Proceedings of the World Congress on Engineering 2013 Vol III,
WCE 2013, July 3 - 5, 2013, London, U.K.

Discriminative Prosodic Features to Assess the Dysarthria Severity Levels

K.L. Kadi, S.-A. Selouani, B. Boudraa, and M. Boudraa

***Abstract*—In this paper, Linear Discrimination Analysis (LDA) is combined with two approaches of automatic classification, Gaussian Mixture Model (GMM) and Support Vector Machine (SVM) to perform an automatic assessment of dysarthric speech. The front-end processing uses a set of prosodic features selected with LDA on the basis of their discriminative ability. The Nemours database of American dysarthric speakers is used throughout experiments. Results show a best classification rate with LDA/SVM system of 93% that was achieved over four severity levels of dysarthria: no dysarthric L0, mild L1, severe L2 and severe L3. This tool can help clinicians to assess dysarthria, can be used in remote diagnosis and may reduce some of the costs associated with subjective tests.**

***Index Terms*—Dysarthria, gmm, lda, nemours-database, prosodic-features, severity-level-assessment, svm**

I. Introduction

DYSARTHRIA is a disease that affects millions of people across the world; it is due to disturbances of brain and nerve stimuli of muscles involved in the production of speech. This disorder induces perturbation in timing and accuracy of movements that are needful for a normal prosody and intelligible speech [1].

Depending on the severity of the dysarthria, the intelligibility of speech can range from near normal to unintelligible [2]. Usually, a large battery of tests is necessary to assess the intelligibility that measures the disease severity or a treatment's progress. Actually, automatics methods of assessments can aid clinicians in the diagnosis and monitoring of dysarthria.

Diverse methods have been performed for automatic assessment of dysarthric speech. In [3], a combination of statistical method Gaussian Mixture Model (GMM) and soft computing technique Artificial Neural Network (ANN) along MFCC and speech rhythm metrics based front-end, achieved 86.35% performance over four severity levels of dysarthria. Feed forward ANN and SVM have been used successfully to design discriminative models for dysarthric speech with phonological features in [4]. In [5], a Mahalanobis distance based discriminant analysis classifier was proposed to classify the dysarthria severity by using a set of acoustic features. In this latter study, the classification achieved 95% accuracy over two level (mid to low and mid to high) by considering an improved objective intelligibility assessment of spastic dysarthric speech.

This paper presents an approach for assessing the severity levels of dysarthria by combining Linear Discriminant Analysis (LDA) with two classification methods: the Gaussian Mixture Model (GMM) and the Support Vector Machine (SVM). The discriminant analysis is used to select a pool of relevant prosodic features having a prominent discrimination capacity. We compare the performance of two combinations: LDA-GMM and LDA-SVM. The task consists of classifying four severity levels of dysarthria by using the Nemours speech database [6].

The original contribution reported in this paper lies in the selection of the most relevant prosodic features that can be used in the front-end processing of the discriminant analysis to achieve a better performance when compared to existing dysarthria severity level classification systems. Furthermore, the proposed approach reduces the processing time since it represents each observation (sentence) by only one vector of eleven prosodic features, instead of using many acoustic vectors for each observation.

The remainder of the paper is structured as follows. Section 2 gives some definitions related to the prosodic features used by the proposed system. Section 3 presents the discriminant function analysis. In section 4, the experiments and results are presented and discussed. Section 5 concludes this paper.

II. Speech Prosodic Features

Speech is primarily intended to transmit a message through a sequence of sound units in a language. Prosody is defined as a branch of linguistics devoted to the description and representation of speaking elements. Prosodic cues include intonation, stress and rhythm; each of them is a complex perceptual entity, expressed fundamentally using three acoustic parameters: pitch, duration and energy [7]. The stress, timing and intonation in speech that are closely related to the speech prosody, enhance the intelligibility of conveyed message allowing listeners to segment continuous speech into words and phrases easily [8].

In dysarthria, usually, a neurological damage affects the nerves that control the articulatory muscle system involved in speech causing weakness, slowness and incoordination.

Manuscript received March 24, 2013. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC).

K. L. Kadi is with the Faculty of Electronics and Computer Science, University of Sciences and Technology Houari Boumediene, Algiers, Algeria (phone: +213-554-074-873; e-mail: kkadi@usthb.dz).

