

Distances et Critères de Contraste dans le Plan Temps-Fréquence

Manuel DAVY et Christian DONCARLI

IRCyN, UMR CNRS 6597
1 rue de la Noë, BP 92101, 44321 Nantes Cedex 03, France
Manuel.Davy@ircyn.ec-nantes.fr, Christian.Doncarli@ircyn.ec-nantes.fr

Résumé – La classification des signaux non stationnaires dans le plan temps-fréquence permet de résoudre les cas où les méthodes avec modèle sont en défaut. Nous proposons ici de déterminer la meilleure combinaison distance temps-fréquence / Représentation Temps-Fréquence en regard de plusieurs critères de contraste.

Abstract – The classification of non-stationary signals in the time-frequency plane is an efficient approach when parametric methods are not available. We propose in this paper to determine the best combination time-frequency distance / kernel with respect to several contrast criteria.

1 Introduction

La classification de signaux non stationnaires est un problème courant qui n'est pas complètement résolu par les approches paramétriques [4]. Une alternative naturelle consiste alors à utiliser les représentations temps-fréquence (RTF), initialement prévues pour l'analyse [6], comme outils de décision. La détection optimale d'une onde connue noyée dans du bruit a conduit aux filtres adaptés temps-fréquence [5] mais le problème général de la classification supervisée à partir d'un ensemble d'apprentissage nécessite la définition de distances dans l'ensemble de représentation (les RTF). Le premier objectif de ce travail est de proposer une revue des outils (distances L_q , corrélation, distances entre lois de probabilité, divergences spectrales) utilisables dans ce contexte. Le deuxième point concernera la proposition de critères de contraste permettant d'apprécier et d'optimiser les performances de la classification (en termes de mal classés). Des résultats obtenus sur de larges familles de signaux viennent illustrer ce travail.

2 Distances pour la comparaison de RTF

La détection optimale dans le plan temps-fréquence d'une onde connue conduit à l'écriture d'un filtre adapté (ou encore corrélateur temps-fréquence), ce qui peut être vu, dans le cas général de la classification avec apprentissage supervisé, comme une restriction à la classe des fonctions discriminantes linéaires. Les procédures générales (non linéaires) nécessitent alors la définition de distances (on utilisera ce vocabulaire aussi bien pour les divergences que pour les vraies distances) et la première idée consiste

à utiliser des outils issus de la comparaison des images. Toutefois, les RTF du groupe de Cohen, objet de cette étude, ayant été construites à l'origine comme des distributions de probabilité conjointes, il semble tout aussi naturel d'utiliser des distances entre lois de probabilité. Enfin, les RTF marginalisables pouvant être interprétées en termes de distributions spectrales instantanées, on s'intéressera aussi aux divergences spectrales.

2.1 Distances L_q , distance quadratique et corrélation

Pour comparer deux RTF, de nombreux auteurs ont tout simplement choisi la distance Euclidienne [1] qui s'intègre dans la famille des distances L_q dont l'expression générale est :

$$d_{L_q}(\mathcal{C}_{x_1}^\phi, \mathcal{C}_{x_2}^\phi) = \left[\iint |\mathcal{C}_{x_1}^\phi(t, f) - \mathcal{C}_{x_2}^\phi(t, f)|^q dt df \right]^{\frac{1}{q}} \quad (1)$$

où \mathcal{C}_x^ϕ désigne la RTF du signal x de noyau ϕ . On peut particulariser :

$q = 1$ distance absolue moyenne ou distance de Manhattan,

$q = 2$ distance quadratique moyenne (distance Euclidienne),

$q \rightarrow \infty$ déviation maximale.

La distance par corrélation, très utilisée, se déduit facilement de la distance quadratique :

$$d_{\text{corr}}(\mathcal{C}_{x_1}^\phi, \mathcal{C}_{x_2}^\phi) = \frac{(d_{L_2}(\mathcal{C}_{x_1}^\phi, \mathcal{C}_{x_2}^\phi))^2}{\iint |\mathcal{C}_{x_1}^\phi(t, f)|^2 dt df + \iint |\mathcal{C}_{x_2}^\phi(t, f)|^2 dt df} \quad (2)$$

TAB. 1: Quelques φ -divergences.

