistiquement indépendantes, ce qui minimise la redondance d'information entre les axes. D'une certaine façon, cela correspond à la recherche d'un espace dans lequel les données sont optimalement représentées.

Le second avantage théorique de l'ACI réside dans sa capacité à redistribuer l'information dans l'espace de représentation, alors que les axes principaux issus d'une ACP sont hiérarchisés par ordre de représentativité : le premier axe est celui qui représente le mieux l'ensemble des données, le second est celui qui représente le mieux les données dans l'espace orthogonnal à l'axe principal, etc. L'ACI n'extrait pas de direction privilégiée, mais réorganise l'ensemble de l'espace de représentation de manière à ce que chacun des axes porte une information originale.

Malheureusement, on a vu que les méthodes de SAS étaient fondées généralement sur des hypothèses qui peuvent restreindre leur champ d'application. D'une part, on suppose que la solution existe, c'est à dire qu'il y a une base dans laquelle les données sont statistiquement indépendantes. Cette hypothèse correspond à celle d'indépendance des signaux pour le problème de séparation de sources. Le fait que les points d'équilibre d'un algorithme correspondent à un état d'indépendance des sorties ne signifie pas que l'algorithme converge vers un état stable si la solution n'existe pas. Par ailleurs, de nombreuses techniques reposent sur des hypothèses au sujet des distributions statistiques des signaux.

Ces limitations ne signifient pas que l'ACI ne peut pas être utilisée, dans la pratique, comme une technique d'analyse de données, mais plutôt qu'il faut étudier systématiquement l'adéquation de la méthode choisie et du problème à traiter.

5.7 Conclusion

Les modèles neuromimétiques présentés dans ce chapitre permettent de dissocier des signaux qui sont mélangés linéairement. Cette opération est faite par une auto-organisation du réseau qui conduit à minimiser une fonction de coût qui est une mesure de dépendance statistique des signaux de sortie. Ces techniques permettent donc de restituer une organisation sous-jacente des données qui est masquée par le mélange des sources.

Le problème traité dépasse le cadre du problème initial de séparation de sources pour aborder celui, plus général, de la représentation de l'information dans des espaces multi-dimensionnels.

Dans cette optique, nous avons utilisé les résultats récemment obtenus en séparation de sources pour traiter du problème de l'adaptation de notre classifieur de bruits impulsifs. Le chapitre suivant présente, dans le détail, un processus de réorganisation des classes dérivé de l'algorithme de séparation de sources de Matsuoka.

Chapitre 6

Détection et classification des mixtures d'événements acoustiques

Ce chapitre décrit une règle d'auto-organisation du réseau projectif présenté au chapitre 4. Il s'agit d'un processus d'adaptation qui permet de retrouver l'image spectrale moyenne d'une classe à partir du mélange des signaux.

Nous commençons par montrer que la séparation de classes est un problème qui peut être traité par des méthodes de séparation de sources. Ce résultat nous permet d'implanter, dans l'architecture décrite au chapitre 4, une règle d'apprentissage inédite basée sur un algorithme de séparation de sources : les prototypes sont déplacés vers des positions minimisant la dépendance statistique des classes.

Les performances de l'algorithme sont évaluées à la fois en détection de rupture et en classification. Les expériences ont été réalisées sur des signaux artificiellement mélangés. Nous étudions l'évolution des performances en fonction du niveau de bruit et de la diversité des bruits impulsifs.

6.1 Introduction

6.1.1 Mixtures des prototypes

Le réseau neuronal décrit au chapitre 4 permet la détection d'événements acoustiques simultanés, à condition que les prototypes des classes correspondantes soient connus. Malheureusement, cette dernière condition n'est réalisée que lorsque la première occurrence d'un événement d'une classe apparaît seule, sans superposition avec d'autres bruits, qu'il s'agisse de bruits impulsifs ou d'un bruit de fond. En effet, la classe créée à la suite d'une détection de rupture possède, comme prototype, le vecteur stimulus qui a provoqué la rupture. Si cet événement survient simultanément à l'apparition d'un événement d'une autre classe ou mélangé à un bruit de fond, alors le prototype ne sera pas correct, puisqu'il sera le résultat d'une superposition de plusieurs formes acoustiques.

Le pire scénario qui peut se produire est l'apparition de nouveautés simultanées : dans ce cas, deux classes sont créées, mais chacune d'elles représente un peu des deux événements acoustiques et aucune d'elles n'est caractéristique d'une seule catégorie. La mauvaise initialisation d'une classe provoque ensuite une mauvaise formation de l'ensemble des classes intégrées au système et, d'une façon générale, une mauvaise qualité de la classification. Voici un exemple de sénario *catastrophe*, dans lequel trois vecteurs stimuli surviennent successivement. Ces trois vecteurs sont en réalité trois mélanges de quatres événements acoustiques indépendants notés $P_{c_0}, P_{c_1}, P_{c_2}, P_{c_3}$:

$t = 0$; $X(0) = P_{c_0} + P_{c_1}$; $\mathrm{P}_0 = X(0)$ est intégré
$t = 1$; $X(1) = P_{c_1} + P_{c_2}$; $\mathrm{P}_1 = X(1)$ est intégré ; P_0 et P_1 sont actifs
$t = 2$;$X(2) = P_{c_0} + P_{c_1} + P_{c_2} + P_{c_3}$;$\mathrm{P}_2 = X(2)$ est intégré ; P_0 et P_1 et P_2 sont actifs
etc..

On voit qu'une telle situation a de multiples effets pervers :

- le partage des catégories "réelles" en multiples classes réduit le pouvoir discriminant d'une cellule, qui est sa capacité à distinguer une classe d'une autre,
- la mauvaise formation d'une classe peut lui permettre de se maintenir, par des

activations plus fréquentes, à une température suffisante pour qu'elle ne soit pas supprimée. Ceci peut se produire malgré une représentativité très partielle de la classe (dans notre exemple, $P(0)$ est toujours actif),

- la mauvaise formation des classes réduit la dynamique des cellules correspondantes, ce qui augmente la difficulté de la décision d'appartenance à une classe.

Cependant, la mauvaise formation des classes n'influence pas directement l'activité de la cellule de nouveauté. En effet, bien que les prototypes soient mal positionnés, l'espace qu'ils engendrent correspond globalement bien à l'espace-signal : chaque prototype incorrect s'exprime comme une combinaison linéaire de prototypes corrects. Par ailleurs, les prototypes forment toujours une famille libre de l'espace-système, puisqu'ils ont été intégrés à cause de l'importance de leur composante orthogonale à l'espace-prototype. L'espace engendré par les prototypes corrects et celui engendré par leur mixtures sont donc identiques.

L'espace nouveauté n'est donc pas immédiatement influencé par le mauvais positionnement des prototypes. Cependant, ce phénomène perturbe le processus de classification et celui d'élimination des classes. À terme, c'est donc l'espace des prototypes lui-même qui subit les conséquences de la mauvaise qualité de la classification apprise. Ceci perturbe l'ensemble du processus de détection et de classification.

La figure 6.1 illustre les effets du mauvais positionnement des prototypes dans le cas de ruptures simultanées du modèle. La première ligne de ce shéma représente le signal capté, la seconde, le spectrogramme correspondant. La troisième et la quatrième ligne représentent les activités de deux des classes produites par le système. Le signal est composé de deux événements récurrents, dont les premières occurences sont simultanées. La première cellule créée correspond donc à une rupture provoquée par un stimulus composé de la superposition de deux événements de types différents. On notera ces deux types dévénements e_1 et e_2. e_1 est le premier à être détecté. La cellule correspondante modélise le mélange des classes. La seconde cellule correspond à la détection des événements e_2. On peut observer que la cellule C_1 est faiblement réactive aux deux types de stimuli. Par contre, elle réagit plus fortement au mélange, qui correspond à la configuration du signal au moment de la création de la classe. C_2 modélise la classe des événements de type e_2. Par contre, aucune cellule ne modélise les événements e_1. Cette situation est typique du problème posé par le mélange des classes : un cellule modélise mal différents événements acoustiques, tout en réduisant la sensibilité du détecteur à un certain type de stimuli. Cette situation provoque donc des confusions dans les détections (une cellule est réactive à plusieurs classes), et, globalement, une mauvaise représentation des composantes du signal. Par ailleurs, on peut observer une corrélation des activités des deux cellules.

Le risque de mauvais positionnement des prototypes peut devenir très pénalisant dans un milieu acoustique composé de sources fréquemment actives. Résoudre ce problème pourrait consister à retrouver une représentation spectrale moyenne des événements acoustiques à partir de l'observation de leurs mixtures. Ce chapitre traite de ce sujet et décrit une méthode d'adaptation des poids synaptiques permettant un fonctionnement correct du système dans les situations critiques qui ont été décrites.

6.1.2 Indépendance statistique des classes

Soient P_{c_j} les prototypes qui représentent effectivement l'image spectrale type d'une catégorie de bruits impulsifs. Soient P_i les prototypes des classes acquis par le réseaux projectif. Ces vecteurs sont des mixtures des prototypes corrects :

$$P_i = \sum_{j=0}^{j=n} m_{i,j}.P_{c_j}$$

ou encore, en notation matricielle :

$$P = M.P_c$$

où P est la matrice dont chaque ligne est un vecteur prototype appris, P_c^t la transposée de la matrice des prototypes corrects et $M = (m_{i,j})$ la matrice de mélange. L'activité temporelle des cellules de sortie du réseau projectif peut donc être exprimée en fonction de la matrice des prototypes corrects Pc et des stimuli $V(t)$ par :

$$A(t) = (P_c.M^t)^{-t}.V(t)$$

Notre problème est de retrouver P_c à partir de P et des observations $A(t)$, sans connaître M.

On peut considérer les $a_i(t)$ comme n signaux temporels. Soit $a_{c_i}(t)$ les activités des cellules qui auraient comme prototypes les vecteurs Pc. Il s'agit des activités des cellules de sortie du réseau projectif dont les prototypes seraient correctement positionnés :

$$A_c(t) = P_c^{-t}.V(t)$$

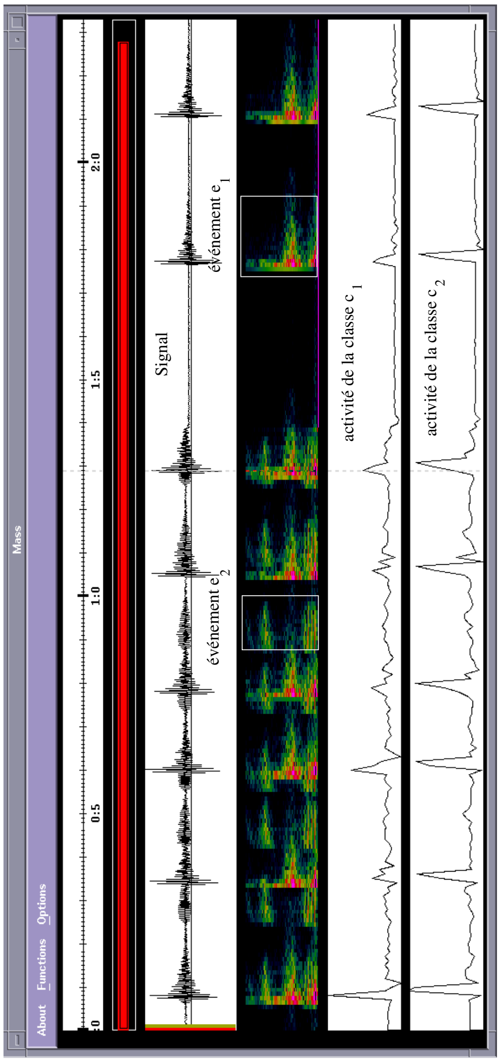

FIG. 6.1: *Erreur de classification provoquée par des ruptures simultanées du modèle. Deux événements récurrents e_1 et e_2 apparaîssent au même instant dans le signal. Ceci provoque la création de deux cellules c_1 et c_2 modélisant respectivement un mélange de e_1 et e_2. La cellule c_1 ne caractérise aucune classe mais insensibilise le système et empêche la modélisation de e_1.*

Si la matrice de mélange est inversible, alors l'équation précédente peut s'écrire :

$$A(t) = P^{-t}.V(t) = (P_c^{+}.M^{-1})^t.V(t) = M^{-t}.P_c^{-t}.V(t) = M^{-t}.A_c(t)$$

On voit que le mélange des prototypes provoque le mélange des activités des cellules de sortie.

