INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS
ISSN 2320-7345

FEATURE DIMENSIONALITY REDUCTION THROUGH GENETIC ALGORITHMS FOR FASTER SPEAKER RECOGNITION

E.Malathi[1], P.Elango[2]

[1]M.Phil Research Scholar, PG & Research Dept. of Computer Science
Gobi Arts & Science College (Autonomous), Gobichettipalayam - 638 453
malathyeswaran@gmail.com
[2]Associate Professor of Comp. Science, PG & Research Dept. of Computer Science
Gobi Arts & Science College (Autonomous), Gobichettipalayam - 638 453
pitchaielango@gmail.com

Abstract:

Mel-Frequency Cepstral Coefficients and their derivatives square measure commonly used as acoustic options for speaker recognition. Reducing the amount of options results in additional strong estimates of model parameters, and hastens the classification task, that is crucial for period speaker recognition applications running on low-resource devices. During this paper, a feature choice procedure supported Genetic Algorithms (GA) is given and compared to two well-known spatiality reduction techniques, specifically PCA and LDA. Analysis is dispensed for two speech databases, containing laboratory browse speech and telephone spontaneous speech, applying a customary speaker recognition system. Results counsel that dynamic options square measure less discriminant than static ones, since the low-size best subsets found by the GA failed to embody dynamic options. GA-based feature choice outperformed PCA and LDA once coping with clean speech, whereas PCA and LDA outperformed GA-based feature choice for telephone speech, most likely attributable to some reasonably noise compensation inexplicit linear trans-forms, that can't be accomplished simply by choosing a set of options.

Keywords: Speech Recognition, PCA, LDA, Genetic algorithm.

1. Introduction

Smart environments with pervasive computing capabilities need automatic adaptation/customization of the product and services they supply identification may be a natural method of customizing several services, by 1st characteristic and so retrieving information concerning shoppers. Once identification, consumer selections or activities are often half tracked and hold on, up their profiles for any interactions. In these applications, transparency and naturalness are critical. Shoppers ought to remember that their activities are being half-tracked, their voice recorded and their profile info hold on, but these actions shouldn't interfere with the service itself counting on the interface, speakers can be unendingly tracked, or simply identified once the service is started. In any case, they can't be asked even for

simply some seconds of speech knowledge to form correct profiles, nor interactions delayed thanks to a computationally costly search of their profiles. Moreover, shoppers could also be accessing the service through a conveyable or embedded device with low storage and computational capabilities. During this case, real-time processing becomes the foremost crucial issue to permit natural interactions. For these latter applications, high dimensional feature vectors don't appear appropriate, and a few reasonably spatial property reduction technique should be applied to avoid wasting the maximum amount time as attainable with no or very little performance degradation.

State-of-the-art speaker recognition systems use short spectrum options, the Mel-Frequency Cepstral constant s (MFCC) [4], as a result of they convey not solely the frequency distribution distinguishing sounds, however conjointly speaker specific options .Additionally, it's been shown that dynamic info improves the performance of recognizers, so MFCC, energy and their initial and second derivatives area unit usually used as options. The ensuing feature vectors might carries with it up to fifty parts, all of them conveyance of title a precise quantity of relevant information.

Literature is many comparative studies that think about various feature extraction techniques then choose that yielding the simplest performance in a very target task (see [15] for reference). During this work, instead, we tend to aim to neatly cut back feature dimension to hurry up computation whereas keeping performance. MFCC, energy and their 1st and second derivatives square measure taken because the baseline acoustic options. Then, the K options most relevant to the classification task square measure extracted. In different words, the D-dimensional feature area is reworked into a K-dimensional mathematical space (K & lt; D) that minimizes the loss of relevant info. A hardiness issue is additionally concerned, since a restricted quantity of knowledge is accessible to estimate speaker models.

