# COMPLEMENTARY FEATURES FOR SPEAKER VERIFICATION BASED ON GENETIC ALGORITHMS

*C. Charbuillet, B. Gas, M. Chetouani, J. L. Zarader*

Université Pierre et Marie Curie-Paris6, EA2385 Groupe Perception et Réseaux Connexionnistes (PRC),
Ivry sur Seine, F-94200 France

## ABSTRACT

Speech recognition systems usually need a feature extraction stage aiming at obtaining the best signal representation. State of the art speaker verification systems are based on cepstrals features like MFCC, LFCC or LPCC. In this article, we propose to use a genetic algorithm to provide new features able to complete the LFCC's.

We present an adaptation of the common LFCC feature extractor which consists in designing a filter bank, optimized for a high level fusion issue.

Experiments are carried out using a state of the art speaker verification system. Results show that the proposed method improves the system performances on the 2006 Nist SRE Database.

***Index Terms—*** Speaker recognition, Feature extraction, Genetic algorithms

## 1. INTRODUCTION

Speech feature extraction plays a major role in speaker recognition systems. State of the art speaker recognition systems front end are based on the estimation of the spectral envelope of the short term signal, e.g., Mel-scale Filterbank Cepstrum Coefficients (MFCCs), Linear-scale Filterbank Cepstrum Coefficients (LFCCs), or Linear Predictive Cepstrum Coefficients (LPCCs). Even if these extraction methods achieve good performances on the speaker verification task, feature extraction from the spectrum could still be improuved. M. Zhiyou and al. [1] proposed to combine LPCCs and MFCCs to improve system's performances. C. Myajima and al. [2] proposed a data driven cepstrum warping function adapted to the speaker identification task.

The aim of this study is to provide complementary features that describe information not captured by the conventional LFCC features for the speaker verification tasks. To achieve this goal, we propose to use a Genetic Algorithm (GA) to optimize a cepstrum based feature extractor. This optimization consists of designing a filter bank, able to complete the LFCCs.

Genetic algorithms (GA) were first proposed by Holland in 1975 [3] and became widely used in various disciplines as a new means of complex systems optimization. In recent years these have been successfully applied to the speech processing domain. Chin-Teng Lin and al. [4] proposed to apply a GA to the feature transformation problem for speech recognition and M. Demirkler and al. [5] worked on a GA based feature selection for speaker recognition.

GAs most attractive quality is certainly their aptitude to avoid local minima. However, our study relies on another quality which is the fact that GAs are an unsupervised optimization method. So they can be used as an exploration tool, free to find the best solution without any constraint. In a previous work [6] we used this approach to show the importance of specific spectral information for the speaker diarization task.

Data driven feature extractor optimizations can be problematic because they are generally adapted to specifics classifiers and can not be directly transposed to conventional speaker verification systems. For example, experiments carried out using a Vector Quantization or Hidden Markov Model can not always be transposed on a classical GMM based speaker verification system.

In this paper, all experiments are done on a state-of-the-art GMM-UBM speaker verification system, and tested on the whole 2006 NIST SRE corpus. The obtained results have been submitted to the 2006 NIST Speaker Recognition Evaluation Campaign [7] and ranked our system 16th/36.

In the first part, filter bank based feature extractors are presented. Afterwards, we describe the genetic algorithm we used, followed by its application to the feature extractor optimization. Then, the experiments we made and the obtained results are presented.

## 2. FILTER BANK BASED CEPSTRUM FEATURE EXTRACTORS

The conventional MFCC and LFCC feature extractor process mainly consists of modifying the short-term spectrum by a filter bank. This process has four steps:

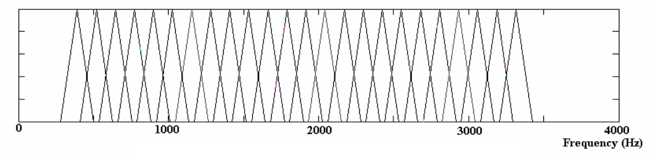

**Fig.1. Linear scaled filter bank**

- Compute the power spectrum of the analyzed frame.
- Sum the power spectrum for each triangular filter of the bank
- Applie the log operator to the obtained coefficients
- Compute the Discrete Cosine Transform (DCT)

Figure 1 presents the linear scaled filter bank used for the LFCC's computation. This feature extractor is known to be the most robust for the short band signals representation.

