

# Localisation et discrimination de signaux acoustiques de bulles d'air par des techniques statistiques

## Statistical Methods for Localization and Discrimination of Acoustical Signals Backscattered by Air Bubbles

par Ali KHELIL\*, Claude MANTE, Pascal DAVID

Centre d'Océanologie de Marseille  
Campus de Luminy, Case 901  
F-13288 Marseille cedex 9  
\*Seconde Adresse : Ecole Normale Supérieure  
BP 358, Route de Constantine  
DZ-04000 Oum El Bouaghi

### *résumé et mots clés*

On étudie cinq enregistrements d'échos acoustiques de « monobulles », de tailles distinctes, produites expérimentalement. Ces enregistrements sont découpés en blocs afin de détecter les signaux de bulles. On procède ensuite à l'analyse des inter-distances entre blocs, exprimées soit au moyen de la distance usuelle, soit au moyen de distances entre classes d'équivalence de blocs (distances orbitales). La détection des bulles est facilitée par un filtrage des enregistrements par analyse en composantes principales. Les blocs sélectionnés sont soumis à une série d'analyses discriminantes en cinq, quatre et trois classes, les variables discriminantes étant soit les coefficients de Fourier du signal, soit les facteurs issus de l'analyse de diverses tables d'inter-distances orbitales. Les résultats obtenus sont discutés et validés par une technique de rééchantillonnage.

**Bulle d'air, discrimination, analyse en composantes principales, distances orbitales, validation croisée.**

### *abstract and key words*

Five records of acoustic echoes backscattered by experimentally produced distinct size « monobubbles » are studied. These records were splitted into several blocks so as to detect bubble signals. We proceeded to the analysis of distances between blocks, which were expressed either by usual distances or by distances between equivalence classes of blocks (orbital distances). The bubble detection is rendered easier by filtering records through principal component analysis. The selected blocks were then submitted to a series of discriminant analyses in five, four or three classes. The discriminant variables were either Fourier coefficients or scores stemming from various orbital distances analyses. The obtained results are discussed and validated by a resampling method.

**Air bubble, discrimination, principal component analysis, orbital distances, cross validation.**

## 1. introduction

L'étude des bulles de gaz présentes dans la mer est d'une grande importance, que l'on s'intéresse aux échanges océan-atmosphère [1; 2], à la chimie de surface, au transport vertical ou encore à la

migration d'organismes marins [3]. En acoustique sous-marine, il est connu [4-6] que les bulles d'air ont une très grande influence sur la propagation acoustique dans l'eau à cause d'une part, des différences de densité et de compressibilité entre l'air et l'eau, et d'autre part du caractère résonant des bulles. On sait également que l'onde acoustique est très sensible à tout ce qui se trouve dans le cône insonifié (bulles de gaz, particules solides, organismes

vivants etc...). Ainsi, sur les enregistrements effectués en mer par notre équipe [3; 7], on trouve les réponses acoustiques de diverses cibles mais leur identification ou leur reconnaissance restent difficiles, voire souvent impossibles. L'idée de discriminer ces cibles à partir de leurs signatures acoustiques constitue une étape fondamentale dans les études par acoustique sous-marine de la migration d'organismes marins tels que le zooplancton par exemple.

Nous disposons de cinq enregistrements réalisés [8] sur banc d'essais et renfermant les réponses acoustiques de monobulles calibrées. Sur de tels enregistrements apparaissent les échos de bulles distinctes, de même taille, occupant différentes positions-temps. Notre objectif est de rechercher et d'expliquer les différences entre les classes d'échos associées à différentes tailles de bulles. C'est donc typiquement un problème d'Analyse Discriminante.

Il faut d'abord extraire les échos de chaque enregistrement, découpé au préalable en une suite de blocs chevauchants. Nous isolons les blocs contenant un écho par analyse statistique de divers tableaux recensant les distances entre blocs. Une fois extraits, ils sont soumis à une série d'Analyses Factorielles Discriminantes (AFD) faisant intervenir différents regroupements des classes de taille initiales.

Dans ce travail, nous testerons l'efficacité de diverses distances pour résoudre aussi bien le problème de l'extraction des échos que celui de leur discrimination. L'utilité d'un filtrage préliminaire sera également établie.

## 2. prétraitements

### 2.1. les données

Les bulles d'air étudiées ont été produites par une technique [8] utilisant une pompe péristaltique qui aspire simultanément de l'air et de l'eau et donne un flux d'eau transportant des bulles d'air. Ce flux vient se déverser dans une colonne d'eau qui constitue le milieu où évoluent les bulles. La mesure des rayons de bulles est réalisée par méthode indirecte, à partir des mesures des vitesses limites d'ascension [9].

Le sondage acoustique est réalisé avec un transducteur large bande fréquentielle ( $500 \pm 200$  kHz) fonctionnant en mode mixte : émetteur-récepteur, avec une émission de  $20 \mu\text{s}$ , et insonifiant une colonne d'eau d'environ 1 mètre. L'acquisition des données est assurée par un oscilloscope numérique avec un pas d'échantillonnage de 100 ns.

Les cinq enregistrements constituant les données à traiter correspondent chacun à une taille de bulle donnée et sont codés comme suit : od 6 est l'enregistrement renfermant les réponses acoustiques de bulles dont le rayon est  $325 \mu\text{m}$ , od 4 correspond à une taille de bulle de  $345 \mu\text{m}$ , od 9 à  $365 \mu\text{m}$ , od 16 à  $410 \mu\text{m}$  et od 20 à  $485 \mu\text{m}$ .

### 2.2. choix de l'unité statistique

Chaque enregistrement est découpé en blocs composés d'un nombre de points déterminé. Chacun de ces blocs doit satisfaire simultanément à deux contraintes : sa longueur doit être supérieure à la durée moyenne d'un écho de bulle ( $8 \mu\text{s}$ , soit 80 points) et le nombre de points correspondant doit être une puissance de 2, afin que la Transformée de Fourier Rapide (FFT) de chaque bloc soit calculable par l'algorithme de Cooley-Tukey.

