



RECONNAISSANCE DES FORMES DANS LE PLAN TEMPS FRÉQUENCE PAR MODÉLISATION PARAMÉTRIQUE.

J.F. TROUILHET - I. COUSTAUX - J.P. GUILHOT

Laboratoire d'Acoustique de Métrologie et d'Instrumentation
38, Rue des 36 Ponts 31400 TOULOUSE - FRANCE.

RÉSUMÉ

La reconnaissance automatique de signaux nécessite une étape de classification à partir de paramètres caractéristiques de ce signal. Si ces signaux sont non stationnaires, une prise en compte de l'évolution temporelle devient nécessaire. Pour des raisons de performances, on a tout intérêt à travailler sur un nombre restreint de paramètres. Ainsi, la méthode présentée ici permet de respecter ces impératifs puisqu'elle fournit des informations sur l'évolution temporelle des paramètres issus du signal à reconnaître à partir d'une modélisation paramétrique. Afin de situer les performances d'une telle modélisation, nous présentons les résultats obtenus pour un problème de reconnaissance.

1. Introduction

Nous présentons dans cet article la méthode de modélisation retenue pour la réalisation d'un dispositif de reconnaissance automatique de signature acoustique [7].

Le but de ce dispositif est de permettre de quantifier aisément la fréquentation des frayères de l'alose (*Alosa alosa* L. : Un poisson migrateur). Pour cela, il n'existe que la possibilité de compter sur place (ou en temps différé à partir d'enregistrements) le nombre de pontes, caractérisées par un son particulier produit par la nage des poissons venus momentanément en surface.

Près d'une quinzaine de sites devraient être étudiés chaque nuit de la saison migratoire. Jusqu'à présent les études étaient limitées à deux sites, pendant les phases de fraie (de minuit au lever du jour pendant les mois de mai à juillet), un magnétophone fonctionnait sur les lieux de ponte. Ensuite un chercheur comptabilisait les bruits pour estimer la fréquentation des frayères moyennant la correction des pertes d'information liées à l'enregistrement.

Le dispositif réalisé permet un dépouillement automatique en temps réel des enregistrements et permet donc d'étudier l'ensemble des sites.

Les sons provoqués par ces poissons correspondent à des signaux difficiles à reconnaître, puisque leurs dates d'apparition sont aléatoires, ils surviennent dans un environnement bruité par un très grand nombre de sons différents et ils sont non stationnaires. De plus ce n'est jamais le même couple de poissons qui produit le son, il n'y a aucune raison pour que les sons soient identiques d'une ponte à l'autre. Les bruits sont dus, en majeure partie, aux véhicules routiers, aux trains, et enfin, aux animaux nocturnes : Oiseaux, grillons, grenouilles ...

ABSTRACT

The automatic signal recognition require a classification step from the signal characteristic parameters. If these signals are non stationary it is necessary to consider their temporal evolution. For performance reasons we have to work on a limited number of parameters. The shown method allow to expect these imperatives as it gives the informations on the temporal evolution of the parameters coming from the signal from a parametric modelisation.

In order to situate the performances of a such modelisation we show the obtained results for a recognition problem.

2. Reconnaissance de signature acoustique.

La reconnaissance automatique de signatures acoustiques complexes peut être, dans certains cas, efficacement résolue par l'observation de la représentation temps-fréquence du signal. Les méthodes de modélisation généralement utilisées en traitement du signal s'appliquent à la classe des signaux stationnaires. Toutefois, la modélisation d'un signal non stationnaire à partir des méthodes précédentes est possible durant l'intervalle de quasi-stationnarité du signal, mais au lieu d'un vecteur modèle unique on doit en prendre plusieurs en compte afin de ne pas perdre l'évolution temporelle de ceux-ci.

Conformément à ce principe, à partir des vecteurs modèles successifs calculés à l'aide d'un algorithme de modélisation, on peut obtenir une représentation temps-fréquence du signal à reconnaître. Cette représentation, entièrement définie par une matrice contenant les vecteurs modèles correspondants à chaque tranche de temps, se prête bien à la classification. Mais lorsque le temps de réalisation du phénomène devient long devant l'intervalle de quasi-stationnarité du signal, le nombre de paramètres participant à la classification devient trop important.

