CHAPITRE 10. APPLICATIONS.

Il eut été dommage de ne pas conclure cette description d'une nouvelle méthodologie pour le traitement des signaux non-stationnaires, sans décrire son application à un certain nombre de signaux issus de situations réelles et concrètes. Ce but ne pourra malheureusement pas être complètement atteint ici, car les applications mettent aussi en jeu d'autres outils que ceux de la modélisation, ce qui accroît le délai de mise au point de l'ensemble. Au jour où j'écris ces lignes, deux applications sont suffisamment avancées pour que des résultats préliminaires, mais vivement encourageants puissent être présentés. Ce sont d'une part la synthèse de la parole par unités supra-phonémiques, et d'autre part la reconnaissance de mots isolés, mono-locuteur. Le travail sur la synthèse visait d'abord une synthèse par diphones et avait fait l'objet des stages de D.Zone et J.Fang. Il se poursuit maintenant vers une synthèse syllabique, en collaboration avec G.Chollet et constitue le sujet de thèse de M.C.Chevalier. Quant au travail sur la reconnaissance de mots il a été entamé avec D.Aboutajdine (LEESA, Rabat) puis avec G.Ahlbom (KTH, Stockholm) lors de leurs séjours à l'ENST en 1982 pour le premier et 1983-1984 pour le second.

1. Expériences en synthèse de parole.

Puisque les modèles évolutifs sont à même de représenter une transition entre contenus spectraux différents, il nous a semblé naturel de tenter de réaliser une synthèse de parole où chaque modèle serait associé à la transition d'un phonème à un autre, ce qui correspond à une synthèse par diphones. Il y a derrière cela l'hypothèse qu'existent les phonèmes, en utilisant ce

mot non pas comme l'entendent les phonéticiens, mais au sens de zone stable du signal, constituant une réalisation d'un phonème. On peut alors en détectant ces zones à contenu spectral stable au cours du temps, segmenter le signal en leur milieu. L'analyse se fera sur chaque segment commençant et finissant à un des points de segmentation. A la synthèse le problème sera de concaténer correctement les divers segments ou diphones synthétisés. C'est la difficulté à réaliser cette opération qui nous a conduits à nous tourner maintenant vers une synthèse utilisant des entités plus longues que les diphones, en interdisant la segmentation et donc la juxtaposition, au milieu des voyelles. Mais commençons par décrire la première de ces approches.

1.1. Synthèse par diphones.

Il ne s'agit pas d'une véritable synthèse, mais dans un premier temps d'une analyse du signal suivie par une restitution de ce signal à partir des paramètres issus de l'analyse. En court-circuitant ainsi la longue phase d'acquisition du catalogue des modèles, on occulte un certain nombre de difficultés associées au choix du bon représentant de chaque élément du dictionnaire, à la concaténation de modèles issus de contextes très différents ... Mais esquiver ces problèmes permet de se concentrer plus vite sur ceux liés à la détermination des modèles, et qui sont les premiers à résoudre quand on souhaite, comme ici, mettre en oeuvre une méthodologie nouvelle dont on ignore le comportement sur un signal de parole.

Chaque expérience comportera les mêmes phases:

A) Segmentation, c'est-à-dire détection des limites entre diphones. Cela se fera ici en coupant à chaque instant où la courbe de l'énergie du

signal calculée sous une fenêtre glissante passe par un extremum.

B) Analyse de chaque segment, de façon à obtenir un modèle évolutif pour chaque segment. C'est à cette étape que se situe la variable essentielle dans cette expérience, à savoir le choix de la structure du modèle. Celle-ci se caractérise par:

B1) Le choix des ordres de la partie autorégressive et de la partie MA. On peut s'aider pour celui-ci de la connaissance acquise dans la modélisation stationnaire sous fenêtres brèves où l'ordre AR est compris entre 10 et 16, et l'ordre MA le plus souvent pris égal à 0 alors que certains sons (nasales) réclameraient un ordre MA de 4 ou 6.

B2) Le choix de la base de fonctions et de son degré m. Deux bases seront utilisées ici: Legendre et Fourier, mais ceci ne préjuge pas de l'existence d'une base de fonctions idéalement adaptée à la parole. Le degré m est la grande inconnue, il devra être au moins 1, pour avoir avec les deux fonctions $f_0(t)$ et $f_1(t)$ une valeur moyenne et une transition. Nous n'avons pas dans cette expérience dépassé m=5.

B3) Le choix de l'algorithme et du type de modèle. Le modèle peut avoir deux formes, du moins pour sa partie autorégressive qui peut être transverse avec les prédicteurs $a_i(t)$ ou en treillis avec les coefficients de réflexion $k_i(t)$. Un seul algorihme est disponible dans ce second cas alors que trois au moins existent dans le premier, sans compter la variante "à la Prony" dont l'emploi sur un signal de parole est plus délicat. On retiendra pour son compromis

qualité de l'estimateur / coût de calcul, la méthode de "corrélation".

