

# *Projet SIGMA2*

*Signaux, modèles et algorithmes*

*Rennes*

THÈME 4A



*R*apport  
*d'Activité*

2000



## Table des matières

<b>1</b>	<b>Composition de l'équipe</b>	<b>3</b>
<b>2</b>	<b>Présentation et objectifs généraux</b>	<b>5</b>
<b>3</b>	<b>Fondements scientifiques</b>	<b>6</b>
3.1	Identification et approximation stochastique . . . . .	6
3.2	Approche asymptotique locale pour la surveillance et le diagnostic des systèmes continus . . . . .	9
3.3	Modèles partiellement stochastiques pour la surveillance des systèmes distribués	13
<b>4</b>	<b>Domaines d'applications</b>	<b>15</b>
4.1	Analyse de structures vibrantes en ambiance de travail . . . . .	16
4.2	Télécommunications : diagnostic de pannes en gestion de réseaux, turbo-codes, décodage conjoint source-canal . . . . .	17
<b>5</b>	<b>Logiciels</b>	<b>18</b>
5.1	Nonlinear Modelling Matlab Toolbox . . . . .	18
<b>6</b>	<b>Résultats nouveaux</b>	<b>19</b>
6.1	Surveillance et estimation de systèmes dynamiques non-linéaires . . . . .	19
6.2	Méthodes de sous-espace, application à la surveillance de structures soumises à vibrations en opération . . . . .	20
6.3	Surveillance des systèmes distribués . . . . .	22
6.4	Codes correcteurs et modèles graphiques . . . . .	24
6.5	Modèles de Markov cachés généraux . . . . .	25
6.6	Synthèse de filtres numériques . . . . .	27
6.7	Traitement de la parole et du langage . . . . .	28
<b>7</b>	<b>Contrats industriels (nationaux, européens et internationaux)</b>	<b>32</b>
7.1	Analyse de structures vibrantes — Projet Eurêka SINOPSYS . . . . .	32
7.2	Diagnostic de pannes dans les réseaux de télécommunications — Projet exploratoire RNRT MAGDA . . . . .	33
7.3	Mesure de la qualité de service en réseau WDM optique transparent — Contrat Alcatel . . . . .	33
7.4	Identification du comportement dynamique d'un véhicule routier — Contrat Renault . . . . .	34
7.5	Indexation sonore — Projet pré-compétitif RNRT AGIR . . . . .	34
7.6	Gestion d'archives audio-visuelles — Projet Esprit DIVAN . . . . .	34
7.7	Vérification du locuteur — Projet Telematics PICASSO . . . . .	35
7.8	Reconnaissance du locuteur — Contrat Bull . . . . .	35
7.9	Compression de parole — Contrat Innovaction . . . . .	35

---

<b>8</b>	<b>Actions régionales, nationales et internationales</b>	<b>36</b>
8.1	Système générique de reconnaissance de parole — Action de recherche coopérative de l'INRIA SIROCCO . . . . .	36
8.2	Consortium ELISA . . . . .	36
8.3	Identification des systèmes — Réseau TMR SI . . . . .	36
8.4	Méthodes statistiques pour les systèmes dynamiques stochastiques — Réseau IHP DYNSTOCH . . . . .	36
8.5	Autres actions internationales . . . . .	37
8.6	Visites, et invitations de chercheurs . . . . .	37
<b>9</b>	<b>Diffusion de résultats</b>	<b>37</b>
9.1	Animation de la communauté scientifique . . . . .	37
9.2	Enseignement . . . . .	38
9.3	Participation à des colloques, séminaires, cours, etc. . . . .	39
<b>10</b>	<b>Bibliographie</b>	<b>39</b>

# 1 Composition de l'équipe

## Responsable scientifique

François Le Gland [DR INRIA]

## Assistante de projet

Marie-Noëlle Georgeault [TR INRIA]

## Personnel INRIA

Albert Benveniste [DR, à temps partiel]

Frédéric Cérou [CR]

Éric Fabre [CR]

Rémi Gribonval [CR, à partir du 1<sup>er</sup> décembre 2000]

Laurent Mevel [CR]

Qinghua Zhang [CR]

## Personnel CNRS

Michèle Basseville [DR]

Frédéric Bimbot [CR]

## Personnel université de Rennes 1

Jean-Jacques Fuchs [professeur]

Annie Morin [maître de conférences]

## Collaborateur extérieur

Bernard Delyon [professeur à l'université de Rennes 1]

## Ingénieur expert INRIA

Bruno Jacob [jusqu'au 31 juillet 2000]

**Ingénieur associé**

Vincent Pigourier [depuis le 1<sup>er</sup> octobre 2000]

**Chercheurs doctorants**

Samy Abbes [bourse MENRT]

Laurent Benaroya [bourse MENRT]

Raphaël Blouet [bourse INRIA]

Arnaud Clavel [bourse Cifre Renault]

Arnaud Guyader [bourse MENRT]

Lorcan McDonagh [bourse INRIA, depuis le 1<sup>er</sup> novembre 2000]

Nadia Oudjane [bourse MENRT/Onera, jusqu'au 30 septembre 2000]

Olivier Perrin [bourse Cifre Renault, depuis le 1<sup>er</sup> décembre 2000]

Rodolphe Priam [bourse MENRT]

Mouhamadou Seck [bourse INRIA, jusqu'au 30 septembre 2000]

Benjamin Torno [bourse MENRT, jusqu'au 29 février 2000]

Aiping Xu [bourse INRIA]

**Chercheurs post-doctorants INRIA**

Guillaume Gravier [du 1<sup>er</sup> février 2000 au 31 janvier 2001]

Ivan Magrin-Chagnolleau [jusqu'au 15 octobre 2000]

Laurie Ricker [jusqu'au 31 mai 2000, en commun avec le projet PAMPA]

Mark Smith [jusqu'au 15 août 2000, en commun avec le projet PAMPA]

**Chercheur post-doctorant université de Rennes 1**

Bo Wang [depuis le 1<sup>er</sup> septembre 2000]

**Chercheurs invités**

George Moustakides [université de Patras, du 1<sup>er</sup> septembre 2000 au 28 février 2001]

Nigel Newton [université d'Essex, du 1<sup>er</sup> mai au 15 août 2000]

## 2 Présentation et objectifs généraux

**Mots clés :** identification, surveillance, diagnostic, méthode de sous-espace, approche locale, automobile, vibration, analyse modale, système distribué, système à événements discrets, HMM, filtrage particulaire, réseau de Petri, modèle partiellement stochastique, réseau de télécommunications, gestion d'alarmes, turbo-code, décodage conjoint source-canal, traitement de la parole, analyse de données textuelles.

Les objectifs du projet SIGMA2 sont la conception, l'analyse et la mise en œuvre d'algorithmes statistiques basés sur l'utilisation de modèles, pour l'identification, la surveillance et le diagnostic de systèmes industriels complexes. Les modèles considérés sont d'une part les modèles d'état de l'automatique stochastique, avec une importance croissante des modèles non-linéaires dans les applications industrielles, et d'autre part des modèles partiellement stochastiques (HMM, réseaux de Petri, réseaux d'automates, etc.) sur des structures discrètes (arbres, graphes, etc.), par exemple pour modéliser les systèmes distribués à événements discrets. Les contributions méthodologiques les plus importantes du projet, et qui constituent les bases scientifiques des activités actuelles, concernent l'utilisation de l'approche asymptotique locale pour la surveillance et le diagnostic des systèmes continus, et la conception d'algorithmes répartis de reconstruction d'état, de type Viterbi, pour la surveillance et le diagnostic des systèmes distribués à événements discrets. Les principales applications considérées sont la surveillance et le diagnostic des structures mécaniques en vibration (automobile, aéronautique, génie civil), la surveillance et le diagnostic d'organes de véhicules automobiles, et le diagnostic de pannes dans les réseaux de télécommunications. Les activités en traitement de la parole et du signal audio seront prochainement menées dans un autre projet.

### Axes de recherche

- observateurs et filtres pour la surveillance et le diagnostic des systèmes dynamiques non-linéaires,
- méthodes de sous-espaces pour l'analyse et la surveillance modale,
- surveillance et diagnostic des systèmes distribués à événements discrets,
- algorithmes approchés d'estimation d'état dans les modèles graphiques et les réseaux bayésiens, par exemple pour le décodage,
- statistique des HMM généraux, et méthodes numériques associées,
- synthèse de filtres numériques,
- traitement de la parole et du signal audio,
- analyse de données textuelles.

## Relations internationales et industrielles

- conventions de recherche : avec Renault sur l'identification du comportement dynamique d'un véhicule routier, et sur la surveillance et le diagnostic d'organes de véhicules automobiles, avec Alcatel sur l'estimation transparente du taux d'erreur dans un réseau optique, avec Bull sur la reconnaissance du locuteur,
- projets multi-partenaires : au niveau national sur le diagnostic de pannes dans les réseaux de télécommunications (RNRT), sur l'indexation sonore de données multimédia (RNRT), et au niveau européen sur l'identification et la surveillance de structures vibrantes en ambiance de travail (Eurêka), sur l'indexation sonore dans la gestion d'archives audiovisuelles (Esprit), sur la vérification du locuteur à travers le réseau téléphonique (Telematics),
- réseaux de recherche académiques : au niveau national sur un système générique de reconnaissance de parole (ARC INRIA), et au niveau européen sur l'identification des systèmes (TMR), sur la dynamique des structures (COST), sur les méthodes statistiques pour les systèmes dynamiques stochastiques (IHP).

## 3 Fondements scientifiques

Le projet SIGMA2 s'intéresse aux techniques de modélisation, à partir de principes physiques, mais surtout à partir de données d'observation. Les problèmes centraux sont donc l'estimation et l'identification, mais aussi la validation de modèle, le test et le diagnostic, qui permettent de reconnaître et d'expliquer un désaccord entre modèle et mesures. Ces questions sont examinées sur différents types de modèles de systèmes dynamiques : linéaires, non-linéaires, et plus récemment, distribués à événements discrets. Nous avons choisi de détailler les trois points ci-après, où le projet a apporté des contributions importantes, et qui constituent les bases scientifiques et méthodologiques des activités actuelles.

### 3.1 Identification et approximation stochastique

**Mots clés** : identification, fonction d'estimation, système adaptatif, approximation stochastique, poursuite.

**Résumé** : *On se donne une suite d'observations  $(Z_0, \dots, Z_n, \dots)$  de loi paramétrée par  $\theta \in \Theta$ . Le problème de l'identification paramétrique consiste à retrouver la vraie valeur  $\theta^*$  du paramètre, à partir de la donnée d'une suite d'observations  $(Z_0, \dots, Z_n, \dots)$  gouvernée par ce paramètre.*

*L'identification peut être «hors-ligne», ou «non réursive». Dans ce cas, on se donne un échantillon  $(Z_0, \dots, Z_N)$  de longueur finie  $N$ , et on cherche à construire un estimateur  $\hat{\theta}_N$ , c'est-à-dire une fonction de l'échantillon  $(Z_0, \dots, Z_N)$  à valeurs dans  $\Theta$ . Les questions qui se posent sont alors : lorsque  $N$  tend vers l'infini, 1)  $\hat{\theta}_N$  converge-t-il vers la «vraie» valeur  $\theta^*$  du paramètre ? 2) si oui, à quelle vitesse ?*



existe-t-il une borne optimale pour cette vitesse, borne qui dépend de la vraie valeur (inconnue)  $\theta^*$ , et qui permet de parler de l'optimalité d'un estimateur ? sait-on construire des estimateurs optimaux ?

L'identification peut être «en-ligne», ou «réursive». Dans ce cas, on lit à la volée l'échantillon  $(Z_0, \dots, Z_n, \dots)$ , et on construit récursivement une suite d'estimateurs  $\hat{\theta}_n$ . Cette suite est en général calculée incrémentalement, ce qui signifie que  $\hat{\theta}_{n+1}$  est calculé en fonction de  $\hat{\theta}_n$  et des nouvelles données reçues. On parle alors souvent d'approximation stochastique, ou de système adaptatif. Outre les questions précédentes, on se pose alors le problème de la «poursuite» : supposons que la «vraie» valeur  $\theta_n^*$ , au lieu d'être fixe, soit «lentement» variable, est-il possible que la suite d'estimateurs  $\hat{\theta}_n$  poursuive les variations de  $\theta_n^*$  ?

Nous avons, au cours des dix dernières années, apporté diverses contributions sur l'ensemble de ces questions, et tout récemment sur le problème de la poursuite.

**Fonctions d'estimation** On suppose donné un modèle paramétrique  $\{\mathbf{P}_\theta, \theta \in \Theta\}$  décrivant le système à identifier, où la loi des observations  $(Z_0, \dots, Z_N)$  est caractérisée par le paramètre  $\theta \in \Theta$ , et une «fonction d'estimation» de la forme

$$U_N(\theta) = \frac{1}{N} \sum_{n=0}^N H(\theta, Z_n),$$

telle que  $\mathbf{E}_\theta[U_N(\theta)] = 0$  pour tout  $\theta \in \Theta$ . Dans un grand nombre de situations,  $H$  est le gradient d'une fonction à minimiser (erreur quadratique de prédiction, log-vraisemblance (au signe près), etc.). Pour identifier le modèle au vu des observations  $(Z_0, \dots, Z_N)$ , un estimateur du paramètre inconnu sera alors [MS88,Hey97,DJB97]

$$\hat{\theta}_N = \arg\{\theta \in \Theta : U_N(\theta) = 0\}.$$

Si  $\theta^*$  est la «vraie» valeur du paramètre, et si  $\mathbf{E}_{\theta^*}[U_N(\theta)] = 0$  si et seulement si  $\theta = \theta^*$  (avec  $\theta^*$  fixé), alors  $\hat{\theta}_N$  converge vers  $\theta^*$ . Par le théorème-limite central, le vecteur  $U_N(\theta^*)$  est asymptotiquement gaussien centré, de matrice de covariance  $\Sigma$  que l'on sait calculer ou estimer. Si de plus la matrice  $M_N = -\mathbf{E}_{\theta^*}[U_N'(\theta^*)]$  est inversible, alors en utilisant un développement de Taylor et la contrainte  $U_N(\hat{\theta}_N) = 0$ , on obtient la normalité asymptotique de l'estimateur, i.e.  $\sqrt{N}(\hat{\theta}_N - \theta^*) \approx M_N^{-1} \sqrt{N} U_N(\theta^*)$ .

Dans de nombreuses applications, une approche de ce type est insuffisante, et doit être améliorée dans les directions suivantes :

- récursivité : possibilité de calculer  $\hat{\theta}_{N+1}$  simplement à partir de  $\hat{\theta}_N$  — ceci se fait au prix de certaines approximations dont le coût doit être d'un ordre de grandeur inférieur à l'erreur d'estimation  $1/\sqrt{N}$ ,

- 
- [MS88] D. L. McLEISH, C. G. SMALL, *The Theory and Application of Statistical Inference Functions, Lecture Notes in Statistics, 44*, Springer-Verlag, Berlin, 1988.
- [Hey97] C. C. HEYDE, *Quasi-Likelihood and its Applications, Springer Series in Statistics*, Springer-Verlag, Berlin, 1997.
- [DJB97] B. DELYON, A. JUDITSKY, A. BENVENISTE, « On the Relationship between Identification and Local Tests », *Publication Interne n° 1104*, IRISA, mai 1997, <ftp://ftp.irisa.fr/techreports/1997/PI-1104.ps.gz>.

- adaptativité : il arrive que le vrai paramètre  $\theta$  varie au cours du temps (par exemple le transfert d'un canal radio-moblie) — un bon algorithme doit être capable de «poursuivre» cette valeur.

La théorie de l'approximation stochastique permet d'aborder de manière adéquate les problèmes de récursivité et d'adaptativité en identification, dans un cadre statistique.

**Approximation stochastique** L'usage des algorithmes stochastiques est très répandu, et recouvre des domaines tels que la commande adaptative, les systèmes de transmission, le filtrage adaptatif, certains algorithmes d'apprentissage en reconnaissance des formes, etc. De nombreux exemples se trouvent dans [Sar74,BMP90,LS83].

Le but de ces algorithmes est l'estimation récursive d'un paramètre inconnu et invariant dans le temps (ou lentement variable) traditionnellement noté  $\theta$ . Dans le cadre présenté ci-dessus, où  $H$  est le gradient d'une fonction à minimiser, la structure générale de l'algorithme sera

$$\theta_n = \theta_{n-1} - \gamma_n H(\theta_{n-1}, Z_n)$$

où  $\gamma_n$  est une suite décroissante, typiquement  $1/n$  ou une constante, et  $\theta_n$  est l'estimée de  $\theta^*$  au temps  $n$ . L'étude théorique de ces algorithmes est généralement faite dans un cadre markovien assez large [BMP90] donnant une forme explicite à la dépendance en  $\theta$  de la loi du processus  $Z_n$ .