On peut considérer ces sorties comme des signaux temporels. Après les avoir centrées (moyenne nulle), ces signaux ont une forme qui pourrait être celle de signaux acoustiques. Chaque canal de sortie du classifieur peut donc être assimilé à un microphone qui capterait des signaux émis par différentes sources. Dans ce cas, le mélange des activités des classes peut être considéré comme le mélange des signaux-sources. Inversement, dissocier les contributions de ces différentes sources virtuelles revient à annuler la dépendance des activités des classes correspondantes. La recherche de l'indépendance statistique des activités peut être considérée comme un problème de séparation aveugle de sources. Mais pour appliquer une méthode de SAS, il faut que les signaux satisfassent certaines contraintes.

La première d'entre elles porte sur la distribution des signaux sources. Ils sont fréquemment supposés tous sur-gaussiens ou tous sous-gaussiens, ce qui correspond à une condition sur le signe du Kurtosis J_{kurth} défini par :

$$J_{kurth}(S) =_{i=1}^{M} |E(s_i^4) - 3E^2(s_i^2)|$$

Matsuoka propose de s'affranchir de cette hypothèse en adoptant celle de non-stationarité des signaux (5.5.3). Cette dernière contrainte nous paraissant mieux adaptée à notre problème, nous nous sommes orienté vers cette méthode. L'hypothèse de non-stationnarité correspond à supposer que la variance de l'activité des cellules doit être mobile, ce qui semble vraisemblable dans des conditions réelles. Néanmoins, cette supposition pourrait être mise en défaut si les signaux sont périodiques et d'une période correspondant à la durée d'estimation des paramètres statistiques du séparateur. Cette situation étant relativement marginale, on n'en tiendra pas compte ; de plus, on peut envisager d'adapter la durée d'estimation de façon à éviter ce type de problème.

Par contre, l'inactivité d'une classe pendant une durée significative est une situation très probable, et qui conduirait à la stationnarité locale du signal, ainsi qu'à une déstabilisation du réseau. Ce problème est abordé dans la section (cf. 6.2.2). Une solution y est décrite sous la forme d'une modification de la règle initialement proposée par Matsuoka.

Par ailleurs, les signaux doivent avoir une moyenne nulle, ce qui peut être fait par un simple centrage. Cette normalisation en ligne des données introduira une erreur d'estimation qui sera négligée ici.

Le nombre de sources détermine la structure du réseau et doit être connu à priori. Quelques auteurs ont proposé des méthodes permettant de s'affranchir de cette condition souvent difficile à satisfaire dans des conditions réelles ([Jutten 87], [Comon 90] [Gaeta 90], [Parashiv 97]).

Enfin, les méthodes de SAS supposent que la solution existe, c'est à dire qu'il existe n signaux indépendants qui sont mélangés selon un opérateur linéaire de rang n. Cette condition en regroupe deux : la matrice de mélange doit être inversible et les signaux doivent être statistiquement indépendants. Ceci correspond, dans notre cas, à une hypothèse sur la dimension de l'espace signal effectif : si n classes d'événements acoustiques indépendants composent effectivement l'environnement, alors les n signaux de sortie résultent du mélange des activités selon l'opérateur M, qui est l'opérateur de mélange des prototypes. On a vu que les vecteurs de base et leurs mélanges engendrent le même sous-espace ; M est donc inversible de rang n. Cette contrainte est satisfaite systématiquement ici, alors qu'elle dépend des positions relatives des sources et des capteurs dans le cas général de la séparation de sources. L'hypothèse d'indépendance statistique des signaux est moins facile à traiter ; elle correspond, dans notre analogie, à l'indépendance des activités des classes. Pratiquement, cela signifie qu'on considère que deux phénomènes physiques qui doivent être dissociés sont indépendants ; s'ils sont liés, alors les signaux peuvent être regroupés dans une même classe, puisqu'ils véhiculent la même information. Bien que deux événements acoustiques de classes différentes puissent être ponctuellement simultanés, on admettra qu'il n'y a généralement pas de lien statistique entre les classes de bruits impulsifs (ce qui n'exclut pas les corrélations temporelles).

Sous ces hypothèses, on peut envisager l'utilisation d'un réseau séparateur de sources pour dissocier les activités des cellules de sortie du réseau projectif. La section suivante montre comment une telle architecture peut être compilée pour permettre une modification directe des classes qui aboutit à leur indépendance statistique.

6.2 Adaptation des poids synaptiques par A.C.I.

6.2.1 Principe

Nous avons vu que le mélange des prototypes conduisait au mélange des signaux de sortie du réseau. En plaçant un réseau séparateur de sources à la sortie du classifieur, on obtiendrait les activités d'un classifieur virtuel possédant des prototypes correctement positionnés. La figure 6.7 montre le schéma de principe d'un classifieur à deux niveaux : le premier déterminant de manière dynamique et adaptative l'espace-signal effectif, le second réalisant un mélange inverse des activités normalisées des cellules de sortie de façon à minimiser leur dépendance.

La fonction de transfert d'un tel réseau serait :

$$S(t) = D.(A(t) - < A >)$$

avec

$$A(t) = P^{-t}.V(t)$$

où $< S_1 >$ est le vecteur d'activité moyenne et D est la matrice des poids du réseau séparateur de sources. D peut s'écrire en fonction de la matrice C des connexions récurrentes du réseau, conformement aux notations définies dans le chapitre précédent :

$$D = (I + C)^{-1}$$

On peut donc exprimer $S(t)$ en fonction de P et V par :

$$\begin{aligned} S(t) &= D.P^{-t}.V(t) + \zeta \\ S(t) &= (D^{-t}.P)^{-t}.V(t) + \zeta \end{aligned}$$

où ζ est un vecteur constant.

Cette équation montre qu'appliquer le réseau séparateur de sources à la sortie du réseau classifieur est équivalent à appliquer l'opérateur de décorrélation inverse à la matrice des prototypes, à une constante additive (ζ) près. On peut donc utiliser la règle d'apprentissage des poids synaptiques du réseau séparateur de sources comme une loi d'évolution des prototypes : en appliquant D^{-t} à P, on déplace les prototypes de façon à minimiser la dépendance statistique des cellules de sortie.

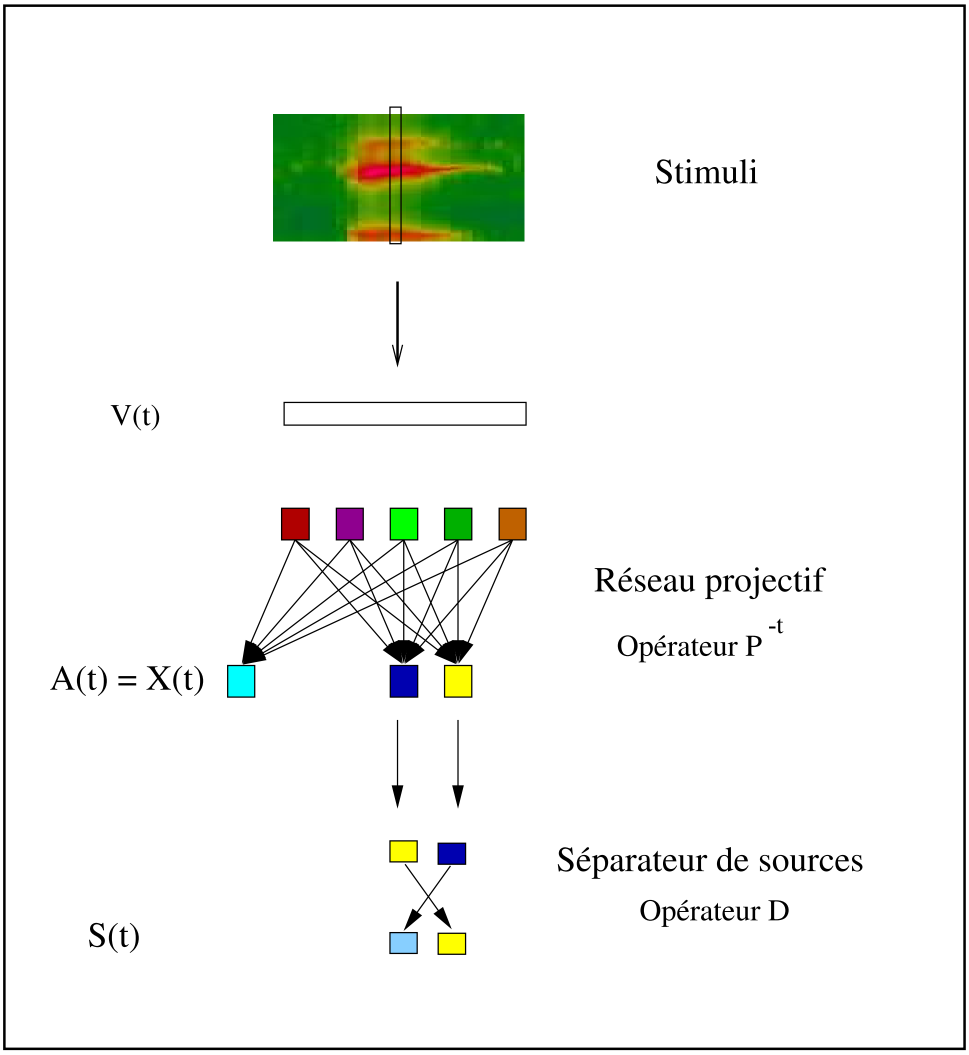

FIG. 6.2: *Recalcul des activités par un réseau séparateur de sources : les activités des cellules de sortie sont considérées comme des signaux temporels. Ces signaux constituent les entrées d'un réseau séparateur de sources. Ceci permet de dissocier les activités des classes dont les prototpyes représentent le mélange d'images spectrales moyennes de différentes classes d'événements acoustiques.*

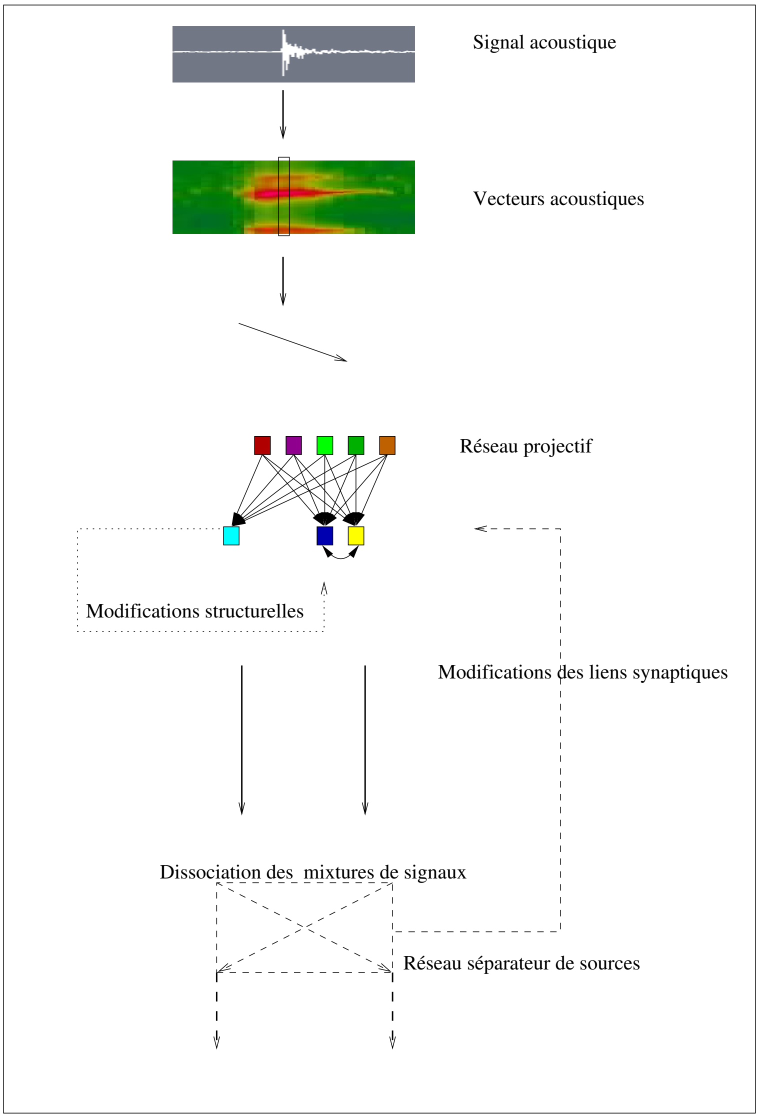

FIG. 6.3: *Schéma de principe de l'architecture du classifieur en composantes indépendantes : le signal acoustique est paramétrisé par une analyse temps fréquence classique. Les vecteurs acoustiques issus de cette analyse sont les stimuli du réseau projectif. Le réseau s'adapte aux évolutions des entrées par modification de sa structure. Les activités des cellules de sorties permettent de capturer un opérateur neuronal de décorrélation. Une règle d'évolution des prototypes est déduite de cet opérateur. Les classes ainsi modifiées sont celles pour lesquelles la dépendance statistiques des sorties est minimale.*

La figure 6.3 montre le schéma de principe du système de détection et de classification global. Les deux phases d'apprentissage du modèle — adaptation de la structure du réseau et modification des poids — y sont dissociées. La figure 6.7 décrit l'architecture définitive qui est adoptée.

