DEUX APPROCHES DE NORMALISATION DES ENTREES POUR LA RECONNAISSANCE DE MOTS ISOLES

Abderezak BENCHENIEF[*], Salim SBAA[*], Abdelmalik TALEB-AHMED[**]

[*] Déprtement Génie Electrique, Université de Biskra, Biskra, Algérie
[**] LAMIH UMR CNRS-UVHC, Valenciennes, France,
benchenief@gmail.com

RESUME

Dans cet article, nous allons présenter deux systèmes de reconnaissance de chiffres parlés anglais, en mode indépendant du locuteur, basé sur les deux stratégies principales de la classifi- cation binaire SVM multi-classes. Cependant, les techniques SVM exigent des vecteurs d'entrée de taille fixe. Pour lever cette difficulté, nous avons utilisé deux approches différentes norma- lisation des entées basées sur le fenêtrage fixe et variable des vecteurs acoustiques des énoncés d'entrées. Le but est de réduire le temps de calcul pendant la phase d'apprentissage et de test des deux stratégies et déterminer ainsi celle qui donne le meilleur taux de reconnaissance. Les résultats trouvées montrent que la chaîne de reconnaissance utilisant la stratégie un contre un comme moteur de reconnaissance et l'approche fenêtre de taille fixe pour normaliser les entrées est beaucoup plus satisfaisante par rapport aux autres chaînes présentées dans notre article. Ce système de reconnaissance atteint un taux de 98,95%, tout en utilisant seulement 13 vecteurs caractéristiques par énoncé en entrées du classifieur, ce qui réduit considérablement le temps d'apprentissage et de test.

MOTS CLES: Normalisation; Fenêtrage fixe et variable; MFCC ; SVM ; Classification

1 INTRODUCTION

La reconnaissance automatique de la parole (RAP) est une branche de la reconnaissance des formes. Grâce à cette technologie, on peut communiquer oralement avec la machine au lieu d'utiliser les gestes ou les commandes des automatismes, ce qui facilite considérablement l'interaction homme/ machine.

Parmi les méthodes à noyaux, inspirées de la théorie statistique de l'apprentissage de Vladimir Vapnik, les SVM (Support Vector Machines) constituent la forme la plus connue. SVM est une méthode de classification binaire par apprentissage supervisé, elle fut introduite par Vapnik en 1995. Cette méthode est donc une alternative récente pour la classification. Cette méthode repose sur l'existence d'un classificateur linéaire dans un espace approprié. Puisque c'est un problème de classification à deux classes, cette méthode fait appel à un jeu de données d'apprentissage pour apprendre les paramètres du modèle. Elle est basée sur l'utilisation de fonctions dites noyau (kernel) qui permettent une séparation optimale des données. Dans la présentation des principes de fonctionnements, nous schématiserons les données par des « points » dans un plan.

Les SVM sont des classifieurs binaires et statistique. C'est à dire qu'ils permettent de créer une surface de décision entre deux classes définies dans un même espace. Pour cela, ils construisent une frontière de décision par projection des caractéristiques provenant d'un espace d'origine dans un espace de caractéristiques de dimension supérieure (voir infini) dans le but de rendre les classes linéairement séparables. Ces techniques présentent des résultats meilleurs par comparaison avec d'autres méthodes comme le maximum-likelyhood ou les réseaux neuronaux ([1] [2], [3], [4] et [5] sont quelques exemples). Les SVM ont un comportement plus performant pour un ensemble d'entraînement très réduit. Mais l'intérêt vient aussi de la possibilité d'agir sur la façon comme sont classifiées les données. Les SVM ont été appliqués avec succès sur plusieurs problèmes pratiques. Parmi ceux-ci, la reconnaissance des unités de courte durée, comme phonème isolé et la classification de lettres [6], [7] et [8], la compression d'image ou la reconnaissance d'objets 3D [9], la reconnaissance de chiffres parlés [10] et la reconnaissance de voyelles [11].

Les SVM sont des classifieurs qui réalisent un apprentissage statistique. Il existe de nombreux problèmes à leur utilisation pour la reconnaissance de la parole. L'analyse de la parole génère des séquences de longueurs variables qui modélisent les variations de séquences des mots, l'accent, la vitesse d'élocution. Cependant, les classifieurs SVM nécessitent des vecteurs de longueurs fixes.