Nous avons choisi, après divers essais, une longueur de 256 points par bloc avec un recouvrement de 12 points (6 points en début et 6 à la fin du bloc). Les blocs sont numérotés séquentiellement en fonction de l'enregistrement d'où ils ont été extraits. Par exemple, si od 6 est l'enregistrement à découper, « 615 » représentera le quinzième bloc de l'enregistrement od 6.

### 2.3. passage dans le domaine fréquentiel

Tous les traitements numériques sont réalisés dans le domaine fréquentiel, qui présente un bon comportement vis-à-vis des translations. Tout bloc est représenté par 256 harmoniques :  $H_0$  à  $H_{255}$ . Le calcul des transformées de Fourier discrètes se fait après multiplication par une fenêtre de Hanning trapézoïdale [10]. Cette fenêtre a deux avantages : elle a un coût très faible en temps de calcul et laisse invariante la majorité (ici 95%) des mesures d'un bloc, ce qui nous permettra ultérieurement de comparer les blocs sans faire intervenir de fonction-poids.

Afin de réduire le volume des calculs, nous avons procédé à une sélection préalable des variables. Nos analyses ne faisant intervenir que les moments d'ordre 1 et 2, nous n'avons conservé que les harmoniques contribuant de manière non négligeable à la variance du signal [11]; leur sélection repose sur un test de Fisher. Les variables conservées sont la moyenne  $H_0$  et les harmoniques  $H_7$  à  $H_{18}$ , qui expliquent du reste la quasi-totalité de l'énergie moyenne de chaque bloc [8]. A l'issue de cette étape, chaque enregistrement est représenté par un tableau croisant  $n$  blocs et  $p$  variables ( $p = 2q - 1$ ,  $q$  étant le nombre d'harmoniques sélectionnées).

## 3. filtrage par analyse en composantes principales

Nous avons d'abord fait appel à l'Analyse en Composantes Principales (ACP) [12; 13] pour faire une typologie des blocs afin

d'extraire les échos. Cette tentative [8], que nous ne ferons qu'évoquer ici, a eu pour mérite essentiel de montrer qu'un filtrage préliminaire du signal était indispensable. En effet, étant donné que le rapport signal/bruit se dégrade avec la profondeur à cause des phénomènes d'absorption et de diffraction [5; 6], les bulles situées à une trop grande profondeur ne pouvaient être mises en évidence ni par cette méthode, ni par la méthode des distances orbitales dont nous parlerons plus loin. Le filtrage par ACP consiste à reconstituer approximativement chaque enregistrement à partir des premiers facteurs issus de son analyse [13]. Les données sont d'abord reconstituées par blocs dans le domaine fréquentiel puis, par transformée de Fourier inverse, on les exprime à nouveau dans le domaine temporel.

Pour reconstituer l'enregistrement, il suffit de relier les blocs ainsi filtrés, avec un traitement particulier des zones de recouvrement.

L'utilisation de cette méthode sous-entend les hypothèses suivantes : la variance du bruit est petite devant celle du signal, et tous deux sont non corrélés. Le signal sera alors restauré par les premiers facteurs, le bruit étant associé à des composantes de variance moindre.

Le nombre  $k$  de facteurs retenus pour la reconstitution des différents enregistrements se répartit comme suit (dans l'ordre de taille) : 3, 5, 4, 2 et 3.

**Remarques :** les harmoniques conservées étant de basse fréquence (cf. § 2.3), l'étape de sélection peut être vue comme un filtrage passe-bas adapté à la structure du signal traité. Le filtre linéaire correspondant à l'ACP opère quant à lui dans le domaine des basses fréquences choisies : il complète donc le premier.

Sur la figure (figure 1), sont représentés les cinq enregistrements avant et après filtrage. Globalement, le traitement a permis d'éliminer une grande partie du bruit : le signal relatif aux bulles est très bien reconstitué et quelques nouveaux échos de faible amplitude sont mieux visibles sur les enregistrements filtrés.

## 4. utilisation de distances orbitales

L'ACP peut être présentée [12] comme la recherche d'une image euclidienne aussi simple que possible d'un ensemble d'objets décrits par leurs inter-distances. Or, comme la distance usuelle entre deux blocs renfermant des bulles dépend de la position de celles-ci, elle se trouve en général arbitrairement surévaluée. Il s'en suit qu'un écho de faible puissance ne sera pas toujours mis en évidence par l'ACP. La difficulté provient du fait que celle-ci n'est pas invariante par translation : plusieurs facteurs pourraient par exemple être associés à des échos semblables décalés dans le temps [11]. Afin de mettre en évidence l'identité de tels blocs, mieux vaut raisonner en termes de formes que de courbes.

D'après R. Thom [14], «...si  $E$  est un espace topologique,  $G$  un groupe (ou pseudo-groupe) opérant dans  $E$ , une  $G$ -forme est par définition une classe d'équivalence de fermés dans  $E$  modulo l'action de  $G$  ». L'adoption d'une notion de forme est donc équivalente au choix du groupe de transformations dont on veut neutraliser l'effet, une unique forme étant alors associée à toute une orbite.

De tels principes sont appliqués par Gourd *et al.* [15] qui appliquent la théorie de l'analyse harmonique abstraite à la détermination du système complet d'invariants associé à l'action d'un groupe commutatif de transformation. Les auteurs utilisent leurs résultats pour détecter l'apparition de perturbations dans des signaux. Dans le même esprit Ghorbel [16] construit une famille complète et stable d'invariants pour les contours fermés dans le plan.

Notre approche est davantage géométrique; elle s'inspire plutôt de travaux ayant trait à la typologie des formes [17–19] qu'à la reconnaissance de certaines formes particulières.

