

PARTIE II.

MODELISATION.

PARTIE II: MODELISATION.

CHAPITRE 4. MODELES NON-STATIONNAIRES.

La partie I était bâtie autour du concept de relief d'un signal. Il s'agit là d'un outil d'analyse qui a pour vocation première la représentation du signal, c'est à dire la mise en évidence de ses propriétés. Cette motivation n'est pas la seule que peut avoir celui qui est amené à traiter un signal, et qui peut aussi vouloir transmettre ce signal, le synthétiser ou y reconnaître une forme. Rien ne s'oppose en théorie à ce qu'un relief soit à même d'accomplir ces diverses tâches, mais dans la pratique, ces dernières doivent être parcimonieuses, et ménager la place mémoire, d'encombrement sur disque ou en mémoire morte. Ce souci d'économie pousse à se tourner vers les techniques de modélisation qui, elles, permettent une réduction importante du nombre d'éléments représentant le signal.

Cette seconde partie est ainsi consacrée à la présentation d'une méthodologie de modélisation applicable aux signaux non-stationnaires et qui est adaptée aux diverses motivations rencontrées dans le traitement du signal. D'autres méthodologies existent, mais ne sont pas étudiés ici. On peut, de façon assez brutale et approximative, voir dans la littérature portant sur la modélisation non-stationnaire quatre approches. La plus richement représentée est la modélisation adaptative qui charge l'algorithme d'identification de prendre en compte la non-stationnarité. Une deuxième approche limite les non-stationnarités à des sauts en des instants isolés

entre lesquels le modèle reste invariant, on peut dire qu'il s'agit de modèles stationnaires par morceaux. Les deux autres approches sont plus explicitement non-stationnaires car elles incluent dans le modèle le mode de variation des paramètres du signal, dans un cas, déterministe, les paramètres s'exprimant linéairement sur une base de fonctions connues a priori donnant des modèles que l'on appellera évolutifs, dans l'autre cas aléatoire, les paramètres du modèle étant alors supposés markoviens. Ce chapitre a pour but une description rapide de ces quatre manières de s'attaquer au problème du signal non-stationnaire et de sa modélisation. Ceci permettra de situer l'étude décrite dans les chapitres suivants.

1. Modèles adaptatifs.

L'approche adaptative est celle qui a donné lieu à la plus importante littérature, supérieure en nombre à celle des autres approches. L'essor de ces méthodes s'est fait plutôt dans le domaine de l'automatique, pour l'identification des systèmes (SAGE, MELSA, 1971, EYKHOFF, 1974, MENDEL, 1973). C'est postérieurement que le transfert s'est fait vers le domaine du traitement du signal, essentiellement sous la forme d'algorithmes d'estimation d'un modèle autorégressif, et ensuite d'un modèle ARMA.

Les méthodes se répartissent en fonction de leur complexité suivant une gradation qui débute avec les méthodes du premier ordre n'utilisant que le gradient du critère à optimiser. Celles-ci sont donc apparentées aux méthodes d'approximation stochastique, et se caractérisent par un coût de calcul faible, mais un temps de convergence plus long que pour des méthodes de second-ordre. Dans ces dernières, l'optimisation est effectuée par un algorithme du type Newton-Raphson, combinant l'emploi du gradient du critère avec celui du Hessien. On rencontre ce Hessien au sens strict de la matrice

des dérivées secondes, dans la méthode des moindres carrés, tandis que les méthodes plus élaborées de maximum de vraisemblance ou de filtre de Kalman étendu donnent à cette matrice une signification plus large.

Les modèles à identifier sont le modèle autorégressif ou AR en (2-1) et le modèle autorégressif à moyenne ajustée ou ARMA en (2-2), dans lesquels y_t est le signal, ϵ_t est un bruit blanc centré, de variance σ^2 dans le modèle AR et 1 dans le modèle ARMA.

$$(2-1) \quad y_t + a_1 y_{t-1} + \dots + a_p y_{t-p} = \epsilon_t$$

$$(2-2) \quad y_t + a_1 y_{t-1} + \dots + a_p y_{t-p} = b_0 \epsilon_t + \dots + b_q \epsilon_{t-q}$$

On considérera également le modèle MA obtenu à partir de (2-2) quand $p=0$, mais ce modèle est surtout vu comme une étape dans l'estimation du modèle ARMA plutôt que comme un modèle autonome.

L'élaboration d'une méthode de modélisation adaptative peut être vue comme un processus en deux étapes: la première consiste à rendre récursive une méthode globale, la seconde à rendre oublieuse la méthode ainsi obtenue, pour lui permettre de suivre le signal dans ses non-stationarités. Explicite-tons ce processus dans le cadre de la méthode des moindres carrés. Lorsque le signal est connu sur l'intervalle $[0, t]$, en écrivant θ_t le vecteur contenant les paramètres $a_1 \dots a_p$ du modèle estimé sur cet intervalle, la relation d'optimalité s'écrit:

$$(2-3) \quad \theta_t = -(R_t)^{-1} \xi_t \quad \text{avec:}$$

$$(2-4) \quad \xi_t = \sum_{\tau=0}^t y_{\tau} \begin{bmatrix} y_{\tau-1} \\ \cdot \\ y_{\tau-p} \end{bmatrix}$$

$$(2-5) \quad R_t = \sum_{\tau=0}^t \begin{bmatrix} y_{\tau-1} \\ \cdot \\ y_{\tau-p} \end{bmatrix} \begin{bmatrix} y_{\tau-1} & \dots & y_{\tau-p} \end{bmatrix}$$