Si l'on s'intéresse aux gains en  $1/n$ , il est acquis [HH80,NK76,BMP90,KC78,Del96b] que sous certaines hypothèses, peu restrictives mais pas toujours faciles à vérifier,  $\theta_n$  converge presque sûrement vers  $\theta^*$ , et les variables  $\sqrt{n}(\theta_n - \theta^*)$  convergent en loi vers une variable normale de variance  $V$  satisfaisant une certaine équation de Lyapunov [Del96a]. On vérifie que  $V$  peut être améliorée par l'introduction d'une matrice de gain  $\Gamma$  dans l'algorithme :

$$\theta_n = \theta_{n-1} - \gamma_n \Gamma H(\theta_{n-1}, Z_n)$$

et que le choix optimal du gain permet d'atteindre la borne de Cramér–Rao. Malheureusement, cette matrice optimale est généralement inconnue. Cette difficulté peut être évitée par l'usage

- 
- [Sar74] G. N. SARIDIS, «Stochastic Approximation Methods for Identification and Control – A Survey», *IEEE Transactions on Automatic Control AC-19*, 6, décembre 1974, p. 798–809.
  - [BMP90] A. BENVENISTE, M. MÉTIVIER, P. PRIOURET, *Adaptive Algorithms and Stochastic Approximations, Applications of Mathematics, 22*, Springer–Verlag, New York, 1990.
  - [LS83] L. LJUNG, T. SODERSTRÖM, *Theory and Practice of Recursive Identification, Signal Processing, Optimization and Control Series, 4*, The MIT Press, Cambridge, MA, 1983.
  - [HH80] P. HALL, C. C. HEYDE, *Martingale Limit Theory and its Applications, Probability and Mathematical Statistics*, Academic Press, New York, 1980.
  - [NK76] B. M. NEVELSON, R. Z. KHASHINSKII, *Stochastic Approximation and Recursive Estimation, Translations of Mathematical Monographs, 47*, AMS, Providence, 1976.
  - [KC78] H. J. KUSHNER, D. S. CLARK, *Stochastic Approximation Methods for Constrained and Unconstrained Systems, Applied Mathematical Sciences, 26*, Springer–Verlag, New York, 1978.
  - [Del96b] B. DELYON, «General Results on the Convergence of Stochastic Algorithms», *IEEE Transactions on Automatic Control AC-41*, 9, septembre 1996, p. 1245–1256.
  - [Del96a] B. DELYON, *Études d'Algorithmes d'Estimation Adaptative*, Habilitation à diriger des recherches, université de Rennes I, juin 1996.

de l'algorithme de Polyak–Ruppert <sup>[Pol90,DJ95b]</sup> :

$$\begin{aligned}\theta_n &= \theta_{n-1} - \gamma_n H(\theta_{n-1}, Z_n) \\ \bar{\theta}_n &= \bar{\theta}_{n-1} + \frac{1}{n} [\theta_{n-1} - \bar{\theta}_{n-1}] .\end{aligned}$$

avec un gain  $\gamma_n$  typiquement d'ordre  $n^{-2/3}$ . On peut prouver l'optimalité de cet algorithme (convergence de  $\bar{\theta}_n$  vers  $\theta^*$  à même vitesse que  $\theta_n$  dans le cas où  $\Gamma$  est choisi au mieux) dans des circonstances assez générales.

Les algorithmes à gain constant ( $\gamma_n = \gamma$ ) sont utilisés en revanche lorsque le vrai paramètre varie lentement dans le temps (situation de poursuite). Les études se concentrent dans ce cas sur une approche asymptotique où  $\gamma$  et la vitesse de variation de  $\theta_n^*$  sont petits. Un premier aspect est la convergence dans la période transitoire ( $1 \leq n \leq 1/\gamma$ ), période pendant laquelle  $\theta_n$  passe de la valeur initiale  $\theta_0$  à un voisinage de  $\theta_n^*$ . Cette approche dite de l'équation différentielle moyenne est étudiée de façon approfondie dans <sup>[BMP90]</sup> et conduit à la conclusion suivante : si  $\theta_n^*$  varie régulièrement à vitesse  $v$ , le gain doit être choisi d'ordre  $v^{2/3}$ , mais si  $\theta_n^*$  suit une marche aléatoire, le gain doit être proportionnel à l'amplitude moyenne de  $|\theta_{n+1}^* - \theta_n^*|$ . Un deuxième angle d'attaque correspond à l'étude de la distribution limite de  $\theta_n$  lorsque  $\theta_n^*$  a une distribution donnée et  $\gamma$  est fixé <sup>[DJ95a]</sup>, et conduit au même type de conclusion.

Pour l'estimation directe d'un bon gain sans connaissance a priori sur les variations de  $\theta_n^*$ , i.e. la véritable adaptativité, on peut utiliser une méthode statistique inspirée de la solution d'un problème d'estimation non-paramétrique dans un cadre min-max <sup>[Lep91]</sup>. Cette méthode a été mise en œuvre dans un cadre applicatif, pour un problème d'analyse d'images météorologiques, voir le rapport d'activité 1997 du projet VISTA.

### 3.2 Approche asymptotique locale pour la surveillance et le diagnostic des systèmes continus

Voir modules 3.1, 6.1 et 6.2.

**Mots clés** : détection de panne, identification, approche locale, diagnostic de panne, alarme intelligente.

#### Glossaire :

**Approche asymptotique locale** Technique statistique permettant de comparer l'adéquation de deux modèles différents à un même échantillon de données, lorsque la longueur  $N$  de l'échantillon tend vers l'infini. Pour éviter alors des situations singulières, on renormalise l'écart entre ces deux modèles en le rendant proportionnel à  $1/\sqrt{N}$ . Des résultats du type

- 
- [Pol90] B. T. POLYAK, « New Stochastic Approximation Type Procedures », *Automation and Remote Control* 51, 7, 1990, p. 937–946.
- [DJ95b] B. DELYON, A. JUDITSKY, « Stochastic Approximation with Averaging », *Publication Interne n° 952*, IRISA, 1995, <ftp://ftp.irisa.fr/techreports/1995/PI-952.ps.gz>.
- [DJ95a] B. DELYON, A. JUDITSKY, « Asymptotical Study of Parameter Tracking Algorithms », *SIAM Journal on Control and Optimization* 33, 1, janvier 1995, p. 323–345.
- [Lep91] O. V. LEPSKII, « Asymptotically Minimax Adaptive Estimation. I : Upper Bounds. Optimally Adaptive Estimators », *Theory of Probability and its Applications* 36, 4, 1991, p. 682–697.

théorème–limite central montrent que la statistique de test permettant de décider entre ces modèles est asymptotiquement gaussienne, avec une moyenne différente selon que l'échantillon de données provient de l'un ou l'autre modèle.

**Maintenance conditionnelle** Au lieu d'une inspection systématique, il s'agit d'effectuer une surveillance continue de l'installation considérée (machine, structure, procédé, etc.) à partir de données recueillies à la sortie des capteurs, de façon à prévenir l'apparition d'un dysfonctionnement ou d'un endommagement avant qu'il ait pu avoir des conséquences trop graves.

**Alarmes intelligentes** Indicateurs de panne, porteurs d'informations relatives au diagnostic, sous la forme des composants le plus probablement responsables de la panne détectée. Ces indicateurs réalisent automatiquement le compromis entre l'amplitude des changements détectés et la précision de l'identification du modèle de référence d'une part, et le niveau de bruit présent sur les mesures d'autre part. Ces indicateurs sont peu coûteux, et peuvent donc être embarqués.

**Résumé :** *Nous avons développé une méthode statistique de portée générale permettant de confronter un modèle à des données mesurées sur un procédé, et de détecter de façon précoce une éventuelle inadéquation entre modèle et mesures, même si celle-ci est imperceptible de prime abord. Décider de manière précise d'une telle inadéquation nécessite de comparer l'effet prédit par un changement éventuel du procédé, avec les incertitudes que l'on a sur les mesures. L'approche dite «asymptotique locale» introduite dans les années 70 par Le Cam <sup>[Rou72, Le 86]</sup>, et que nous avons étendue et adaptée [5, 2], permet de fonder une telle démarche [3, 8].*

Cette activité se situe dans le prolongement de l'activité ancienne dans l'équipe en détection de changements. La contribution majeure est le développement d'une démarche générale et originale, reposant sur l'approche asymptotique locale en statistique. Cette démarche se trouve confortée par l'intérêt croissant porté dans un grand nombre d'applications industrielles, à la maintenance conditionnelle. L'approche proposée consiste à détecter de façon précoce des petites déviations par rapport au comportement sain du système, en conditions de travail usuelles, c'est-à-dire sans introduire d'excitation artificielle, sans ralentissement ni arrêt de l'installation. Le principe est de construire un «résidu», idéalement nul en fonctionnement nominal, peu sensible aux bruits et aux perturbations, et très sensible aux pannes.

On suppose donné un modèle paramétrique  $\{\mathbf{P}_\theta, \theta \in \Theta\}$  décrivant le système à surveiller, où la loi des observations est caractérisée par le paramètre  $\theta \in \Theta$ , et où le comportement sain correspond à la valeur  $\theta_0$  du paramètre : cette valeur peut représenter un comportement «type» normalisé pour le système considéré, ou bien être obtenue à l'issue d'une étape préliminaire d'identification utilisant des données de référence, par exemple

$$\theta_0 = \arg\{\theta \in \Theta : U_N(\theta) = 0\} ,$$

---

[Rou72] G. G. ROUSSAS, *Contiguity of Probability Measures : Some Applications in Statistics*, Cambridge Tracts in Mathematics, 63, Cambridge University Press, Cambridge, 1972.

[Le 86] L. LE CAM, *Asymptotic Methods in Statistical Decision Theory*, Springer Series in Statistics, Springer-Verlag, New York, 1986.

où la «fonction d'estimation»

$$U_N(\theta) = \frac{1}{N} \sum_{n=0}^N H(\theta, Z_n) ,$$

est telle que  $\mathbf{E}_\theta[U_N(\theta)] = 0$  pour tout  $\theta \in \Theta$ , voir module 3.1. Etant donnée une nouvelle suite de données recueillies en continu à la sortie des capteurs, on se pose les questions suivantes :

- le nouvel échantillon correspond-il toujours à  $\mathbf{P}_{\theta_0}$  ? (il s'agit de savoir si l'échantillon observé est bien en conformité avec le modèle nominal),
- sinon, quelles sont les composantes du paramètre  $\theta$  qui sont responsables de ce changement ? (il s'agit là d'effectuer un diagnostic sur la nature du changement de comportement) — dans le cas où le paramètre  $\theta$  a une signification physique, on obtient ainsi un diagnostic sur l'origine du changement de comportement.

**Détection** L'approche asymptotique locale permet de réduire de façon générique le problème de validation de modèle formulé pour un «système dynamique», en un problème universel et «statique» de détection dans la moyenne d'un «vecteur» aléatoire gaussien. Etant donné un échantillon de longueur  $N$ , on introduit une hypothèse alternative proche de l'hypothèse nominale, et pour décider entre  $\theta = \theta_0$  et  $\theta = \theta_0 + \Delta/\sqrt{N}$ , où  $\Delta$  est un changement inconnu mais fixe, on génère un «résidu» de la forme

$$\zeta_N = \frac{1}{\sqrt{N}} \sum_{n=0}^N H(\theta_0, Z_n) = \sqrt{N} U_N(\theta_0) .$$

On suppose que  $\mathbf{E}_\theta[U_N(\theta_0)] = 0$  si et seulement si  $\theta = \theta_0$  (avec  $\theta_0$  fixé). Si la matrice  $M_N = -\mathbf{E}_{\theta_0}[U'_N(\theta_0)]$ , qui est aussi la dérivée au point  $\theta_0$  de l'application  $\theta \mapsto \mathbf{E}_\theta[U_N(\theta_0)]$ , converge vers une limite  $M$ , alors le théorème-limite central montre [DJB97] que le résidu est asymptotiquement gaussien

$$\zeta_N \implies \begin{cases} \mathcal{N}(0, \Sigma) & \text{sous } \mathbf{P}_{\theta_0} , \\ \mathcal{N}(M \Delta, \Sigma) & \text{sous } \mathbf{P}_{\theta_0 + \Delta/\sqrt{N}} , \end{cases}$$

où la matrice de covariance asymptotique  $\Sigma$  peut être évaluée ou estimée [8]. Décider entre  $\Delta = 0$  et  $\Delta \neq 0$  dans le modèle «limite» se traduit par le test du  $\chi^2$  suivant (en supposant que  $M$  est de rang plein et  $\Sigma$  inversible)

$$t = \bar{\zeta}^T \mathcal{I}^{-1} \bar{\zeta} \geq \lambda .$$

où  $\bar{\zeta} = M^T \Sigma^{-1} \zeta_N$  et  $\mathcal{I} = M^T \Sigma^{-1} M$ .

Cette approche permet de décider (avec un niveau d'erreur quantifiable) si une valeur du résidu est significativement non-nulle pour qu'on puisse déclarer qu'une panne a eu lieu. Il

---

[DJB97] B. DELYON, A. JUDITSKY, A. BENVENISTE, « On the Relationship between Identification and Local Tests », *Publication Interne n° 1104*, IRISA, mai 1997, <ftp://ftp.irisa.fr/techreports/1997/PI-1104.ps.gz>.

est important de remarquer qu'aussi bien le résidu que les matrices de sensibilité  $M$  et de covariance  $\Sigma$  sont évalués (ou estimés) pour le modèle nominal, c'est-à-dire pour la vraie valeur du paramètre. En particulier, il n'est pas nécessaire de ré-identifier le modèle, et le calcul des matrices de sensibilité et de covariance peut être effectué à l'avance.

**Diagnostic** En supposant pour simplifier que la question «quel type de panne a eu lieu» puisse se réduire à la question «quelle composante du vecteur  $\theta$  a changé», le problème de diagnostic peut alors être résolu par des techniques standard d'élimination de paramètres de nuisance. On partitionne le vecteur de panne  $\Delta$ , ce qui induit un partitionnement de la matrice de sensibilité  $M$ , de la matrice d'information de Fisher  $\mathcal{I} = M^T \Sigma^{-1} M$ , et du résidu normalisé  $\bar{\zeta} = M^T \Sigma^{-1} \zeta_N$

$$\Delta = \begin{pmatrix} \Delta_a \\ \Delta_b \end{pmatrix}, \quad \mathcal{I} = \begin{pmatrix} \mathcal{I}_{aa} & \mathcal{I}_{ab} \\ \mathcal{I}_{ba} & \mathcal{I}_{bb} \end{pmatrix}, \quad M = \begin{pmatrix} M_a & M_b \end{pmatrix}, \quad \bar{\zeta} = \begin{pmatrix} \bar{\zeta}_a \\ \bar{\zeta}_b \end{pmatrix}.$$

On peut adopter une approche par «sensibilité», pour décider entre  $\Delta_a = \Delta_b = 0$  et  $\Delta_a \neq 0$ ,  $\Delta_b = 0$  dans le modèle «limité», et on obtient la statistique de test suivante

$$t_a = \bar{\zeta}_a^T \mathcal{I}_{aa}^{-1} \bar{\zeta}_a,$$

où  $\bar{\zeta}_a$  est le résidu «partiel». Si  $t_a \geq t_b$ , on décidera que la panne est causée par la composante  $a$  plutôt que par la composante  $b$ . Alternativement, on peut adopter une approche «minimax», pour décider entre  $\Delta_a = 0$  et  $\Delta_a \neq 0$ , dans le modèle «limité», avec  $\Delta_b$  inconnu, et on obtient la statistique de test suivante

$$t_a^* = \bar{\zeta}_a^{*T} \mathcal{I}_a^{*-1} \bar{\zeta}_a^*,$$

où  $\bar{\zeta}_a^* = \bar{\zeta}_a - \mathcal{I}_{ab} \mathcal{I}_{bb}^{-1} \bar{\zeta}_b$  est le résidu «effectif», résultant de la régression du résidu partiel  $\bar{\zeta}_a$  sur le résidu de nuisance  $\bar{\zeta}_b$ , et le complément de Schur  $\mathcal{I}_a^* = \mathcal{I}_{aa} - \mathcal{I}_{ab} \mathcal{I}_{bb}^{-1} \mathcal{I}_{ba}$  est la matrice d'information de Fisher associée. Si  $t_a^* \geq t_b^*$ , on décidera que la panne est causée par la composante  $a$  plutôt que par la composante  $b$ .

**Diagnostic physique** Cette approche permet également de réaliser un diagnostic physique, c'est-à-dire en termes d'un paramètre  $\phi$  ayant une signification physique (plutôt qu'en termes du paramètre  $\theta$  utilisé pour l'identification et la surveillance, et provenant éventuellement d'une modélisation boîte noire), et ceci sans résoudre le problème inverse. Il suffit pour cela de constater que la matrice de sensibilité par rapport au paramètre physique s'écrit  $M \mathcal{J}$ , où  $\mathcal{J}$  désigne la matrice jacobienne au point  $\phi_0$  de l'application  $\phi \mapsto \theta(\phi)$ , et de mettre en œuvre la procédure de diagnostic sur les composantes du paramètre  $\phi$ .

La démarche systématique présentée ci-dessus fournit un cadre général pour la surveillance et le diagnostic des installations industrielles continues. La question cruciale qui reste à traiter dans chaque classe de modèles, est le choix du résidu, ou de façon équivalente le choix de la fonction d'estimation. Les activités actuelles de l'équipe concernent deux grandes classes de modèles :

- les systèmes dynamiques non-linéaires, voir module 6.1,
- les systèmes linéaires décrivant des structures mécaniques en vibrations, voir module 6.2.