C) Détection du fondamental, soit par une mesure de sa hauteur F_0, soit par une localisation temporelle des impulsions.

D) Concaténation de la succesion des modèles pour reconstituer la continuité de l'évolution du modèle sur toute la phrase.

E) Synthèse (ou plutôt restitution) du signal par passage d'une excitation (impulsions ou bruit blanc selon le voisement) dans le système reconstitué par concaténation.

Quelles étapes de ce processus ont engendré des difficultés specifiques aux modèles non-stationnaires? Sûrement pas les étapes A ou C, indépendantes de la modélisation. L'étape B ne présente pas plus de difficultés que dans les simulations décrites au chapitre précédent. On prendra seulement la précaution de pré-accentuer le signal de parole, pour réduire sa dynamique spectrale et améliorer le conditionnement de la matrice de covariance du signal Y_t des projections sur la base. La pré-accentuation retenue a été la même que dans la LPC classique: on remplace le signal y_t par sa différence première $y^*_t = y_t - y_{t-1}$ (le signal est échantillonné à 8 Khz, sur 12 e.b). Les ordres et le degré de la base ont été fixés pour chaque expérience, en les choisissant dans les limites données ci-dessus. Les meilleurs résultats avaient été obtenus pour la base de Legendre, un modèle AR(12) sur les fonctions $f_0(t)...f_4(t)$. Mais les difficultés essentielles se sont situées dans les phases D et E, de concaténation et de synthèse.

1.1.1. Stabilisation du modèle.

Dans la phase de synthèse, le problème rencontré est un problème de stabilité du modèle non-stationnaire. Aucun résultat de la partie théorique ne garantit en effet la stabilité du modèle autorégressif évolutif que l'on estime sur un signal. Il est même facile de construire des signaux tels que si on en sélectionne une portion, le modèle estimé sur celle-ci est instable. La stabilité se définit, rappelons le comme une absolue sommabilité de la réponse impulsionnelle h(t,s) du modèle, et ceci pour tout t, la sommation se faisant sur s. Dans la pratique, en analysant un signal qui s'annulle aux deux extrèmités de la fenètre, on est assuré de trouver un modèle stable. Le modèle obtenu par concaténation des modèles estimés sur une phrase répond à cette condition, puisque le signal part du silence et y revient. Pourtant, il existe une autre conception de la stabilité, qu'il s'avère nécessaire d'introduire pour rendre compte des phénomènes observés dans la synthèse, il s'agit de la stabilité locale décrite en annexe 8. Elle impose une forme de stabilité pour chacun des modèles tangents au modèle non-stationnaire, à l'instant t quand t parcourt l'intervalle complet où se déroule le signal.

Pourquoi doit-on imposer cette stabilité locale ? La raison en est que si le modèle tangent devient instable sur un petit intervalle de temps (t_1,t_2), et même si la stabilité du modèle non-stationnaire n'en est pas affectée, le signal synthétisé présentera sur l'intervalle (t_1,t_2) un comportement explosif, son amplitude divergera, ce qui créera une petite boufée d'énergie à un niveau plusieurs fois supérieur à l'énergie moyenne avant et après cet incident. Il s'avère que l'oreille est très sensible à ces variations brèves d'amplitude qu'elle analyse comme des bruits de choc, ou des

explosions. Il est donc nécessaire d'éliminer ces accidents.

La méthode pour forcer la stabilité du modèle sans déformer son contenu spectral consiste à remplacer à tout instant où il est instable le modèle tangent par un modèle équivalent où les pôles stables sont inchangées, mais où les pôles instables sont stabilisés par une inversion qui les renvoie à l'intérieur du cercle unité. Pour être en accord avec la notion de stabilité locale, il est indispensable de travailler avec la forme observable de l'équation d'état. Si A_t est la matrice de dynamique dans cette équation, on travaille sur le polynôme $A_t(z)=Det(zI-A_t)$, et on le factorise:

$$(2\text{-}203)\quad A_t(z)=\prod_{i=1}^{p}(z-p_i(t))$$

Puis on transforme chaque pôle $p_i(t)$ par une fonction M qui renvoie en miroir les pôles du dehors vers le dedans du cercle unité:

$$(2\text{-}204)\quad M(p)=\begin{cases} p & \text{si } |p|<1 \\ p^{-1} & \text{si } |p|>1 \end{cases}$$

On forme alors le polynôme stabilisé $A_t^*(z)$:

$$(2\text{-}205)\quad A_t^*(z)=\prod_{i=1}^{p}(z-M(p_i(t)))$$

La matrice A_t sous forme compagne s'en déduit par une simple écriture. La factorisation (2-203) s'obtient par toute procédure standard de calcul des racines d'un polynôme.