Distance	$d(\mathcal{C}_{x_1}^\phi, \mathcal{C}_{x_2}^\phi) =$
<i>Kolmogorov</i>	$\int \int \mathbf{N}\mathcal{C}_{x_1}^\phi - \mathbf{N}\mathcal{C}_{x_2}^\phi $
<i>Küllback</i>	$\int \int (\mathbf{N}\mathcal{C}_{x_1}^\phi - \mathbf{N}\mathcal{C}_{x_2}^\phi) \log \frac{\mathbf{N}\mathcal{C}_{x_1}^\phi}{\mathbf{N}\mathcal{C}_{x_2}^\phi}$
<i>Chernoff</i>	$-\log \left[\int \int \mathbf{N}\mathcal{C}_{x_1}^{\phi m} \cdot \mathbf{N}\mathcal{C}_{x_2}^{\phi 1-m} \right]$
<i>Matusita</i>	$\sqrt{\int \int \int \left \sqrt{\mathbf{N}\mathcal{C}_{x_1}^\phi} - \sqrt{\mathbf{N}\mathcal{C}_{x_2}^\phi} \right ^2}$

2.2 Distances entre lois de probabilité

Une loi conjointe $P(a, b)$ définie sur un domaine \mathcal{D} doit être non négative :

$$\forall (a, b) \in \mathcal{D}, P(a, b) \geq 0 \quad (3)$$

et unitaire :

$$\int_{\mathcal{D}} P(a, b) da db = 1 \quad (4)$$

Les spectrogrammes à fenêtre unitaire étant les seules RTF bilinéaires satisfaisant ces conditions, on propose pour toute RTF la procédure de normalisation suivante :

$$\mathbf{N}\mathcal{C}_x^\phi(t, f) = \frac{|\mathcal{C}_x^\phi(t, f)|}{\int \int |\mathcal{C}_x^\phi(s, \nu)| ds d\nu} \quad (5)$$

Cette normalisation conserve l'intégralité du contenu informatif de la RTF et doit être préférée à tout autre (troncature des termes négatifs, suppression des interférences ...). En effet l'objectif de comparaison poursuivi ici est totalement différent de celui de lisibilité habituellement considéré.

On peut alors s'intéresser à une large famille de distances entre lois de probabilité, appelées φ -divergences, qui tiennent compte du rapport de vraisemblance et de sa dispersion [2]. Etant donnée une fonction continue strictement convexe $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}$ et une fonction croissante $\psi : \mathbb{R} \rightarrow \mathbb{R}$, l'expression générale des φ -divergences est :

$$d_{\varphi\text{-div.}}(P_1, P_2) = \psi \left\{ \mathbf{E}_1 \left[\varphi \left(\frac{P_2(a, b)}{P_1(a, b)} \right) \right] \right\} \quad (6)$$

où \mathbf{E}_1 représente l'espérance mathématique respectant la loi P_1 . Les distances sont utilisées ici comme fonctions discriminantes, ce qui fait que toutes les distances utilisant la même fonction φ conduisent au même classifieur quelle que soit la fonction croissante ψ . Quelques exemples sont donnés dans le tableau 1.

