

# Segmentation temps-fréquence statistique : Analyse de la répartition spatiale des détections sur spectrogramme N fois seuillé.

Florian DADOUCHI, Cédric GERVAISE, Cornél IOANA, Julien HUILLERY, Jérôme I. MARS

<sup>1</sup> Université Grenoble Alpes, Gipsa-Lab, 38402 Grenoble Cedex

<sup>1</sup> CNRS, Gipsa-Lab, 38402 Grenoble Cedex

<sup>2</sup> Ecole Centrale de Lyon, Laboratoire Ampère, 36 Avenue Guy de Collongue, 69130 Écully, France

<sup>3</sup> Chaire CHORUS, Fondation Grenoble INP, 46, Avenue Felix Viallet 38031 Grenoble Cedex 1, France

<sup>1</sup>[florian.dadouchi, cedric.gervaise, cornel.ioana, jerome.mars]@gipsa-lab.grenoble-inp.fr

<sup>2</sup>julien.huillery@ec-lyon.fr

**Résumé** – L’estimation de pistes temps-fréquence est classiquement réalisée par association de détections dans le plan temps-fréquence et ainsi se heurte au compromis « fausses détections/ détections manquées » lors de la détection des pics d’énergie. Nous proposons une méthode reposant sur un double test d’hypothèse binaire à probabilités de fausses alarmes constantes. Le premier test se concentre sur les pics d’énergie dans le spectrogramme, le second analyse la répartition spatiale des fausses détections dans le plan. Une approche à seuillage multiple permet de gagner en robustesse. La méthode est illustrée sur données de dauphins communs.

**Abstract** – The Estimation of time-frequency tracks is classically performed by the gathering of detections of peaks in time-frequency representations. It is then necessary to deal with the trade-off “false detections/ missed detections”. We propose a double hypothesis test with constant false detection probability. The first hypothesis test handles peak detection in the spectrogram, the second one focuses on the spatial organization of false detections. A multiple thresholding approach enhances the robustness of the method. The proposed method is illustrated on common dolphin data.

## 1 Introduction

La forte pression anthropique sur l’environnement marin, et en particulier sur les mammifères marins, provenant notamment du bruit rayonné dans l’océan (trafic maritime, sonar, prospection pétrolière, etc.) a motivé de nombreuses études scientifiques. Les conditions d’observations visuelles, difficiles et coûteuses, combinées au caractère loquace de nombreux mammifères marins font de l’acoustique passive un outil incontournable de ces études. L’analyse des signaux émis par les mammifères marins (modulations de fréquence non linéaire, impulsions large bande et « bursts ») est cependant difficile, et exige le développement d’outils de traitement du signal spécifiques. De nombreuses difficultés s’additionnent : 1. Non-stationnarité des signaux émis ; 2. Signaux multi-composantes/ Nombre de signaux inconnu/ Croisement des signaux dans le plan temps-fréquence ; 3. Grande variabilité des rapports

signal à bruit en fonction des individus et des enregistrements, 4. Bruit de fond coloré et/ou non stationnaire ; 5. Grandes bases de données ; 6. Forte directivité des sources/Fading ; 7. Peu d’*a priori* sur les signaux étudiés.

Un enjeu majeur est l’estimation des modulations de fréquence non linéaire présentes dans ces bases et ayant une grande variabilité [1]. Plusieurs approches développées dans la littérature visent d’abord à détecter les pics d’une représentation temps-fréquence (généralement le spectrogramme) par seuillage adaptatif ; puis à gérer le compromis détections manquées/ fausses détections par l’utilisation de filtres de Kalman [2], de filtres particuliers [3], de recherche dans des graphes [3], graphes minimaux [4] ou encore par l’utilisation d’opérateurs de morphologie mathématique [5]. Cependant, la détection des pics temps-fréquence est fréquemment réalisée de manière empirique. Ce papier introduit une méthode statistique robuste de segmentation du spectrogramme tirant parti de la différence

d'organisation spatiale entre vraies et fausses détections. Cette méthode est complémentaire de celles précédemment citées, et permet d'orienter ces méthodes vers des régions d'intérêt du plan temps-fréquence.