Une autre façon de stabiliser le modèle consiste à factoriser en partie à phase minimale et partie à phase maximale le produit $A_t(z)A_t(z^{-1})$, en utilisant la décomposition de Schur (décrite au chapitre 7). On obtient ainsi $A_t^*(z)$ tel que $A_t^*(z)A_t^*(z^{-1})=A_t(z)A_t(z^{-1})$, et le module de la fonction de transfert est le même entre A_t et A_t^*, mais A_t^* possède tous ses pôles à

l'intérieur du cercle unité. Cependant cette factorisation échoue si un ou plusieurs des pôles de $A_t(z)$ se situe sur le cercle unité. Il faut alors recourir à la procédure précédente ou faire une transformation du genre chirp-z. De toute manière, ce sont des cas où le contenu spectral du modèle devra être modifié pour respecter la propriété de ρ-contraction (annexe 8): un pôle sur le cercle unité devra être remplacé par un pôle ayant même argument mais un module de $1-\epsilon$ où par exemple ϵ sera la précision du calculateur mis en jeu.

1.1.2. Concaténation des modèles.

La concaténation des modèles estimés (phase D) ne peut pas se faire sans précautions. On risquerait en effet de raccorder des modèles évolutifs en créant une discontinuité dans le modèle résultant si la valeur finale d'un modèle n'était pas égale à la valeur initiale du modèle qui suit. L'effet de cette rupture sera désastreux lors de la synthèse, en introduisant à partir de l'instant de commutation vers le second modèle un régime transitoire, le temps que s'établisse un accord (c'est là une description imagée mais qualitativement juste) entre les modes du nouveau modèle et ceux qui étaient présents dans l'état du filtre par suite de l'action du modèle antérieur. Ceci provoque dans le signal la présence d'un accident qui est très audible, d'où la nécessité de trouver un moyen pour effacer cette difficulté.

Une première méthode a été essayée et consistait à créer une zone intermédiaire où les deux modèles à raccorder étaient valides simultanément. Ainsi si le premier modèle était utilisé jusqu'à l'instant T, le second l'était à partir de T-N. Le nombre d'échantillons sur lesquels se faisait le recouvrement a varié de N=50 à N=400. Le modèle était alors interpolé du

premier au second sur l'intervalle (T-N,T). Ainsi le coefficient $a_i(t)$ s'obtenait à partir de $a_{i1}(t)$ et de $a_{i2}(t)$ grâce à une fonction g(t) telle que:

(2-206) $a_i(t)=g(t)a_{i1}(t)+(1-g(t))a_{i2}(t)$

où g est défini par:

$$(2\text{-}207)\quad g(t)=\begin{cases}1 & t<T-N\\ \cos(\frac{\pi}{N}(t-T-N)+1)/2 & T-N\leq t\leq T\\ 0 & T<t\end{cases}$$

Une telle méthode semble mal adaptée pour plusieurs raisons. Tout d'abord interpoler comme en (2-206) entre deux modèles stables peut conduire à des trajectoires surprenantes pour les pôles du modèle tangent, qui peut devenir instable et qui le devient fréquemment. Ensuite le contenu spectral, ou si l'on veut le relief ne s'obtient pas par une interpolation entre les deux reliefs, ce qui fait intervenir temporairement des résonances parasites à des fréquences nuisibles. Enfin le compromis à réaliser pour fixer N est peu satisfaisant, si N est trop faible, la transition est brutale et donc riche en accidents, si pour éliminer ces accidents on accroît N, on les voit effectivement se raréfier, mais l'évolution temporelle des modèles devient de moins en moins fidèle au signal, ce qui dégrade la qualité de la synthèse.

Une seconde méthode a été mise en oeuvre pour réaliser ce lissage des trajectoires du modèle lors de la concaténation. Puisque les ennuis viennent des erreurs d'estimation des paramètres aux deux extrémités de la fenêtre d'analyse, l'idée était de contraindre le modèle estimé à prendre des valeurs données à priori à l'une ou aux deux extrémités de l'intervalle temporel. Si par exemple la base de fonctions est $f_j(t)=(t/T)^i$, alors a_{ik}

peut s'interpréter comme la dérivée k-ième de $a_i(t)$, à l'instant t=0:

$$(2\text{-}208)\quad \left[\frac{\partial^k}{\partial t^k}a_i(t)\right]_{t=0}=\left[\sum_{j=0}^{m} a_{ij}\frac{\partial^k}{\partial t^k}(t/T)^j\right]_{t=0}=\sum_{j=0}^{m} a_{ij}j(j-1)\ldots(j-k+1)T^{-j}\delta_{j,k}$$

$$=k(k-1)\ldots2.1.T^{-k}a_{ik}$$

Si alors on fixe les k premières dérivées de $a_i(t)$ à l'instant t=0, ceci fixe $a_{i0}\ldots a_{ik}$, et l'identification du modèle autorégressif se fait en minimisant la variance de ϵ_t:

$$(2\text{-}209)\quad \epsilon_t=y_t+\left[Y_{t-1}^{T(1)}\ldots Y_{t-p}^{T(1)}\right]\theta_1+\left[Y_{t-1}^{T(2)}\ldots Y_{t-p}^{T(2)}\right]\theta_2$$