### 3.3 Modèles partiellement stochastiques pour la surveillance des systèmes distribués

**Mots clés :** système distribué, système à événements discrets, concurrence, réseau de Petri, réseau d'automates, modèle partiellement stochastique, HMM, algorithme réparti.

**Glossaire :**

**Système hybride** Désigne génériquement des systèmes combinant de manière intime des aspects traités habituellement de manière disjointe. Par exemple : état continu/état discret, temps continu (ou échantillonné)/événements, équations ou contraintes/variables aléatoires.

**HMM** Modèle de Markov caché (en anglais, hidden Markov model) : automate stochastique (ou chaîne de Markov) dont l'état interne n'est pas observé directement, et ensemble des techniques permettant d'inférer, à partir d'observations «bruitées», cet état caché. Techniques très utilisées en reconnaissance de la parole, et plus récemment dans le diagnostic de systèmes dynamiques distribués à événements discrets.

**Concurrence** Dans un système à événements discrets, deux événements sont dits concurrents s'ils peuvent survenir dans n'importe quel ordre, ou même simultanément, sans changer le comportement ultérieur du système.

**Résumé :** *Ce thème d'étude est en grande partie mené conjointement avec le projet PAMPA (Claude Jard). En vue de l'application au diagnostic des réseaux de télécommunications, nous avons entrepris d'étendre les techniques dites de HMM aux systèmes distribués. Ces systèmes sont modélisés à l'aide de réseaux de Petri (RdP) saufs, ou plus généralement par des réseaux d'automates stochastiques, munis d'une dynamique adaptée aux systèmes distribués, c'est-à-dire fondée sur un modèle du temps du type «ordre partiel».*

*Pour réaliser ceci, il nous a fallu 1) développer une technique d'observateur d'état fondée sur le dépliage [ERV96,ER99] d'un RdP ou d'un produit d'automates, 2) créer une nouvelle notion de système «partiellement stochastique» adaptée à la sémantique en ordres partiels et aux techniques de dépliage, 3) adapter l'algorithme de Viterbi, calculant la trajectoire d'état de vraisemblance maximale d'un HMM, à une notion de temps partiellement ordonné. Les efforts actuels visent à distribuer l'algorithmique d'observation d'état, afin de ne plus manipuler que des états locaux du système surveillé, qui peut être de taille importante.*

**Système distribué** Les systèmes distribués que nous considérons se rangent dans la catégorie des systèmes dynamiques à événements discrets, ou encore des machines à états finis. Le caractère «distribué» souligne une structure particulière de tels systèmes : ils s'obtiennent par

- 
- [ERV96] J. ESPARZA, S. RÖMER, W. VOGLER, « An Improvement of McMillan's Unfolding Algorithm », *in: TACAS'96 : Tools and Algorithms for the Construction and Analysis of Systems*, T. Margaria et B. Steffen (éditeurs), *Lecture Notes in Computer Science, 1055*, Springer-Verlag, Berlin, 1996, p. 87-106.
- [ER99] J. ESPARZA, S. RÖMER, « An Unfolding Algorithm for Synchronous Products of Transition Systems », *in: CONCUR'99 : Concurrency Theory*, J. C. M. Baeten et S. Mauw (éditeurs), *Lecture Notes in Computer Science, 1664*, Springer-Verlag, Berlin, 1999, p. 2-20.

connexion de sous-systèmes élémentaires. Un exemple typique est celui des réseaux de Petri (rdP) : deux rdP peuvent être connectés très simplement par mise en commun de certaines places, qui deviennent ainsi «partagées». Les jetons de ces places peuvent alors circuler d'un réseau à l'autre, ce qui permet de modéliser des interactions comme l'exclusion mutuelle, la transmission de ressource, etc. De façon générale, un système distribué peut se voir comme un réseau d'automates élémentaires, définis chacun sur un ensemble de variables d'état. La connexion se fait par mise en commun de variables (celles-ci jouent le rôle des places dans les rdP). Le diagramme de connexion des différents sous-systèmes définit ainsi un réseau d'interaction.

La modélisation d'un système complexe peut être grandement simplifiée par une approche modulaire. L'intérêt des systèmes distribués est ailleurs cependant : en isolant les points d'interaction entre sous-systèmes, on met aussi en évidence des zones de comportements concurrents. En d'autres termes, tant que deux sous-systèmes n'interagissent pas par le biais de variables partagées, leurs évolutions sont indépendantes. Cette tautologie mérite d'être signalée : elle indique des «zones de concurrence» entre sous-systèmes, dans lesquelles les évolutions peuvent obéir à des horloges indépendantes, sans conséquence sur le comportement du système global. Cela signifie que les événements d'un système distribué ne sont que partiellement ordonnés dans le temps, i.e. qu'une notion de temps global n'est pas pertinente. Les trajectoires de systèmes distribués se décrivent ainsi dans une sémantique dite «d'ordre partiel», ou «de concurrence vraie», qui manipule des ordres partiels d'événements au lieu de séquences d'événements. La taille de l'espace des trajectoires s'en trouve considérablement réduite. Ces espaces de trajectoires se représentent à l'aide de «dépliage» du système, au lieu des graphes de marquages.

**Cadre stochastique** Si l'on s'en tient aux réseaux de Petri, de nombreuses notions de rdP stochastiques ont été proposées. Si l'on excepte le cas des rdP à choix libre, aucune notion de rdP stochastique ne fournit une coïncidence exacte entre «concurrence» et «indépendance stochastique» pour un ensemble de transitions. On souhaiterait en effet que deux transitions concurrentes (qui ne sont pas reliées à une place commune), lorsque leur franchissement est probabilisé, se comportent comme des variables aléatoires indépendantes.

On peut montrer que cette exigence est contradictoire avec une dynamique du rdP décrite sous forme de chaîne de Markov, ce qui est le cas pour les notions habituelles de rdP stochastique. Parler de chaîne de Markov suppose en effet une notion de temps global rythmant les événements : il s'ensuit que la vraisemblance de deux événements concurrents dépend de l'ordre dans lequel ils apparaissent. En revanche, le bon cadre est celui des champs markoviens avec contraintes, qui correspond au modèle dit CSS, développé par A. Benveniste, É. Fabre, B.C. Lévy et P. Le Guernic [4]. L'idée de ce modèle est simple : 1) pour chaque place sujette à conflit (amont et/ou aval), le choix est probabilisé isolément, 2) on considère l'ensemble de ces places, muni de la loi produit (elles sont alors indépendantes), 3) on y adjoint les places non sujettes à conflit, 4) prenant en compte les règles de tir des rdP, on ne conserve que les séquences légales et la loi de probabilité qui en résulte est la loi produit précédemment introduite, «conditionnellement à l'ensemble des séquences légales». On montre alors qu'on a bien parfaite adéquation entre «concurrence» et «indépendance stochastique» pour un ensemble de transitions. En particulier, la dynamique de ces rdP n'est pas décrite par une chaîne de Markov, d'où le nom de rdP «partiellement stochastiques».



Ces idées se généralisent sans difficulté à un réseau d'automates stochastiques. Il suffit de considérer l'espace produit formé des trajectoires de chaque automate. Les trajectoires de sous-systèmes différents sont vues comme indépendantes, et l'on munit donc l'espace produit de la loi produit. On conditionne ensuite cet espace par la contrainte de cohérence des  $t$ -uplets de trajectoires pour la circulation des ressources partagées. On obtient, comme pour les RdP, un automate produit «partiellement stochastique».

**Algorithmique répartie de reconstruction d'état** On dispose donc d'une notion de système (partiellement) stochastique pour laquelle nous voulons généraliser les techniques HMM. En supposant que les transitions du système produisent des événements observables, il s'agit de retrouver la trajectoire du système la plus vraisemblable expliquant les observations, cette trajectoire étant vue comme un ordre partiel d'événements. Nous avons montré qu'il suffit de considérer des «pièces» constituées d'une transition et des variables (ou places) qui lui sont reliées, et d'associer à chaque pièce prise isolément une «vraisemblance» issue de notre modèle stochastique, vraisemblance conditionnée par les observations. Les trajectoires du système peuvent ainsi se voir comme des puzzles formés de ces pièces, et leurs vraisemblances s'obtiennent par produit des vraisemblances des pièces utilisées. Le décodage au maximum de vraisemblance revient alors à construire récursivement les meilleurs puzzles, à la manière d'une programmation dynamique, voir A. Aghasaryan, É. Fabre, A. Benveniste, R. Boubour et C. Jard [1].

Ce schéma est opérationnel lorsque les observations de tout le système sont rassemblées en un seul endroit, appelé «superviseur». Cette situation n'est pas la plus pertinente en pratique : une structure répartie d'observation semble plus naturelle pour un système distribué. Nous avons montré qu'il est inutile de rassembler les observations, et qu'une algorithmique HMM répartie est beaucoup plus judicieuse. Celle-ci se compose de plusieurs agents : chacun traite les observations issues du sous-système qu'il surveille à l'aide des pièces locales à ce sous-système. On reconstruit ainsi une partie du puzzle global. Les communications entre agents ne concernent que les pièces touchant à des ressources partagées. On obtient ainsi une véritable reconstruction parallèle et asynchrone du puzzle global, manipulant uniquement des états locaux, et exploitant au mieux le réseau d'interaction et la concurrence entre sous-systèmes.

## 4 Domaines d'applications

Les domaines d'application du projet sont divers, nous citons ici ceux qui ont donné ou donnent lieu à des applications avec partenaire industriel et données réelles à traiter. Ce sont : la mécanique des vibrations, l'électronique embarquée pour l'industrie automobile, le diagnostic dans les réseaux de télécommunications, le contrôle d'accès dans les communications mobiles, la reconnaissance de la parole, les problèmes d'indexation en multimedia, le traitement d'antenne, l'énergie, la géophysique, et la robotique sous-marine. Nous avons choisi de détailler notre action en mécanique des vibrations, qui constitue clairement notre investissement cumulé majeur. Puis nous décrivons notre activité dans le domaine des télécommunications, dont l'augmentation résulte d'un choix du projet.

## 4.1 Analyse de structures vibrantes en ambiance de travail

Voir modules 3.2, 6.2 et 7.1.

**Mots clés :** vibration, structure mécanique, analyse modale, méthode de sous-espace.

### Glossaire :

**Analyse modale** Identification des «modes de vibration», consistant en 1) les fréquences de vibration et amortissements associés, et en 2) la partie observée des vecteurs propres associés.

**Méthodes de sous-espace** Désigne génériquement un algorithme pour l'identification des systèmes linéaires à partir d'une suite de matrices de covariance de la sortie, dans lequel un rôle essentiel est joué par différents sous-espaces de vecteurs aléatoires gaussiens [vD96].

**Résumé :** *Dans une série d'études entreprises depuis 1980, le projet SIGMA2 a développé une technologie originale offrant les services suivants, «pour une structure en ambiance de travail» : 1) analyse modale, 2) corrélation entre mesures et modèle, 3) détection et diagnostic de fatigues. Le fait, pour ces méthodes, d'opérer en ambiance de travail, impose les contraintes suivantes : 1) l'excitation est naturelle, résultant des conditions mêmes de fonctionnement de la structure, elle est souvent non stationnaire, 2) l'excitation n'est pas mesurée.*

Les applications industrielles de la surveillance vibratoire en fonctionnement sont diversifiées, que ce soit pour des structures mécaniques complexes (plate-formes offshore, ponts, barrages, bâtiments, avions) ou les machines (turbo-alternateurs, systèmes d'engrenage). Des outils de détection et diagnostic de petits changements de caractéristiques vibratoires sont particulièrement utiles pour la mise en place de politiques de maintenance préventive basées sur l'évolution effective de l'état de la machine ou de la structure surveillée, par opposition à une programmation a priori systématique.

Les méthodes classiques d'analyse et de surveillance vibratoires des structures mécaniques et machines tournantes sont essentiellement dédiées au traitement de mesures prises soit sur banc d'essai, soit dans des conditions d'excitation ou de vitesse de rotation spécifiques. L'objet du projet Euréka SINOPSYS «Model Based Structural Monitoring Using in-Operation System Identification» coordonné par LMS (Leuven Measurements Systems, Leuven, Belgique) est précisément le développement et l'intégration de logiciels d'analyse et de surveillance vibratoires dédiés au traitement de mesures prises pendant le fonctionnement usuel de la structure ou machine considérée, sans excitation artificielle, ni ralentissement, ni arrêt de machine. Les projets SIGMA2 et METALAU sont conjointement engagés dans SINOPSYS.

La principale contribution de l'INRIA à SINOPSYS consiste en une algorithmique originale de traitement de signaux multi-capteurs (d'accélérométrie par exemple), fournissant des alarmes intelligentes, c'est-à-dire des alarmes donnant des causes profondes des défauts ou fatigues subis par la structure ou la machine. Ces logiciels peuvent être embarqués et fonctionner en ligne. Parmi les données réelles que l'INRIA traite avec les logiciels développés dans SINOPSYS, figurent les données des vols d'essai d'Ariane 5.

---

[vD96] P. VAN OVERSCHEE, B. DE MOOR, *Subspace Identification for Linear Systems*, Kluwer Academic Publishers, Boston, 1996.

Dans la première étape du projet SINOPSYS, focalisée sur l'analyse modale et l'identification des modes de vibration, nous avons amélioré nos procédures interactives de sélection et de validation de modes, et développé un module de «validation de modèle», permettant une validation croisée du résultat d'une identification sur un jeu de données de validation. Dans une seconde étape, nous avons développé un outil de détection de fatigue, en évaluant le degré d'importance de la modification du comportement modal, pour chaque mode. Ceci fonctionne, tant sur données de laboratoire avec excitation mesurée, que sur données en fonctionnement sans mesurer l'excitation. L'ensemble des fonctionnalités de ces deux étapes est intégré dans l'outil CADA\_X de LMS d'une part, et dans une boîte à outils pour le «freeware» Scilab d'autre part. La dernière étape a porté sur le développement d'un outil de diagnostic des fatigues, où l'on cherche à expliquer les fatigues en termes de modification de la masse volumique ou du module d'Young, avec localisation de ces changements sur la structure.

Un objectif important de notre coopération avec LMS est le diagnostic dans le domaine du génie civil. C'est toutefois dans le domaine de l'aéronautique que nous avons pu construire une nouvelle coopération, portant sur l'utilisation des données de vol sous excitation naturelle pour les essais en vol, voir module 7.1.

Une telle algorithmique de surveillance et de diagnostic a été généralisée à des modèles plus complexes que ceux liés aux vibrations, et peut être utilisée dans le cadre de la surveillance pour l'aide à la conduite de procédés industriels (turbine à gaz, centrale électrique ou thermique, etc.) ou pour le diagnostic embarqué (pot catalytique, etc.).

## 4.2 Télécommunications : diagnostic de pannes en gestion de réseaux, turbo-codes, décodage conjoint source-canal

Voir modules 3.3, 6.3, 6.4, 7.2 et 7.3.

**Mots clés** : réseau de télécommunication, gestion de réseau, supervision, gestion d'alarmes, diagnostic, turbo-code, décodage conjoint source-canal.

### Glossaire :

**Gestion de réseau** Désigne la couche haute de gestion d'un réseau de télécommunications, c'est-à-dire les opérations de supervision : surveillance, maintenance, etc.

**Gestion d'alarmes** Opérations de traitement, de filtrage et d'interprétation des alarmes circulant sur le réseau. Dans ce contexte, le diagnostic désigne l'interprétation des alarmes en vue des opérations de reconfiguration et de maintenance.

HMM voir module 3.3.

**Turbo-code** Codes correcteurs d'erreurs récemment introduits par Berrou, Glavieux et Thitimajshima <sup>[BGT93]</sup>, alliant deux codes convolutifs via un entrelaceur. Le nom provient de l'algorithme itératif de décodage, manipulant une information «soft», de nature probabiliste.

Un investissement important, à l'heure actuelle, concerne le «diagnostic des pannes» en gestion de réseaux. Nous cherchons, à partir d'une modélisation du réseau à un niveau convenable d'abstraction, à engendrer automatiquement l'algorithmique de suivi de comportement, et en

---

[BGT93] C. BERROU, A. GLAVIEUX, P. THITIMAJSHIMA, «Near Shannon Limit Error-Correcting Coding and Decoding : Turbo Codes», in : *Proceedings of the IEEE International Conference on Communications, Geneva 1993*, 2, IEEE-CS, p. 1064-1070, mai 1993.

particulier de diagnostic. Cette technique permet d'envisager une mise à jour plus aisée des logiciels de diagnostic lorsque le réseau évolue. L'originalité de cette approche est que nous cherchons d'emblée une modélisation modulaire du réseau, dans le but de distribuer l'algorithmique de diagnostic. Ce travail se fait en collaboration avec le projet PAMPA, et fait l'objet du projet exploratoire RNRT MAGDA, voir module 7.2.

Avec un point de vue plus proche de la couche physique, nous avons commencé à nous intéresser à la surveillance, et à terme au diagnostic, d'une ligne de transmission de données binaires par fibre optique, voir module 7.3. Il s'agit en quelque sorte d'un retour aux méthodes analogiques, pour assurer la transparence, mais avec des contraintes de très haut débit.