La mise en œuvre sur des données réelles passe par l'usage de librairies type LIBSVM et par la possibilité de faire de la classification multiple. Un algorithme SVM ne pouvant séparer que deux classes, il est alors nécessaire de choisir une stratégie multi-classes adaptée.

Dans cet article, nous allons présenter deux systèmes de reconnaissance de chiffres parlés spécifiques à la langue anglaise en utilisant les deux stratégies principales de la classification binaire SVM multi classe comme techniques de reconnaissance de forme (les deux approches classiques, ''un contre un" et "un contre tous" ont été mises en oeuvre pour éviter des ambiguïtés). Nous avons donc utilisé deux approches de normalisation des entrées des deux classifieurs pour que les entrées des classifieurs soient de taille fixe. Le but est de déterminer l'approche qui donne le meilleur taux de reconnaissance. Autrement dit, quel est le système le plus performant et le plus puissant dans la reconnaissance automatique de chiffres parlés.

Cet article est organisé comme suit: tout d'abord nous abordons la problématique des de Séparateurs à Vaste Marge (SVM) (section II). La description du système de reconnaissance de référence est présentée en section III. Dans la section IV, nous évaluons notre système par un corpus de test des locuteurs qui n'ont pas participé à l'apprentissage. Finalement, une conclusion et des perspectives seront introduites dans la section V.

2 MACHINES A VECTEURS DE SUPPOR

La théorie de l'apprentissage statistique de Vapnik et de Chervonenkis [12] a conduit au développement d'une classe d'algorithmes connue sous le nom de Support Vector Machines (SVM) qui se traduit par machine à vecteurs de support. Une des originalités de la méthode est de produire une fonction décision binaire qui n'utilise qu'un sous ensemble de la base d'apprentissage. Les éléments de ce sous ensemble sont nommés Vecteurs de Support (VS). Soit une base d'apprentissage $S_a = \{(x_1, y_1),\ldots, (x_m, y_m)\}$ composée de m couples (vecteur d'attributs, classe) avec $x_i \in \Re^n$ et $y_i \in \{-1, +1\}$. L'algorithme des SVM projette les vecteurs x_i dans un espace de travail H à partir d'une fonction non linéaire $\Phi : \Re^n \rightarrow H$. L'hyperplan optimal de séparation des deux classes dans l'espace H est ensuite recherché. Cet hyperplan (w, b) matérialise la frontière de séparation entre les deux classes. La classe y d'un nouvel exemple x est définie par l'expression (1):

$$y = sign(\langle w, \Phi(x) \rangle + b) \quad (1)$$

L'hyperplan est optimal s'il maximise la distance qui le sépare des exemples dont il est le plus proche. Cette distance est usuellement appelée marge du classificateur. Il a été démontré [12] que maximiser cette marge correspond à maximiser le «pouvoir» généralisateur du classificateur. Vapnik [12] a proposé une version duale du problème qui revient à minimiser l'expression (2):

$$w(\alpha) = 1/2.\sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^{m} \alpha_i \quad (2)$$

sous les contraintes $\forall i \in [1,..., m] : \sum_{i=1}^{m} y_i \alpha_i = 0,\ 0 \leq \alpha_i \leq C$. L'avantage de cette formulation est qu'elle utilise une fonction noyau K respectant l'égalité $K(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle$. Ce qui évite de travailler directement dans l'espace de projection H. La solution optimale α^* définit l'ensemble V_S des vecteurs de support $\forall i \in [1,...,m], i \in V_S : \alpha_i > 0$ et permet de définir la fonction de décision (expression (3)) [12]:

$$f(x) = \sum_{i \in Vs} \alpha_i^* y_i \langle x_i, x \rangle + b^* \quad (3)$$

Le compromis entre erreur de classification et complexité du modèle dépend du paramètre C.

Face à d'autres méthodes de classification, les SVM ont un comportement plus performant pour un ensemble d'entraînement très réduit. L'intérêt vient aussi de la possibilité d'agir sur la façon comme sont classifiées les données [13]. L'inconvénient des SVM est le choix empirique de la fonction noyau adaptée au problème. Un deuxième inconvénient est le temps de calcul qui croît de façon cubique en fonction du nombre de données à traiter. Si le nombre de données d'apprentissage est n la dimension des données à classer est d, la complexité est alors en $O(dn^3)$ [14].