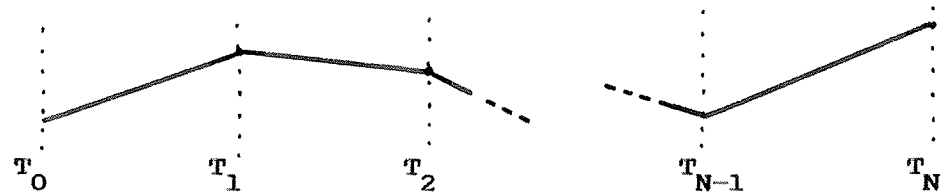
Ceci conduit, en décomposant le vecteur Y_t des projections de y_t sur la base en ses composantes de 0 à k soit $Y_t^{(1)}$ et celles de k+1 à m, soit $Y_t^{(2)}$ et en prenant pour θ_1 le vecteur des contraintes a_{ij}, j=0...k, tandis que θ_2 est le vecteur des inconnues a_{ij}, j=k+1...m, à une équation d'optimalité:

$$(2\text{-}210)\quad \sum_{t\in\Gamma}\begin{bmatrix}Y_{t-1}^{(2)}\\ .\\ Y_{t-p}^{(2)}\end{bmatrix}\left[Y_{t-1}^{T(2)}\ \ldots\ Y_{t-p}^{T(2)}\right]\theta_2=-\sum_{t\in\Gamma}\begin{bmatrix}Y_{t-1}^{(2)}\\ .\\ Y_{t-p}^{(2)}\end{bmatrix}\left(y_t+\left[Y_{t-1}^{T(1)}\ldots Y_{t-p}^{T(1)}\right]\theta_1\right)$$

On voit ainsi rendue possible une analyse où les conditions initiales de chaque modèle sont fixées par les conditions finales du modèle précédent, en progressant dans le sens du temps, mais il n'est pas difficile de retourner cette approche en travaillant de façon rétrograde sur le temps, les conditions initiales d'un modèle fixant le modèle précédent à son instant final. La première de ces éventualités a été seule employée, mais elle souffre d'un grave défaut qui est une accumulation des erreurs d'une fenêtre d'analyse à la suivante. En effet si la valeur initiale d'un modèle

est erronée du fait de l'estimation du précédent modèle, l'erreur de prédiction sera forte au début de la fenêtre, et non modifiable à cause des contraintes. L'estimateur tendra alors à privilégier les zones centrales de la fenêtre où l'erreur de prédiction sera minimale, mais en contre-coup, sauf si la base de fonctions compense ce phénomène, le modèle en fin de fenêtre réintroduira une erreur en sens inverse de l'erreur initiale. Quand on visualise les trajectoires des $a_i(t)$ on voit ainsi apparaître des écarts de plus en plus grands aux instants où les modèles se raccordent. Ce que l'on gagne en qualité de synthèse en forçant les trajectoires à se raccorder, on le reperd donc par les erreurs qui s'introduisent, ce que confirme l'écoute.

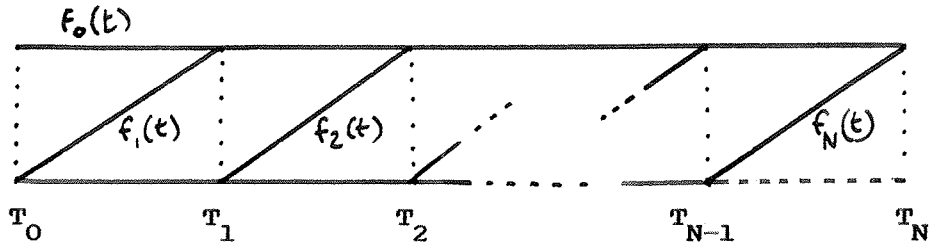
Un meilleur emploi des contraintes serait assurément en combinant l'emploi des contraintes en sens direct et en sens rétrograde par une procédure de programmation dynamique travaillant sur toute la phrase à synthétiser. Pour faire un tel calcul, il est en réalité plus efficace de travailler non pas fenêtre par fenêtre, mais globalement sur l'ensemble de la phrase. On peut en effet réorganiser les bases de fonctions de chaque intervalle, en les fondant en une seule pour les rendre compatibles avec les contraintes. Montrons comment, sur un exemple où on supposera que (T_0,T_N) est l'intervalle total d'analyse, segmenté aux instants $T_1 \ldots T_{N-1}$. On supposera aussi que les paramètres $a_i(t)$ évoluent linéairement entre les instants T_k et sont continus.