Par ailleurs, nous nous intéressons également aux turbo-codes et à leurs extensions, en raison du lien récemment établi [MMC98] entre ces codes et les réseaux bayésiens, sujet traité dans le projet depuis plusieurs années. Les techniques de décodage «soft» peuvent en effet se lire comme un problème d'estimation de variables cachées dans un réseau bayésien (ou un champ de Markov). Ce problème classique en traitement du signal et des images a été abordé sous de nombreux angles dans le projet, ce qui suggère de nouvelles pistes de conception de codes et d'algorithmes de décodage.

Les techniques turbo sont aussi mises à contribution dans un problème de décodage conjoint source-canal, en collaboration avec le projet TEMICS. Il s'agit d'exploiter au mieux toute l'information a priori contenue dans une chaîne de transmission comportant une source markovienne, un codeur de source (de type Huffman) utilisant un code de longueur variable, et enfin un code correcteur d'erreur. Ces trois éléments sont habituellement utilisés indépendamment (dans l'ordre inverse) lors de la réception. Nous avons montré que des gains de performance significatifs peuvent être obtenus par des méthodes itératives, utilisant alternativement les trois modèles. Ce principe semble général et est actuellement exploré dans de nombreux problèmes de traitement du signal pour les télécommunications (par exemple, turbo-égalisation).

## 5 Logiciels

### 5.1 Nonlinear Modelling Matlab Toolbox

**Mots clés :** identification, identification boîte-noire, non-linéarité, non-paramétrique, réseau de neurones, réseau d'ondelettes, Matlab toolbox.

**Résumé :** *En coopération avec Lennart Ljung, de l'université de Linköping, et Anatoli Juditsky, de l'université Joseph Fourier, nous avons développé une boîte à outils Matlab. Cette boîte à outils est conçue comme une extension de la SITB (System Identification Toolbox) de Lennart Ljung pour la modélisation de systèmes dynamiques non linéaires. En plus des techniques classiques pour l'estimation de modèles paramétriques, les algorithmes sont basés sur l'estimation non-paramétrique, avec notamment les réseaux d'ondelettes et les réseaux de neurones. Les modèles*

---

[MMC98] R. J. McELIECE, D. J. C. MacKAY, J.-F. CHENG, «Turbo Decoding as an Instance of Pearl's Belief Propagation Algorithm», *IEEE Journal on Selected Areas in Communications SAC-16*, 2, février 1998, p. 140-152.

*proposés sont, pour l'essentiel, de type régression ou auto-régressif non-linéaire (NARX), avec quelques extensions spécifiques pour lesquelles on dispose de bons algorithmes.*

**Participants :** Albert Benveniste, Bernard Delyon, Qinghua Zhang [correspondant].

Sur la base d'un travail conduit en coopération avec l'université de Linköping [6, 7], nous avons réalisé une boîte à outils Matlab pour la modélisation de systèmes dynamiques non linéaires, prolongeant la SITB «System Identification Toolbox» de Lennart Ljung. L'interface de dialogue ainsi que l'interface graphique sont très largement communes avec la SITB.

En ce qui concerne les modèles offerts, ce sont d'une part, des modèles paramétriques sous forme d'état, d'autre part, des modèles boîte-noire de type 1) régression non-linéaire, 2) NARX (non-linéaire, autorégressif avec entrée exogène), 3) Wiener (ARMA linéaire précédé d'une non-linéarité statique), et 4) Hammerstein (ARMA linéaire suivi d'une non-linéarité statique). Pour les modèles boîte-noire, l'originalité consiste en l'utilisation intensive d'algorithmes non itératifs, ne faisant pas appel à la rétropropagation ni à des méthodes de gradient. On gagne ainsi en vitesse d'identification de manière spectaculaire, et l'on évite les écueils liés à l'accrochage d'une méthode d'optimisation (comme la rétropropagation) sur un optimum local. Pour ces méthodes, voir les articles [6, 7]. Ces méthodes sont en outre complétées par des techniques de rétropropagation, étendues à certaines catégories de systèmes NARMAX (ARMAX non-linéaires). Outre les services d'identification proprement dite, sont également offerts des moyens de valider une modélisation conduite avec une classe restreinte de modèles (par exemple, on peut tester si une modélisation linéaire est ou non suffisante).

La programmation de la boîte à outils exploite largement les fonctionnalités de Matlab en matière de programmation objet.

## 6 Résultats nouveaux

### 6.1 Surveillance et estimation de systèmes dynamiques non-linéaires

*Voir module 3.2.*

**Mots clés :** observateur à grand gain, observateur adaptatif, observateur implicite.

**Participants :** Qinghua Zhang, Aiping Xu, Bernard Delyon, Michèle Basseville, Olivier Perrin.

Les systèmes industriels qui requièrent un dispositif de surveillance ont souvent un comportement significativement non-linéaire. La linéarisation autour d'un point de fonctionnement est souvent inadaptée pour les besoins de la surveillance, par conséquent il est important de développer des méthodes de surveillance pour les systèmes non-linéaires. Le problème est formulé comme la détection et le diagnostic des variations du vecteur de paramètres  $\theta$  dans un système obéissant à un modèle d'état non-linéaire. La présence de variables d'état non mesurées dans un système non-linéaire rend le problème particulièrement difficile.

Nous avons déjà étudié plusieurs méthodes pour résoudre ce problème, qui sont basées sur un observateur et sa dérivée par rapport au vecteur  $\theta$ , sur l'élimination des variables

d'état <sup>[ZB99]</sup>, ou sur un ensemble d'observateurs adaptatifs [28]. La conception d'observateurs, adaptatifs ou non, pour des systèmes non-linéaires est en général un problème difficile. Cette année nous avons étudié des observateurs non-linéaires adaptatifs avec convergence globale. Cette activité se situe dans le cadre de la thèse de Aiping Xu qui a débuté à la rentrée 1999. Inspirés par des résultats connus sur les observateurs à grand gain et par des observateurs implicites <sup>[Nik98]</sup> introduits à l'origine par Ramine Nikoukhah du projet METALAU à Rocquencourt, nous avons développé une méthode pour concevoir des observateurs adaptatifs non-linéaires et implicites. C'est le premier résultat connu sur la conception systématique d'observateurs adaptatifs pour une classe de systèmes intrinsèquement non-linéaires et avec une convergence globale. Les observateurs adaptatifs ainsi conçus permettent d'estimer conjointement l'état et certains paramètres du système. Avec cette méthode, le nombre de paramètres estimés est limité par le nombre de sorties mesurées du système, ce qui signifie que, pour les systèmes mono-sortie, un seul paramètre peut être estimé conjointement avec l'état. Cette restriction est cependant compatible avec les besoins de la méthode de surveillance basée sur des observateurs adaptatifs présentée dans [28].

L'étude de la pertinence de ces différentes méthodes pour la surveillance d'organes de véhicules automobiles à basse consommation débute fin 2000, dans le cadre de la thèse Cifre d'Olivier Perrin, en collaboration avec le projet SOSSO à Rocquencourt et la direction de la recherche de Renault.

## 6.2 Méthodes de sous-espace, application à la surveillance de structures soumises à vibrations en opération

Voir modules 3.2, 4.1 et 7.1.

**Mots clés** : vibration, structure mécanique, analyse modale, méthode de sous-espace, polyréférence.

**Participants** : Michèle Basseville, Albert Benveniste, Laurent Mevel.

Ce travail est conduit en coopération avec Maurice Goursat du projet METALAU à Rocquencourt, dans le cadre du projet Euréka SINOPSYS coordonné par la société LMS, voir modules 4.1 et 7.1.

Le problème traité est celui de l'identification et de la surveillance de la structure propre (valeurs propres de  $F$  et partie observée des vecteurs propres associés) d'un système de la forme :

$$\begin{cases} X_{k+1} &= F X_k + V_k \\ Y_k &= H X_k + W_k \end{cases}$$

où la dimension de l'observation  $Y_k$  est beaucoup plus petite que celle de l'état  $X_k$ . Nous exploitons, de manière différente pour l'identification et la surveillance, le fait que la matrice d'observabilité  $\mathcal{O}$  du couple  $(H, F)$  et la matrice de Hankel  $\mathcal{H}$  des covariances empiriques

---

[ZB99] Q. ZHANG, M. BASSEVILLE, « Local Approach to FDI in Nonlinear Dynamical Systems », in : *Proceedings of the 5th European Control Conference (ECC'99), Karlsruhe, 1999.*

[Nik98] R. NIKOUKHAH, « A new methodology for observer design and implementation », *IEEE Transactions on Automatic Control AC-43*, 2, février 1998, p. 229-234.

des observations ont même noyau à gauche, propriété dite de sous-espace. Notre méthode de surveillance est basée sur l'application de l'approche locale, voir module 3.2, sur cette propriété de sous-espace [12].

Les résultats nouveaux obtenus au cours de l'année concernent l'identification et la surveillance, et sont d'ordres théorique et expérimental.

**Identification par combinaison de mesures prises sur des jeux de capteurs différents.** Il arrive fréquemment, en analyse modale, que l'on cherche à utiliser quelques capteurs fixes (qui servent ainsi de référence) et un nombre relativement petit de capteurs mobiles, dans le but d'imiter une situation où un nombre plus important de capteurs est disponible. En général, l'analyse modale est alors effectuée par identification sur chaque enregistrement pris séparément, puis fusion des résultats. Se pose toutefois la question de savoir s'il est possible de fusionner d'abord l'ensemble des données prises pour les différents emplacements de capteurs, et d'effectuer l'analyse modale ensuite, en utilisant toujours l'algorithme sous-espace évoqué plus haut, basé sur la factorisation de la matrice de Hankel empirique.

Il est facile de mettre en évidence que, lorsque l'excitation est stationnaire, empiler les matrices de covariance des divers enregistrements et remplir la matrice de Hankel avec ces covariances empilées est une solution pertinente. Dans le cas de l'identification en fonctionnement, le caractère non-stationnaire de l'excitation n'autorise plus cet empilement de covariances. Nous avons proposé une normalisation non triviale du facteur droit de la factorisation de la matrice de Hankel pour chaque enregistrement, qui réalise la normalisation des covariances nécessaire en préalable à leur empilement. Cette solution au problème de fusion présente en outre, sur le plan algorithmique, l'avantage d'une mise en œuvre économique, par simple entrelacement des blocs-colonnes ou des blocs-lignes des matrices de Hankel des différents enregistrements [29, 30]. Nous pouvons donc maintenant proposer une version dite «polyréférence» pour la méthode d'identification par sous-espace en ambiance de travail.

Cette année, la méthode a été effectivement mise en œuvre sur données réelles, voir ci-dessous. En outre, un théorème de consistance en non-stationnaire de l'estimateur de structure propre ainsi obtenu a été démontré [65], qui généralise le théorème démontré dans le passé par A. Benveniste et J.J. Fuchs [BF85].

**Expérimentations.** Diverses expérimentations sur structures simulées ou réelles ont été conduites par Laurent Mevel et Luc Hermans (LMS).

- Des données expérimentales ont été collectées sur un pont désaffecté en Suisse, situé au dessus d'une autoroute, au cours d'un processus de dégradation volontaire de la structure [HBvM99,MBH99]. Les premières études suggéraient une approche ponctuelle du test

---

[BF85] A. BENVENISTE, J.-J. FUCHS, «Single Sample Modal Identification of a Non-Stationary Stochastic Process», *IEEE Transactions on Automatic Control AC-30*, 1, janvier 1985, p. 66-74.

[HBvM99] L. HERMANS, M. BRUGHMANS, H. VAN DER AUWERAER, L. MEVEL, «Damage Diagnosis in a Concrete Three-Span Bridge : A Case Study», in : *Proceedings of the 1999 SEM Annual Conference on Theoretical, Experimental, and Computational Mechanics, Cincinnati*, SEM, Inc., juin 1999. Paper 192.

[MBH99] L. MEVEL, M. BASSEVILLE, L. HERMANS, «Diagnostic Vibratoire Embarqué», in : *17ème Colloque GRETSI, Vannes*, p. 143-146, septembre 1999.

de détection. Le développement d'un module de détection automatique embarquable a permis une meilleure compréhension des propriétés statistiques du test en utilisation sur données réelles [42, 54]. Cette étude a montré la grande stabilité du test au cours du temps et sa réactivité aux procédures d'endommagement. On en a déduit que cette technique se prêtait plus à une utilisation de type surveillance continue [54] plutôt qu'à une approche ponctuelle [HvM99,MHv99]. Ces derniers résultats ont donc tempéré les réticences sur la méthode énoncées précédemment.

- Les techniques de détection polyréférence ont été testées sur les mêmes jeux de données que le test de détection [42, 54]. Il a été montré [55] que cette approche permet :
  - de fusionner les données et d'obtenir des identifications modales de bien meilleure qualité, tant au niveau du nombre de modes extraits que de la qualité et de la stabilité de ceux-ci (diagramme plus stable, amortissements mieux estimés),
  - une meilleure robustesse par rapport au rejet des «faux» modes — correspondant à l'excitation ou introduits par la méthode d'identification,
  - l'obtention de diagramme de stabilité «propre» suggérant l'utilisation de cette méthode dans une procédure d'identification automatique.

Les deux applications ont montré l'utilité et le champ d'application de ces méthodes. Elles sont donc considérées comme satisfaisantes du point de vue de la validation expérimentale des théories développées par ailleurs.

### 6.3 Surveillance des systèmes distribués

*Voir modules 3.3, 4.2 et 7.2.*

**Mots clés** : système distribué, système à événements discrets, réseau de Petri, réseau d'automates, réseau bayésien, HMM, multi-agents.

**Participants** : Éric Fabre, Albert Benveniste, Vincent Pigourier.

La surveillance des systèmes distribués était initialement décrite dans le formalisme des réseaux de Petri à capacité un. Ce cadre a été étendu à des compositions d'automates définis sur plusieurs variables, l'interaction s'opérant par mise en commun de variables (de même que les transitions d'un RdP interagissent par mise en commun de places). Les notions de pièces

---

[HvM99] L. HERMANS, H. VAN DER AUWERAER, L. MEVEL, «Health Monitoring and Detection of a Fatigue Problem of a Sports Car», *in: Proceedings of the 17th International Modal Analysis Conference (IMAC-XVII)*, Kissimmee, SEM, Inc., p. 42-48, février 1999.

[MHv99] L. MEVEL, L. HERMANS, H. VAN DER AUWERAER, « On the Application of Subspace-Based Fault Detection Methods to Industrial Structures», *Mechanical Systems and Signal Processing 13*, 6 (Special section on Model-Based Structural Identification and Monitoring Using in-Operation Data), novembre 1999, p. 823-838.



et de dépliages [ERV96,ER99] s'étendent naturellement à ces systèmes. De même, les techniques de randomisation partielle, assurant une cohérence entre la notion de concurrence et celle d'indépendance stochastique, y trouvent aussi un prolongement naturel.

Ce formalisme plus général permet de définir des «réseaux bayésiens de systèmes dynamiques». Les réseaux bayésiens permettent de visualiser (et d'exploiter) la structure des relations de dépendance d'un ensemble de variables aléatoires. Plus précisément, ce sont les relations d'indépendance conditionnelle qui sont pertinentes. La construction d'un système distribué par composition d'automates élémentaires permet de voir le système obtenu comme un réseau d'interaction entre systèmes dynamiques, dont les points de contact sont les ressources partagées. Le graphe ainsi obtenu décrit non pas des dépendances stochastiques, mais des contraintes sur les trajectoires des systèmes élémentaires ainsi reliés. Nous avons montré que le parallèle va plus loin : les réseaux bayésiens traduisent les propriétés de factorisation d'une distribution jointe de probabilité. De même ici, le réseau d'interaction traduit les propriétés de factorisation de l'ensemble des trajectoires du système global, dans une sémantique de concurrence vraie.

Ce résultat remarquable permet de transcrire directement pour les systèmes distribués toute une gamme d'algorithmes d'inférence efficaces conçus pour les réseaux bayésiens et les modèles graphiques, par exemple les algorithmes d'inférence sur les arbres. L'avantage principal de ces méthodes est qu'elles ne manipulent que la forme factorisée de l'ensemble des trajectoires compatibles avec les observations. Ainsi, il n'est jamais nécessaire de manipuler l'ensemble des états globaux du système, qui serait beaucoup trop gros : on n'utilise à la place que des états locaux de sous-systèmes, en assurant leur cohérence avec les états des sous-systèmes voisins. Qui plus est, ces facteurs, ou morceaux locaux de trajectoires, peuvent être distribués et maintenus par des agents distincts et asynchrones (un par sous-système). Ceci permet de définir une véritable architecture distribuée de supervision [38, 61]. Cette algorithmique est en cours de prototypage, en vue de développer un superviseur distribué pour une maquette de réseau SDH, voir le projet MAGDA, module 7.2. On démontre par ailleurs que, dans ce cadre, les fonctions de maintien de la cohérence des meilleures trajectoires locales, et de recherche de la meilleure histoire globale sont toutes deux assurées par des algorithmes d'arbre.