S.-A. Selouani is with the Department of Information Management, University of Moncton, Campus of Shippagan, NB ,Canada (e-mail: selouani@umcs.ca).

B. Boudraa is with the Faculty of Electronics and Computer Science, University of Sciences and Technology Houari Boumediene, Algiers, Algeria (e-mail: bboudraa@usthb.dz).

M. Boudraa is with the Faculty of Electronics and Computer Science, University of Sciences and Technology Houari Boumediene, Algiers, Algeria (e-mail: mboudraa@usthb.dz).

ISBN: 978-988-19252-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

This disturbance affects the prosody depending on the severity level of dysarthria.

The extraction of a reasonably limited, informative and meaningful set of features is an important step towards the automatic dysarthria severity classification. In this work, we use a discriminant analysis with Wilk's lambda measure to select the prosodic features that will be well adapted to dysarthria classification.

The proposed front-end processes the speech waveform at the sentence level; patients are able to repeat individual unit (phoneme or word) of speech with a fairly normal consistency [9]. For each sentence uttered by each speaker, eleven features are considered: Jitter, Shimmer, mean Pitch, standard deviation of Pitch, number of Periods, standard deviation of Period, proportion of the Vocalic duration (%V), Harmonics to Noise Ratio (dB), Noise to Harmonics Ratio (%), Articulation Rate, and degree of voice Breaks.

A. Mean pitch

The physical correlate of pitch is the fundamental frequency (*F0*) estimated by the vibration rate of the vocal folds during phonation of voiced sounds [7]. The ensemble of pitch variations during an utterance is defined as intonation [10]. The typical range of male speaker is 80-200 Hz (approximately for conventional speech) depends on the mass and length of the vocal chords [11]. In this work, mean pitch is calculated by averaging the fundamental frequency across one sentence by using the autocorrelation method. Mean pitch value in dysarthric speech can help to detect a glottic signal abnormality.

B. Jitter

Jitter represents the variations of fundamental frequency within the time evolution of an utterance. It indicates the variability or perturbation of the time period *(T0)* across several cycles of oscillation. Jitter is mainly affected by a deficiency in control of vocal fold vibration [12]. The threshold of comparison normal/pathologic is 1.04% given by the Multi-Dimensional Voice Processing Program (MDVP) designed by Kay Elemetrics Company [13]. The raw jitter and the normalized jitter are defined by:

$$Jitter(seconds) = \sum_{i=1}^{N-1} |T_i - T_{i+1}| \Big/ N-1 \quad (1)$$

$$Jitter(\%) = Jitter(seconds) \Big/ \frac{1}{N} \sum_{i=1}^{N} T_i \quad (2)$$

where T is the period and N the number of periods.

C. Shimmer

Shimmer indicates the perturbation or variability of the sound amplitude. It is related to the variations of the vocal emission intensity and it is partially affected by the reduction of glottic resistance [12]. MDVP gives 3.81% as a threshold for pathology. Shimmer is estimated similarly as jitter but by using the amplitude as a parameter.

D. Articulation rate

The articulation rate is given by the number of syllables pronounced per second by excluding the pauses [14]. In our study, the more the severity level of dysarthria is high the more the articulation rate decreases.

E. Proportion of the vocalic duration

Vocalic duration is what separates the release from the constriction framing a vowel [15]. Proportion of the vocalic duration (%V) is the fraction of utterance duration which is composed of vocalic intervals [14]. The trouble that maintains voicing over a sustained vowel can be considered as a sign of pathology [16].

F. HNR

Harmonics to Noise Ratio (HNR) represents the degree of acoustic periodicity. Harmonicity is measured in dB, calculated by the ratio of the energy of the periodic part related to the noise energy. Harmonics to Noise Ratio can be used as a measure of voice quality. For example, a healthy speaker can produce a sustained "*a*" with HNR around 20dB [16]. HNR is defined by:

$$HNR(dB) = 10log\left(\frac{E_p}{E_n}\right) \quad (3)$$

where E_p is the energy of the periodic part and E_n is the energy of the noise.

G. Degree of voice breaks

Degree of voice breaks is the total duration of the breaks over the signal, divided by the total duration, excluding silence at the beginning and the end of the sentence [14]. A voice break can occur with sudden stoppage of the air stream due to a transient deficiency in the control of the phonation mechanism [17].