The problem of spatiality reduction is typically formulated as a linear remodel that comes feature vectors on a reworked topological space outlined by relevant directions. Given a D-dimensional feature vector X , a K • D matrix A is applied to induce a K-dimensional vector Y of reworked options (K < D). The matrix A is calculable in order that, from the purpose of read of classification, redundancy is removed and relevant info preserved. this could, at least, optimize the performance for the target price of K, however it should even surpass the baseline feature set, attributable to the removal of harmful or confusing options and, a lot of in all probability, to raised (more robust) estimates of model parameters. the subsequent strategies are planned (among others)

Principal part Analysis (PCA) [12], associate degree recent technique of variable applied mathematics analysis, consists of computing the eigenvectors of the D • D variance matrix □, then sorting them per the corresponding eigenvalues, in falling order, and eventually building the projection matrix A (called Karhunen-Loeve Trans-form, KLT) with the most important K eigenvectors (i.e. the K directions of greatest variance). Every feature vector X is then pre-processed per the expression Y = A(X − □), wherever □ represents the mean feature vector. KLT decor relates the features and provides the smallest possible reconstruction error among all linear trans-forms, i.e. the smallest possible mean-square error between the data vectors in the original *D*-dimensional space and the data vectors in the projected *K*-dimensional subspace. Unfortunately, this does not guarantee minimizing classification error.

Linear Discriminant Analysis (LDA) [7] makes an attempt to seek out the transform A that maximizes a criterion of sophistication disjuncture. this can be done by computing the within-class and between-class variance matrices, □wc and □bc, then finding the eigenvectors of □−wc1 □bc, sorting them in line with the eigenvalues in descendent order, and finally building the projection matrix A with the primary K eigenvectors (which outline the K most discriminant hyperplanes). LDA assumes that everyone the classes share a common within-class covariance matrix, and that each class is modeled by a single Gaussian distribution. LDA also assumes that classes are linearly separable. Additionally, as any supervised approach, it requires labeling samples with class identities.

Linear transforms combine in an elegant way feature extraction and feature selection. However, these two steps can be also applied in an uncoupled way. Strictly speaking, feature selection consists of determining an optimal subset of features by exhaustively exploring all the 2^D possible combinations. Most feature selection procedures use

the classification error as the evaluation function. This makes exhaustive search computationally unfeasible in practice, even for moderate values of *D*. The simplest method consists of evaluating the *D* features individually and selecting the *K* most discriminative ones, but it does not take into account dependencies among features. So a number of suboptimal heuristic search techniques have been proposed in the literature, which essentially trade-off the optimality of the selected subset for computational efficiency [11].

Genetic Algorithms (GA) fitly work this sort of complicated optimization issues. Candidate solutions are pictured as individuals during a massive population. An initial solution (which is also randomly generated) are iteratively driven by the GA to associate degree best purpose consistent with a posh metric that measures the performance of the people during a target task. The fittest people are selected and their chromosomes mixed, mutated or taken unchanged to subsequent generation. A serious advantage of GA over alternative heuristic search techniques is that they are doing not believe suppose deem trust admit accept have confidence have faith in place confidence in} any assumption about the properties of the analysis perform. Multi objective evaluation functions (e.g. combining the accuracy and also the value of classification) is outlined and employed in a natural approach [18] [14]. GA will simply encrypt selections regarding choosing or not choosing options as sequences of Boolean values, permit to neatly explore the feature area by retentive those selections that profit the classification task, and at the same time avoid native optima as a result of their intrinsic randomness. GA are recently applied to feature extraction [3], feature weight associate degree feature choice [5] in speaker recognition. In [5], a reduced set of features was determined on a speaker-by-speaker basis by applying GAs to maximize a measure of discrimination between each speaker and her/his two closest neighbors. Speaker recognition performance was measured on a small dataset containing only 15 speakers, and using a very simple speaker identification algorithm. GAs was applied to search for feature weights maximizing speaker recognition performance on a validation dataset. Speaker models were based on empirical distributions of acoustic labels, obtained through vector quantization. Finally, features were sorted according to their weights and the *K* features with greatest average ranks were retained and evaluated.

In this paper, a feature choice procedure supported a GA-driven search is bestowed and compared to PCA and LDA in an exceedingly speaker recognition task. Experiments are applied for 2 speech databases, containing laboratory browse speech and phone spontaneous speech, severally. A customary GMM-based speaker recognition system is applied the remainder of the paper is organized as follows. The speaker recognition system and therefore the feature choice approach are delineating in Sections a pair of and three, severally. The experimental setup is printed in Section four, together with details regarding the speech databases, the computation of MFCC, the speaker models and therefore the implementations of GA, PCA and LDA. Section five presents the results of the GA-based feature choice approach in speaker recognition experiments, and compares them to those of PCA and LDA. Finally, conclusions are summarized in Section half-dozen.