Les efforts actuels, en matière d'algorithmique, portent sur le traitement des pertes d'observations. Un prototypage est en cours, prenant en compte des topologies quelconques de systèmes, ce qui posera des problèmes nouveaux pour traiter de cycles d'interaction. Sur le plan du formalisme, les points d'intérêt concernent les cas de modèles incomplets, et l'identification des probabilités dans un système partiellement stochastique. Ce dernier point demande le développement d'une théorie complète des chaînes de Markov cachées (hidden Markov models, HMM), pour les systèmes distribués asynchrones [13].

---

[ERV96] J. ESPARZA, S. RÖMER, W. VOGLER, « An Improvement of McMillan's Unfolding Algorithm », in : *TACAS'96 : Tools and Algorithms for the Construction and Analysis of Systems*, T. Margaria et B. Steffen (éditeurs), *Lecture Notes in Computer Science*, 1055, Springer-Verlag, Berlin, 1996, p. 87-106.

[ER99] J. ESPARZA, S. RÖMER, « An Unfolding Algorithm for Synchronous Products of Transition Systems », in : *CONCUR'99 : Concurrency Theory*, J. C. M. Baeten et S. Mauw (éditeurs), *Lecture Notes in Computer Science*, 1664, Springer-Verlag, Berlin, 1999, p. 2-20.

## 6.4 Codes correcteurs et modèles graphiques

Voir module 4.2.

**Mots clés** : turbo-code, modèle graphique, réseau bayésien, décodage conjoint source-canal.

**Participants** : Éric Fabre, Arnaud Guyader.

Il a récemment été établi par McEliece, MacKay et Cheng [MMC98] que les algorithmes itératifs de décodage «soft» pour les turbo-codes peuvent se lire comme des algorithmes de propagation de croyance dans un réseau bayésien décrivant les interactions entre les variables aléatoires du code (les bits). En ce sens, le décodage «soft» se ramène à un problème d'estimation de variables cachées dans un réseau bayésien, les autres variables de ce champ étant parfaitement observées. Posé en ces termes, le problème de décodage se rapproche de sujets très classiques en traitement du signal et des images, et l'algorithme de propagation de croyance peut se lire comme une généralisation d'un lisseur de Kalman.

Pour les codes correcteurs, le réseau bayésien (ou champ de Markov) décrivant les interactions entre variables peut admettre une structure simple. C'est le cas des codes convolutifs, qui ont une structure de chaîne de Markov. On sait que l'algorithme de propagation de croyance converge alors en temps fini vers la solution optimale, et permet donc un décodage au maximum de vraisemblance «exact». Dans le cas des turbo-codes, entrelaçant deux codes convolutifs, le graphe sous-jacent est beaucoup plus compliqué. En particulier, il n'a pas de structure d'«arbre», et la propagation de croyance devient donc une méthode «approchée» de décodage. On peut démontrer que cette méthode est d'autant meilleure que le graphe compte peu de cycles, et que leur longueur est grande.

Ces constatations ont ramené l'attention de la communauté de codage vers la construction de codes directement à partir d'une représentation graphique, et notamment les codes de Gallager, dits «sparse parity checks codes», se sont révélés aussi performants que les turbo-codes. Leur inconvénient reste de n'offrir ces performances que pour des longueurs de bloc importantes.

Nos travaux en ce domaine portent sur deux points. Le premier concerne la construction de codes correcteurs à partir de leur représentation sous forme de champ de Markov. Il s'agit en particulier de construire des codes offrant de bonnes propriétés de protection des bits et résistant aux méthodes de décodage par propagation de croyance. L'expérience montre que de tels codes sont difficiles à construire pour de petites longueurs. Nous nous intéressons à des longueurs intermédiaires, très inférieures à celles des turbo-codes ou des «sparse parity check codes». Nous adoptons une construction récursive du graphe, par incorporation de contrainte de façon à optimiser le degré de protection tout en maintenant des propriétés graphiques intéressantes. Nous nous intéressons ensuite à la conception d'algorithmes approchés de décodage exploitant plus précisément la structure du code. Ces algorithmes étendent et combinent des méthodes d'estimation classiques pour les réseaux bayésiens et le traitement d'images, et permettent d'étendre la portée de la simple propagation de croyance. En particulier, ils permettent

---

[MMC98] R. J. McELIECE, D. J. C. MACKAY, J.-F. CHENG, «Turbo Decoding as an Instance of Pearl's Belief Propagation Algorithm», *IEEE Journal on Selected Areas in Communications SAC-16*, 2, février 1998, p. 140-152.

de s'affranchir des difficultés liées à la présence de cycles courts. Des résultats positifs ont été obtenus sur des codes courts combinant les aspects algébriques et graphiques : ils sont conçus comme des codes algébriques élémentaires mis en interaction via une structure graphique [39].

## 6.5 Modèles de Markov cachés généraux

**Mots clés :** HMM, filtrage particulaire, approche locale.

### Glossaire :

HMM voir module 3.3.

**Filtrage particulaire** Méthode numérique permettant d'approcher la distribution de probabilité conditionnelle de l'état caché sachant les observations, au moyen de la distribution empirique d'un système de particules, qui se déplacent selon des réalisations indépendantes de l'équation d'état, et qui sont redistribuées en fonction de leur cohérence (quantifiée par la fonction de vraisemblance) avec les observations.

**Approche locale** voir module 3.2.

**Résumé :** *On s'intéresse à la fois aux chaînes de Markov cachées classiques (avec un espace d'état fini) et aux chaînes de Markov cachées générales (avec un espace d'état continu), et dans ce dernier cas on étudie aussi des méthodes d'approximation numérique.*

### Statistique asymptotique des HMM à état fini

**Participants :** Samy Abbes, François Le Gland, Laurent Mevel.

Ce travail est une extension de celui précédemment accompli sur la statistique des processus de Markov partiellement observés (HMM). L'étude comporte plusieurs volets, menés en collaboration avec Lorenzo Finesso (LADSEB/CNR, Padoue, Italie), dans le cadre des réseaux européens TMR SI et IHP DYNSTOCH, voir modules 8.3 et 8.4. On a montré que la vitesse de convergence presque sûre de l'estimateur du maximum de vraisemblance pour les HMM est  $\sqrt{n^{-1} \log \log n}$ . Le résultat s'appuie sur la preuve de logarithme itéré pour la classe de fonctions étudiées par Laurent Mevel dans sa thèse [Mev97], et sur l'extension au cas des HMM du théorème de Shannon–MacMillan sur la convergence uniforme de la log-vraisemblance et de ses dérivées [57]. On a étudié ensuite les HMM comme approximants d'une classe de processus ergodiques, et on a étendu à ce modèle plus général les résultats obtenus pour la statistique des HMM à modèle connu. La principale différence technique se situe sur la prise en compte du caractère non compact du processus ergodique approximé. Une étude sur le comportement du rapport de vraisemblance a été menée d'une part dans le cadre d'un test d'hypothèses simples, et d'autre part dans le cadre d'un test asymptotique local, avec application à la validation de modèles et à la surveillance [46].

---

[Mev97] L. MEVEL, *Statistique Asymptotique pour les Modèles de Markov Cachés*, Thèse de doctorat, Université de Rennes 1, Rennes, novembre 1997, <ftp://ftp.irisa.fr/techreports/theses/1997/mevel.ps.gz>.

En s'appuyant sur les résultats obtenus par Laurent Mevel dans sa thèse <sup>[Mev97]</sup> pour l'identification des HMM, on s'intéresse aussi à l'identification d'une certaine classe de chaînes de Markov non-homogènes, dont les transitions dépendent d'évènements extérieurs, qui représentent l'effet du système global sur le sous-système considéré. L'objectif est de parvenir à estimer les lois de probabilité dans un système partiellement stochastique, voir module 6.3. Ce travail fait l'objet de la thèse de Samy Abbès.

### Stabilité et approximation particulière

**Participants** : François Le Gland, Nadia Oudjane.

Des résultats généraux ont été obtenus dans le travail de thèse de Nadia Oudjane pour la stabilité du filtre optimal par rapport à la condition initiale, et par rapport au modèle. On s'est intéressé à la manière dont l'erreur locale d'approximation du modèle se propageait au cours du temps, pour différentes hypothèses sur le noyau de transition du modèle caché, et pour différents types d'estimation de l'erreur locale d'approximation du modèle (au sens de la distance projective de Hilbert, en variation totale, ou dans un sens faible). Ces estimations de l'erreur globale d'approximation ont été ensuite appliquées à différents types de filtres approchés (filtre particulière pondéré avec oubli exponentiel, — avec ré-échantillonnage, — régularisé, etc.) [47, 9].

### Génération de résidus pour les HMM à état continu

**Participants** : Frédéric Cérou, François Le Gland, Nigel Newton, Bo Wang.

L'objectif est de contribuer à la surveillance des systèmes dynamiques non-linéaires, voir module 6.1, en mettant en œuvre d'autres outils.

La première phase, à laquelle Nigel Newton a participé lors de son séjour dans le projet, a consisté à proposer un algorithme pour le calcul approché de résidus, voir module 3.2. Comme dans le cas des HMM à état fini, une grande classe de fonctions d'estimation peut s'exprimer à l'aide du filtre de prédiction (distribution de probabilité conditionnelle de l'état caché, sachant les observations passées) et de son gradient par rapport au paramètre inconnu. Dans le cas particulier où l'état caché vérifie une équation différentielle stochastique avec un coefficient de dérive dépendant d'un paramètre inconnu, on a montré que le gradient du filtre de prédiction est une mesure signée, absolument continue par rapport au filtre de prédiction, pourvu que le gradient du coefficient de dérive appartienne à l'image du coefficient de diffusion, c'est-à-dire pourvu que le bruit soit non-dégénéré dans les directions où le coefficient de dérive varie lorsque le paramètre varie. Cette observation a permis de proposer un algorithme particulière efficace pour le calcul approché d'une grande classe de fonctions d'estimation [34, 17]. L'approximation proposée pour le gradient du filtre de prédiction est une combinaison linéaire pondérée de masses de Dirac, utilisant le même système de particules que pour l'approximation du filtre de prédiction. En particulier, on évite ainsi toute référence au système gradient.

---

[Mev97] L. MEVEL, *Statistique Asymptotique pour les Modèles de Markov Cachés*, Thèse de doctorat, Université de Rennes 1, Rennes, novembre 1997, <ftp://ftp.irisa.fr/techreports/theses/1997/mevel.ps.gz>.

La seconde phase, qui fait l'objet du projet de recherche post-doctoral de Bo Wang, consiste à étudier le comportement asymptotique du résidu sous l'hypothèse d'un fonctionnement nominal, et sous l'hypothèse alternative d'une petite déviation du paramètre par rapport à la valeur nominale, voir module 3.2. Cette approche asymptotique locale sera mise en œuvre dans le cadre d'une asymptotique de type «petits bruits».

## 6.6 Synthèse de filtres numériques

**Participant** : Jean-Jacques Fuchs.

Les télécommunications numériques ont créé de nouveaux besoins en matière de filtrage numérique. On peut citer notamment les filtres interpolateurs, qui jouent un rôle important dans la synchronisation, et les filtres à complexité variable qui s'adaptent à chaque instant à la puissance de calcul disponible.

**Interpolateur de Lagrange** Pour retarder un signal afin de mieux le décoder, il suffit de savoir reconstruire à la demande le signal en un point quelconque entre deux échantillons, on parle de retard fractionnaire. Si le retard est susceptible de changer, il faut que les paramètres du filtre (les valeurs de la réponse impulsionnelle de longueur finie) soient une fonction simple du retard fractionnaire afin de limiter la taille mémoire requise ou même réaliser de la poursuite en temps réel.

Il existe une infinité de façons de reconstruire exactement un signal à bande étroite suréchantillonné en un point quelconque. Dans la pratique, on va se contenter d'une reconstruction approchée utilisant au mieux les  $2m$  échantillons (une dizaine) du voisinage immédiat et on va limiter à sa partie centrale la fonction de reconstruction exacte de support infini. Il est donc intéressant de savoir synthétiser les fonctions de reconstruction les plus localisées possibles au voisinage de l'origine. Par échantillonnage, on en déduit alors la réponse impulsionnelle finie d'un filtre interpolateur.

On peut montrer assez simplement que la formule de reconstruction de Lagrange devient celle de Shannon quand on prend des points équirépartis et que l'on fait tendre leur nombre vers l'infini. Le sinus cardinal est le «polynôme» qui vaut 1 en zéro et 0 sur tous les autres entiers. Si le signal est suréchantillonné, l'information disponible est redondante et on peut décider de ne pas utiliser tous les échantillons disponibles, d'en garder un nombre (infini) minimal. La fonction de reconstruction est alors à nouveau unique et l'interpolateur de Lagrange permet de la trouver de façon analytique ou numérique. En choisissant judicieusement les échantillons à utiliser, on obtient des fonctions de reconstruction fortement localisées autour de l'origine, et en ne gardant que cette partie centrale, on obtient d'excellents filtres interpolateurs.

**Filtre numérique à complexité adaptative** Dans de nombreuses situations, il est intéressant de pouvoir adapter les performances aux ressources disponibles à chaque instant. Dans le contexte du filtrage numérique, il s'agit alors de développer des filtres à réponse impulsionnelle finie dont les performances s'améliorent avec la longueur : la partie centrale de la réponse complète, quelle que soit sa longueur, doit correspondre à un filtre aux performances acceptables. Les méthodes usuelles de synthèse de filtres donnent des filtres qui ne possèdent pas

cette propriété : quand on augmente l'ordre du filtre, la partie centrale de la réponse impulsionnelle change. En partant d'un développement original du signal carré périodique, nous avons développé des filtres possédant cette propriété.

## 6.7 Traitement de la parole et du langage

*Voir modules 7.5, 7.6, 7.7, et 8.2.*

**Mots clés** : traitement de la parole, vérification du locuteur, indexation sonore, description audio pour l'indexation, reconnaissance de parole, analyse de données textuelles.

**Résumé** : *Les signaux de communication linguistique (parlée et écrite) constituent un domaine d'application de nombreuses techniques de traitement et de modélisation du signal. Dans ce contexte, nous nous intéressons à plusieurs aspects liés au contenu linguistique (analyse de données textuelles, reconnaissance de la parole) et para-linguistique (indexation de gros corpus de textes, indexation sonore, caractérisation et vérification du locuteur) de ces signaux. Nous étendons nos travaux à d'autres types de signaux sonores, notamment les signaux musicaux.*

*Les techniques utilisées reposent essentiellement sur des approches statistiques (analyse des correspondances, mélanges de gaussiennes, modèles HMM), des algorithmes d'alignement dynamique (Viterbi, Smith et Waterman) et des méthodes statistiques de détection de rupture (un des thèmes privilégiés du projet SIGMA2). Nous nous intéressons également aux approches par décomposition sur bases redondantes, de type «matching pursuit». Ces différentes techniques sont adaptées aux spécificités du problème traité, notamment les faibles volumes d'apprentissage ou le besoin de traitements en ligne (méthodes adaptatives).*

*Plusieurs projets, nationaux et européens, apportent des débouchés à ces travaux, notamment l'intégration de nos algorithmes à des prototypes (projets AGIR, DiVAN et PICASSO). Nous participons également à des consortiums de laboratoires qui mettent leurs efforts en commun pour participer à des campagnes d'évaluation nationale et internationale (consortium ELISA et action de recherche coopérative SIROCCO).*

### Vérification du locuteur

**Participants** : Frédéric Bimbot, Raphaël Blouet, Bruno Jacob, Ivan Magrin-Chagnolleau, Mouhamadou Seck.

La vérification du locuteur peut être comprise comme un problème de détection. Il s'agit en effet de décider, à partir d'un échantillon sonore, si celui-ci a été prononcé ou non par un locuteur particulier. L'état-de-l'art dans le domaine repose sur l'utilisation de modèles de Markov cachés (HMM) associés à un test de rapport de vraisemblance. Néanmoins, la robustesse des HMM, des algorithmes d'apprentissage et des stratégies de décision est souvent imparfaite, compte tenu du faible volume de données d'entraînement en regard des nombreuses sources de variabilité (dépendance au texte prononcé, variations du canal de transmission, fluctuations de la voix au cours du temps, etc.).

Dans un contexte applicatif, les méthodes d'apprentissage doivent fonctionner en présence de faibles volumes de données et pouvoir s'accommoder d'un mot de passe choisi par l'utilisateur (projet PICASSO, voir module 7.7, dans la continuité du projet CAVE [14]). Les travaux effectués par Bruno Jacob s'inscrivent dans ce contexte.

Parmi les contributions de l'équipe au sein du projet PICASSO, citons la mise au point de la technique d'alignement synchrone, en collaboration avec l'IDIAP, qui consiste à décoder conjointement la séquence d'état optimale correspondant à un ensemble d'observations dans deux modèles de Markov cachés (en l'occurrence, un modèle du locuteur et un modèle indépendant du locuteur) [52, 53]. Une autre contribution de l'équipe, en collaboration avec l'ENST, concerne les travaux en personnalisation du mot de passe par inférence de typologie et «désadaptation» d'un HMM d'un locuteur donné vers un modèle indépendant du locuteur [43].