2.1 Le classifieur multi classes

Les SVM étant des classificateurs binaires, c'est-à-dire, un problème à deux classes. La résolution d'un problème multi-classes est réalisée en le transformant en une combinaison de problèmes binaires [15], ceci restant cependant un domaine de recherche très ouvert. Jusqu'à maintenant la meilleure méthode de construire les SVM multi classes n'est pas claire [16]. Schölkopf et al. [17] ont proposé un classificateur de type "*un contre tous*" : l'idée consiste simplement à transformer le problème à k classes en k classifieurs binaires. Le classement est donné par le classifieur qui répond le mieux. Clarkson et Moreno [16] ont proposé un classifieur de type "*un contre un*". Cette fois le problème est transformé en $k \cdot (k-1)/2$ classifieurs binaires, chaque classe i étant en effet comparée à chaque classe j. Le classement est donné par le vote majoritaire.

3 DESCRIPTION DU SYSTÈME DE LA RECONNAISSANCE DE LA PAROLE

Dans notre expérience, nous avons combiné les approches de normalisation des entrées (*Fenêtre de Taille variable*, *Fenêtre de Taille fixe*) aux les deux stratégies principales de la classification binaire multi classes (SVM Multi classes), "*un contre un*" (one vs. one) et "*un contre tous*" (one vs. all). Notre but est de faire une comparaison du point de vue performance de la classification de chiffres parlés anglais, en mode indépendant du locuteur.

Le système fonctionnel de reconnaissance de la parole, basé sur SVM Multi class, peut se schématiser comme indiqué

dans la figure

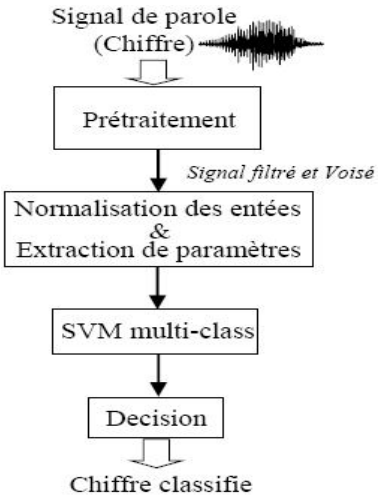

Figure 1: Architecture fonctionnelle du système de reconnaissance de la parole basée sur la classification SVM Multi class

Afin de réaliser notre expérience, nous avons tenu compte des étapes suivantes:

3.1 Pré-traitement des données

Durant cette phase, le signal vocal (segments phonémiques) est pré-accentué pour lever les hautes fréquences qui sont moins énergétiques que les basses fréquences, ce qui permet de compenser le niveau plus faible des sons. On utilise généralement un filtre passe-haut, dit de préaccentuation qui est de la forme de l'expression (4):

$$H(z) = 1\text{-}a.z^{-1} \; ; \; a = 0.95, \qquad 4$$

3.2 Extraction de caractéristiques et normalisation des entrées

Le même message prononcé deux fois par un même locuteur dans des conditions identiques produit deux formes spectrales différentes. Cette variabilité est dite *intra-locuteur*. La qualité de la voix, le débit de parole, le degré d'articulation sont tous des facteurs à la base de variations acoustiques pour un signal donné. Ces variations entraînent des transformations non linéaires dans le temps du signal parole. La non linéarité vient du fait que les transformations affectent plus les parties stables du signal que les phases de transitions [18].

Après la préaccentuation, le signal vocal qui est fortement non stationnaire est décomposé en une succession de tranches élémentaires supposées stationnaires. Ces tranches sont appelées fenêtres d'analyse ou trames. Typiquement, une analyse est appliquée toutes les f_p =10 [ms] (période de trame), une fenêtre, w_s de taille fixe de 20 - 30 [ms] est extraite du signal. A cette fenêtre, on applique ensuite une fenêtre de Hamming, une analyse acoustique à chaque fenêtre à l'aide de la technique MFCC (*mel-frequency cepstral coefficients*) pour extraire les 14 premiers coefficients MFCCs pertinents du signal vocal. Ces coefficients sont calculés à partir des énergies extraites de 20 filtres triangulaires répartis sur l'échelle de Mel. La formule de conversion du hertz en *mel* la plus utilisée est la suivante :

$$mel = 2595 * \log_{10}(1 + \frac{f_{HZ}}{700}) \qquad 5$$

La première qualité est sa résistance reconnue au bruit. La deuxième qualité majeure de la méthode MFCC est sa plausibilité biologique puisqu'elle utilise une échelle psycho acoustique des fréquences similaires à celles de l'oreille interne [19], [20].