Ainsi, afin de réaliser une compression de l'information et de prendre en compte l'évolution temporelle du spectre, nous proposons de réaliser une approximation polynomiale du spectre, au sens des moindres carrés, pour certaines valeurs de la fréquence. Après élimination des coefficients non pertinents pour la classification désirée, on obtient une représentation peu encombrante pour la résolution du problème.



3. Modélisation paramétrique et espace temps-fréquence

A condition de respecter le compromis entre la durée minimale de la fenêtre d'observation permettant une estimation avec une variance encore réduite, et la durée maximale sous laquelle l'hypothèse de stationnarité du signal reste raisonnable, l'estimation des coefficients du modèle auto-régressif par l'algorithme de Levinson-Durbin peut donner d'excellents résultats si le signal ne nécessite pas un ordre de modélisation trop élevé. On ne peut pas calculer le biais et la variance de l'estimation. Mais dans le cas de l'analyse spectrale, on peut montrer expérimentalement que l'on obtient une meilleure résolution pour l'estimation du spectre à partir de cette modélisation qu'à l'aide de la transformée de Fourier [1]. La relation entre le spectre du signal et les coefficients AR est :

$$S_i(t_i, f) = \frac{\sigma_{B_i}^2}{\left| 1 + \sum_{m=1}^p a_{mi} e^{j2\pi f m} \right|^2} \quad (3.1)$$

où $V_i(t_i) = [\sigma_{B_i}^2, a_{1i}, a_{2i}, \dots, a_{pi}]$ (3.2)

En outre la modélisation auto-régressive permet d'effectuer un lissage du spectre en jouant sur l'ordre de la modélisation.

La figure ci-dessous représente l'évolution spectrale d'un signal en fonction du temps. Si la modélisation choisie est AR, la valeur du spectre pour un instant t_i donné est :

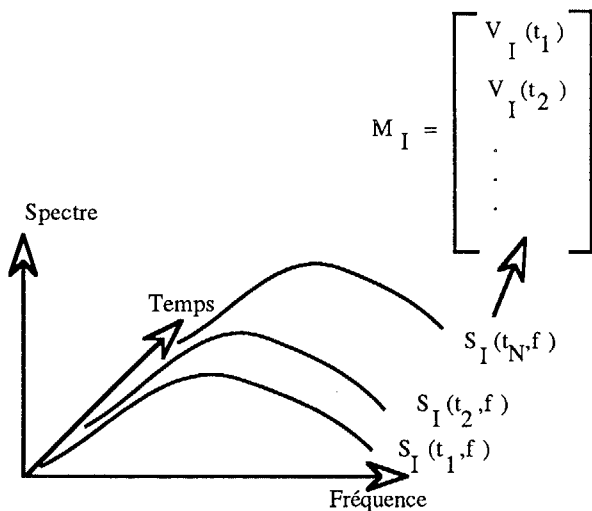


Figure n°1 : Représentation temps fréquence à partir d'une modélisation paramétrique.

M_I est la matrice contenant la totalité des vecteurs modèles. L'utilisation d'un tel modèle pour la classification devient rapidement utopique. En effet, dès que le temps devient grand devant l'intervalle de quasi-stationnarité, le nombre de vecteurs du modèle devient prohibitif.

Nous présentons une alternative à ce problème en utilisant une autre forme de modélisation qui fournit $S_{II}(t, f_i)$, c'est-à-dire le spectre en fonction du temps pour une fréquence donnée. Il

suffit à partir des vecteurs de la première modélisation V_I et par approximation polynomiale, au sens des moindres carrés, de calculer les vecteurs modèles V_{II} correspondant à la nouvelle modélisation.

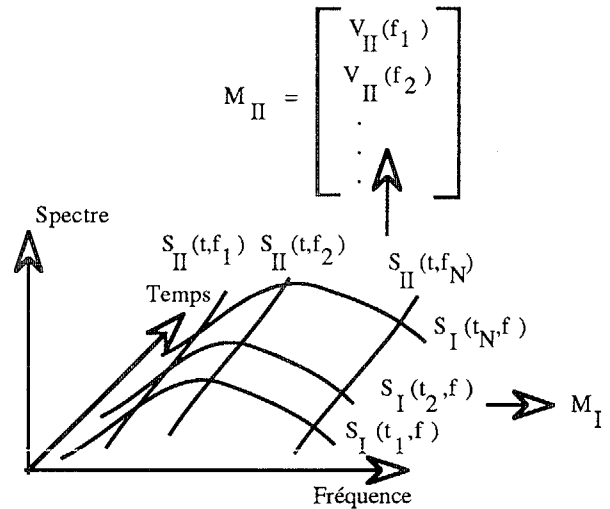


Figure 2 : Correspondance entre les deux types de modélisation dans le plan temps - fréquence.