6.2.2 Règle de séparation

La règle de mise à jour des poids du réseau séparateur de sources proposée par Matsuoka est :

$$T\frac{dC}{dt} = (I + C^t)^{-1}(diag < S(t).S(t)^t >)^{-1} < S(t)S(t)^t > -I$$

Dans le cas d'une mise à jour incrémentale du réseau, le terme

$$(diag < S(t), S(t)^t >)^{-1} < S(t)S(t)^t >$$

doit être estimé en ligne. Ceci peut être fait dynamiquement en utilisant une fonction Φ estimant la variance des sorties du réseau :

$$\Phi(t + dt) = \alpha\Phi(t) + (1 - \alpha)S_i(t)^2$$

Avec cette règle, l'inactivité d'une source déstabilise le réseau. En effet, $\frac{dc_{ij}}{dt}$ est d'ordre -1 en $y_i(t)$.

Afin de réduire la contrainte de non-stationnarité des sources et de pouvoir appliquer le réseau à la séparation de signaux impulsifs, nous avons rajouté un terme exponentiel régularisant la dynamique des poids. Ce terme module la modification des poids lorsque l'écart-type des signaux de sortie est quasiment nul et est peu influencé lorsque cette valeur augmente. Cette modulation est faite par la fonction :

$$f(x) = \frac{1 - e^{-2|x|}}{1 + e^{-2|x|}}$$

Cette fonction est équivalente à x autour de 0 et tend rapidement vers 1 lorsque x augmente ; la figure 6.4 montre sa courbe représentative.

La règle définitive est donc :

$$T\frac{dC}{dt} = (I + C^t)^{-1} f(S(t)$$

.diag(¡S(t),S(t)^t $>)^{-1} < S(t), S(t)^t > -I$

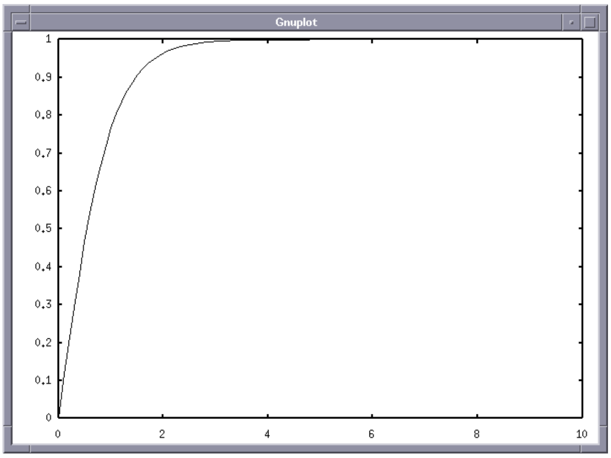

FIG. 6.4: *Modulation du pas d'apprentissage : fonction représentative du terme régulant la dynamique des poids synaptiques dans le cas de périodes stationnaires du signal. Cette fonction est équivalente à x dans le voisinage de zéro, et tend rapidement vers 1 quand x augmente. Elle permet une coupure de l'apprentissage dans les périodes stationnaires.*

De façon à tester cette modification, nous avons effectué des tests sur des signaux acoustiques comportant des silences de durées significatives. Il s'agit de lettres de l'alphabet prononcées isolement. Nous avons observé l'influence de la nouvelle règle sur la dynamique des poids du réseau.

Les figures 6.5 et 6.6 montrent l'apport de la nouvelle règle sur la séparation de 2 signaux de parole artificiellement mélangés. Les trois premières lignes représentent un canal de sortie, le signal effectivement émis par la source, puis le signal reçu par un des micros. Les trois lignes suivantes représentent les trois autres signaux (sortie, source et micro). Enfin, les deux dernières lignes montrent l'instabilité des poids du réseau.

On voit que, dans le premier cas, les modifications des liens sont brutales et deséquilibrées par les périodes de silence des sources. Cette instabilité est fortement réduite dans le second cas, ce qui conduit à une accélération de la convergence.

6.2.3 Adaptation en ligne du réseau

En appliquant la règle de séparation présentée ci-dessus, on exploite les liens statistiques entre les sorties du réseau pour déplacer les prototypes. Cette règle complète le processus d'adaptation par ajout et suppression de cellules. En réalité,

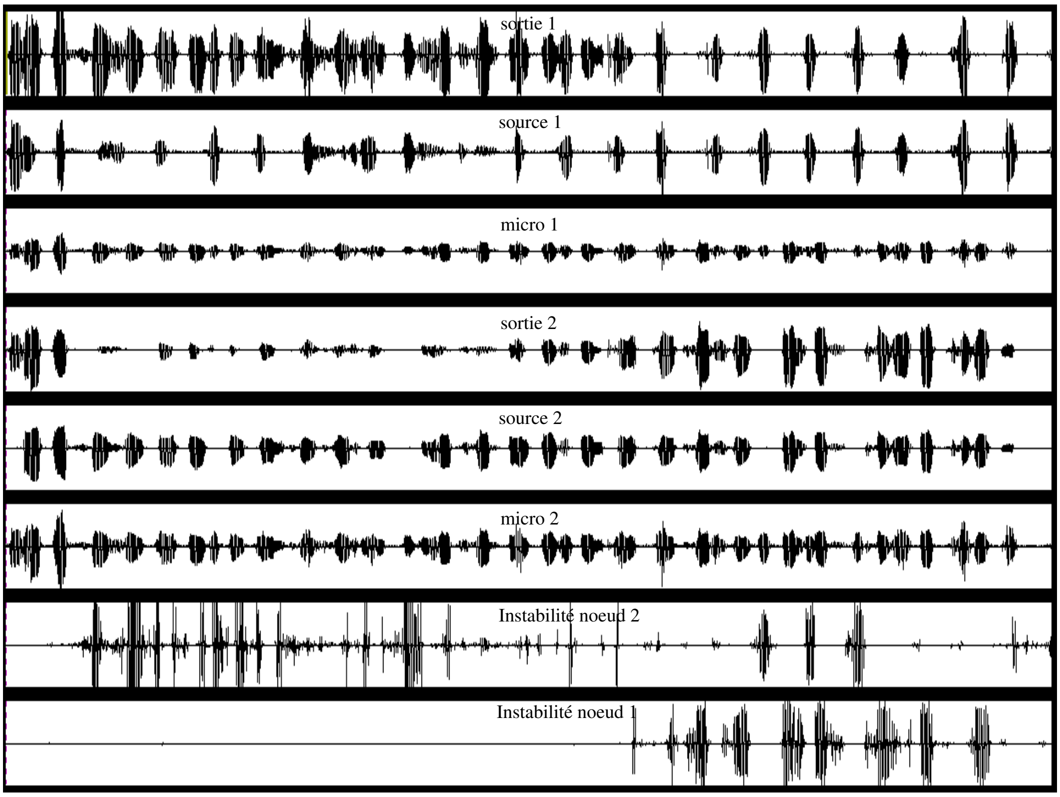

FIG. 6.5: *Séparation de sources de bruits transitoires avec l'ancienne règle d'apprentissage : des mélanges des signaux-sources 1 et 2 sont perçus par deux capteurs (signaux micro 1 et micro 2), puis dissociés en deux signaux de sorties. Les deux dernières lignes du schéma montre la modification des poids synaptiques obtenue sans modulation de l'apprentissage. Les moments de silences ponctuels d'une des sources provoque une instablité des poids qui déséquilibre et ralentit fortement l'apprentissage.*

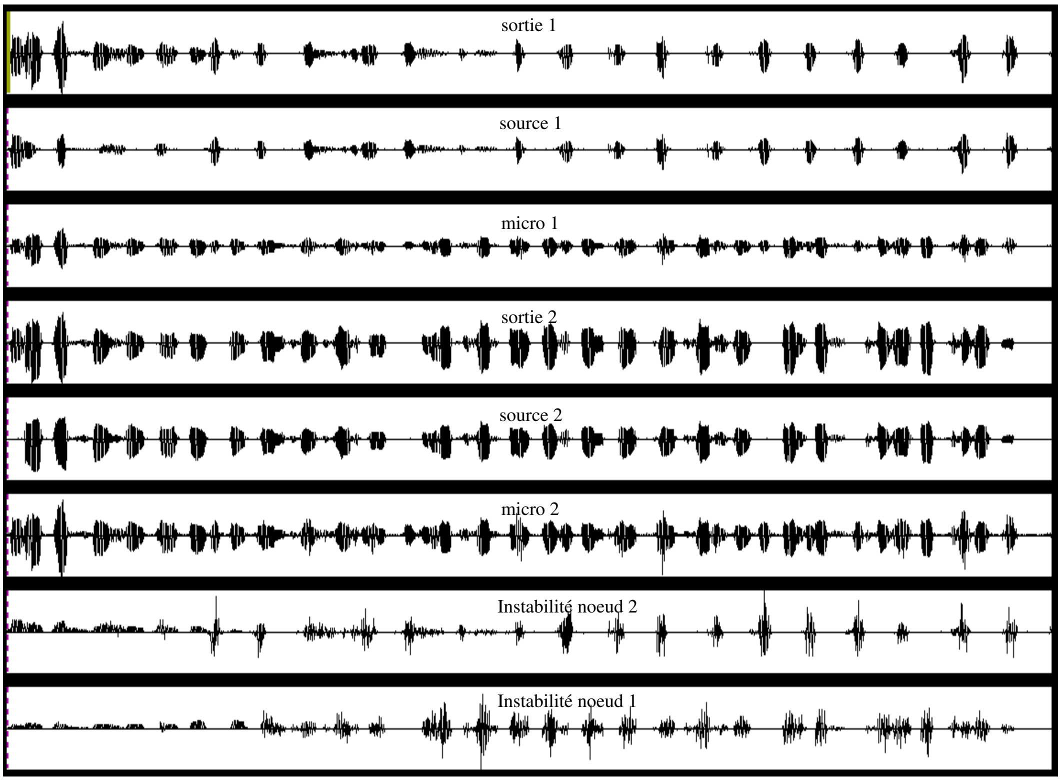

FIG. 6.6: *Séparation de sources de bruits transitoires avec modulation de la dynamique d'apprentissage : ce schéma correspond à la même situation que celle de la figure 6.5, mais la dynamique des poids est modulée. On voit que l'apprentissage n'est plus perturbé par les moments de silence d'une des sources. Un état stable est obtenu plus rapidement.*

ces deux processus ne sont pas concurrents ; en effet, tous les événements ne sont pas récurrents et l'estimation des paramètres statistiques des classes de faible population peut devenir problématique. Néanmoins, les classes d'événements isolés auront, à terme, une température faible qui suscitera leur élimination du modèle. Le processus de création et de suppression de classes garantit donc une activité minimale des classes représentées et une certaine quantité de données pour l'estimation des paramètres statistiques. Malheureusement, ceci ne signifie pas qu'on disposera systématiquement des données nécessaires à la séparation de classes mélangées. Cependant, la probabilité de superposition d'événements est d'autant plus faible que leur fréquence est faible.