Généralement, les valeurs de w_s et f_p sont maintenues constantes pour toutes les prononciations et chacune d'elles présentes une durée différente. Par conséquent, l'analyse de la parole génère des vecteurs caractéristiques ou (*features :* en anglais) de taille variables. Cependant, les classifieurs SVM nécessitent des vecteurs de longueurs fixes. Pour lever cette difficulté, On a utilisé deux méthodes, pour but de passer d'une séquence de longueur variable à des vecteurs de dimension fixe, qui représentent des séquences de taille variables. Deux méthodes (fenêtrage fixe, fenêtrage variable) sont effectuées avant la phase d'extraction des paramètres (*mfcc*), afin de garder une taille fixe pour les vecteurs spectraux. Les valeurs sélectionnées pour ces paramètres (w_s *et* f_p) dans notre expérience seront discutées par la suite.

3.2.1 Fenêtre de Taille variable

Dans ce cas, comme dans les procédures de paramétrage traditionnel, la taille de la fenêtre est choisie pour être proportionnelle à la période de trame (i.e. $w_s = K * f_p$), avec K (le facteur de recouvrement) étant constant pour tous les énoncés d'entrées. Notez que K détermine le degré de chevauchement (recouvrement) entre les fenêtres d'analyse adjacentes. Néanmoins, nous choisissons d'abord la valeur de w_s de manière dynamique en fonction de la longueur de chaque énoncé, calculée par la formule: $w_s = L / N$, tel que, L est la longueur de la séquence acoustique d'entrée et N est un entier donné, étant constant pour tous les énoncés d'entrées afin d'obtenir des vecteurs caractéristiques de même taille de tous les échantillons de la base de donnée. Ensuite, f_p est trouvée comme suit, $f_p = w_s / K$. la figure 2 illustre cette procédure. De cette façon, nous sommes en mesure de munir les SVM par des vecteurs d'entrée de dimension fixe. Toutefois, lorsque la durée de w_s est trop longue, le segment de parole analysée ne satisfait pas les propriétés de stationnarité requise et un résultat est obtenu en moyenne. Par conséquent, il nous manque quelques détails pertinents du signal de parole.

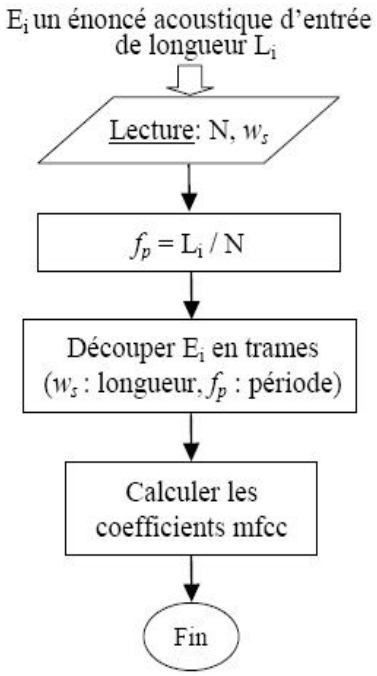

Figure 2:Déroulement de l'algorithme de la normalisation des entrées par le fenêtrage variable

3.2.2Fenêtre de Taille fixe

Dans ce cas, la valeur de la fenêtre w_s est fixé "a priori" et maintenue constante pour tous les énoncés. Toutefois, pour obtenir un vecteur d'entrée de dimension fixe, nous avons besoin de sélectionner de manière dynamique la période de trame f_p en fonction de la longueur L_i de chaque énoncé, f_p calculé comme suit : $f_p = L_i / N$, avec N est un entier donné, étant constant pour tous les énoncés d'entrées, représente le nombre des fenêtres à extraire de chacune d'elles. La figure 3 montre un exemple extrême dans laquelle les fenêtres d'analyse ne sont pas plus se recouvrées, la figure 4 explique l'algorithme. Par conséquent, certaines informations sont manquantes pour des énoncés de longue durée.