On représentera avantageusement ce modèle sur la base définie de T_0 à T_N

$$(2\text{-}211)\quad f_j(t)=\begin{cases}0 & t<T_{j-1}\\ \dfrac{t-T_{j-1}}{T_j-T_{j-1}} & T_{j-1}\leqslant t\leqslant T_j\\ 1 & T_j<t\end{cases}$$

pour j=0...N.



Il faut noter que la méthodologie présentée à la faveur d'un exemple est très générale. Il est toujours possible de réorganiser les fonctions $f_i(t)$ par un changement de base dont la matrice de passage soit invariante avec t, et qui isole les contraintes comme certaines des composantes sur cette base, puis par un réarrangement de l'ensemble des fonctions sur la réunion de tous les intervalles de ramener le problème de minimisation à un problème sans contrainte, excepté aux deux bornes de l'intervalle total. Une seconde remarque à faire est que le calcul du modèle concaténé ne requiert que les autocorrélations (ou les covariances) du vecteur Y_t sur chaque intervalle, ce qui permet même cette procédure de concaténation sous contrainte dans le cadre d'une synthèse segmentale où le dictionnaire consisterait non pas en modèles a_{ij} mais en corrélations des Y_t. Les opérations de changement de base par une matrice P transforment une corrélation R_k en PR_kP^T, tandis que la concaténation transforme les corrélations en les sommant tout simplement. Il faut encore remarquer que ce recours aux corrélations évite de réaliser effectivement le filtrage de Y_t qui apparait dans le second membre de (2-210) où y_t est corrigé par θ_1.

1.2. Synthèse par syllabes.

Cette approche par contraintes n'a pas été poursuivie, puisqu'elle s'apparente à un accroissement simultané du nombre des fonctions et de la longueur de l'intervalle traité. Il a semble plus direct de réaliser cet accroissement par un choix d'une unité plus longue que le diphone. La segmentation est donc modifiée pour détecter de plus longues entités. Ceci s'obtient en ne détectant que les minima de l'énergie, et pas les maxima. On obtient ainsi une segmentation approximativement syllabique. L'intérêt essentiel est de permettre de faire une concaténation brutale, en espérant que les accidents dus à une continuité, s'ils se produisent soient peu audibles, se situant dans des zones de faible énergie du signal. La stabilisation reste inchangée. Les résultats, à l'écoute de cette synthèse semblent beaucoup plus satisfaisants que précédemment. Les fenêtres ont une durée moyenne comprise entre 800 et 2000 échantillons, les ordres des modèles sont en moyenne de 12, avec 3 à 6 fonctions dans la base de Fourier. Ceci assure un codage du contenu spectral du signal sur environ 250 paramètres par seconde, ce qui est à mettre en parallèle avec les 500 paramètres par seconde que représent la LPC-10 (prédiction linéaire à l'ordre 10, un modèle toutes les 20 ms). Ceci fait espérer un gain probable d'un facteur 2 sur cette méthode, avec une qualité analogue sinon meilleure, et encourage à poursuivre ces expériences.

2. Reconnaissance de mots isolés

La deuxième série d'expérience mise en place pour tester l'adéquation de cette méthodologie aux signaux de parole est une reconnaissance de mots isolés. L'idée en est que si un modèle évolutif peut représenter une succession de phonèmes avec les transitions de l'un à l'autre, on doit pouvoir

se servir du vecteur des paramètres calculés comme d'une forme représentative du groupe de phonèmes prononcés, exactement comme un modèle stationnaire sur une petite portion de signal est représentatif du phonème étudié. La validation de cette idée a été entreprise dans un cadre assez restreint qui laisse subsister un doute sur la généralité du résultat obtenu, mais dans ce cadre, les résultats ont été excellents. Il s'agissait d'une expérience de reconnaissance de chiffres de "zéro" à "neuf" prononcés par un seul locuteur, à six reprises, ce qui donne 60 occurrences. Chaque occurrence est testée en reconnaissance sur un système appris sur les 59 autres, avec une procédure de plus proche voisin, qui a tendance, il faut le reconnaître à surévaluer les scores obtenus (aucune erreur). Une fois fixée la représentation, ici modèle évolutif, deux questions se posent dans ce contexte de reconnaissance de mots isolés: quelle mesure de distance entre modèles doit-on adopter, et comment prendre en compte la variation de durée des occurrences, ou des phonèmes au sein de chaque occurrence ?