The purpose of our study is to find a new filter bank, able to provide some complementary information. To this end, we propose to use a genetic algorithm.

## 3. GENETIC ALGORITHM

A genetic algorithm is an optimization method. Its aim is to find the best values of the system's parameters in order to maximize its performance. The basic idea is that of "natural selection", i.e. the principle of "the survival of the fittest". A GA operates on a population of systems. In our application, each individual of the population is a filter bank defined by its bandwidth and center frequency parameters.

### 3.1. The S.E.V. algorithm

The algorithm used is called the Selection Evaluation Variation (SEV) [8] and is illustrated in figure 2. To evolve the desired filter bank, we consider a population **p(t)** of **Np** filter banks undergoing a variation-evaluation-selection loop, i.e. **p(t+1) = S.E.V p(t)**. First, a random initialization is done for each individual of the population **p(0)**. The *variation operator* **V** consists in a random variation of each filter bank's parameters. The *evaluation operator* **E** is defined problem specific, and is usually given in terms of a fitness function. It consists in evaluating the performances of each individual of the population. The performance criterion used in our experiment will be detailed on the section 3.2. After evaluating the performance of each individual filter bank, the *selection operator* **S** selects the **Ns** best individuals. These individuals are then cloned according to the evaluation results to produce the new generation **p(t+1)** of **Np** filter banks. As a consequence of this selection process, the average of the performance of the population tends to increase and in our application adapted filter banks tend to emerge.

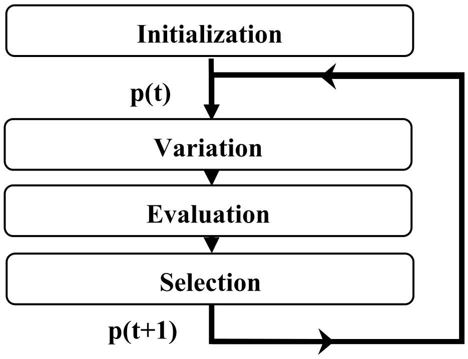

**Fig.2. The S.E.V. algorithm**

### 3.2. Application to complementary feature extraction

The objective is to obtain a filter bank able to complete the LFCCs information. To orient the filter bank evolution in this direction, we need to select the most complementary filter bank.

Our measure of this quality consists of measuring the Equal Error Rate (EER) produced by the arithmetic fusion between the baseline system and the system including the filter bank to be evaluated. A schema of this method is presented in figure 3.

## 4. EVOLUTION SIMULATION

This section presents the evolution simulation we made and the obtained filter bank.

### 4.1. Evolution database

The database used for the evolution phase is extract from the 2004 Nist SRE corpus. This corpus is composed of conversational telephone speech signals passed through different channels, (landline, cordless or cellular) and sampled to 8 kHz. We used 20 male and 20 female with one utterance of 5 minutes per speaker for the train and for the test. The number of tests involved for each filter bank evaluation was of 800.

### 4.2. Speaker verification system used for the evolution

The principal drawback of GA is certainly the computation cost. This is the reason why we used a reduced system for the filter bank evolution.

This system is derived from the system presented on the section 5.2. These characteristics are:

- GMM-UBM models using 32 Gaussian with diagonal covariance matrix.
- No Tnorm normalization