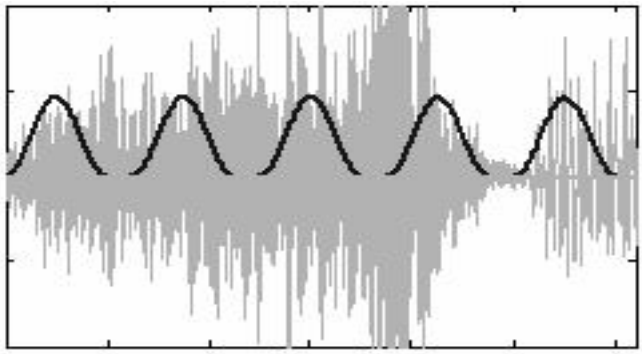

Figure 3: Fenêtrage fixe: Un exemple extrême dans laquelle les fenêtres d'ana lyse ne sont pas plus se recouvrées

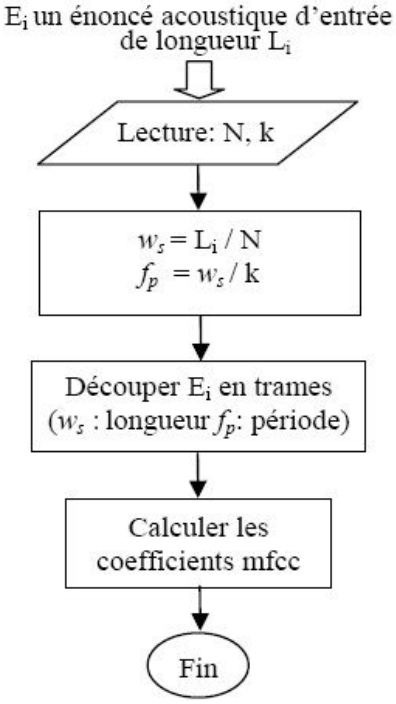

Figure 4:Déroulement de l'algorithme de la normalisation des entrées par le fenêtrage fixe.

3.2.3 La base de données utilisée

TI-DIGITS [21] est un corpus anglais. Ce dernier est composé d'enregistrements de 207 locuteurs, répartis en deux sous-ensembles : un ensemble d'apprentissage (94 locuteurs) et un de test (113 locuteurs). Chaque locuteur a prononcé deux fois chacun des onze chiffres (de 1 à 9, plus "oh" et "zéro"). Le sous-ensemble d'apprentissage (respectivement de test) contient donc 2068 occurrences de chiffres (respectivement, 2486 occurrences de chiffres). Tous les réglages ont été effectués sur l'ensemble d'apprentissage seul, celui-ci étant divisé en un ensemble d'apprentissages de 1386 points et un ensemble de validation de 682 points.

L'apprentissage et le test

L'apprentissage et le test des SVM ont été réalisés à l'aide du logiciel LIBSVM dont les algorithmes sont décrits dans [16]. Nous avons choisi d'utiliser le C-SVM avec un noyau gaussien RBF, puisqu'il a des meilleures performances face à d'autres noyaux et non sensibles à l'échelle des données [22]. comme indiqué dans l'expression (6) :

$$K(x, y) = \exp\left(-\|x - y\|^2 / 2\sigma^2\right) \qquad 6$$

TI DIGITS comprenant 11 classes, il a fallut choisir une méthode de classification multi-classes. Les deux principales, facilement mises en oeuvre à partir de classifieurs binaires, sont le "un contre un" et le "un contre tous". Dans le premier cas, nous obtenons 55 classifieurs binaires et la classification est donnée par un vote majoritaire sur l'ensemble des classifieurs. Dans l'autre cas,

nous obtenons 11 classifieurs qui apprennent chacun sur tous les points de la base d'apprentissage et la classification est donnée par le classifieur qui affiche la plus grande valeur de la fonction de décision.

Les hyper paramètres $Gamma = 1/(2.\sigma^2)$ et C ont été déterminés empiriquement en cherchant à minimiser le taux d'erreur sur la base de validation [23].