### 4.1. définitions

Soit  $(E, D_{is})$  un espace métrique et  $G$  un groupe d'isométries de  $E$ , vérifiant donc :

$$\forall g \in G, \forall B, B' \in E, D_{is}(gB, gB') = D_{is}(B, B')$$

L'orbite de  $B$  sous l'action de  $G$  est l'ensemble des transformées de  $B$  par  $G$  :

$$\text{orb}(B) = GB = \{gB; g \in G\}$$

Les orbites ne sont autres que les classes d'équivalence dans  $E$  de la relation :

$$B \sim B' \Leftrightarrow \exists g \in G; gB = B'$$

Dans notre cas  $E$  désignera soit l'espace euclidien usuel  $\mathbb{R}^p$  associé aux fréquences sélectionnées, soit la sphère unité de cet espace. Les analyses sur la sphère compareront entre eux les blocs normés au sens de l'espace  $L^2$  des fonctions de carré intégrable.

Le groupe  $G$  opérant sur  $E$  sera soit celui des translations soit le groupe produit [20] de celui-ci avec le groupe à deux éléments des antipodias (changements de signe). Si l'emploi du groupe des translations va de soi, celui des antipodias, qui identifient les blocs  $B$  et  $-B$ , se justifie ici dans la mesure où il n'existe pas d'« anti-écho ». Cette transformation ne peut donc rapprocher entre eux que des blocs de bruit, ce qui est souhaitable.

On peut montrer que la distance orbitale entre  $B$  et  $B'$ , définie par :

$$D_{or}(B, B') = \min_{g \in G} \{D_{is}(B', g(B))\}$$

est la distance géodésique sur la variété quotient  $E/G$  [11].

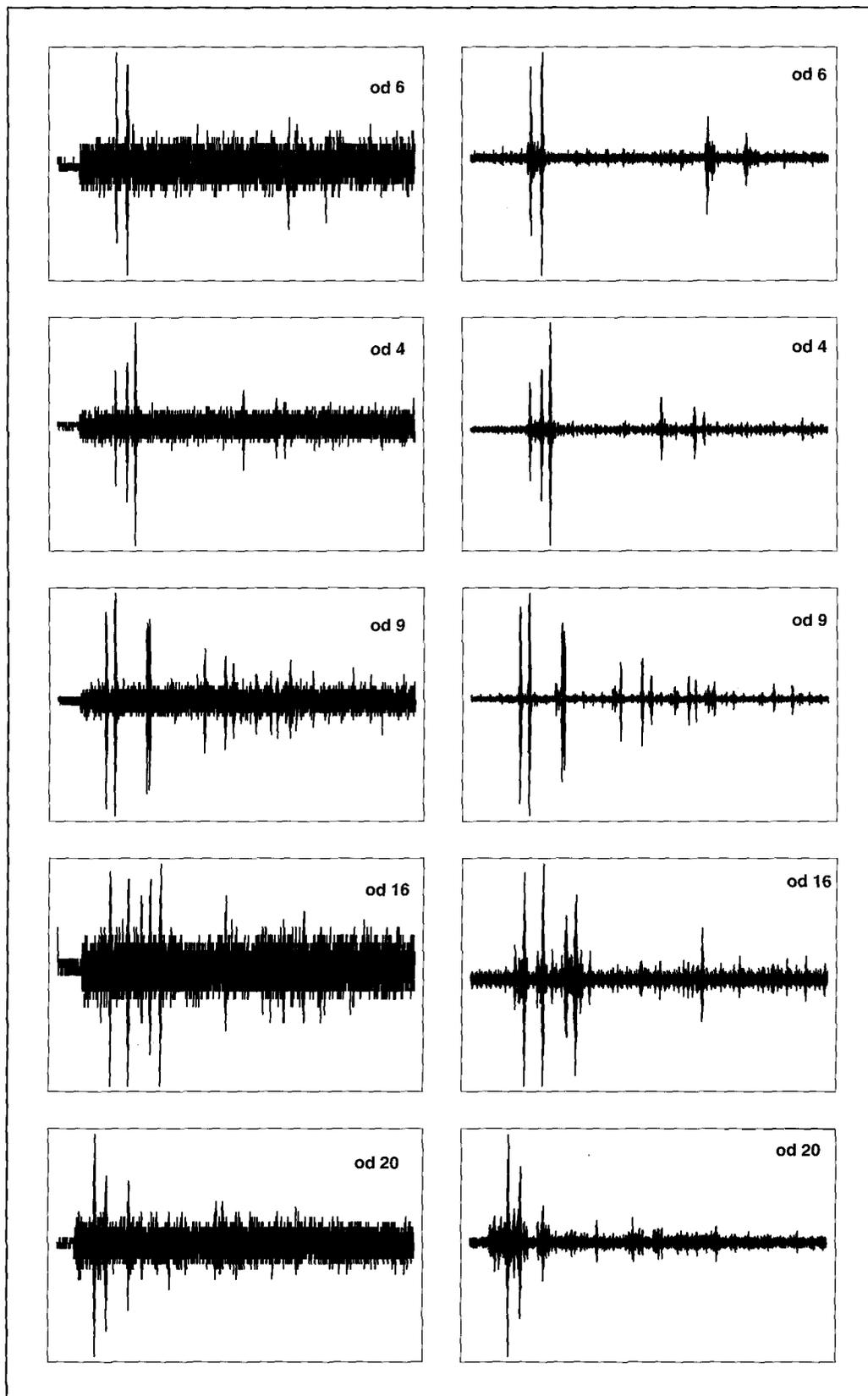


Figure 1. – Les cinq enregistrements étudiés avant (à gauche) et après (à droite) le filtrage par ACP.

Nous utiliserons quatre distances orbitales :  $D_{or1}$  (invariance par translation) et  $D_{or2}$  (invariance par translation et antipodie) dans  $R^p$ , et  $D_{or1S}$  et  $D_{or2S}$ , ayant les mêmes propriétés d'invariance, pour comparer les blocs après projection sur la sphère unité.

## 4.2. analyse factorielle de tableaux de distances

Nous ne décrivons ici que les traitements d'enregistrements filtrés. Pour une comparaison avec des résultats issus d'autres analyses (ACP, distances orbitales sur données brutes), nous renvoyons le lecteur à la thèse de Khelil [8].