La méthode proposée consiste à réaliser un double test d'hypothèses binaires, appliqué sur chaque point temps-fréquence puis sur des sous-régions du plan. Une stratégie de seuillage multiple appliquée au premier test permet à la fois de gagner en robustesse et en souplesse d'utilisation. Le seuil du premier test d'hypothèse binaire est en effet choisi pour maximiser la probabilité de détection à probabilité de fausse alarme constante du second test d'hypothèse binaire. A partir d'un spectrogramme du signal étudié, l'algorithme de segmentation se décompose en 4 étapes : 1. Plusieurs spectrogrammes binaires sont générés en suivant une stratégie de Neyman-Pearson (N-P) [6] pour plusieurs probabilités de fausses alarmes ; 2. Une fenêtre glissante compte le nombre de détections dans le spectrogramme binaire; 3. Un second test d'hypothèses binaires est appliqué sur le nombre de détections compté dans la fenêtre glissante [7] ; 4. La segmentation retenue est celle qui maximise le nombre de sous-régions détectées à l'étape 3. La méthode est validée sur données réelles de vocalises de dauphins communs.

## 2 Méthode de segmentation

### 2.1 Seuillage multiple du spectrogramme

Un spectrogramme  $\mathcal{S}_x = \{s(n,k)\}$  est formé à partir de la série temporelle  $x(t)$  d'un signal  $e(t)$ , noyé dans un bruit additif Gaussien. La distribution des coefficients du bruit dans  $\mathcal{S}_x$  est localement modélisée par une loi du Chi2 à deux degrés de liberté – i.e. sur un voisinage temps-fréquence – afin de tenir compte de la non-stationnarité et de la non-blancher du bruit [6] :

$$p(s) = \frac{1}{\gamma_b(n,k)} \exp\left(-\frac{s}{\gamma_b(n,k)}\right) U(s), \quad (1)$$

avec  $\gamma_b(n,k)$  la densité spectrale de puissance du bruit au point  $(n,k)$  et  $U(s)$  la fonction de Heaviside. Nous estimons  $\gamma_b(n,k)$  par les Z-statistiques minimales [8]. Une stratégie de

Neyman-Person est utilisée pour seuiller le spectrogramme pour N valeurs de probabilité de fausse alarme  $\{p_{FA}^{(i)}\}_{i=1,\dots,N}$  (Figure 1.b.), produisant N spectrogrammes binaires  $\{\mathcal{S}_{bin}^{(i)}\}_{i=1,\dots,N}$  (Figure 1.c.).

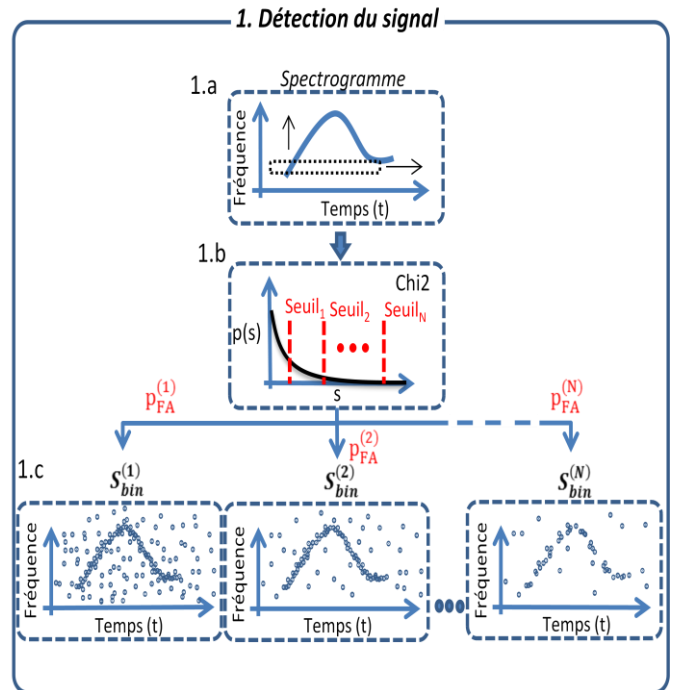


FIG. 1 - Méthode de détection statistique des points temps-fréquence dans le plan : a) Spectrogramme à segmenter, b) estimation de la loi du chi2 correspondante sur un voisinage temps-fréquence glissant (rectangle-tirets noir), définition de N seuils adaptatifs à probabilité de fausse alarme constante, et c) création de N spectrogrammes binaires.

