Application des machines à vecteurs support mono-classe à l'indexation en locuteurs de documents audio

Belkacem FERGANI [1,2]*, Manuel DAVY* [2] *et Amrane HOUACINE* [1]

[1] LCPTS - USTHB, B.P. 32,El Alia, Bab Ezzouar, Alger, ALGERIE
[2] LAGIS/CNRS, BP 48, Cité Scientifique, 59651 Villeneuve d'Ascq cedex, FRANCE
bfergani2001@yahoo.fr,a_houacine@lycos.fr, Manuel.Davy@ec-lille.fr

ABSTRACT

This paper addresses a new approach based on the Kernel Change Detection algorithm introduced recently by Desobry et al. This new algorithm is applied to the speaker change detection and clustering tasks, which are the key issues in any audio indexing process. We show the efficiency of the method through several experiments using RT'03S NIST data. We discuss also the parameters tuning and compare the results to the well known GLR-BIC algorithm.

1. INTRODUCTION

Avec la multiplication de sources de données multimédia et le développement des techniques de numérisation de l'information, nous assistons à une explosion des bases de données d'archivage. De ce fait se pose un problème crucial : Comment accéder facilement et rapidement à l'information recherchée ? Ces deux critères (rapidité et facilité d'accès) sont incontournables pour toute requête d'utilisateur.

L'indexation en locuteur d'un signal sonore consiste à "structurer" ce signal selon l'information véhiculée par le locuteur. Dans ce contexte, la segmentation en locuteur constitue une étape préalable et déterminante pour la suite du processus d'indexation. Elle consiste à d'abord découper le signal audio en zones homogènes contenant uniquement les informations relatives à un seul locuteur. Cette étape est suivie du regroupement (clustering) de ces segments afin d'assembler les zones appartenant à un seul locuteur.

Ce problème a déjà fait l'objet de nombreuses études (voir par exemple [4, 5] et dans les références qui y sont indiqués. Les techniques standard utilisent généralement des descripteurs acoustiques (souvent des MFCC et leurs dérivées) puis appliquent deux fenêtres d'analyse glissantes sur les données de part et d'autre de l'instant courant. Etant donné les vecteurs acoustiques $\mathbf{x}(n)$, $n = 1, 2, \ldots$, la fenêtre d'analyse située avant l'instant d'analyse n définit l'ensemble passé immédiat $\mathsf{X}_\mathrm{p}(n) = \{\mathbf{x}(n - m_\mathrm{p}), \ldots, \mathbf{x}(n-1)\}$ de m_p vecteurs acoustiques, tandis que l'autre fenêtre contient m_f vecteurs acoustiques représentant l'ensemble futur immédiat $\mathsf{X}_\mathrm{f}(n) = \{\mathbf{x}(n + 1), \ldots, \mathbf{x}(n + m_\mathrm{f})\}$. L'objectif de ces techniques classiques est de comparer les ensembles $\mathsf{X}_\mathrm{p}(n)$ et $\mathsf{X}_\mathrm{f}(n)$. Ceci est réalisé au moyen de méthodes à base de rapport de vraisemblance généralisé (RVG) soit directement [5] ou indirectement comme dans l'approche par critère d'information bayésien (BIC) notée RVG-BIC et adoptée comme référence dans ce papier [2]. Les méthodes à base de RVG nécessitent la connaissance d'un modèle de la distribution de probabilité des données $\mathbf{x}(n)$. Les modèles gaussien ou mélange de gaussiennes ont été largement exploités dans ce cadre.

Dans cette communication, nous proposons d'appliquer l'algorithme basé sur une méthode à noyau introduit dans [3] aux tâches de détection de ruptures et de regroupement dont le résultat d'ensemble est connu sous le vocable de segmentation en locuteurs. Différemment des approches citées précédemment, notre algorithme exploite une méthode à base de Machines à Vecteurs de Support mono-classe (SVM-1) dont la finalité est de comparer les ensembles $\mathsf{X}_\mathrm{p}(n)$ et $\mathsf{X}_\mathrm{f}(n)$ à chaque instant d'analyse au moyen d'une mesure de similarité. Dans ce sens, notre méthode reste semblable aux méthodes classiques à base de RVG, néanmoins notre approche exploite les informations extraites de l'entraînement de deux (SVM-1), dont l'avantage principal est de contrôler la complexité du modèle ajusté aux données et de prendre en compte l'information paramétrée selon diverses configurations et tailles des vecteurs acoustiques.