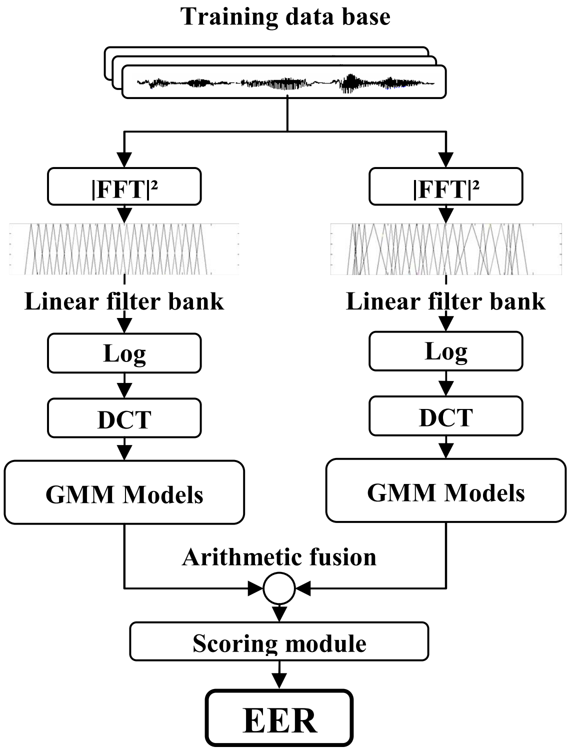

Fig.3. Filter bank evaluation method

### 4.3. Simulation parameters

In this experiment, we used the **S.E.V.** algorithm to optimize the bandwidth parameters of the filter banks. The overlapping between filters was preserved, imposing the center frequency of the filter bank by using the following rule:

$$C_{i+1} = C_i + \frac{B_{i+1}}{2}$$

$C_\mathrm{i}$ and $B_\mathrm{i}$ being the center frequency and the bandwidth respectively of the i$^{th}$ filter in the bank, and $C_0$=300 Hz

The bandwidth and center frequency are discrete parameters varying from 0 to 255 and coding the [0 4000]Hz frequency domain. Before starting the **S.E.V.** algorithm, a random initialization of the bandwidths is done using a uniform random distribution of 25 units (~ 392 Hz).

To evolve the desired filter bank, the following algorithm parameters were:

- *Number of filters in the bank*: 24
- *Feature vector dimension*: 16 cepstral + 16Δ = 32
- *Population size* **Np**: 50
- *Number of selected individuals* **Ns:** 20
- *Variation operator*: uniform random variation of 3 units (~100Hz) for each bandwidth.

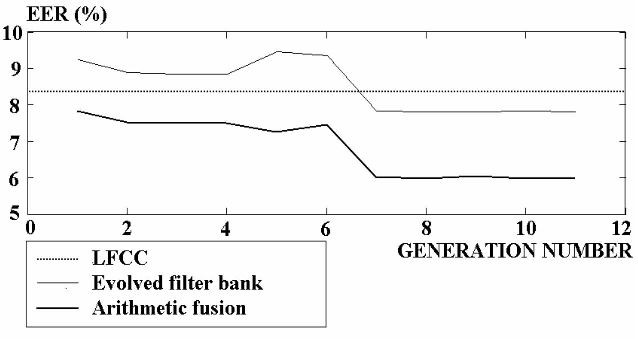

Fig.4. Equal error rates evolution

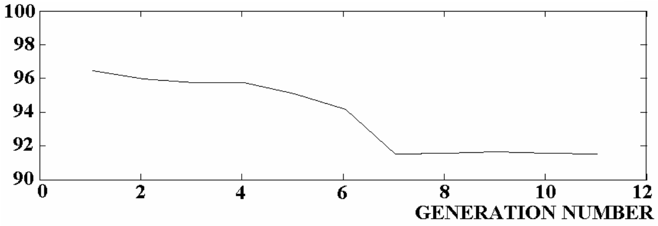

Fig.5. Correlation evolution

### 4.4. Evolution simulation

The evolution simulation was done using the previously defined parameters. The figure 4 reports the evolution of the Equal Error Rate (EER) provided by the arithmetic fusion between the baseline system and the system based on the evolved filter bank. Figure 5 represents the evolution of the correlation between these two systems. This correlation measure is done on the LLR produced by the two systems for all the tests segments of the evolution base.

First, we notice that the EER obtained by the fusion decreases from 8.3% (baseline) to 6.0%, proving that complementary features have been found.

The second important thing to notice is the fact that the selection pressure on the performance of the fusion involves the increase of the performance of the evolved filter bank alone and the decrease of the correlation between the two systems.

Figure 6 shows the filter bank obtained on the 11$^{th}$ generation.

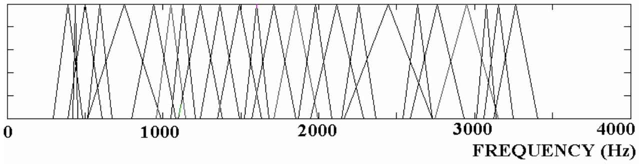

Fig.6. The Obtained filter bank

## 5. RESULTS

This section presents the results obtained by the evolved filter bank on the 2006 Nist databases with a state-of-the-art speaker verification system.