Lors de la transformation, les vecteurs $V_{II}(f_i)$ doivent être suffisamment nombreux pour ne pas perdre d'information. Ce point n'est vraiment gênant que dans le cas de signaux comportant une fréquence pure. D'autant plus que la phase d'analyse du signal à reconnaître aura permis de déterminer les bandes de fréquence caractéristiques.

4. Reconnaissance des formes par classification des modèles

La supériorité de la deuxième modélisation pour la classification étant démontrée, on peut encore comprimer l'information en fonction de l'univers des signatures à classer. En effet avant l'étape d'apprentissage, les résultats d'une analyse factorielle discriminante permettront d'éliminer les coefficients non pertinents pour la classification. La répartition des éléments utilisée est le classement de la base d'apprentissage effectué par un expert humain.

Ainsi, le fait d'affecter le modèle obtenu après projection à la classe la plus proche au sens de la distance choisie, revient à effectuer le classement par reconnaissance du spectre dans le plan temps-fréquence.

5. Application à la reconnaissance de signatures acoustiques

Afin d'illustrer la méthode présentée ci-dessus, nous allons décrire son implémentation dans un cas pratique sur une machine compatible IBM PC équipée d'une carte de traitement numérique du signal à base d'ADSP 2100 : La reconnaissance du son produit par la fraie de l'aloise.

L'étape d'analyse du signal a révélé la présence de deux bandes principales d'énergie dans le spectre.

Le traitement en vue de la reconnaissance peut être décomposé en trois étapes dont voici un schéma :

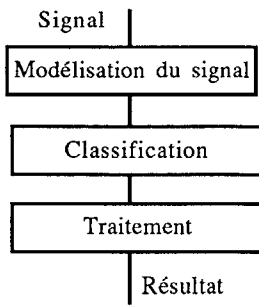


Fig. n°3: Traitement effectué pour la reconnaissance de signature acoustique

A chaque étape correspond le choix d'une solution. Ainsi les performances obtenues en fonction des diverses solutions doivent être évaluées afin de réaliser le choix optimal. Pour cela, il a été nécessaire d'adopter une méthodologie de test rigoureuse afin que la comparaison ait un sens. Les bandes magnétiques ont été recopiées avec l'adjonction de marques de présence sur la deuxième voie. Ainsi les résultats fournis par le système pourront être comparés avec la référence que constitue le marquage sur la deuxième voie.

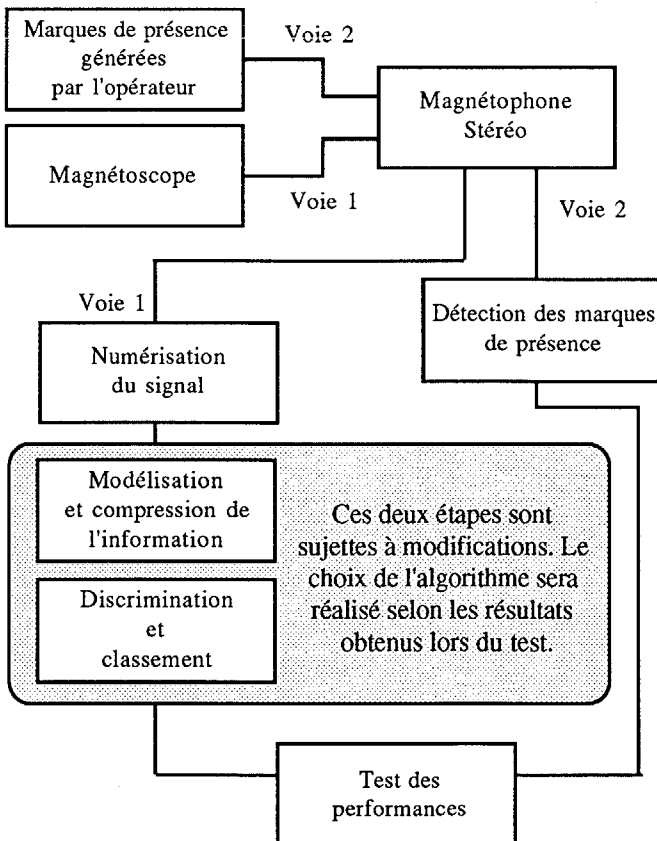


Fig. n°4: Méthodologie adoptée pour le test des performances

En fonction des résultats obtenus lors du test des performances, nous avons retenu les solutions suivantes :