Dans la section suivante, nous rappelons le principe de l'algorithme de détection de rupture basé (SVM-1), puis la section 3 présente la méthodologie d'application à la segmentation en locuteurs en détaillant le choix des paramètres de détection de rupture et de regroupement. La section 4 présente les résultats d'application de notre méthode aux signaux de la base de données NIST RT'03S [6], en comparaison avec la méthode RVG-BIC, selon le critère d'erreur définie par NIST, finalement la dernière section 5 présente les conclusions et perspectives.

2. UN ALGORITHME POUR LA DÉTECTION DE RUPTURES BASÉ SVM-1

Nous partons de l'hypothèse que les vecteurs acoustiques $\mathbf{x}_1, \ldots, \mathbf{x}_m$ sont générées identiquement et indépendemment par une distribution de probabilité (ddp) inconnue $\mathsf{p}(\mathbf{x})$. Le principe de l'algorithme développé dans [3] est de comparer les ensembles $\mathsf{X}_\mathrm{p}(n)$ et $\mathsf{X}_\mathrm{f}(n)$ au travers de la comparaison de leurs support de ddp. On définit le support d'une ddp S^λ par l'ensemble des points de l'espace des vecteurs acoustiques $\mathcal{X}$ telle que $\mathsf{p}(\mathbf{x}) \geq \lambda$, avec λ une constante positive quelconque.

2.1. *Les Machines à Vecteurs Support mono-classe (SVM-1)*

Soit une fonction réelle symétrique appelée noyau définie dans $\mathcal{X}$. Dans la suite, de cette communication, nous considérons un noyau de forme gaussien, comme suit :

$$\mathsf{k}(\mathbf{x}_1, \mathbf{x}_2) \;=\; \exp -\frac{1}{2\sigma^2}\|\mathbf{x}_1 - \mathbf{x}_2\|^2_{\mathcal{X}} \qquad (1)$$

avec $\|\cdot\|^2_{\mathcal{X}}$ est une norme définie sur $\mathcal{X}$. Le modèle SVM-1 estime le support de la ddp comme suit :

$$\mathsf{S}^{\lambda} \;=\; \{\mathbf{x} \in \mathcal{X} | \mathsf{f}^{\lambda}(\mathbf{x}) + b \geq 0\} \qquad (2)$$

Ce problème d'estimation du support de la ddp revient à estimer une fonction dans l'espace augmenté $\mathcal{H}$ (hilbertien et à noyau reproductible induit par $\mathsf{k}(\mathbf{x}_1, \mathbf{x}_2)$) , proche du support de la ddp recherchée. On montre dans [7] que les fonctions minimisant le risque régularisé s'écrivent en fonction de $\mathbf{x}$ comme :

$$\mathsf{f}^{\lambda}(\mathbf{x}) + b \;=\; \sum_{i=1}^{m} \alpha_i \mathsf{k}(\mathbf{x}, \mathbf{x}_i) + b \qquad (3)$$

les coefficients de pondération α_i sont dénommés les multiplicateurs de lagrange. Le paramètre ν (réel positif) joue un rôle de contrôle des vecteurs supports. Ainsi, choisir $\nu = 0.2$ équivaut à admettre 20% de vecteurs acoustiques dans X comme "outliers". A ce problème d'optimisation correpond un problème dual plus simple à résoudre puisque quadratique avec des contraintes linéaires :

$$\begin{array}{ll} \text{Minimiser} & \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j \mathsf{k}(\mathbf{x}_i, \mathbf{x}_j) \text{ w.r.t. } \{\alpha_1, \ldots, \alpha_m\} \\ \text{avec} & 0 \leq \alpha_i \leq \frac{1}{\nu m} \text{ pour } i = 1, \ldots, m \\ \text{et} & \sum_{i=1}^{m} \alpha_i = 1 \end{array} \qquad (4)$$