III. Discriminant Function Analysis

Discriminant analysis is used to model a dependent categorical variable's value based on its relationship to one or more predictors. From a set of independent variables, discriminant analysis try to find linear combinations of those variables that best discriminate the classes. These combinations are called discriminant functions and are defined by [18]:

$$d_{ik} = b_{0k} + b_{1k}\, x_{i1} + \dots + b_{pk}\, x_{ip} \quad (4)$$

where d_{ik} is the value of the k^{th} discriminant function for the i^{th} class

p is the number of predictors (independent variables)

b_{jk} is the value of the j^{th} coefficient of the k^{th} function

x_{ij} is the value of the i^{th} class of the j^{th} predictor

The number of functions equals *min(number of classes-1, number of predictor).*

The procedure automatically chooses a first function that will separate the classes as much as possible. It then selects a second function that is both uncorrelated with the first

ISBN: 978-988-19252-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

function and provides as much further discrimination as possible. The procedure continues adding functions in this way until achieving the maximum number of functions, as determined by the number of predictors and categories in the dependent variable. For selecting the best variables to use in the model, the stepwise method can be used [18].

Wilks' lambda is a method of variable selection for stepwise discriminant analysis that selects variables on the basis of their capacity to minimize Wilks' lambda. At each step, the variable that reduces the overall Wilks' lambda is entered [18]. The Wilks' lambda method needs a discrimination capacity measure.

To measure the discriminant capacity of every variable X_p, we use the univariate ANOVA (Analysis of Variance). Its decomposition formula is [19]:

$$\underbrace{\sum_{i=1}^{I}\sum_{n=1}^{N_i}(X_{jin}-\bar{X}_j)^2}_{Total\ covariance} = \underbrace{\sum_{i=1}^{I}N_i(\overline{X_{ji}}-\bar{X}_j)^2}_{Separate-groups\ covariance} + \underbrace{\sum_{i=1}^{I}\sum_{n=1}^{N_i}(X_{jin}-\overline{X_{ji}})^2}_{Within-groups\ covariance} \quad (5)$$

We consider a dataset with N observations constituted by $X_1,\dots,X_p$ variables. These observations are partitioned by a qualitative variable into I classes having the sizes: $N_1,\dots,N_I$.

X_{jin} is the value of X_j for the n^{th} observation of the class i

$\overline{X_{ji}}$ is the average of X_j on the class i

$\bar{X}_j$ is the average of X_j

For each variable X_p, Wilks' Lambda is calculated by the ratio of the within-groups-covariance and total-covariance. Smaller value of Lambda indicates greater ability of discrimination [19].

$$\Lambda_p = \frac{within\ group\ sum\ of\ squares}{total\ sum\ of\ squares} \quad (6)$$

In this work, we created a discriminant model that classifies dysarthric speakers into one of the four predefined "severity level of dysarthria" groups. This model uses eleven prosodic features that have been selected by the Wilks' lambda method through the use of a discriminant analysis. To determine the model of the relationship between a categorical dependent variable (severity level) and independent variables (eleven features), we use a linear regression procedure.

IV. Experiments and Results

A. Speech material

Nemours is one of the few databases of recorded dysarthric speech. It contains 814 short nonsense sentences spoken by 11 American patients with varying degrees of dysarthria. Additionally, the database includes two connected-speech paragraphs: the "Grandfather" passage and the "Rainbow" passage produced by each of the 11 speakers. Each sentence in the database is of the form "*The X is Ying the Z*" and generated by randomly selecting *X* and *Z* from a set of 74 monosyllabic nouns without replacement and selecting *Y* from a set of 37 disyllabic verbs without replacement. This process generated 37 sentences from which another 37 sentences were produced by swapping the *X* and *Y* [6]. Therefore, each noun and verb was produced twice by each patient over the complete set of 74 sentences. The whole database has been marked at the word level; sentences for 10 of the 11 talkers have been marked at the phoneme level also. The entire speech corpus was recorded by one non-dysarthric speaker as a control. Speech pathologist conducted the recording session, he was considered as the healthy control (HC). All speech materials were recording using a 16 kHz sampling rate and 16 bit sample resolution after low pass filtering at 7500 Hz cutoff frequency with 90 dB/Octave filter [6].