Les tableaux inter-distances ont été chacun soumis à une Analyse Factorielle de Tableau de Distances (AFTD) [12]; les quatre distances orbitales aboutissent toutes à un choix identique [8] des blocs même si cela est plus difficile à faire dans le cas des distances sphériques. En effet, la puissance des signaux ayant été éliminée par normalisation, seule leur forme permet de les distinguer l'un de l'autre.

Commentons brièvement l'analyse des tableaux correspondant aux distances non normées  $D_{or1}$  et  $D_{or2}$  avant de donner la liste des blocs sélectionnés.

Les plans factoriels montrent chacun un amas de points groupés à proximité de l'origine et quelques rares blocs excentrés éloignés les uns des autres. Dans chacun des cinq enregistrements l'axe 1 semble lié à la profondeur, donc indirectement à la puissance des échos. A titre d'exemple, nous reproduisons le premier plan issu de l'analyse de od 9 (figure 2). Le contraste entre les blocs ne contenant pas d'écho, massés dans une petite zone, et ceux en contenant un, situés à la périphérie du nuage, y est beaucoup plus important que sur la figure (figure 3) représentant le premier plan factoriel issu de l'ACP du même enregistrement. De plus, le taux d'inertie (ou de variance) associé au premier plan étant plus élevé dans le cas de  $D_{or1}$  (86%) que dans le cas de l'ACP (64%), la représentation plane du nuage des blocs y est plus fidèle dans ce cas.

Le tableau 1 regroupe les blocs sélectionnés par cette méthode et qui seront, désormais, les représentants des bulles pour la suite de

Tableau 1. – Liste des blocs sélectionnés par AFTD. Ce sont les blocs renfermant un signal de bulle.

Enregistrement	Blocs représentant des bulles
od 6	608; 609; 627; 631
od 4	408; 409; 410; 422; 425; 426
od 9	907; 908; 911; 917; 920; 921
od 16	1607; 1609; 1612; 1613; 1626
od 20	2006; 2007; 2008; 2010; 2016

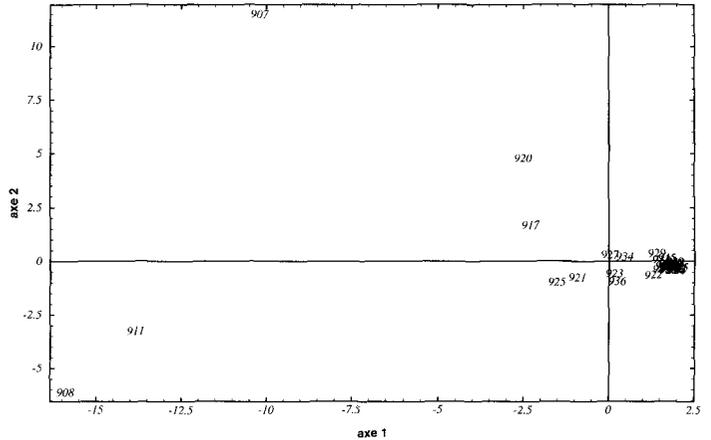


Figure 2. – Plan factoriel 1/2 de l'AFTD sur les blocs de od 9 représentés par les distances  $D_{or1}$ .

Valeurs propres :  $vp1 = 17.50$ ;  $vp2 = 5.72$ .  
Inertie :  $axe1 \rightarrow 64.83\%$ ;  $axe2 \rightarrow 21.20\%$ .

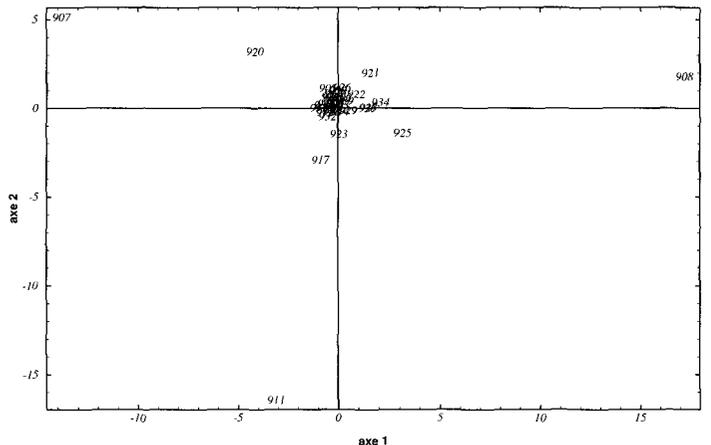


Figure 3. – Plan factoriel 1/2 de l'ACP sur od 9.

Valeurs propres :  $vp1 = 14.215$ ;  $vp2 = 8.841$ .  
Inertie :  $axe1 \rightarrow 39.44\%$ ;  $axe2 \rightarrow 24.53\%$ .

l'étude. Le bloc 925 devrait normalement faire partie des sélectionnés mais, en vue de la discrimination, nous voulions garder approximativement le même nombre de blocs par enregistrement.

## 5. analyse discriminante

### 5.1. objectif et méthode

Le problème posé est maintenant le suivant : étant donné un ensemble de blocs renfermant chacun un écho, peut-on prédire la taille de chaque bulle à partir des caractéristiques du bloc contenant son écho? Pour répondre à cette question nous avons d'abord cherché à voir si l'existence de cinq classes de taille se manifestait spontanément, en utilisant des méthodes exploratoires telles que l'AFTD, la Classification Hiérarchique Ascendante ou les Nuées

Dynamiques [12; 21]. Le fait qu'aucune de ces méthodes ne les ait mises en évidence montre la difficulté du problème : elles ne sont pas *visibles*.

Nous avons donc fait appel à une variante pas à pas de l'analyse discriminante de Fisher [22]. Les données sont représentées par un tableau constitué de vingt-six individus décrits de deux façons différentes :

- soit par vingt et une variables qui sont les coefficients de Fourier à dix fréquences individuelles, auxquelles s'ajoute la moyenne  $H_0$ ;
- soit par les sept premiers facteurs de l'AFTD sur distances orbitales.