Par ailleurs, la règle de séparation ne modifie que les prototypes qui sont statistiquement dépendants. Le problème de l'estimation des corrélations (non-linéaires) ne se pose que dans le cas d'événements peu fréquents et qui sont apparus sous forme de mixtures. D'une certaine façon, les deux processus d'adaptation du modèle à l'environnement sont complémentaires.

L'algorithme d'adaptation en-ligne peut être décrit en trois processus :

1. propagation du stimulus dans les cellules de sorties,
2. mise à jour des températures,
3. si la cellule de nouveauté est supérieure au seuil de vigilance, création d'une cellule,
4. élimination des cellules dont la température est trop faible,
5. calcul du terme séparateur suivant la règle :
$$T\frac{dC}{dt} = (I + C^t)^{-1}(f(S(t)).diag(< S(t), S(t)^t >))^{-1} < S(t)S(t)^t > -I)$$
,
6. mise à jour des prototypes par $P_{t+dt} = P.(I + C)^{-1}$,
7. calcul des poids synaptiques : $W_{t+dt} = P_{t+dt}^{-t}$.

La figure 6.7 décrit l'architecture neuronale par laquelle cet algorithme est mis en œuvre.

6.2.4 Classification de signaux multi-dimensionnels

Le premier test est simplement destiné à illustrer les capacités du modèle à dissocier des formes spectrales superposées.

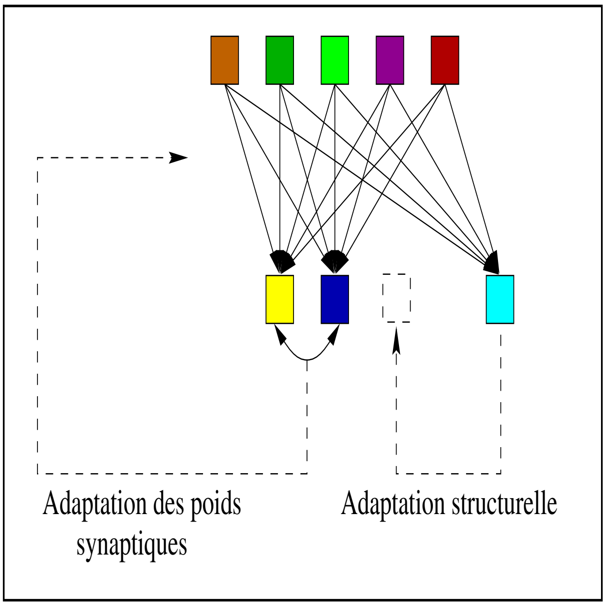

Fig. 6.7: *Architecture réelle du réseau classifieur : deux couches sont totalement interconnectées. Un double processus d'adaptation modifie le réseau, soit en modifiant sa structure, soit en modifiant les poids synaptiques de façon à minimiser la dépendance statistique des sorties.*

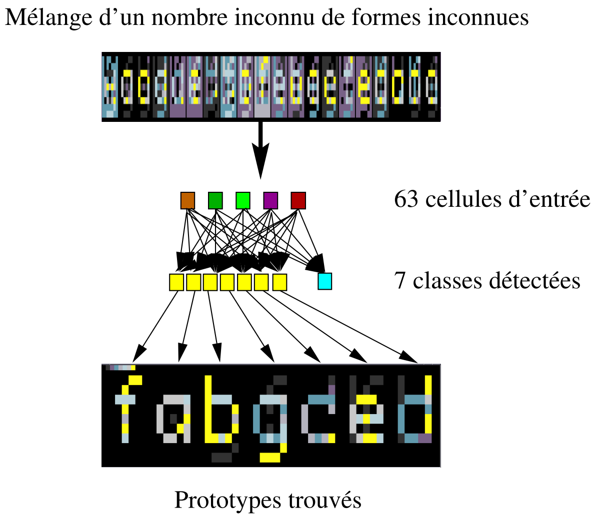

Fig. 6.8: *Test sur des signaux artificiels mélangés aléatoirement : des superpositions d'images de lettres de l'alphabet sont présentés au réseau. La contribution de chaque motif au mélange est choisie aléatoirement pour chaque exemple. Après stabilisation, les prototypes trouvés correspondent bien aux 7 images mélangées ([Linares 97a]).*

Nous avons extrait de l'alphabet les motifs des 7 premières lettres. Ces motifs $(P_i)_{1\leq i\leq 7}$ ont été mélangés aléatoirement en 500 mixtures, qui constituent les entrées successives du réseau. Les vecteurs d'entrée sont obtenus par concaténation des lignes d'une matrice de dimension 7x9. Ces vecteurs sont de dimension 63 et représentent des superpositions des 7 motifs. Les motifs des lettres sont linéairement indépendants et tous leurs pixels ont la même couleur (c'est à dire qu'on leur attribue la même valeur). Ce dernier point n'est cependant pas nécessaire. Bien entendu, la forme et le nombre des motifs sont inconnus du système.

Les entrées du réseau sont les vecteurs $X(t)$ de la forme :

$$X(t) = \sum_{i=0}^{i=N} r_i(t) P_i$$

Ici, toutes les formes sont systématiquement superposées. La figure 6.8 montre quelques entrées et les classes trouvées.

Le nombre de formes mélangées a été correctement trouvé, ce qui était évident dans ce cas. Les motifs ont été retrouvés alors qu'ils ne sont apparus que sous forme de mixtures. Cet exemple illustre le processus de séparation de classes opéré par le modèle qui est proposé ici.

6.3 Expériences

6.3.1 Introduction

Dans ce travail, nos choix ont été guidés par le fait que les milieux réels sont variables et imprévisibles. Ces caractéristiques rendent leur modélisation difficile et ont dirigé nos études vers des systèmes dynamiques et auto-adaptatifs. Dans ce contexte, l'évaluation précise des performances du système pose plusieurs problèmes fondamentaux ou plus pratiques.

En premier lieu, il peut paraître contradictoire d'évaluer les performances sur un corpus réalisé dans des conditions de laboratoire, alors que l'approche dynamique qui a été adoptée est justifiée par l'impossibilité de constuire *a priori* une image suffisamment fidèle du signal réel. Cependant, il est délicat d'extraire des informations fiables de tests réels dont les conditions ne peuvent pas être précisement décrites. De plus, s'il est difficile de construire un corpus d'apprentissage, il est aussi difficile de construire un corpus d'évaluation. La diversité des situations réelles susceptibles d'être rencontrées rendrait les résultats obtenus en conditions réelles dépendants des données de test et difficilement généralisables.

Par ailleurs, pour mesurer les performances du système, il faut disposer de données de tests et des résultats attendus, de façon à comparer les sorties du réseau et celles que l'on désire. Cela pose le problème pratique de la collection des données et de leur étiquetage, ainsi que celui, plus général, du choix éventuellement subjectif des cibles du classifieur : comment dissocier ce qui doit être détecté de ce qui ne doit pas l'être, à partir de quel niveau d'énergie un bruit doit-il être détecté, quelle est la finesse de la classification qui est attendue, etc.

Deux séries d'expériences ont été réalisées.

La première série d'expériences permet l'étude du comportement du système dans des situations bien définies. Il s'agit de tests sur des signaux artificiellement mélangés. Nous mesurons les performances du modèle en classification et en détection de ruptures en fonction du niveau de bruit, puis en fonction de la densité et de la diversité des bruits impulsifs.

La seconde porte sur des signaux enregistrés à bord de sous-marins de la marine nationale. Ces signaux sont composés de bruits transitoires d'origines biologiques ou mécaniques. Cette étude a été réalisée au sein de l'entreprise *DCN Ingénierie Sud*. Elle s'insère dans un projet de système d'aide à la veille acoustique qui réunit plusieurs approches. Cette étude étant confidentielle, nous ne décrirons pas le détail des expériences menées, ni les résultats obtenus. Cependant, la démarche adoptée et la façon dont nous avons fixé le processus expérimental peuvent être présentées.

6.3.2 Expériences dans des conditions contrôlées

Signaux de tests

De façon à étudier l'impact de certains paramètres sur les performances du système, nous avons collecté des signaux impulsifs réels qui ont été utilisés pour composer des signaux correspondant aux situations étudiées. 10 formes acoustiques différentes forment la base de bruits impulsifs. Les images temps-fréquences de ces bruits sont données dans l'annexe.

Pour chaque signal de test, on extrait aléatoirement n sons $P_{1 \leq i \leq n}$ parmi les dix de cette base. Ces bruits constituent les classes qui doivent être détectées et reconnues par le système.

ν_i occurrences de chacun des bruits sélectionnés sont insérés aléatoirement entre les instants $t_0 = 3s$ et t_f.

Enfin, on rajoute au signal ainsi constitué un bruit blanc de façon à avoir un rapport de puissance bruit impulsifs / bruit de fond de l'ordre de r décibels. À chaque fichier de signal créé, on associe un fichier d'étiquettes qui décrit les classes insérées et les moments des insertions.

Paramétrisation du réseau

Pour ces expériences, nous avons fixé les différents seuils du réseau. Ces réglages ont été faits à partir d'un observation empirique de signaux à traité. Ils ne seront pas modifiés au cours des différents tests. Le seuil de vigilance, qui détermine l'activité limite de la cellule de nouveauté provoquant la création d'une classe, est fixé à 0.3,

La valeur minimale de la température d'une cellule est fixé à 0.5 (seuil d'élimination).

Le coefficient calorimétrique γ des cellules est à 1. Le paramètre α de pondération température / activité pour la mise à jour de la température est fixé à 0.9.

Pour des raisons pratiques, le traitement des signaux est fait en deux passes : la première détermine le nombre de classes détectées, l'autre réorganise les classes suivant la règle de minimisation de la dépendance statistique des cellules.

Préparation des sorties du réseau

Le système présenté est non-supervisé. Il n'y a donc pas d'étiquetage des classes, celles ci étant créées automatiquement. Pour comparer les sorties du réseau et les sorties désirées, il reste à mettre en place un mécanisme d'association entre les classes construites et celles effectivement insérées.

Ceci pose un problème car il n'y a pas de bijection entre ces deux ensembles, pour deux raisons :

- bien que les événements acoustiques soient impulsifs et donc de durées restreintes, ils ont quand même une structure temporelle qui rend leur modélisation par une seule classe difficile. En réalité, dans la majorité des cas, plusieurs cellules vont caractériser une classe. Les prototypes de ces classes correspondent à des parties différentes de la même forme acoustique. La présence d'une forme référencée se traduira alors par les activations successives de plusieurs cellules. Laquelle de ces cellules doit-on choisir pour caractériser une classe de bruits ? Plusieurs solutions peuvent être envisagées. Par exemple, on peut associer toutes les cellules créées lors de la détection de la première occurrence d'un événement de la classe. Le score de la classe serait alors la moyenne des scores, ou la moyenne d'un pourcentage des meilleurs scores, etc.
- le réseau modélise l'ensemble de l'environnement acoustique. Certaines classes représentent donc des bruits impulsifs et d'autres sont dédiées au bruit de fond. Même si plusieurs classes représentent un même événement acoustique, une classe ne doit pas être forcément associée à un événement transitoire mais peut modéliser le bruit de fond.

En dehors du problème de l'absence de bijection entre les classes introduites dans le signal et celles créées par le réseau, l'étiquetage d'une cellule représentant une classe fait nécessairement intervenir l'opérateur humain.

Nous avons choisi d'associer la cellule i à la classe j qui minimise la somme du nombre de fausses détections et celui de non-détections, ces deux critères étant définis dans la section suivante.

Une fois la cellule choisie, il faut un critère permettant de décider à partir de quelle valeur on doit considérer qu'un événement de la classe a été détecté. En effet, l'activité $a_i(t)$ de la cellule est continue et non bornée. À chaque instant, il faut décider, en fonction de la valeur de la cellule, s'il y a eu détection. Cette étape, où les classes sont mises en relation avec les événements étiquetés du corpus de test,

est donc supervisée.