On a ici posé $y_\tau=0$ pour $\tau < 0$. Rendre récursif cet estimateur consiste à déterminer θ_{t+1} en fonction de θ_t . Si on note Y_t le vecteur des échantillons $y_{t-1} \dots y_{t-p}$, le passage de t à $t+1$ se fait pour les quantités ξ_t et R_t par:

$$(2-6) \quad \xi_{t+1} = \xi_t + Y_{t+1} y_{t+1}$$

$$(2-7) \quad R_{t+1} = R_t + Y_{t+1} Y_{t+1}^T$$

Le calcul de θ_{t+1} se fait alors par application du lemme d'inversion matricielle à (2-7), ce qui donne:

$$(2-8) \quad (R_t)^{-1} = R_t^{-1} - R_t^{-1} Y_{t+1} (1 + Y_{t+1}^T R_t^{-1} Y_{t+1})^{-1} Y_{t+1}^T R_t^{-1}$$

En combinant (2-6) et (2-8), et en posant:

$$(2-9) \quad \epsilon_t = y_t + \theta_{t-1} Y_t$$

il vient:

$$(2-10) \quad \theta_{t+1} = \theta_t + R_t^{-1} Y_{t+1} (1 + Y_{t+1}^T R_t^{-1} Y_{t+1})^{-1} \epsilon_{t+1}$$

La seconde étape consiste ensuite à transformer l'algorithme donné par (2-8) et (2-10) de façon à ce que le modèle suive les évolutions du système supposé avoir engendré le signal. Il faut pour cela faire en sorte que la matrice R_t qui représente une covariance du signal soit adaptée localement au signal, aux environs de l'instant t , c'est à dire que l'estimateur R_t ait oublié les valeurs de y_τ situées dans un passé suffisamment lointain. A des variantes près, deux méthodes existent pour cela. La première introduit un facteur d'oubli λ dans l'estimateur récursif (2-7):

$$(2-11) \quad R_{t+1} = \lambda R_t + Y_{t+1} Y_{t+1}^T$$

Le facteur λ est pris voisin de 1, mais inférieur. Ceci revient à pondérer dans le critère d'estimation l'erreur à l'instant τ par $\lambda^{t-\tau}$ d'où l'oubli du

passé, avec une décroissance exponentielle. Dans la seconde méthode, une fenêtre de longueur constante est déplacée sur le signal ce qui revient à utiliser pour R_t l'estimateur récursif suivant:

$$(2-12) \quad R_{t+1} = R_t + Y_{t+1} Y_{t+1}^T - Y_{t-h} Y_{t-h}^T$$

La mémoire de l'algorithme est alors représentée par la durée h de l'intervalle où sont sélectionnés les échantillons. Dans le premier cas, oubli exponentiel, l'algorithme découle directement de (2-8) et (2-10) en remplaçant R_t par λR_t . Dans le second cas, mémoire finie, une double application à (2-12) du lemme d'inversion matricielle donne des équivalents à (2-8) et (2-10) où interviennent deux corrections, l'une à $t+1$ pour l'échantillon entrant dans la fenêtre, l'autre à $t-h$ pour l'échantillon sortant. Le coût du calcul en est donc doublé.

Les méthodes du second ordre, dans la formulation qui vient d'être donnée par (2-8) et (2-10), ont certes une rapidité de convergence appréciable, mais leur coût de calcul est de l'ordre de p^2 opérations par échantillon, contrairement aux méthodes du premier ordre avec un coût en p opérations. Cet inconvénient a été supprimé par LJUNG, MORF, FALCONER, 1978, qui ont donné un algorithme récursif de résolution de l'équation (2-3) dont le coût de calcul est de l'ordre de p opérations par échantillon. Le principe repose sur le fait que la matrice R_t est le produit d'une matrice Y_t qui est de Toeplitz, par sa transposée:

$$(2-13) \quad Y_t = \begin{bmatrix} y_{p-1} & & y_{t-p} & y_{t-1} \\ y_0 & y_{p-1} & & y_{t-p} \end{bmatrix}$$

On voit donc que la dernière colonne de Y_t qui est le vecteur Y_t n'est pas indépendant de la matrice R_{t-1} , ses $p-1$ dernières composantes ayant déjà

servi au calcul de R_{t-1} . Ce fait a été utilisé par LJUNG, MORF, FALCONER, 1978, et leur permet de constater que la matrice $R_{t+1}(p+1)$ associée au modèle d'ordre $p+1$ contient à la fois la matrice $R_{t+1}(p)$ (dans son coin supérieur gauche) et la matrice $R_t(p)$ dans son coin inférieur droit. L'algorithme repose entièrement dans une utilisation efficace de cette constatation par une alternance entre les ordres p et $p+1$. Cette démarche a ensuite été étendue aux diverses méthodes adaptatives, que ce soit à oubli exponentiel ou à mémoire finie, pour l'identification des systèmes. Le concept qui est au coeur de cette extension a été formalisé par FRIEDLANDER, KAILATH, MORF, LJUNG, 1978, et les mêmes auteurs en 1979 sous le nom de rang de déplacement. Il s'agit d'une mesure algébrique de la distance d'une matrice R à une matrice de Toëplitz. Une abondante littérature a été consacrée à ces algorithmes rapides par Kailath, Morf, Ljung, Friedlander, Delosme, Lev-Ari, Porat etc ... (en voir une synthèse dans LANDAU, 1982). Les aspects les plus récents de cette démarche concernent l'identification du filtre en treillis AR ou ARMA, et l'écriture d'algorithmes dits en racine carrée où est propagée la racine carrée au sens d'une décomposition de Cholesky de la matrice R , pour garantir la positivité des covariances de l'erreur, malgré la propagation des arrondis dus aux calculs. Ces algorithmes suscitent actuellement des efforts prometteurs pour leur réalisation en circuits intégrés VLSI.