Les mesures de distance possibles sont nombreuses. Parmi celles déjà citées il faut retenir l'écart quadratique moyen entre les reliefs issus des deux modèles à comparer, avec transformation logarithmique de ces reliefs pour obtenir un écart en dB. La métrique définie par (2-199), c'est-à-dire la moyenne temporelle sur la norme euclidienne du vecteur d'erreur entre les modèles, est aussi utilisable dans ce problème. Ces deux métriques ont été employées dans cette expérience avec un succès égal, ce qui tient à la corrélation importante entre elles. Une troisième métrique ou plutôt une mesure de distance, envisageable est calquée sur l'idée de vocodeur à canaux adaptés (GUEGUEN, FARJAUDON, LE CHEVALIER, CARAYANNIS, 1975). Lorsqu'on veut comparer une occurence à plusieurs modèles, et déterminer lequel en est le plus proche, il est efficace dans le cas stationnaire de filtrer en

inverse le signal par chaque modèle. Le résidu obtenu, interprété comme une erreur de prédiction est minimal pour le modèle qui est le mieux adapté, et donc le plus proche. Dans le cas de modèles et de signaux non-stationnaires, il est légitime de penser que ce critère lié à l'énergie résiduelle devrait fonctionner aussi bien, vu la qualité des modèles obtenus par la minimisation de cette énergie résiduelle. Mais ceci est une conjecture qu'aucun résultat expérimental n'est encore venu confirmer ni infirmer.

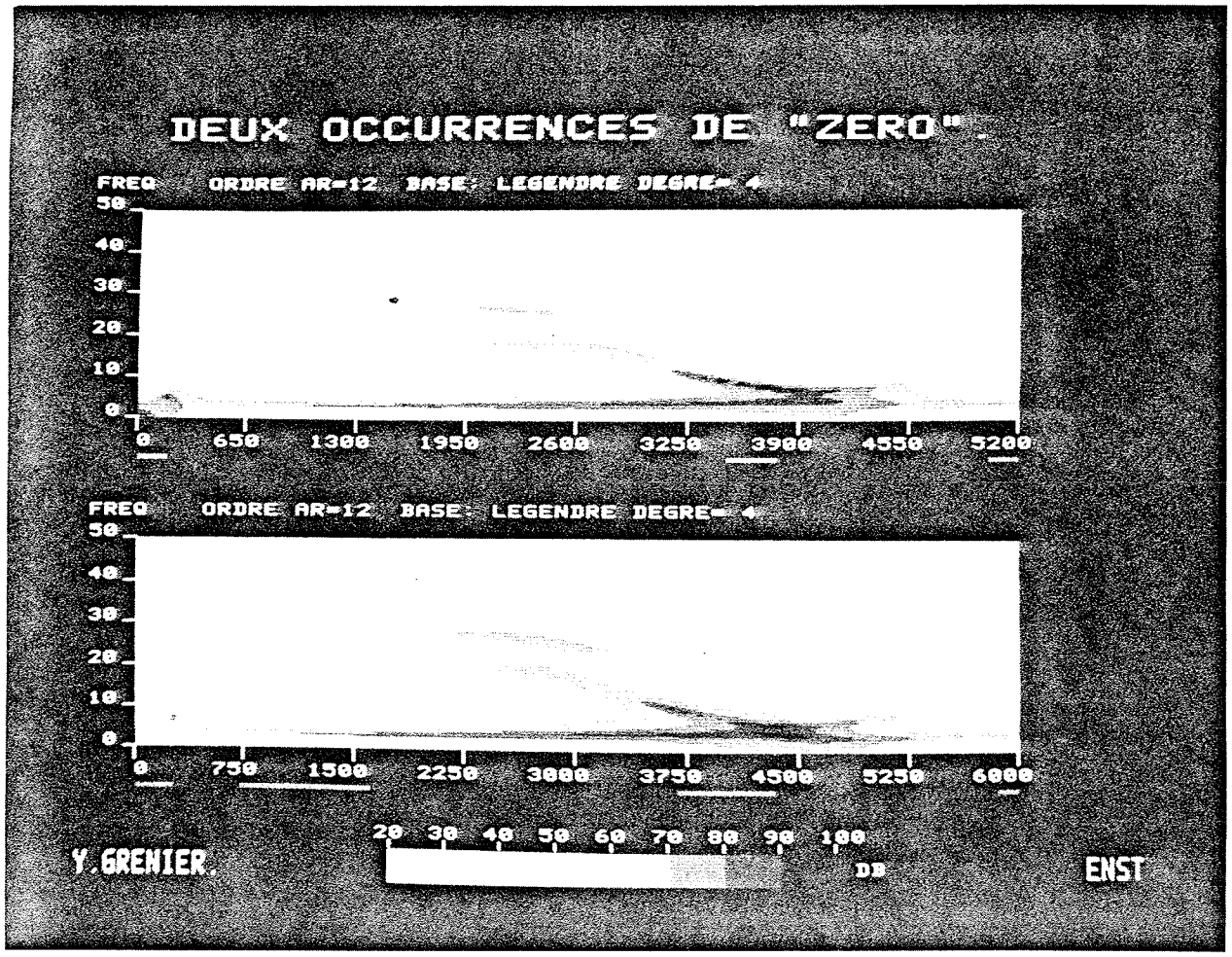


Figure 2.39. Reliefs de deux occurences du mot "zéro", modèle AR(12), base de Legendre, 5 fonctions.