le modèle SVM-1 admet une simple interprétation géométrique dans l'espace augmenté $\mathcal{H}$: premièrement, les vecteurs acoustiques dans $\mathcal{X}$ sont projetés vers $\mathcal{H}$ au moyen de l'application :$\mathbf{x} \;\rightarrow\; \mathsf{k}(\mathbf{x}, \cdot)$. Deuxièment, les vecteurs acoustiques dans $\mathcal{H}$ sont de norme unitaire lorsque le noyau gaussien est choisi, car $\|\mathsf{k}(\mathbf{x}, \cdot)\|^2_{\mathcal{H}} = \langle \mathsf{k}(\mathbf{x}, \cdot), \mathsf{k}(\mathbf{x}, \cdot)\rangle_{\mathcal{H}} = \mathsf{k}(\mathbf{x}, \mathbf{x}) = 1$ (propriété du noyau reproductible dans l'espace hilbertien), ansi ces vecteurs sont situés sur la surface d'une hypersphère de rayon unité. Troisièment, la résolution de Eq. (4) peut se ramener à trouver dans $\mathcal{H}$ un hyperplan orthogonal à $\mathsf{f}(\cdot)$ tel que celui-ci serait le plus éloigné de l'origine, séparant ainsi les données d'apprentissage $\mathsf{k}(\mathbf{x}_i, \cdot)$ entre deux classes,— voir figure 1.

2.2. *une mesure de similarité basée sur SVM-1*

Cette mesure est construite sur le principe que les ensembles $\mathsf{X}_{\mathsf{p}}(n)$ and $\mathsf{X}_{\mathsf{f}}(n)$ sont similaires si et seulement si les supports de densités estimés sont similaires selon un certain critère de similarité. Notons, que dans l'espace initial des données $\mathcal{X}$, la forme des contours de décision

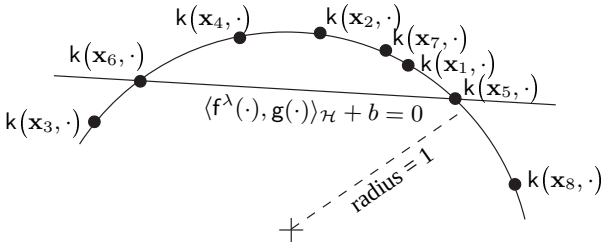

FIG. 1: Interprétation géométrique du modèle SVM-1 dans $\mathcal{H}$. Les vecteurs acoustiques projetés $\mathsf{k}(\mathbf{x}_i, \cdot), i = 1, \ldots, m$ sont situés sur une hypersphère de rayon unité. La fonction $\mathsf{f}^{\lambda}(\cdot)$ et b définissent un hyperplan d'équation $\langle \mathsf{f}^{\lambda}(\cdot), \mathsf{g}(\cdot)\rangle_{\mathcal{H}} + b = 0$. La majorité des données est située du côté de l'hyperplan ne comprenant pas l'origine de l'hypersphère. les coefficients α_i correspondants sont nulles ($i \in \{1, 2, 4, 7\}$), tandis que les points marginaux $\mathsf{k}(\mathbf{x}_3, \cdot)$ et $\mathsf{k}(\mathbf{x}_8, \cdot)$ sont situés du côté de l'hyperplan comprenant l'origine $\alpha_3 = \alpha_8 = 1/\nu m$. Les points situés sur l'intersection de l'yperplan et de l'hypersphère vérifient $0 < \alpha_i < 1/\nu m$ ($i \in \{5, 6\}$).