### 2.2 Segmentation des N spectrogrammes binaires et choix de la probabilité de détection maximale.

Les hypothèses de ce second test sont : «  $H_0$  : région support de fausses détections uniquement » et «  $H_1$  : région support de signal ». Les détections associées aux fausses alarmes sont supposées dispersées sur tout le plan temps-fréquence alors que celles associées au signal sont supposées concentrées. Une fenêtre glissante est utilisée sur chacun des N spectrogrammes binaires  $\mathcal{S}_{bin}^{(i)}$  au sein de laquelle

les nombres de détections  $\Sigma^{(i)}$  sont comptés (Figure 2.a.). Le second test d'hypothèse est appliqué sur le nombre de détection dans cette sous-région glissante du spectrogramme binaire. La densité de probabilité de la variable aléatoire  $\Sigma^{(i)}$  est estimée par calcul de l'histogramme. Une valeur élevée de  $\Sigma^{(i)}$  est associée à l'hypothèse  $H_1$ .

La répartition sous-jacente, i.e. celle du nombre de fausses détections  $\Sigma_{FA}^{(i)}$  est modélisée par une loi Binomiale  $p_{\Sigma_{FA}^{(i)}}$  de paramètres  $A$  (i.e. l'aire des sous-régions) et  $p^{(i)} = P(\{s_{bin}^{(i)}(n, k) = 1\} | H_0)$  (Figure 2.b.) [7]. Une stratégie de Neyman-Pearson est utilisée pour résoudre le test d'hypothèse. Il est donc seulement nécessaire d'estimer le paramètre  $p^{(i)}$  de chaque distribution. L'estimation des  $p^{(i)}$  est obtenue par minimisation de la divergence de Kullback-Leibler ( $D_{KL}$ ) entre la distribution observée  $p_{\Sigma^{(i)}}$  et le modèle Binomial  $p_{\Sigma_{FA}^{(i)}}$  de paramètre  $A$  fixe et  $p_m$  variable:

$$\hat{p}^{(i)} = \operatorname{argmin}_{p_m} D_{KL}(p_{\Sigma_{FA}^{(i)}}(p_m) || p_{\Sigma^{(i)}}).$$

Une fois les distributions  $\{p_{\Sigma_{FA}^{(i)}}\}_{i=1, \dots, N}$  estimées, les sous-régions contenant du signal sont détectées à probabilité de fausse alarme constante  $p_{FA}^{Région}$  (Neyman-Pearson) et unique pour l'ensemble des spectrogrammes à segmenter. Les  $N$  spectrogrammes segmentés  $\{S_{seg}^{(i)}\}_{i=1, \dots, N}$  sont générés (Figure 2.c).

Parmi les  $N$  spectrogrammes segmentés (régions grisées de la Figure 2.c.), la segmentation finale  $S_{seg}^{(k_0)}$  est celle qui maximise le nombre de détections en sortie du second test d'hypothèse. Ce choix permet d'augmenter la probabilité de détection tout en garantissant une probabilité de fausse alarme  $p_{FA}^{Région}$ .

### 3 Validation sur données réelles

La méthode est validée sur données réelles d'enregistrements de vocalises de dauphins de 86835 points, échantillonné à 32.768kHz. Le spectrogramme est généré avec une fenêtre de Hanning de 512 points et un recouvrement de 50% et calculé pour 512 points fréquentiels.