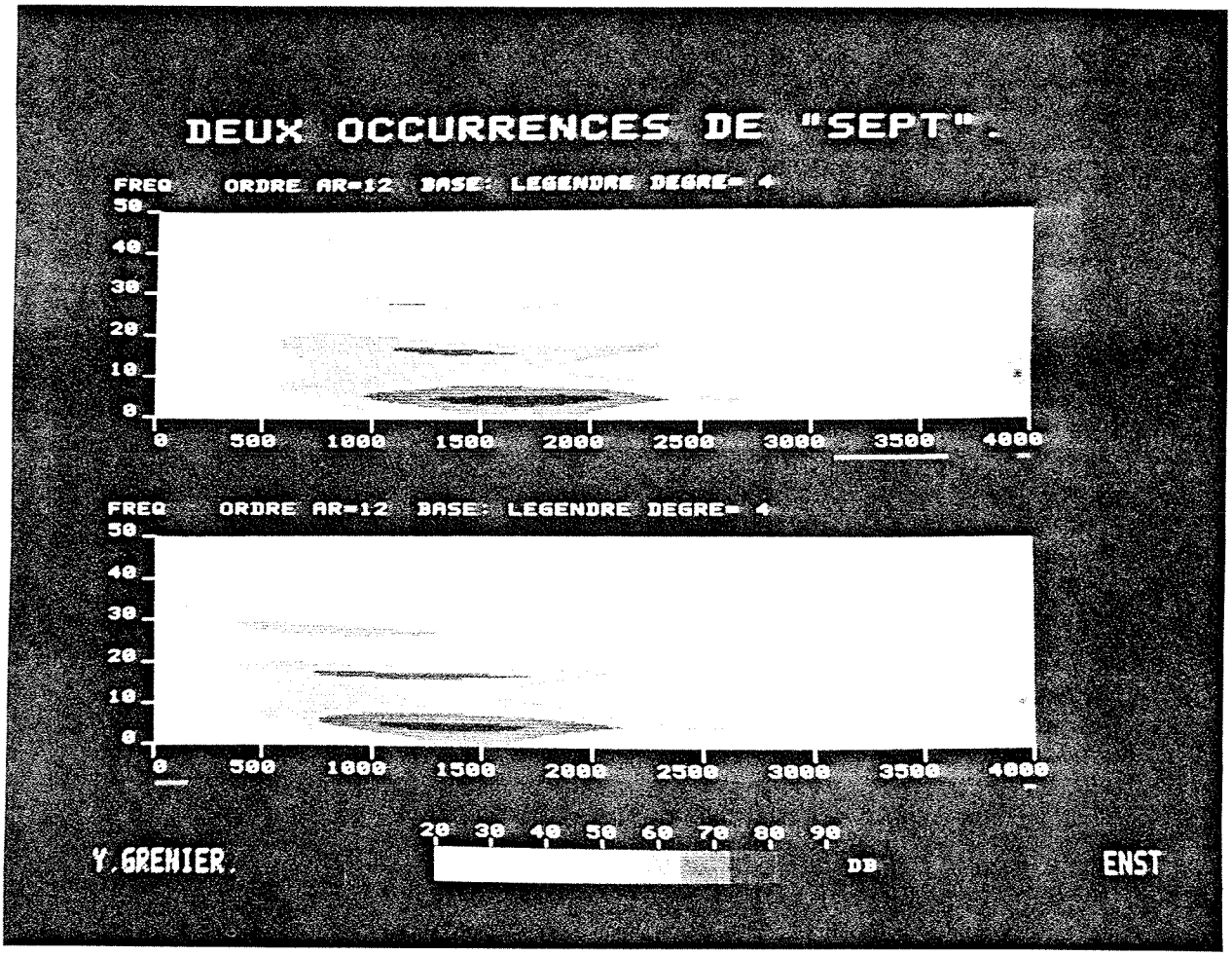


Figure 2.40. Relief de deux occurences du mot "sept", même analyse que pour la figure 2.39.

Les figures 2.39 et 2.40 donnent deux exemples de reliefs obtenus par modélisation évolutive sur des mots issus du corpus traité. On y constate l'analogie entre occurrences du même mot, et la dissemblance entre occurrences de mots différents. Mais on y observe aussi les déformations temporelles sur des occurrences du même mot, par variation soit de la durée totale, soit du rythme à l'intérieur de cette durée. Une comparaison correcte entre modèles se doit d'inclure ces variations dans le calcul de la distance pour les exclure du résultat. Pour ce qui est de la durée totale, il suffit simplement de se souvenir que les fonctions des bases (Legendre, Fourier...) sont définies avec une variable sans dimension t/T où T est la durée totale. Il devient clair alors que le vecteur θ des paramètres a_{ij} spécifie un modèle de durée arbitraire et que si on veut faire coïncider les

extrèmités de deux occurences par une homothétie sur l'échelle des temps, il n'y a pas à modifier les vecteurs θ, mais uniquement à recalculer les bases de fonctions, la normalisation de la durée est automatique.