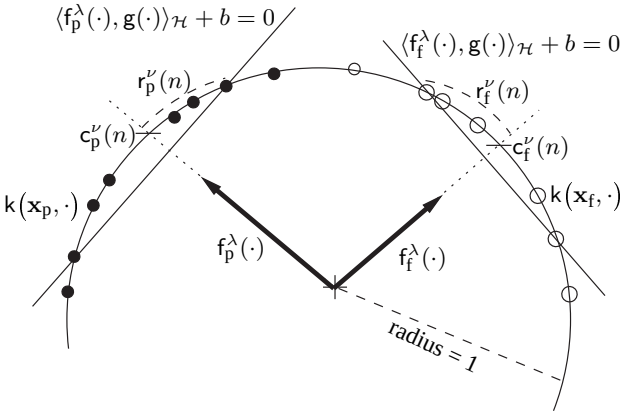

FIG. 2: Interprétation géométrique de l'algorithme basé SVM-1. La situation représentée ici correspond à une détection de ruptures, car les hyperplans représentés par $\mathsf{f}^{\lambda}_{\mathsf{p}}(\cdot)$ (correspondant à l'ensemble passé immédiat – cercles pleins) et $\mathsf{f}^{\lambda}_{\mathsf{f}}(\cdot)$ (correspondants à l'ensemble futur immédiat – cercles vides) sont distinctement séparés, et la distance $\mathsf{d}(\mathsf{c}^{\nu}_{\mathsf{p}}(n), \mathsf{c}^{\nu}_{\mathsf{f}}(n))$ est grande par rapport aux arcs $\mathsf{r}^{\nu}_{\mathsf{p}}(n)$ et $\mathsf{r}^{\nu}_{\mathsf{f}}(n)$.

représentant $\mathsf{S}^{\nu}_{\mathsf{p}}(n)$ et $\mathsf{S}^{\nu}_{\mathsf{f}}(n)$ peuvent être complexes et discontinus, rendant ainsi la définition d'une mesure de similarité dans cet espace très difficile. Heureusement, l'interprétation géométrique des SVM-1 dans l'espace augmenté $\mathcal{H}$ permet d'en déduire une mesure très intuitive et simple à mettre en oeuvre : les quantités $\mathsf{S}^{\nu}_{\mathsf{p}}(n)$ et $\mathsf{S}^{\nu}_{\mathsf{f}}(n)$ correspondent géométriquement aux hypercercles résultants de l'intersection de l'hypersphère avec l'hyperplan, voir figure 1. Ainsi, la comparaison des quantités $\mathsf{S}^{\nu}_{\mathsf{p}}(n)$ et $\mathsf{S}^{\nu}_{\mathsf{f}}(n)$ dans l'espace initial des données $\mathcal{X}$ se ramène à une comparaison dans $\mathcal{H}$ en comparant les hypercercles correspondants dont les centres sont notés $\mathsf{c}^{\nu}_{\mathsf{p}}(n)$ et $\mathsf{c}^{\nu}_{\mathsf{f}}(n)$ et les arcs de cercle $\mathsf{r}^{\nu}_{\mathsf{p}}(n)$ et $\mathsf{r}^{\nu}_{\mathsf{f}}(n)$, voir figure 2.

La mesure de similarité basée sur notre algorithme est définie comme suit [3] :

$$\mathsf{D}^{\nu}(n) \;=\; \frac{\mathsf{d}(\mathsf{c}^{\nu}_{\mathsf{p}}(n), \mathsf{c}^{\nu}_{\mathsf{f}}(n))}{\mathsf{r}^{\nu}_{\mathsf{p}}(n) + \mathsf{r}^{\nu}_{\mathsf{f}}(n)} \qquad (5)$$

Pratiquement, $\mathsf{D}^{\nu}(n)$ est calculée à partir des coefficients α_i de chaque support de densité $\mathsf{S}^{\nu}_{\mathsf{p}}(n)$ et $\mathsf{S}^{\nu}_{\mathsf{f}}(n)$,— voir l'article de F. Desobry et M. Davy dans [3] pour les détails de calcul et de développement de cette mesure.