1) La modélisation du signal :

La solution retenue est exposée au paragraphe n°3, le spectre est estimé à partir d'une modélisation autorégressive obtenue à l'aide de l'algorithme de Levinson-Durbin. Nous avons choisi deux valeurs de la fréquence pour lesquelles la valeur du spectre permettra l'identification du signal ($f_1=1,2\text{kHz}$, $f_2=1,8\text{kHz}$).

Si l'on examine le spectre moyen du signal, ces fréquences correspondent au premier maximum et au creux entre les deux maxima.

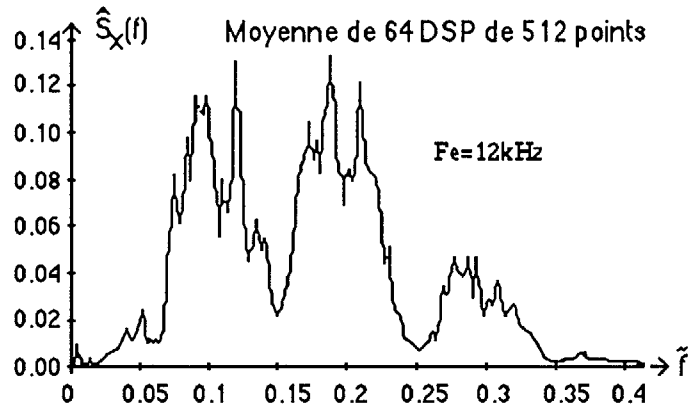


Fig. n°5: Spectre moyen du signal à reconnaître.

Les caractéristiques de la modélisation auto-régressive sont les suivantes :

- Fréquence d'échantillonnage : 12kHz
- Nombre de points pour l'estimation : 1024
- Ordre de la modélisation : 6

Ces valeurs correspondent à l'obtention d'un nouveau modèle toutes les 0,1 seconde.

2) La classification comporte deux phases :

- la phase d'apprentissage basée sur une méthode de classification non supervisée (Algorithme d'agrégation autour de centres mobiles) précédée d'une analyse factorielle discriminante.
- la phase de classification qui réalise la projection du modèle dans l'espace défini par l'analyse factorielle discriminante suivie du calcul des distances Euclidiennes aux diverses classes afin de déterminer l'appartenance du modèle à classer.

3) Traitement de l'information :

Il consiste ici en la mémorisation de la durée et de la date des événements en vue d'une étude statistique.

Afin de situer les performances atteintes, nous présentons les résultats obtenus à partir de la classification de deux séquences du signal, leurs caractéristiques sont les suivantes :

Séquence n°1 : P(Alose)=0,50 P(Bruit)=0,10
Longueur 10 minutes.



Séquence n°2 : P(Alose)=0,12 P(Bruit)=0,65
Longueur 10 minutes.

Le choix du nombre de classes à considérer pour la classification ne saurait être décidé de manière arbitraire et systématique. Des essais doivent être effectués pour chaque cas de figure. On peut montrer expérimentalement [7] que dans le cas qui nous préoccupe les résultats sont sensibles au nombre de classes demandées pour la classification, d'une manière générale, plus le nombre de classes augmente et plus les taux de détection $P(dA/A)$ et de fausse détection $P(dA/\bar{A})$ augmentent. Nous avons choisi de travailler sur douze classes pour cette partie, car au-delà, le taux de fausse détection devient trop important.

Si l'on ne tient pas compte de l'évolution temporelle, les meilleurs résultats ont été obtenus pour la classification à partir du paramètre obtenu à l'aide d'une analyse factorielle discriminante : $P = 0,17.k_{xx1} + \sigma_b^2 - 0,1.Cp_5$

	$P(dA/A)$	$P(dA/\bar{A})$	TRéussite
S1	0,75	0,11	82%
S2	0,59	0,02	92%

Tableau n°1 : Résultats sans prise en compte du temps.