Plusieurs vocalises de dauphins sont visibles sur le spectrogramme. Le voisinage temps-fréquence pour l'estimation de  $\gamma_b(n, k)$  est [10 s x 500 Hz]. Le spectrogramme est seuillé pour 32 valeurs de probabilité de fausses alarmes logarithmiquement réparties entre  $10^{-1}$  et  $10^{-5}$ . La fenêtre glissante est de taille 5x5 bins. Les résultats sont relativement stables pour une assez large gamme de tailles de fenêtres (de 3x3 bins jusqu'à et 15x15), une grande taille de fenêtre aboutira à une segmentation plus « grossière ». La segmentation est réalisée avec  $p_{FA}^{Région} = 0.01$ . La figure 3 est un exemple de résultat de segmentation et montre les différentes étapes de segmentation : binarisation du spectrogramme à probabilité de fausse alarme constante, carte de comptage des détections, puis décision sur la segmentation. Les vocalises sont correctement segmentées, peu de fausses détections sont présentes.

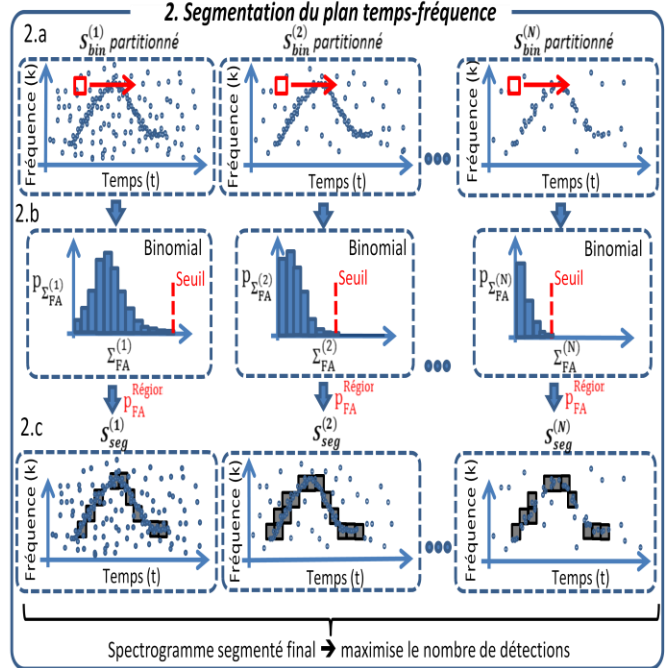


FIG. 2 – Méthode de segmentation statistique du plan temps-fréquence binaire : a) Comptage du nombre de détections dans une fenêtre glissante pour chacun des  $N$  spectrogrammes binaires (exemple avec une fenêtre carrée), b) tracé des histogrammes du nombre de détections dans cette fenêtre et c) définition d'un seuil à probabilité de fausse alarme  $p_{FA}^{Région}$  unique pour toutes les segmentations.