Nous utilisons un seuil normalisé par la dispersion des $a_i(t)$. L'activité continue a_i est transformée en une fonction binaire $F_i(t)$; lorsque cette fonction est à 1, la classe est détectée, lorsqu'elle est à 0, il n'y a pas de détection.

$$F_i(t) = \begin{cases} 1 & si \ |a_i(t) - < a_i > | > 2.\theta(a_i), \\ 0 & sinon \end{cases}$$

où $\theta(x)$ est l'écart-type de la variable x, et $< a_i >$ la moyenne de $a_i(t)$.

Mesures de performances

Les résultats sont obtenus en créant un grand nombre de signaux de test, puis en mesurant pour chacun d'eux les quatre paramètres suivants :

- **taux de non détections (suppressions)** : il s'agit du nombre d'événements acoustiques qui n'ont pas été détectés par le système. Ce nombre est exprimé en fonction du nombre total d'événements ajoutés au signal, quelle que soit leur classe.
- **taux de fausses détections (insertions)** : il s'agit du nombre de détections qui ne correspondent pas aux classes associées (un événement a été détecté alors qu'il n'y en a pas, ou qu'il n'est pas de la bonne classe). Ce nombre est exprimé en pourcentage relativement au nombre total d'événements à détecter. Ce pourcentage peut dépasser les 100 %.
- **taux de bonnes alarmes** : tandis que les paramètres précédents mesuraient les capacités du modèle en classification, ce critère concerne la détection de ruptures : il indique la proportion des ruptures bien détectées.
- **taux de fausses alarmes** : ce critère représente le nombre de ruptures détectées qui ne correspondent pas à l'apparition d'une nouvelle classe d'événements. Ce nombre est exprimé en pourcentage, relativement au nombre de ruptures à détecter.

6.3.3 Évolution des performances en fonction du bruit

Cette expérience est destinée à évaluer le comportement du système en présence d'un bruit ambiant uniformement réparti sur le domaine fréquentiel. Les signaux créés comportent deux classes d'événements, représentées par 20 occurrences chacune. Pour chacun des niveaux de bruit, 50 signaux sont construits suivant la

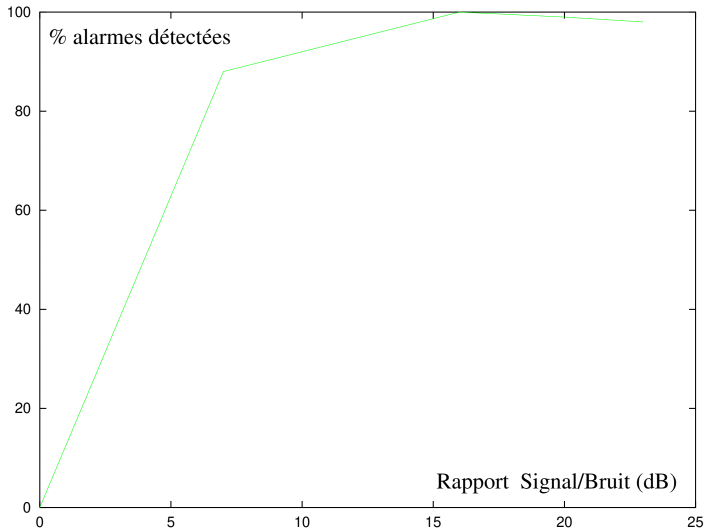

FIG. 6.9: *Proportion des alarmes émises qui correspondent bien à l'apparition d'une nouvelle classe de buits impulsifs. On constate de mauvaises performances à 0 dB, et des performances quasiment maximales dès 7 dB.*

procédure décrite plus haut. Les classes sont tirées aléatoirement parmi les dix disponibles, 3 secondes étant réservées en début de signal pour l'amorce du système. Chaque test dure 23 secondes. Le niveau de bruit est estimé en fonction de l'énergie moyenne d'un événement acoustique inséré et de celle du bruit blanc ajouté. Nous avons fait varier le rapport signal/bruit r de 0 dB à 27 dB et étudié l'influence de ce paramètre sur les taux de bonnes détections, de fausses détections, de ruptures correctes et de fausses alarmes.

La figure 6.9 montre l'évolution du taux de bonnes alarmes en fonction du bruit. À 0 dB, rien n'est détecté. Par ailleurs, aucune fausse alarme n'est produite. On peut conclure de cette observation que toutes les classes créées l'ont été durant les 3 secondes d'initialisation. Le niveau élevé de bruit blanc a conduit à la création d'un grand nombre de classes, ce qui a fait perdre au système toute sa sensibilité. Ce comportement illustre un défaut de l'architecture proposée, particulièrement lorsque le bruit est blanc : un tel bruit étant uniformement réparti sur l'ensemble du domaine fréquentiel, il n'y a pas de sous espace de l'espace de représentation dans lequel il peut être circonscrit. Toutes les dimensions seraient donc nécessaires à sa modélisation.

Les performances, en terme de détection de ruptures, augmentent jusqu'à atteindre quasiment la valeur maximale dès la seconde mesure à 7 dB.

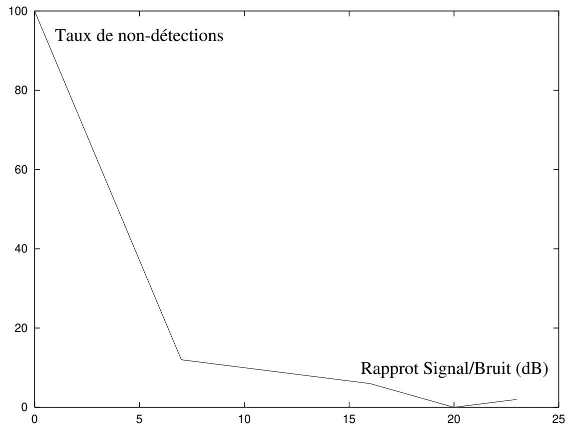

FIG. 6.10: *Taux de non détection en fonction du niveau de bruit : proportion des événements présents dans le signal qui n'ont pas été détectés.*

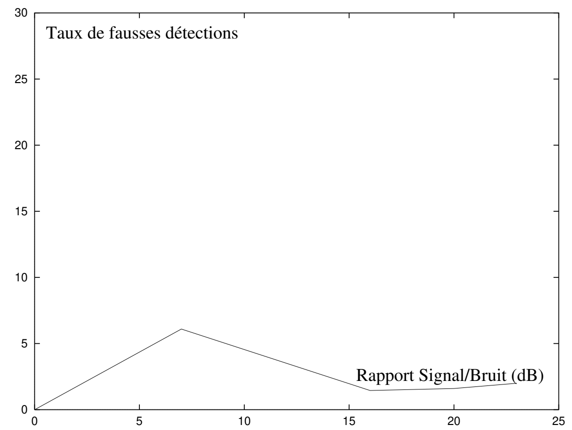

FIG. 6.11: *Taux de fausses détections en fonction du niveau de bruit : proportion des événements détectés qui ne correspondent pas à des bruits impulsifs. On observe un taux nul pour 0 dB : aucun des événements n'a été détecté. Le taux de fausses détections reste en deçà des 5*

Les figures 6.11 et 6.10 illustrent les performances en classification dans les mêmes situations. Globalement, on observe un comportement similaire : disfonctionnement à 0 dB et augmentation très rapide des performances ensuite.

Ces résultats montrent les difficultés du modèle en face de bruits complètement destructurés, mais aussi une certaine sensibilité au seuil de vigilance : en augmentant la valeur de ce seuil, on aurait retardé l'acquisition des classes. Au risque de ne pas détecter une classe, ou de mal reconnaitre certains événements, on aurait sans doute obtenu un système moins sensible mais plus robuste au bruit blanc. Ce problème est donc à la fois un problème de sensibilité à un bruit déstructuré et un problème de réglage statique du seuil de vigilance. Ici, aucune fausse alarme n'a été produite.

6.3.4 Évolution des performances en fonction du nombre de classes

L'expérience précédente portait sur l'influence d'un bruit déstructuré sur le système. Nous évaluons maintenant l'influence du nombre de classes sur les performances.

En effet, en augmentant le nombre de classes, on augmente la possibilité d'ambiguïtés et la difficulté de la détection et de la discrimination des événements acoustiques. Par ailleurs, comme dans l'expérience précédente, on n'a pas contraint les signaux de tests à éviter les superpositions d'événements. La diversité des bruits augmente la diversité des mélanges et complique la tâche de séparation. De façon à évaluer seulement l'influence du nombre de classes, nous avons augmenté la durée des signaux de tests en proportion du nombre de classes. On a ainsi une densité d'événements par unité de temps qui reste constante, ce qui permet de comparer les résultats avec une probabilité de superposition constante. Les événements ont une durée comprise entre 100 et 300 ms. On a en moyenne 2 événements par seconde. 20 occurrences de chaque classe sont insérées dans un signal de test. Tous les signaux de ce test ont été bruités par un bruit blanc avec un rapport signal/bruit de 23 dB. Pour chaque classe, les tests ont été réalisés sur 50 signaux de durée variant de 23 à 93 secondes.

Quel que soit le nombre de classes, aucune fausse alarme n'a été émise ; par ailleurs, la proportion de ruptures correctement détectées reste quasiment maximale dans toutes les configurations. Bien que le nombre de classes augmente, la perte de sensibilité du système ne provoque pas une dégradation des performances sensible à ce niveau de couverture de l'espace de représentation par l'espace-système.

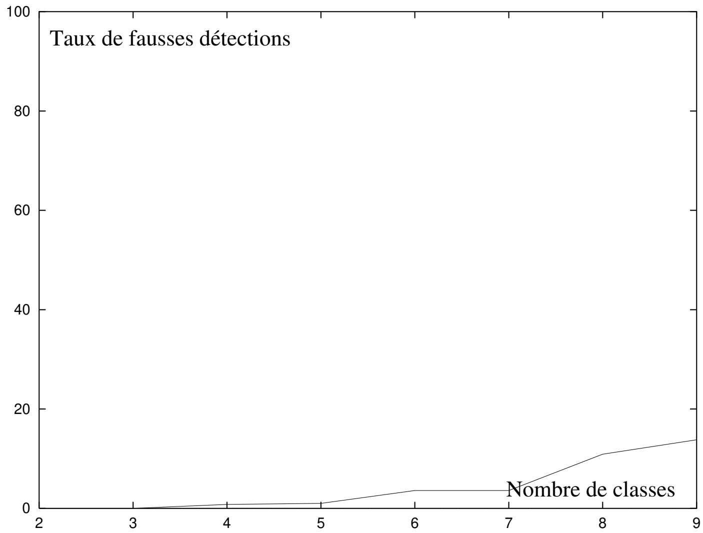

FIG. 6.12: *Taux de fausses détections en fonction du nombre de classes : on observe une augmentation du nombre des confusions avec l'augmentation du nombre de classes.*

Comme dans l'exemple précédent, aucune fausse alarme n'a été détectée.

Les performances en classification sont moins bonnes ; on observe une augmentation régulière du nombre de fausses détections et de celui de non détection, comme le montrent les courbes 6.12 et 6.13.

6.3.5 Influence de la densité d'événements acoustiques

Un point essentiel de la méthode proposée est sa capacité à détecter et classifier des événements acoustiques éventuellement simultanés. Pour évaluer le comportement du système dans des situations de fort recouvrement des formes acoustiques, nous avons augmenté, dans les signaux de tests, le nombre moyen de bruits impulsifs par unité de temps. Les signaux utilisés sont composés de 20 occurrences de deux classes d'événements acoustiques. Ces événements sont ajoutés aléatoirement dans un signal bruité par du bruit blanc à 23 dB. Les signaux ont des durées de 30, 20, 15, et 10 secondes, ce qui correspond à une densité d'événements qui varie de 4 à 1.3 événements par seconde. Le tableau 6.1 montre les résultats obtenus.