2. Speaker Recognition

In this work, the distribution of feature vectors extracted from a speaker's speech is delineate by a linear combination of M multivariate Gaussian densities, referred to as Gaussian Mixture Model (GMM) [16]. GMM parameters are estimated from speaker samples by applying the *Maximum Likelihood* (ML) criterion. Each sample *X* consists of a sequence of *D*-dimensional feature vectors: $X = (x_1, x_2, \ldots, x_T)$. The conditional probability of a feature vector *x*, given the speaker model $\lambda = \{w_j, \mu_j, \Sigma_j \mid j = 1, \ldots, M\}$, is computed as follows:

$$p(x|\lambda) = \sum_{j=1}^{M} w_j N(x; \mu_j, \Sigma_j) \qquad (1)$$

Where $\mathbf{N}(x; \mu, \Sigma)$ denotes the *D*-dimensional normal density function of mean vector μ and covariance matrix Σ, and the mixture weights satisfy the constraint $\sum_{j=1}^{M} w_j = 1$.

We assume that input utterances are produced by *S* known speakers, represented by their corresponding models $\lambda_1, \lambda_2, \ldots, \lambda_S$. Then, for any input utterance $X = (x_1, x_2, \ldots, x_T)$, the most likely speaker i(x) is selected

according to the following expression:

$$i(X) = \arg\max_{i=1,\ldots,S} \log p(\mathrm{l}i|X) \qquad (2)$$

Applying the Bayes rule, taking into account that maximizing over the set of speakers does not depend on the acoustic sequence, assuming that all the speakers have equal *a priori* probabilities and that acoustic vectors are statistically independent, it follows:

According to Eq. 3, the computational cost of speaker recognition depends linearly on the number of speakers (S) and on the length of the input utterance (T). Since GMM are used as speaker models, the computational cost also depends linearly on the number of mixtures (M) and on the dimension of the feature space (D).

3. Feature Selection Using Genetic Algorithms

In this study, the well-known *Simple Genetic Algorithm* (SGA) [10] is applied to search for the optimal feature set. The evaluation of feature sets (i.e. the fitness function used by the GA) is base d on the classification accuracy obtained in speaker recognition experiments for development data.

The GA-driven selection process begins by fixing the target size K of the reduced feature subspace. Then, an initial population of candidate solutions (K-feature subsets) is randomly generated. To evaluate the K-feature subset □ = { f_1, f_2, . . . , f_K }, the following steps are carried out: (1) the acoustic vectors of the whole speech database are reduced to the components enumerated in □;(2) speaker models are estimated using a training corpus; (3) utterances in a development corpus are classified by applying the speaker models; and (4) the speaker recognition accuracy obtained for the development corpus is used to evaluate □.

Each candidate solution is represented by a D-dimensional vector of positive integers $R = \{r_1 , r_2, \ldots , r_D\}$, the K highest values determining what features are selected. Note that the same feature set □ may be represented by different vectors, that is, modifications to a given candidate solution R might not change the selection of features. This redundancy in representation makes the genetic algorithm to evolve smoothly and facilitates its convergence.

At the tip of every iteration/generation, on balance the K-feature subsets within the population square measure evaluated, a number of them (usually the fittest ones), square measure elite, mixed and mutated so as to induce the population for following generation. Mutation is employed to introduce little variations that facilitate decrease the possibilities of obtaining native optima. On the opposite hand, political theory (copying a number of the fittest people to following generation) is applied to ensure that the fitness function increases monotonically with sequent generations. If that increase is smaller than a given threshold, or a most variety of generations is reached, the rule stops and therefore the best K-feature set □ˆ = is came back. Finally, □ˆ is evaluated on a take a look at corpus. The 3 datasets utilized in this procedure: coaching, development and take a look at, square measure freelance and composed of disjoint sets of utterances.

4. Experimental Setup

4.1 Speech databases

Two series of experiments were carried out for two different databases, emphAlbayz´ın and *Dihana*, each partitioned in three sets: (1) the training set, used to estimate the speaker models; (2) the development set, used by the GA to compute the fitness function; and (3) the test set, used to evaluate the performance of the optimal K-feature subset.