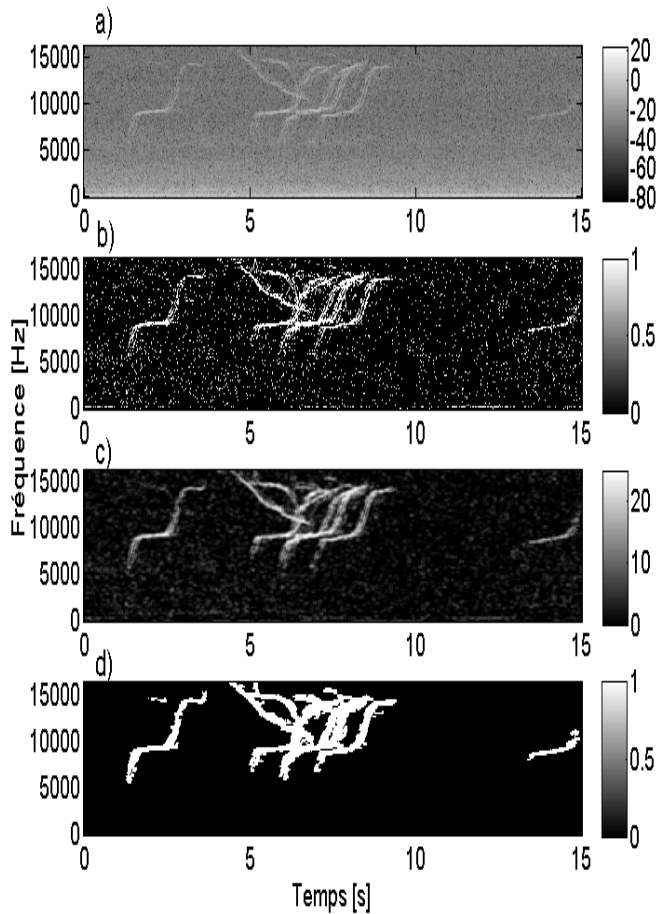


FIG. 3 - Segmentation du spectrogramme, a) Signaux de dauphins communs, spectrogramme (en dB) calculé sur 512 points, recouvrement de 50% entre les fenêtres de hanning de 512 points ; b) Détection des pics temps-fréquence, la probabilité de fausse alarme est 0.1, ce qui correspond à la probabilité choisie maximisant le nombre de détections en sortie du processus ; c) Carte temps-fréquence dénombrant le nombre de détections au sein de la fenêtre glissante (carré 5x5 bins) pour la carte binaire montrée au dessus, les valeurs sont comprises strictement entre 0 et 25. La distribution des fausses alarmes est supposée binomiale ; d) Segmentation de la carte de dénombrement, la probabilité de fausse alarme est fixée égale à 0.01.

## 4 Conclusion

Une méthode de segmentation robuste basée sur un double test d'hypothèse binaire est proposé. Le spectrogramme est d'abord analysé et seuillé de manière locale, l'organisation spatiale des

fausses détections est ensuite analysée. Une décision basée sur un test statistique sur les régions permet alors de faire la distinction entre régions abritant de vraies détections et celles ne contenant que des fausses alarmes. La méthode est peu paramétrée et donne de bons résultats pour une assez large gamme de paramètres. Aucun processus de pré-conditionnement du spectrogramme classiquement utilisé n'est requis (égalisation, soustraction du bruit etc.)

## 5 Remerciements

Ce travail est supporté par la DGA par l'intermédiaire d'une Bourse. Les auteurs remercient la DGA pour le financement des ces travaux dans le cadre de l'appel ANR ASTRID référencé ANR-12-ASTR-0021-03 "MER CALME".

## 6 Références

- [1] Lampert, T. A., and O'Keefe, S. EM. "A survey of spectrogram track detection algorithms." *Applied acoustics* 71.2 (2010): 87-100.
- [2] Mallawaarachchi, A., Ong, S. H., Chitre, M., & Taylor, E. (2008). Spectrogram denoising and automated extraction of the fundamental frequency variation of dolphin whistles. *The Journal of the Acoustical Society of America*, 124, 1159.
- [3] Roch, M. A., Brandes, T. S., Patel, B., Barkley, Y., Baumann-Pickering, S., & Soldevilla, M. S. (2011). Automated extraction of odontocete whistle contours. *The Journal of the Acoustical Society of America*, 130, 2212.
- [4] Michel, Olivier, Patrick Flandrin, and Alfred O. Hero III. "Automatic extraction of time-frequency skeletons with minimal spanning trees." *Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on*. Vol. 1. IEEE, 2000.
- [5] Simard, Y., Roy, N., Giard, S., Gervaise, C., Conversano, M., & Ménard, N. (2010). Estimating whale density from their whistling activity: Example with St. Lawrence beluga. *Applied Acoustics*, 71(11), 1081-1086.
- [6] Huillery, J., Millioz, F., & Martin, N. (2008). On the description of spectrogram probabilities with a chi-squared law. *Signal Processing, IEEE Transactions on*, 56(6), 2249-2258.
- [7] Dadouchi, F., Gervaise, C., Ioana, C., Huillery, J. & Mars, J. I. "Automated segmentation of linear time frequency representations of marine mammal sounds", *The Journal of the Acoustical Society of America* (In press)
- [8] Huillery J. (2008), "Support temps-fréquence d'un signal en présence de bruit additif gaussien", PHD thesis, Grenoble INP, France