3. DÉTECTION DE CHANGEMENT DE LOCUTEURS

3.1. Algorithme de détection de ruptures

1. **Parametrisation acoustique** : Celle-ci est effectuée au moyen d'outils conventionnels tel que HTK Tools [8]. Cette étape est effectuée pour l'ensemble du signal.
2. **Entraînement des SVM-1** : Pour chaque instant n et consécutivement à la formation des ensembles $\mathsf{X}_\mathrm{p}(n)$ et $\mathsf{X}_\mathrm{f}(n)$ deux modèles SVM-1 sont entraînés en résolvant le problème (4) pour les deux ensembles. Pour réduire les charges de calculs et accélerer la procédure, la technique développée dans [1] peut être utilisée.
3. **évaluation de la mesure de similarité** : Comme dans toute technique de segmentation (détection de rupture), une rupture est détectée lorsque l'indice $\mathsf{D}^\nu(n)$ définie dans Eq. (5) dépasse un seuil prédéterminé, qui peut être fixe pour tout les instants. Une autre approche consiste à calculer toute la courbe des distances puis choisir un seuil donné permettant d'estimer les instants de ruptures.

3.2. Le regroupement hiérarchique

Suite à la détection de ruptures, nous sommes en présence d'une collection d'objets (segments de paroles homogènes/locuteurs) et nous devons regrouper ces objets par classe (les locuteurs). Nous avons choisi de mettre en oeuvre le regroupement hiérarchique agglomératif qui consiste à considérer au départ chaque segment comme étant une classe et à chaque itération on réunit deux classes les plus proches au sens d'un critère, appelé critère de regroupement [2]. Dans ce cas ce critère est la mesure de similarité définie dans la section 2.2. Ce processus est réitéré jusqu'à l'obtention d'une classe unique. Nous obtenons à l'issue du regroupement un arbre de classification appelé dendrogramme. C'est la manière de parcourir l'arbre qui définit la partition finale.

4. EXPÉRIENCES ET RÉSULTATS

4.1. La Base de Données

Les signaux utilisés sont issus d'enregistrements d'émissions d'informations radio-diffusés (Broadcast News en abrégé bnews) de diverses stations américaines fournies par NIST [6]. Ces fichiers se divisent en deux catégories : 6 fichiers de développement (dry run files) de 10 mn environ chacun dédiés au réglage des paramètres et 3 fichiers d'évaluation (Eval files) de 30 mn chacun.

4.2. Expériences sur les signaux de développement

Afin de régler les paramètres de notre algorithme pour les tâches de détection de ruptures et de regroupement nous utilisons les fichiers de développement pris séparément et le score global est la moyenne globale sur l'ensemble des fichiers. Le critère de performance établi et fourni par l'institut NIST est le "Speaker Diarization Error" ou "Diarrization Error Rate" (DER) fourni par un script en langage perl (voir [4] pour d'amples détails).Pour les expériences concernant cette catégorie de fichiers, la parametrisation utilisée est de 16 coefficients MFCC.

TAB. 1:
Evolution du "DER" en fonction
de la taille des fenêtres glissantes m pour ν=0.1s, σ=0.51 et $\Delta_n = 0.2$s.

m (s)	0.5	0.7	0.9	1.5	2.5	3.5
DER (%)	14.70	13.83	19.21	10.73	16.65	25.36

Sélection de paramètres Les paramètres pertinents de notre algorithme objet de cette étude de sélection sont : $m = m_\mathrm{p}$=m_f (taille de l'ensemble $\mathsf{X}_\mathrm{p}(n)$ et de $\mathsf{X}_\mathrm{f}(n)$ que nous supposons égaux) et le pas de progression Δ_n de ces ensembles appelés communément fenêtres glissantes précédent et succédant à l'instant d'analyse n. Ces deux paramètres m et Δ_n sont communs avec la méthode de référence RVG-BIC. Aux fins de comparaison des deux méthodes nous avons fixé les mêmes valeurs pour ces deux paramètres. Le jeu de paramètres initialement testé dans la table 1 n'est pas fortuit, mais s'inspire des réglages des paramètres de la méthode de référence RVG-BIC, en conformité avec les expériences menées dans [2].