4 RÉSULTATS ET COMMENTAIRES

4.1 Performance globale des systèmes

La Table 1 montre la précision des systèmes pour la reconnaissance des chiffres globalement sur le corpus de test. La performance globale des systèmes est 98.75% pour la stratégie "un contre tous" et la méthode normalisation des entrées fenêtre de taille variable et 98.67% pour la stratégie "un contre un", en utilisant que 9 vecteurs caractéristiques par énoncé. Dans le cas de l'utilisation de la fenêtre de taille fixe, les classifieurs fournis un taux de 98.75% par le classifieur "un contre tous", la classification par "un contre un" donne le meilleur taux de reconnaissance de l'ordre de 98.95%, en utilisant 13 vecteurs caractéristiques par énoncé en entrées des classifieurs.

Table 1:Taux de reconnaissance des chiffres par les deux stratégies du classification, tout en utilisant a la fois fenêtre de taille fixe et variable

Méthode de normalisation des entrées	Classifieur de chiffres parlés	Taux (%)
Fenêtre de Taille variable (9 trames optimales)	SVM un contre tous	98.75
	SVM un contre un	98.67
Fenêtre de Taille fixe (13 trames optimales)	SVM un contre tous	98.75
	SVM un contre un	98.95

Les taux de reconnaissance fixés à la Table I montrent que la classification utilisant les deux stratégies principales de la classification binaire SVM multi classe est homogène dans le cas de l'utilisation de la fenêtre de taille variable. En revanche, l'utilisation de la fenêtre de taille fixe et ces classificateurs, dans ce cas, la classification utilisant l'algorithme SVM "un contre un" est plus performante que la classification utilisant l'algorithme SVM "un contre tous".

Néanmoins, il devrait être pris en compte que les résultats rapportés ne sont que des explorations préliminaires et il y a encore plusieurs questions à être améliorés et les alternatives à être étudiés.

La Table 2 montre l'effet du nombre N sur le nombre des trames de sortie de chaque approche de normalisation des entrées par (Fenêtrage variable, Fenêtrage fixe).

Table 2:L'effet du nombre entier N sur le nombre des trames de sortie de chaque approche de normalisation des entrées

	Trames obtenues								
Le nombre 'N'	2	4	5	6	7	9	10	12	14
Fenêtrage variab	3	7	9	11	13	17	19	23	27
Fenêtrage fixe	2	4	5	6	7	9	10	12	14

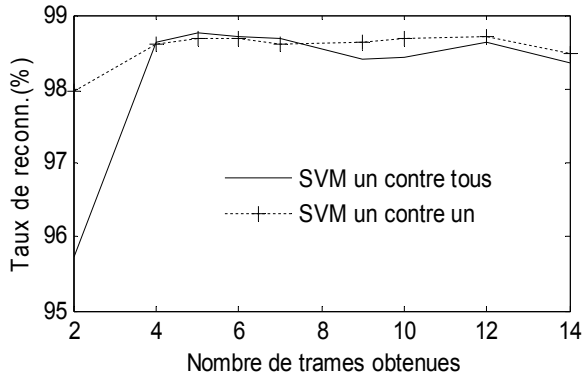

Figure 5:La précision des systèmes en fonction du nombre fixe de trames caractéristiques fournies par la méthode normalisation des entrées (fenêtre de taille variable) considérées comme l'entrée pour le SVM.

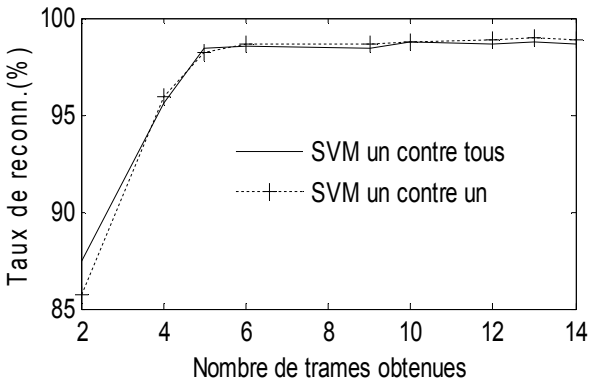

Figure 6:La précision des systèmes en fonction du nombre fixe de trames caractéristiques fournies par la méthode normalisation des entrées (fenêtre de taille fixe) considérées comme l'entrée pour le SVM.