Disposant de quatre distances orbitales, nous pouvons procéder à autant de variantes d'analyse sur les facteurs. Cela nous donne la possibilité de discriminer soit par la forme ( $D_{or}S$ ), soit simultanément par la forme et la puissance du signal ( $D_{or}$ ).

Par ailleurs, l'analyse discriminante sur les spectres (coefficients de Fourier) permet de déterminer les fréquences les plus discriminantes, ce qui constitue une application pratique importante en acoustique.

**Notations :** Pour éviter de surcharger les représentations graphiques, les cinq classes définies a priori sont notées  $A$ ,  $B$ ,  $C$ ,  $D$  et  $E$  (dans l'ordre croissant des rayons de bulles).

Nous effectuerons d'abord une analyse en cinq ( $A, B, C, D$  et  $E$ ), puis quatre ( $AB, C, D$  et  $E$  ou  $A, B, C$  et  $DE$ ) et enfin trois classes ( $AB, C$  et  $DE$ ).

Les raisons ayant motivé cette définition des classes a priori de l'analyse discriminante sont :

- $A$  et  $B$  étant de tailles très voisines (325  $\mu\text{m}$  et 345  $\mu\text{m}$ ), elles peuvent être regroupées en une classe unique  $AB$ ;
- au delà de 400  $\mu\text{m}$  de rayon, les bulles produites changent de comportement durant leur mouvement ascensionnel et adoptent un comportement de grosses bulles, c'est-à-dire que leur mouvement ne serait pas tout à fait rectiligne [8; 23].  $D$  et  $E$  peuvent donc être groupées en une classe unique  $DE$ .

### 5.2. discrimination sur coefficient de fourier

Le tableau 2 résume les résultats de cette analyse. Rappelons que dans le domaine fréquentiel un bloc est décrit par un certain nombre d'harmoniques :  $H_0$  à  $H_k$ . Chacune des harmoniques  $H_1$  à  $H_k$  correspond à une fréquence  $f_k = \frac{k}{\Delta T}$  (kHz),  $\Delta T$  étant la longueur d'un bloc (soit  $255 \cdot 10^{-7}$  s).

Connaissant les indices des harmoniques retenues on peut donc remonter aux fréquences intervenant dans la discrimination, ce que nous mettrons à profit pour coder les variables : la paire de variables réelles associée à  $H_k$  sera notée ( $Cf_k, Sf_k$ ).

Tableau 2. – Fréquences les plus discriminantes des signaux de bulles.

Harmoniques retenues	Nombre de classes	$\tau$ (%)	Variables (fréquences) discriminantes (kHz)	
$H_0$	5	61.54	$H_0$	
	4	$A \equiv B$	61.54	$H_0$
		$D \equiv E$	69.23	$H_0$
$\&$ $H_8-H_{17}$	3	88.46	$H_0$ C667 & S667 C510 S313 C431	

Pour une discrimination en cinq ou quatre classes, l'analyse donne un pourcentage de bien-classés  $\tau$  allant de 60 à 70 %, la seule variable discriminante étant  $H_0$ . Le cas de trois classes est plus intéressant puisque d'une part, on obtient un pourcentage  $\tau$  supérieur à 88 % et d'autre part, la discrimination est le fait de six variables. Cela nous permet de déterminer les fréquences acoustiques possédant le meilleur pouvoir discriminant.

La discrimination est en grande partie le fait du niveau moyen du signal ( $H_0$ ), mais les autres variables contribuent à une meilleure séparation des classes, comme le montre l'examen des coordonnées des deux premiers vecteurs propres.

	$H_0$	C667	S667	C510	S313	C431
1 <sup>er</sup> vecteur propre	<b>-0.780</b>	.157	.157	-.006	.155	-.070
2 <sup>ème</sup> vecteur propre	.170	<b>-1.530</b>	<b>-.858</b>	.256	<b>.593</b>	-.074

La première composante du premier vecteur propre est de loin la plus grande, ce qui veut dire que  $H_0$  est la variable la plus importante. De même, les composantes du deuxième vecteur propre montrent que les fréquences 667 kHz et éventuellement 313 kHz sont celles qui ont l'apport le plus important au deuxième facteur discriminant. En conclusion, s'il fallait faire la même étude expérimentale avec le même objectif (discrimination entre des groupes de bulles dans la même gamme de taille), mais en travaillant en monofréquence et non en large bande, il faudrait utiliser la fréquence 667 kHz.

### 5.3. discrimination à partir des distances orbitales

Dans le tableau 3, sont résumés les principaux résultats des différentes analyses discriminantes réalisées avec les sept premiers

facteurs issus de l'analyse des quatre types de tableaux de distances orbitales. Les taux d'inertie expliquée sont plus petits dans le cas des distances orbitales sphériques (environ 57% au lieu de 82%). Par contre, pouvoirs discriminants (valeurs propres) et pourcentages de bien-classés augmentent avec l'analyse des distances orbitales sphériques.

**Tableau 3. – Résultats de l'analyse discriminante réalisée à partir des facteurs de l'AFTD. Le pas désigne le nombre de variables utilisées pour la construction des règles de décision de la discrimination. Inertie désigne le taux d'inertie expliqué par les sept premiers facteurs de l'AFTD;  $\tau$  désigne le pourcentage de blocs bien-classés.  $\lambda_1$  et  $\lambda_2$  sont les pouvoirs discriminants des deux premiers facteurs.**