### 5.1. The 2006 Nist Database

The 2006 Nist corpus is composed of conversational telephone speech signals passed through different phone channels (landline, cordless or cellular) and sampled to 8 kHz. The 1conv4w_conv4w task involved 608 speakers and 51448 tests. Each segment of the train and the test is composed of 5 min of conversational speech.

### 5.2. Speaker verification system

The system we used for the test is the LIA SpkDet [9] baseline system. This system is a state of the art GMM-UBM based on the EM algorithm for the world training and the MAP algorithm for the clients adaptation. We used a 512 Gaussian models with diagonal covariance matrix. A Tnorm normalization was applied on the LLR.

### 5.3. Results

Figure 7 shows the EER obtained by the arithmetic fusion according to the fusion weight. This fusion is defined by:

$$F = \alpha \times EFB + (1 - \alpha) \times LFB$$

Where $F$ is the system resulting of the fusion, $\alpha$ is the fusion weight, *EFB* represents the system based on the *Evolved Filter Bank* and *LFB* the *Linear Filter Bank* based system.

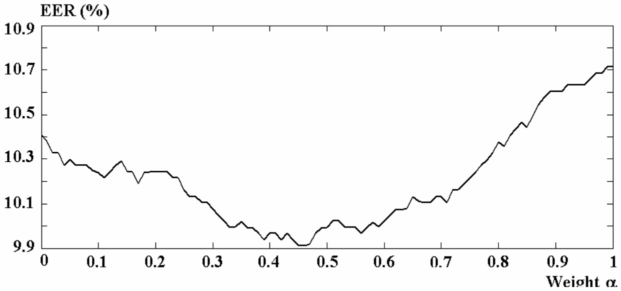

**Fig.7. Arithmetic fusion**

As we can see from this graph, the fusion of these two systems improves the EER from 10.4% to 9.9%. We can conclude that the evolved filter bank is able to complete LFCCs, providing new efficient information. Moreover, this filter bank presents similar performances with an EER of 10.7%.

## 6. CONCLUSION

In this paper, we proposed to use genetic algorithms to optimize the filter bank of a cepstrum based feature extractor in order to provide complementary information. The experiments we made showed that the obtained filter bank provides efficient information, not captured by the conventional LFCC feature extractor.

This first experiment shows that there is room for improvement for filter bank based feature extraction.

Our future work will consist of searching a new spectrum representation, based on a multi filter bank approach.

## 7. REFERENCES

[1] M. Zhiyou, Y. Yingchun, W Zhaohui, "Further feature extraction for speaker recognition," *IEEE International Conference on Systems, Man and Cybernetics*, *vol.5*, pp. 4153-4158, 2003

[2] C. Miyajima, H. Watanabe, K Tokuda, T. Kitamura, S. Katagiri, "A new approach to designing a feature extractor in speaker identification based on discriminative feature extraction," *Speech Communication 35*, No.3-4, pp. 203-218, 2001

[3] Holland, J. H., *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, MI, 1975

[4] Chin-Teng L. Hsi-Wen N. and Jiing-Yuan H, "GA-based noisy speech recognition using two-dimensional cepstrum," In *Proc. Conf. Intl. IEEE Transactions on Speech and Audio Processing, .vol. 8*, pp 664-675, 2000

[5] M. Demirekler, A. Haydar, "Feature selection using genetics-based algorithm and its application to speaker identification," In *Proc. Con.f Intl. IEEE International Conference on Acoustics, Speech, and Signal Processing*, *vol 1*, pp 329-332, 1999

[6] C. Charbuillet, B. Gas, M, Chetouani and J.L. Zarader, "Filter bank design for speaker diarization based on genetic algorithms," In *Proc. Conf. Intl. IEEE International Conference on Acoustics, Speech, and Signal Processing*, *vol. 1* pp 673-676, 2006

[7] 2006 NIST Speaker Recognition Evaluation site, www.nist.gov/speech/tests/spk/2006/

[8] F. Pasemann, U. Dieckmann, and U. Steinmetz, "Evolving Structure and Function of Neurocontrollers," In *Proc. Conf. Congress on Evolutionary Computation Journal*, IEEE Press US, Piscataway, pp. 1937-1978, 1999

[9] LIA SpkDet system web site, http://www.lia.univ-avignon.fr/heberges/ALIZE/LIA RAL