Notre algorithme utilise un paramètre additionnel relatif au noyau, assurant le contrôle de la corrélation des points voisins dans l'espace augmenté $\mathcal{H}$. Dans le cas du noyau gaussien ce paramètre est l'ecart-type σ.

La figure 3 montre l'évolution du DER en fonction de la variation de σ. Le minimum d'erreur est atteint pour σ =0.51, et vaut 10.73%. Les autres paramètres utilisés pour cette expérience sont fixés comme suit : m = m_p=m_f=win=1.5s et Δ_n=0.2s, qui sont des paramètres adéquats pour les deux méthodes en comparaison. Le paramètre ν =0.1 traduit une tolérance maximale de 10% des vecteurs support.

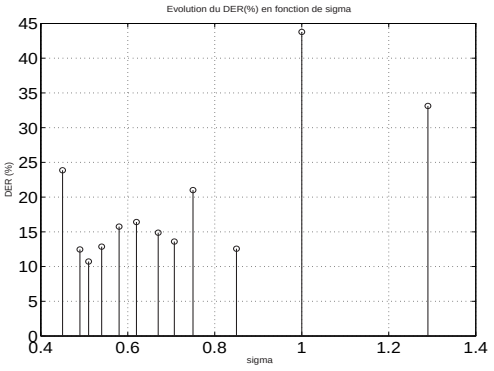

FIG. 3: Evolution du DER en fonction du paramètre du noyau σ.

Les tables 1 et 2 résument les variations de la taille des fenêtres adjacentes glissantes m = m_p=m_f et leurs pas de progression Δ_n. On confirme, que les valeurs de $m = 1.5$s et $\Delta_n = 0.2$s sont des réglages adéquats au regard de la l'erreur obtenue (10.73%).

Evaluations de stratégies de parametrisation acoustique Nous présentons dans cette sous-section l'impact des diverses stratégies de parametrisation acoustiques décrites dans la table 3. Pour chaque configuration, nous

TAB. 2:
Evolution du "DER" en fonction du pas de progression Δ_n, pour $m = 1.5$s, $\nu = 0.1$ et $\sigma = 0.51$.

Δ_n (s)	0.1	0.2	0.3	0.4	0.5	0.6
DER (%)	24.42	10.73	11.93	15.27	12.85	22.68

TAB. 3: Parametrisations acoustiques testées.

Configuration	composition du vecteur acoustique
C_0	16 MFCCs
C_1	16 MFCCs et 10 LPCCs
C_2	C_1 et 10 coefficients de reflection
C_3	C_2 et 10 coefficients banc de filtre
C_4	16 MFCCs et 16 ΔMFCCs
C_5	C_4 et 16 $\Delta\Delta$MFCCs

avons optimisé la sélection des paramètres, tels que mentionnée dans la section ci-dessous. Nous reportons dans les tables 4 et 5 les erreurs minimales obtenues avec le jeu de paramètres sélectionné.

L'estimation du nombre de locuteurs présents dans la conversation analysée est obtenue en parcourant le dendrogramme selon une coupe horizontale, ce qui revient à faire une hypothèse sur le nombre de locuteurs désiré puis de vérifier celle-ci selon la performance obtenue. Les travaux de synthèse reportés dans [4] offrent une explication claire sur la détermination automatique du nombre de locuteurs.

4.3. Validation sur les fichiers d'évaluation

La table 6 montre que notre méthode obtient des taux d'erreurs DER bien inférieurs à la méthode de référence RVG-BIC. Le meilleur résultat obtenu est la moyenne sur ces trois fichiers, soit un taux d'erreur de 13.63%. Ces résultats sont d'autant plus prometteurs car comparés à ceux publiés récemment dans la littérature constituant l'état de l'art [4] (page 22,Table 7) et dans laquelle les méthodes présentées ont été optimisées indépendamment de notre algorithme.