B. Subjects

The speakers are eleven young adult males suffering different types of dysarthrias resulting from either Cerebral Palsy (CP) or head trauma (HT) and one male adult control speaker. Seven of the talkers had CP, among whom three had spastic CP with quadriplegia and two had athetoid CP (one quadriplegic), two had a mixture of spactic and athetoid CP with quadriplegia. The remaining four subjects were victims of head trauma. The speech from one of the patients (head trauma, quadriplegic) was extremely unintelligible and so poor, it was not marked at the phoneme level, and perceptual data were not collected for this patient. A code of two letters was assigned to each patient: BB, BK, BV, FB, JF, KS, LL, MH, RK, RK and SC. The patients can be divided into three subgroups due to Frenchay Dysarthria Assessment scores (see Table I): one 'mild L1', including patients FB, BB, MH, and LL; the second subgroup 'severe L2' include patients RK, RL, and JF and the third subgroup 'severe L3' is severe and includes patients KS, SC, BV, and BK. The speech assessment and the perceptual data did not take into consideration the too mild case (subject FB) and the too severe case (KS) [1], [2].

TABLE I
Frenchay Dysarthria Assessment Scores of Dysarthric Speakers of Nemours Database [6]

Patients	KS	SC	BV	BK	RK	RL	JF	LL	BB	MH	FB
Severity (%)	-	49.5	42.5	41.8	32.4	26.7	21.5	15.6	10.3	7.9	7.1

C. LDA

In this part, we present results of discriminant analysis using the stepwise method, the linear regression procedure, and the Wilks' lambda.

Table II shows the eleven selected prosodic metrics and their ability to discriminate the four severity levels of dysarthria. Wilks' lambda varies from 0 to 1; smaller values of lambda reveal greater ability of discrimination.

Discriminant analysis generated three discriminant functions to distinguish four severity level of dysarthria. The first two functions are more meaningful for the classification. Figure 1 represent the four classes discriminated by the first two discriminant functions.

TABLE II
WILK'S LAMBDA OF THE ACOUSTICS FEATURES

Feature	Wilks' lambda
Articulation rate	0.565
Number of period	0.595
Mean pitch	0.701
Voice breaks	0.835
%V	0.861
HNR	0.864
Jitter	0.925
Shimmer	0.962
Std Pitch	0.979
Std Period	0.984
NHR	0.989

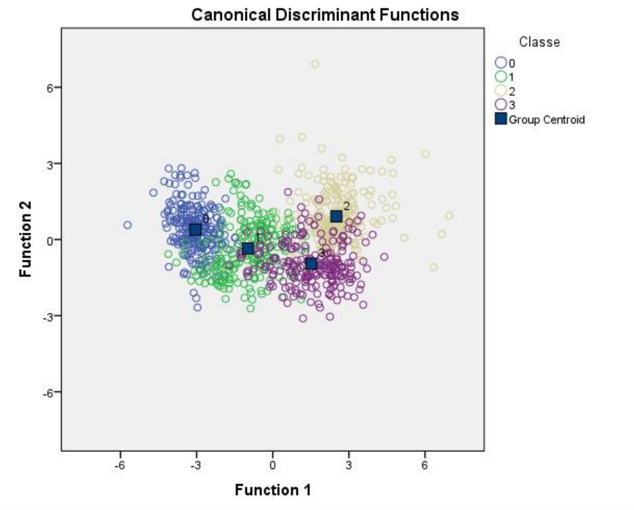

Fig. 1. Representation of combined groups

Classification summary of discriminant analysis with the rate of correct classification are presented in Table3:

TABLE III
CLASSIFICATION RESULTS

		Predicted Group Membership				
	Class	0	1	2	3	Total
Count	0	215	7	0	0	222
	1	27	185	1	9	222
	2	0	1	193	28	222
	3	0	31	29	162	222
%	0	96,8	3,2	0,0	0,0	100
	1	12,2	83,3	0,5	4,1	100
	2	0,0	0,5	86,9	12,6	100
	3	0,0	14,0	13,1	73,0	100

The largest number of misclassification occurs for the classification of the level 'sever L3' (60/222), which also appears in Figure 1 where the representation of L3 is scatter. Most the other errors occur between two nearby classes. To reduce the misclassification rate automatic classifiers having a high discrimination capacity are used.

D. *Automatic classifiers of severity level*

We compare the two approach of automatic classification, GMM and SVM using as front-end the eleven prosodic features selected in the linear discriminant analysis. The two methods perform training and classification. We divided the entire set of sentences of the corpus into two subsets: the training subset that contains 70% of the sentences with different severity levels of dysarthria and the test subset that contains 30% of the sentences. The training subset includes 459 sentences of dysarthtic speech + 153 sentences of non-dysarthric speech (HC) ; the test subset contains 207 sentences of dysarthtic speech + 69 sentences of non-dysarthric speech.