Albayz´ın is a phonetically and gender-balanced database in Spanish, recorded at 16 KHz in laboratory

conditions [2]. It contains 204 speakers, each speaker contributing 25 utterances in a single session, each utterance lasting an average of 3.55 seconds. The 25 utterances corresponding to each speaker are distributed as follows: 10 are taken for training, 7 for development and 8 for testing. So, the training, development and test sets are composed of 2040, 1428 and 1632 utterances, respectively.

The coaching set consists of two dialogues and eight phonetically balanced scan utterances per speaker, and each the event and check sets carries with it one dialogue and four task-specific scan utterances per speaker . The coaching, development and check sets contain 4598, 2379 and 2897 utterances, severally.

4.2 Speech processing

Speech is analyzed in 25-millisecond frames, at intervals of ten milliseconds. A overacting window is applied associated an FFT computed, whose length depends on the sampling frequency: 256 points for signals sampled at eight kilohertz and 512 points for signals sampled at sixteen kilohertz. FFT amplitudes area unit then averaged in twenty (8 kHz) or twenty four (16 kHz) overlapped triangular filters, with central frequencies and bandwidths outlined in keeping with the Mel scale. To extend strength against channel distortion, Cepstral Mean standardization (CMN) [17] is applied on associate utterance-by-utterance basis. the primary and second derivatives of the MFCC, the frame energy (E) and its 1st and second spinoff s are computed, so yielding a 33-dimensional (8 kHz) or a 39-dimensional (16 kHz) feature vector.

4.3 Speaker models

The baseline system uses 32-mixture diagonal variance GMM as speaker models. The amount of mixtures was tuned in preliminary experiments, progressing to get an acceptable trade-off between computational load and performance. metric capacity unit estimates of model parameters are computed from speaker samples by applying the unvaried Expectation-Maximization (EM) algorithmic rule [6], ranging from random values. Although little iteration is enough for the model parameters to converge, the random nature of format implies that totally different |completely different} runs of the EM algorithmic rule will result in different parameter estimates.

4.4 GA implementation

The genetic rule was enforced by means that of ECJ [8], a Java-based biological process Computation and Genetic Programming analysis System, developed at American Revolutionary leader University's Evolutionary Computation Laboratory and free beneath a special open supply license. Preliminary experimentation was dole out to ad-just the parameters that management the performance and convergence of the GA. Population size is one in all the foremost crucial parameters: high volume populations create the convergence of the rule too slow, whereas too tiny populations might limit the search performance. The second parent was chosen within the same approach, however solely from 2 arbitrarily chosen people, to permit diversity and avoid native optima. One-point crossover was applied and therefore the mutation likelihood was set to zero.01. Finally, the best case of political orientation was applied by keeping the fittest individual for following generation.

4.5 PCA and LDA implementations

LNKnet [13], public domain software developed at MIT Lincoln Laboratory, was used to perform PCA. Regarding LDA, a custom implementation was developed in Java. It computes the within-class covariance matrix $\square_{wc}$ as a weighted average of the covariance matrices of speakers, using the fraction of training samples corresponding to each speaker as weight. Covariance matrices of speakers are estimated from training data, assuming a single Gaussian density model. The between-class covariance matrix is computed by subtracting the within-class covariance matrix from the global covariance matrix: $\square_{bc} = \square_{g} - \square_{wc}$.

5. Results and Discussion

5.1 Performance of the feature sets provided by the GA

Table one shows the mean speaker recognition error rates and therefore the ninety fifth confidence intervals obtained with the best feature sets provided by the GA. Recognition results for 3 reference sets (MFCC, MFCC+E and therefore the full feature vector) area unit shown too. For instance the consistency of the best sets provided by the GA, error rates for each the event and take a look at sets area unit shown. Note that the GA appearance for the most effective K-dimensional feature set by acting speaker recognition experiments on a development dataset; whereas the performance of the best set is measured on AN independent take a look at set. The shut correlation between the rates for each sets supports the utilization of genetic algorithms for this sort of optimization issues.