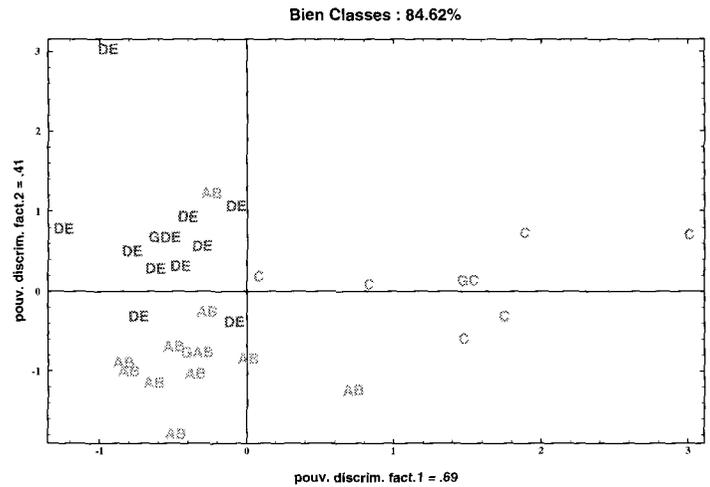
		5 classes A, B, C, D, E	4 classes AB, C, D, E	4 classes A, B, C, DE	3 classes AB, C, DE
$D_{or1}$	Pas	5	2	7	6
(Inertie :	$\tau$ %	65.38	57.69	57.69	69.23
82.6 %)	$\lambda_1$ et $\lambda_2$	.47; .35	.32; .07	.56; .27	.55; .17
$D_{or2}$	Pas	2	5	7	1
(Inertie :	$\tau$ %	50	69.23	61.54	73.08
81.6 %)	$\lambda_1$ et $\lambda_2$	.29; .28	.60; .27	.59; .22	.58
$D_{or1S}$	Pas	6	6	7	6
(Inertie :	$\tau$ %	84.62	88.46	84.62	69.23
56.8 %)	$\lambda_1$ et $\lambda_2$	.80; .63	.67; .55	.79; .53	.55; .16
$D_{or2S}$	Pas	6	4	5	3
(Inertie :	$\tau$ %	84.62	88.46	84.62	84.62
58.56 %)	$\lambda_1$ et $\lambda_2$	.73; .63	.73; .61	.68; .53	.69; .41

**Discrimination en 5 classes :** on aboutit à des pourcentages  $\tau$  de 50% et 65.38% et ceux-ci augmentent lorsque l'on projette les blocs sur la sphère ( $\tau = 84.62\%$ ). Il en va de même pour le pouvoir discriminant de chacun des deux premiers facteurs, qui passe du simple au double.

**Discrimination en 4 Classes :** globalement, quelle que soit la façon dont sont définies les quatre classes, la discrimination à partir des facteurs de l'AFTD réalisée à partir des  $D_{orS}$  donne les meilleurs résultats en terme de pourcentage  $\tau$ . De même, les pouvoirs discriminants des deux premiers facteurs doublent quand on norme les blocs.

**Discrimination en 3 classes :** là aussi, ce sont les  $D_{orS}$  qui fournissent les meilleurs résultats que ce soit en terme de pourcentage de blocs bien-classés ou de pouvoir discriminant.

En conclusion, le dépouillement du tableau 3 montre que les facteurs issus de l'analyse de la forme des échos sont les plus discriminants. Pour illustrer cette étape, nous donnons en (figure 4) la représentation dans le premier plan factoriel de nos blocs discriminés en trois classes.



**Figure 4. – Résultat graphique de l'analyse discriminante (en trois classes) sur les facteurs de l'AFTD sur  $D_{or2S}$ . Nombre de facteurs utilisés : 3.**

## 6. test de la validité des résultats de discrimination

### 6.1. les différentes techniques

Parmi les critères de mesure de la qualité d'une analyse discriminante, le pourcentage empirique d'individus bien classés  $\tau$  est généralement une estimation optimiste de la valeur réelle de ce taux. Il s'avère donc nécessaire de le corriger.

Pour cela, la procédure la plus courante consiste à séparer aléatoirement les individus en deux échantillons : un échantillon d'apprentissage qui sert à la construction des règles de décision et un échantillon test qui permet d'éprouver leur validité. Pour cela, il faut disposer d'un nombre suffisant d'individus pour constituer deux échantillons de taille raisonnable.

Lorsque l'échantillon est de petite taille, comme c'est le cas dans cette étude, il existe des techniques de rééchantillonnage permettant de mesurer la qualité d'une discrimination [21] :

- la méthode du bootstrap [24; 25] consiste à construire, par tirages aléatoires avec remise dans l'échantillon initial, un nombre très grand d'échantillons de même taille que l'échantillon initial, à discriminer chaque échantillon, puis à exploiter statistiquement les résultats : intervalles de confiances pour  $\tau$ , calcul de la probabilité d'affectation correcte de chaque individu, distribution statistique des coordonnées des vecteurs discriminants, etc...

- la méthode de validation croisée [21; 25] consiste à partager l'échantillon initial en plusieurs parties disjointes. Chacune de celles-ci sert à son tour d'échantillon test, tandis que le reste des individus sert d'échantillon d'apprentissage. La moyenne des

pourcentages  $\tau$  obtenus sur les échantillons tests successifs fournit une estimation réaliste du pourcentage véritable d'individus bien classés.

### 6.2. mise en œuvre de la validation croisée

Notre échantillon initial noté  $ECI$  est constitué de 26 individus répartis en cinq classes, définies a priori, comme suit : 4 individus  $A$ , 6 individus  $B$ , 6 individus  $C$ , 5 individus  $D$  et 5 individus  $E$ .

La première étape consiste à découper aléatoirement l'échantillon  $ECI$  en parties disjointes qu'on notera  $EC_i$  (dans notre cas, nous avons choisi  $i = 1, 4$ ). Cela a donné la répartition suivante :

- $EC_1$  composée de 6 blocs :  $1A + 1B + 2C + 1D + 1E$ .
- $EC_2$  composée de 6 blocs :  $1A + 2B + 1C + 1D + 1E$ .
- $EC_3$  composée de 7 blocs :  $1A + 2B + 1C + 1D + 2E$ .
- $EC_4$  composée de 7 blocs :  $1A + 1B + 2C + 2D + 1E$ .