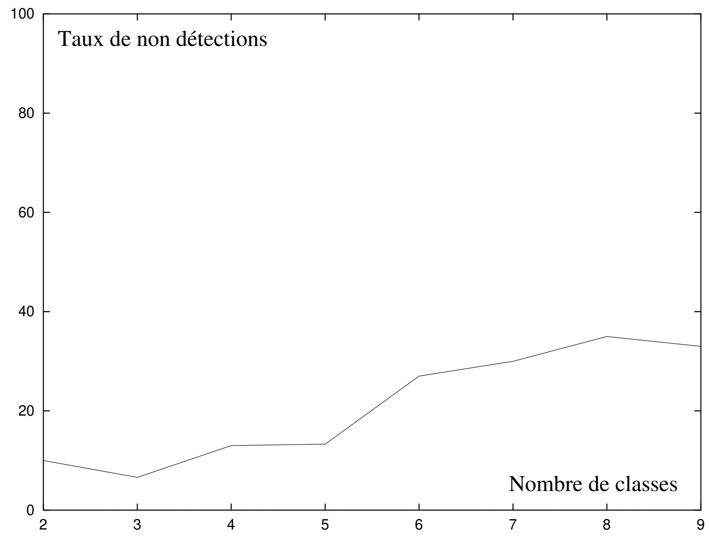

Fig. 6.13: *Taux de non détections en fonction du nombre de classes : on observe une perte de sensibilité avec l'augmentation de la couverture de l'espace de représentation par l'espace des classes.*

Événements / sec	4	2.66	2	1.33
Non-détections	0 %	0 %	2 %	6 %
Fausses détections	0.4 %	0.5 %	0.75 %	3.15 %
Bonnes détections	100 %	100 %	98 %	94 %
Fausses alarmes	0 %	0 %	0 %	0%
Bonnes alarmes	100 %	99 %	100 %	99 %

Tab. 6.1: Influence de la densité de bruits impulsifs sur les performances du système : on constate une légère amélioration des résultats avec l'augmentation de la probabilité de superposition des bruits impulsifs. Ceci peut être attribué à une faible activité moyenne des cellules qui provoque la suppression de certaines classes.

Contrairement à ce qu'on pourrait attendre, les résultats sont légèrement meilleurs pour un fort taux de recouvrement que pour les signaux où les bruits impulsifs sont moins fréquents. Ceci montre d'abord le bon comportement du système dans les situations de superposition de formes acoustiques.

La légère dégradation des résultats observée lorsque la densité de bruits impulsifs diminue peut être attribuée à l'augmentation de la probabilité de suppression d'une cellule. En effet, cette probabilité se trouve augmentée par la durée plus importante du signal, ainsi que par la diminution de l'activité moyenne d'une cellule. Ceci aurait put être limité par le choix d'une persistance plus importante de la mémoire du réseau.

6.3.6 Expériences en milieux réels

Plutôt que de chercher à établir une mesure de performance globale de l'algorithme sur un corpus exhaustif qui aurait été extrêmement difficile à constituer, nous avons cherché, pour les tests en conditions réelles, à selectionner un ensemble de signaux typiques des différentes situations qui peuvent être rencontrées. Le dépouillement de ces signaux a été fait par des experts qui ont proposé un étiquetage des bruits transitoires détectés dans le signal. La comparaison de l'analyse de l'expert et des résultats du détecteur de ruptures a ensuite été faite manuellement. Cette phase a mis en évidence une première difficulté liée à l'approche non-supervisée adoptée ici. L'absence d'étiquetage de sorties du réseau conduit l'opérateur à faire un effort d'interprétation qui nuit légérement à l'autonomie du système. Cet aspect, qui concerne essentiellement ses capacités de classification, induirait en phase d'exploitation un mode d'utilisation assez particulier : le système ne peut pas répondre à une requète explicite (du type "ce bruit est il une bruit de mamifère marin ?"), mais il peut permettre d'associer plusieurs événements et ainsi de suivre la piste d'un phénomène physique particulier.

Les premiers résultats obtenus sont prometteurs. Ils montrent cependant la nécessité de mettre en œuvre des pré-traitements destinés à réduire le niveau de bruit ambiant et la variabilité des classes de bruits impulsifs. Globalement, les difficultés rencontrées correspondent aux points que nos tests en laboratoire ont montrés : la capacité du système à rester opérationnel dans les situations de forte densité de bruits se paye, dans les zones moins "turbulentes", par des performances moyennes. Cette conclusion provisoire suggère l'utilisation de l'algorithme proposé ici en alternance ou en coopération avec d'autre méthodes correspondant à des approches très différentes.

6.3.7 Conclusion

La règle d'apprentissage proposée s'intègre au modèle décrit au chapitre 4. L'évaluation des performances de cette méthode confirme les avantages et les défauts qui avaient été préssentis : le système peut se trouver désorganisé par un bruit déstructuré d'un niveau suffisant. En deça du niveau critique, ces performances sont bonnes. Par ailleurs, l'augmentation du nombre de classes diminue la sensibilité du réseau et dégrade donc ses performances essentiellement en classification. La capacité de l'algorithme à détecter des ruptures du modèle n'a été mise en défaut que dans le cas d'un fort bruit blanc. Ce phénomène est atténué dans le cas d'un bruit aléatoire coloré ([Linares 97b]). Généralement, les performances sont plus fiables en détection de rupture qu'en classification.

Cependant, l'aspect essentiel qui a fortement orienté la conception du modèle est la capacité à modéliser et à détecter des classes d'événements superposés. Dans le cadre des expériences menées, le réseau s'est comporté correctement dans les situations de fort taux de recouvrement des formes acoustiques. Globalement, les expériences montrent que les performances en classification sont moins bonnes que celles en détection de ruptures, mais que le classifieur reste opérationnel dans les situations de forte densité de bruits impulsifs.

6.4 Conclusion

Dans ce chapitre, nous avons montré que le mélange des prototypes des classes conduisait au mélange des activités des cellules. Or la recherche d'un opérateur neuronal qui effectuerait le mélange inverse des activités est un problème de séparation de sources. Nous avons montré qu'un tel opérateur permet non seulement de séparer les activités, mais aussi de dissocier les prototypes des classes. Cette observation nous a conduit à implanter, dans l'architecture développée auparavant, une règle de réorganisation des classes fondée sur le principe auto-organisateur d'un réseau séparateur de sources.

Le principe d'une séparation de classes linéairement mélangées n'a pas été auparavant utilisé dans un système de classification automatique non-supervisée. L'opération effectuée dans ce cadre consiste à construire et à mettre à jour dynamiquement une base de l'espace de représentation qui met en évidence l'indépendance statistique des données. Ceci correspond à la recherche des vecteurs propres de la matrice des dépendances statistiques. Le problème de la recherche de la base indépendante a été abordé par ailleurs, dans des perspectives très différentes des nôtres ; en particulier, [Oja 93] propose un algorithme d'estimation en ligne de ces vecteurs.

Les performances du modèle ainsi constitué sont évaluées en fonction de divers paramètres. Les expériences réalisées montrent les bonnes performances du modèle en détection de rupture, ainsi que sa capacité à détecter et à classifier des événements dans des situations de très fort taux de recouvrement des formes acoustiques. Ces expériences ont aussi mis en évidence les défauts du modèle, en particulier la perte de sensibilité due soit à un fort bruit blanc, soit à un nombre élevé de classes référencées.

Chapitre 7

Conclusion et perspectives

7.1 Conclusion

Nous nous sommes interessés au problème de la supervision d'environnements réels. La problématique a été décrite d'abord de façon générale, puis en terme de détection des ruptures dans un signal composite. Nous avons montré que la difficulté de la tâche tenait essentiellement à la diversité et à l'imprévisibilité du signal engendré par les phénomènes physiques du monde réel. Cette observation nous a conduits à orienter notre étude vers des méthodes dynamiques, qui privilégient la plasticité d'un modèle plutôt que ses performances dans des conditions de laboratoire qui ne couvrent pas la diversité des situations susceptibles d'être rencontrées.

L'étude bibliographique des méthodes de détection de ruptures montre que si ces algorithmes sont adaptés à la surveillance de signaux fortement réguliers, leur utilisation n'est pas envisageable pour la surveillance de milieux intrinsèquement transitoires. Le problème initial a donc dû être abordé avec une approche de plus haut niveau, qui relève de l'analyse de scènes auditives plutôt que de la détection des changements brutaux d'un signal aléatoire. Cette approche suggère la modélisation en ligne de l'environnement par un système adaptatif capable de distinguer les événements "anormaux" par leur inadéquation au modèle appris.

La modélisation de l'environnement peut être réalisée par un classifieur capable de regrouper automatiquement les événements acoustiques en classes homogènes. Nous avons étudié les différentes méthodes de classification non supervisées qui ont été développées ces dernières années. Cette étude insiste sur les aspects qui sont essentiels dans le cadre de la problématique abordée ici. En particulier, nous avons mis l'accent sur l'incapacité des méthodes décrites à gérer des situations de superposition de formes acoustiques, situations probables dans des environnements complexes. Néanmoins, différents aspects de ces méthodes nous ont permis d'envisager une solution originale à ce problème.

Nous avons développé un modèle neuromimétique capable de modéliser l'environnement en regroupant en classes ses différentes composantes. Il s'agit d'un réseau qui s'adapte dynamiquement au signal par une auto-modification de sa structure. Ce processus d'apprentissage consiste à détecter les événements qui ne correspondent pas au modèle et à adapter le modèle en intégrant ces événements nouveaux à un ensemble des événements référencés. Un mécanisme d'élimination des classes complète ce processus d'adaptation en ligne.

L'originalité du modèle proposé réside à la fois dans sa plasticité et dans le principe d'une classification non-compétitive :

- plasticité : le système se construit dynamiquement suivant deux processus mutuellements dépendants, de détection des ruptures du modèle et d'adaptation structurelle du classifieur
- classification non compétitive : plutôt que de chercher à accentuer artificiellement le contraste entre les cellules, le modèle préserve la structuration des données en espace vectoriel. Ceci permet la détection des formes acoustiques superposées.

Cette capacité du réseau projectif à reconnaître des sons simultanés est mise en défaut lorsqu'une rupture du modèle survient simultanément à d'autres événements acoustiques. Une telle situation provoque le mélange des formes de référence. Nous avons montré que le problème de la dissociation des références peut être formulé comme un problème de séparation aveugle de sources. Ce résultat nous a amené à étudier les développements récents dans ce domaine. Reformuler le problème de classification d'événements superposés en termes de séparation de classes nous a permis d'implanter, dans le modèle présenté, une règle de modification des poids synaptiques basée sur une méthode de SAS. Ce processus permet d'extraire du signal composite des vecteurs caractérisant une classe d'événements, alors qu'ils n'ont été rencontrés que sous forme de mélanges d'images spectrales. La seule hypothèse adoptée est celle de l'indépendance statistique des classes à dissocier.

Les expériences réalisées sur un corpus de signaux réels artificiellement mélangés confirment la validité théorique de notre approche. Les performances montrent un comportement satisfaisant dans le cadre spécifique de notre procédure d'évaluation.

Plusieurs remarques doivent cependant être faites au sujet des limitations du modèle présenté.

D'une part, la dimension temporelle des formes acoustiques est négligée. Pourtant, la structure temporelle des sons participe à leur bonne perception et porte une information qui pourrait se réveler essentielle. Ceci n'est pas un handicap majeur pour la détection de bruit impulsifs, mais limite le pouvoir de représentation du modèle ainsi que son champ d'application.

Bien qu'on ait cherché à limiter les hypothèses sur les propriétés du signal traité, la méthode proposée n'est pas fondée uniquement sur des "universels", mais doit être paramétrée par des coefficients dont les valeurs influenceront son comportement : seuil de vigilance, persistance de la mémoire du réseau, pas de la modification de l'opérateur de décorrélation, etc.

Le critère de rupture repose sur le principe d'une régularité à moyen terme de l'espace-signal : la rupture a lieu lorsqu'une dimension nouvelle est visitée par le signal. Cela signifie que tout événement se décomposant dans l'espace des prototypes ne suscite pas d'alarme. Ceci comporte un premier inconvénient, qui réside dans la limitation du nombre de classes par la dimension de l'espace de représentation. Cependant, ceci ne doit pas devenir problématique : si tout l'espace est couvert, il n'est pas nécessaire d'ajouter une classe puisqu'il n'y a pas d'événement nouveau au sens du critère retenu. De plus, nous supposons que le mélange des classes ne porte pas d'informations concernant la présence d'un événement inattendu ou improbable dans le contexte modélisé.