Les Tables 3 et 4 montrent la précision de reconnaissance de chaque classe indépendante, la première remarque que l'on peut faire, c'est que les classes ont des taux de classifications homogènes et assez bon particulièrement pour les classes «0» et «Oh» où atteignent les systèmes de reconnaissance à 100%.

Table 3: la précision des classes fournie par le classifieur SVM un contre tous ainssi la decision GLOBALE DU SYSTEME DE RECONNAISSANCE.

	SVM un contre tous	
Classe	Fenêtre deTaille variable (9 trames optimales)	Fenêtre de Taille fixe (13 trames optimales)
«1»	98.67	98.23
«2»	98.67	98.23
«3»	99.12	99.11
«4»	97.35	97.34
«5»	99.12	98.67
«6»	98.23	99.11
«7»	98.67	97.78
«8»	98.23	98.67
«9»	98.23	97.78
«Oh»	100 .00	100.00
«zero»	100.00	100.00
Total	98.75	98.75

Table 4: la précision des classes fourie par le classifieur SVM un contre un ainssi la decission blobale du systemme de reconnaissance.

	SVM un contre un	
Classe	Fenêtre de Taille variable (9 trames optimales)	Fenêtre de Taille fixe (13 trames optimales)
«1»	98.23	98.23
«2»	98.23	98.67
«3»	99.11	99.56
«4»	97.34	97.35
«5»	98.67	99.12
«6»	99.11	98.67
«7»	97.78	99.12
«8»	98.67	98.23
«9»	97.78	99.56
«Oh»	100.00	100.00
«zero»	100.00	100.00
Total	98.67	98.95

5 CONCLUSION ET PERSPECTIVES

Malgré les efforts et les travaux intensifs réalisés dans le domaine de la Reconnaissance Automatique de la Parole, aucun système RAP n'est jugé fiable à 100%. Mais au fur et à mesure les auteurs essayent d'améliorer les scores pour de meilleurs résultats.

Dans cet article nous avons présenté deux systèmes de reconnaissance de chiffres parlés anglais en mode indépendant du locuteur, basé sur les deux stratégies principales de la classification binaire multi classe (SVM Multi classe) et deux approche de normalisation des entées (en utilisant à la fois la taille de fenêtre fixe et variable) pour éviter le problème des séquences de longueurs variables qui modélisent les variations de séquences des mots, l'accent, la vitesse d'élocution.

Lors de la conception de notre système, nous avons utilisé une analyse acoustique très robuste, représentative et discriminante (l'analyse MFCC) afin d'avoir une bonne modélisation du signal vocal. Cette dernière donne un meilleur taux de reconnaissance et reste plus performante dans la reconnaissance de la parole.

Tous les vecteurs de dimension fixe considérés à l'entrée des classifieurs sont manquants d'informations pertinentes du signal de parole, soit parce que utilisent des fenêtres d'analyse importante (fenêtre de taille variable) qui conduisent à des résultats moyennés ou parce que certaines parties du signal sont ignorés (fenêtre de taille fixe).

Les résultats obtenus par rapport aux classifieurs à base des SVMs utilisés sont très satisfaisants du fait que la technique SVM, a la capacité à généraliser la classification. Ainsi, cette méthode est adaptée aux applications présentant une grande variation intra-classes, comme par exemple la parole. Les taux de reconnaissance obtenus montrent clairement que la reconnaissance par un classifieur utilisant la stratégie "un contre un" est beaucoup plus satisfaisante par rapport que le classifieur utilisant la stratégie "un contre tous".

Dans [15] les auteurs montrent que la méthode "un contre un" a une meilleure précision que la méthode "un contre tous" dans 60% des cas, mais dans toutes les comparaisons les taux de précision restent proches de 2%. Même si la différence de précision est faible, un argument plus important en faveur de la stratégie "un contre un" est le temps nécessaire à l'apprent-issage. Cette dernière est de 2 à 6 fois plus rapide que la méthode "un contre tous".

Comme perspectives, nous proposons d'étudier des algorithmes de normalisation des entrées, basés sur les valeurs des coefficients (Kirtosis, Skewness… ext.) des trames acoustiques dont le but est de réduire le temps nécessaire à l'apprentissage des classifieurs et fiabiliser la classification des chiffres parlés et autres.