2.3 Divergences spectrales

Une extension immédiate consiste à appliquer les distances L_q aux RTF normalisées et à définir :

$$d_{\mathbf{N}L_q}(\mathcal{C}_{x_1}^\phi, \mathcal{C}_{x_2}^\phi) = d_{L_q}(\mathbf{N}\mathcal{C}_{x_1}^\phi, \mathbf{N}\mathcal{C}_{x_2}^\phi) \quad (7)$$

Le cas $q = 1$ permet de retrouver la distance de Kolmogorov et dans le même ordre d'idées on peut définir la déviation spectrale logarithmique :

$$d_{\mathbf{N}LSD_{\text{Dev}_q}}(\mathcal{C}_{x_1}^\phi, \mathcal{C}_{x_2}^\phi) = d_{L_q}(\log[\mathbf{N}\mathcal{C}_{x_1}^\phi], \log[\mathbf{N}\mathcal{C}_{x_2}^\phi]) \quad (8)$$

dont l'utilisation pour discriminer des signaux bruités est toutefois à éviter [9].

3 Règles de décision

Dans le domaine temps-fréquence, l'écriture de la règle de décision optimale (Bayes) n'est pas envisageable, faute d'avoir accès aux densités de probabilité des pixels temps-fréquence. Nous proposons l'étude de deux types de règles : affectation d'un individu à la classe dont l'élément représentatif est le plus proche, ou k plus proches voisins (k -ppv). Dans ce papier, nous nous intéressons à la première, l'expérience ayant révélé que les k -ppv ne conduisent pas à des résultats supérieurs, tout en étant beaucoup plus coûteux en temps de calcul. L'élément représentatif de chaque classe sera la moyenne des représentations temps-fréquence.

4 Critères de contraste

L'optimisation des performances de la classification en fonction du type de distance et du type de noyau est basée sur la définition d'un critère de contraste entre classes d'apprentissage dont la maximisation doit conduire au plus bas taux de mal classés dans l'ensemble de test. On se propose de discuter trois critères possibles. Le problème comporte N classes dont chacune est constituée de n_i signaux d'apprentissage notés x_i^k , $k \in [1, n_i]$.

4.1 Critère du premier ordre (CPO)

Le choix le plus simple consiste à maximiser la distance entre les éléments représentatifs de chaque classe d'apprentissage :

$$\mathcal{K}_{\text{CPO}}(\phi) = \sum_{i=2}^N \sum_{j=1}^{i-1} \frac{n_i + n_j}{n} d(\bar{\mathcal{C}}_i^\phi, \bar{\mathcal{C}}_j^\phi) \quad (9)$$

où $n = \sum_{i=1}^N n_i$ et $\bar{\mathcal{C}}_i^\phi$ est l'élément représentatif de la classe i , obtenu par moyenne de son ensemble d'apprentissage.

Ce critère, utilisé dans [8] ne tient pas compte de la dispersion des classes et on peut proposer de compacter les individus autour de leur représentant.

4.2 Critère de type Fisher (CTF)

Dans le but de séparer les classes de l'ensemble d'apprentissage tout en réduisant leur dispersion, on propose de calculer les distances séparant chaque élément de l'ensemble d'apprentissage des éléments représentatifs des classes. Soit \mathbf{d}_{ij}^ϕ la variable aléatoire correspondant à la distance séparant un élément de la classe i du représentant

de la classe j . Son espérance mathématique sera notée $\langle \mathbf{d}_{ij}^\phi \rangle$. La probabilité d'erreur de classification \mathbf{P}_e est donnée par :

$$\mathbf{P}_e = \sum_{i=1}^N \sum_{j \neq i}^N \mathbf{P} \left[\mathbf{d}_{ij}^\phi < \mathbf{d}_{ii}^\phi \right] \quad (10)$$

La minimisation de cette probabilité peut-être obtenue en maximisant $\langle \mathbf{d}_{ij}^\phi \rangle$, $j \neq i$ et en minimisant $\langle \mathbf{d}_{ii}^\phi \rangle$, ce qui conduit au critère de type Fisher (à maximiser) :

$$\mathcal{K}_{\text{CTF}}(\phi) = \frac{\sum_{i=1}^N n_i \left[\sum_{j \neq i}^N \langle \mathbf{d}_{ij}^\phi \rangle \right]}{\sum_{i=1}^N n_i \langle \mathbf{d}_{ii}^\phi \rangle} \quad (11)$$