Dans la seconde étape, on procède à quatre analyses discriminantes en construisant les fonctions de décision à l'aide des échantillons  $ECI - EC_i$ , et en faisant la prévision sur les échantillons  $EC_i$ . Autrement dit, on effectue la  $i^{\text{ème}}$  discrimination avec pour échantillon de base  $ECI - EC_i$  et pour échantillon test  $EC_i$ . On obtient alors quatre pourcentages  $\tau_{EC_i}$  d'individus bien classés, calculés sur les échantillons tests, dont la moyenne  $\tau_{est}$  fournit une estimation du pourcentage théorique. L'importance de la dégradation de qualité :  $\Delta\tau = \tau - \tau_{est}$  permet de juger de la stabilité des résultats.

### 6.3. résultats et commentaires

La procédure décrite dans le paragraphe précédent sera appliquée d'une part à la discriminante sur les facteurs de l'AFTD des deux distances  $D_{or1S}$  et  $D_{or2S}$  et d'autre part à la discriminante sur les spectres. Le tableau 4 résume les pourcentages  $\tau$  et  $\tau_{est}$  obtenus. Examinons ces résultats selon la nature des variables prédictrices :

**1<sup>er</sup> cas : facteurs de l'AFTD sur  $D_{or1S}$  :** Le pourcentage  $\tau_{est}$  est de l'ordre de 65% pour cinq classes et vaut 69% pour une discrimination en trois classes. Le risque d'erreur est donc sensiblement identique dans les deux cas. Cependant, les résultats de la discrimination en trois classes sont les plus stables ( $\Delta\tau = 0.18$ ).

**2<sup>ème</sup> cas : facteurs de l'AFTD sur  $D_{or2S}$  :** Que ce soit en cinq ou en trois classes, les résultats sont sensiblement équivalents aux précédents, avec un pourcentage  $\tau_{est}$  de l'ordre de 65 à 68%. Du point de vue de la stabilité, sauf dans le cas de 3 classes, on ne relève pas non plus de différence significative :  $\Delta\tau$  est de l'ordre de 17 à 20%. On préférera donc la première distance orbitale à la seconde, la qualité des résultats ne justifiant pas la complexité plus grande des traitements.

Tableau 4. – Résultats du test de validation croisée sur l'analyse discriminante.

		5 classes <i>A, B, C, D, E</i>	4 classes <i>AB, C, D, E</i>	4 classes <i>A, B, C, DE</i>	3 classes <i>AB, C, DE</i>
$D_{or1S}$	Pas	6	6	7	6
	$\tau$ (%)	84.62	88.46	84.62	69.23
	$\tau_{est}$ (%)	<b>64.88</b>	<b>53.47</b>	<b>57.74</b>	<b>69.05</b>
	$\Delta\tau$ (%)	19.74	34.99	26.88	0.18
$D_{or2S}$	Pas	6	4	5	3
	$\tau$ (%)	84.62	88.46	84.62	84.62
	$\tau_{est}$	<b>64.29</b>	<b>59.53</b>	<b>53.57</b>	<b>67.26</b>
	$\Delta\tau$ (%)	20.33	28.93	31.05	17.36
Spectres	Pas	1	1	1	4
	$\tau$ (%)	61.54	61.54	69.23	80
	$\tau_{est}$	<b>60.71</b>	<b>48.81</b>	<b>68.45</b>	<b>53</b>
	$\Delta\tau$ (%)	0.83	12.73	0.78	27

Globalement, la discrimination par les facteurs de l'AFTD sur les distances orbitales sphériques aboutit à des résultats sensiblement identiques. Les cas de discrimination en quatre classes sont à écarter à cause des pourcentages  $\tau_{est}$  relativement faibles et de l'instabilité des résultats. On choisira donc, de discriminer soit en cinq classes, soit en trois classes avec l'une ou l'autre des distances  $D_{or1S}$  ou  $D_{or2S}$ .

**3<sup>ème</sup> cas : coefficients de Fourier (spectres) :** Si l'on se base sur l'examen de  $\tau_{est}$  et  $\Delta\tau$ , le meilleur cas est celui des quatre groupes ( $A, B, C$  et  $DE$ ), qui aboutit à un pourcentage de bien classés équivalent à celui obtenu avec les distances orbitales (70%), et ce avec une valeur bien moindre de  $\Delta\tau$ . Ceci dit, on obtient aussi de bonnes performances avec 5 classes.

## 7. discussion et conclusion

Le filtrage par ACP donne des résultats satisfaisants du point de vue de l'amélioration du rapport signal/bruit. On a pu noter que les résultats de l'AFTD et de l'AFD se sont affinés après le filtrage. Toutefois, le choix du nombre  $k$  de facteurs initiaux pour la reconstitution est délicat, et a été fait empiriquement en fonction de l'allure des fonctions propres et de la distribution des composantes principales. On pourra rendre cette importante étape plus rigoureuse en faisant appel à des tests de sphéricité sur les valeurs propres [26; 27], destinés à déterminer la dimension de l'espace latent. D'autre part, nos échos étant transitoires, il serait sans doute possible de réaliser ce prétraitement par un filtrage de Wiener.

L'introduction de distances orbitales a permis une meilleure localisation des blocs contenant nos échos. L'intérêt de ce type de

distance réside ici dans le caractère d'invariance par translation et antipodie qu'elle confère à l'AFTD. Elles permettent notamment de diminuer l'inertie du nuage de blocs.

Concernant la discrimination, on conclura en trois points :

– les facteurs de l'AFTD sur distances orbitales sphériques permettent une discrimination en 3 ou 5 classes par la **forme** du signal;

– les coefficients de Fourier permettent une discrimination essentiellement basée sur le niveau moyen du signal : les blocs se différencient alors par un effet de **taille**;

– les deux précédents commentaires conduisent à penser que la meilleure manière de discriminer nos échos serait de compléter les facteurs de l'AFTD sur  $D_{or}1S$  par  $H0$  et éventuellement les fréquences discriminantes  $C667$  et  $S667$ .