L'approche adaptative qui vient d'être brièvement évoquée se signale donc par une grande souplesse algorithmique manifeste dans le nombre des variantes possibles comme dans la facilité de mise en oeuvre ou d'implantation. Un de ses avantages essentiels est de ne pas faire d'hypothèses a priori sur le type de non-stationnarité que présente le signal tant que celui-ci, du moins, reste centré. Ceci est en réalité assez

illusoire car chacune de ces méthodes postule un modèle stationnaire sur la durée de l'analyse exprimée par la longueur h de la mémoire, ou une durée équivalente dans le cas de l'oubli. C'est alors ce paramètre de l'algorithme qui fixe la non-stationnarité "maximale" que peut suivre la méthode. Pour une précision fixée, la pente maximale de chaque paramètre du modèle est approximativement inversement proportionnelle à la mémoire h . Ceci exclut en particulier les discontinuités dans le modèle. Celles-ci ne peuvent être suivies qu'avec un certain retard par les méthodes adaptatives. Un second inconvénient de ces méthodes réside dans l'absence de paramétrisation de l'évolution temporelle du signal: si le modèle AR à l'instant t condense en p paramètres la totalité de l'information concernant le spectre du signal, il faut sur un intervalle $[0, t]$ conserver les $t+1$ valeurs successives du modèle ou bien faire intervenir après la modélisation et indépendamment d'elle, une réduction tenant compte de la redondance entre modèles successifs (par exemple par un lissage des trajectoires des paramètres comme le pratique ATAL, 1983, ou par une interpolation polynomiale ...). Cette balance entre cet aspect négatif et l'aspect positif qu'est sa souplesse, est cause que la démarche adaptative connaît un grand succès dans le domaine de la transmission des signaux (parole, images), et peu de succès dans la reconnaissance de formes, ou la synthèse.

2. Modèles stationnaires par morceaux.

La seconde façon d'aborder la modélisation de signaux non stationnaires introduit une contrainte forte sur le type de non-stationnarité: elle suppose qu'à certains instants, le modèle présente des sauts d'un jeu de valeurs à un autre, et qu'entre ces instants, le modèle demeure constant. Mais cette hypothèse peut avoir deux lectures, l'une stricte, l'autre plus

relâchée. Strictement, on considérera que c'est réellement dans la production du signal que se situent ces sauts, et le problème sera d'identifier les instants de leur occurrence, puis d'estimer la valeur des modèles sur chaque intervalle. Dans une lecture plus relâchée, l'hypothèse des sauts n'est vue que comme une astuce d'analyse, permettant de modéliser l'évolution lente mais continue d'un signal comme une succession de marches d'escalier. Les instants de sauts sont alors arbitraires, ne dépendent plus du signal lui-même et peuvent être fixés par une durée constante pour chaque intervalle séparant deux sauts successifs. Ce schéma est très utilisé en reconnaissance de la parole, donnant lieu à ce qu'on dénomme analyse "centiseconde" où un modèle est estimé toutes les dix millisecondes.

Dans l'hypothèse stricte, le travail d'identification se décompose le plus souvent en deux phases: segmentation du signal en zones stationnaires, puis identification d'un modèle sur chaque zone, ce qui ne pose pas de problème grâce à la stationnarité. Les méthodes de segmentation sont nombreuses, et ont fait l'objet d'une littérature abondante. Si plusieurs d'entre elles utilisent les outils de la modélisation, ce n'est pas en général leur but aussi ne seront elles pas étudiées ici. Du point de vue de la modélisation d'un signal stationnaire par morceaux, seules sont à retenir les méthodes de maximum de vraisemblance qui réalisent un test sur les deux hypothèses: H_0 , le signal ne présente pas de saut de modèle dans l'intervalle d'étude $[0, t]$ et H_1 , le signal présente un saut à un instant τ , $0 < \tau < t$. La vraisemblance s'estime à partir des variances des résidus: σ_0^2 pour le modèle dans H_0 et σ_1^2 , σ_2^2 pour les deux modèles dans H_1 . C'est ce test que mettent en oeuvre OSAKI, TONG, 1975, et KITAGAWA, AKAIKE, 1978, après l'avoir modifié pour choisir simultanément l'ordre de chaque modèle, et bien sûr les paramètres du modèle. Ils se servent pour cela du critère

d'ordre de AKAIKE, 1971, et associent à H_0 et H_1 les critères C_0 et C_1 :

$$(2-14-1) \quad C_0 = t \cdot \text{Log}(\sigma_0^2) + 2(p_0 + 2)$$

$$(2-14-2) \quad C_1 = \tau \cdot \text{Log}(\sigma_1^2) + (t - \tau) \text{Log}(\sigma_2^2) + 2(p_1 + p_2 + 4)$$

Dans cette relation, p_0 désigne l'ordre du modèle sous H_0 , p_1 et p_2 les ordres des deux modèles sous H_1 . Cette méthode peu répandue, semble-t-il a été utilisée par GERSCH, BROTHERTON, 1982, sur des mesures de vibration d'un immeuble.