Ce critère est utilisé dans [8] avec une distance par corrélation et dans [3] avec une distance de Kolmogorov. Utilisant l'inégalité triangulaire (pour les distances la vérifiant), on peut aisément montrer, dans le cas de deux classes d'apprentissage de même cardinal :

$$\mathcal{K}_{\text{CTF}}(\phi) \leq 1 + \frac{2 d(\bar{\mathcal{C}}_1^\phi, \bar{\mathcal{C}}_2^\phi)}{\langle \mathbf{d}_{11}^\phi \rangle + \langle \mathbf{d}_{22}^\phi \rangle} \quad (12)$$

Toutefois ce critère manipule des distances moyennes, ce qui peut laisser prévoir la mauvaise classification de certains éléments excentrés, comme on le montrera dans les essais sur signaux synthétiques.

Une autre forme de ce critère, traduit dans le plan des ambiguïtés, a été proposée dans [7] et consiste à sélectionner les pixels les plus discriminants, mais sans tenir compte de la connexité. La classification utilise alors la distance de Mahalanobis dans l'espace des pixels sélectionnés pour chaque signal. Cette technique de "noyau binaire" est extrêmement rapide et conduit à de bons résultats.

4.3 Critère de probabilité d'erreur (CPE)

On propose maintenant de considérer la variable "différence de distances" : $\mathbf{d}_{ij}^\phi - \mathbf{d}_{ii}^\phi$, $j \neq i$. Le théorème de la limite centrale permet d'approcher la distribution de cette variable par une loi gaussienne (ce qui sera vérifié dans la partie 5) et on peut donc classiquement majorer la probabilité d'erreur :

$$\begin{aligned} \mathbf{P}_e(i, j) &= \mathbf{P} \left[\mathbf{d}_{ij}^\phi - \mathbf{d}_{ii}^\phi < 0 \right] \\ &\leq \frac{1}{2} \cdot e^{-\frac{1}{2} \frac{\langle \mathbf{d}_{ij}^\phi - \mathbf{d}_{ii}^\phi \rangle^2}{\langle \mathbf{d}_{ij}^\phi - \mathbf{d}_{ii}^\phi \rangle^2}} \end{aligned} \quad (13)$$

ce qui conduit, pour des classe équiprobables *a priori* au critère suivant :

$$\mathcal{K}_{\text{CPE}}(\phi) = -\log \left(\frac{1}{2} \sum_{i=1}^N \sum_{j \neq i}^N e^{-\frac{1}{2} \frac{\langle \mathbf{d}_{ij}^\phi - \mathbf{d}_{ii}^\phi \rangle^2}{\langle \mathbf{d}_{ij}^\phi - \mathbf{d}_{ii}^\phi \rangle^2}} \right) \quad (14)$$

5 Résultats

Le comportement des critères et distances présentés est étudié sur un exemple de signaux synthétiques.

5.1 Etude des critères de contraste

Les signaux sont constitués d'une composante continue de fréquence réduite 0.25, d'un bruit blanc additif et d'une composante chirp de pente décroissante aléatoire (se reporter à la figure 1) différente selon la classe et l'individu. L'ensemble d'apprentissage comporte 50 individus de chaque classe, l'ensemble de test 2×1000 autres signaux.

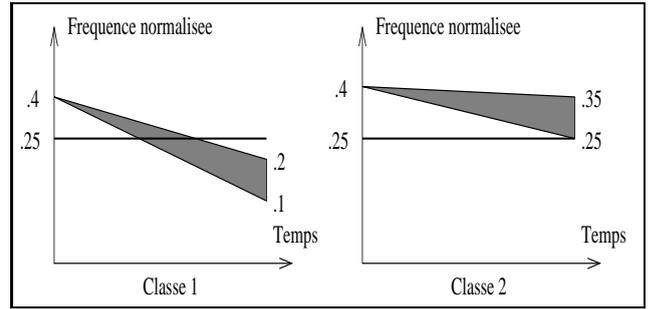


FIG. 1: Signaux synthétiques en deux classes. Les signaux sont constitués d'une composante stationnaire, un chirp décroissant et un bruit additif (RSB = 0 dB). La différence entre les classes est la pente (aléatoire dans la zone grisée) du chirp.