On s'intéresse également à la vérification du locuteur indépendamment du texte, qu'elle s'effectue à travers le réseau téléphonique (consortium ELISA pour les évaluations NIST) ou directement sur le terminal (convention de recherche avec Bull, voir module 7.8). Dans ce contexte, on travaille à la fois sur les approches de l'état de l'art (modèles GMM avec apprentissage MAP) et sur des procédés alternatifs (notamment l'utilisation d'arbres de décision) [24, 26, 27, 49, 50, 51].

## Description audio pour l'indexation

**Participants :** Frédéric Bimbot, Bernard Delyon, Ivan Magrin-Chagnollean, Mouhamadou Seck.

L'accroissement permanent de la masse de documents sonores (enregistrements radiophoniques, bandes sonores de films, archives audio, etc.) rend indispensable le développement d'outils de repérage et de navigation dans ce type de données. L'indexation sonore a pour but de représenter de façon plus structurée le matériau audio afin d'en faciliter l'accès par le contenu ou selon des critères de similarité.

Nos travaux portent essentiellement sur la détection de segments sonores homogènes (détection de changement de morceau de musique, suivi de locuteur, etc.) et sur la classification de ces segments (en musique et parole, par exemple). Ils sont l'occasion d'une collaboration étroite avec le projet VISTA, dans le contexte de projets nationaux et européens, AGIR et DiVAN, voir modules 7.5 et 7.6. Nous travaillons également, mais dans une moindre mesure, sur la reconnaissance de la langue parlée [36, 37].

Ces travaux reposent sur la combinaison de techniques de détection de ruptures et de méthodes de classification statistique. Nous utilisons des modèles de mélanges de gaussiennes pour caractériser la distribution statistique des paramètres acoustiques des différentes classes de signaux considérées (parole, musique, applaudissements, locuteurs, langue, etc.). Des critères de décision basés sur le rapport de vraisemblance sont utilisés pour détecter des changements de statistiques (dans le cas de la détection de rupture) et pour décider de la présence d'un type de son dans un document sonore (pour la classification de documents). Parmi nos contributions originales, mentionnons le modèle de mélange classe/non-classe (travaux de thèse de Mouhamadou Seck [10]) et l'utilisation de composantes principales temps-fréquence pour représenter le signal audio (travaux d'Ivan Magrin-Chagnollean [48]).

## Analyse de signaux sonores

**Participants** : Laurent Benaroya, Frédéric Bimbot, Rémi Gribonval, Lorcan McDonagh.

Dans le contexte du multimédia, les besoins en analyse, modélisation et représentation du signal sonore (et plus particulièrement du signal musical) sont l'objet d'une croissance soutenue. Nous nous intéressons à cette thématique depuis maintenant plus d'un an, nos travaux portant d'une part sur la séparation de voies et d'autre part sur la recherche et l'extraction de sons-clés dans des enregistrements audio.

La séparation de voies s'apparente à un problème de séparation de sources, dans le cas sous-déterminé où le nombre de sources excède le nombre de capteurs. L'approche utilisée consiste à caractériser chacune des sources à séparer par une base (redundante) adaptée, puis à utiliser des techniques de projection adaptative (de type «matching pursuit») pour sélectionner la représentation la «mieux» adaptée aux composantes du signal analysé. C'est un des axes privilégiés des travaux de thèse de Laurent Benaroya.

Un second aspect de nos travaux consiste à utiliser les techniques de décomposition de signaux par addition de molécules élémentaires, afin d'extraire des sons clés d'un signal audio. La thèse de Lorcan McDonagh, qui vient de débiter, consiste à généraliser ce modèle en introduisant une modélisation probabiliste des molécules, plus à même de rendre compte de la variabilité d'un même son dans ses diverses réalisations.

Outre des applications en analyse et en compression du son, on étudie la possibilité de réaliser automatiquement des résumés sonores, pour accéder rapidement au contenu d'un enregistrement, tout en préservant des éléments essentiels de sa structure. L'arrivée de Rémi Gribonval, dont les travaux de thèse ont porté sur les représentations redondantes du signal, constitue un renfort de premier plan pour développer l'ensemble de cette thématique.

Mentionnons aussi des travaux avec l'IDIAP sur les modèles de prédiction par treillis à topologie non-uniforme [44].

## Reconnaissance de la parole

**Participants** : Frédéric Bimbot, Guillaume Gravier, Bruno Jacob.

Nos efforts en reconnaissance de parole s'inscrivent dans la participation à l'action de recherche coopérative (ARC) de l'INRIA SIROCCO visant à mettre au point une plate-forme multisite pour la reconnaissance de parole grand-vocabulaire. Notre contribution consiste à mettre en œuvre des techniques de décodage rapide de séquences d'états dans les modèles de Markov cachés («beam-search»).

L'ARC est sur le point d'aboutir, et les algorithmes développés par l'équipe sont en cours d'évaluation. Les temps de calcul rendent possible la multiplication des expériences permettant le réglage fin du système de reconnaissance de parole. Les performances en termes de précision de la reconnaissance de mots sont en cours d'évaluation sur la base de données de la dernière campagne d'évaluation de l'AUF [19].

Le moteur de décodage pourra également être utilisé pour l'indexation par le contenu linguistique (décodage total ou partiel de bandes son).



## Analyse de données textuelles

**Participants :** Annie Morin, Rodolphe Priam.

Il existe actuellement des outils permettant de mesurer la similarité, notamment entre des documents disponibles sur le web. Les questions que l'on se pose sont, entre autres, comment ne pas se perdre en navigant dans ces gisements et comment dégager des informations pertinentes de cette masse de données. Nous nous plaçons délibérément dans un contexte de données textuelles. L'information arrive par flux et peut être soit numérique, soit textuelle. Les données numériques sont structurées ou marquées. Notre hypothèse de travail est qu'il s'agit d'une grande masse de données qui possède une structure minimale constituée par un marqueur de séparation de message, un message étant une unité textuelle autonome. C'est le cas de textes saisis directement sur ordinateur par un opérateur, mais c'est aussi le cas de données résultant d'enquêtes répétitives par sondage, comme celles qui sont organisées pour étudier le comportement et la satisfaction d'utilisateurs de services divers ou encore de documents dans des bases de données documentaires.

Les réponses apportées actuellement dans la littérature au traitement de données textuelles tournent autour de plusieurs sujets à savoir :

- la récupération d'information de provenances et de formats divers,
- l'indexation automatique de gros corpus textuels et leur classement par thèmes,
- l'interrogation et l'utilisation des bases d'information (classement par thèmes),
- la mise à jour des index et la détection de nouveaux thèmes.

Nous travaillons à développer une stratégie pour la classification d'unités textuelles autonomes et pour procéder à leur indexation dynamique. D'une part, il s'agit de décrire les données avec étude de fréquence et constitution de dictionnaire. D'autre part, il s'agit de déterminer le vocabulaire minimal caractérisant un domaine. Ce vocabulaire minimal, que nous appellerons les sélecteurs, servira a posteriori à coder des documents. Plusieurs méthodes sont utilisées.

- La première méthode est basée essentiellement sur l'analyse factorielle des correspondances (AFC) et sur la classification ascendante hiérarchique (CAH). Dans une première étape, on fait le dictionnaire des mots, avec calcul de la fréquence des mots et fréquence de leur occurrence. Un tableau de correspondances entre les 1000 premiers mots classés par ordre décroissant de fréquence et les documents (il peut s'agir de messages sur le web ou de résumés d'articles) de l'échantillon est constitué. Ce tableau, dont on élimine les documents ayant moins d'un certain nombre d'occurrences de mots, est soumis à une AFC, puis à une CAH sur les facteurs, afin de segmenter l'échantillon en classes relativement homogènes. L'analyse factorielle systématique de ces classes permet de récupérer, pour chaque facteur, deux groupes de mots ayant une contribution à la valeur propre du facteur supérieure à 2 fois la valeur moyenne des contributions. Ces mots sont appelés des «méta clés» quand on a pour objectif l'indexation des textes.
- Nous utilisons aussi le «latent semantic indexing» (LSI), méthode concurrente de l'AFC très utilisée aux États-Unis. Cette technique voisine de l'AFC est basée sur la décomposition en valeurs singulières de la matrice de fréquences croisant les textes à indexer et les mots.

- Une troisième méthode basée sur l'utilisation des cartes auto-organisatrices (SOM) de Kohonen paraît très prometteuse en termes de fiabilité de résultats (qualité de l'indexation) et de rapidité des calculs, et permet donc de traiter de très grands fichiers de données textuelles.

Actuellement, dans le cadre de la thèse de Rodolphe Priam, nous travaillons sur les deux dernières méthodes LSI et SOM. Nous comparons ces méthodes à l'AFC et aux autres méthodes plus classiques d'analyse des données tout en développant les programmes correspondants. Par ailleurs, nous travaillons en étroite collaboration avec Michel Kerbaol et Jean-Yves Bansard du Laboratoire de Santé Publique de l'université de Rennes 1 pour le développement du logiciel et nous participons à un programme de coopération avec le Portugal autour du développement de méthodes pour la fouille de données numériques et textuelles.

## 7 Contrats industriels (nationaux, européens et internationaux)

### 7.1 Analyse de structures vibrantes — Projet Eurêka SINOPSYS

*Voir modules 4.1 et 6.2. Contrat INRIA 2 96 C 295 00 31301 00 6 — novembre 1996/février 2000.*

**Participants :** Laurent Mevel, Albert Benveniste, Michèle Basseville.

Le projet SIGMA2 a participé, en collaboration avec le projet METALAU (Maurice Goursat), au projet Eurêka SINOPSYS sur l'analyse de structures vibrantes en ambiance de travail. Outre l'INRIA, sont membres de ce projet LMS (maître d'œuvre), ISMC et KUL (Belgique), Sopemea et Centrale Recherche (France), VUM (Royaume-Uni), université de Cracovie et PZL-Swidnik (Pologne), Saab-Scania (Suède).

Ce projet s'est achevé en février 2000. Les travaux sur l'identification et la surveillance modale des structures, en ambiance de travail et sous excitation naturelle non observée, se sont traduits par un transfert à la société belge LMS (un des deux leaders mondiaux du domaine) d'une version pré-alpha d'un prototype industriel développé par Laurent Mevel chez LMS, ainsi que par une boîte à outils pour Scilab développée par Maurice Goursat [42]. Le module d'analyse de structure en opération a été en particulier testé sur les données du vol d'essai Ariane 501, et les résultats en seront présentés à IMAC'2001.

Une version dite «polyréférence», pour la méthode d'identification par sous-espace en ambiance de travail, a été développée [29, 30, 65], et l'expérimentation sur données réelles collectées sur un pont sera présentée à IMAC'2001, voir module 6.2.

Un objectif important de notre coopération avec LMS est le diagnostic dans le domaine du génie civil. C'est toutefois dans le domaine de l'aéronautique que nous avons pu construire une nouvelle coopération, en participant activement à la préparation de la proposition Eurêka FLITE (Flight Test Easy), avec le laboratoire d'essais français Sopemea, leader du projet, les constructeurs Dassault-Aviation et AeroMatra Airbus (France), LMS et KUL (Belgique), université de Cracovie et le constructeur PZL-Swidnik (Pologne).

Cette proposition a pour objet de permettre l'utilisation des données de vol sous excitation naturelle (et non plus seulement avec les ailerons), pour les essais en vol en aéronautique. Elle est fortement soutenue par les trois pays proposant et devrait être approuvée en janvier 2001, pour démarrage immédiat. En cas de succès, nous avons en ligne de mire l'application à l'A3XX.

Par ailleurs et toujours dans le domaine de l'aéronautique, Maurice Goursat poursuit une série de contrats pour l'identification en ambiance de travail du lanceur Ariane 5.

## 7.2 Diagnostic de pannes dans les réseaux de télécommunications — Projet exploratoire RNRT MAGDA

*Voir modules 3.3 et 6.3. Contrat INRIA 2 98 C 561 00 MPR 01 1 — novembre 1998/août 2001*

**Participants** : Éric Fabre, Albert Benveniste, Mark Smith, Vincent Pigourier.

Cette activité est partagée avec le projet PAMPA (Claude Jard). Elle se situait initialement dans le cadre de la CTI-CNET 95 1B 151, et se prolonge par le projet exploratoire RNRT MAGDA «Modélisation et Apprentissage pour une Gestion Distribuée des Alarmes» labellisé en septembre 1998, qui se propose de tester «en grand» cette approche sur un cadre plus industriel. Outre les projets PAMPA et SIGMA2, et également le projet AIDA de l'IRISA, les participants de ce projet sont : France Télécom R&D (maître d'œuvre), Alcatel, ILOG, et l'université de Paris-Nord. Les deux industriels utilisateurs doivent apporter un cas d'étude de réseau, Alcatel apporte en outre une plate-forme de gestion qui inclut déjà des outils de ILOG (sur base de ILOG-rules).

Il s'agit de développer une approche systématique pour le diagnostic de pannes dans les réseaux de télécommunications, avec les objectifs suivants :

- prendre en compte explicitement le caractère distribué des réseaux,
- suivre une approche «modèle», modèle dont découlera automatiquement l'algorithme de diagnostic,
- prendre en compte les aléas (perte d'alarmes, confusions possibles, etc.),
- viser une mise en œuvre du logiciel de diagnostic qui soit répartie sur le réseau.

Une technologie originale, fondée sur une notion nouvelle de réseau de Petri stochastique ou plus généralement de réseau d'automates stochastiques, conforme au point de vue dit de la «concurrence vraie», a été développée, voir module 6.3. Cette technologie nous permet naturellement de prendre en compte toutes les contraintes qui résultent du contexte distribué où nous nous situons.

## 7.3 Mesure de la qualité de service en réseau WDM optique transparent — Contrat Alcatel

*Contrat INRIA 1 00 C 0294 00 31321 01 2 — avril 2000/octobre 2000.*

**Participants** : Frédéric Céro, Albert Benveniste, Bernard Delyon, George Moustakides, Qinghua Zhang.

Il s'agit d'une activité nouvelle, qui relève de l'accord-cadre entre Alcatel et l'INRIA. Pour mesurer le taux d'erreur et évaluer ainsi la qualité de service en un point d'un réseau optique, les méthodes actuelles ont recours à la couche électrique (SDH ou SONET) ce qui limite le développement du concept de réseau optique transparent («all optical network»). Nous avons

développé une technique permettant d'estimer ce taux d'erreur à partir d'un échantillon asynchrone du signal, en n'ayant accès qu'à la couche analogique optique. La seule donnée nécessaire est une connaissance, tolérant une certaine incertitude, du débit. Le temps bit est estimé de façon précise, puis la synchronisation est effectuée de façon logicielle, par construction d'un diagramme de l'œil à partir de l'échantillon. Enfin, une technique d'estimation de mélanges de gaussiennes permet d'estimer le taux d'erreur. On a constaté un très bon accord avec les résultats de simulation du logiciel OCEAN développé par Alcatel à Marcoussis, et qui calcule le taux d'erreur en utilisant toute la connaissance du message transmis [69].

#### 7.4 Identification du comportement dynamique d'un véhicule routier — Contrat Renault

*Contrat INRIA 1 99 C 271 00 31321 01 2 — mars 1999/janvier 2002.*

**Participants** : Arnaud Clavel, Qinghua Zhang, Bernard Delyon.

Les véhicules routiers, automobiles et poids-lourds, présentent des comportements parfois instables en fonction notamment de la vitesse, de l'adhérence de la route, ou de leur chargement. Afin de corriger le comportement dynamique de ces véhicules et d'éviter des accidents, de nouveaux systèmes électroniques sont actuellement à l'étude chez Renault et chez ses équipementiers. Dans le cadre d'une convention Cifre, cette étude a pour objectif de concevoir et de mettre en œuvre des méthodes d'identification hors-ligne, puis en-ligne, répondant aux besoins d'estimation de certains paramètres essentiels pour les lois de commande d'un véhicule, tels que l'adhérence, la masse du véhicule, la hauteur du centre de gravité, ainsi que certains paramètres concernant le comportement latéral des pneumatiques. Cette année, nous avons développé un modèle de suspension avec des ressorts à lame dont la non-linéarité principale est causée par le frottement à sec. La pertinence du modèle et l'efficacité de la méthode d'identification associée ont été démontrées par des expérimentations sur un banc d'essai et sur un camion [35].

#### 7.5 Indexation sonore — Projet pré-compétitif RNRT AGIR

*Voir module 6.7. Contrat INRIA 2 99 C 006 00 00 MPR 01 1 — septembre 1999/février 2001.*

**Participants** : Frédéric Bimbot, Bernard Delyon, Ivan Magrin-Chagnolleau.

Le projet SIGMA2 participe, en collaboration avec les projets VISTA et MOVI, au projet pré-compétitif RNRT AGIR, sur le thème de l'indexation sonore. La contribution de l'équipe consiste à mettre au point des outils de description du contenu sonore pour l'annotation automatique et l'indexation, basés sur des approches statistiques. Outre l'INRIA, les partenaires du projet sont : INA (maître d'œuvre), Arts Video Interactive, CS Systèmes d'Information, Mémodata, INT, IRIT, LIP6, UJF.