Enfin, la séparation des classes est linéaire. Cette caractéristique est nécessaire à la dissociation des classes mélangées et à leur détection, mais elle limite la capacité de représentation du modèle. Néanmoins, cet inconvénient doit être relativisé par le fait qu'on cherche à extraire une organisation naturelle des classes et non pas à forcer une séparation qu'on ne saurait pas mettre en évidence par le choix d'une distance adaptée. Par ailleurs, la notion même d'axe séparateur est en contradiction avec le principe d'un étiquetage multiple d'une donnée : l'activité simultanée de plusieurs cellules est interprétée comme significative de la superposition d'événements des classes correspondantes. La frontière des classes n'est pas une limite qui permet de décider de l'appartenance d'un élément à une classe. Par ailleurs, l'organisation des prototypes est faite de façon à regrouper les composantes statistiquement liées, et à positionner les classes dans un état minimisant leur dépendance statistique.

Cette observation nous incite à re-situer la méthode de classification proposée dans la problématique générale des méthodes de *clustering* : d'ordinaire, ces algorithmes sont caractérisés par une métrique et éventuellement une mesure de complexité de la représentation souvent liée au nombre de classes. Ici, le nombre de classes est déterminé par la dynamique d'apprentissage du réseau projectif; elle dépend du processus de création et de suppression des classes. La définition des distances utilisée est plus complexe; il faut en distinguer deux types : la distance des classes entre elles, et celle des éléments aux classes. La distance inter-classe est une mesure approchée de leur dépendance statistique. La distance d'un prototype à une classe est, plus classiquement, une mesure angulaire normalisée par les distances entre prototypes. L'algorithme d'optimisation utilisé cherche à maximiser la distance inter-classes, au sens d'une mesure de dépendance statistique.

Du point de vue de la séparation de sources acoustiques, on est passé de la dissociation de signaux mono-dimensionnels mélangés à celle d'images temps-fréquence superposées. À la place de l'onde acoustique elle-même, ce sont les variations des niveaux d'énergie auquel les formes spectrales sont perçues qui constituent le signal analysé.

7.2 Perspectives

Le travail présenté dans ce mémoire ouvre de nombreuses perspectives dans l'amélioration du modèle proposé, dans son extension à des types de données plus variés, ou encore dans son application à des domaines différents.

7.2.1 Améliorations du modèle

Une première amélioration du système pourrait être la détermination des seuils de vigilance et de suppression des classes en fonction d'indices de contrôle du classifieur; par exemple, à partir d'une mesure de stabilité du réseau (création / suppression de classes par unité de temps), ou à partir de règles sur la dispersion des données d'entrée, d'un niveau de bruit moyen, etc. On peut imaginer de nombreux processus qui permettraient le "pilotage" du modèle par un système de règles utilisant une description globale du signal.

Cependant, la principale amélioration envisagée concerne l'intégration de la dimension temporelle. Notre algorithme échantillonne le signal en images spectrales de durée fixe. Ce sont les données issues de cette paramétrisation qui sont ensuite classifiées. Un événement acoustique de durée plus importante et dont l'image spectrale évolue dans le temps sera modélisé par plusieurs classes. Ces classes seront actives successivement, chacune représentant un moment particulier du son. La figure 7.1 illustre ce "découpage" d'une forme acoustique structurée en plusieurs images spectrales.

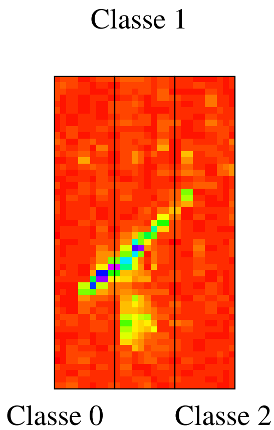

FIG. 7.1: *Segmentation d'une forme acoustique temporellement structurée : le son est modélisé par trois classes centisecondes, qui correspondent à différents moments de l'événement acoustique.*

Si cet événement est récurrent, l'enchaînement de ces classes se répétera, et des séquences récurrentes d'activité des classes apparaitront. *A priori*, les différentes classes qui participent à la modélisation d'une même forme acoustique sont inconnues. Cependant, l'étude des corrélations temporelles des activités des classes pourrait faire apparaître des séquences récurrentes et permettre de reconstituer les unités acoustiques à partir des classes élémentaires. Les classes centisecondes seraient alors les données élémentaires du système de regroupement des classes en méta-classes modélisant des formes acoustiques temporellement structurées et de durée variable. Bien entendu, une telle extension de la méthode pose des problèmes importants : comment associer une classe à deux méta-classes si elle correspond à une des similarités momentanées des images spectrales ? Quel critère permet d'extraire une séquence caractéristique, quelle quantité de données est nécessaire à l'estimation de paramètres statistiques qui pourraient mesurer la vraisemblance d'une séquence, etc. L'utilisation d'un algorithme de ce type pourrait permettre l'application de la méthode à la modélisation d'environnements acoustiques divers, ou encore à la séparation de sources de signaux récurrents avec un seul capteur.

7.2.2 Application à d'autres problèmes

Le modèle neuro-mimétique qui a été developpé ici pour la modélisation de bruits impulsifs pourrait être appliqué à d'autres types de données. Ceci est d'autant plus vrai que la structure temporelle des sons est perdue, alors que c'est une caractéristique majeure des problèmes du traitement des formes acoustiques. Les contraintes qui doivent être respectées sont essentiellement la structure vectorielle des données. De plus, le principe de séparation repose sur le mélange linéaire des données. Ce point limite l'usage de ce type de méthode pour la séparation d'images : les images superposées comportent des parties masquées ; il ne s'agit donc pas d'un mélange additif. Par contre, on peut imaginer son application au traitement du langage écrit, où les documents peuvent être représentés par des vecteurs dont chaque coefficient correspond au nombre d'occurrences d'un mot. Cet algorithme pourrait permettre la détection de ruptures de thèmes ou le regroupement de mots statistiquement liés.

Par ailleurs, indépendamment de l'utilisation du modèle en détection de ruptures ou en classification, la modélisation dynamique du signal peut être exploitée comme un pré-traitement susceptible d'extraire automatiquement un sous-espace dans lequel les données sont bien représentées. Le réseau jouerait alors le rôle d'un quantificateur vectoriel dynamique.

Bibliographie

[Anderson 84] T. W. Anderson. An introduction to multivariate statistical analysis. New York, second edition, 1984.

[Attneave 54] F. Attneave. *Informational aspects of visual perception.* Psychological Review, vol. 61, pages 183–193, 1954.

[Barflow 89] H.B. Barflow, T.P. Kaushal & G.J. Mitchison. *Finding minimum entropy codes.* Neural Computation, vol. 1, pages 412–423, 1989.

[Basseville 89] M. Basseville. *Distance measures for signal processing and pattern recognition.* Signal Processing, vol. 18, pages 349–369, 1989.

[Basseville 93] M. Basseville & I.V. Nikiforov. Detection of abrupt changes : Theory and application. Englewood Cliffs, New Jersey 07632, 1993.

[Bell 95] A.J. Bell & T.J. Sejnowski. *An information maximisation approach to blind separation and blind deconvolution.* Neural Computation, vol. 7, no. 6, pages 1129–1159, 1995.

[Brandt 83a] A.V. Brandt. *Detecting and estimating parameter jumps using ladder algorithms and likehood ratio tests.* In Proc. of ICASSP'83, pages 1017–1020, 1983.

[Brandt 83b] A.V. Brandt. *An efficient GLR algorithm for adaptative segmentation of non stationary signals.* In Proc. of EUSIPCO, pages 763–766, Erlangen, RFA, Septembre 1983.

[Buhmann 92a] J. Buhmann & H. Kuhnel. *Complexity Optimized Vector Quantification : A Neural Network Approach.* In Data Compression Conference, Snowbird, March 1992.

[Buhmann 92b] J. Buhmann & H. Kuhnel. *UNsupervised and Supervised Data Clustering with Competitive Neural Networks.* In IJCNN'92, Baltimore, 1992. IEEE press.

[Cardoso 89] J.F. Cardoso. *Source separation using higher order moments.* In Proceedings of ICASSP'89, vol. 4, pages 2109–2112, Glasgow, Scotland, May 1989. IEEE.

[Cardoso 94] J.F. Cardoso & B. Laheld. *Adaptative source separation with uniform performance.* In EUSIPCO, Edinburgh, 1994.

[Carpenter 76] G.A. Carpenter & S. Grossberg. *Adaptative Patern Classification and Universal Recoding.* Biological Cybernetics, vol. 23, pages 187–202, 1976.

[Carpentier 92] E. Le Carpentier, C. Doncarli & B. Vozel. *Optimal bayesian detection of changes in random signals.* In proc. of EUSIPCO 92, Bruxelles, 1992.

[Castanié 83] F. Castanié & P. Soule. *Méthode de détection de non stationnarités par analyse des propriétés de franchissement d'un seuil par l'erreur de prédiction linéaire.* In Actes du 9ème colloque GRETSI, pages 105–109, Nice, 1983.

[Cichocki 92] A. Cichocki & L. Mosczynski. *New learning algorithms for blind separation of sources.* Electronic letters, vol. 28, pages 1986–1987, 1992.

[Cichocki 94] A. Cichocki & R. Unbehaunen. *Robust neural networks with on-line learning for blind indentification and blind separation of sources.* In IEEE Transations on Circuits and Systems, volume 28, June 1994.

[Comon 89] P. Comon. *Séparation of Stochastic Process whose a linear Mixture is observed.* In ONR-NSF-IEEE Workshop on Higher-Order Spectral Analysis, volume 7, pages pp. 174–179, Vail (colorado), June 1989.

[Comon 90] P. Comon. *Analyse en composantes indépendantes et identification aveugle.* Traitement du signal, vol. 7, no. 5, pages pp. 435–450, 1990.

[Comon 91a] P. Comon. *Independent Component Analysis.* In International Workshop On Higher Order Statistics, volume 7, pages pp. 111–120, Chanrousse, France, 1991.

[Comon 91b] P. Comon, C. Jutten & J. Herault. *Blind separation of sources, part II : problems statement.* Signal Processing, vol. 24, pages 11–21, 1991.

[Comon 94] P. Comon. *Independent component analysis - a new concept ?* Signal Processing, vol. 36, pages 287–314, 1994.

[Daymier 85] E. Daymier & F. Castanié. *Analyse d'une méthode de détection des sauts de moyenne de variance.* In Actes du 10ème colloque GRETSI- TDSI, Nice, Mai 1985.

[Deco 95] G. Deco & W. Brauer. *Non-linear highter-order statistical decorrelation by volume-conserving neural-architectures.* Neural-Networks, 1995.

[Diday 71] E. Diday. *La méthode des nuées dynamiques.* Revue de statistique appliquée, vol. 2, pages 19–34, 1971.

[Doncarli 83] C. Doncarli. *H(X—Y)Un algorithme séquentiel de détection de ruptures d'une modèle ARMA.* In Actes du 9ème colloque GRETSI, pages 1035–1039, Nice, 1983.

[Féty 88] F. Féty. *Méthodes de traitement d'antenne adaptées aux radiocommunications.* PhD thesis, ENST, Paris, 1988.

[Fort 91] J.C. Fort. *Stabilité de l'algorithme de séparation de sources de Jutten et Hérault.* Traitement du Signal, vol. 8, no. 1, pages 35–42, 1991.

[Gaeta 90] M. Gaeta & J.L. Lacoume. *Source separation without a priori knowledge : the ML solution.* In proc of EUSIPCO, pages 621–624, 1990.

[Gersho 92] A. Gersho & R. M. Gray. Vector quantization and signal compression. Kluwer Academic Publishers, Boston, 1992.

[Grossberg 76] S. Grossberg. *Adaptative Pattern Classification and Universal Recoding. I. Parallel Development and Coding of Neural Feature Detectors.* Biological Cybernetics, no. 23, pages pp 187–202, 1976.

[Haykin 96] S. Haykin. Adaptative filter theory, third ed. New Jersey, prentice hall edition, 1996.

[Hebb 49] D. O. Hebb. The organization of behavior. New York, 1949.