Confidence intervals allow significant performance comparisons among different feature sets. Preliminary experimentation showed that, fixed the set of features and the training database, random initializations led to slightly different model parameters after convergence, and therefore slight differences in speaker recognition performance were observed. This uncertainty can be taken into account in performance comparisons by computing the mean error rate and the corresponding confidence interval in a significant number of experiments. In this stud y, the whole process of training speaker models and carrying out speaker recognition experiments on the test set was repeated 20 times, and the 95% confidence interval was computed, assuming a Gaussian distribution of error rates.

In the experiments for clean speech, the recognition error rate decreases consistently as the number of features increases from 6 to 12, but performance improvements become relatively smaller for $K > 12$. Since optimal feature sets for $K \leq 12$ consist exclusively of a number of MFCCs plus the frame energy, this suggests that,

Table 1: Mean error rates and 95% confidence intervals in speaker recognition experiments for clean and telephone speech using the optimal *K*-dimensional feature subsets provided by the GA, for $K = 6, 8, 10, 11, 12, 13, 20$ and 30. Results using MFCC, MFCC+E and the full feature vector are shown too, for reference.

K	Clean speech		Telephone speech	
	Development	Test	Development	Test
6	7.64±0.12	5.71±0.09	31.76±0.16	34.23±0.12
8	2.86±0.12	1.81±0.09	21.99±0.13	23.90±0.14
10	2.24±0.11	0.94±0.04	17.91±0.16	19.70±0.12
11	0.81±0.06	0.35±0.04	17.64±0.11	19.32±0.14
12	1.23±0.07	0.30±0.04	17.37±0.09	19.27±0.14
13	1.05±0.06	0.36±0.03	17.30±0.12	19.12±0.14
20	0.67±0.09	0.16±0.02	17.59±0.09	19.99±0.11
30	0.57±0.05	0.13±0.02	16.05±0.14	19.10±0.14
MFCC	1.27±0.08	0.40±0.06	17.91±0.16	19.70±0.12
MFCC+E	0.90±0.05	0.22±0.04	19.76±0.14	22.34±0.10
Full feature vector	0.77±0.09	0.20±0.03	15.66±0.16	18.69±0.15

When dealing with clean laboratory speech, the information about speaker characteristics contained in dynamic features (first and second derivatives) is less relevant than that contained in static features. It does not mean that dynamic features are useless. Many studies have demonstrated that including them improves performance. It only means that when reduced sets must be defined, static features are the best choice.

Results for telephone speech also support this conclusion: the optimal feature sets for $K \leq 10$ are composed exclusively by MFCCs, and performance improvements for $K > 10$ are very small. It is worth noting the case of the reference subset composed of 10 MFCC and the frame energy, whose performance is 1.85 absolute points worse than that of the subset composed exclusively by 10 MFCC. This result reveals the lack of robustness of the frame energy when dealing with telephone speech, an issue that was already discovered by the GA in the selection experiments, since the optimal feature subsets for $K \leq 13$ did not include the frame energy.

5.2 Comparing GA to PCA and LDA

GA-based feature choice comes the first D-dimensional feature house into a reduced K-dimensional topological space by simply choosing K options. PCA and LDA not solely cut back however additionally scale and rotate the first feature house, through a metamorphosis matrix A that optimizes a given criterion on the coaching information. From this time of read, PCA and LDA generalize feature choice, however the standards applied to calculate A (the

highest variance in PCA, and therefore the highest quantitative relation of between to inside category variances in LDA) don't match the criterion applied in analysis. This is often the subject of GA, since feature choice is performed so as to maximise the speaker recognition rate on associate degree freelance development corpus.

GA-based feature selection, PCA and LDA were tested in speaker recognition experiments on clean and telephone speech. First, *D*-dimensional feature vectors were transformed into reduced *K*-dimensional feature vectors, according to the optimal sub-set/transformation given by GA, PCA or LDA, then speaker models were estimated on the training corpus and finally speaker recognition experiments were carried out on the test corpus. Again, the mean error rate and the 95% confidence interval in 2 0 different experiments are given, to account for the uncertainty intrinsic to the estimation of GMM parameters.

In the case of clean speech, neither PCA nor LDA outperformed GA. PCA yielded lower error rates than LDA for $K > 12$. For $K \leq 12$, LDA outperformed PCA. However, the error rates are too low and the differences in performance too small for these conclusions to be statistically significant.