#### 7.6 Gestion d'archives audio-visuelles — Projet Esprit DIVAN

*Voir module 6.7. Contrat INRIA 1 97 C 696 00 00 MPR 00 5 — septembre 1997/juin 2000.*

**Participants** : Frédéric Bimbot, Bernard Delyon, Mouhamadou Seck.

Le projet SIGMA2 a participé, en collaboration avec le projet VISTA, au projet Esprit DIVAN «Distributed Audio–Visual Archives Network». L’objectif de ce projet était la mise au point d’un prototype de système de gestion d’archives audio–visuelles. Notre contribution a consisté à développer des traitements permettant la détection et la localisation de classes ou d’évènements sonores dans un enregistrement audio, notamment la détection de présence de parole ou de musique et la détection de jingles. Le projet était coordonné par Intracom (Grèce). Outre l’INRIA, les partenaires du projet étaient : Tecmath et THD–ITO (Allemagne), INA (France), RAI (Italie), ERT et ICS–Forth (Grèce),

## 7.7 Vérification du locuteur — Projet Telematics PICASSO

*Voir module 6.7. Contrat INRIA 1 98 C 457 00 31321 — janvier 1998/juin 2000*

**Participants** : Frédéric Bimbot, Bruno Jacob.

Le projet SIGMA2 a participé au projet Telematics PICASSO sur la vérification du locuteur à travers le réseau téléphonique, principalement en mode dépendant du texte. Nos contributions scientifiques ont plus précisément porté sur les techniques d’apprentissage, de décodage et de calcul de score pour la vérification du locuteur par modèles de Markov cachés (inférence de typologie et de paramètres, alignement synchrone, normalisation par état, etc.). Outre l’INRIA, les partenaires du projet étaient : KPN Telecom (maître d’œuvre), KPN Research et KUN (Pays–Bas), Fortis (Pays–Bas et Belgique), KTH (Suède), NBS et Vocalis (Royaume–Uni), ENST (France), Swisscom, IDIAP et UBS Ubilab (Suisse).

## 7.8 Reconnaissance du locuteur — Contrat Bull

*Voir module 6.7. Contrat INRIA 1 99 C 138 00 31321 01 2 — décembre 1998/novembre 2001.*

**Participants** : Frédéric Bimbot, Raphaël Blouet.

Une convention de recherche a été signée avec Bull pour développer et perfectionner, dans le cadre d’une thèse, des algorithmes et des procédés de vérification du locuteur dans le cas d’architectures de traitement distribuées, certains traitements pouvant être réalisés sur le terminal, d’autres étant centralisés. Ceci introduit des contraintes spécifiques de rapidité de traitement et de modularité.

## 7.9 Compression de parole — Contrat Innovaction

*Voir module 6.7. Contrat INRIA 1 99 C 1781 00 31321 01 2 — janvier 2000/décembre 2000.*

**Participants** : Laurent Benaroya, Frédéric Bimbot.

Etude bibliographique et présentation didactique d’approches en codage de parole. Le but de cette étude était d’aider une start–up à définir son positionnement et fixer ses objectifs, en terme de faisabilité technologique et de créneau applicatif à occuper.

## 8 Actions régionales, nationales et internationales

### 8.1 Système générique de reconnaissance de parole — Action de recherche coopérative de l'INRIA SIROCCO

*Voir module 6.7. ARC INRIA 39007 — février 2000/janvier 2001.*

**Participants** : Guillaume Gravier, Bruno Jacob, Frédéric Bimbot.

Commencé fin 1997 à l'initiative de plusieurs laboratoires, le projet SIROCCO a pour objectif de mettre en place une plate-forme commune de reconnaissance de parole grand-vocabulaire. Ce projet, qui regroupe ENST, IRISA, IRIT, LIA et LORIA est devenu en 1999 une action de recherche coopérative de l'INRIA. Notre équipe est plus particulièrement concernée par le module de décodage rapide de parole continue par recherche en faisceaux («beam-search»).

### 8.2 Consortium ELISA

*Voir module 6.7. Consortium sans financement.*

**Participants** : Frédéric Bimbot, Raphaël Blouet, Guillaume Gravier, Ivan Magrin-Chagnolleau, Mouhamadou Seck.

Le consortium ELISA a été formé en 1998 par plusieurs laboratoires (ENST, EPFL, IDIAP, IRISA, LIA), pour partager les efforts nécessaires à la participation aux campagnes d'évaluation NIST en vérification du locuteur. Pour la troisième année consécutive, ELISA a participé à ces évaluations en 2000. Ce projet ne bénéficie actuellement d'aucun soutien financier extérieur.

### 8.3 Identification des systèmes — Réseau TMR SI

*Contrat université de Rennes 1 MW 33 — mars 1998/février 2002.*

**Participants** : Michèle Basseville, Albert Benveniste, Jean-Jacques Fuchs, François Le Gland, Laurent Mevel, Qinghua Zhang.

Le projet SIGMA2 participe au réseau européen SI «System Identification» qui regroupe neuf équipes de recherche européennes : CWI (coordinateur, Pays-Bas), Technische Universität Wien (Autriche), université catholique de Louvain (Belgique), INRIA Sophia-Antipolis et IRISA/université de Rennes 1 (France), University of Cambridge (Royaume-Uni), LADSEB/CNR et Università degli Studi di Padova (Italie), KTH et Linköpings Universitet (Suède), dans le cadre du programme TMR. Le séminaire annuel s'est tenu à Linköping du 25 au 27 septembre. Notre contribution concerne l'identification non-paramétrique, l'identification des HMM, et l'identification pour la surveillance.

### 8.4 Méthodes statistiques pour les systèmes dynamiques stochastiques — Réseau IHP DYNSTOCH

*Voir module 6.5. Rattachement à l'équipe de Paris VI — septembre 2000/août 2004.*

**Participants :** Frédéric Céro, François Le Gland, Laurent Mevel.

Nous participons également au réseau européen DYNSTOCH «Statistical Methods for Dynamical Stochastic Models», qui regroupe neuf équipes de recherche européennes : Københavns Universitet (coordinateur, Danemark), Universiteit van Amsterdam (Pays-Bas), Humboldt Universität zu Berlin et Albert Ludwigs Universität Freiburg (Allemagne), Universidad Politécnica de Cartagena (Espagne), Helsingin Yliopisto (Finlande), University College London (Royaume-Uni), LADSEB/CNR (Italie), université de Paris VI (France), dans le cadre du programme IHP. Le premier séminaire annuel s'est tenu à Padoue du 21 au 23 septembre. Notre contribution au sein de l'équipe française concerne la statistique asymptotique des HMM (à espace d'état fini ou continu).

## 8.5 Autres actions internationales

M. Basseville est membre du groupe de travail WG2 «Health Monitoring and Damage Detection», du COST F3 «Structural Dynamics».

## 8.6 Visites, et invitations de chercheurs

Nigel Newton (université d'Essex) a séjourné trois mois dans le projet, et a travaillé avec Frédéric Céro et François Le Gland sur la génération de résidus pour les systèmes dynamiques non-linéaires, voir module 6.5.

George Moustakides (université de Patras) effectue un séjour de six mois dans le projet, et travaille avec Frédéric Céro sur la surveillance et le diagnostic en réseau WDM optique transparent, voir module 7.3.

Johnny Mariéthoz (IDIAP) a séjourné pendant un mois dans le projet, pour travailler sur la personnalisation et l'apprentissage incrémental du mot de passe du locuteur, dans le cadre du projet Telematics PICASSO, voir module 7.7.

# 9 Diffusion de résultats

## 9.1 Animation de la communauté scientifique

M. Basseville est membre du comité de direction du GDR-PRC ISIS (Information, Signal, Images). Elle est membre des comités techniques IFAC IAF-TC «Fault Detection, Supervision and Safety of Technical Processes», relevant du comité de coordination IA-CC «Industrial Applications», et SSM-TC «Modelling, Identification and Signal Processing», relevant du comité de coordination SSM-CC «Systems and Signals». Elle participe au GT «Aide au Diagnostic» du club CRIN ITI (Ingénierie du Traitement de l'Information).

F. Bimbot est président du GFCEP (Groupe Francophone en Communication Parlée), groupe spécialisé de la SFA (Société Française d'Acoustique) et de l'ISCA (International Speech Communication Association). Il est également membre du bureau de l'ISCA, chargé de la promotion et des relations avec les membres. Il est aussi membre du groupe spécialisé de l'ISCA sur la caractérisation du locuteur et de la langue.

A. Morin est responsable du groupe enseignement de la statistique de la SFds (Société Francophone de Statistique).

M. Basseville est éditeur associé du journal IFAC «Automatica». Elle a été éditeur invité, en collaboration avec Thomas Parisini (Politecnico di Milano), d'un numéro spécial de IJACSP (International Journal of Adaptive Control and Signal Processing) intitulé : «Fault Detection and Isolation», et paru en novembre 2000, et en collaboration avec Étienne Balmès (École Centrale Paris), d'une section spéciale d'un numéro de MSSP (Mechanical Systems and Signal Processing), intitulée : «Model-based Structural Identification and Monitoring Using in-Operation Data», et parue en novembre 1999.

A. Benveniste est membre du comité éditorial des revues «European Journal of Control», «Discrete Event Dynamic Systems» et «Proceedings of the IEEE».

M. Basseville est membre des comités internationaux de programme de DAMAS'2001 (Damage Assessment in Structures), SYSID'2000, et SAFEPROCESS'2000.

J.-J. Fuchs est membre du comité technique IEEE SAM «Sensor Array and Multichannel» qui a organisé le workshop ASAP 2000 et SAM 2000 du 14 au 17 mars 2000 à Cambridge, MA.

F. Bimbot a été président du comité scientifique des XXIII<sup>es</sup> Journées d'Etude sur la Parole (Aussois, 2000). Il est membre du comité scientifique d'Eurospeech 2001-Scandinavie (Aalborg, 2001) et membre du comité d'organisation du workshop «2001- A speaker Odyssey» (Jérusalem, 2001).

A. Morin est membre du comité scientifique de la conférence ITI (Information Technology Interfaces) qui a lieu à Pula, Croatie, en juin 2001. Elle a organisé une école Modulad sur le thème «Bases de données et statistique» du 17 au 18 novembre 1999 au Croisic avec le parrainage de l'INRIA et de la SFds.

M. Basseville a été sollicitée pour l'évaluation d'un Senior Individual Grant de la SFSR (Swedish Foundation for Strategic Research).

A. Benveniste est régulièrement invité pour l'évaluation des projets TFR, NUTEK, et SSF, agences de moyens dans le domaine des STIC en Suède : il a en particulier participé, conjointement avec John Proakis et Michel Gevers, à l'évaluation du programme «Signal and Systems» du TFR en 2000 (équivalent suédois de la NSF).

## 9.2 Enseignement

Le projet intervient de façon importante dans le DEA-STIR (Signal, Télécommunications, Images, Radar) de l'école doctorale MATHISSE, université de Rennes 1 : J.-J. Fuchs y enseigne l'«Optimisation» et l'«Estimation spectrale». F. Le Gland y enseigne le «Filtrage de Kalman et les chaînes de Markov cachées».

M. Basseville enseigne les «Méthodes statistiques pour la surveillance en fonctionnement» dans le cadre du module «Outils d'aide au diagnostic», de l'option «Automatique et Informatique Industrielle», de la dernière année de l'École des Mines de Nantes.

F. Bimbot enseigne le «Traitement de la parole» au DEA-TIS de l'université de Cergy-Pontoise, ainsi qu'à l'ESIEA de Laval, à l'EISTI, et à l'INT.

É. Fabre et Q. Zhang participent aux enseignements d'«Optimisation» au DIIC de l'IFSIC, université de Rennes 1.



É. Fabre enseigne la «Théorie de l'information et les codes correcteurs d'erreurs» à l'ESIEA de Laval.

Q. Zhang enseigne la «Modélisation boîte-noire» à l'Institut National d'Horticulture d'Angers.

A. Morin enseigne dans différentes formations de 3ème cycle de l'université de Rennes 1 : l'«Analyse des données» en DESS de mathématiques, la «Maîtrise statistique des processus» en DESS composants micro-électroniques, les «Plans d'expérience et la fiabilité» en DEA d'électronique.

### 9.3 Participation à des colloques, séminaires, cours, etc.

Outre les communications faisant l'objet d'une publication dans les actes, et qui sont listées en fin de document, les membres de l'équipe ont effectué les présentations suivantes.

Lors de deux réunions du COST F3 «Structural Dynamics», M. Basseville a présenté des travaux en commun avec L. Mevel, A. Benveniste et M. Goursat : «On-board vibration monitoring — Application to the Z24 bridge», à Liège en janvier, et «Polyreference output-only subspace-based modal analysis under nonstationary excitation», à Kassel en novembre, et a donné un exposé «Détection et diagnostic de petites déviations dans les signaux et systèmes dynamiques», à la journée IFREMER/IRISA, à Brest en mars.

F. Bimbot a été conférencier invité aux Journées Thématiques 2000 du GDR-PRC ISIS, consacrées au thème : «Signal et Image au service de la Sécurité dans la Société de l'Information».

É. Fabre a donné un exposé «Inference on dynamic Bayesian networks», au workshop «Partial Knowledge and Uncertainty», à Rome en mai.

L. Mevel a donné un exposé «Rate of almost sure convergence for estimators in hidden Markov chains», au premier workshop du réseau IHP «Statistical Methods for Dynamical Stochastic Models» qui a eu lieu en septembre à Padoue, voir module 8.4.

## 10 Bibliographie

### Ouvrages et articles de référence de l'équipe

- [1] A. AGHASARYAN, E. FABRE, A. BENVENISTE, R. BOUBOUR, C. JARD, «Fault Detection and Diagnosis in Distributed Systems: An Approach by Partially Stochastic Petri Nets», *Journal of Discrete Events Dynamical Systems* 8, 2 (Special issue on Hybrid Systems), juin 1998, p. 203–231.
- [2] M. BASSEVILLE, I. V. NIKIFOROV, *Detection of Abrupt Changes — Theory and Applications*, *Information and System Sciences Series*, Prentice Hall, Englewood Cliffs, 1993, <http://www.irisa.fr/sigma2/kniga/>.
- [3] M. BASSEVILLE, «On-Board Component Fault Detection and Isolation Using the Statistical Local Approach», *Automatica* 34, 11, novembre 1998, p. 1391–1416.
- [4] A. BENVENISTE, B. C. LÉVY, E. FABRE, P. LE GUERNIC, «A Calculus of Stochastic Systems: Specification, Simulation, and Hidden State Estimation», *Theoretical Computer Science* 152, 2, 1995, p. 171–217.

- [5] A. BENVENISTE, M. MÉTIVIER, P. PRIOURET, *Adaptive Algorithms and Stochastic Approximations, Applications of Mathematics, 22*, Springer Verlag, New York, 1990.
- [6] A. JUDITSKY, H. HJALMÄRSSON, A. BENVENISTE, B. DELYON, L. LJUNG, J. SJÖBERG, Q. ZHANG, «Non-Linear Black-Box Modelling in System Identification: Mathematical Foundations», *Automatica* 31, 12, décembre 1995, p. 1725–1750.
- [7] J. SJÖBERG, Q. ZHANG, L. LJUNG, A. BENVENISTE, B. DELYON, P.-Y. GLORENNEC, H. HJALMÄRSSON, A. JUDITSKY, «Non-Linear Black-Box Modelling in System Identification: A Unified Overview», *Automatica* 31, 12, décembre 1995, p. 1691–1724.
- [8] Q. ZHANG, M. BASSEVILLE, A. BENVENISTE, «Early Warning of Slight Changes in Systems», *Automatica* 30, 1 (Special issue on Statistical Methods in Signal Processing and Control), janvier 1994, p. 95–113.

### Thèses et habilitations à diriger des recherches

- [9] N. OUDJANE, *Stabilité et Approximation Particulaire en Filtrage Non-Linéaire. Application en Pistage*, Thèse de doctorat, université de Rennes 1, décembre 2000.
- [10] M. SECK, *Détection de Ruptures et Suivi de Classes de Sons pour l'Indexation Sonore*, Thèse de doctorat, université de Rennes 1, janvier 2001.
- [11] L. TROMP, *Surveillance et Diagnostic de Systèmes Industriels Complexes : une Approche Hybride Numérique / Symbolique*, Thèse de doctorat, université de Rennes 1, janvier 2000, <ftp://ftp.irisa.fr/techreports/theses/2000/tromp.ps.gz>.