La figure 2 illustre l'efficacité des critères présentés. Il ressort nettement que le critère de probabilité d'erreur est le plus efficace, car fortement lié à l'erreur de classification. Par ailleurs, la gaussianité des variables aléatoires est confirmée par les résultats présentés figure 3.

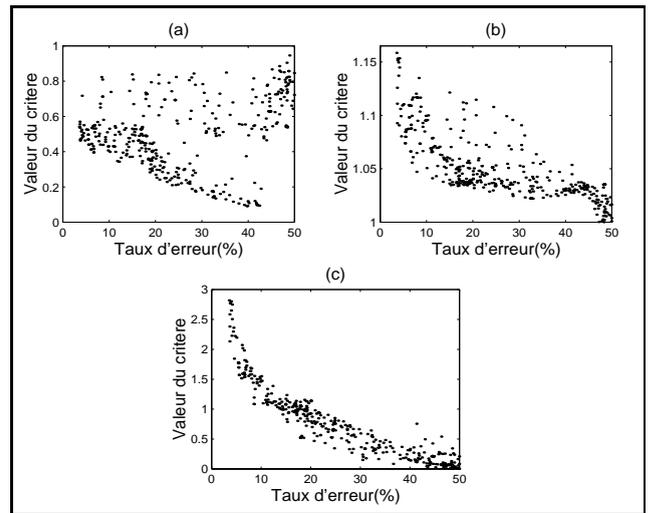


FIG. 2: Etude des critères. 400 noyaux différents sont testés. Pour chaque noyau, les trois critères sont calculés, ainsi que le taux de mal classés. (a) CPO, (b) CTF, (c) CPE, pointillés : $-\log(\text{taux d'erreur})$.

TAB. 2: Résultats de classification obtenus par différentes méthodes.

Noyau	Distance	Taux d'erreur
Wigner-Ville	Corrélation	11.8 %
Wigner-Ville	Kolmogorov	19.8 %
Pseudo Wigner-Ville Lissé	Corrélation	12.6 %
Noyau binaire	Mahalanobis	4.4%
Noyau Radial. gaussien optimisé (Crit. de Fisher)	Kolmogorov	3.1%
Cet article	Optimisation noyau + distance	2.05%

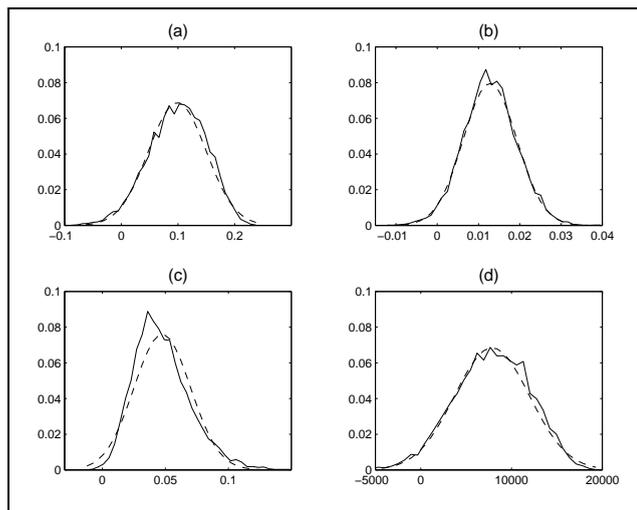


FIG. 3: (Trait plein) Distribution de la variable $d_{12}^{\phi} - d_{11}^{\phi}$ pour 2×10000 réalisations. (pointillés) : la distribution gaussienne obtenue avec les mêmes moyenne et variance. (a) Distance de Kolmogorov + Noyau radialement gaussien (b) Corrélation + Wigner-Ville. (c) Divergence de Jensen et noyau Choi-Williams généralisé à une branche (d) distance L_1 + Noyau radialement gaussien.