Les variations de rythme internes à une occurrence, sont en général prises en compte par une procédure de comparaison dynamique, qui calcule sous forme d'une ligne polygonale le graphe de la fonction $t'=g(t)$ associant l'échelle de temps t d'une occurrence à l'échelle de temps t' de la seconde occurrence. Une telle procédure s'applique sans transposition dans le cas présent, mais elle est peu satisfaisante car elle néglige totalement l'information sur l'évolution temporelle des modèles, véhiculée par les fonctions de la base. On peut plutôt chercher à localiser la fonction $g(t)$ à l'intérieur d'une famille de fonctions, indicée par un ou plusieurs paramètres, paramètres que l'on optimisera pour trouver l'anomorphose temporelle $t'=g(t)$ minimisant la distance de deux occurrences. Ceci a une grand intérêt si il est possible d'exprimer les fonctions $f_j(t')$ sur la base des $f_i(t)$, en écrivant:

$$(2\text{-}212)\quad f_j(t')=f_j(g(t))=\sum_{k=0}^{M}\phi_{jk}f_k(t)$$

Les poids ϕ_{jk} dépendront des paramètres de g. Les bases ne se prètent pas toutes à ce jeu. Les seuls exemples pour l'instant concernent la base de Legendre, et $g(t)$ y est une fonction quadratique ou cubique de t. Les fonctions $g(t)$ polynômes de degré supérieur à 3 sont envisageables, mais peut-être trop complexes.

Dans le premier exemple, on supposera que la variable temps réduite est comprise entre -1 et $+1$, ce qui simplifie l'écriture des polynômes de Legendre. On pose alors:

(2-213) $t'=g(t)=t+\lambda(t+1)(t-1)$

Ceci assure que $g(1)=1$ et $g(-1)=-1$, afin de laisser le début et la fin des occurrences en coïncidence. La fonction $g(t)$ n'est acceptable que si elle est monotone, c'est-à-dire si sa dérivée reste positive. Le calcul est direct, et montre que ceci impose $-1/2\leqslant\lambda\leqslant 1/2$.

Le calcul des ϕ_{jk} n'est pas difficile à partir de l'expression des polynômes de Legendre, par simple développement de $f_j(t')$. On remarque que M doit être pris égal à $M=2m$, et que $\phi_{jk}=0$ si $k>2j$.

Le second exemple met en jeu une foncton cubique $g(t)$. En imposant les conditions $g(-1)=-1$ et $g(1)=1$, on restreint la classe des polynômes de degré 3 à la sous-classe indicée par les deux paramètres λ et μ:

(2-214) $t'=g(t)=t+(\lambda t+\mu)(t^2-1)$

Imposer à $g(t)$ d'être monotone croissante sur l'intervalle $(-1,+1)$ fixe des bornes pour le domaine admissible de λ et μ:

$-1/2\leqslant\lambda\leqslant 1$

$-(\lambda+1/2)\leqslant\mu\leqslant\lambda+1/2$ si $-1/2\leqslant\lambda\leqslant 1/4$

$\mu^2\leqslant 3\lambda(1-\lambda)$ si $1/4\leqslant\lambda\leqslant 1$

Une fonction cubique du type (2-214) est intéressante car elle permet d'avoir un mélange complexe de contractions et d'expansions du temps (C-E, C-E-C, E-C, E-C-E) d'une occurrence relativement à l'autre. Le calcul des ϕ_{ij} n'offre pas plus de difficulté et n'est pas moins fastidieux que dans le cas précédent. Ici $M=3m$ et $\phi_{jk}=0$ si $k>3j$.

Le calcul de l'anamorphose optimale nécessite ensuite la minimisation

de la distance (on prendra celle définie par (1-199) de préférence) relativement à λ ou λ et μ. Puisque les ϕ_{jk} sont des polynômes en λ et μ de degré au plus égal à m en chaque variable, la distance sera aussi un polynôme en λ et μ, mais de degré au plus 2m en chaque variable. On pourra alors avoir recours à toute procédure standard (par exemple programmation non linéaire) pour minimiser cette distance. Cette partie de l'expérience n'a pas encore été réalisée, son urgence ne se faisant sentir que pour le passage à une reconnaissance multilocuteurs.

Ces deux expériences en synthèse et en reconnaissance de la parole montrent, en dépit de leur aspect préliminaire, combien les modèles évolutifs peuvent se révéler efficaces sur des signaux non-stationnaires fort complexes. D'autres applications font apparaître de tels signaux, caractérisés par un contenu spectral variant sans ruptures, mais éventuellement rapidement, sur des signaux géophysiques, sismiques, en radar ... La méthodologie décrite dans cette seconde partie devrait y trouver de futur champs d'investigation.