Error rates for telephone speech were much higher than those obtained for clean speech. Besides considering the presence of channel and environment noise, it can be argued that a large part of that corpus consists of spontaneous speech. The presence of noise makes PCA and LDA more suitable than GA, because feature selection cannot compensate for noise, whereas linear transforms can do it to a certain extent. This may explain why either PCA or LDA outperformed GA in all cases but for $K = 8$. LDA was the best approach in most cases (for $K = 10$, 11, 12, 13 and 20), whereas GA was the second best approach for $K = 6$, 10, 11, 12 and 13. On the other hand, the lowest error rate (15.97%) was obtained for $K = 30$ using PCA.

In summary, the GA-based feature choice theme projected during this paper appears to be competitive only if addressing clean speech, although it performs quite well even for telephone-channel speech once the target K is tiny. Authors that argue against GA optimization say that it's too expensive, since it needs iteratively evaluating candidate solutions in classification experiments over a development dataset. It should be noted, however, that GA optimization is completed off-line, therefore the procedure price isn't a problem in practice. Moreover, throughout recognition, feature choice is a smaller amount expensive than feature transformation.

5.3 Empirical time savings

To check empirically the time savings that could be attained by reducing the number of features, recognition times were recorded for several values of *K* in two different computers (see Figure 1). As expected, the running time *t* grew linearly with *K*. In the case of Albayz´ın (clean/laboratory/read speech), using 13-dimensional feature vectors took on average around 40% the time of using full 39-dimensional feature vectors. In the case of Dihana (telephone/office/spontaneous speech), similar savings were observed when comparing the running times of 10-dimensional and 33-dimensional feature vectors.

6. Conclusion

In this work, genetic algorithms were applied to look for the sub-set of K options maximizing the popularity performance. Alternatively, 2 well-known feature spatial property reduction techniques, PCA and LDA, were applied and their performance compared to it of the GA-based feature choice approach. Experiments were applied for 2 speech databases in Spanish, containing scan speech in laboratory conditions and spontaneous speech through telephone lines, severally, applying a customary GMM-based speaker recognition system.

Feature choice supported GA suggests that static options are a lot of discriminant than dynamic options for speaker recognition applications. If a reduced set of options had to be elite (due to storage or process restrictions), MFCC would be the simplest alternative, increased with the frame energy once managing clean-laboratory speech. Within the case of telephone speech, the littlest feature subsets ($K \leq 13$) didn't embrace the frame energy, that reveals that channel and/or surroundings noise is distorting the information it conveys. Concerning the methodology, the consistency of the feature choice results across the event and take a look at datasets validates the utilization of GA for this sort of improvement issues.

At the end of this study, we were tempted to combine the strong points of GA and linear transforms by applying GA to search for the linear transform that maximized the speaker recognition rate on a development set. However,

such an approach was found unfeasible in practice, because determining $K \cdot D$ floating-point transform coefficients (instead of just K feature indices) require a huge amount of training and development data (and a shocking amount of processing time) for the GA to converge and provide a robust transform.

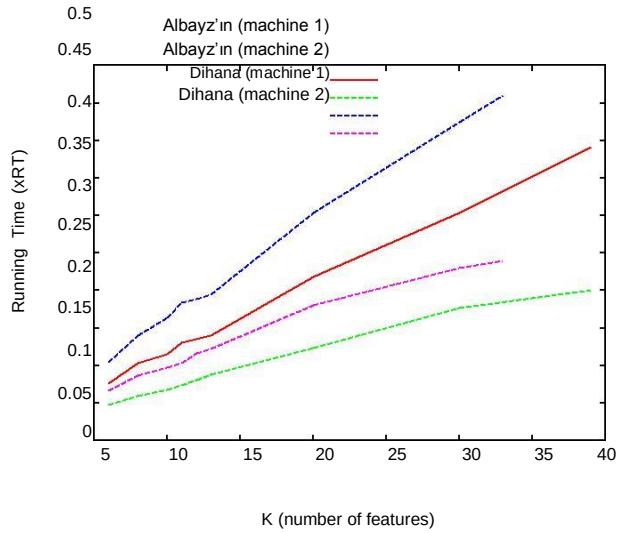

Figure 1: Average running times (real-time factor, xRT) for several values of K, in speaker recognition experiments carried out in two different computers (machine 1: 2 x Quad Core Intel Xeon E5320, 1.86GHz, 1066MHZ FSB, 4GB FB 667MHz; machine 2: 2 x AMD Opteron270 64bit Dual Core 2.0Ghz, 4GB).