GMM

The issue of automatic classification of observed vectors into one of the I classes can be performed using the Gaussian Mixture Model method.

Training: For each class C_i from the corpus, the training is initiated to obtain a model containing the characteristics of each Gaussian distribution m of the class: the average vector $\mu_{i,m}$, the covariance matrix $\sum_{i,m}$, and the weight of the Gaussian $w_{i,m}$. These parameters are calculated after performing a certain number of iterations of the expectation maximization (EM) algorithm [20]. One model is generated for each severity level of dysarthria.

Recognition: Each extracted signal X is represented by the acoustical vector x of p components. The size of acoustical vector d is the number of acoustical parameters extracted from the signal. The likelihood of each acoustical vector given for a class C_i is estimated. The likelihood is defined by (M is the number of Gaussians) [21]:

$$p(x \backslash C_i) = \sum_{m=1}^{M} w_{i,m} \cdot \frac{1}{\sqrt{(2\pi)^d |\sum_{i,m}|}} \cdot e^{A_{i,m}}$$

$$A_{i,m} = \left(-\frac{1}{2} \left(x - \mu_{i,m}\right)^T \cdot \frac{1}{\sum_{i,m}} \cdot \left(x - \mu_{i,m}\right) \right) \qquad (7)$$

Each sentence is represented by one acoustical vector contained eleven prosodic features (not by one vector for each frame), the likelihood of the signal is denoted by $p(x \backslash C_i)$. The algorithm estimates that the signal X will belong to the group C_i in which $p(x \backslash C_i)$ is greater. The highest rate was achieved by using eight Gaussians (M=8): 88.89% of correct classification of dysarthria severity levels.

SVM

The theory of Support Vector Machine (SVM) was proposed by Vapnik as a new method of machine learning, via introduction of the kernel function [22]. The kernel function projects the (non-linearly separable) data to a new high dimension space where a linear separation is possible. SVM is a supervised binary linear classifier which finds the linear hyperplan separator that maximizes the margin between two classes of data.

The key technology of SVM is Kernel function; choice the type of kernel function will affect learning ability and generalization capacity of machine learning [23]. In our

ISBN: 978-988-19252-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

experiments, the Radial Basis Function (RBF) as Kernel function of SVM is used. RBF properties depend of the Gaussian width σ and the error penalty parameter C. The RBF Kernel is defined by:

$$k(x, y) = exp\left(\frac{\|x - y\|^2}{\sigma^2}\right) \quad (8)$$

Multiclass-SVM using the 'One against one' method is set to perform the classification of severity levels of dysarthria. Binary classifiers are build to differentiate classes C_i and C_j, $0 < i \leq I$ and $0 < j < i$, I is the number of classes [24]. The number of binary classifiers (SVM) necessary to classify I classes is $\frac{I\,(I-1)}{2}$.

The multiclass-SVM includes six SVMs and a decision function based on majority voting (best candidate) using all classifiers. For each of the six SVMs, a cross-validation was carried out over four subset of the corpus to determine the most relevant pair (C, σ) of the RBF Kernel function. This method of automatic assessment achieves **93%** correct rate of dysarthria severity level classification.

V. Conclusion

In this paper, we proposed and compared GMM and SVM discriminative approaches to perform an assessment of the dysarthria severity levels. A reliable front-end processing using relevant prosodic feature is proposed. These features have been selected after performing a discriminant linear analysis. We believe that the proposed system could constitute an appropriate objective test for the automatic evaluation of dysarthria severity. For the clinicians, this tool might be useful and it can be used to prevent wrong subjective diagnosis.