REFERENCES

[1] F. Melgani, L. Bruzzone,."Classification of hyperspectral remote sensing images with support vector machines". In IEEE Transactions on Geoscience and Remote Sensing, vol. 42, nº 8, pp. 1778-1790 ,2004.

[2] H. Sakoe, R. Isotani, K. Yoshida, K. Iso, T. Watanabe, , "Speaker-Independent Word Recognition using Dynamic Programming Neural Networks"; Proc. ICASSP-89, pp. 29-32; Glasgow, Scotland; 1989.

[3] K. Iso, T. Watanabe, , "Speaker-Independent Word Recognition using a Neural Prediction Model", Proc. ICASSP-90, pp. 441-444; Alburquerque, New México, USA, 1990.

[4] J. Tebelskis, A. Waibel, B. Petek, O. Schmidbauer, "Continuous Speech Recognition using Predictive Neural Networks", Proc. ICASSP-91, pp. 61-64; Toronto, Canada; 1991.

[5] Y. Bengio, "Neural networks for speech and sequence recognition", London International Thomson Computer Press, 1995.

[6] P. Clarkson, P..J. Moreno, "On the use of support vector machines for phonetic classification", Proc. ICASSP, Vol. 2 , pp.585 –588, 1999.

[7] A. Ganapathiraju, "Support vector machines for speech recognition" PhD Thesis, Mississipi State University, 2002.

[8] N. D. Smith, M.J.F. Gales, "Using SVMs and discriminative models for speech recognition", IEEE International Conference on Acoustics, Speech, and Signal Processing, 2002.

[9] Y. Bazi, F. Melgani, "Toward an optimal svm classification system for hyperspectral remote sensing images". IEEE Transactions on geoscience and remote sensing, 44:3374-3385 , 2006.

[10] I. Bazzi, D. Katabi, "Using support vector machines for spoken digit recognition,", In /ICSLP-2000, vol.1, 433-436.

[11] W. Xuechuan, K, P. Kuldip, "Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition". Pattern Recognition (PR) 36(10):2429-2439, 2003.

[12] V. Vapnik, " Statistical Learning Theory , ", Book, Wiley, New York ,1998.

[13] L. Mercier, ''Les machines à vecteurs support pour la classification en imagerie hyperspectrale' : implémentation et mise en œuvre'', UE ENG111 - Epreuve TEST.

[14] O. Bousquet, "Introduction aux Support Vector machine". (SVM), Orsay , 2001.

[15] C.-W. Hsu, C.-J. Lin, "A Comparison of Methods for Multi-class Support Vector,", Article In: IEEE Transactions on Neural Networks, Vol. 13, Nr. 2 (2002) , p. 415--425.

[16] P. Clarkson, J. P. Moreno, "On the use of support vector machines for phonetic classification,". Article. In: Proceedings of International Conference on Acoustics, Speech, Signal Processing, pp. 585–588, 1999.

[17] B. Scholkopf, C. Gurges, V. Vapnik, Extracting support data for a given task, Proceedings of First International Conference on Knowledge Discovery and Data Mining, Menlo Park, 1995, pp. 252–257

[18] F. Melgani, L. Bruzzone,."Classification of hyperspectral remote sensing images with support vector machines". In IEEE Transactions on Geoscience and Remote Sensing, vol. 42, nº 8, pp. 1778-1790 ,2004.

[19] B. A. Mellor, A. P. Varga."Noise masking in a transform domain". Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Proces -sing, vol. 2, pp 87–90, 1993.

[20] R.G. Leonard, ''A database for speaker-independent digit recognition'', In proceedings of ICASSP, volume 3, San Diego, 1984.

[21] B. D. Steven, P. Mermelstein,"Comparison of parametric representtations for monosyllabic word recognition in continuously spoken sentences,". Journal. IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 28, no 4, pp 357-366, 1980.

[22] C. Burges ,''A tutorial on support vector machines for pattern recognition'', Data Mining and Knowledge Discovery, 2-2, 1998.

[23] O. Chapelle, V. N. Vapnik, O. Bousquet, and S. Mukherjee Choosing ,''multiple parameters for support vectors machines''. Machine Learning, vol. 46, nº1, 2002.