Il existe une méthode récente qui repose aussi sur l'idée de signaux stationnaires par morceaux. Elle est due à SHIRAI, MINOWA, 1983, et est assez originale. Elle considère d'abord un modèle AR sur l'ensemble du signal, et le résidu ϵ_t qui en découle, puis découpe l'intervalle en parties. Un modèle AR est alors identifié sur le résidu précédent, dans chaque sous-intervalle. La même chose est répétée un nombre fini de fois m , ce qui conduit à 2^m sous-intervalles, et à $m+1$ modèles sur chacun, ces modèles étant en cascade. Pour séduisante que paraisse cette idée, il n'est pas sûr pourtant qu'elle apporte plus qu'une série de modèles estimés indépendamment les uns des autres sur les 2^m intervalles finaux. Le nombre des paramètres est certes moindre dans la démarche hiérarchique, mais la modélisation en cascade est moins précise qu'une minimisation directe du résidu final.

3. Modèles à coefficients aléatoires.

Le troisième cadre possible pour la modélisation des signaux non-stationnaires est celui où les coefficients du modèle sont eux-mêmes considérés comme aléatoires. On a là un modèle à deux niveaux, le premier est celui du système engendrant le signal, le second est celui de la dynamique du système au niveau 1. Cette dynamique est elle-même exprimée par un

modèle qui peut aller (par ordre de complexité croissante) du modèle markovien (FURUTA, PAQUET, 1969, HARVEY, 1978), au modèle AR (CARAYANNIS, 1973, KAHUSHO, YANAGIDA, 1982, SHIRAI, KOBAYASHI, 1983, GRENIER, 1975) au modèle d'état (KITAGAWA, GERSH, 1983), et au modèle ARMA (INDJEHAGOPIAN, 1981). Quant à BANKA, 1979, il ne fait aucune hypothèse sur le second niveau et identifie le modèle du premier niveau par une procédure complexe, maximisant la vraisemblance du modèle au moyen d'une méthode de gradients conjugués. L'absence de contraintes au second niveau entraîne une énorme dispersion des paramètres estimés, alors que leur biais semble faible. Ceci est probablement à mettre au compte de la non-identifiabilité du modèle (trop de paramètres par rapport aux mesures) que BANKA, 1979, ne semble pas remarquer.

L'identification simultanée des deux niveaux est très délicate, puisqu'il s'agit en général d'un problème non-linéaire: au second niveau l'entrée comme la sortie du modèle sont inconnues, et multipliées par les paramètres eux aussi inconnus du modèle de niveau 2. Si on suppose par exemple que le vecteur θ_t des paramètres d'un modèle autorégressif de y_t (1er niveau) est markovien (2ème niveau), on obtient les équations (2-15):

$$(2-15) \quad \begin{cases} y_t + [y_{t-1} \dots y_{t-p}] \theta_t = \epsilon_t \\ \theta_t = A \theta_{t-1} + v_t \end{cases}$$

Dans ce modèle, y_t est le signal, engendré par le modèle AR θ_t excité par le bruit blanc ϵ_t , et le modèle θ_t a pour matrice de transition A (éventuellement A_t) et pour innovation le bruit blanc vectoriel v_t . Les équations (2-15) forment alors une équation d'état pour y_t et le problème est bien celui, non-linéaire, de l'estimation conjointe de l'état θ_t et des paramètres A du

systeme.

Les auteurs déjà cités qui ont considéré ces modèles à deux niveaux ont contourné la difficulté en considérant tous l'un des deux niveaux résolu afin de pouvoir identifier l'autre. Lorsque les paramètres du second modèle, c'est à dire la matrice A, et la covariance de v_t dans (2-15) sont connus, l'identification de θ_t est un problème d'estimation de l'état d'un système, qui se réalise au moyen du filtre de Kalman (dans une approximation gaussienne pour y_t). Si la matrice de transition A doit être bien connue, il est par contre à remarquer qu'il suffit pour les covariances de v_t et de ϵ_t de disposer d'une estimation sur-évaluée pour que le filtre donne une estimation non biaisée de θ_t . Utilisant cette méthode, FURUTA, PAQUET, 1969, identifient un modèle autorégressif d'ordre 1 dont le coefficient $a_1(t)$ est lui-même autorégressif d'ordre 1, ou markovien. HARVEY, 1978, considère un modèle tel que (2-15) pour le paramètre θ_t d'une régression linéaire, mais θ_t n'est pas centré et possède une moyenne θ :

$$(2-16) \quad \begin{cases} y_t = x_t^T \theta_t \\ \theta_t - \theta = A(\theta_{t-1} - \theta) + v_t \end{cases}$$