### Articles et chapitres de livre

- [12] M. BASSEVILLE, M. ABDELGHANI, A. BENVENISTE, «Subspace-Based Fault Detection Algorithms for Vibration Monitoring», *Automatica* 36, 1, janvier 2000, p. 101–109.
- [13] A. BENVENISTE, F. LE GLAND, E. FABRE, S. HAAR, «Distributed Hidden Markov Models», in: *Optimal Control and PDE's — Innovations and Applications. In honor of A. Bensoussan on the occasion of his 60th birthday*, J.-L. Menaldi, E. Rofman, et A. Sulem (éditeurs), IOS Press, Amsterdam, 2001, p. 211–220.
- [14] F. BIMBOT, M. BLOMBERG, L. BOVES, D. GENOUD, H.-P. HUTTER, C. JABOULET, J. KOOLWAALJ, J. LINDBERG, J.-B. PIERROT, «An Overview of the CAVE Project Research Activities in Speaker Verification», *Speech Communication* 31, 2–3, juin 2000, p. 155–180.
- [15] L. J. BOË, F. BIMBOT, J.-F. BONASTRE, P. DUPONT, «De l'Évaluation des Systèmes de Vérification du Locuteur à la Mise en Cause des Expertises Vocales en Identification Juridique», *Cahiers / Langues* 2, 4, décembre 1999, p. 270–288.
- [16] F. CAMPILLO, Y. A. KUTOYANTS, F. LE GLAND, «Small Noise Asymptotics of the GLR Test for Off-Line Change Detection in Misspecified Diffusion Processes», *Stochastics and Stochastics Reports* 70, 1–2, 2000, p. 109–129.
- [17] F. CÉROU, F. LE GLAND, N. J. NEWTON, «Stochastic Particle Methods for Linear Tangent Filtering Equations», in: *Optimal Control and PDE's — Innovations and Applications. In honor of A. Bensoussan on the occasion of his 60th birthday*, J.-L. Menaldi, E. Rofman, et A. Sulem (éditeurs), IOS Press, Amsterdam, 2001, p. 231–240.

- [18] F. CÉROU, « Long Time Behaviour for some Dynamical Noise-Free Nonlinear Filtering Problems », *SIAM Journal on Control and Optimization* 38, 4, 2000, p. 1086–1101.
- [19] J.-M. DOLMAZON, F. BIMBOT, G. ADDA, J.-C. CAËROU, J. ZEILIGER, M. ADDA-DECKER, « Première Campagne AUPELF d'Évaluation des Systèmes de Dictée Vocale : Organisation et Résultats », in : *Ressources et Évaluation en Ingénierie des Langues*, K. Chibout, J. Mariani, N. Masson, et F. Néel (éditeurs), *Champs Linguistiques*, De Boeck université, Louvain-la-Neuve, 2000.
- [20] J.-J. FUCHS, B. DELYON, « Why the  $\ell_1$  Norm Leads to Sparseness and High Resolution in Time Delay Estimation », *IEEE Transactions on Information Theory IT-46*, 4, juillet 2000, p. 1666–1673.
- [21] A. JUDITSKY, B. DELYON, « On Minimax Prediction of Nonparametric Autoregressive Models », *Probability Theory and Related Fields* 116, 1, 2000, p. 21–39.
- [22] F. LE GLAND, L. MEVEL, « Basic Properties of the Projective Product, with Application to Products of Column-Allowable Nonnegative Matrices », *Mathematics of Control, Signals and Systems* 13, 1, 2000, p. 41–62.
- [23] F. LE GLAND, L. MEVEL, « Exponential Forgetting and Geometric Ergodicity in Hidden Markov Models », *Mathematics of Control, Signals and Systems* 13, 1, 2000, p. 63–93.
- [24] I. MAGRIN-CHAGNOLLEAU, G. DUROU, « Application of Time-Frequency Principal Component Analysis to Speaker Verification », *Digital Signal Processing* 10, 1–3 (Special issue on the NIST'99 Evaluation Workshop), janvier/avril/juillet 2000, p. 226–236.
- [25] C. MUSSO, N. OUDJANE, F. LE GLAND, « Improving Regularized Particle Filters », in : *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. de Freitas, et N. Gordon (éditeurs), *Statistics for Engineering and Information Science*, Springer Verlag, New York, 2001, ch. 12.
- [26] M. SECK, R. BLOUET, F. BIMBOT, « The IRISA / ELISA Speaker Detection and Tracking System for the NIST'99 Evaluation Campaign », *Digital Signal Processing* 10, 1–3 (Special issue on the NIST'99 Evaluation Workshop), janvier/avril/juillet 2000, p. 154–171.
- [27] The ELISA Consortium, « The ELISA Systems for the NIST'99 Evaluation in Speaker Detection and Tracking », *Digital Signal Processing* 10, 1–3 (Special issue on the NIST'99 Evaluation Workshop), janvier/avril/juillet 2000, p. 143–153.
- [28] Q. ZHANG, « A New Residual Generation and Evaluation Method for Detection and Isolation of Faults in Nonlinear Systems », *International Journal of Adaptive Control and Signal Processing* 14, 7 (Special issue on Fault Detection and Isolation), novembre 2000, p. 759–773.

### Communications à des congrès, colloques, etc.

- [29] A. BENVENISTE, M. BASSEVILLE, M. GOURSAT, L. MEVEL, L. HERMANS, H. VAN DER AUWEREAER, « Polyreference Version of Subspace Method for Structural Identification », in : *Proceedings of the 18th International Modal Analysis Conference (IMAC-XVIII)*, San Antonio, SEM, Inc., février 2000.
- [30] A. BENVENISTE, M. BASSEVILLE, M. GOURSAT, L. MEVEL, « Subspace Algorithms for Data from Varying Sensor Locations », in : *Proceedings of the 12th Symposium on System Identification and System Parameter Identification (SYSID)*, Santa Barbara, IFAC / IFORS, juin 2000. Paper WeAM3–6.

- 
- [31] A. BENVENISTE, M. BASSEVILLE, L. MEVEL, «Convergence Rates for Eigenstructure Identification Using Subspace Methods», *in: Proceedings of the 39th Conference on Decision and Control (CDC), Sydney, IEEE-CSS*, p. 1550–1554, décembre 2000.
- [32] A. BENVENISTE, B. DELYON, «Frequency Domain Local Tests for Change Detection», *in: Proceedings of the 12th Symposium on System Identification and System Parameter Identification (SYSID), Santa Barbara, IFAC / IFORS*, juin 2000. Paper ThAM4–2.
- [33] A. BENVENISTE, B. DELYON, «Using Local Tests to Estimate Convergence Rates for Identification», *in: Proceedings of the 39th Conference on Decision and Control (CDC), Sydney, IEEE-CSS*, p. 1985–1990, décembre 2000.
- [34] F. CÉROU, F. LE GLAND, «Efficient Particle Filters for Residual Generation in Partially Observed SDE's», *in: Proceedings of the 39th Conference on Decision and Control (CDC), Sydney, IEEE-CSS*, p. 1200–1205, décembre 2000.
- [35] A. CLAVEL, M. SORINE, «Modelling and Identification of a Leaf Spring System», *in: Proceedings of the 3rd IFAC Workshop on Advances in Automotive Control, Karlsruhe, IFAC*, mars 2001.
- [36] M. DUTAT, I. MAGRIN-CHAGNOLLEAU, F. BIMBOT, «Analyse en Composantes Principales Temps-Fréquence: Application à la Reconnaissance de la Langue», *in: Actes des XXIIIèmes Journées d'Etudes sur la Parole (JEP), Aussois*, p. 49–52, juin 2000.
- [37] M. DUTAT, I. MAGRIN-CHAGNOLLEAU, F. BIMBOT, «Language Recognition Using Time-Frequency Principal Component Analysis and Acoustic Modeling», *in: Proceedings of the 6th International Conference on Spoken Language Processing (ICSLP), Beijing*, octobre 2000.
- [38] E. FABRE, A. BENVENISTE, C. JARD, L. S. RICKER, M. SMITH, «Distributed State Reconstruction for Discrete Event Systems», *in: Proceedings of the 39th Conference on Decision and Control (CDC), Sydney, IEEE-CSS*, p. 2252–2257, décembre 2000.
- [39] E. FABRE, A. GUYADER, «Dealing with Short Cycles in Graphical Codes», *in: Proceedings of the IEEE International Symposium on Information Theory (ISIT), Sorrento, IEEE-ITS*, juin 2000. Résumé.
- [40] J.-J. FUCHS, «The Global Matched Filter: Application to DOA Estimation with a Uniform Circular Array», *in: Proceedings of the Sensor Array and Multichannel Signal Processing Workshop (SAM'2000), Cambridge, MA, IEEE-SPS*, mars 2000.
- [41] J.-J. FUCHS, «On the Application of the Global Matched Filter to DOA Estimation with Uniform Circular Arrays», *in: Proceedings of the 25th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Istanbul, Volume V, IEEE-SPS*, p. 3089–3092, juin 2000.
- [42] M. GOURSAT, M. BASSEVILLE, A. BENVENISTE, L. MEVEL, «A SCILAB Toolbox for Output Only Modal Analysis and Diagnosis», *in: Proceedings of the 18th International Modal Analysis Conference (IMAC-XVIII), San Antonio, SEM, Inc.*, février 2000.
- [43] B. JACOB, J. MARIÉTHOZ, G. GRAVIER, F. BIMBOT, «Robustesse de la Vérification du Locuteur par Mot de Passe Personnalisé», *in: Actes des XXIIIèmes Journées d'Etudes sur la Parole (JEP), Aussois*, p. 357–360, juin 2000.
- [44] S. KRSTULOVIĆ, F. BIMBOT, «Inverse Lattice Filtering of Speech with Adapted Non-Uniform Delays», *in: Proceedings of the 6th International Conference on Spoken Language Processing (ICSLP), Beijing*, octobre 2000.

- 
- [45] Y. A. KUTOYANTS, F. LE GLAND, R. RAKOTOZAFY, «Identification d'un Système Non-Linéaire Partiellement Observé par la Méthode de la Distance Minimale», *in: Actes du 5ème Colloque Africain sur la Recherche en Informatique (CARI), Antananarivo*, octobre 2000.
- [46] F. LE GLAND, L. MEVEL, «Fault Detection in HMM's: A Local Asymptotic Approach», *in: Proceedings of the 39th Conference on Decision and Control (CDC), Sydney, IEEE-CSS*, p. 4686–4690, décembre 2000.
- [47] F. LE GLAND, N. OUDJANE, «Stability and Approximation of Nonlinear Filters using the Hilbert Metric, and Applications to Particle Filters», *in: Proceedings of the 39th Conference on Decision and Control (CDC), Sydney, IEEE-CSS*, p. 1585–1590, décembre 2000.
- [48] I. MAGRIN-CHAGNOLLEAU, F. BIMBOT, «Indexing Telephone Conversations by Speakers Using Time-Frequency Principal Component Analysis», *in: Proceedings of the International Conference on Multimedia Education (ICME), New York*, juillet 2000.
- [49] I. MAGRIN-CHAGNOLLEAU, G. DUROU, «Application of Vector Filtering to Pattern Recognition», *in: Proceedings of the 15th International Conference on Pattern Recognition (ICPR), Barcelona, 3 (Image, Speech and Signal Processing), IAPR*, p. 437–440, septembre 2000.
- [50] I. MAGRIN-CHAGNOLLEAU, G. GRAVIER, M. SECK, O. BOEFFARD, R. BLOUËT, F. BIMBOT, «A Further Investigation on Speech Features for Speaker Characterization», *in: Proceedings of the 6th International Conference on Spoken Language Processing (ICSLP), Beijing*, octobre 2000.
- [51] I. MAGRIN-CHAGNOLLEAU, «Utilisation du Filtrage Vectoriel en Reconnaissance des Formes», *in: Actes du 12ème congrès francophone de Reconnaissance des Formes et Intelligence Artificielle (RFIA), Paris, AFRIF / AFIA*, février 2000.
- [52] J. MARIÉTHOZ, F. BIMBOT, «Adaptation Robuste de Modèles HMM pour la Vérification du Locuteur Dépendante du Texte», *in: Actes des XXIIIèmes Journées d'Etudes sur la Parole (JEP), Aussois*, p. 349–352, juin 2000.
- [53] J. MARIÉTHOZ, J. LINDBERG, F. BIMBOT, «A MAP Approach, with Synchronous Decoding and Unit-Based Normalization, for Text-Dependent Speaker Verification», *in: Proceedings of the 6th International Conference on Spoken Language Processing (ICSLP), Beijing*, octobre 2000.
- [54] L. MEVEL, A. BENVENISTE, M. BASSEVILLE, M. GOURSAT, «In-Operation Structural Damage Detection and Diagnosis», *in: Proceedings of the European COST F3 Conference on System Identification and Structural Health Monitoring, Madrid*, juin 2000.
- [55] L. MEVEL, A. BENVENISTE, M. BASSEVILLE, M. GOURSAT, «Polyreference Subspace Based Modal Identification of a Concrete Three-Span Bridge», *in: Proceedings of the 19th International Modal Analysis Conference (IMAC-XIX), Kissimmee, SEM, Inc.*, février 2001.
- [56] L. MEVEL, L. FINESSO, «Bayesian Estimation of Hidden Markov Models», *in: Proceedings of the 14th International Symposium on Mathematical Theory of Networks and Systems (MTNS), Perpignan*, 2000.
- [57] L. MEVEL, L. FINESSO, «Convergence Rates of the Maximum Likelihood Estimator of Hidden Markov Models», *in: Proceedings of the 39th IEEE Conference on Decision and Control (CDC), Sydney, IEEE-CSS*, p. 4691–4696, décembre 2000.

- [58] L. MEVEL, L. HERMANS, H. VAN DER AUWERAER, «Health Monitoring of a Concrete Three-Span Bridge», *in: Proceedings of the 18th International Modal Analysis Conference (IMAC-XVIII), San Antonio*, SEM, Inc., février 2000.
- [59] A. MORIN, M. KERBAOL, J.-Y. BANSARD, «Etude des Résumés en Français des Rapports de Recherche d'un Institut d'Informatique Publiés de 1989 à 1998», *in: Actes des 5èmes journées internationales d'Analyse Statistique des Données Textuelles (JADT), Lausanne*, mars 2000. Résumé.
- [60] A. MORIN, «Utilisation d'Analyses Factorielles des Correspondances Successives à partir d'un Tableau de Données pour des Objectifs Différents en Analyse des Données Textuelles», *in: Actes des 7èmes journées de Classification et d'Analyse des Données (JOCLAD), Lisbonne*, 2000.
- [61] L. S. RICKER, E. FABRE, «On the Construction of Modular Observers and Diagnosers for Discrete Event Systems», *in: Proceedings of the 39th Conference on Decision and Control (CDC), Sydney*, IEEE-CSS, p. 2240–2244, décembre 2000.
- [62] H. VAN DER AUWERAER, L. HERMANS, L. MEVEL, A. BENVENISTE, M. BASSEVILLE, «Model Based Approach to Structural Damage Detection and Localization», *in: Proceedings of the 13th International Congress and Exhibition on Condition Monitoring and Diagnostic Engineering Management (COMADEM'2000), Houston*, décembre 2000.
- [63] Q. ZHANG, «A Nonlinear Adaptive Observer Based Method for Fault Detection and Isolation», *in: Proceedings of the 12th Symposium on System Identification and System Parameter Identification (SYSID), Santa Barbara*, IFAC / IFORS, juin 2000. Paper ThAM4–1.
- [64] Q. ZHANG, «Some Implementation Aspects of Sliding Window Least Squares Algorithms», *in: Proceedings of the 12th Symposium on System Identification and System Parameter Identification (SYSID), Santa Barbara*, IFAC / IFORS, juin 2000. Paper ThPM2–6.

## Rapports de recherche et publications internes

- [65] L. MEVEL, A. BENVENISTE, M. BASSEVILLE, M. GOURSAT, «Blind Subspace-Based Eigenstructure Identification under Nonstationary Excitation Using Moving Sensors», *Publication Interne n° 1354*, IRISA, octobre 2000, <ftp://ftp.irisa.fr/techreports/2000/PI-1354.ps.gz>.

## Divers

- [66] L. BARATCHART, Q. ZHANG, «De l'Observation au Modèle», INédit 27 (Spécial mathématiques), décembre 2000, <http://www.inria.fr/actualites/inedit/pdf/inedit27.fr.pdf>.
- [67] M. BASSEVILLE, G. CELEUX, «Les Statistiques au Secours de la Sûreté Industrielle», INédit 27 (Spécial mathématiques), décembre 2000, <http://www.inria.fr/actualites/inedit/pdf/inedit27.fr.pdf>.
- [68] M. BASSEVILLE, «In-Operation Modal Analysis and Vibration Monitoring», ERCIM News 40 (Special theme: Control and Systems Theory), janvier 2000, [http://www.ercim.org/publication/Ercim\\_News/enw40/basseville.html](http://www.ercim.org/publication/Ercim_News/enw40/basseville.html).
- [69] F. CÉROU, «Diagnostic Transparent : Surveillance en Ligne et Asynchrone du Taux d'Erreur dans un Réseau Optique», Rapport de fin d'étude, convention Alcatel / Inria 100C0294, novembre 2000.

- 
- [70] E. FABRE, « MAGDA — Alarm Supervision in Telecommunication Networks », ERCIM News 40 (Special theme: Control and Systems Theory), janvier 2000, [http://www.ercim.org/publication/Ercim\\_News/enw40/fabre.html](http://www.ercim.org/publication/Ercim_News/enw40/fabre.html).
- [71] F. LE GLAND, « Statistics of Hidden Markov Models », ERCIM News 40 (Special theme: Control and Systems Theory), janvier 2000, [http://www.ercim.org/publication/Ercim\\_News/enw40/le\\_gland.html](http://www.ercim.org/publication/Ercim_News/enw40/le_gland.html).
- [72] J. LÉVY-VÉHEL, A. BENVENISTE, « Ondelettes et Fractales », INédit 27 (Spécial mathématiques), décembre 2000, <http://www.inria.fr/actualites/inedit/pdf/inedit27.fr.pdf>.