References

[1] S.-A. Selouani, H. Dahmani, R. Amami, and H. Hamam, "Dysarthric speech classification using hierarchical multilayer perceptrons and posterior rhythmic features," *6th international conference, Soft Computing Models in Industrials and Environmental Applications, SOCO*, 2011.
[2] J. B. Polikoff and H. T. Bunnel, "The Nemours database of dysarthric speech: a perceptual analysis," *14th International Congress of Phonetic Sciences, ICPhS*, pp. 783-786, 1999.
[3] S.-A. Selouani, H. Dahmani, R. Amami, and H. Hamam, "Using speech rhythm knowledge to improve dysarthric speech recongnition," *International Journal of Speech Technology*, Vol. 15, no. 1 , pp. 57-64, 2012.
[4] F. Rudzicz, "Phonological features in discriminative classification of dysarthric speech," *ICASSP*, 2009.
[5] M. S. Paja and T. H. Falk, "Automated dysarthria severity classification for improved objective intelligibility assessment of spastic dysarthric speech," *INTERSPEECH*, 2012.
[6] X. Menendez-Pidal, J. B. Polikoff, S. M. Peters, J. E. Leonzio, and H. T. Bunnell, "The Nemours database of dysarthric speech," *Fourth International Conference on Spoken Language, ICSLP*, vol. 3, pp. 1962-1965, 1996.
[7] L. Mary, and B. Yegnanarayana, "Extraction and representation of prosodic features for language and speaker recognition," *Elsevier, Speech Communication*, vol. 50, no. 10, pp. 782-796, 2008.
[8] E. Shriberg, A. Stolcke, and D. Hakkani, "Prosody-based automatic segmentation of speech into sentences and topics," *Speech Communication. Special Issue on Accessing Information in Spoken Audio*, vol .32, no. 1-2, pp. 127-154, 2000.
[9] R. Kent, H. Peters, P. Van-Lieshout, and W. Hulstijn, "Speech motor control in normal and disordered speech," *Oxford University Press*, 2004.
[10] J. T. Hart, R. Collier, and A. Cohen , "A perceptual study of intonation", *Cambridge University Press*., 1990.
[11] L. Mary, *Extraction and representation of prosody for speaker, speech and language recognition*, Springer Briefs in Speech Technology, 2012, ch. 1.
[12] H. F. Westzner, S. Schreiber, and L. Amaro, "Analysis of fundamental frequency, jitter, shimmer and vocal intensity in children with phonological disorders," *Brazilian Journal of Orthinolaryngology*, vol. 71, no. 5, pp. 582-588, 2005.
[13] Multi-Dimensional Voice Processing Program (MDVP), Kay Elemetrics Company: http://www.kayelemetrics.com
[14] J. M. Liss, L. White, S. L. Mattys, K. Lansford, A. J. Lotto, S. M. Spitzer, and J. N. Caviness, "Quantifying speech rhythm abnormalities in the dysarthrias," *Journal of Speech, Language, and Hearing Research*, vol. 52, pp. 1334-1352, 2009.
[15] Calliope, *La parole et son traitement automatique*, Dunod, 1989.
[16] P. Boersma and D. Weenink, "Praat, a system for doing phonetics by computer," *Glot International*, vol. 5, no. 9-10, pp. 341-345, 2001.
[17] C. E. Guerra and D. F. Lovey, "A modern approach to dysarthria classification," *Engineering in Medecine and Biology Society, Proceedings of the 25th Annual International Conference of the IEEE, EMBS*, 2003.
[18] Copyright IBM Corporation., 1989-2012. Available: http://www.ibm.com
[19] A. El Ouardighi, A. El Akadi, and D. Aboutadjine, "Feature selection on supervised classification using Wilk's lambda statistic," *International Symposium on Computational Intelligence and Intelligent Informatics, ISCCIII*, pp. 51-55, 2007.
[20] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum-likelihood from incomplete data via the EM algorithm," *Journal of the Acoustical Society of America*, vol. 39,no. 1, pp. 1-38, 1977.
[21] D. Istrate, E. Castelli, M. Vacher, L. Besacier and J. Serignat, "Information extraction from sound for medical telemonitoring," *IEEE Trans. Information Technology in Biomedicine*, vol. 10, no. 2, pp. 264-274, 2006.
[22] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Trans. Neural Networks*, vol. 10, no. 5, pp. 988-999, 1999.
[23] H. Gao, A. Guo, X. Yu, and C. Li, "RBF-SVM and its application on network security risk evaluation," *4th International Conference on Wireless Communication, Networking and Mobile Computing (WiCOM'08)*, 2008.
[24] A. Fleury, M. Vacher, and N. Noury, "SVM-Based multimodal classification of Activities of daily living in health smart homes: sensors, algorithms, and first experimental results," *IEEE Trans. Information Technology in Biomedicine*, vol. 14, no. 2, pp. 274-283, 2010.

ISBN: 978-988-19252-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)