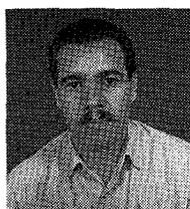
## BIBLIOGRAPHIE

- [1] Y.H.L. Hsu, P.A. Hwang, J. Wu, « Bubbles produced by breaking wind waves in gas transfer at water surfaces », W. Brutsaert and G.H. Jirka, D. Reidel, New York, 1984, p. 221–227.
- [2] S.A. Thorpe, D.N. Humphries, « Bubbles and breaking waves », *Nature*, Vol. 283, 1980, p. 463–465.
- [3] O. Guerin–Ancey, P.M. David, « Use of a multibeam–multifrequency sounder to study the distribution of small zooplankton », *Deep–Sea Res.*, Vol. 40, 1993, p. 119–128.
- [4] H. Medwin, C. Albert, Jr. Daniel, « Acoustical measurements of bubble production by spilling breakers », *J. Acoust. Soc. Amer.*, Vol. 88, n°1, 1990, p. 408–412.
- [5] R.J. Ulrich, *Principles of Underwater Sound for Engineers*, McGraw–Hill, New York, 1967, p. 337.
- [6] T.G. Leighton, *The Acoustic Bubble*, Academic Press, London, 1994.
- [7] J.P. Van Cuyck, *Etude de la réflexion acoustique d'une lame d'eau de mer contenant des organismes planctoniques*, Thèse, Univ. Aix–Marseille II, 1995.
- [8] A. Khelil, *Acoustique de bulles d'air dans une colonne d'eau. Identification et discrimination*, Thèse Univ. Méditerranée, 1995.
- [9] C. Cachard, *Mouvement de bulles dans un champ ultrasonore intense. Débulage d'un liquide*, Thèse, I.N.S.A. Lyon, 1988.
- [10] D.R. Brillinger, *Time series – Data analysis and theory*, McGraw–Hill, New York, 1981.
- [11] C. Mante, « Analyse en composantes principales d'un processus multiple non stationnaire : une application à des données météorologiques », *Statistique et Analyse des Données*, Vol. 14, n°2, 1989, p. 25–53.
- [12] F. Cailleux, J.P. Pages, *Introduction à l'analyse des données*, Smash, Paris, 1976.
- [13] B. Escofier, J. Pages, *Analyses factorielles simples et multiples : objectifs, méthodes et interprétation*, Dunod, Paris, 1990.
- [14] R. Thom, *Stabilité structurelle et morphogénèse*, Interéditions, 1977.
- [15] F. Gourd, J.P. Gauthier, H. Younes, « Une méthode d'invariants de l'analyse harmonique en reconnaissance de formes », *Traitement du Signal*, Vol. 6, n°3, 1989, p. 161–178.
- [16] F. Ghorbel, « Distance et invariance pour la classification d'objets dans les images », *International meeting on distance analysis DISTANCIA 92*, 1992, Rennes (F.).
- [17] P. Lütz, D. Maïti, « Classification automatique d'après la distance entre orbites : application à la physique corpusculaire », *Cahiers de l'Analyse des Données*, Vol. 3, n°4, 1978, p. 449–458.
- [18] I.L. Dryden, K.V. Mardia, « Multivariate shape analysis », *Sankhya*, Series A, Vol. 55, n°3, 1993, p. 460–480.
- [19] D.G. Kendall, « Mathematical statistics in the humanities, and some related problems in astronomy », *The ISI, Centenary Volume*, p. 392–408.
- [20] S. Maclane, G. Birkhoff, *Algèbre – Structures fondamentales*, Gauthier–Villars, Paris, 1970, p. 408.
- [21] G. Celeux, E. Diday, Coll., *Classification automatique des données*, Dunod, Paris, 1989.
- [22] J.M. Romeder, *Méthodes et programmes d'analyse discriminante*, Dunod, Paris, 1976, p. 274.
- [23] J. Hadamard, « Mouvement permanent lent d'une sphère liquide et visqueuse dans un milieu visqueux », *C. R. A. Sci.*, Vol. 152, 1911, Vol. 154, 1912, p. 1735–1738 et p. 109.
- [24] P. Diaconis, B. Efron, « Méthodes de calculs statistiques intensifs sur ordinateurs », *Pour la Science*, Juillet, 1983, p. 46–59.
- [25] B. Efron, « The Jackknife, the bootstrap and other resampling plans », *Soc. Ind. App. Math.*, CBMS, 1982, Philadelphia, Pennsylvania.
- [26] P. Gonzales Vicente, « Recherche de la dimension de l'espace latent en ACP », *Statistiques et Analyse de Données*, Vol. 11, n°3, 1986, p. 19–29.
- [27] I. Anastassakos, G. D'Aubigny, « L'utilisation des tests de sphéricité pour la recherche de la dimension de l'espace latent en analyse factorielle classique et en analyse en composantes principales », *Revue de Statistiques Appliquées*, Vol. XXXII, n°2, 1984, p. 45–57.

Manuscrit reçu le 19 mars 1996.

## LES AUTEURS

### Ali KHELIL



Ali Khelil est enseignant de physique à l'Ecole Normale Supérieure d'Oum–El–Bouaghi. Il a soutenu sa thèse à l'Université de la Méditerranée en 1995 sur le sujet « Acoustique de bulles d'air dans une colonne d'eau. Identification et discrimination ». Son domaine d'activité demeure l'acoustique pour l'océanographie et l'analyse des données acoustiques par des techniques statistiques.

### Claude MANTÉ



Claude Manté est ingénieur au CNRS depuis 1980. Il a soutenu en 1981, à l'Université Pierre et Marie Curie, une thèse de 3ème cycle en Statistique portant sur l'étude par l'Analyse des Données de la mémoire d'un champ météorologique. Ses domaines d'activité sont le traitement des séries chronologiques multivariées et le traitement statistique de l'image.

### Pascal DAVID



Pascal David, Dr, HdR, Ingénieur de Recherche au CNRS, est responsable de l'équipe « Dynamique des Particules et Acoustique Sous–Marine » au Centre d'Océanologie de Marseille depuis 1987. Il développe l'Acoustique en vue d'applications océanographiques et plus particulièrement pour l'étude de particules marines comme le plancton. Ses travaux concernent l'océanographie et le traitement du signal en temps réel.