INDJEHAGOPIAN, 1981, étudie le cas où θ_t est un vecteur ARIMA (autorégressif intégré à moyenne ajustée) d'ordre (p,q,m):

$$(2-17-1) \quad (A_0 + A_1 \sigma + \dots + A_p \sigma^p)(1 - \sigma)^m \theta_t = (B_0 + B_1 \sigma + \dots + B_q \sigma^q) v_t$$

$$(2-17-2) \quad y_t = x_t^T \theta_t + \epsilon_t$$

Dans cette équation, σ représente l'opérateur de décalage: $\sigma \theta_t = \theta_{t-1}$. Les A_i , B_i et les ordres p, q, m sont supposés connus, la solution fait également appel à un filtre de Kalman. C'est le cas particulier ARIMA (0,0,m)

qu'étudient KITAGAWA, GERSCH, 1983, où (2-17-1) est remplacé par (2-18):

$$(2-18) \quad (1-\sigma)^m \theta_t = v_t$$

Dans ce dernier cas, le modèle se met sous la forme (2-15) avec une matrice A très simple, sous une forme bloc compagne où les α_i sont les coefficients du binôme:

$$(2-19-1) \quad y_t = [y_{t-1} \dots y_{t-p} \ 0 \dots 0] \theta_t = \epsilon_t$$

$$(2-19-2) \quad \theta_t = \begin{bmatrix} \alpha_m I & \dots & \alpha_1 I \\ I & & 0 \\ & \diagdown & \\ 0 & & I & 0 \end{bmatrix} \theta_{t-1} + \begin{bmatrix} I \\ 0 \\ 0 \end{bmatrix} v_t$$

Si on souhaite estimer le modèle du second niveau, il est nécessaire que le modèle du premier niveau ait déjà été identifié, par une des autres méthodes de modélisation non-stationnaire, soit une méthode adaptative, soit une méthode du type "centiseconde". Cette dernière solution est retenue par KAKUSHO, YANAGIDA, 1982, qui estiment sur chaque coefficient $a_i(t)$ un modèle autorégressif scalaire. CARAYANNIS, 1973, utilise un filtre de Kalman étendu pour estimer le modèle AR scalaire de niveau 1, puis le modèle AR vectoriel (A_i matriciels) du niveau 2. Le suivi de l'innovation lui permet de segmenter le signal de parole en phonèmes. J'ai présenté un exemple analogue (GRENIER, 1980), où le vecteur θ_t est modélisé comme un signal vectoriel autorégressif dont les coefficients matriciels $A_1 \dots A_p$ permettent de reconnaître l'identité du locuteur ayant prononcé le signal scalaire de parole.

Un autre usage de ce modèle à deux niveaux où le premier est estimé classiquement, a été fait par moi-même (GRENIER, 1975), et SHIRAI, KOBAYASHI, 1983. L'idée est celle de "cibles" sous-jacentes à la suite des

modèles θ_t du niveau 1. Ceux-ci, dans le traitement de la parole représentent la position à l'instant t des organes phonatoires. Cette position résulterait de commandes issues du cerveau qui indiqueraient la position idéale à prendre par les organes phonatoires pour produire le phonème ou le son élémentaire suivant. Les commandes sont ensuite filtrées par le système mécanique (musculaire) qui amène les organes phonatoires vers cette position. Il s'agit donc de cibles vers lesquelles tend θ_t , mais qui ne sont pas en général atteintes, car le phonème suivant survient trop tôt, ce qui peut être une façon de se figurer le phénomène de co-articulation. Du point de vue du modèle de niveau 2, les cibles sont des entrées en échelons, et le système est un système résonant d'ordre faible, choisi égal à 2 dans les deux travaux cités. Comme dans le cas stationnaire par morceaux, l'identification peut se faire en deux temps: détection des sauts de l'entrée, puis estimation du modèle (GRENIER, 1975) ou conjointement, par une procédure de programmation dynamique (SHIRAI, KOBAYASHI, 1983). Ces derniers auteurs montrent d'ailleurs la validité de l'hypothèse des cibles en se servant des cibles estimées pour effectuer une reconnaissance des phonèmes.

En marge de ces modèles à coefficients markoviens, on peut aussi citer la démarche de PORITZ, 1982, qui considère le signal découpé en tranches successives. Les modèles autorégressifs de ces tranches forment une chaîne de Markov dont les n états possibles sont n modèles AR. PORITZ, 1982, identifie simultanément ces n états et leurs probabilités de transition de manière itérative.

4. Modèles évolutifs.

La quatrième classe de modèles envisageable est celle dans laquelle

prend place cette étude. Les paramètres du modèle AR ou ARMA du signal non-stationnaire s'y expriment comme combinaisons linéaires de fonctions $f_i(t)$, en nombre fini, connues à priori. Si LIPORACE, 1975, a été le premier à étudier l'estimation d'un modèle AR appartenant à cette classe, l'idée d'introduire ces fonctions $f_i(t)$ était déjà sous-jacente à deux travaux antérieurs (MENDEL, 1969, 1973 et RAO, 1970). Le premier proposait pour l'estimation du vecteur θ_t des paramètres d'un système d'effectuer un changement de représentation $\theta_t = P(t)\beta_t$ où $P(t)$ est une matrice de changement de base connue à priori et β_t , le vecteur transformé évoluera moins vite que θ_t ce qui améliorera son estimation par une méthode adaptative par gradient. Le second approxime chaque paramètre $a_i(t)$ par son développement limité d'ordre 2: $a_i(t) = a_{i0} + a_{i1}t + a_{i2}t^2/2$. L'estimation se fait par un critère du maximum de vraisemblance. Cette méthode s'apparente à celle de LIPORACE, 1975, en prenant $f_0(t)=1$, $f_1(t)=t$ et $f_2(t)=t^2/2$, mais RAO, 1970, n'explicite pas les raisons qui le poussent à limiter le nombre des fonctions $f_i(t)$ à 3 ($i=0,1,2$).