[Hérault 85] J. Hérault, C. Jutten & B. Ans. *Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non-supervisé.* volume 24, pages 1017–1020, Nice, France, 1985. GRETSI.

[Hopfield 82] J. J. Hopfield. *Neural networks and physical systems with emergent collective computational facilities.* vol. 79, pages 2554–2558, 1982.

[Jardine 66] N. Jardine & R. Sibson. Mathematical taxonomy. London, 1966.

[Jutten 87] C. Jutten. *Calcul Neuromimétique et Traitment du Signal. Analyse en Composantes Indépendantes, Thèse d'état.* PhD thesis, INPG-USMG, Grenoble, 1987.

[Karhunen 94] J. Karhunen & J. Joutsensalo. *Representation and separation of signals using nonlinear PCA type learning.* Neural Networks, vol. 7, no. 1, pages 113–127, 1994.

[Karhunen 95a] J. Karhunen, E. Ojja, L. Wang, R. Vigario & J. Joutsensalo. *A class of neural networks for independant component analysis.* Technical REport A-28, October 1995.

[Karhunen 95b] J. Karhunen, L. Wang & R. Vigario. *Non-linear PCA type approaches for source separation and independant component analysis.* pages 995–1000, Perth, Australia, November 1995. ICNN.

[Karhunen 96] J. Karhunen. *Neural Approaches yo Independent Component Analysis and Source Separation.* In Eurpean Symposium on Artificial Neural Networks, 1996.

[Kendall 66] M.G. Kendall. In multivariate analysis. Academic Press, London, 1966.

[Kohonen 89] T. Kohonen. Self organisation and associative memory. Springer Series in Information Sciences, third edition, 1989.

[Laughlin 81] S. Laughlin. *A simple coding procedure enhances a neuron's information capacity.* Naturforsh, vol. 36, pages 910–912, 1981.

[Linares 97a] G. Linares, P. Nocera & H. Meloni. *Mixed acoustic events classification using Subspace Classifier and ICA.* In Proc. 1997 IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, Munich, Germany, April 1997.

[Linares 97b] G. Linares, P. Nocera & H. Meloni. *Model Breaking Detection using Independant Component Classifier.* In Proc. 1997 ENNS International Conference on Artificial Neural Networks, Lausanne, Switzerland, September 1997.

[Linsker 89] R. Linsker. *An application of the principle of maximum information preservation to linear systems.* In Touretzky D.S., editeur, Advances in Neural Information Processing Systems, volume 4. Morgan-Kauffman, 1989.

[Linsker 92] R. Linsker. *Local synaptic learning rules suffie to maximise mutual information in a linear network.* Neural Computation, vol. 4, pages 691–702, 1992.

[Ljung 83] L. Ljung & T. Sodertrom. Theory and practice of recursive identification. MIT Press, Cambridge, Massachusetts, 1983.

[Lloyd 82] S. P. Lloyd. *Least squares quantization in PCM.* vol. 28, pages 128–137, 1982.

[Malvar 90] H. Malvar. *Lapped transforms for efficient transform/subband coding.* IEEE Transactions on Acoustic, Speech and Signal Processing, pages 969–978, 1990.

[Matsuoka 95] K. Matsuoka, M. Ohya & M. Kawamoto. *Neural Net for Blind Separation of Nonstationary Signals.* Neural Networks, vol. 8, no. 3, pages 441–419, 1995.

[Nadal 94] J.P. Nadal & N. Parga. *Non-linear neurons in the low noise limit : a factorial code maximizes information transfert.* Networks, vol. 5, pages 565–581, 1994.

[Nadal 97] J.P. Nadal. *Redundancy Reduction and Independant Component Analysis : Conditions on Cumulants and Adaptatives Approaches.* Neural Computation, vol. to appear, 1997.

[Nguyen 92] T. Nguyen. *Séparation aveugle de sources à bande large dans un mélange convolutif. Application au rehaussement de la parole.* PhD thesis, Institut National Polytechnique de Grenoble, Grenoble, 1992.

[Nguyen 95] H.L. Nguyen & C. Jutten. *Blind source separation for convolutives mixtures.* Signal Processing, vol. 45, pages 209–229, 1995.

[Nikiforov 86] I.V. Nikiforov. Sequential detection of changes in stochastic systems. detection of abrupt changes in signals and dynamical systems, chapitre 7. Lectures notes in Control and Information Sciences. 1986.

[Oja 85] E. Oja & J. Karhunen. *On Stochastic Approximation of the Eigenvectors and Eigenvalues of the Expectation of a Random Matrix.* J. Math. Analysis and Appl., vol. 106, pages 69–84, 1985.

[Oja 92] E. Oja. *Principal Components, minor components, and linear neural networks.* Neural Networks, vol. 5, pages 927–935, 1992.

[Oja 93] E. Oja & J. Karhunen. *Nonlinear PCA : Algorithms and Applications.* Rapport technique A18, Helsinki University of Technology Laboratory of Computer and Information Sciences, Rakentajanaukio 2 C, SF-02150 Espoo, Finland, September 1993.

[Parashiv 97] A. Parashiv, C. Jutten & G. Bouvier. *Neural Network Based Processing for Smart Sensor Arrays.* In Proc. of ICANN'97, pages 565–569, Lausanne, Switzerland, 1997. ENNS.

[Parra 95] L. Parra, G. Deco & S. Miesbach. *Statistical independence and novelty detection with information preserving non-linear maps.* Neural processing letters, 1995.

[Peretto 84] P. Peretto. *Collective Properties of Neural Networks : a Statistical Approach.* Biological Cybernetics, vol. 50, pages pp 51–62, 1984.

[Pican 95] N. Pican. *Approches Statique et Dynamique de la Modulation des Efficacités Synaptiques dans les Réseaux de Neurones.* PhD thesis, Université Henri Poincaré, Nancy, 1995.

[Platt 91] J. Platt & F. Faggin. Networks for the separation of sources that superimposed and delayed,, volume 4 of *Advances in Neural Information Processing Systems*, pages 730–737. Morgan Kauffman, San Mateo, California, 1991.

[Quelavoine 97] R. Quelavoine. *Étude de l'apprentissage et des structures des réseaux de neurones multicouches pour l'analyse de données.* PhD thesis, Université d'Avignon et des Pays de Vaucluse, 1997.

[Redlich 93] A.N. Redlich. *Redundancy Reduction as a strategy for unsupervised learning.* Neural Computation, vol. 5, pages 289–304, 1993.

[Rosenblatt 62] F. Rosenblatt. Principles of neurodynamics. Spartan Books, Washington, DC, 1962.

[Schuster 92] H.G. Schuster. *Learning by maximizing the information transfert thrpough nonlinear noisy neurons and "noise breakdown"*. Physical Review, vol. A, no. 46, pages 2131–2138, 1992.

[Sorouchyari 91] E. Sorouchyari. *Blind Separation of Sources, Part III : stability Analysis*. Signal Processing, vol. 24, no. 1, pages 11–20, 1991.

[Taleb 97] A. Taleb & C. Jutten. *Entropy Optimisation*. In Proc. of ICANN'97, pages 529–533, Lausanne, Switzerland, 1997. ENNS.

[Therrien 92] C. Therrien. Discrete random signals and statistical signal processing. Englewood Cliffs, 1992.

[Theys 93] C. Theys, G. Alengrin & J. Ménez. *Détection d'un saut de phase ou d'amplitude dans une sinusoïde bruitée*. In Actes du 14ème colloque GRETSI, Juans les pins, septembre 1993.

[Wang 95] L. Wang, J. Karhunen & E. Oja. *A bigradient optimization approach for robust PCA, MCA, and source separation*. In IEEE International Conference on Neural Networks, Perth, Austrlia, November 1995.

[Wesfreid 92] E. Wesfreid & M.V. Wickerhauser. *Etude des signaux vocaux par ondelettes de Malvar*. In Reconnaissance des Formes et Intelligence Artificielle, 1992.

[Widrow 60] B. Widrow & M. E. Hoff Jr. *Adaptive switching circuits*. IRE WESCON Convention Record, vol. 4, pages 96–104, 1960.

[Zhang 92] Zhang, Basseville & Beneviste. *Signaux non stationnaires. Analyse temps-fréquence et segmentation*. Traitement du signal., vol. 9, 1992.

Table des figures

Table des matières

Annexe

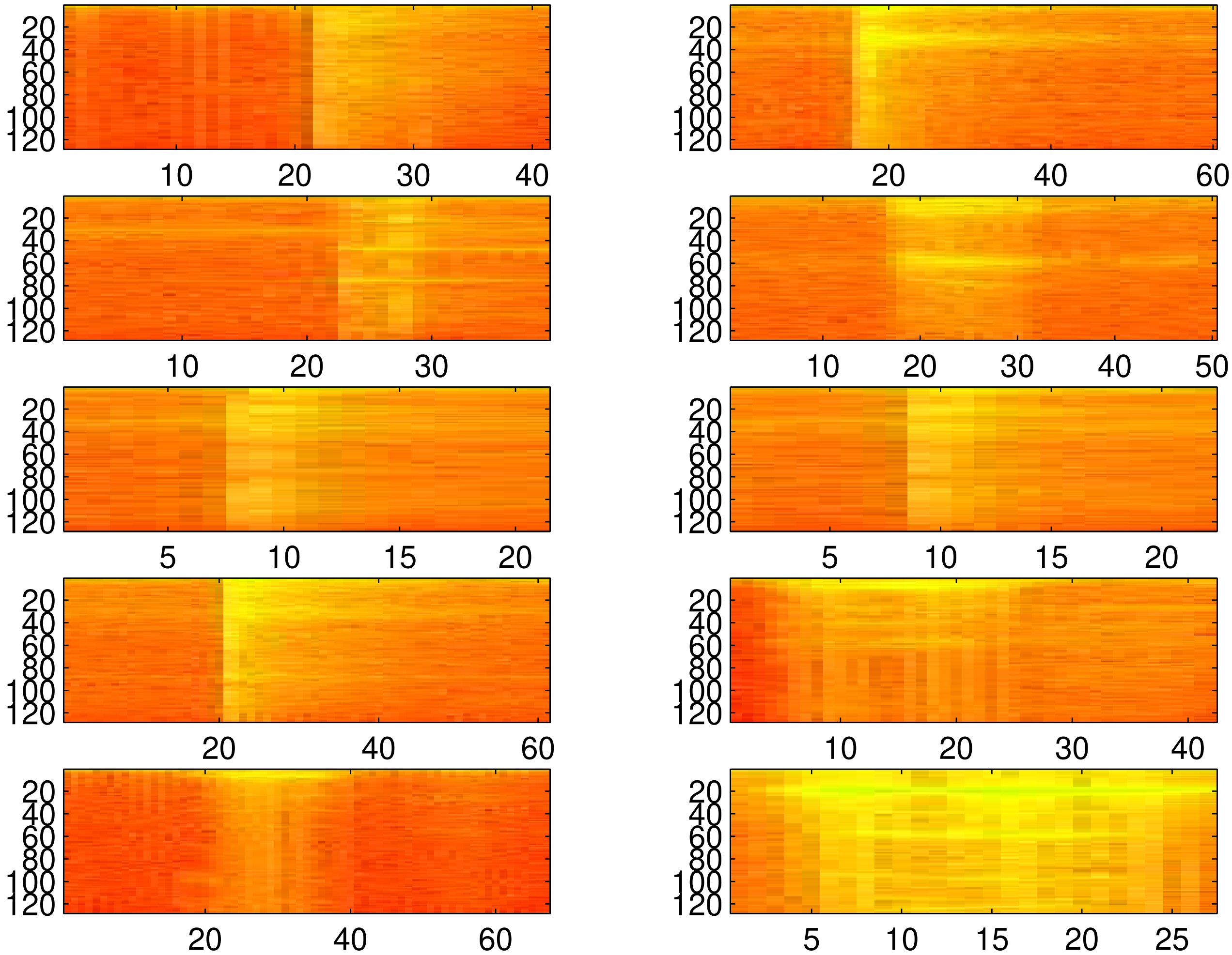

FIG. 7.2: *Image spectrale des bruits impulsifs utilisés pour l'évaluation des performances en détection de ruptures et en classification.*