5. CONCLUSIONS ET PERSPECTIVES

Les résultats présentés dans cette communication montrent clairement que notre approche basée sur un algorithme à base des machines à vecteurs support monoclasse ouvre une voie de recherche très prometteuse pour l'indexation en locuteurs de discours multi-locuteurs. Nous imputons cette performance, à un meilleur processus de segmentation acoustique plutôt qu'à un meilleur regroupement, du fait que c'est la première phase qui conditionne la seconde, néanmoins, une affirmation rigoureuse ne peut découler que d'une étude comparative détaillée de la pureté moyenne des segments obtenues à la suite du processus de regroupement. Ce travail est une perspective déja entammée. Un autre résultat intéressant concerne l'étude comparée des méthodes RVG-BIC et SVM-1 en fonction des diverses stratégies de parametrisation. Ainsi, il apparaît, qu'une paramétrisation même redondante améliore les résultats globalement pour les deux méthodes mais que c'est notre méthode qui assure une nette supériorité.

TAB. 4: Performances comparées (KCD/RVG) en fonction des parametrisations acoustiques décrites dans Table 3.

Config	DERmin (%)		estim. # loc.	
	KCD	RVG-BIC	KCD	RVG-BIC
C_0	10.73	26.38	17	9
C_1	8.37	21.18	17	9
C_2	7.95	15.08	17	11
C_3	10.90	15.26	9	9
C_4	11.44	20.91	13	11
C_5	8.63	14.30	19	11

TAB. 5: Jeu de paramètres en fonction des parametrisations acoustiques testes.

Configuration	Jeu de paramètres selectionné
C_0	$m = 1.5$s,$\sigma = 0.51$, $\Delta_n = 0.2$s.
C_1	$m = 2.0$s,$\sigma = 1$, $\Delta_n = 0.3$s
C_2	$m = 1.5$s,$\sigma = 1$, $\Delta_n = 0.3$s
C_3	$m = 2.5$s,$\sigma = 0.707$, $\Delta_n = 0.2$s
C_4	$m = 0.7$s,$\sigma = 0.707$, $\Delta_n = 0.2$s
C_5	$m = 0.9$s,$\sigma = 0.85$, $\Delta_n = 0.3$s

RÉFÉRENCES

[1] M. Davy, F. Desobry, A. Gretton, and C. Doncarli. An online support vector machine for abnormal events detection. *Signal Processing*, 2006. to appear.

[2] P. Delacourt and C. Wellekens. Distbic : a speaker-based segmentation for audio data indexing. *Speech Communication*, 32(1) :111–126, September 2000.

[3] F. Desobry, M. Davy, and C. Doncarli. An online kernel change detection algorithm. *IEEE Trans. Sig. Proc.*, 53(5), August 2005.

[4] S. Meignier, D. Moraru, C. Fredouille, J.-F. Bonastre, and L. Besacier. Step-by-step and integrated approaches in broadcast news speaker diarization. *Computer Speech and Language* , pages 1–28, 2005. in Press.

[5] D. Moraru, S. Meignier, C. Fredouille, L. Besacier, and J.-F. Bonastre. The ELISA consortium approaches in broadcast news speaker segmentation during the NIST 2003 rich transcription evaluation. In *IEEE ICASSP'04*, Montreal, Canada, 2004.

[6] NIST RT03S. The rich transcription spring 2003 (rt-03s) evaluation plan http ://www.nist.gov/speech/tests/rt/rt2003/spring/docs/rt-03-spring-eval-plan-v4.pdf,. 2003.

[7] B. Schölkopf and A. Smola. *Learning with Kernels*. MIT Press, Cambridge, USA, 2002.

[8] S. Young and all. *The HTK Book (for HTK Version 3.2.1)*. cambridge.

TAB. 6: Résultats obtenus sur les signaux d'évaluation. Les paramètres sont $m = 1.5$s, $\nu = 0.1$, $\sigma = 0.51$ et $\Delta_n = 0.2$s. La parametrisation choisie est C_2. Les fichiers sont (a) 20010228.2100-2200-MNB-NBW, (b) 20010217.1000-1030-VOA-ENG and (c) 20010220.2000-2100-PRI-TWD.

Fichier	RVG-BIC	KCD	Estimation du Nbre de Locuteurs
(a)	25.60	14.34	23
(b)	20.17	12.28	25
(c)	22.69	14.27	22