Le modèle autorégressif (2-1) et le modèle ARMA (2-2), seront dits évolutifs lorsque leurs coefficients sont combinaisons linéaires des fonctions $f_i(t)$:

$$(2-20) \quad a_i(t) = \sum_{j=0}^m a_{ij} f_j(t)$$

La première question qui se pose est celle du choix des fonctions $f_i(t)$ pour une application donnée. LIPORACE, 1975, ne précise pas ce point, mais contraint les fonctions $f_i(t)$ à être orthogonales. Ceci n'est pas nécessaire pour l'identification du modèle, mais améliore l'estimation par un meilleur conditionnement numérique des systèmes à résoudre. La base de fonctions la

plus simple est celle des puissances du temps, issue du développement limité des $a_i(t)$:

$$(2-21) \quad f_m(t) = \frac{t^m}{m!}$$

Cette base est utilisée par plusieurs auteurs (tableau 2-1), de même que son équivalent après orthogonalisation, où les fonctions $f_i(t)$ sont des polynômes de Legendre. HALL, OPPENHEIM, WILLISKY, 1977, 1983, ont introduit un second jeu de fonctions, issues de la décomposition en série de Fourier des coefficients $a_i(t)$:

$$(2-22) \quad \begin{cases} f_0(t) = 1 \\ f_{2m-1}(t) = \cos(mt), \quad m > 0 \\ f_{2m}(t) = \sin(mt), \quad m > 0 \end{cases}$$

Une troisième base introduite par GRENIER, 1981-c, est celle des fonctions sphéroidales aplaties (annexe 5). Sur une durée T , un signal dont la largeur de bande est B peut se représenter par $m = \text{Ent}(BT)$ fonctions sphéroidales aplaties où $\text{Ent}(x)$ est la partie entière de x . Pour obtenir une famille de $(m+1)$ fonctions $f_i(t)$, on utilisera les $(m+1)$ premières fonctions sphéroidales aplaties associées à la durée T et la largeur de bande $B = \frac{T}{m+1}$.

Cette diversité dans le choix des bases possibles se retrouve bien entendu aussi dans les méthodes d'estimation (tableau 2-1). La méthode la plus simple est la méthode du gradient ou approximation stochastique, utilisée par MENDEL, 1969, 1973. Plus complexe mais plus efficace, est la méthode de moindres carrés utilisée par RAO, 1970, et plusieurs autres auteurs. La méthode du maximum de vraisemblance a été mise en oeuvre avec un calcul toujours approché de cette vraisemblance par RAO, 1970, KOZIN, 1977,

ABDRABBO, 1979, CLERGEOT, 1982. On trouve aussi chez HINICH, ROLL, 1977, l'emploi d'une méthode robuste et non paramétrique d'estimation des $a_i(t)$ suivie par leur approximation par des polynômes de Legendre. Certains auteurs utilisent une estimation en plusieurs étapes du modèle, à chaque étape est estimé une partie du modèle, en principe les composantes des $a_i(t)$ sur une des fonctions de la base. Cette partie est estimée sur le résidu obtenu par filtrage du signal avec les parties déjà identifiées du modèle. NAKAJIMA, SUZUKI, 1980, et 1982, le font de façon optimale, et MARTINELLI, 1983, d'une manière sous-optimale. PEARSON, 1982, est le seul auteur à avoir appliqué ces idées pour des signaux à temps continu. GINN, 1982, identifie un modèle autorégressif complexe sur une sinusoïde modulée, sa procédure d'estimation est non-linéaire. GERSCH, KITAGAWA, 1982, étendent ces modèles au cas vectoriel et résolvent les équations d'optimalité par une transformation de Householder. XIANYA, EVANS, 1984-a et -b, font de ces modèles une estimation adaptative par fenêtre glissante.

La mise en pratique de ces modèles évolutifs suppose que des procédures permettent une prise de décision purement mécanique entre deux bases de fonctions possibles, ou entre plusieurs ordres, plusieurs dimensions de la base de fonctions. Ce point a été très peu étudié. Seuls KOZIN, NAKAJIMA, 1980, puis CLERGEOT, 1984, ont considéré le problème de l'estimation de l'ordre du modèle, et ceci lorsque les paramètres sont estimés au sens du maximum de vraisemblance. Ils montrent alors que la consistance et la normalité asymptotique des paramètres assurent la validité du critère d'AKAIKE. Il semble alors raisonnable d'utiliser les divers critères d'ordre existants, même dans le cas des estimateurs de moindre carrés, bien que la validité de ces critères n'ait pas été montrée. On retiendra essentiellement

les critères AIC d'AKAIKE, 1971, celui de SCHWARZ, 1978, et de RISSANEN, 1978, celui de HANNAN, QUINN, 1979. Ces trois critères ont une expression simple à partir de la vraisemblance du modèle, comme l'a montré LAKEHAL, 1980. Il est clair que ces critères permettent de choisir la meilleure structure "ordre + dimension de la base" pour un modèle dans une base donnée, sur un ensemble d'échantillons $[y_0 \dots y_T]$ d'un signal non stationnaire. Le choix entre deux ou plusieurs bases possibles se fait alors simplement en retenant pour chaque base la structure minimisant le critère d'ordre, puis la base associée au plus petit des critères ainsi obtenus.