7. Acknowledgement

This work has been partially funded by the Spanish MEC, under Plan National de I+D+i, project TSI2006-14250-C02-01; the Government of the Basque Country, under program SAIOTEK, projects S-PE06UN48, S-PE06IK1, S-PE07UN43 and S-PE07IK3; and the University of the Basque Country, under project EHU06/96.

REFERENCE

[1] N. Alcocer, M. J. Castro, I. Galiano, R. Granel, S. Grau, and D. Griol. Adquisici´on de un Corpus de Di´alogo: DIHANA. In *Actas de las III Jornadas en Tecnolog´ıa del Habla (in Span-ish)*, pages 131–134, Valencia (Spain), November 2004.

[2] F. Casacuberta, R. Garc´ıa, J. Llisterri, C. Nadeu, J. M. Pardo, and A. Rubio. Development of Spanish Corpora for Speech Research (Albayz´ın). In *G. Castagneri Ed., Proceedings of the Workshop on International Cooperation and Standardiza-tion of Speech Databases and Speech I/O Assessment Meth-ods*, pages 26–28, Chiavari, Italy, September 1991.

[3] C. Charbuillet, B. Gas, M. Chetouani, and J. L. Zarader. Filter Bank Design for Speaker Diarization Based on Genetic Al-gorithms. In *Proceedings of the IEEE ICASSP'06*, Toulouse, France, 2006.

[4] S. B. Davis and P. Mermelstein. Comparison of Parametric Representations for Monosyllabic Word Recognition in Con-tinuously Spoken Sentences. *IEEE Transactions on Acoustic, Speech and Signal Processing*, 28(4):357–366, 1980.

[5] M. Demirekler and A. Haydar. Feature Selection Using a Genetics-Based Algorithm and its Application to Speaker Identification. In *Proceedings of the IEEE ICASSP'99*, pages 329–332, Phoenix, Arizona, 1999.

[6] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum like-lihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society – Series B* , 39(1):1–38, September 1977.

[7] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification (Second Edition)*. Wiley Interscience, 2000.

[8] ECJ 16. http://cs.gmu.edu/ eclab/projects/ecj/.

[9] N. M. Fraser and G. N. Gilbert. Simulating speech systems. *Computer Speech and Language*, 5:81–99, 1991.

[10] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, 1989.

[11] A. K. Jain, R. P. W. Duin, and J. Mao. Statistical Pattern Recognition: A Review. *IEEE Transactions on Pattern Anal-ysis and Machine Intelligence*, 22(1):4–37, January 2000.

[12] I. T. Jolliffe. *Principal Component Analysis (Second Edition)*. Springer, 2002.

[13] R. P. Lippmann, L. Kukolich, and E. Singer. LNKnet: Neural Network, Machine Learning and Statistical Software for Pattern Classification. *Lincoln Laboratory Journal*, 6(2):249– 268, 1993.

[14] L. S. Oliveira, R. Sabourin, F. Bortolozzi, and C. Y. Suen. A Methodology for Feature Selection Using Multiobjective Genetic Algorithms for Handwritten Digit String Recognition.

International Journal of Pattern Recognition and Artificial l Inteligence, 17(6):903–929, 2003.

[15] D. A. Reynolds. Experimental Evaluation of Features for Ro-bust Speaker Identification. *IEEE Transactions on Speech and Audio Processing*, 2(4):639–643, October 1994.

[16] D. A. Reynolds and R. C. Rose. Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Mod-els. *IEEE Transactions on Speech and Audio Processing*, 3(1):72–83, January 1995.

[17] A. E. Rosenberg, C. H. Lee, and F. K. Soong. Cepstral Channel Normalization Techniques for HMM-Based Speaker Verification. In *Proceedings of the ICSLP'94*, pages 1835–1838, Yokohama, Japan, 1994.

[18] J. Yang and V. Hanover. Feature subset selection using a genetic algorithm. *IEEE Intelligent Systems*, 13(2):44–49, March 1998.