La prise en compte du temps est effectuée à l'aide d'une approximation polynomiale au sens des moindres carrés. Le point délicat de cette méthode réside dans le choix de l'ordre du polynôme et de la durée d'observation des paramètres. En effet, si la durée d'observation est trop courte devant le temps de variation des caractéristiques du signal, elle ne rend pas compte de l'évolution du phénomène. Si elle est trop longue, les différences dues à la non-reproductibilité du signal sont prises en compte et augmentent le taux d'échec de la classification.

Les meilleurs résultats ont été obtenus à partir d'une approximation polynomiale du premier ordre calculée à partir de vingt valeurs consécutives du spectre selon les équations suivantes :

$$\hat{S}_x(t) \Big|_{f=0,1} = A_1.t + A_2 \quad \text{et} \quad \hat{S}_x(t) \Big|_{f=0,15} = A_3.t + A_4$$

On obtient, avec une classification à partir des paramètres A1 et A3 les résultats suivants:

	$P(dA/A)$	$P(dA/\bar{A})$	TRéussite
S1	0,78	0,12	83%
S2	0,79	0,08	90%

Tableau n°2: Résultats avec prise en compte du temps.

Au-delà du taux de réussite brut, on peut remarquer que l'on est en présence d'un dilemme entre la probabilité de détection et la probabilité de fausse alarme. Pour le problème présenté, un taux de fausse alarme de l'ordre de 10% est acceptable. Ainsi on obtient un taux de détection d'environ 79% pour la solution retenue.

Au premier abord, cette valeur peut paraître faible, mais si un opérateur expérimenté arrive à 90% de reconnaissance lors des premières minutes d'écoute, sa vigilance décroît rapidement

au cours du temps si bien qu'au bout de quelques heures il devient nettement moins performant que le dispositif automatique. La principale source d'erreur de la part de l'opérateur (y compris pour le marquage des bandes pour la phase de test) est son temps de réaction. D'autant plus qu'il n'est pas simple de définir rigoureusement le signal car il apparaît et disparaît progressivement si bien qu'il est difficile de faire la différence entre le bruit de la rivière et celui de la fraie de l'alse lorsqu'il est de faible amplitude.

6. Conclusions

L'originalité de la méthode présentée réside dans l'utilisation d'une modélisation paramétrique du signal fournissant un vecteur modèle $V_I(t)$ (décrivant le spectre en fonction de la fréquence à une date donnée) pour obtenir un autre vecteur modèle $V_{II}(f)$ (décrivant le spectre en fonction du temps pour une fréquence donnée). Ce dernier type de modèle fournit en effet une alternative très intéressante pour les problèmes de représentation sous forme condensée de l'information et par là même, pour la classification.

L'application décrite étant maintenant exploitée à l'extérieur du laboratoire, déjà d'autres problèmes de reconnaissance similaires ont été abordés avec succès. Le choix de la durée d'approximation, pour le modèle temps-fréquence, ne peut être fait que par essais successifs. Nous cherchons actuellement une méthode permettant un choix plus systématique des paramètres devant participer à la classification.

7. Bibliographie

- [1] J. Max et Collaborateurs.
Méthodes et techniques du traitement du signal.
Éditions MASSON.
- [2] K.M. Prabhu, K.B. Bagan.
Resolution capability of non linear spectral-estimation methods for short data lengths.
I.E.E. Proceedings Vol 136 n°3 pp 143-151, june 1989.
- [3] J.L. Lacoume - M. Gharbi - C. Latombe - J.L. Nicolas
Close frequency resolution by maximum entropy spectral estimation.
IEEE Trans. on ASSP vol ASSP 32 Oct. 84
- [4] A.H. Gray Jr, J.D. Markel.
Distance measures for speech processing
I.E.E.E. trans. on ASSP, ASSP 24(5), pp 380-391, 1976
- [5] Calliope. La parole et son traitement automatique.
Éditions MASSON série CNET-ENST.
- [6] C. Doncarli, M.F. Lucas, P. Guihéneuc.
Modélisation statistique et classification automatique de signaux physiologiques. Méthodes et premiers résultats.
GRETSI 87 pp 619-622
- [7] J.F. Trouilhet. Étude et réalisation d'un dispositif de reconnaissance automatique de signature acoustique.
Thèse de doctorat de l'INP Toulouse Juin 1992