5.2 Comparaison des méthodes

Des méthodes présentées par le passé sont comparées ici. Le tableau 2 donne les résultats obtenus. On peut noter que le résultat est fortement lié :

- à la distance (lignes 1 et 2),
- au noyau (lignes 2 et 5),
- à la combinaison distance-noyau (lignes 2 et 6)

Pour confirmer ce résultat optimal, on optimise 80 noyaux (10 distances et 8 types de noyaux paramétriques) sur un ensemble d'apprentissage de 2×50 signaux.

Ces 80 noyaux optimisés sont ensuite utilisés pour classer un ensemble de test de 2×10000 signaux. Le critère est estimé sur 2×10000 autres signaux. La figure 4 rassemble les résultats par distance utilisée. Les φ -divergences et les distances entre RTF normalisées sont clairement les plus efficaces.

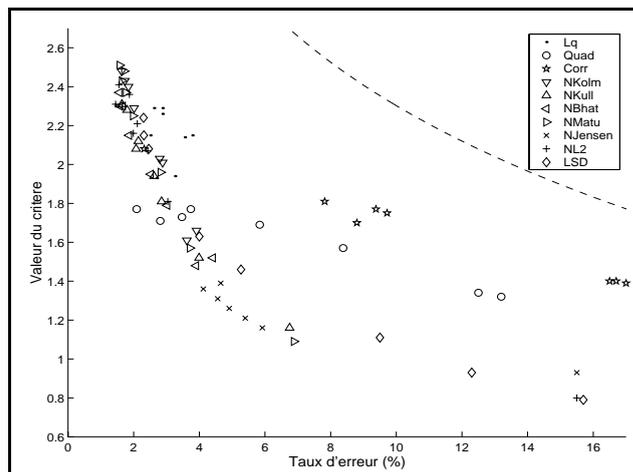


FIG. 4: Résultats de classification en utilisant les noyaux optimisés avec 2×50 signaux. Les noyaux paramétriques sont : Radialement gaussien à 3,5 ou 7 descripteurs [3], Choi-Williams généralisés à 1, 2 ou 3 branches et noyau exponentiel multiforme symétrique ou non.

6 Conclusion

La technique proposée permet une amélioration significative des performances des classificateurs de signaux non-stationnaires, notamment par l'utilisation de φ -divergences et a été appliquée avec succès à des signaux réels (détection de défaut de haut-parleurs, reconnaissance de locuteur, surveillance d'engrenages, ...).

Références

- [1] L. Atlas, J. Droppo, and J. McLaughlin. Optimizing time-frequency distributions for automatic classification. In *SPIE*, 1997.
- [2] M. Basseville. Distance measures for signal processing and pattern recognition. *Signal Processing*, 18(4):349–369, 1989.
- [3] M. Davy and C. Doncarli. Optimal kernels of time-frequency representations for signal classification. In *Int. Symp. on TFTS*, pages 581–584, 1998.
- [4] C. Doncarli and E. Le Carpentier. An optimal approach for random signals classification. *IEEE Trans. on PAMI*, 13:1192–1196, 1991.
- [5] P. Flandrin. A time-frequency formulation of optimal detection. *IEEE Trans. on Acous., Speech and Sig. Proc.*, 36(9):1337–1384, September 1992.
- [6] P. Flandrin. *Temps-Fréquence*. Hermès, 1993.
- [7] B. W. Gillespie and L. E. Atlas. Data-driven optimization of time and frequency resolutions for radar transmitter identification. In *SPIE*, 1998.
- [8] C. Heitz. Optimum time-frequency representations for the classification and detection of signals. *Applied Signal Processing*, (3):124–143, 1995.
- [9] I. Vincent. *Classification de Signaux Non-stationnaires*. PhD thesis, Ecole Centrale de Nantes, France, 1995.