Auteurs.	Année.	Modèle			Estimateur				Base			
		A	M	R	G	C	V	A	T	L	F	S
MENDEL,	69,73			*	*							
RAO,	70	*				*	*		*			
LIPORACE,	75	*				*				*		
HALL et al.	77,83	*				*			*		*	
HINICH, ROLL,	77			*				*		*		
KOZIN,	77	*					*			*		
ABDRABBO,	79	*					*		*			
NAKAJIMA, SUZUKI,	80,80,82	*				*			*			
CLERGEOT,	82,84,	*					*					
GERSCH, KITAGAWA,	82	*				*				*		
GINN,	82	*										
PEARSON,	82		*			*			*			
MARTINELLI,	83	*						*			*	
SHARMAN et al.	84	*				*					*	
XIANYA, EVANS,	84	*				*			*			
Cette étude,		*	*			*	*			*	*	*

Tableau 2-1:

Comparaison des méthodes évolutives publiées, vis-à-vis:
du modèle choisi (A=autorégressif, M=mixte autorégressif à moyenne ajustée, R=régression linéaire),
de l'estimateur (G=gradient, C=moindres carrés, V=maximum de vraisemblance, A=approximations successives)
et de la base (T=puissances du temps t ou développement limité, L=polynômes de Legendre, F=série de Fourier, S=fonctions sphéroïdales aplaties).

Dans cette classe de modèles évolutifs peuvent s'incorporer nombre de cas particuliers qui ne semblent pas à priori en faire partie, mais qui pourvu qu'on formule convenablement la base des fonctions $f_i(t)$, rentrent dans ce formalisme. Ainsi les modèles stationnaires par morceaux, à condition de prendre pour base les fonctions indicatrices des intervalles entre sauts. Ainsi les modèles autorégressifs périodiques. Ceux-ci ont des coefficients périodiques, de période n . KAPUSTINSKAS, 1977, a montré que leur estimation pouvait se faire par n systèmes linéaires qui sont des équations du type Yule-Walker, décalées dans le temps. PAGANO, 1978, a établi que ces équations avaient une interprétation vectorielle: le signal scalaire est remplacé par le vecteur $Y_t^T = [y_t \ y_{t-1} \ \dots \ y_{t-n+1}]$. Les n modèles se déduisent alors du modèle autorégressif de Y_t . SAKAI, 1982, donne un algorithme rapide adapté à ce cas. On peut alors remarquer que si on définit les fonctions $f_i(t)$ par $f_i(t) = (t-i)$ modulo n pour $i=0 \dots n-1$, on obtient un modèle évolutif sur cette base avec $m=n-1$. L'analogie est d'autant plus grande que l'estimation du modèle évolutif fera apparaître dans le cas général le passage d'un modèle scalaire non-stationnaire à un modèle vectoriel stationnaire.

1. Développements de la méthodologie évolutive.

Il reste maintenant, le cadre étant posé à préciser la position relativement à lui des résultats élaborés dans cette étude (dernière ligne du tableau 2-1). Le modèle autorégressif évolutif a été bien étudié par LIPORACE, 1975, et par HALL, OPPENHEIM, WILLSKY, 1977, 1973. Cependant les algorithmes d'estimation qu'ils donnent ne sont pas les plus efficaces possibles. Ils se sont intéressés essentiellement à la réduction du nombre d'opérations pour le calcul des covariances du signal et de ses dérivés,

mais ils n'ont pas étudié l'algorithme de résolution des équations d'optimalité, en vue de réduire son coût. Il est pourtant possible de déduire d'algorithmes rapides existants, des algorithmes qui pour le modèle non-stationnaire auront un coût en $p^2 m^3$ opérations où p est l'ordre du modèle, m la dimension de la base de fonctions, (GRENIER, 1981-c, 1982-b) au lieu des $p^3 m^3$ opérations de l'algorithme donné par LIPORACE, 1975. Comme le produit pm sera pour un signal de parole en général compris entre 30 et 80, on voit que le gain en vitesse de calcul est appréciable. Ceci s'obtient grâce à la structure des matrices de covariance qui sont de Toeplitz ou proches de Toeplitz.

A côté de cette amélioration des algorithmes existants, il a été possible d'étendre à d'autres modèles l'idée d'origine de cette méthodologie qui est de postuler l'existence et la connaissance d'une famille de fonctions $f_i(t)$ exprimant les évolutions des paramètres du modèle. L'extension la plus directe, mais pas nécessairement la plus facile à obtenir, du modèle autorégressif est donnée par le modèle mixte autorégressif à moyenne ajustée ou ARMA. L'identification de ces modèles peut se faire en deux temps. Dans le premier, c'est la partie autorégressive du modèle qui est estimée, par une équation du type Yule-Walker, rendue non-symétrique (GRENIER, 1981-c, 1983-c). La solution de cette équation met en jeu, comme dans le modèle AR un algorithme rapide en $p^2 m^3$ opérations (GRENIER, 1982-b). De même que, pour le modèle stationnaire, la partie autorégressive peut être estimée par une variante correspondant à un système d'équations surdéterminé, dans la méthode dite de "covariance sur la corrélation" ou la méthode pondérée due à Cadzow, de même dans le cas non-stationnaire peut s'écrire un système d'équations sur-déterminé (GRENIER, 1983-c), mais il semble alors que tout

algorithme rapide soit à exclure.

Le second temps, dans l'identification ARMA, sera celui de l'estimation de la partie MA, après qu'ait été éliminée, par filtrage inverse, la contribution de la partie AR du modèle. Deux solutions à ce problème peuvent être proposées. Dans l'une, le modèle MA est approximé par un modèle autorégressif d'ordre élevé, nettement supérieur à l'ordre du modèle MA. Le retour de ce modèle autorégressif au modèle MA se fait par l'estimation de l'entrée (filtrage inverse) et régression du signal MA sur cette entrée (GRENIER, 1981-c, 1983-c). Dans l'autre résolution du modèle MA, on réalise une factorisation de la covariance du signal qui peut être vue comme une extension de la paramétrisation de Schur d'une covariance stationnaire, extension au cas non-stationnaire, avec la contrainte que la solution soit exprimée sur la base des fonctions $f_i(t)$. Cette solution comporte une factorisation d'un signal vectoriel, suivi par une renormalisation pour obtenir le modèle scalaire non-stationnaire (GRENIER, 1983-a).

Plutôt que cette approche en deux temps de la modélisation ARMA, on peut aussi concevoir une identification simultanée de la partie autorégressive et de la partie à moyenne ajustée. Une méthode de maximum de vraisemblance exacte a pu être proposée, assortie d'un calcul du gradient et du Hessien de la vraisemblance, permettant d'utiliser pour sa maximisation une méthode de Newton-Raphson (GRENIER, 1981-a, 1981-b).

Le modèle ARMA ne pouvant à lui seul répondre à l'ensemble des situations survenant dans le traitement du signal, d'autres modèles sont à envisager. L'un d'eux, que par abus de langage ou simplement par analogie on pourrait appeler modèle ARMA déterministe, permet de modéliser un signal

comme la sortie d'un système linéaire excité par une entrée certaine mais inconnue, l'élément aléatoire n'apparaissant que comme un bruit additif sur la sortie du système. Cette modélisation due à J. Le Roux, s'apparente à la méthode de Prony en ce qu'elle fait apparaître une réponse libre d'un système, c'est-à-dire une somme d'exponentielles complexes. Elle trouve son extension dans le cadre non-stationnaire à base de fonctions $f_i(t)$ et se décompose en la solution d'un problème de vecteurs propres généralisés et la solution d'un filtrage de Wiener (GRENIER, 1983-c, 1983-d, 1984).

Une autre variante du modèle ARMA est donnée par le filtre en treillis. Quand le modèle est stationnaire il s'agit uniquement d'une structure de filtre différente de l'habituelle structure transverse: même réponse impulsionnelle, même fonction de transfert, mais implantation différente. Pour un modèle non-stationnaire, l'équivalence entre filtre transverse et filtre en treillis ne se conserve plus à cause de la variation temporelle des paramètres du modèle, ou plutôt cette équivalence s'exprime dans une transformation d'une base de fonctions en une autre. Ce résultat conduit à élaborer une méthode d'estimation adaptée à cette structure de filtre en treillis, par une extension non-stationnaire de l'algorithme de Burg (GRENIER, 1982-a, 1983-c).

Après l'édification de cette méthodologie de modélisation évolutive, le prolongement naturel et imposé de cette étude était la validation des modèles trouvés, validation qui a été menée sur deux plans: dans des simulations et dans une application à des signaux réels. Les simulations ont porté sur des modèles simples (AR, treillis) dans une configuration où la difficulté d'identification vient d'un croisement dans le plan temps-fréquence entre crêtes énergétiques ("formants" si il s'était agi de parole), les

comparaisons portant entre les diverses méthodes évolutives et les méthodes adaptatives (non publié). Cette confrontation a été élargie aux estimateurs non-paramétriques de représentations temps-fréquence ou "reliefs", les modèles évolutifs ou adaptatifs en constituent des estimateurs paramétriques. Les signaux tests ont été des signaux déterministes, sinusoïdes modulées en fréquence (GRENIER, 1983-b), et des signaux aléatoires engendrés par des modèles présentant des sauts et invariants entre les instants de saut (GRENIER, ABOUTAJDINE, 1984).

L'application de cette méthodologie a été faite dans le champ du signal de parole, en poursuivant deux buts, la synthèse, d'une part, et la reconnaissance d'autre part. La synthèse visée à l'origine était une synthèse par diphones et par règles, au sens où un texte phonétique aurait été converti en suite de diphones synthétisés puis concatenés. Ce but était trop ambitieux pour l'état encore balbutiant de la méthodologie "évolutive", et l'étude s'est réorientée vers un stockage sur un nombre réduit d'éléments binaires et une restitution de cette parole. L'entité de base retenue n'a plus été le diphone, mais la syllabe. La reconnaissance de parole a été étudiée dans le cadre d'une reconnaissance de mots isolés, d'abord monolocuteurs, puis dans les prolongements futurs de cette étude, multilocuteurs. La particularité du système mis en oeuvre est de représenter chaque mot par un unique modèle évolutif au lieu des dizaines ou centaines de modèles représentant chaque mot dans les travaux analogues !

L'ordre dans lequel viennent d'être décrites les phases de cette étude est aussi celui dans lequel elles vont être développées dans la suite de ce document: modélisation AR, ARMA, modèles déterministes, en treillis, simulations des méthodes, applications au signal de parole.